Statistica Sinica **34** (2024), 1521-1543
doi:https://doi.org/10.5705/ss.202022.0086

AN RKHS APPROACH FOR PIVOTAL INFERENCE IN FUNCTIONAL LINEAR REGRESSION

Holger Dette* and Jiajun Tang

Ruhr-Universität Bochum

Abstract: We use a reproducing kernel Hilbert space approach to develop a methodology for testing hypotheses about the slope function in a functional linear regression for time series. In contrast to most existing studies, which tests for the exact nullity of the slope function, we are interested in the null hypothesis that the slope function vanishes only approximately, where deviations are measured with respect to the L^2-norm. We propose an asymptotically pivotal test that does not require estimating nuisance parameters or long-run covariances. The key technical tools that we use to prove the validity of our approach include a uniform Bahadur representation and a weak invariance principle for a sequential process of estimates of the slope function. Lastly, we demonstrate the potential of our methods using a small simulation study and a data example.

Key words and phrases: Functional linear regression, functional time series, m-approximability, relevant hypotheses, reproducing kernel Hilbert space, self-normalization.

1. Introduction

Numerous statistical methods exist for analyzing functional data; see Ramsay and Silverman (2005), Ferraty and Vieu (2010), Horváth and Kokoszka (2012), Hsing and Eubank (2015), and Wang, Chiou and Müller (2016). Because of its good interpretability, the functional linear regression model

$$Y_i = \int_0^1 X_i(s)\, \beta_0(s)\, ds + \varepsilon_i, \quad i \in \mathbb{Z}, \tag{1.1}$$

has become a useful tool for functional data analysis (e.g., see Cardot, Ferraty and Sarda (1999); Müller and Stadtmüller (2005); Yao, Müller and Wang (2005); Hall and Horowitz (2007); Yuan and Cai (2010)). In our study, $\{(X_i, \varepsilon_i)\}_{i\in\mathbb{Z}}$ denotes a strictly stationary time series, where X_i is a mean zero square-integrable random function on the interval $[0, 1]$, and ε_i is a real-valued centered random noise.

Because the slope function β_0 characterizes the dependence between the predictor and the response, many studies have focused on its estimation and corresponding statistical inference. A popular method for analyzing the slope function in model (1.1) is to use functional principle components (FPCs) (e.g.,

*Corresponding author.

see Yao, Müller and Wang (2005); Hall and Horowitz (2007); Horváth and Kokoszka (2012); Hilgert, Mas and Verzelen (2013)). Others use a reproducing kernel Hilbert space (RKHS) approach to develop inference tools for β_0 and the corresponding theoretical results on consistency and optimality. Yuan and Cai (2010) and Cai and Yuan (2012) study an RKHS estimator and its prediction risk in a scalar-on-function linear regression model, and Shin and Lee (2016) use an RKHS approach for a robust functional linear regression. Shang and Cheng (2015) propose an RKHS inference framework for a generalized functional linear regression, and Hao et al. (2021) consider the functional Cox model. These authors also suggest tests for the nullity of the slope function. Recently, Dette and Tang (2021) used an RKHS approach to develop a statistical inference methodology in the function-on-function linear model, measuring deviations from the null hypothesis with respect to the sup-norm focusing on confidence bands. Statistical inference for functional time series has also been studied extensively (e.g., see Chen and Song (2015); Kokoszka, Rice and Shang (2017); van Delft and Eichler (2018); Dette, Kokot and Aue (2020); Dette, Kokot and Volgushev (2020); Cui and Zhou (2022)).

A common feature of the statistical theory for the functional linear regression model is that the proposed methodologies depend on the knowledge of nuisance parameters that appear in the asymptotic variance of the estimators of the slope function. As discussed in Section 3, these parameters are related to the long-run covariance structure of the data, and describe the behavior of a sequence of solutions of a system of estimated integro-differential equations induced by the covariance operator of the predictor. As a result, their estimation is not an easy problem. In the case of independent data (as considered in all works that use the RKHS approach), several estimators have been proposed and studied. On the other hand, for time series data, these nuisance parameters have an even more complicated structure, because of the dependencies in the data, making their estimation yet more difficult.

The purpose of this study is to develop pivotal statistical inference tools for the slope function β_0 in the functional linear regression model (1.1) by using an RKHS approach, which avoids needing to estimate nuisance parameters. Most existing works focus on testing hypotheses of the form

$$H_0 : d_0 := \int_0^1 |\beta_0(s)|^2 \, ds = 0 \quad \text{versus} \quad H_1 : d_0 \neq 0, \tag{1.2}$$

which is the classical hypothesis of the null effect ($\beta_0 \equiv 0$) of the functional covariate; see, for example, Cardot et al. (2003); García-Portugués, González-Manteiga and Febrero-Bande (2014); Lei (2014); Kong, Staicu and Maity (2016); Su, Di and Hsu (2017); Tekbudak et al. (2019) among man others. In contrast, we develop a pivotal test for the hypotheses

$$H_0 : d_0 = \int_0^1 |\beta_0(s)|^2 \, ds \le \Delta \quad \text{versus} \quad H_1 : d_0 > \Delta. \tag{1.3}$$

Here, $\Delta > 0$ is a (small) prespecified threshold that represents the maximal acceptable deviation (measured with respect to the L^2-distance) of β_0 from the null function. Note that in contrast to (1.2), the hypotheses in (1.3) are symmetric, in the sense that the null and the alternative can be interchanged. This allows us to investigate, at a controlled type-I error, whether the effect of the covariate on the response is negligible by testing the hypotheses

$$H_0 : d_0 > \Delta \quad \text{versus} \quad H_1 : d_0 \le \Delta. \tag{1.4}$$

Throughout this paper, we refer to hypotheses of the form (1.2) as"classical", and to those of the form (1.3) or (1.4) as "relevant" hypotheses. We discuss the pros and cons of these hypotheses in more detail in Section 2.

Our aim is to develop a pivotal methodology for testing relevant hypotheses (1.3) (or (1.4)), with no need to estimate nuisance parameters. Our approach is based on an RKHS and a novel self-normalization technique, recently introduced by Dette, Kokot and Volgushev (2020) in the context of testing relevant hypotheses about the mean and covariance functions of stationary time series. As such, our approach differs substantially from the common self-normalization approaches for testing classical hypotheses about finite-dimensional parameters (see Lobato (2001); Shao (2010); Shao and Zhang (2010), among many others). Because a statistical inference about the slope function is an inverse problem, it cannot be addressed directly using classical methods. In Section 3, we introduce a sequential RKHS estimator for the slope function in model (1.1). Section 4 is devoted to the development of our self-normalization methodology for the relevant hypotheses (1.3). As a by-product, we also construct (asymptotically) pivotal confidence intervals for the L^2-norm of the slope function. Here, the crucial result is a weak invariance principle for the process of estimators $\{\widehat{\beta}(\nu)\}_{\nu \in [\nu_0, 1]}$, where $\nu_0 \in (0, 1]$ is a constant and $\widehat{\beta}(\nu)$ denotes the estimator of β_0 calculated from the data $\{(X_i, Y_i)\}_{i=1,\dots,\lfloor n\nu \rfloor}$ (see Theorem 2 and the discussion in the subsequent paragraph). In Section 5, we provide details for the numerical implementation of our approach, and present simulated data experiments and a real-data example. The proofs of our theoretical results are given in the online Supplementary Material. The Matlab program for implementing our method is available at `https://github.com/jttang/SN_RKHS`.

To the best of our knowledge, testing relevant hypotheses about the slope function has only recently been considered, recently by Kutta, Dierickx and Dette (2022), who investigate a normal equation corresponding to the linear model (1.1), which they solve by applying a regularized inverse based on a spectral-cut-off series estimator. Although their approach has some theoretical advantages, its practical usefulness is limited, because it requires an estimation of the spectral decomposition of the regularized inverse. In contrast, the proposed estimator is defined as the minimizer of a regularized loss function in an appropriate RKHS. In Section 5, we demonstrate the advantages of our approach.

2. Classical and Relevant Hypotheses

Our particular interest in hypotheses of the form (1.3) and (1.4) stems from the fact that in many cases it is rare, and perhaps impossible, to have a null hypothesis that can be exactly modeled as $\beta_0 \equiv 0$ (see Berger and Delampady (1987) for a detailed discussion). More precisely, in most applications, such as in our data example in Section 5.3, the covariate X has some (possibly small) effect on the response Y. Thus, a more reasonable question is whether this effect is small and negligible. We address this point by testing the relevant hypotheses in (1.3), where we measure the size of the effect by the (squared) L^2-norm of the the function β_0, although other norms can be considered as well.

Although relevant hypotheses have only recently been considered in the context of functional data (see Fogarty and Small (2014); Dette, Kokot and Aue (2020); Dette, Kokot and Volgushev (2020), among others), they have a long history in (mathematical) statistics. Early references include the paper of Hodges and Lehmann (1954) and the textbook by Lehmann (1959). Testing relevant hypotheses (in particular those of the form (1.4)) for real-valued (or finite-dimensional) parameters has found considerable interest in the biostatistics community (see the mongraphs of Chow and Liu (1992); Wellek (2010)). Moreover, in the context of drug development, several authors have considered relevant hypotheses for comparing dose response files (see Liu, Hayter and Wynn (2007); Liu et al. (2009), among others), where they estimate parametric curves from real-valued data. On the other hand, hypotheses of the form (1.3) have found considerable interest in mathematical statistics; see Spokoiny (1996) and Lepski and Spokoiny (1999) for some early works and Blanchard and Fermanian (2021) and Brutsche and Rohde (2022) for some more recent references.

From a statistical point of view, whether to use classical or relevant hypotheses is often a subjective decision. Relevant hypotheses should be preferred if there is clear evidence that exact equality cannot hold (otherwise we are testing a hypothesis that we know in advance is not true). As pointed out by Berger and Delampady (1987), this situation appears rather frequently. On the other hand, using hypotheses (1.3) and (1.4) means we have to specify the threshold Δ, which is often not an easy task. This choice is case dependent, and requires a careful investigation of the scientific problem and a discussion with scientists from other fields to define what is considered relevant in the specific application. Note that for bioequivance testing, these discussions have already been completed (see Chow and Liu (1992); Wellek (2010)). Here, regulators such as the EMA or FDA have have defined thresholds for specific applications.

Moreover, even if the choice of the threshold is difficult, the proposed methodology still provides useful alternatives to testing classical hypotheses. On the one hand, it is possible to test for relevant differences for a finite number of thresholds simultaneously, and to determine for a fixed level α, the largest threshold such that the null hypothesis is rejected. On the other hand, we can construct confidence intervals for the measure $d_0 = \int_0^1 |\beta_0(s)|^2 ds$ (see Remark 2).

3. The RKHS Approach to Functional Linear Regression

We first introduce the notation used throughout this article. Let $L^2([0,1])$ denote the Hilbert space of square-integrable functions on $[0,1]$ equipped with the usual L^2-inner product $\langle \cdot, \cdot \rangle_{L^2}$ and the corresponding L^2 norm $\|\cdot\|_{L^2}$. Let $\ell^\infty([0,1])$ denote the set of all bounded real-valued functions on $[0,1]$ with corresponding norm $\|\cdot\|_\infty$, let "$\rightsquigarrow$" denote weak convergence in $\ell^\infty([0,1])$, and let "$\xrightarrow{d}$" denote the usual convergence in distribution in $\mathbb{R}^k$ (for some positive integer k). Write $a_n \asymp b_n$ if there exist constants $c_1, c_2 > 0$ such that $c_1 \leq a_n/b_n \leq c_2$, for all n. For $a \in \mathbb{R}$, let $\lfloor a \rfloor$ denote the largest integer smaller than or equal to a.

Suppose a sample of n observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ generated by the functional linear regression model (1.1) is available, and let $\nu_0 \in (0,1]$ be an arbitrary, but fixed constant. For any $\nu \in [\nu_0, 1]$, we first define an estimator of β_0 based on the first $\lfloor n\nu \rfloor$ observations $(X_1, Y_1), \ldots, (X_{\lfloor n\nu \rfloor}, Y_{\lfloor n\nu \rfloor})$. For this purpose, let

$$\mathcal{H} = \Big\{\beta : [0,1] \to \mathbb{R} \,\big|\, \partial^{(\theta)}\beta \text{ is absolutely continuous, for } 0 \leq \theta \leq m-1\,;\partial^{(m)}\beta \in L^2([0,1])\Big\} \tag{3.1}$$

denote the Sobolev space of order $m > 1/2$ of functions defined on $[0,1]$ (e.g., see Wahba (1990)), and define for $\nu \in [\nu_0, 1]$, the estimator

$$\widehat{\beta}_{n,\lambda}(\cdot, \nu) = \underset{\beta \in \mathcal{H}}{\operatorname{argmin}} \left[\frac{1}{2\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \Big\{ Y_i - \int_0^1 X_i(s)\,\beta(s)\,ds \Big\}^2 + \frac{\lambda}{2} J(\beta, \beta) \right], \tag{3.2}$$

for the function β_0. Here, $\lambda > 0$ is a regularization parameter, and for $\beta_1, \beta_2 \in \mathcal{H}$,

$$J(\beta_1, \beta_2) = \int_0^1 \beta_1^{(m)}(s)\,\beta_2^{(m)}(s)\,ds \tag{3.3}$$

defines the penalty functional. In (3.2), we use the notation $\widehat{\beta}_{n,\lambda}(\cdot,\nu)$ to reflect the dependence of the estimator on the parameters λ and ν. We emphasize that $\widehat{\beta}_{n,\lambda}(\cdot,\nu)$ is the estimator based on the first $\lfloor n\nu \rfloor$ observations $(X_1, Y_1), \ldots, (X_{\lfloor n\nu \rfloor}, Y_{\lfloor n\nu \rfloor})$, and that the parameter $\nu \in [\nu_0, 1]$ stands for the proportion of the sample $\{(X_i, Y_i)\}_{i=1}^n$ used to obtain $\widehat{\beta}_{n,\lambda}(\cdot, \nu)$. The case $\nu = 1$ corresponds to using the full sample $\{(X_i, Y_i)\}_{i=1}^n$ to estimate the slope function β_0, and we use the statistic

$$\widehat{\mathbb{T}}_n = \int_0^1 |\widehat{\beta}_{n,\lambda}(s, 1)|^2 \, ds \tag{3.4}$$

as an estimate for its squared L^2-norm. It can be shown that $\widehat{\mathbb{T}}_n$ defines a consistent estimator of $d_0 = \int_0^1 |\beta_0(s)|^2 ds$, such that the null hypothesis in (1.3) should be rejected if $\widehat{\mathbb{T}}_n$ is large. In fact, it follows from Theorem 3 that, under suitable conditions,

$$\sqrt{n}\lambda^{(2a+1)/(2D)}(\widehat{\mathbb{T}}_n - d_0) \xrightarrow{d} N(0, 4\sigma_d^2), \tag{3.5}$$

where

$$\sigma_d^2 = \lim_{\lambda \downarrow 0} \int_0^1 \int_0^1 C_{U,\lambda}(s, t) \, \beta_0(s) \, \beta_0(t) \, ds \, dt, \tag{3.6}$$

$$C_{U,\lambda}(s, t) = \lambda^{(2a+1)/D} \sum_{\ell=-\infty}^{+\infty} \operatorname{cov}\{\varepsilon_0 \, \tau_\lambda(X_0)(s), \varepsilon_\ell \, \tau_\lambda(X_\ell)(t)\}, \tag{3.7}$$

τ_λ is an operator defined by

$$\tau_\lambda(z) = \sum_{k=1}^{\infty} \frac{\langle z, \varphi_k \rangle_{L^2}}{1 + \lambda \rho_k} \, \varphi_k \ , \tag{3.8}$$

$\{(\rho_k, \varphi_k)\}_{k \geq 1}$ is the eigensystem of certain integro-differential equations defined by the covariance operator of the predictor X, and the constants a and D in (3.5) depend on the maximum norm of the eigenfunctions φ_k and on the eigenvalues ρ_k (see Assumption 2 and the subsequent discussion).

As a result, in practice, the normalizing factor $\sqrt{n}\lambda^{(2a+1)/(2D)}$, the long-run covariance $C_{U,\lambda}$ in (3.7), and the asymptotic variance σ_d^2 in (3.6) are often either intractable or difficult to estimate. This is because σ_d^2 is defined as the limit of a series, which in turn relies on the operator τ_λ in (3.8), and therefore depends on the eigensystem $\{(\rho_k, \varphi_k)\}_{k \geq 1}$ of the integro-differential equations. Moreover, the normalizing factor in (3.5) and the operator $C_{U,\lambda}$ defined in (3.7) depend on the unknown nuisance parameters a and D, making estimation even more challenging. These difficulties motivate us to propose a self-normalization approach so that pivotal tests can be constructed for the relevant hypotheses (1.3), even without knowledge of σ_d^2 in (3.6) and the nuisance parameters a and D. The approach is described in detail in Section 4, but a heuristic argument ignoring the technical details follows. For a fixed value $\nu_0 \in (0, 1]$ and a given probability measure ω on the interval $[\nu_0, 1]$, we define the statistic

$$\widehat{\mathbb{V}}_n = \left[\int_{\nu_0}^1 \left| \nu^2 \int_0^1 \{ \widehat{\beta}_{n,\lambda}^2(s,\nu) - \widehat{\beta}_{n,\lambda}^2(s,1) \} \, ds \right|^2 \omega(d\nu) \right]^{1/2}, \tag{3.9}$$

where $\widehat{\beta}_{n,\lambda}$ is the estimator of the slope function β_0 from the sample $(X_1, Y_1), \ldots, (X_{\lfloor n\nu \rfloor}, Y_{\lfloor n\nu \rfloor})$ defined in (3.2). From Theorem 3, we have

$$\begin{aligned} &\sqrt{n}\lambda^{(2a+1)/(2D)} \big(\widehat{\mathbb{T}}_n - d_0, \widehat{\mathbb{V}}_n \big) \\ &\xrightarrow{d} \left(2\sigma_d \, \mathbb{B}(1), 2\sigma_d \left\{ \int_{\nu_0}^1 |\nu \, \mathbb{B}(\nu) - \nu^2 \mathbb{B}(1)|^2 \, \omega(d\nu) \right\}^{1/2} \right), \end{aligned}$$

where $\mathbb{B}$ denotes the standard Brownian motion. In particular, the ratio $(\widehat{\mathbb{T}}_n - d_0)/\widehat{\mathbb{V}}_n$ is asymptotically free. Therefore, we can obtain a consistent and asymptotic level α-test for the hypotheses (1.3) by comparing the statistic $(\widehat{\mathbb{T}}_n - \Delta)/\widehat{\mathbb{V}}_n$ with the $(1-\alpha)$-quantile of the limiting distribution. The details are provided in the following section.

4. Self-Normalization and Pivotal Inference

We first establish a uniform Bahadur representation of the sequential process of estimators of the slope function $\{\widehat{\beta}_{n,\lambda}(\cdot, \nu)\}_{\nu \in [\nu_0, 1]}$, which is crucial for our approach. For this purpose, we define

$$L_{n,\lambda,\nu}(\beta) = \frac{1}{2\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \left\{ Y_i - \int_0^1 X_i(s) \, \beta(s) \, ds \right\}^2 + \frac{\lambda}{2} J(\beta, \beta)$$

as the objective functional in (3.2), and note that its Fréchet derivatives are given by

$$\begin{aligned} \mathcal{D}L_{n,\lambda,\nu}(\beta)\beta_1 &= -\frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \left\{ Y_i - \int_0^1 X_i(s_1)\beta(s_1) ds_1 \right\} \\ &\qquad \int_0^1 X_i(s_2)\beta_1(s_2) ds_2 + \lambda J(\beta, \beta_1) \, ; \\ \mathcal{D}^2 L_{n,\lambda,\nu}(\beta)\beta_1\beta_2 &= \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \int_0^1 X_i(s_1)\beta_1(s) ds_1 \int_0^1 X_i(s_2)\beta_2(s_2) ds_2 + \lambda J(\beta_1, \beta_2), \end{aligned} \tag{4.1}$$

and $\mathcal{D}^3 L_{n,\lambda,\nu}(\beta) \equiv 0$. If $C_X(s,t) = \mathrm{cov}\{X_1(s), X_1(t)\}$ denotes the covariance kernel of the predictor, then a simple calculation shows that $\mathrm{E}\{\mathcal{D}^2 L_{n,\lambda,\nu}(\beta)\beta_1\beta_2\} = \langle \beta_1, \beta_2 \rangle_K$, where the mapping $\langle \cdot, \cdot \rangle_K : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is defined by

$$\langle \beta_1, \beta_2 \rangle_K = V(\beta_1, \beta_2) + \lambda J(\beta_1, \beta_2), \qquad \beta_1, \beta_2 \in \mathcal{H}, \tag{4.2}$$

J is the functional in (3.3), and

$$V(\beta_1, \beta_2) = \int_0^1 \int_0^1 C_X(s,t)\, \beta_1(s)\, \beta_2(t)\, ds\, dt. \tag{4.3}$$

For our theoretical analysis, we first make the following mild assumption on the kernel C_X.

Assumption 1. The covariance kernel C_X is continuous on $[0,1]^2$. For any $\gamma \in L^2([0,1])$, $\int_0^1 C_X(s,t)\gamma(s)ds = 0$ for any $t \in [0,1]$ implies that $\gamma \equiv 0$.

Assumption 1 is a common condition in the literature (e.g., see Yuan and Cai (2010); Shang and Cheng (2015)), and implies that the mapping $\langle \cdot, \cdot \rangle_K$ in (4.2) defines an inner product on $\mathcal{H}$ with corresponding norm $\|\cdot\|_K$. In addition, $\mathcal{H}$ is an RKHS equipped with the inner product $\langle \cdot, \cdot \rangle_K$. We follow Shang and Cheng (2015) and assume that there exists a sequence of functions in $\mathcal{H}$ that diagonalize the operators V in (4.3) and J in (3.3) simultaneously.

Assumption 2 (Simultaneous diagonalization). There exists a sequence of functions $\{\varphi_k\}_{k\geq 1}$ in $\mathcal{H}$, such that $\|\varphi_k\|_\infty \leq c\, k^a$, $V(\varphi_k, \varphi_{k'}) = \delta_{kk'}$, and $J(\varphi_k, \varphi_{k'}) = \rho_k\, \delta_{kk'}$, for any $k, k' \geq 1$, where $a \geq 0$, $c > 0$ are constants, $\delta_{kk'}$ is the Kronecker delta, and the sequence $\{\rho_k\}_{k\geq 1}$ satisfies $\rho_k \asymp k^{2D}$, for some constant $D > a+1/2$. Furthermore, any $\beta \in \mathcal{H}$ admits the expansion $\beta = \sum_{k=1}^{\infty} V(\beta, \varphi_k)\varphi_k$ with convergence in $\mathcal{H}$ w.r.t. the norm $\|\cdot\|_K$.

In their Proposition 2.2, Shang and Cheng (2015) prove that Assumption 2 is satisfied for the eigensystem $\{(\rho_k, \varphi_k)\}_{k\geq 1}$ of the following integro-differential equations with boundary conditions:

$$\begin{cases} \rho \int_0^1 C_X(s,t)\, x(t)\, dt = (-1)^m\, x^{(2m)}(s), \\ x^{(\theta)}(0) = x^{(\theta)}(1) = 0, \qquad \text{for } m \leq \theta \leq 2m-1. \end{cases} \tag{4.4}$$

For the inner product $\langle \cdot, \cdot \rangle_K$ in (4.2), it follows from Assumption 2 that $\langle \varphi_k, \varphi_{k'} \rangle_K = V(\varphi_k, \varphi_{k'}) + \lambda J(\varphi_k, \varphi_{k'}) = (1 + \lambda\rho_k)\, \delta_{kk'}$, for $k, k' \geq 1$, such that $\langle \beta, \varphi_k \rangle_K = (1+\lambda\rho_k) V(\beta, \varphi_k)$, for any $\beta \in \mathcal{H}$, which implies the representation

$$\beta = \sum_{k=1}^{\infty} \frac{\langle \beta, \varphi_k \rangle_K}{1 + \lambda\rho_k}\, \varphi_k. \tag{4.5}$$

Recalling the definition of the penalty J in (3.3), we denote by $W_\lambda : \mathcal{H} \to \mathcal{H}$ the operator such that $\langle W_\lambda(\beta_1), \beta_2 \rangle_K = \lambda J(\beta_1, \beta_2)$, for $\beta_1, \beta_2 \in \mathcal{H}$. By definition, we have, for the eigenfunctions $\{\varphi_k\}_{k\geq 1}$ in Assumption 2, $\langle W_\lambda(\varphi_k), \varphi_{k'} \rangle_K = \lambda J(\varphi_k, \varphi_{k'}) = \lambda\rho_k\, \delta_{kk'}$, for any $k, k' \geq 1$. Thus, from (4.5), we have

$$W_\lambda(\varphi_k) = \sum_{k'=1}^{\infty} \frac{\langle W_\lambda(\varphi_k), \varphi_{k'} \rangle_K}{1+\lambda\rho_{k'}} \varphi_{k'} = \frac{\lambda\,\rho_k\,\varphi_k}{1+\lambda\rho_k}. \tag{4.6}$$

In addition, note that $\mathfrak{G}_z(\beta) = \int_0^1 \beta(s)z(s)ds$ is a bounded linear functional on $\mathcal{H}$, for any $z \in L^2([0,1])$ and $\beta \in \mathcal{H}$. By the Riesz representation theorem, there exists a unique element $\tau_\lambda(z) \in \mathcal{H}$ such that $\langle \tau_\lambda(z), \beta \rangle_K = \mathfrak{G}_z(\beta)$. In particular, $\langle \tau_\lambda(z), \varphi_k \rangle_K = \langle z, \varphi_k \rangle_{L^2}$, such that we obtain the representation (3.8) for the operator τ_λ. Now, for any $\beta, \beta_1, \beta_2 \in \mathcal{H}$, define

$$\begin{aligned} S_{n,\lambda,\nu}(\beta) &= -\frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \tau_\lambda(X_i) \left\{ Y_i - \int_0^1 X_i(s)\beta(s)ds \right\} + W_\lambda(\beta) \\ &= -\frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \varepsilon_i \tau_\lambda(X_i) + W_\lambda(\beta_0), \\ \mathcal{D}S_{n,\lambda,\nu}(\beta)\beta_1 &= \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \tau_\lambda(X_i) \int_0^1 X_i(s)\beta_1(s)ds + W_\lambda(\beta_1), \end{aligned} \tag{4.7}$$

such that $\mathcal{D}L_{n,\lambda,\nu}(\beta)\beta_1 = \langle S_{n,\lambda,\nu}(\beta), \beta_1 \rangle_K$ and $\mathcal{D}^2 L_{n,\lambda,\nu}(\beta)\beta_1\beta_2 = \langle \mathcal{D}S_{n,\lambda,\nu}(\beta)\beta_1, \beta_2 \rangle_K$. Here, the term $S_{n,\lambda,\nu}$ is the dominating term in the expansion of $\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0$, that is,

$$\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 \approx -S_{n,\lambda,\nu}(\beta_0) = \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \varepsilon_i\,\tau_\lambda(X_i) - W_\lambda(\beta_0). \tag{4.8}$$

A rigorous statement of this approximation is given in Theorem 1, and requires several assumptions, which are stated next. We begin by characterizing the dependence structures of the functional time series, where we use the concept of m-approximability (e.g., Hörmann and Kokoszka (2010); Berkes, Horváth and Rice (2013)).

Assumption 3. For $i \in \mathbb{Z}$, (X_i, Y_i) is generated from model (1.1) and satisfies the following:

3.1. $X_i = g(\ldots, \xi_{i-1}, \xi_i)$ and $\varepsilon_i = h(\ldots, \eta_{i-1}, \eta_i)$, for $i \in \mathbb{Z}$ and some deterministic measurable functions $g : \mathcal{S}^\infty \to L^2([0,1])$ and $h : \mathbb{R}^\infty \to \mathbb{R}$, where $\mathcal{S}$ is some measurable space and $\xi_i = \xi_i(t,\omega)$ is jointly measurable in (t,ω); ξ_i and η_i are independent and identically distributed (i.i.d.).

3.2. For any $s \in [0,1]$, $\mathrm{E}\{X_0(s)\} = \mathrm{E}(\varepsilon_0) = 0$. For some $\delta \in (0,1)$, $\mathrm{E}|\varepsilon_0|^{2+\delta} < \infty$.

3.3. The sequences $\{X_i\}_{i\in\mathbb{Z}}$ and $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ can be approximated by ℓ-dependent sequences $\{X_{i,\ell}\}_{i,\ell\in\mathbb{Z}}$ and $\{\varepsilon_{i,\ell}\}_{i,\ell\in\mathbb{Z}}$, respectively, in the sense that, for some $\kappa > 2+\delta$,

$$\sum_{\ell=1}^{\infty} \left(\mathrm{E}\|X_i - X_{i,\ell}\|_{L^2}^{2+\delta}\right)^{1/\kappa} < \infty, \qquad \sum_{\ell=1}^{\infty} \left(\mathrm{E}|\varepsilon_i - \varepsilon_{i,\ell}|^{2+\delta}\right)^{1/\kappa} < \infty.$$

Here, $X_{i,\ell} = g(\xi_i, \xi_{i-1}, \ldots, \xi_{i-\ell+1}, \boldsymbol{\xi}_{i,\ell}^*)$ and $\varepsilon_{i,\ell} = h(\eta_i, \eta_{i-1}, \ldots, \eta_{i-\ell+1}, \boldsymbol{\eta}_{i,\ell}^*)$, where $\boldsymbol{\xi}_{i,\ell}^* = (\xi_{i,\ell,i-\ell}^*, \xi_{i,\ell,i-\ell-1}^*, \ldots)$ and $\boldsymbol{\eta}_{i,\ell}^* = (\eta_{i,\ell,i-\ell}^*, \eta_{i,\ell,i-\ell-1}^*, \ldots)$, and where $\xi_{i,\ell,k}^*$ and $\eta_{i,\ell,k}^*$ are independent copies of ξ_0 and η_0, respectively, and are independent of $\{\xi_i\}_{i\in\mathbb{Z}}$ and $\{\eta_i\}_{i\in\mathbb{Z}}$, respectively.

Assumption 4 (Regularity conditions).

4.1. There exists a constant $\varpi > 0$ such that $\mathrm{E}\{\exp(\varpi\|X_0\|_{L^2}^2)\} < \infty$.

4.2. For any $\beta \in \mathcal{H}$, $\mathrm{E}\big(\langle X_0, \beta\rangle_{L^2}^4\big) \le c_0\big\{\mathrm{E}\big(\langle X_0, \beta\rangle_{L^2}^2\big)\big\}^2$, for some constant $c_0 > 0$.

4.3. The true slope function β_0 is such that $\sum_{k=1}^{\infty} \rho_k^2 V^2(\beta_0, \varphi_k) < \infty$.

4.4. For $s, t \in [0,1]$ and $C_{U,\lambda}$ in (3.7), the limit $C_U(s,t) = \lim_{\lambda\downarrow 0} C_{U,\lambda}(s,t)$ exists.

Assumption 5. The constants a and D in Assumption 2 and the regularization parameter λ in (3.2) satisfy $\lambda = o(1)$, $n^{-1}\lambda^{-(2a+1)/D} = o(1)$, and $n\lambda^{2+(2a+1)/(2D)} = o(1)$ as $n \to \infty$. In addition, $n^{-1}\lambda^{-2\varsigma}\log n = o(1)$ and $\lambda^{-2\varsigma+(2D+2a+1)/(2D)} \log n = o(1)$ as $n \to \infty$, where $\varsigma = (2D-2a-1)/(4Dm) + (a+1)/(2D) > 0$.

Remark 1. Assumption 4 requires an exponential tail of $\|X_0\|_{L^2}$. This condition is satisfied for any stochastic process with an almost surely bounded L^2-norm, and can also be satisfied for Gaussian processes with a square-integrable mean function if we take $\varpi \in (0, 1/4)$; this is proved in Proposition 3.2 in Shang and Cheng (2015). Assumption 4 is a common condition in linear regression models for functional data; see, for example, Cai and Yuan (2012) and Shang and Cheng (2015). Assumption 4 corresponds to the so-called undersmoothing scenario in Shang and Cheng (2015); see their Remark 3.2. Finally, Assumption 5 specifies the conditions for the regularization parameter λ in (3.2).

Our first main result justifies the approximation (4.8), and is proved in Section S1.1 of the Supplementary Material.

Theorem 1 (Uniform Bahadur representation). Suppose Assumptions 1–5 are satisfied. Then, for any fixed (but arbitrary) $\nu_0 \in (0,1]$,

$$\sup_{\nu\in[\nu_0,1]} \left\|\nu\big\{\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + W_\lambda(\beta_0)\big\} - \frac{1}{n}\sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i\, \tau_\lambda(X_i)\right\|_K = O_p(v_n), \tag{4.9}$$

where for the constant $\varsigma > 0$ in Assumption 5, $v_n = n^{-1/2}\lambda^{-\varsigma}(\lambda^{1/2} + n^{-1/2}\lambda^{-(2a+1)/(4D)})(\log n)^{1/2}$.

Next, we define for $i \in \mathbb{Z}$ and $\tau_\lambda(\cdot)$ in (3.8), the random variables

$$U_i = \lambda^{(2a+1)/(2D)}\, \varepsilon_i\, \tau_\lambda(X_i) = \lambda^{(2a+1)/(2D)} \varepsilon_i \sum_{k=1}^{\infty} \frac{\langle X_i, \varphi_k \rangle_{L^2}}{1+\lambda\rho_k} \varphi_k. \tag{4.10}$$

Theorem 1 shows that, under suitable conditions, the approximation

$$\nu\big\{\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + W_\lambda(\beta_0)\big\} \approx n^{-1}\lambda^{-(2a+1)/(2D)} \sum_{i=1}^{\lfloor n\nu \rfloor} U_i$$

holds uniformly in $\nu \in [\nu_0, 1]$ with respect to the $\|\cdot\|_K$-norm, where $\nu_0 \in (0,1]$ is an arbitrary, but fixed value. We now verify the weak invariance principle of the process $\{n^{-1/2}\sum_{i=1}^{\lfloor n\nu \rfloor} U_i\}_{n\in\mathbb{N}}$, and define for this purpose the class

$$\mathcal{F} = \left\{ g : [0,1]\times[0,1] \to \mathbb{R} \,\middle|\, \sup_{\nu\in[0,1]} \int_0^1 |g(s,\nu)|^2\, ds < \infty \right\}. \tag{4.11}$$

The following theorem is proved in Section S1.2 of the online Supplementary Material.

Theorem 2 (Weak invariance principle). Suppose Assumptions 1–5 hold. Then, there exists a mean-zero Gaussian process $\{\Gamma(s,\nu)\}_{s,\nu\in[0,1]}$ in $\mathcal{F}$ defined in (4.11), with covariance function $\mathrm{cov}\big\{\Gamma(s_1,\nu_1), \Gamma(s_2,\nu_2)\big\} = \min\{\nu_1,\nu_2\} C_U(s_1,s_2)$, for C_U in Assumption 4.4, such that

$$\sup_{\nu\in[0,1]} \int_0^1 \Big\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\nu \rfloor} U_i(s) - \Gamma(s,\nu) \Big\}^2 ds = o_p(1), \qquad \text{as } n\to\infty.$$

Theorem 2 shows that the partial sum $n^{-1/2}\sum_{i=1}^{\lfloor n\nu \rfloor} U_i$ can be approximated by a Gaussian process Γ in the L^2-sense, uniformly in $\nu \in [0,1]$. Thus, we obtain from Theorem 1, we obtain the approximation

$$\sup_{\nu\in[\nu_0,1]} \int_0^1 \Big[\sqrt{n}\lambda^{(2a+1)/(2D)} \nu\big\{\widehat{\beta}_{n,\lambda}(s,\nu) - \beta_0(s) + W_\lambda(\beta_0)\big\} - \Gamma(s,\nu) \Big]^2 ds = o_p(1).$$

Next, in order to propose our self-normalization methodology, we define a useful quantity related to the difference between the L^2-norms of the estimator $\widehat{\beta}_{n,\lambda}(\cdot,\nu)$ defined in (3.2) and the true slope function β_0, that is,

$$\widehat{\mathbb{G}}_n(\nu) = \sqrt{n}\lambda^{(2a+1)/(2D)}\, \nu^2 \int_0^1 \big\{ \widehat{\beta}^{\,2}_{n,\lambda}(s,\nu) - \beta_0^2(s) \big\}\, ds\,, \tag{4.12}$$

where $\nu \in [\nu_0, 1]$. The following theorem establishes the weak convergence of the process $\{\widehat{\mathbb{G}}_n(\nu)\}_{\nu\in[\nu_0,1]}$, and is proved in Section S1.3 of the online Supplementary Material.

Theorem 3. If Assumptions 1–5 hold, then the process $\widehat{\mathbb{G}}_n$ defined in (4.12) satisfies

$$\big\{\widehat{\mathbb{G}}_n(\nu)\big\}_{\nu\in[\nu_0,1]} \rightsquigarrow \big\{2\sigma_d\,\nu\mathbb{B}(\nu)\big\}_{\nu\in[\nu_0,1]} \qquad \text{in } \ell^\infty([\nu_0,1]),$$

where $\mathbb{B}$ denotes the standard Brownian motion and σ_d is defined in (3.6).

Recalling the definition of the statistics $\widehat{\mathbb{T}}_n$ and $\widehat{\mathbb{V}}_n$ in (3.4) and (3.9), respectively, we obtain from the continuous mapping theorem and Theorem 3 that

$$\sqrt{n}\lambda^{(2a+1)/(2D)}\big(\widehat{\mathbb{T}}_n - d_0, \widehat{\mathbb{V}}_n\big) = \left(\widehat{\mathbb{G}}_n(1), \left\{\int_{\nu_0}^1 \big|\widehat{\mathbb{G}}_n(\nu) - \nu^2\widehat{\mathbb{G}}_n(1)\big|^2\,\omega(d\nu)\right\}^{1/2}\right)$$
$$\xrightarrow{d} \left(2\sigma_d\,\mathbb{B}(1), 2\sigma_d\left\{\int_{\nu_0}^1 |\nu\,\mathbb{B}(\nu) - \nu^2\mathbb{B}(1)|^2\,\omega(d\nu)\right\}^{1/2}\right). \tag{4.13}$$

In particular, the ratio $(\widehat{\mathbb{T}}_n - d_0)/\widehat{\mathbb{V}}_n$ is asymptotically free, as stated in following theorem, which is proved in Section S1.4 of the Supplementary Material.

Theorem 4. Suppose Assumptions 1–5 are satisfied and assume that $\sigma_d^2 > 0$. For the $\widehat{\mathbb{T}}_n$, d_0, and $\widehat{\mathbb{V}}_n$ defined in (3.4), (1.3), and (3.9), respectively, we have

$$\frac{\widehat{\mathbb{T}}_n - d_0}{\widehat{\mathbb{V}}_n} \xrightarrow{d} \mathbb{W} = \frac{\mathbb{B}(1)}{\{\int_{\nu_0}^1 |\nu\,\mathbb{B}(\nu) - \nu^2\,\mathbb{B}(1)|^2\,\omega(d\nu)\}^{1/2}}. \tag{4.14}$$

Theorem 4 reveals a self-normalized statistic $(\widehat{\mathbb{T}}_n - d_0)/\widehat{\mathbb{V}}_n$ that converges weakly to a pivotal random variable $\mathbb{W}$, because its distribution does not depend on the nuisance parameters (i.e., a and D in Assumption 2, and σ_d^2 in (3.6)) or the eigensystem $\{(\rho_k, \varphi_k)\}_{k\geq 1}$. Moreover, the distribution of $\mathbb{W}$ in (4.14) can be simulated easily from computer-generated sample paths of standard Brownian motions. Therefore, we propose rejecting the null hypothesis in (1.3) at the nominal level α if

$$\widehat{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta, \tag{4.15}$$

where $\mathcal{Q}_{1-\alpha}(\mathbb{W})$ denotes the $(1-\alpha)$-quantile of the distribution of $\mathbb{W}$ in (4.14). Our final result, proved in Section S1.5 of the Supplementary Material, provides a theoretical justification for the consistency of the test defined in (4.15) at the nominal level α.

Theorem 5. Assume $\Delta > 0$. Under Assumptions 1–5, we have

$$\lim_{n\to\infty} \mathrm{P}\big\{\widehat{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta\big\} = \begin{cases} 0 & \text{if } d_0 < \Delta \\ \alpha & \text{if } d_0 = \Delta \text{ and } \sigma_d^2 > 0\,. \\ 1 & \text{if } d_0 > \Delta \end{cases}$$

Remark 2 (Further statistical consequences).

(1) The choice of the threshold Δ in the relevant hypotheses in (1.3) should be discussed with experts from the field of application. We noted in Section 2 that this is not an easy problem. However, we argue that instead of testing a null hypothesis, which we believe not to be true, one should think carefully about the effect, which is of real scientific interest.

If this is not possible, we recommend constructing a confidence interval for the (squared) L^2-norm of the slope function. Specifically, for the statistics $\widehat{\mathbb{T}}_n$ and $\widehat{\mathbb{V}}_n$ defined in (3.4) and (3.9), respectively, the set

$$\widehat{\mathcal{I}}_n := \big[0, \widehat{\mathbb{T}}_n + \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n\big] \tag{4.16}$$

defines an asymptotic $(1-\alpha)$-confidence interval for the squared L^2-norm $d_0 = \int_0^1 |\beta_0(s)|^2 ds$ of the unknown slope function. To see this, note that it follows in the case $d_0 > 0$ from Theorem 4 that

$$\mathrm{P}_{d_0>0}\big(d_0 \in \widehat{\mathcal{I}}_n\big) = \mathrm{P}_{d_0>0}\bigg\{\frac{\widehat{\mathbb{T}}_n - d_0}{\widehat{\mathbb{V}}_n} \geq -\mathcal{Q}_{1-\alpha}(\mathbb{W})\bigg\} \to 1-\alpha \tag{4.17}$$

as $n \to \infty$, where we use the fact that the distribution of the random variable $\mathbb{W}$ in (4.14) is symmetric, that is $-\mathcal{Q}_{1-\alpha}(\mathbb{W}) = \mathcal{Q}_{\alpha}(\mathbb{W})$. In the case $d_0 = 0$, because $\widehat{\mathbb{T}}_n, \widehat{\mathbb{V}}_n \geq 0$ almost surely, it follows that $\mathrm{P}_{d_0=0}\big(d_0 \in \widehat{\mathcal{I}}_n\big) = \mathrm{P}_{d_0=0}\big\{\widehat{\mathbb{T}}_n + \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n \geq 0\big\} = 1$. Moreover, if it is reasonable to assume that the quantity $d_0 = \int_0^1 |\beta_0(s)|^2 ds$ is positive, an asymptotic two-sided confidence interval for $d_0 > 0$ is given by

$$\Big(\max\big\{0, \widehat{\mathbb{T}}_n - \mathcal{Q}_{1-\alpha/2}(\mathbb{W})\widehat{\mathbb{V}}_n\big\}, \widehat{\mathbb{T}}_n + \mathcal{Q}_{1-\alpha/2}(\mathbb{W})\widehat{\mathbb{V}}_n\Big], \tag{4.18}$$

which follows by Theorem 4, observing that, by (4.13), $\widehat{\mathbb{T}}_n = d_0 + o_p(1)$ and $\widehat{\mathbb{V}}_n = o_p(1)$ as $n \to \infty$, and $\widehat{\mathbb{V}}_n \geq 0$ almost surely.

Alternatively, it is also possible to test the relevant hypotheses for a finite number of thresholds $\Delta^{(1)} < \cdots < \Delta^{(L)}$ simultaneously, for some $L \in \mathbb{N}_+$. In particular, a rejection of a $\Delta^{(L_0)}$ means rejecting for all smaller thresholds. In this sense, evaluating the test for several thresholds is logically consistent for the user, and it is possible to determine, for a fixed nominal level α, the largest threshold such that the null hypothesis is rejected.

(2) Theorem 4 also allows us to construct a consistent and asymptotic level-α test for the relevant hypotheses (1.4), defined by, rejecting H_0 if

$$\widehat{\mathbb{T}}_n < \mathcal{Q}_{\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta,$$

where $\widehat{\mathbb{T}}_n$ and $\widehat{\mathbb{V}}_n$ are given in (3.4) and (3.9), respectively, and $\mathcal{Q}_{\alpha}(\mathbb{W})$

denotes the α-quantile of the pivotal distribution of $\mathbb{W}$ in (4.14). The proof is omitted, for brevity.

Remark 3. For the classical hypotheses in (1.2), that is, $H_0 : d_0 = \int_0^1 |\beta_0(s)|^2 ds = 0$, a likelihood ratio-type test was proposed by Shang and Cheng (2015). Because the statistic $\widehat{\mathbb{T}}_n$ in (3.4) defines an estimator of the squared L^2-norm of the function β_0, an alternative test could be obtained by rejecting H_0 in (1.2) for large values of the statistic $\widehat{\mathbb{T}}_n$. However, it follows from the proof of Theorem 4 that for $d_0 = 0$,

$$\frac{\widehat{\mathbb{T}}_n}{\widehat{\mathbb{V}}_n} \xrightarrow{d} \frac{\int_0^1 \Gamma^2(s,1)ds}{\left[\int_{\nu_0}^1 \left|\int_0^1 \Gamma^2(s,\nu)ds - \nu^2 \int_0^1 \Gamma^2(s,1)ds\right|^2 \omega(d\nu)\right]^{1/2}},$$

where Γ is a mean-zero Gaussian process with covariance function $\mathrm{cov}\{\Gamma(s_1,\nu_1), \Gamma(s_2,\nu_2)\} = \min\{\nu_1,\nu_2\} C_U(s_1,s_2)$ and C_U is defined in Assumption 4.4. This limit distribution differs from that of $\mathbb{W}$ in (4.14), and is not pivotal, because it depends on the long-run covariance C_U. As a result, the decision rule defined in (4.15) does not define an asymptotic level-α test for the classical null hypotheses in (1.2).

5. Finite-Sample Properties

5.1. Implementation

In this section, we discuss the details of the implementation of the proposed tests for the relevant hypotheses. To begin with, in practice, we choose the probability measure ω in (3.9) and (4.14) as the discrete uniform distribution on the interval $[\nu_0, 1]$. Specifically, for some positive integer Q, let

$$\nu_q = \nu_0 + \frac{q(1-\nu_0)}{Q}, \quad \text{for } 1 \le q \le Q. \tag{5.1}$$

Then, we define ω as the discrete uniform distribution supported on the set $\{\nu_q\}_{q=1}^Q$, with equal probability mass $1/Q$, such that the pivotal random variable $\mathbb{W}$ in (4.14) is given by

$$\mathbb{W}_Q = \frac{\mathbb{B}(1)}{\left\{Q^{-1}(1-\nu_0)\sum_{q=1}^Q \left|\nu_q \mathbb{B}(\nu_q) - \nu_q^2 \mathbb{B}(1)\right|^2\right\}^{1/2}}, \tag{5.2}$$

and the quantiles of the pivotal distribution of $\mathbb{W}_Q$ can be obtained easily from simulated sample paths of standard Brownian motions. Recall from Section 4 that in order to obtain the statistics $\widehat{\mathbb{T}}_n$ and $\widehat{\mathbb{V}}_n$, we need to compute the RKHS estimator $\widehat{\beta}_{n,\lambda}(\cdot,\nu_q)$ defined in (3.2) using the observations $(X_1,Y_1),\ldots,(X_{n_q},Y_{n_q})$, where $n_q = \lfloor \nu_q n \rfloor$ $(q = 1,\ldots,Q)$. Because $\widehat{\beta}_{n,\lambda}(\cdot,\nu_q)$ is defined as the solution of a penalized minimization problem on an infinitedimensional function space $\mathcal{H}$ defined in (3.1), exact solutions are inaccessible. We circumvent this difficulty by introducing the following finite-sample method, and propose a method for choosing the regularization parameter λ in (3.2). We first observe from Assumption 2 that $J(\varphi_{k\ell}, \varphi_{k'\ell'}) = \rho_{k\ell}\delta_{kk'}\delta_{\ell\ell'}$, such that for $\beta = \sum_{k=1}^{\infty} b_k\varphi_k \in \mathcal{H}$ and $b_k \in \mathbb{R}$, we have $J(\beta, \beta) = \sum_{k=1}^{\infty} b_k^2\rho_k$. Consider the Sobolev space on $[0,1]$ of order $m = 2$. In this case, the penalty functional in (3.2) is $J(\beta, \beta) = \int_0^1 \{\beta''(s)\}^2 ds$. In order to find the empirical eigenfunctions φ_k and eigenvalues ρ_k, we solve the integro-differential equation (4.4)

$$\rho \int_0^1 \widehat{C}_X(s,t)\, x(t)\, dt = x^{(4)}(s), \quad \text{with } x^{(3)}(0) = x^{(3)}(1) = x^{(4)}(0) = x^{(4)}(1) = 0, \tag{5.3}$$

where $\widehat{C}_X$ denotes the empirical covariance function of X computed from the full sample $X_1, \ldots, X_n$. Let $\{\widehat{\varphi}_k\}_{k\geq 1}$ denote the eigenfunctions of (5.3), with the corresponding eigenvalues $\{\widehat{\rho}_k\}_{k\geq 1}$, which can be obtained using `Chebfun`, an efficient open-source Matlab add-on package available at `https://www.chebfun.org/`. This allows us to approximate the Sobolev space $\mathcal{H}$ defined in (3.1) using the r-dimensional subspace $\widetilde{\mathcal{H}} = \{\sum_{k=1}^{r} b_k\widehat{\varphi}_k : b_k \in \mathbb{R}\}$. Here, r is a truncation parameter that depends on the sample size n, which, in practice, can be chosen using cross-validation on the full sample $(X_1, Y_1), \ldots, (X_n, Y_n)$.

For fixed r, and for $1 \leq q \leq Q$, $1 \leq i \leq n_q$, and $1 \leq k \leq r$, let $\omega_{ik} = \int_0^1 X_i(s)\widehat{\varphi}_k(s)ds$; for each $1 \leq q \leq Q$, let $\Omega_{rq} = (\omega_{ik})_{1\leq i\leq n_q, 1\leq k\leq r}$ denote an $n_q \times r$ matrix; and let $\widehat{\Lambda}_r = \text{diag}\{\widehat{\rho}_1, \ldots, \widehat{\rho}_r\}$ denote an $r \times r$ diagonal matrix; let $\widetilde{Y}_q = (Y_1, \ldots, Y_{n_q})^\mathsf{T} \in \mathbb{R}^{n_q}$. If we write $\widetilde{\beta}_r(\cdot, \nu_q) = \sum_{k=1}^{r} \widetilde{b}_k^{(q)}\widehat{\varphi}_k \in \widetilde{\mathcal{H}}$, for $\widetilde{b}_k^{(q)} \in \mathbb{R}$, then, in order to approximate $\widehat{\beta}_{n,\lambda}(\cdot, \nu_q)$ in (3.2), for each $1 \leq q \leq Q$, we can compute the coefficients $\widetilde{b}_1^{(q)}, \ldots, \widetilde{b}_r^{(q)}$ by solving

$$\begin{aligned}
&(\widetilde{b}_1^{(q)}, \ldots, \widetilde{b}_r^{(q)}) \\
&= \underset{b_1^{(q)},\ldots,b_r^{(q)}}{\text{argmin}} \left\{ \frac{1}{2n_q} \sum_{i=1}^{n_q} \left| Y_i - \sum_{k=1}^{r} b_k^{(q)} \int_0^1 X_i(s)\, \widehat{\varphi}_k(s)\, ds \right|^2 + \frac{\lambda}{2} \sum_{k=1}^{r} b_k^{(q)^2} \widehat{\rho}_k \right\} \\
&= \underset{B_r^{(q)}}{\text{argmin}} \left\{ \frac{1}{2n_q} \big(\widetilde{Y}_q - \Omega_{rq}\, B_r^{(q)}\big)^\mathsf{T} \big(\widetilde{Y}_q - \Omega_{rq}\, B_r^{(q)}\big) + \frac{\lambda}{2} B_r^{(q)^\mathsf{T}} \widehat{\Lambda}_r\, B_r^{(q)} \right\},
\end{aligned} \tag{5.4}$$

where we write $B_r^{(q)} = (b_1^{(q)}, \ldots, b_r^{(q)})^\mathsf{T} \in \mathbb{R}^r$. A direct calculation shows that the solution to (5.4) is given by $\widehat{B}_r^{(q)} = \big(\Omega_{rq}^\mathsf{T}\, \Omega_{rq} + n_q\lambda\widehat{\Lambda}_r\big)^{-1}\Omega_{rq}^\mathsf{T}\, \widetilde{Y}_q$. Therefore, we can approximate the estimator $\widehat{\beta}_{n,\lambda}(\cdot, \nu_q)$ in (3.2) using $\widetilde{\beta}_r(\cdot, \nu_q) = \widehat{B}_r^{(q)^\mathsf{T}}\widehat{\varphi}$, where $\widehat{\varphi} = (\widehat{\varphi}_1, \ldots, \widehat{\varphi}_r)^\mathsf{T}$ denotes an r-dimensional vector of functions. Let $\widehat{\Phi}_r$ denote an $r \times r$ matrix with entries $\widehat{\Phi}_{k\ell} = \int_0^1 \widehat{\varphi}_k(t)\widehat{\varphi}_\ell(t)dt$, for $1 \leq k, \ell \leq r$. Then, $\widehat{\mathbb{T}}_n$ and $\widehat{\mathbb{V}}_n$ in (3.4) and (3.9) can be approximated using $\widetilde{\mathbb{T}}_n = \widehat{B}_r^{(Q)^\mathsf{T}}\widehat{\Phi}_r\widehat{B}_r^{(Q)}$ and $\widetilde{\mathbb{V}}_n = \big\{Q^{-1}(1-\nu_0)\sum_{q=1}^{Q} \nu_q^4\big(\widehat{B}_r^{(q)^\mathsf{T}}\widehat{\Phi}_r\widehat{B}_r^{(q)} - \widehat{B}_r^{(Q)^\mathsf{T}}\widehat{\Phi}_r\widehat{B}_r^{(Q)}\big)^2\big\}^{1/2}$, respectively. Finally, the decision rule in the test (4.15) is defined by rejecting the null hypothesis in (1.3) at the nominal level α if

$$\widetilde{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W}_Q)\widetilde{\mathbb{V}}_n + \Delta, \tag{5.5}$$

where $\mathcal{Q}_{1-\alpha}(\mathbb{W}_Q)$ denotes the $(1-\alpha)$-quantile of the pivotal distribution of $\mathbb{W}_Q$ in (5.2).

In order to choose the regularization parameter λ in (5.4) (for fixed r), we propose using a modified version of generalized cross-validation (GCV; e.g., see Wahba (1990)). Specifically, we choose λ as the value that minimizes the modified GCV score

$$\mathrm{GCV}(\lambda) = \sum_{q=1}^{Q} \frac{\|\widehat{Y}_q(\lambda) - \widetilde{Y}_q\|_2^2}{n_q|1 - \mathrm{tr}\{H_{rq}(\lambda)\}/n_q|^2}, \tag{5.6}$$

where $\widehat{Y}_q(\lambda) = \Omega_{rq}(\Omega_{rq}^{\mathsf{T}}\Omega_{rq} + n_q\lambda\widehat{\Lambda}_r)^{-1}\Omega_{rq}^{\mathsf{T}}\widetilde{Y}_q$, and $H_{rq}(\lambda)$ is the so-called hat matrix, with $\mathrm{tr}\{H_{rq}(\lambda)\} = \mathrm{tr}\{\Omega_{rq}(\Omega_{rq}^{\mathsf{T}}\Omega_{rq} + n_q\lambda\widehat{\Lambda}_r)^{-1}\Omega_{rq}^{\mathsf{T}}\}$.

5.2. Simulated data

In this section, we apply the pivotal test (5.5) to various settings of simulated data. In order to evaluate the function X on its domain $[0,1]$, we take 100 equally spaced time points, and for all the settings, we take the nominal level $\alpha = 0.05$. For the true slope function β_0 in functional linear regression (1.1), we consider the following two settings:

(S1) Let $f_1 \equiv 1$, $f_{j+1}(s) = \sqrt{2}\cos(j\pi s)$, for $j \geq 1$, and define $\beta_0 = \sqrt{\delta}\widetilde{\beta}_0/\|\widetilde{\beta}_0\|_{L^2}$, where $\widetilde{\beta}_0(s) = f_1(s) + 4\sum_{j=2}^{50}(-1)^{j+1}j^{-2}f_j(s)$, for $s \in [0,1]$.

(S2) $\beta_0(s) = \sqrt{\delta}\widetilde{\beta}_0(s)/\|\widetilde{\beta}_0\|_{L^2}$, where $\widetilde{\beta}_0(s) = \exp(-s/4)$, for $s \in [0,1]$.

The first setting (S1) is similar to those used in Yuan and Cai (2010). For both settings (S1) and (S2), the slope function is standardized such that $d_0 = \|\beta_0\|_{L^2}^2 = \delta > 0$, where we take various values of δ and Δ in the relevant hypotheses (1.3). For the predictor process $\{X_i\}_{i\in\mathbb{Z}}$, we use a similar setting to that in Dette, Kokot and Volgushev (2020) by generating i.i.d. random variables $\eta_i = \sum_{j=1}^{50} j^{-1}Z_{ij}f_j$, where $Z_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Normal}(0,1)$, and consider the following two settings:

(i) the functional moving average process FMA(1), defined by $X_i = \eta_i + \theta_i\eta_{i-1}$, for $1 \leq i \leq n$, where $\theta_i \overset{\text{i.i.d.}}{\sim} \mathrm{uniform}(-1/\sqrt{2}, 1/\sqrt{2})$.

(ii) The i.i.d. case $X_i = \sqrt{7/6}\,\eta_i$, such that (i) and (ii) have the same point-wise variance.

For the errors ε_i in (1.1), we generate i.i.d. standard normal random variables ξ_i, and take $\varepsilon_i = c_\varepsilon(\xi_i + \upsilon_{i,1}\xi_{i-1} + \upsilon_{i,2}\xi_{i-2})$, where $\upsilon_{i,j} \overset{\text{i.i.d.}}{\sim} \mathrm{uniform}(-1/\sqrt{2}, 1/\sqrt{2})$, for $i \in \mathbb{Z}$ and $j = 1,2$, and the constant $c_\varepsilon > 0$ is chosen such that $\mathrm{var}(\varepsilon_i)/\mathrm{var}\{\int_0^1 \beta_0(s)X(s)ds\} = 0.3$.

We compare the numerical performance of our proposed pivotal test (5.5) (denoted as DT in the following discussion) with that of the method in

Table 1. Left part: decisions of the test (5.5) for the relevant hypotheses (1.3) with different values of Δ and nominal levels using bike-sharing data. "R" stands for rejecting the null hypothesis, and "-" stands for no rejection. Right part: One-sided and two-sided confidence intervals of $d_0 = \int_0^1 |\beta_0(s)|^2 ds$.

Δ	0.41	0.42	1.10	1.11	1.35	1.36	One-sided CI	Two-sided CI
$\alpha = 0.01$	R	-	-	-	-	-	$(0, 4.51]$	$[0.21, 4.79]$
$\alpha = 0.05$	R	R	R	-	-	-	$(0, 3.82]$	$[0.89, 4.11]$
$\alpha = 0.10$	R	R	R	R	R	-	$(0, 3.57]$	$[1.19, 3.82]$

Kutta, Dierickx and Dette (2022) (denoted as KDD), for the relevant hypotheses (1.3). Figures 1 and 2 display the empirical rejection probabilities of both tests calculated from 500 simulation runs, where we vary the values of $d_0 = \delta$ in (S1) and (S2), together with different values of the threshold Δ; we took $\nu_0 = 1/2$ and chose ω as the discrete uniform distribution on $\{\nu_q\}_{q=1}^{Q}$, with $Q = 25$, where ν_q is defined in (5.1); for the sample sizes, we took $n = 50$ and 200 observations; we chose r using cross-validation based on the whole sample, and chose λ using GCV in (5.6). The results confirm our theoretical findings in Theorem 5, and are summarized as follows. (1) Both DT and KDD provide a reasonable approximation of the nominal level α when $\Delta = d_0 = \delta$. (2) For both DT and KDD, the rejection probabilities are close to zero when $d_0 < \Delta$ (interior of the null hypothesis). (3) For both DT and KDD, when $d_0 > \Delta$ (interior of the alternative), the empirical rejection probabilities increase with Δ, and in most cases, larger sample sizes ($n = 200$) attain higher empirical rejection probabilities. (4) In most cases, DT outperforms KDD in terms of empirical power (when $d_0 > \Delta$).

5.3. Data example: bike-sharing

Bike-sharing can potentially alleviate the environmental impact of transport activities, and thus individuals and bike-sharing companies are investigating the effect of environmental factors on bike sharing. In this example, we use our proposed pivotal inference tools to investigate the impact of wind speed on bike rental activities on workdays. We use the bike-sharing data of Captial Bike Sharing (CBS) at Washington, D.C., the United States, for 2011 (Fanaee-T and Gama (2014)), together with hourly measurements of local wind speed, obtained from the R package `ISLR2` (James et al. (2021)). This data set was analyzed in Kim et al. (2018) in the functional response context, where the response curves consist of hourly counts of bike rentals. In our case, Y_i are scalar variables taking values in $[0, 1]$, evaluating the daily frequencies of bike rental, obtained from a linear transformation of the daily count of bike rentals. For each day i, the predictor curves X_i represent the hourly measurements of wind speed. It is known that environmental conditions such as wind speed

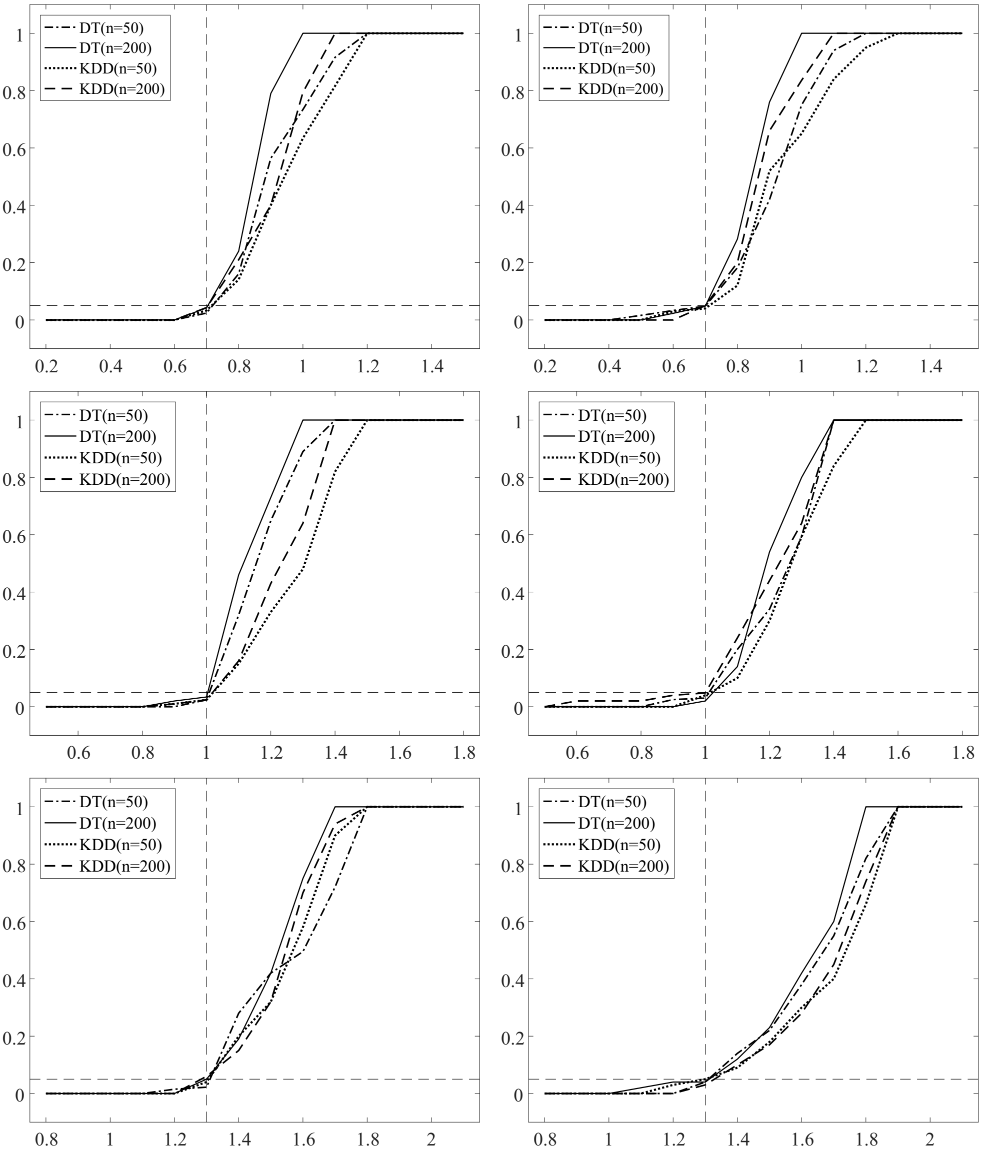

Figure 1. Empirical rejection probabilities of DT and KDD for the relevant hypotheses (1.3) under Setting (S1) with error setting (i) in column 1 and (ii) in column 2, with various δ (x-axis). The horizontal and vertical dashed lines are $\alpha = 0.05$ and $\Delta = 0.7, 1, 1.3$, respectively (first, second, and third row, respectively).

are of temporal dependence, making our pivotal inference approach, which does not depend on long-run variance estimates, attractive. We extracted the data for the 250 workdays in 2011, and removed missing data to obtain $n = 247$ observations. The hourly measurements of wind speed are normalized using a linear transformation, such that the data take values in $[0, 1]$. The curves X_i are obtained by projecting the hourly observations onto the space spanned by

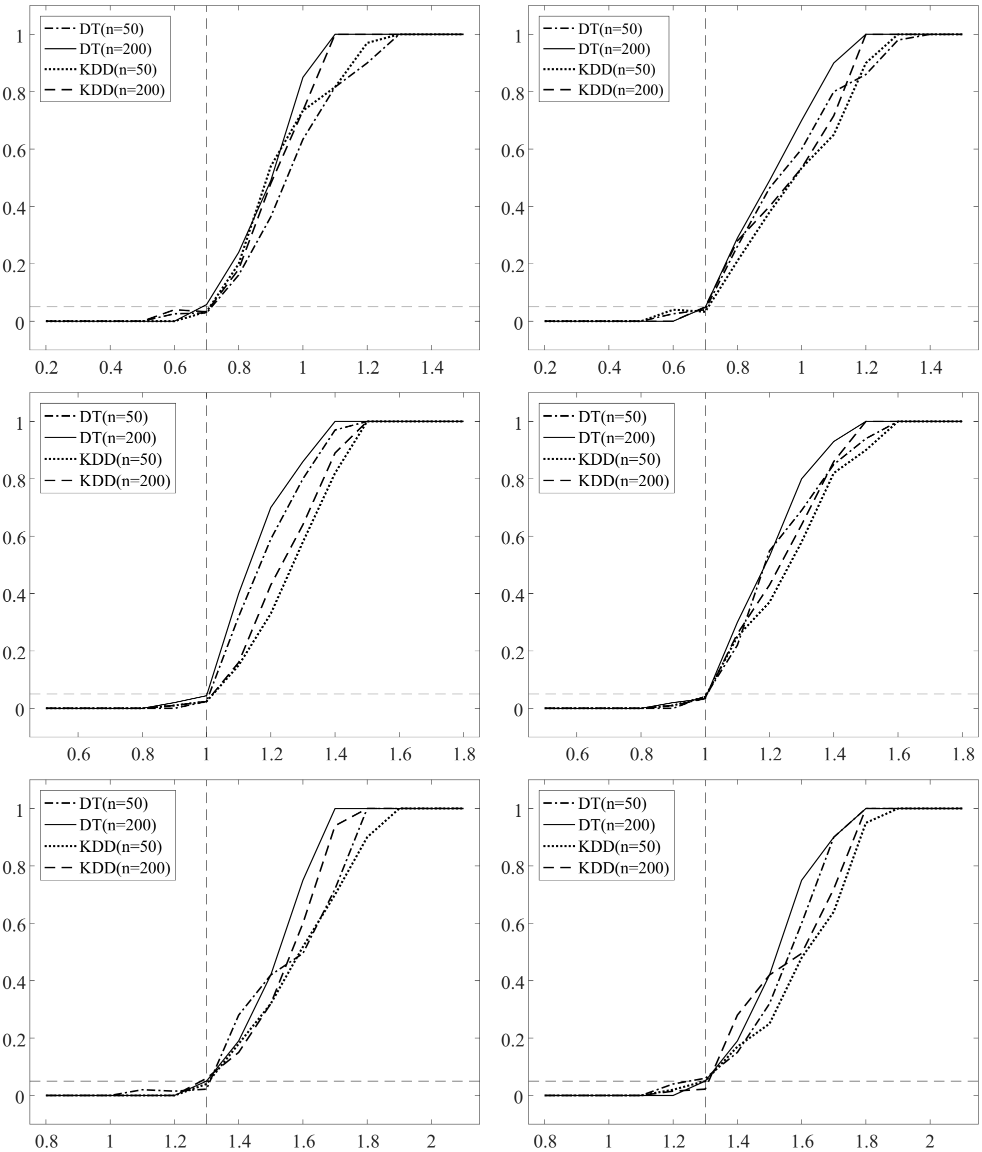

Figure 2. Empirical rejection probabilities of DT and KDD for the relevant hypotheses (1.3) under Setting (S2) with error setting (i) in column 1 and (ii) in column 2 with various δ (x-axis). The horizontal and vertical dashed lines are $\alpha = 0.05$ and $\Delta = 0.7, 1, 1.3$, respectively (first, second, and third row, respectively).

the first seven Fourier basis functions on $[0, 1]$, and are evaluated on an equally spaced grid $t = 0.01, 0.02, \ldots, 1$. Figure 3 displays the histogram of Y_i, together with the wind speed curves.

We centered the data, considered the relevant hypotheses (1.3), and took $\nu_0 = 1/2$ and $Q = 25$ in (5.1). The left part of Table 1 displays the decisions of our test with different values of the threshold Δ and nominal levels $\alpha = 0.10$,

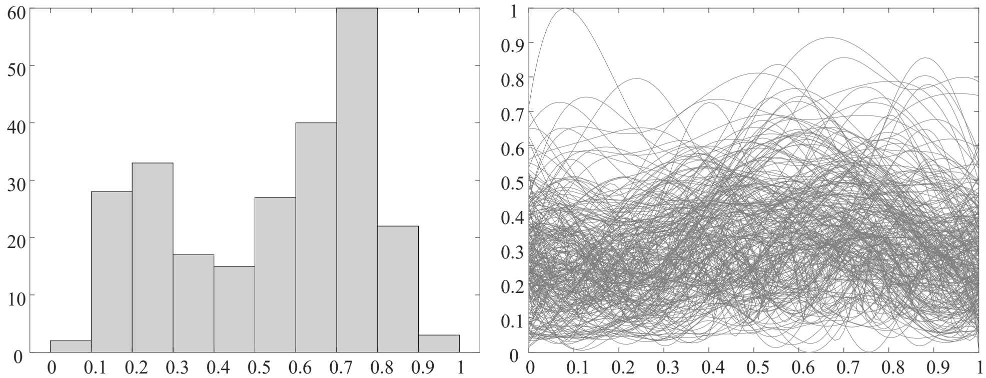

Figure 3. Left panel: histogram of daily bike rentals (scaled to the interval $[0, 1]$). Right panel: corresponding wind speed curves.

0.05, and 0.01. For instance, the largest value of Δ such that the test (5.5) rejects the null hypothesis in (1.3) at level $\alpha = 0.05$ is given by $\Delta = 1.01$, and this value is 1.35 at level $\alpha = 0.10$. If we wish to avoid specifying the threshold Δ for a test (see the discussion in Remark 2), we can construct one-sided or two-sided confidence intervals for $d_0 = \int_0^1 |\beta_0(s)|^2 ds$, which are defined in (4.16) and (4.18), respectively. The results are displayed in Table 1 for various confidence levels.

Supplementary Material

The online Supplementary Material contains the proofs of our theoretical results.

Acknowledgments

This work was partially supported by the DFG Research unit 5381 *Mathematical Statistics in the Information Age*. The authors thank Tim Kutta for helpful discussions and sharing with us the R code for Kutta, Dierickx and Dette (2022). We are also grateful to the reviewers for their constructive comments on an earlier version of this paper.

References

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statist. Sci.* **2**, 317–335.

Berkes, I., Horváth, L. and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Process. Appl.* **123**, 385–403.

Blanchard, G. and Fermanian, J. B. (2021). Nonasymptotic one-and two-sample tests in high dimension with unknown covariance structure. *arXiv:2109.01730.*

Brutsche, J. and Rohde, A. (2022). Sharp adaptive similarity testing with pathwise stability for ergodic diffusions. *arXiv:2203.13776.*

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

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45**, 11–22.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Stat.* **30**, 241–255.

Chen, M. and Song, Q. (2015). Simultaneous inference of the mean of functional time series. *Electron. J. Stat.*, **9**, 1779–1798.

Chow, S. C. and Liu, P. J. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies.* CRC Press.

Cui, Y. and Zhou, Z. (2022). Simultaneous inference for time series functional linear regression. *arXiv:2207.11392.*

van Delft, A. and Eichler, M. (2018). Locally stationary functional time series. *Electron. J. Stat.* **12**, 107–170.

Dette, H., Kokot, K. and Aue, A. (2020). Functional data analysis in the Banach space of continuous functions. *Ann. Statist.* **48**, 1168–1192.

Dette, H., Kokot, K. and Volgushev, S. (2020). Testing relevant hypotheses in functional time series via self-normalization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82**, 629–660.

Dette, H and Tang, J. (2021) Statistical inference for function-on-function linear regression. *arXiv:2109.13603.*

Fanaee-T, H. and Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence* **2**, 113–127.

Ferraty, F. and Vieu P. (2010). *Nonparametric Functional Data Analysis.* Springer.

Fogarty, C. B. and Small, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Ann. Appl. Stat.* **8**, 2002–2026.

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.

Hao, M., Liu, K. Y., Xu, W. and Zhao, X. (2021). Semiparametric inference for the functional Cox model. *J. Amer. Statist. Assoc.* **116**, 1319–1329.

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

Hodges J. L. and Lehmann, E. L. (1954). Testing the approximate validity of statistical hypotheses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **16**, 261–268.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38**, 1845–1884.

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

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

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R.* Springer, New York.

Kim, J. S., Staicu, A. M., Maity, A., Carroll, R. J. and Ruppert, D. (2018). Additive function-on-function regression. *J. Comput. Graph. Statist.* **27**, 234–244.

Kokoszka, P., Rice, G. and Shang, H. L. (2017). Inference for the autocovariance of a functional time series under conditional heteroscedasticity. *J. Multivariate Anal.* **162**, 32–50.

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

Kutta, T., Dierickx, G. and Dette, H. (2022) Statistical inference for the slope parameter in functional linear regression. *Electron. J. Statist.* **16**, 5980–604 . DOI: 10.1214/22-EJS2078.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

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

Lepski, O. V. and Spokoiny, V. G. (1999). Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5**, 333–358.

Liu, W., Hayter, A. J. and Wynn, H. P. (2007a). Operability region equivalence: Simultaneous confidence bands for the equivalence of two regression models over restricted regions. *Biom. J.* **49**, 144–150.

Liu, W., Bretz, F., Hayter, A. J. and Wynn, H. P. (2009). Assessing non-superiority, non-inferiority of equivalence when comparing two regression models over a restricted covariate region. *Biometrics* **65**, 1279–1287

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.* **96**, 1066–1076.

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

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

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

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72**, 343–366.

Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24**, 2477–2498.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105**, 1228–1240.

Shin, H. and Lee, S. (2016). An RKHS approach to robust functional linear regression. *Statist. Sinica* **26**, 255–272.

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

Tekbudak, M. Y., Alfaro-Córdoba, M., Maity, A. and Staicu, A. M. (2019). A comparison of testing methods in scalar-on-function regression. *AStA Adv. Stat. Anal.* **103**, 411–436.

Wahba, G. (1990). *Spline Models for Observational Data*. Society For Industrial and Applied Mathematics, Philadelphia.

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

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. CRC Press.

Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873–2903.

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

Holger Dette

Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum D-44801, Germany.

E-mail: holger.dette@ruhr-uni-bochum.de

Jiajun Tang
Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum D-44801, Germany.
E-mail: jiajun.tang@ruhr-uni-bochum.de

(Received March 2022; accepted November 2022)