AN RKHS APPROACH FOR PIVOTAL INFERENCE IN FUNCTIONAL LINEAR REGRESSION

Holger Dette and Jiajun Tang

Ruhr-Universität Bochum

Abstract: We develop methodology for testing hypotheses regarding the slope function in functional linear regression for time series via a reproducing kernel Hilbert space approach. In contrast to most of the literature, which considers 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. An asymptotically pivotal test is proposed, which does not require the estimation of nuisance parameters and long-run covariances. The key technical tools 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. We also illustrate the potential of our methods by means of 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

Statistical methods for analysing functional data have been extensively developed in the past decades, as reviewed in the monographs Ramsay and Silverman (2005), Ferraty and Vieu (2010), Horváth and Kokoszka (2012), Hsing and Eubank (2015) and the survey article by Wang, Chiou and Müller (2016). Because of its good interpretability, the functional linear regression model

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

has become a useful toolbox for functional data analysis and has gained considerable attention (see, for exemple, Cardot, Ferraty and Sarda, 1999; Müller and Stadtmüller, 2005; Yao, Müller and Wang, 2005; Hall and Horowitz, 2007; Yuan and Cai, 2010, among many others). In this paper $\{(X_i, \varepsilon_i)\}_{i\in\mathbb{Z}}$ denotes a strictly stationary time series, where the X_i's are mean zero square-integrable random functions on the interval $[0,1]$, and the ε_i's are real valued centred random noise.

As the slope function β_0 characterizes the dependence between the predictor and the response, many authors have worked on its estimation and corresponding statistical inference. A popular method for analysing the slope function in model (1.1) are functional principle components (FPC) (see, for example, Yao, Müller and Wang, 2005; Hall and Horowitz, 2007; Horváth and Kokoszka, 2012; Hilgert, Mas and Verzelen, 2013, among many others). Other authors considered a reproducing kernel Hilbert space (RKHS) approach to develop inference tools for β_0 and corresponding theoretical results regarding consistency and optimality. Yuan and Cai (2010) and Cai and Yuan (2012) studied an RKHS estimator and its prediction risk in the scalar-on-function linear regression model. Shang and Cheng (2015) proposed an RKHS framework of inference for the generalized functional linear regression and Hao et al. (2021) considered the functional Cox model. These authors additionally suggested tests for the nullity of the slope function. Recently, Dette and Tang (2021) used an RKHS approach to develop statistical inference methodology in the function-on-function linear model, measuring deviations from the null hypothesis with respect to the sup-norm with a focus on confidence bands. Statistical inference for functional time series has also found considerable interest in the recent literature (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, among many others).

A common feature of most references about statistical theory for the functional linear regression model consists in the fact that the proposed methodology depends on the knowledge of nuisance parameters appearing in the asymptotic variance of the estimators of the slope function. As will be seen in Section 3 below, these parameters are related to the long-run covariance structure of the data and describe the behaviour of the sequence of solutions of a system of estimated integro-differential equations induced by the covariance operator of the predictor, and therefore their estimation is not an easy problem. In the case of independent data (as considered in all references using the RKHS approach), several estimators have been proposed and studied. On the other hand, for time series data these nuisance parameters would be of an even more complicated structure because of the dependencies in the data, which would make their estimation even more difficult.

The purpose of the present paper is to develop pivotal statistical inference tools for the slope function β_0 in the functional linear regression model (1.1) using an RKHS approach, which avoids the estimation of nuisance parameters. While a large part of the literature has its 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 many others), this paper takes a different point of view and develops a pivotal test for the hypotheses

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

Here, $\Delta > 0$ is a (small) pre-specified 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 formulation of the hypotheses in (1.3) are symmetric, in the sense that the null and the alternative can be interchanged. This allows us also to investigate at a controlled type I error that the effect of the covariate on the response is negligible by testing the hypotheses

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

Throughout this paper we will call hypotheses of the form (1.2) "classical" and hypotheses of the form (1.3) or (1.4) "relevant" hypotheses, respectively, and the pros and cons of these hypotheses are discussed in more detail in Section 2.

The aim of this article is the development of pivotal methodology for testing relevant hypotheses (1.3) (or (1.4)) with no need to estimate nuisance parameters. Our approach is based on reproducing kernel Hilbert space (RKHS) and a novel self-normalization technique, which has recently been introduced by Dette, Kokot and Volgushev (2020) in the context of testing relevant hypotheses regarding the mean and covariance functions of stationary time series and differs substantially from the the common self-normalization approaches proposed for testing classical hypotheses regarding finite dimensional parameters (see Lobato, 2001; Shao, 2010; Shao and Zhang, 2010, among many others). As statistical inference regarding the slope function is an inverse problem, it cannot be directly treated by the methods developed in these papers. In Section 3 we introduce a sequential reproducing kernel Hilbert space 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 $\{(X_i, Y_i)\}_{i=1,\ldots,\lfloor n\nu\rfloor}$ (see Theorem 2 and the discussion in the following paragraph). In Section 5 we provide details for the numerical implementation of our approach, present simulated data experiments and a real data example. The proofs of our theoretical results are given in the online supplementary materials. The Matlab program for implementing our method is available at `https://github.com/jttang/SN_RKHS`.

To our best knowledge, testing relevant hypotheses regarding the slope function has only been considered recently by Kutta, Dierickx and Dette (2021). Roughly speaking, these authors investigated a normal equation corresponding to the linear model (1.1), which is then solved by an application of a regularized inverse based on a spectral-cut-off series estimator. Although such an approach has some theoretical advantages, its practical usefulness is limited by the fact that it requires the estimation of the spectral decomposition of the regularized inverse. In contrast, the estimator considered in this paper is defined as the minimizer of a regularized loss function in an appropriate reproducing kernel Hilbert space and we demonstrate in Section 5 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, and a more reasonable question is, if this effect is small and negligible. In this paper 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, but other norms could 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 are 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 bio-statistics community (see the mongraphs of Chow and Liu, 1992; Wellek, 2010). Moreover, in the context of drug development, several authors considered relevant hypotheses for comparing dose response files (see Liu et al., 2007, 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) or Lepski and Spokoiny (1999) for some early and Blanchard and Fermanian (2021); Brutsche and Rohde (2022) for some more recent references.

From a statistical point of view, the use of classical hypotheses or relevant hypotheses is often a subjective decision. Relevant hypotheses should be preferred if there is clear evidence that exact equality cannot hold (otherwise one is testing a hypothesis which is known to be not true in advance). As pointed out by Berger and Delampady (1987), this situation appears rather frequently. On the other hand, the use of the hypotheses (1.3) and (1.4) comes with the price that one has 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 as relevant in the concrete application. We note that for bioequivance testing these discussions have already been completed (see Chow and Liu, 1992; Wellek, 2010). Here regulators such as EMA or FDA have have defined thresholds for concrete applications.

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

3. The RKHS approach to functional linear regression

We first introduce some notations 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 observation $(X_1, Y_1), \ldots, (X_n, Y_n)$ 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

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

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

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

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

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

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 the scenario where we use the full sample $\{(X_i, Y_i)\}_{i=1}^n$ for the estimation of the slope function β_0, and we use the statistic

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

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

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

where

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

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

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 eigen-system 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 eigen-functions φ_k and on the eigen-values ρ_k (see Assumption A2 and the discussion below for more details).

As a consequence, 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 due to the fact that σ_d^2 is defined as the limit of a series, which in turns relies on the operator τ_λ in (3.8) and therefore depends on the eigen-system $\{(\rho_k, \varphi_k)\}_{k\geq 1}$ 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, which makes its 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 the knowledge of σ_d^2 in (3.6) and the nuisance parameters a and D. The details of this approach will be worked out in the following Section 4, but a heuristic argument ignoring the technical details is as 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

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

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). As a consequence of Theorem 3 below, we find

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

where $\mathbb{B}$ denotes the standard Brownian motion. In particular, the ratio $(\widehat{\mathbb{T}}_n-d_0)/\widehat{\mathbb{V}}_n$ will be asymptotically free. A consistent and asymptotic level α-test for the hypotheses (1.3) can therefore be obtained by comparing the statistic $(\widehat{\mathbb{T}}_n-\Delta)/\widehat{\mathbb{V}}_n$ with the $(1-\alpha)$-quantile of the limiting distribution. The details will be worked out 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 will be crucial for our approach. For this purpose, we define

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

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

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

and $\mathcal{D}^3L_{n,\lambda,\nu}(\beta)\equiv0$. 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 $\mathrm{E}\{\mathcal{D}^2L_{n,\lambda,\nu}(\beta)\beta_1\beta_2\}=\langle\beta_1,\beta_2\rangle_K$, 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 A1. 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 A1 is a common condition in the literature (see, for example, 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 a reproducing kernel Hilbert space (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 A2 (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 $\|\cdot\|_K$.

It was proved in Proposition 2.2 in Shang and Cheng (2015) that Assumption A2 is satisfied for the eigen-system $\{(\rho_k, \varphi_k)\}_{k\geq 1}$ 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 under Assumption A2 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 $k, k' \geq 1$, so 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 A2, $\langle W_\lambda(\varphi_k), \varphi_{k'} \rangle_K = \lambda J(\varphi_k, \varphi_{k'}) = \lambda \rho_k\, \delta_{kk'}$, for any $k, k' \geq 1$, so that in view of (4.5),

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

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, $\langle \tau_\lambda(z), \varphi_k \rangle_K = \langle z, \varphi_k \rangle_{L^2}$, so that we obtain the representation (3.8) for the operator τ_λ. Now, for any $\beta, \beta_1, \beta_2 \in \mathcal{H}$ define

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

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

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

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

Assumption A3. For $i \in \mathbb{Z}$, (X_i, Y_i) is generated by the model (1.1) and satisfies:

(A3.1) $X_i = g(\ldots, \xi_{i-1}, \xi_i)$ 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 $\xi_i = \xi_i(t, \boldsymbol{\omega})$ is jointly measurable in $(t, \boldsymbol{\omega})$. The ξ_i's and the η_i's are independent and identically distributed.

(A3.2) For any $s \in [0,1]$, $\mathrm{E}\{X_0(s)\} = \mathrm{E}(\varepsilon_0) = 0$. For some $\delta \in (0,1)$, $\mathrm{E}|\varepsilon_0|^{2+\delta} < \infty$.

(A3.3) 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 the $\xi^*_{i,\ell,k}$'s and the $\eta^*_{i,\ell,k}$'s are independent copies of ξ_0 and η_0, and are independent of $\{\xi_i\}_{i\in\mathbb{Z}}$ and $\{\eta_i\}_{i\in\mathbb{Z}}$, respectively.

Assumption A4 (Regularity conditions)**.**

(A4.1) There exists a constant $\varpi > 0$ such that $\mathrm{E}\{\exp(\varpi\|X_0\|_{L^2}^2)\} < \infty$.

(A4.2) 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 $c_0 > 0$.

(A4.3) The true slope function β_0 is such that $\sum_{k=1}^\infty \rho_k^2 V^2(\beta_0, \varphi_k) < \infty$.

(A4.4) 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 A5. The constants a and D in Assumption A2 and the regularization parameter λ in (3.2) satisfy $\lambda = o(1)$, $n^{-1}\lambda^{-(2a+1)/D} = o(1)$, $n\lambda^{2+(2a+1)/(2D)} = o(1)$ 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 A4.1 requires an exponential tail of $\|X_0\|_{L^2}$. This condition can be satisfied for any stochastic processes with almost surely bounded L^2-norm, and can also be satisfied for Gaussian processes with 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 A4.2 is a common condition in linear regression models for functional data; see, for example Cai and Yuan (2012) and Shang and Cheng (2015). Assumption A4.3 corresponds to the so-called undersmoothing scenario in Shang and Cheng (2015); see their Remark 3.2. Finally, Assumption A5 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 materials.

Theorem 1 (Uniform Bahadur representation)**.** *Suppose Assumptions A1–A5 are satisfied. Then, for any fixed (but arbitrary) $\nu_0 \in (0, 1]$,*

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

where for the constant $\varsigma > 0$ in Assumption A5, $v_n = n^{-1/2}\lambda^{-\varsigma}(\lambda^{1/2}+n^{-1/2}\lambda^{-(2a+1)/(4D)})(\log n)^{1/2}$.

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

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

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 will 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

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

The following theorem is proved in Section S1.2 of the online supplementary materials.

Theorem 2 (Weak invariance principle)**.** *Suppose Assumptions A1–A5 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* (A4.4)*, 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 $\nu \in [0,1]$. As a consequence, we obtain from Theorem 1 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 regarding 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 establishing the weak convergence of the process $\{\widehat{\mathbb{G}}_n(\nu)\}_{\nu\in[\nu_0,1]}$ is proved in Section S1.3 of the online supplementary materials.

Theorem 3. *If Assumptions A1–A5 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 by the continuous mapping theorem and Theorem 3 that

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

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

Theorem 4. *Suppose Assumptions A1–A5 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*

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

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 $\mathbb{W}$, since its distribution does not depend on the nuisance parameters (namely a and D in Assumption A2, and the σ_d^2 in (3.6)) or the eigen-system $\{(\rho_k,\varphi_k)\}_{k\geq1}$. Moreover, the distribution of $\mathbb{W}$ in (4.14) can be easily simulated from computer-generated sample paths of standard Brownian motions. Therefore, we propose to reject the null hypothesis in (1.3) at nominal level α, if

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

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 online supplementary materials provides a theoretical justification of the consistency of the test defined in (4.15) at nominal level α.

Theorem 5. *Assume $\Delta > 0$. Under Assumptions A1–A5, we have*

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

Remark 2 (Further statistical consequences)**.**

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

If this is not possible, we recommend to construct a confidence interval for the (squared) L^2-norm of the slope function. To be precise, 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

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

as $n \to \infty$, where we have used 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 $d_0 = 0$, since $\widehat{\mathbb{T}}_n, \widehat{\mathbb{V}}_n \geq 0$

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 $\Delta^{(1)} < \ldots < \Delta^{(L)}$ simultaneously, for some $L \in \mathbb{N}_+$. In particular, rejection for a $\Delta^{(L_0)}$ means rejection 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), i.e $H_0 : d_0 = \int_0^1 |\beta_0(s)|^2 ds = 0$, a likelihood ratio type test was proposed by Shang and Cheng (2015). As 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 in the case $d_0 = 0$,

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

where Γ is a mean zero Gaussian process with covariance function $\text{cov}\{\Gamma(s_1, \nu_1), \Gamma(s_2, \nu_2)\} = \min\{\nu_1, \nu_2\} C_U(s_1, s_2)$ and C_U is defined in (A4.4). This limit distribution is different from that of $\mathbb{W}$ in (4.14), and is not pivotal since 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 regarding the implementation of the proposed tests for the relevant hypotheses. To begin with, in practice we may choose the probability measure ω in (3.9) and (4.14) as the discrete uniform distribution on the interval $[\nu_0, 1]$. To be precise, for some positive integer Q, let

$$\nu_q = \nu_0 + q(1 - \nu_0)/Q\,, \qquad \text{for } 1 \leq q \leq Q\,. \tag{5.1}$$

Then, we define ω as the discrete uniform distribution supported on the set $\{\nu_q\}_{q=1}^{Q}$ with equal probability mass $1/Q$, so 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 easily obtained 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 $n_q = \lfloor \nu_q n \rfloor$ $(q = 1, \ldots, Q)$. Since $\widehat{\beta}_{n,\lambda}(\cdot, \nu_q)$ is defined as the solution of a penalized minimization problem on an infinite-dimensional 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 to choose the regularization parameter λ in (3.2). We first observe from Assumption A2 that $J(\varphi_{k\ell}, \varphi_{k'\ell'}) = \rho_{k\ell}\delta_{kk'}\delta_{\ell\ell'}$, so that for $\beta = \sum_{k=1}^{\infty} b_k \varphi_k \in \mathcal{H}$ and for $b_k \in \mathbb{R}$, 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)

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

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 by 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) by the r-dimensional subspace $\widetilde{\mathcal{H}} = \big\{ \sum_{k=1}^{r} b_k \widehat{\varphi}_k : b_k \in \mathbb{R} \big\}$. Here, r is a truncation parameter that depends on the sample size n, which in practice can be chosen via cross-validation using 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; let $\widehat{\Lambda}_r = \mathrm{diag}\big\{\widehat{\rho}_1, \ldots, \widehat{\rho}_r\big\}$ denote an $r \times r$ diagonal matrix; let $\widetilde{Y}_q = (Y_1, \ldots, Y_{n_q})^\intercal \in \mathbb{R}^{n_q}$. 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

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

where we write $B_r^{(q)} = (b_1^{(q)}, \ldots, b_r^{(q)})^\intercal \in \mathbb{R}^r$. A direct calculation shows that the solution to (5.4) is given by $\widehat{B}_r^{(q)} = \left(\Omega_{rq}^\mathsf{T}\Omega_{rq} + n_q\lambda\widehat{\Lambda}_r\right)^{-1}\Omega_{rq}^\mathsf{T}\widetilde{Y}_q$. Therefore, we can approximate the estimator $\widehat{\beta}_{n,\lambda}(\cdot,\nu_q)$ in (3.2) by $\widetilde{\beta}_r(\cdot,\nu_q) = \widehat{B}_r^{(q)^\mathsf{T}}\widehat{\varphi}$, 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 $1\le k,\ell\le r$. Then, $\widehat{\mathbb{T}}_n$ and $\widehat{\mathbb{V}}_n$ in (3.4) and (3.9) can be approximated by $\widetilde{\mathbb{T}}_n = \widehat{B}_r^{(Q)^\mathsf{T}}\widehat{\Phi}_r\widehat{B}_r^{(Q)}$ and $\widetilde{\mathbb{V}}_n = \left\{Q^{-1}(1-\nu_0)\sum_{q=1}^Q \nu_q^4\left(\widehat{B}_r^{(q)^\mathsf{T}}\widehat{\Phi}_r\widehat{B}_r^{(q)} - \widehat{B}_r^{(Q)^\mathsf{T}}\widehat{\Phi}_r\widehat{B}_r^{(Q)}\right)^2\right\}^{1/2}$, respectively. The decision rule in the test (4.15) is finally defined by, rejecting the null hypothesis in (1.3) at 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 to use a modified version of the generalized cross-validation (GCV, see, for example, Wahba, 1990). To be precise, we choose λ to be 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

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

(S1) Let $f_1 \equiv 1$, $f_{j+1}(s) = \sqrt{2}\cos(j\pi s)$, for $j\ge 1$, 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 $s \in [0,1]$.

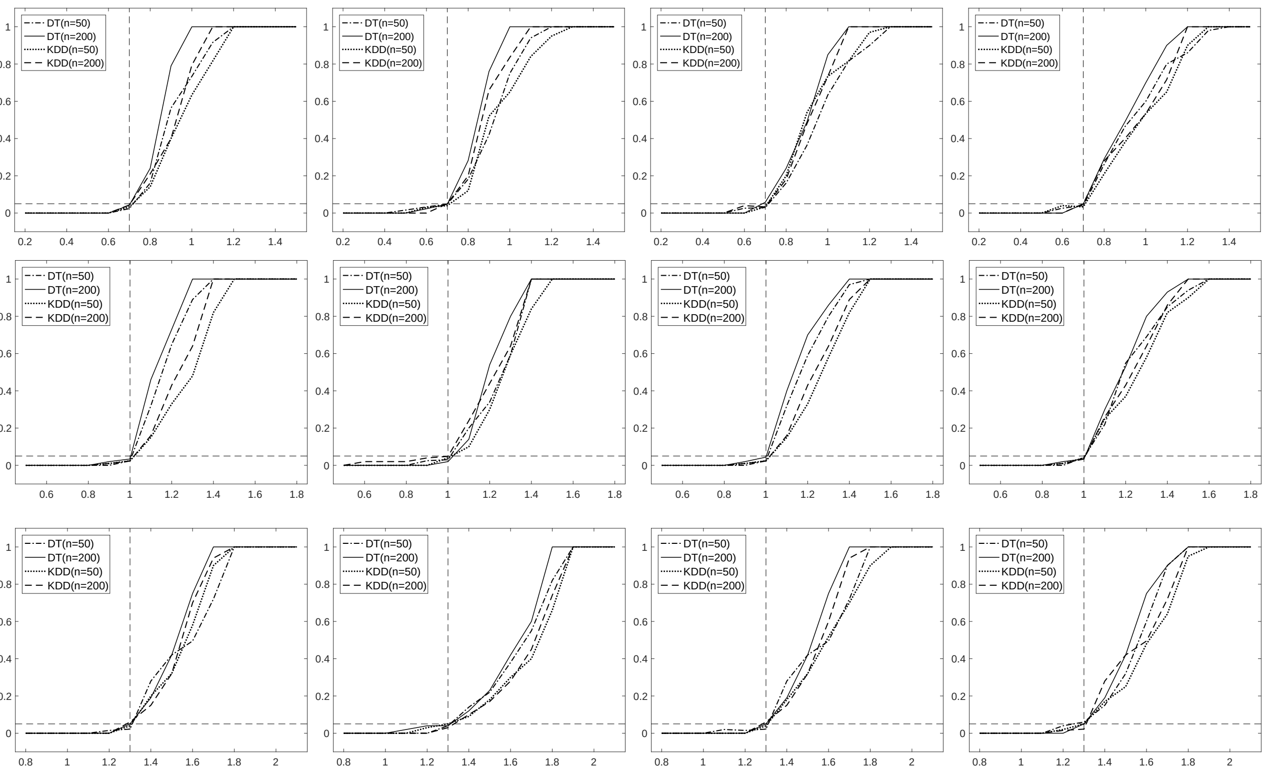
The first setting (S1) is similar to the ones used in Yuan and Cai (2010), and for both settings (S1) and (S2), the slope function is standardized such that $d_0 = \|\beta_0\|_{L^2}^2 = \delta > 0$, where we took various values of δ and Δ in relevant hypotheses (1.3). For the predictor process $\{X_i\}_{i\in\mathbb{Z}}$, we adopted a similar setting as 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 $Z_{ij} \overset{\text{iid}}{\sim} \text{Normal}(0,1)$, and considered 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 $\theta_i \overset{\text{iid}}{\sim} \text{uniform}(-1/\sqrt{2}, 1/\sqrt{2})$.

(ii) The i.i.d. case $X_i = \sqrt{7/6}\,\eta_i$, so that (i) and (ii) has 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 $\upsilon_{i,j} \overset{\text{iid}}{\sim} \text{uniform}(-1/\sqrt{2}, 1/\sqrt{2})$, for $i \in \mathbb{Z}$ and $j = 1, 2$, and the constant $c_\varepsilon > 0$ is chosen such that $\text{var}(\varepsilon_i)/\text{var}\{\int_0^1 \beta_0(s)X(s)ds\} = 0.3$.

We compared the numerical performance of our proposed pivotal test (5.5) (denoted by DT in the following discussion) with the method in Kutta, Dierickx and Dette (2021) (denoted by KDD), for the relevant hypotheses (1.3). Figure 1 displays 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 threshold Δ; we took $\nu_0 = 1/2$ and chose ω as the discrete uniform distribution on the $\{\nu_q\}_{q=1}^{Q}$, with $Q = 25$, where the ν_q is defined in (5.1); for the sample sizes we took $n = 50$ and 200 observations; we chose r by cross-validation using the whole sample, and chose λ through GCV in (5.6). The results confirm our theoretical findings in Theorem 5 and can be summarized as follows. (1) Both DT and KDD provide a reasonable approximation of the nominal level α when $\Delta = d_0 = \delta$.



(S1,i) (S1,ii) (S2,i) (S2,ii)

Figure 1: *Empirical rejection probabilities of DT and KDD for the relevant hypotheses* (1.3) *under 4 settings (columns 1–4) with various δ (x-axis). The horizontal and vertical dashed lines are $\alpha = 0.05$ and $\Delta = 0.7, 1, 1.3$ (first, second and third row, respectively).*

(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 increases with Δ, and in most cases, larger sample size ($n = 200$) attains higher empirical rejection probabilities. (4) In most cases, DT performs better than KDD in terms of empirical power (when $d_0 > \Delta$).

5.3 Data example: bike-sharing

Bike-sharing has received increasing attention in recent years with the initiative to alleviate environmental impact of the transport activities, and it is of interest for individuals and bike

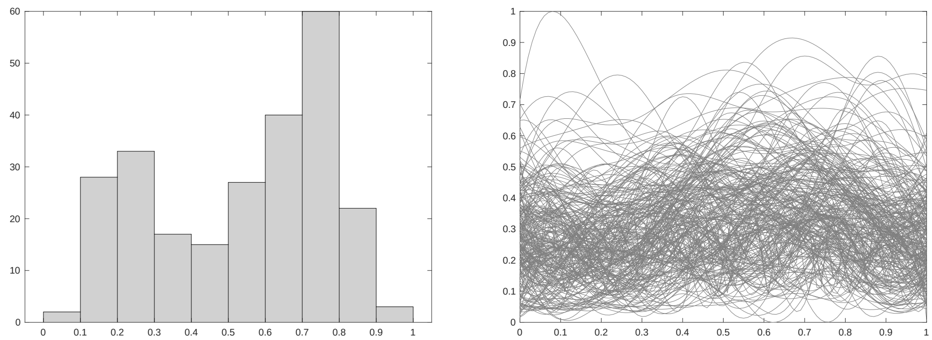


Figure 2: *Left penal: histogram of the daily bike rentals (scaled to the interval* $[0, 1]$*). Right panel: corresponding wind speed curves.*

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	-	$(0, 3.57]$	$[1.19, 3.82]$

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 the 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* $d_0 = \int_0^1 |\beta_0(s)|^2 ds$.

sharing companies to investigate the influence of environmental factors on the bike sharing. In this data example, we aim to use our proposed pivotal inference tools to investigate the impact of wind speed on bike rental activities on workdays. We used the bike-sharing data of Captial Bike Sharing (CBS) at Washington, D.C., USA in the year 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 dataset 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, the Y_i's 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 are of temporal dependence, such that the application of our pivotal inference approach which does not depend on long-run variances estimates, becomes 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 were normalized through a linear transformation so that the data take values in $[0, 1]$, and the curves X_i are obtained by projecting the hourly observations onto the space spanned by the first 7 Fourier basis functions on $[0, 1]$, and are evaluated at a equally spaced grid $t = 0.01, 0.02, \ldots, 1$. Figure 2 displays the histogram of Y_i together with 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 $\alpha = 0.10$, 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 one wants to avoid the specification of the threshold Δ for a test (see the discussion in Remark 2), one 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.

Acknowledgements

This work was partially supported by the DFG Research unit 5381 *Mathematical Statistics in the Information Age.* The authors would like to thank Tim Kutta for helpful discussions and sharing with us the R code for Kutta, Dierickx and Dette (2021). 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 preprint. arXiv:2109.01730.*

Brutsche, J. and Rohde, A. (2022). Sharp adaptive similarity testing with pathwise stability for ergodic diffusions. *arXiv preprint. 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. *Stat. 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. *Electronic Journal of Statistics*, **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 preprint arXiv:2207.11392.*

Dehling, H., Mikosch, T. and Sørensen, M. (2002). *Empirical process techniques for dependent data.* Birkhäuser.

van Delft, A. and Eichler, M. (2018). Locally stationary functional time series. *Electronic Journal of Statistics*, 12:107–170.

Dette, H., Kokot, K. and Aue, A. (2020). Functional data analysis in the Banach space of continuous functions. *Ann. Stat.* **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. Series. B. Stat. Methodol.* **82**, 629–660.

Dette, H and Tang, J. (2021) Statistical inference for function-on-function linear regression. *arXiv preprint.* 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 P. Vieu (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. Stat.* **23**, 761–778.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Stat.* **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. Stat.* **41**, 838–869.

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

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Stat.* **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*. New York: Wiley.

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

Kim, J. S., Staicu, A. M., Maity, A., Carroll, R. J. and Ruppert, D. (2018). Additive function-on-function regression. *Journal of Computational and Graphical Statistics*, **27**, 234–244.

Kokoszka, P., Rice, G. and Shang, H. L. (2017). Inference for the autocovariance of a functional time series under conditional heteroscedasticity. *Journal of Multivariate Analysis*, **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. (2021) Statistical inference for the slope parameter in functional linear regression. *arXiv preprint.* arXiv:2108.07098.

Lavergne, P. (2014). Model equivalence tests in a parametric framework. *Journal of Econometrics*, **178**, 414–425.

Lehmann, E. L. (1959). *Testing statistical hypotheses*. New York: Wiley.

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*, 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. *Biometrical Journal*, 49(1):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. Stat.* **33**, 774–805.

Romano, J. P. (2005). Optimal testing of equivalence hypotheses. *The Annals of Statistics*, **33**, 1036–1047.

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

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

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

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

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

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.

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. Stat.* **33**, 2873–2903.

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

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

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

E-mail: Jiajun.Tang@ruhr-uni-bochum.de

AN RKHS APPROACH FOR PIVOTAL INFERENCE IN FUNCTIONAL LINEAR REGRESSION

Holger Dette and Jiajun Tang

Supplementary Material

Section S1 contains the proofs of the theorems in our main article, and Section S2 contains the supporting lemmas.

S1. Theoretical details of main results

S1.1 Proof of Theorem 1

In the sequel, we use c to denote a generic positive constant that might differ from line to line. We first prove in Lemma 1 below the uniform convergence rate of the sequential RKHS estimator $\widehat{\beta}_{n,\lambda}(\cdot,\nu)$ for the slope function β_0 defined in (3.2) w.r.t. the $\|\cdot\|_K$-norm.

Lemma 1. *Under Assumptions A1–A5, we have, for any fixed (but arbitrary)* $\nu_0 \in (0,1]$,

$$\sup_{\nu\in[\nu_0,1]} \big\|\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0\big\|_K = O_p(\lambda^{1/2} + n^{-1/2}\lambda^{-(2a+1)/(4D)}) .$$

Proof. Define $S_{\lambda,\nu}(\beta) = \mathrm{E}\{S_{n,\lambda,\nu}(\beta)\}$ and $\mathcal{D}S_{\lambda,\nu}(\beta) = \mathrm{E}\{\mathcal{D}S_{n,\lambda,\nu}(\beta)\}$. In view of (4.7),

$$\begin{aligned} S_{\lambda,\nu}(\beta) &= \mathrm{E}\{S_{n,\lambda,\nu}(\beta)\} = -\mathrm{E}\bigg[\bigg\{Y_0 - \int_0^1 X_0(s)\,\beta(s)\,ds\bigg\}\tau_\lambda(X_0)\bigg] + W_\lambda(\beta)\,, \\ \mathcal{D}S_{\lambda,\nu}(\beta)\beta_1 &= \mathrm{E}\{\mathcal{D}S_{n,\lambda,\nu}(\beta)\beta_1\} = \mathrm{E}\bigg[\bigg\{\int_0^1 X_0(s)\,\beta(s)\,ds\bigg\}\tau_\lambda(X_0)\bigg] + W_\lambda(\beta)\,. \end{aligned} \tag{S1.1}$$

Recall from (4.7) that, for $\nu \in [0,1]$ and for any $\beta_1, \beta_2 \in \mathcal{H}$,

$$\langle \mathcal{D}S_{\lambda,\nu}(\beta)\beta_1, \beta_2\rangle_K = \mathrm{E}\big\{\langle \tau_\lambda(X_i), \beta_1\rangle_K\, \langle \tau_\lambda(X_i), \beta_2\rangle_K\big\} + \langle W_\lambda(\beta_1), \beta_2\rangle_K$$

$$= \mathrm{E}\big\{\langle X_i, \beta_1\rangle_{L^2}\, \langle X_i, \beta_2\rangle_{L^2}\big\} + \langle W_\lambda(\beta_1), \beta_2\rangle_K = V(\beta_1,\beta_2) + \lambda J(\beta_1,\beta_2) = \langle \beta_1, \beta_2\rangle_K = \langle id(\beta_1), \beta_2\rangle_K\,,$$

which implies that $\mathcal{D}S_{\lambda,\nu}(\beta) = id$ where id denotes the identity operator on $\mathcal{H}$. Since $\mathcal{D}^2 S_{\lambda,\nu}$ vanishes, there exists a unique solution to the estimating equation $S_{\lambda,\nu}(\beta) = 0$. In addition, by the mean value theorem, for any $\beta \in \mathcal{H}$, $S_{\lambda,\nu}(\beta) = S_{\lambda,\nu}(\beta_0) + \mathcal{D}S_{\lambda,\nu}(\beta)(\beta - \beta_0) = S_{\lambda,\nu}(\beta_0) + (\beta - \beta_0)$. Let $\beta_{\lambda,\nu} = \beta_0 - S_{\lambda,\nu}(\beta_0)$. We deduce that $S_{\lambda,\nu}(\beta_{\lambda,\nu}) = S_{\lambda,\nu}(\beta_0) + (\beta_{\lambda,\nu} - \beta_0) = 0$, so that $\beta_{\lambda,\nu}$ is the unique solution to the estimating equation $S_{\lambda,\nu}(\beta) = 0$. Moreover, since $\mathrm{E}\big[\big\{Y_0 - \int_0^1 X_0(s)\beta_0(s)ds\big\}\tau_\lambda(X_0)\big] = 0$, in view of (S1.1), for any $\nu \in [\nu_0, 1]$,

$$\|\beta_{\lambda,\nu} - \beta_0\|_K = \|S_{\lambda,\nu}(\beta_0)\|_K = \|W_\lambda(\beta_0)\|_K \,. \tag{S1.2}$$

Therefore, by the Cauchy-Schwarz inequality, we deduce that

$$\begin{aligned} &\sup_{\nu \in [\nu_0, 1]} \big\|\beta_{\lambda,\nu} - \beta_0\big\|_K = \|W_\lambda(\beta_0)\|_K = \sup_{\|\gamma\|_K = 1} |\langle W_\lambda(\beta_0), \gamma\rangle_K| = \sup_{\|\gamma\|_K = 1} \lambda |J(\beta_0, \gamma)| \\ &\le \sup_{\|\gamma\|_K = 1} \left\{\sqrt{\lambda J(\beta_0, \beta_0)}\sqrt{\lambda J(\gamma, \gamma)}\right\} \le \sup_{\|\gamma\|_K = 1} \left\{\sqrt{\lambda J(\beta_0, \beta_0)}\, \|\gamma\|_K\right\} = O(\lambda^{1/2})\,. \end{aligned} \tag{S1.3}$$

Since $\|\widehat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0\|_K \le \|\beta_{\lambda,\nu} - \beta_0\|_K + \|\widehat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_{\lambda,\nu}\|_K$, we then proceed to show the rate of $\|\widehat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_{\lambda,\nu}\|_K$. For $\nu \in [\nu_0, 1]$, let $F_{n,\nu}(\beta) = \beta - S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta)$. Observing that $\mathcal{D}S_{\lambda,\nu}(\beta) = id$, since $\mathcal{D}^2 S_{\lambda,\nu}$ vanishes, we obtain

$$F_{n,\nu}(\beta) = \mathcal{D}S_{\lambda,\nu}(\beta_{\lambda,\nu})\beta - S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta) = I_{1,n,\nu}(\beta) + I_{2,n,\nu}(\beta) - S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\,, \tag{S1.4}$$

where $\mathcal{D}S_{n,\lambda,\nu}$ is defined in (4.7) and

$$\begin{aligned} I_{1,n,\nu}(\beta) &= -\{S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta) - S_{n,\lambda,\nu}(\beta_{\lambda,\nu}) - \mathcal{D}S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\beta\}\,, \\ I_{2,n,\nu}(\beta) &= -\{\mathcal{D}S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\beta - \mathcal{D}S_{\lambda,\nu}(\beta_{\lambda,\nu})\beta\}\,. \end{aligned} \tag{S1.5}$$

First, for $I_{1,n,\nu}(\beta)$ in (S1.5), in view of $S_{n,\lambda,\nu}$ and $\mathcal{D}S_{n,\lambda,\nu}$ defined in (4.7), we find

$$I_{1,n,\nu}(\beta) = \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \left(\left[Y_i - \int_0^1 \{\beta_{\lambda,\nu}(s) + \beta(s)\}\, X_i(s)\, ds \right] \tau_\lambda(X_i) \right.$$

$$- \left\{ Y_i - \int_0^1 \beta(s)\, X_i(s)\, ds \right\} \tau_\lambda(X_i) + \left\{ \int_0^1 \beta_{\lambda,\nu}(s)\, X_i(s)\, ds \right\} \tau_\lambda(X_i) \Bigg) = 0\,. \quad \text{(S1.6)}$$

For the second term $I_{2,n,\nu}(\beta)$ in (S1.5), define the event $\mathcal{E}_n(c) = \big\{ \max_{1\le i\le n} \|X_i\|_{L^2} \le c \log n \big\}$. By Assumption A4 and Markov's inequality, if we take $c > 3/\varpi > 0$, we have $\mathrm{P}\{\mathcal{E}_n^{\mathrm{c}}(c)\} \le n\, \mathrm{P}\big(\|X_0\|_{L^2} \le c \log n\big) \le n^{1-c\varpi}\, \mathrm{E}\{\exp(\varpi \|X_0\|_{L^2})\} = o(n^{-2})$. Then, it suffices to confine the proof on the event $\mathcal{E}_n(c)$. In view of (4.7) and (S1.1),

$$I_{2,n,\nu}(\beta) = \mathcal{D}S_{n,\nu}(\beta_{\lambda,\nu})\beta - \mathcal{D}S_\nu(\beta_{\lambda,\nu})\beta = I_{2,1,n,\nu}(\beta) + I_{2,2,n,\nu}(\beta)\,, \quad \text{(S1.7)}$$

where

$$I_{2,1,n,\nu}(\beta) = \mathrm{E}\left[\tau_\lambda(X_i) \int_0^1 \beta(s) X_i(s) ds \times \mathbb{1}\{\mathcal{E}_n^{\mathrm{c}}(c)\} \right], \quad \text{(S1.8)}$$

$$I_{2,2,n,\nu}(\beta) = -\frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \left(\tau_\lambda(X_i) \int_0^1 \beta(s) X_i(s) ds\, \mathbb{1}\{\mathcal{E}_n(c)\} - \mathrm{E}\left[\tau_\lambda(X_i) \int_0^1 \beta(s) X_i(s) ds\, \mathbb{1}\{\mathcal{E}_n(c)\} \right] \right).$$

For the first term $I_{2,1,n,\nu}$ in (S1.7), by the Cauchy-Schwarz inequality and Lemma 4,

$$\begin{aligned}
&\|I_{2,1,n,\nu}(\beta)\|_K = \left\| \mathrm{E}\left[\tau_\lambda(X_0) \int_0^1 \beta(s) X_0(s) ds \mathbb{1}\{\mathcal{E}_n^{\mathrm{c}}(c)\} \right] \right\|_K \\
&\le \sup_{\|\gamma\|_K=1} \left\langle \gamma, \mathrm{E}\left[\tau_\lambda(X_0) \int_0^1 \beta(s) X_0(s) ds \mathbb{1}\{\mathcal{E}_n^{\mathrm{c}}(c)\} \right] \right\rangle_K \\
&= \sup_{\|\gamma\|_K=1} \mathrm{E}\left[\int_0^1 \gamma(s) X_0(s) ds \int_0^1 \beta(s) X_0(s) ds \mathbb{1}\{\mathcal{E}_n^{\mathrm{c}}(c)\} \right] \\
&\le \big[\mathrm{P}\{\mathcal{E}_n^{\mathrm{c}}(c)\}\big]^{1/2} \mathrm{E}\big(\langle X_0, \beta\rangle_{L^2}^4\big)^{1/4} \sup_{\|\gamma\|_K=1} \mathrm{E}\big(\langle X_0, \gamma\rangle_{L^2}^4\big)^{1/4} \\
&\le c\big[\mathrm{P}\{\mathcal{E}_n^{\mathrm{c}}(c)\}\big]^{1/2} \mathrm{E}\big(\langle X_0, \beta\rangle_{L^2}^2\big)^{1/2} \sup_{\|\gamma\|_K=1} \mathrm{E}\big(\langle X_0, \gamma\rangle_{L^2}^2\big)^{1/2} \\
&= c\big[\mathrm{P}\{\mathcal{E}_n^{\mathrm{c}}(c)\}\big]^{1/2} \mathrm{E}\big(\langle \tau_\lambda(X_0), \beta\rangle_K^2\big)^{1/2} \sup_{\|\gamma\|_K=1} \mathrm{E}\big(\langle \tau_\lambda(X_0), \gamma\rangle_K^2\big)^{1/2} \\
&\le c\big[\mathrm{P}\{\mathcal{E}_n^{\mathrm{c}}(c)\}\big]^{1/2} \mathrm{E}\|\tau_\lambda(X_0)\|_K^2 \|\beta\|_K \le o(n^{-1}\lambda^{-1/(2D)}) \|\beta\|_K = o(1)\|\beta\|_K. \quad \text{(S1.9)}
\end{aligned}$$

Therefore, we deduce that

$$\sup_{\nu\in[\nu_0,1]} \|I_{2,1,n,\nu}(\beta)\|_K = o(1)\, \|\beta\|_K\,. \quad \text{(S1.10)}$$

For the second term $I_{2,2,n,\nu}(\beta)$ in (S1.7), for a, D in Assumption A2 and c_K in Lemma 5 in Section S2, let $p_n = c_K^{-2}\lambda^{(2a+1)/(2D)-1}$. In order to apply Lemma 5 in Section S2, we shall rescale β such that the L^2-norm of its rescaled version is bounded by 1, that is

$$\widetilde{\beta} = \begin{cases} \left(c_K\lambda^{-(2a+1)/(4D)}\|\beta\|_K\right)^{-1}\beta & \text{if } \beta \neq 0\,, \\ 0 & \text{if } \beta = 0 \end{cases} \tag{S1.11}$$

where c_K is the constant in Lemma 5. We have $\|\widetilde{\beta}\|_{L^2} \leq c_K\lambda^{-(2a+1)/(4D)}\|\widetilde{\beta}\|_K \leq 1$, since $\|\widetilde{\beta}\|_K \leq (c_K\lambda^{-(2a+1)/(4D)})^{-1}$ in view of Lemma 5. In addition, observing (4.2), it follows that $J(\widetilde{\beta},\widetilde{\beta}) \leq \lambda^{-1}\|\widetilde{\beta}\|_K^2 \leq c_K^{-2}\lambda^{(2a+1)/(2D)-1} = p_n$. Therefore,

$$\widetilde{\beta} \in \mathcal{F}_{p_n} := \{\beta \in \mathcal{H} : \|\beta\|_{L^2} \leq 1, J(\beta,\beta) \leq p_n\}. \tag{S1.12}$$

For the event $\mathcal{E}_n(c)$ defined below equation (S1.6) and for any $\beta \in \mathcal{H}$, let

$$\widetilde{H}_{n,\nu}(\beta) = \frac{1}{\sqrt{\lfloor n\nu\rfloor}}\sum_{i=1}^{\lfloor n\nu\rfloor}\left(\tau_\lambda(X_i)\langle\beta, X_i\rangle_{L^2}\,\mathbb{1}\{\mathcal{E}_n(c)\} - \mathrm{E}\Big[\tau_\lambda(X_i)\langle\beta, X_i\rangle_{L^2}\,\mathbb{1}\{\mathcal{E}_n(c)\}\Big]\right). \tag{S1.13}$$

Note that, for $\nu \in [\nu_0, 1]$,

$$\sup_{\beta\in\mathcal{F}_{p_n}}\|\widetilde{H}_{n,\nu}(\beta)\|_K = \frac{1}{\sqrt{\lfloor n\nu\rfloor/n}}\sup_{\beta\in\mathcal{F}_{p_n}}\|H_{n,\lfloor n\nu\rfloor}(\beta)\|_K \leq c\nu_0^{-1/2}\max_{1\leq k\leq n}\sup_{\beta\in\mathcal{F}_{p_n}}\|H_{n,k}(\beta)\|_K\,, \tag{S1.14}$$

where, for the $\mathcal{E}_n(c)$ defined below equation (S1.6), $H_{n,k}$ is defined by

$$H_{n,k}(\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{k}\left(\tau_\lambda(X_i)\int_0^1\beta(s)X_i(s)ds\,\mathbb{1}\{\mathcal{E}_n(c)\} - \mathrm{E}\Big[\tau_\lambda(X_i)\int_0^1\beta(s)X_i(s)ds\,\mathbb{1}\{\mathcal{E}_n(c)\}\Big]\right). \tag{S1.15}$$

Therefore, observing that $n^{-1/2} = o(p_n^{1/(2m)})$ by Assumption A5, combining (S1.14) and Lemma 8 yields with probability tending to one,

$$\sup_{\nu\in[\nu_0,1]}\sup_{\widetilde{\beta}\in\mathcal{F}_{p_n}}\|\widetilde{H}_{n,\nu}(\widetilde{\beta})\|_K \leq c\left(p_n^{1/(2m)} + n^{-1/2}\right)\left(\lambda^{-1/(2D)}\log n\right)^{1/2} \leq c\,p_n^{1/(2m)}\lambda^{-1/(4D)}(\log n)^{1/2}\,,$$

where $c > 0$ depends on ν_0. In view of (S1.11), we deduce from the above equation that, for the β in (S1.11), with probability tending to one,

$$\sup_{\nu\in[\nu_0,1]}\|\widetilde{H}_{n,\nu}(\beta)\|_K \leq \left(c_K\lambda^{-(2a+1)/(4D)}\|\beta\|_K\right)\sup_{\nu\in[\nu_0,1]}\sup_{\widetilde{\beta}\in\mathcal{F}_{p_n}}\|\widetilde{H}_{n,\nu}(\widetilde{\beta})\|_K$$

$$\leq c\, p_n^{1/(2m)} \lambda^{-(a+1)/(2D)} (\log n)^{1/2} \|\beta\|_K \,.$$

Observing that $p_n = O(\lambda^{(2a+1)/(2D)-1})$ and (S1.8), we have with probability tending to one,

$$\sup_{\nu\in[\nu_0,1]} \|I_{2,2,n,\nu}(\beta)\|_K \leq n^{-1/2} \sup_{\nu\in[\nu_0,1]} \|\widetilde{H}_{n,\nu}(\beta)\|_K \leq c\, n^{-1/2} p_n^{1/(2m)} \lambda^{-(a+1)/(2D)} (\log n)^{1/2} \|\beta\|_K$$

$$\leq c\, n^{-1/2} \lambda^{-\varsigma} (\log n)^{1/2} \, \|\beta\|_K = o(1)\, \|\beta\|_K \,, \tag{S1.16}$$

where we used Assumption A5 in the last step. Therefore, combining (S1.7), (S1.10) and (S1.16) yields that, as $n \to \infty$,

$$\sup_{\nu\in[\nu_0,1]} \|I_{2,n,\nu}(\beta)\|_K = o(1)\, \|\beta\|_K \,. \tag{S1.17}$$

We now consider the term $-S_{n,\lambda,\nu}(\beta_{\lambda,\nu})$ in (S1.4). Recalling the definition of τ_λ in (3.8) and observing that $S_{\lambda,\nu}(\beta_{\lambda,\nu}) = 0$ and $\mathrm{E}\{\varepsilon_0 \tau_\lambda(X_0)\} = 0$, in view of (4.7), we find

$$- S_{n,\lambda,\nu}(\beta_{\lambda,\nu}) = -\{S_{n,\lambda,\nu}(\beta_{\lambda,\nu}) - S_{\lambda,\nu}(\beta_{\lambda,\nu})\} = \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \varepsilon_i\, \tau_\lambda(X_i) - I_{2,n,\nu}(\beta_0 - \beta_{\lambda,\nu}) \,,$$

where $I_{2,n,\nu}$ is defined in (S1.7). We deduce from the above equation and (S1.17) that

$$\sup_{\nu\in[\nu_0,1]} \|S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\|_K^2 \leq 2 \sup_{\nu\in[\nu_0,1]} \left\| \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \varepsilon_i \tau_\lambda(X_i) \right\|_K^2 + o(1)\, \|\beta_0 - \beta_{\lambda,\nu}\|_K^2 \,. \tag{S1.18}$$

For the first term in (S1.18), by direct calculations, we find

$$\left\| \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \tau_\lambda(X_i) \varepsilon_i \right\|_K^2 = \frac{1}{\lfloor n\nu \rfloor^2} \sum_{i_1=1}^{\lfloor n\nu \rfloor} \sum_{i_2=1}^{\lfloor n\nu \rfloor} \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \left\langle \frac{\langle \varepsilon_{i_1} X_{i_1}, \varphi_k \rangle_{L^2}}{1+\lambda\rho_k} \varphi_k, \frac{\langle \varepsilon_{i_2} X_{i_2}, \varphi_\ell \rangle_{L^2}}{1+\lambda\rho_\ell} \varphi_\ell \right\rangle_K$$

$$= \frac{1}{\lfloor n\nu \rfloor^2} \sum_{i_1=1}^{\lfloor n\nu \rfloor} \sum_{i_2=1}^{\lfloor n\nu \rfloor} \sum_{k=1}^{\infty} \frac{1}{1+\lambda\rho_k} \langle \varepsilon_{i_1} X_{i_1}, \varphi_k \rangle_{L^2} \langle \varepsilon_{i_2} X_{i_2}, \varphi_k \rangle_{L^2} = \sum_{k=1}^{\infty} \frac{1}{1+\lambda\rho_k} \left(\frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \langle \varepsilon_i X_i, \varphi_k \rangle_{L^2} \right)^2$$

$$= \frac{1}{\lfloor n\nu \rfloor} \sum_{k=1}^{\infty} \frac{1}{1+\lambda\rho_k} \left\langle \frac{1}{\sqrt{\lfloor n\nu \rfloor}} \sum_{i=1}^{\lfloor n\nu \rfloor} \varepsilon_i X_i, \varphi_k \right\rangle_{L^2}^2 \leq \frac{1}{\lfloor n\nu \rfloor} \left\| \frac{1}{\sqrt{\lfloor n\nu \rfloor}} \sum_{i=1}^{\lfloor n\nu \rfloor} \varepsilon_i X_i \right\|_{L^2}^2 \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{1+\lambda\rho_k} \,. \tag{S1.19}$$

Denote the long-run covariance function $C_{X\varepsilon}(s,t) = \sum_{\ell=-\infty}^{+\infty} \mathrm{cov}\{\varepsilon_0 X_0(s), \varepsilon_\ell X_\ell(t)\}$. Observing Lemmas 6 and 7, we have that $C_{X\varepsilon} \in L^2([0,1]^2)$ and $\int_0^1 C_{X\varepsilon}(s,s) ds < \infty$. By Assumption A1, we have that $C_{X\varepsilon}$ is positive definite. Let $\{\breve{\zeta}_j\}_{j=1}^{\infty}$ and $\{\breve{\psi}_j\}_{j=1}^{\infty}$ denote the eigenvalues and the corresponding eigenfunctions of the covariance kernel $C_{X\varepsilon}$, such that $\sum_{j=1}^{\infty} \breve{\zeta}_j < \infty$. In addition, since the X_i's and the ε_i's are independent, it is easy to see that the series $\{\varepsilon_i X_i\}_{i\in\mathbb{Z}}$ is m-approximable by $\{\varepsilon_{i,\ell} X_{i,\ell}\}_{i,\ell\in\mathbb{Z}}$. By Theorem 1.1 in Berkes, Horváth and Rice (2013), there exists a Gaussian process $\{\Gamma_{X\varepsilon}(s,\nu)\}_{s\in[0,1],\nu\in[0,1]}$ in $\mathcal{F}$ defined in (4.11), given by $\Gamma_{X\varepsilon}(s,\nu) = \sum_{j=1}^{\infty} \breve{\zeta}_j^{1/2}\, W_j(\nu)\, \breve{\psi}_j(s)$, such that $\sup_{\nu\in[0,1]} \big\|\Gamma_{X\varepsilon}(\cdot,\nu) - n^{-1/2}\sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i X_i\big\|_{L^2}^2 = o_p(1)$. Here, $\{W_j\}_{j=1}^{\infty}$ is a series of i.i.d. Wiener processes. Note that $\mathrm{E}\big\{\sup_{\nu\in[0,1]} W_j^2(\nu)\big\} < \infty$, so that $\mathrm{E}\big\{\sup_{\nu\in[0,1]} \|\Gamma_{X\varepsilon}(\cdot,\nu)\|_{L^2}^2\big\} \le \sum_{j=1}^{\infty} \breve{\zeta}_j\, \mathrm{E}\big\{\sup_{\nu\in[0,1]} W_j^2(\nu)\big\} < \infty$. Therefore, in view of (S1.19), we deduce from the above finding that

$$\begin{aligned}
&\sup_{\nu\in[\nu_0,1]} \left\| \frac{1}{\lfloor n\nu\rfloor} \sum_{i=1}^{\lfloor n\nu\rfloor} \tau_\lambda(X_i)\varepsilon_i \right\|_K^2 \le \sup_{\nu\in[\nu_0,1]} \left\{ \frac{1}{\lfloor n\nu\rfloor} \left\| \frac{1}{\sqrt{\lfloor n\nu\rfloor}} \sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i X_i \right\|_{L^2}^2 \right\} \times \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{1+\lambda\rho_k} \\
&\le n^{-1}\nu_0^{-2} \sup_{\nu\in[0,1]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i X_i \right\|_{L^2}^2 \times \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{1+\lambda\rho_k} \times \{1+o_p(1)\} \\
&\le n^{-1}\nu_0^{-2} \left\{ \sup_{\nu\in[0,1]} \left\| \Gamma_{X\varepsilon}(\cdot,\nu) - \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i X_i \right\|_{L^2}^2 + \sup_{\nu\in[0,1]} \|\Gamma_{X\varepsilon}(\cdot,\nu)\|_{L^2}^2 \right\} \\
&\quad \times \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{1+\lambda\rho_k} \times \{1+o_p(1)\} = O_p(n^{-1}\lambda^{-(2a+1)/(2D)}) . \qquad \text{(S1.20)}
\end{aligned}$$

Consequently, combining the above finding and (S1.3) and (S1.18), we obtain that

$$\sup_{\nu\in[\nu_0,1]} \|S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\|_K = O_p\big(n^{-1/2}\lambda^{-(2a+1)/(4D)}\big) + o(\lambda^{1/2}) . \qquad \text{(S1.21)}$$

Next, let $q_n = c(n^{-1/2}\lambda^{-(2a+1)/(4D)} + \lambda^{1/2})$ and denote by $\mathcal{B}(r) = \{\gamma \in \mathcal{H}, \|\gamma\|_K \le r\}$ denote the $\|\cdot\|_K$-ball with radius $r > 0$ in $\mathcal{H}$. In view of (S1.16), for any $\beta \in \mathcal{B}(q_n)$, with probability tending to one, $\|I_{2,n,\nu}(\beta)\|_K \le \|\beta\|_K/2 \le q_n/2$. Therefore, in view of (S1.6), (S1.17) and (S1.21), for $F_n(\beta)$ defined in (S1.4), with probability tending to one, for any $\beta \in \mathcal{B}(q_n)$, $\sup_{\nu\in[\nu_0,1]} \|F_{n,\nu}(\beta)\|_K \le \sup_{\nu\in[\nu_0,1]} \|I_{2,n,\nu}(\beta)\|_K + \sup_{\nu\in[\nu_0,1]} \|S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\|_K \le c\, n^{-1/2}\lambda^{-(2a+1)/(4D)} + q_n/2 \le q_n$, which implies that $F_{n,\nu}\{\mathcal{B}(q_n)\} \subset \mathcal{B}(q_n)$ uniformly in $\nu \in [\nu_0,1]$. Observing (S1.4)–(S1.6), we have, for any $\beta_1,\beta_2 \in \mathcal{B}(q_n)$, $F_{n,\nu}(\beta_1) - F_{n,\nu}(\beta_2) =$

$I_{2,n,\nu}(\beta_1) - I_{2,n,\nu}(\beta_2)$. Due to (S1.17), with probability tending to one,

$$\sup_{\nu\in[\nu_0,1]} \|F_{n,\nu}(\beta_1) - F_{n,\nu}(\beta_2)\|_K = \sup_{\nu\in[\nu_0,1]} \|I_{2,n,\nu}(\beta_1) - I_{2,n,\nu}(\beta_2)\|_K \le \|\beta_1 - \beta_2\|_K/2\,,$$

which implies that $F_{n,\nu}$ is a contraction mapping on $\mathcal{B}(q_n)$ uniformly in $\nu \in [\nu_0, 1]$. By the Banach contraction mapping theorem, there exists a unique element $\beta^*_\nu \in \mathcal{B}_n$ such that $\beta^*_\nu = F_{n,\nu}(\beta^*_\nu) = \beta^*_\nu - S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta^*_\nu)$. Letting $\widehat{\beta}_{n,\lambda}(\cdot,\nu) = \beta_{\lambda,\nu} + \beta^*_\nu$, we have $S_{n,\lambda,\nu}\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)\} = 0$, which implies that $\widehat{\beta}_{n,\lambda}(\cdot,\nu)$ is the estimator defined by (3.2). Moreover, we have, with probability tending to one, $\sup_{\nu\in[\nu_0,1]} \|\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_{\lambda,\nu}\|_K = \sup_{\nu\in[\nu_0,1]} \|\beta^*_\nu\|_K \le q_n$. In view of (S1.3), we conclude

$$\begin{aligned}\sup_{\nu\in[\nu_0,1]} \|\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0\|_K &\le \sup_{\nu\in[\nu_0,1]} \|\beta_{\lambda,\nu} - \beta_0\|_K + \sup_{\nu\in[\nu_0,1]} \|\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_{\lambda,\nu}\|_K \\ &= O_p(\lambda^{1/2} + q_n) = O_p\big(\lambda^{1/2} + n^{-1/2}\lambda^{-(2a+1)/(4D)}\big)\,. \qquad \square\end{aligned}$$

Proof of Theorem 1. Let

$$\begin{aligned} S_{n,\nu}(\beta) &= -\frac{1}{\lfloor n\nu\rfloor}\sum_{i=1}^{\lfloor n\nu\rfloor}\left\{Y_i - \int_0^1 X_i(s)\,\beta(s)\,ds\right\}\tau_\lambda(X_i)\,, \\ S_\nu(\beta) &= -\mathrm{E}\left[\left\{Y_0 - \int_0^1 X_0(s)\,\beta(s)\,ds\right\}\tau_\lambda(X_0)\right]\,, \end{aligned} \tag{S1.22}$$

and $\Delta_\nu\beta = \widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + W_\lambda(\beta_0)$ for the sake of notational convenience. Since $\mathcal{D}^2 S_{\lambda,\nu}$ vanishes and $\mathcal{D}S_{\lambda,\nu}(\beta_0) = id$, we have $S_{\lambda,\nu}\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)\} - S_{\lambda,\nu}(\beta_0) = \mathcal{D}S_{\lambda,\nu}(\beta_0)\Delta_\nu\beta = \Delta_\nu\beta$. Since $S_{n,\lambda,\nu}\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)\} = 0$, we deduce from this equation that

$$\begin{aligned} &\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + S_{n,\lambda,\nu}(\beta_0) = \Delta_\nu\beta + S_{n,\lambda,\nu}(\beta_0) \\ &= -S_{n,\nu}\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)\} + S_{n,\nu}(\beta_0) + S_\nu\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)\} - S_\nu(\beta_0)\,, \end{aligned} \tag{S1.23}$$

where $S_{n,\nu}$ and S_ν are defined in (S1.22). Let $r_n = \lambda^{1/2} + n^{-1/2}\lambda^{-(2a+1)/(4D)}$. For $c_1 > 0$, consider the event $\mathcal{M}_n = \big\{\sup_{\nu\in[\nu_0,1]} \|\Delta_\nu\beta\|_K \le c_1 r_n\big\}$. By Lemma 1, we obtain that $\mathrm{P}(\mathcal{M}_n)$

tends to one if the constant $c_1 > 0$ is chosen sufficiently large. For $c_K > 0$ in Lemma 5, let $q_n = c_1 c_K \lambda^{-(2a+1)/(4D)} r_n$ and let $p_n = c_1^2\, q_n^{-2} \lambda^{-1} r_n^2 = c_K^{-2}\, \lambda^{(-2D+2a+1)/(2D)}$. Note that $p_n \geq 1$ for n large enough. In order to apply Lemma 8, we shall rescale $\Delta_\nu \beta$ such that the L^2-norm of its rescaled version is bounded by 1. Let $\widetilde{\Delta}_\nu \beta = q_n^{-1} \Delta_\nu \beta$. By Lemma 5, we have that, on the event $\mathcal{M}_n$,

$$\|\widetilde{\Delta}_\nu \beta\|_{L^2} \leq c_K \lambda^{-(2a+1)/(4D)} \|\widetilde{\Delta}_\nu \beta\|_K \leq c_K q_n^{-1} \lambda^{-(2a+1)/(4D)} \|\Delta_\nu \beta\|_K \leq c_1 c_K q_n^{-1} \lambda^{-(2a+1)/(4D)} r_n \leq 1\,.$$

In addition, since $J(\Delta_\nu \beta, \Delta_\nu \beta) \leq \lambda^{-1} \|\Delta_\nu \beta\|_K^2$, we have

$$J(\widetilde{\Delta}_\nu \beta, \widetilde{\Delta}_\nu \beta) \leq q_n^{-2} J(\Delta_\nu \beta, \Delta_\nu \beta) \leq q_n^{-2} \lambda^{-1} \|\Delta_\nu \beta\|_K^2 \leq c_1^2\, q_n^{-2} \lambda^{-1} r_n^2 = p_n\,.$$

Hence, we have shown that $\widetilde{\Delta}_\nu \beta \in \mathcal{F}_{p_n}$, where $\mathcal{F}_{p_n}$ is defined in (S1.12).

Recall from (S1.13) that, for the event $\mathcal{E}_n(c)$ defined below equation (S1.6), for any $\beta \in \mathcal{H}$,

$$\widetilde{H}_{n,\nu}(\beta) = \frac{1}{\sqrt{\lfloor n\nu \rfloor}} \sum_{i=1}^{\lfloor n\nu \rfloor} \Big(\tau_\lambda(X_i) \langle \beta, X_i \rangle_{L^2}\, \mathbb{1}\{\mathcal{E}_n(c)\} - \mathrm{E}\Big[\tau_\lambda(X_i) \langle \beta, X_i \rangle_{L^2}\, \mathbb{1}\{\mathcal{E}_n(c)\} \Big] \Big)\,.$$

In view of (S1.22) and (S1.23) we thus obtain on the event $\mathcal{E}_n(c)$ that

$$\begin{aligned}
\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + S_{n,\lambda,\nu}(\beta_0) &= -S_{n,\nu}\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)\} + S_{n,\nu}(\beta_0) + S\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)\} - S_\nu(\beta_0) \\
&= \frac{1}{\lfloor n\nu \rfloor} \sum_{i=1}^{\lfloor n\nu \rfloor} \left[\tau_\lambda(X_i) \int_0^1 X_i(s)\, \Delta_\nu \beta(s)\, ds - \mathrm{E}\left\{ \tau_\lambda(X_i) \int_0^1 X_i(s)\, \Delta_\nu \beta(s)\, ds \right\} \right] \\
&\leq \frac{1}{\sqrt{\lfloor n\nu \rfloor}}\, \widetilde{H}_{n,\nu}(\Delta_\nu \beta) - \mathrm{E}\left\{ \tau_\lambda(X_i) \int_0^1 X_i(s)\, \Delta_\nu \beta(s)\, ds\, \mathbb{1}\{\mathcal{E}(c)^{\mathrm{c}}\} \right\}\,. \qquad \text{(S1.24)}
\end{aligned}$$

Note that following arguments similar to the ones used in (S1.9), we deduce that

$$\left\| \mathrm{E}\left[\mathbb{1}\{\mathcal{E}(c)^{\mathrm{c}}\} \tau_\lambda(X_i) \int_0^1 X_i(s) \Delta_\nu \beta(s) ds \right] \right\|_K \leq o(n^{-1} \lambda^{-1/(2D)}) \|\Delta_\nu \beta\|_K = o(v_n)\,. \qquad \text{(S1.25)}$$

Since $\widetilde{\Delta}_\nu \beta \in \mathcal{F}_{p_n}$, by applying Lemma 8, observing (S1.14), we deduce that

$$\sup_{\nu \in [\nu_0, 1]} \sup_{\widetilde{\Delta}_\nu \beta \in \mathcal{F}_{p_n}} \|\widetilde{H}_{n,\nu}(\widetilde{\Delta}_\nu \beta)\|_K = O_p\big\{ \big(p_n^{1/(2m)} + n^{-1/2} \big) \big(\lambda^{-1/(2D)} \log n \big)^{1/2} \big\} = O_p\big\{ p_n^{1/(2m)} \lambda^{-1/(4D)} (\log n)^{1/2} \big\}\,.$$

Consequently, for the $\Delta_\nu\beta$ in (S1.24), it follows with probability tending to one,

$$n^{-1/2}\sup_{\nu\in[\nu_0,1]}\|\widetilde{H}_{n,\nu}(\Delta_\nu\beta)\|_K \le n^{-1/2}q_n \sup_{\nu\in[\nu_0,1]}\sup_{\widetilde{\Delta}_\nu\beta\in\mathcal{F}_{p_n}}\|\widetilde{H}_{n,\nu}(\widetilde{\Delta}_\nu\beta)\|_K$$

$$\le cn^{-1/2}q_n\, p_n^{1/(2m)}\lambda^{-1/(4D)}(\log n)^{1/2} \le cn^{-1/2}(\lambda^{-(2a+1)/(4D)}r_n)\lambda^{(-2D+2a+1)/(4Dm)}\lambda^{-1/(4D)}(\log n)^{1/2}$$

$$= c\,n^{-1/2}\lambda^{-\varsigma}\big(\lambda^{1/2}+n^{-1/2}\lambda^{-(2a+1)/(4D)}\big)(\log n)^{1/2}\,,$$

for the constant $\varsigma > 0$ in Assumption A5. Combining this with (S1.24) and (S1.25) yields

$$\sup_{\nu\in[\nu_0,1]}\big\|\widehat{\beta}_{n,\lambda}(\cdot,\nu)-\beta_0+W_\lambda(\beta_0)+S_{n,\lambda,\nu}(\beta_0)\big\|_K$$
$$= O_p\big\{n^{-1/2}\lambda^{-\varsigma}(\lambda^{1/2}+n^{-1/2}\lambda^{-(2a+1)/(4D)})(\log n)^{1/2}\big\} = O_p(v_n)\,. \quad \text{(S1.26)}$$

Observing $S_{n,\lambda,\nu}(\beta_0)$ defined in (4.7), we therefore deduce from the above equation that

$$\sup_{\nu\in[\nu_0,1]}\left\|\nu\{\widehat{\beta}_{n,\lambda}(\cdot,\nu)-\beta_0+W_\lambda(\beta_0)\} - \frac{1}{n}\sum_{i=1}^{\lfloor n\nu\rfloor}\varepsilon_i\tau_\lambda(X_i)\right\|_K$$
$$\le \sup_{\nu\in[\nu_0,1]}\Big\{\nu\big\|\widehat{\beta}_{n,\lambda}(\cdot,\nu)-\beta_0+W_\lambda(\beta_0)+S_{n,\lambda,\nu}(\beta_0)\big\|_K\Big\}$$
$$+\sup_{\nu\in[\nu_0,1]}\left\{\left|\frac{\nu}{\lfloor n\nu\rfloor}-\frac{1}{n}\right|\times\left\|\sum_{i=1}^{\lfloor n\nu\rfloor}\varepsilon_i\,\tau_\lambda(X_i)\right\|_K\right\}+\sup_{\nu\in[\nu_0,1]}\Big\{\nu\|W_\lambda(\beta_0)\|_K\Big\} \quad \text{(S1.27)}$$
$$\le \sup_{\nu\in[\nu_0,1]}\big\|\widehat{\beta}_{n,\lambda}(\cdot,\nu)-\beta_0+W_\lambda(\beta_0)+S_{n,\lambda,\nu}(\beta_0)\big\|_K + n^{-1}\sup_{\nu\in[\nu_0,1]}\left\|\frac{1}{\lfloor n\nu\rfloor}\sum_{i=1}^{\lfloor n\nu\rfloor}\varepsilon_i\,\tau_\lambda(X_i)\right\|_K+\|W_\lambda(\beta_0)\|_K\,.$$

Observing (S1.20) and Assumption A5 we find

$$n^{-1}\sup_{\nu\in[\nu_0,1]}\left\|\frac{1}{\lfloor n\nu\rfloor}\sum_{i=1}^{\lfloor n\nu\rfloor}\varepsilon_i\,\tau_\lambda(X_i)\right\|_K = O_p(n^{-3/2}\lambda^{-(2a+1)/(4D)}) = o_p(v_n)\,.$$

The proof is therefore complete by combining the above equation with (S1.26) and (S1.27).

S1.2 Proof of Theorem 2

The proof is now performed in two steps. First, in Lemma 2, we will show that the U_i's are L^2-m-approximable (see Assumptions (1.1)–(1.4) in Berkes, Horváth and Rice, 2013). Second, in Lemma 3 we will show that $C_{U,\lambda}$ defined in (3.7) satsifies $\sup_{\lambda>0}\int_0^1\int_0^1 |C_{U,\lambda}(s,t)|dsdt < \infty$. Then, Theorem 2 is proved by the arguments as given in the proof of Theorem 1.1 in Berkes, Horváth and Rice (2013), which are omitted for the sake of brevity.

Lemma 2. *Under the assumptions of Theorem 2, the series $\{U_i\}_{i\in\mathbb{Z}}$ defined in (4.10) is L^2-m-approximable w.r.t. the series $\{U_{i,\ell}\}_{i,\ell\in\mathbb{Z}}$ uniformly in $\lambda > 0$, where*

$$U_{i,\ell} = \lambda^{(2a+1)/(2D)}\,\varepsilon_{i,\ell}\,\tau_\lambda(X_{i,\ell}) = \lambda^{(2a+1)/(2D)}\varepsilon_{i,\ell}\sum_{k=1}^{\infty}\frac{\langle X_{i,\ell},\varphi_k\rangle_{L^2}}{1+\lambda\rho_k}\varphi_k\,, \quad (i,\ell\in\mathbb{Z})\,.$$

Proof. By Lemmas 4 and 5 in Section S2, there exists a constant $c > 0$ such that

$$\|\tau_\lambda(X_i)\|_{L^2} \le c\lambda^{-(2a+1)/(4D)}\|\tau_\lambda(X_i)\|_K \le c\lambda^{-(2a+1)/(2D)}\|X_i\|_{L^2}\,. \tag{S1.28}$$

This together with the fact that $\|U_i\|_{L^2} \le \lambda^{(2a+1)/(2D)}|\varepsilon_i|\cdot\|\tau_\lambda(X_i)\|_{L^2}$ implies that $U_i \in L^2([0,1])$ uniformly in $\lambda > 0$. In addition, $\mathrm{E}(U_i)\equiv 0$, and, by (S1.28), for any $\delta\in(0,1)$,

$$\mathrm{E}\|U_i\|_{L^2}^{2+\delta} \le \lambda^{(2a+1)(2+\delta)/(2D)}\mathrm{E}|\varepsilon_i|^{2+\delta}\,\mathrm{E}\|\tau_\lambda(X_i)\|_{L^2}^{2+\delta} \le c\,\lambda^{(2a+1)(1+\delta)/(2D)}\mathrm{E}|\varepsilon_i|^{2+\delta}\,\mathrm{E}\|X_i\|_{L^2}^{2+\delta} < \infty\,,$$

where in the last step we have used Assumptions A4.1 and A3.2. Moreover, note that m-approximable series are strictly stationary (see, for example, Hörmann and Kokoszka, 2010). Hence, by applying Assumption A3 and (S1.28), we find that, uniformly in $\lambda > 0$,

$$\begin{aligned}
&\sum_{\ell=1}^{\infty}\left(\mathrm{E}\|U_i-U_{i,\ell}\|_{L^2}^{2+\delta}\right)^{1/\kappa} = \sum_{\ell=1}^{\infty}\left\{\lambda^{(2a+1)(2+\delta)/(2D)}\,\mathrm{E}\big\|\varepsilon_i\,\tau_\lambda(X_i)-\varepsilon_{i,\ell}\,\tau_\lambda(X_{i,\ell})\big\|_{L^2}^{2+\delta}\right\}^{1/\kappa}\\
&\le 2^{(1+\delta)/\kappa}\sum_{\ell=1}^{\infty}\left\{\lambda^{(2a+1)(2+\delta)/(2D)}\times\mathrm{E}|\varepsilon_i-\varepsilon_{i,\ell}|^{2+\delta}\times\mathrm{E}\|\tau_\lambda(X_i)\|_{L^2}^{2+\delta}\right\}^{1/\kappa}\\
&\quad + 2^{(1+\delta)/\kappa}\sum_{\ell=1}^{\infty}\left\{\lambda^{(2a+1)(2+\delta)/(2D)}\times\mathrm{E}|\varepsilon_{i,\ell}|^{2+\delta}\times\mathrm{E}\|\tau_\lambda(X_i)-\tau_\lambda(X_{i,\ell})\|_{L^2}^{2+\delta}\right\}^{1/\kappa}\\
&= 2^{(1+\delta)/\kappa}\left\{\lambda^{(2a+1)(2+\delta)/(2D)}\mathrm{E}\|\tau_\lambda(X_i)\|_{L^2}^{2+\delta}\right\}^{1/\kappa}\sum_{\ell=1}^{\infty}\left(\mathrm{E}|\varepsilon_i-\varepsilon_{i,\ell}|^{2+\delta}\right)^{1/\kappa}\\
&\quad + 2^{(1+\delta)/\kappa}\left(\mathrm{E}|\varepsilon_0|^{2+\delta}\right)^{1/\kappa}\times\left\{\lambda^{(2a+1)(2+\delta)/(2D)}\sum_{\ell=1}^{\infty}\mathrm{E}\|\tau_\lambda(X_i-X_{i,\ell})\|_{L^2}^{2+\delta}\right\}^{1/\kappa}
\end{aligned}$$

$$\le 2^{(1+\delta)/\kappa}\Big(\mathrm{E}\|X_i\|_{L^2}^{2+\delta}\Big)^{1/\kappa}\sum_{\ell=1}^{\infty}\big(\mathrm{E}|\varepsilon_i-\varepsilon_{i,\ell}|^{2+\delta}\big)^{1/\kappa}+2^{(1+\delta)/\kappa}\big(\mathrm{E}|\varepsilon_0|^{2+\delta}\big)^{1/\kappa}\Big(\sum_{\ell=1}^{\infty}\mathrm{E}\|X_i-X_{i,\ell}\|_{L^2}^{2+\delta}\Big)^{1/\kappa}<\infty\,.$$

Now, we have shown that the series $\{U_i\}_{i\in\mathbb{Z}}$ is L^2-m-approximable uniformly in $\lambda > 0$. □

Lemma 3. *Under the assumptions of Theorem 2, we have* $\sup_{\lambda>0}\int_0^1\int_0^1\{C_{U,\lambda}(s,t)\}^2\,ds\,dt < \infty$*, where* $C_{U,\lambda}$ *is defined in* (3.7).

Proof. Note that by Assumption A3, for $\ell \ge 1$, $\varepsilon_{0,\ell}$ and $\varepsilon_{-\ell}$ are independent; $\tau_\lambda(X_{0,\ell})$ and $\tau_\lambda(X_{-\ell})$ are independent. Note that $\mathrm{E}(\varepsilon_\ell) = 0$ for any $\ell \in \mathbb{Z}$. Hence we deduce that, for $\ell \ge 1$, $\mathrm{E}(\varepsilon_{0,\ell}\varepsilon_{-\ell}) = 0$, so that $\mathrm{E}(\varepsilon_0\varepsilon_{-\ell}) = \mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\} + \mathrm{E}(\varepsilon_{0,\ell}\varepsilon_{-\ell}) = \mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}$. In addition, for $\ell \ge 1$, we have

$$\begin{aligned}\mathrm{E}\{\tau_\lambda(X_0)(s)\tau_\lambda(X_{-\ell})(t)\} &= \mathrm{E}\big[\{\tau_\lambda(X_0)(s)-\tau_\lambda(X_{0,\ell})(s)\}\tau_\lambda(X_{-\ell})(t)\big]+\mathrm{E}\{\tau_\lambda(X_{0,\ell})(s)\tau_\lambda(X_{-\ell})(t)\}\\ &= \mathrm{E}\big[\{\tau_\lambda(X_0)(s)-\tau_\lambda(X_{0,\ell})(s)\}\tau_\lambda(X_{-\ell})(t)\big]+\mathrm{E}\{\tau_\lambda(X_0)(s)\}\mathrm{E}\{\tau_\lambda(X_0)(t)\}.\end{aligned}$$

Since $\mathrm{E}\{\varepsilon_\ell\,\tau(X_\ell)\}\equiv 0$, this equation implies that, for $\ell \ge 1$,

$$\begin{aligned}\mathrm{cov}\big\{\varepsilon_0\tau_\lambda(X_0)(s),\varepsilon_{-\ell}\tau_\lambda(X_{-\ell})(t)\big\} &= \mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\mathrm{E}\big[\big\{\tau_\lambda(X_0)(s)-\tau_\lambda(X_{0,\ell})(s)\big\}\tau_\lambda(X_{-\ell})(t)\big]\\ &\quad+\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\mathrm{E}\big\{\tau_\lambda(X_0)(s)\big\}\times\mathrm{E}\big\{\tau_\lambda(X_0)(t)\big\}.\end{aligned}$$

Therefore, we deduce from the above equation that

$$\begin{aligned}\int_0^1\int_0^1\{C_{U,\lambda}(s,t)\}^2dsdt &= \lambda^{2(2a+1)/D}\int_0^1\int_0^1\Big[\mathrm{cov}\big\{\varepsilon_0\,\tau_\lambda(X_0)(s)\,,\varepsilon_0\,\tau_\lambda(X_0)(t)\big\}\\ &\quad+2\sum_{\ell=1}^{+\infty}\mathrm{cov}\big\{\varepsilon_0\,\tau_\lambda(X_0)(s)\,,\varepsilon_{-\ell}\,\tau_\lambda(X_{-\ell})(t)\big\}\Big]^2dsdt\le 3I_1+12I_2+12I_3\,,\end{aligned}\tag{S1.29}$$

where

$$I_1=\lambda^{2(2a+1)/D}\{\mathrm{E}(\varepsilon_0^2)\}^2\int_{[0,1]^2}\Big[\mathrm{E}\big\{\tau_\lambda(X_0)(s)\times\tau_\lambda(X_0)(t)\big\}\Big]^2dsdt\,,$$

$$I_2=\lambda^{2(2a+1)/D}\int_{[0,1]^2}\left(\sum_{\ell=1}^{+\infty}\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\,\mathrm{E}\Big[\big\{\tau_\lambda(X_0)(s)-\tau_\lambda(X_{0,\ell})(s)\big\}\tau_\lambda(X_{-\ell})(t)\Big]\right)^2dsdt\,,$$

$$I_3 = \lambda^{2(2a+1)/D}\left[\sum_{\ell=1}^{+\infty} \mathrm{E}\big\{(\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\right]^2 \int_{[0,1]^2} \Big[\mathrm{E}\big\{\tau_\lambda(X_0)(s)\big\} \times \mathrm{E}\big\{\tau_\lambda(X_0)(t)\big\}\Big]^2 dsdt\,.$$

For the first term I_1, note that

$$\begin{aligned}
&\mathrm{E}\big\{\tau_\lambda(X_0)(s) \times \tau_\lambda(X_0)(t)\big\} = \mathrm{E}\left[\left\{\sum_{k_1=1}^{\infty} \frac{\langle X_0, \varphi_{k_1}\rangle_{L^2}}{1+\lambda\rho_{k_1}}\varphi_{k_1}(s)\right\}\left\{\sum_{k_2=1}^{\infty} \frac{\langle X_0, \varphi_{k_2}\rangle_{L^2}}{1+\lambda\rho_{k_2}}\varphi_{k_2}(t)\right\}\right] \\
&= \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} \frac{\varphi_{k_1}(s)\varphi_{k_2}(t)}{(1+\lambda\rho_{k_1})(1+\lambda\rho_{k_2})} \mathrm{E}\Big(\langle X_0, \varphi_{k_1}\rangle_{L^2}\langle X_0, \varphi_{k_2}\rangle_{L^2}\Big) \\
&= \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} \frac{\varphi_{k_1}(s)\varphi_{k_2}(t)}{(1+\lambda\rho_{k_1})(1+\lambda\rho_{k_2})} \int_0^1\int_0^1 C_X(t_1,t_2)\varphi_{k_1}(t_1)\varphi_{k_2}(t_2)dt_1dt_2 = \sum_{k=1}^{\infty} \frac{\varphi_k(s)\varphi_k(t)}{(1+\lambda\rho_k)^2}\,.
\end{aligned}$$

Hence, by the Cauchy-Schwarz inequality, we find

$$\begin{aligned}
I_1 &= \lambda^{2(2a+1)/D}\{\mathrm{E}(\varepsilon_0^2)\}^2 \int_0^1\int_0^1 \left\{\sum_{k=1}^{\infty} \frac{\varphi_k(s)\varphi_k(t)}{(1+\lambda\rho_k)^2}\right\}^2 dsdt \\
&\le \lambda^{2(2a+1)/D}\{\mathrm{E}(\varepsilon_0^2)\}^2\left\{\sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{(1+\lambda\rho_k)^2}\right\}^2 = O(\lambda^{(2a+1)/D})\,.
\end{aligned}$$

For the second term I_2, by a direct application of the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\left(\sum_{\ell=1}^{+\infty} \mathrm{E}\big\{(\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\,\mathrm{E}\Big[\big\{\tau_\lambda(X_0)(s) - \tau_\lambda(X_{0,\ell})(s)\big\} \times \tau_\lambda(X_{-\ell})(t)\Big]\right)^2 \\
&\le \mathrm{E}(\varepsilon_0)^2\left\{\sum_{\ell=1}^{+\infty} \mathrm{E}(\varepsilon_0 - \varepsilon_{0,\ell})^2\right\}\left[\sum_{\ell=1}^{+\infty} \mathrm{E}\big\{\tau_\lambda(X_0)(s) - \tau_\lambda(X_{0,\ell})(s)\big\}^2\right]\mathrm{E}\{\tau_\lambda(X_0)(t)\}^2\,.
\end{aligned}$$

Therefore, observing (S1.28) and Assumption A3, we deduce from the above equation that

$$\begin{aligned}
I_2 &\le \lambda^{2(2a+1)/D}\,\mathrm{E}(\varepsilon_0)^2\left\{\sum_{\ell=1}^{+\infty} \mathrm{E}(\varepsilon_0 - \varepsilon_{0,\ell})^2\right\}\int_0^1\left[\sum_{\ell=1}^{+\infty} \mathrm{E}\big\{\tau_\lambda(X_0)(s) - \tau_\lambda(X_{0,\ell})(s)\big\}^2\right]ds\int_0^1 \mathrm{E}\{\tau_\lambda(X_0)(t)\}^2dt \\
&= \lambda^{2(2a+1)/D}\,\mathrm{E}(\varepsilon_0)^2\left\{\sum_{\ell=1}^{+\infty} \mathrm{E}(\varepsilon_0 - \varepsilon_{0,\ell})^2\right\}\left[\sum_{\ell=1}^{+\infty} \mathrm{E}\|\tau_\lambda(X_0) - \tau_\lambda(X_{0,\ell})\|_{L^2}^2\right]\mathrm{E}\|\tau_\lambda(X_0)\|_{L^2}^2 \\
&\le c\,\mathrm{E}(\varepsilon_0)^2 \times \left\{\sum_{\ell=1}^{+\infty} \mathrm{E}(\varepsilon_0 - \varepsilon_{0,\ell})^2\right\} \times \mathrm{E}\|X_0\|_{L^2}^2 \times \left(\sum_{\ell=1}^{+\infty} \mathrm{E}\|X_0 - X_{0,\ell}\|_{L^2}^2\right) < \infty\,.
\end{aligned}$$

For the third term I_3, by the Cauchy-Schwarz inequality, (S1.28) and Assumption A3,

$$I_3 \le \lambda^{2(2a+1)/D}\big(\mathrm{E}\|\tau_\lambda(X_0)\|_{L^2}^2\big)^2\mathrm{E}(\varepsilon_0)^2\sum_{\ell=1}^{+\infty} \mathrm{E}(\varepsilon_0 - \varepsilon_{0,\ell})^2 \le c\big(\mathrm{E}\|X_0\|_{L^2}^2\big)^2\mathrm{E}(\varepsilon_0)^2\sum_{\ell=1}^{+\infty} \mathrm{E}(\varepsilon_0 - \varepsilon_{0,\ell})^2 < \infty\,.$$

In conclusion, we deduce from (S1.29) that $\sup_{\lambda>0}\int_0^1\int_0^1\{C_{U,\lambda}(s,t)\}^2dsdt < \infty$. □

S1.3 Proof of Theorem 3

We first deal with the bias term $W_\lambda(\beta_0)$ in (4.9). Observing (4.6), we deduce that

$$W_\lambda(\beta_0) = \sum_{k=1}^{\infty} V(\beta, \varphi_k)\, W_\lambda(\varphi_k) = \lambda \sum_{k=1}^{\infty} \frac{\rho_k\, \varphi_k\, V(\beta_0, \varphi_k)}{1+\lambda\rho_k}\,,$$

and, using Assumption A4.3, we conclude

$$\|W_\lambda(\beta_0)\|_K = \lambda \left\{ \sum_{k=1}^{\infty} \frac{\rho_k^2\, V^2(\beta_0, \varphi_k)}{1+\lambda\rho_k} \right\}^{1/2} \leq \lambda \left\{ \sum_{k=1}^{\infty} \rho_k^2\, V^2(\beta_0, \varphi_k) \right\}^{1/2} = O(\lambda)\,.$$

Note that by Lemma 5 in Section S2 and Assumption A5,

$$\sqrt{n}\lambda^{(2a+1)/(2D)} \|W_\lambda(\beta_0)\|_{L^2} \leq c\sqrt{n}\lambda^{(2a+1)/(4D)} \|W_\lambda(\beta_0)\|_K = o(1)\,. \tag{S1.30}$$

Next, applying Theorem 1 and Lemma 5 in Section S2, we find

$$\begin{aligned}
&\sup_{\nu\in[\nu_0,1]} \left\| \sqrt{n}\lambda^{(2a+1)/(2D)} \left[\nu\{\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + W_\lambda(\beta_0)\} - \frac{1}{n}\sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i \tau_\lambda(X_i) \right] \right\|_{L^2}^2 \\
&\leq c_K\, n\, \lambda^{(2a+1)/(2D)} \sup_{\nu\in[\nu_0,1]} \left\| \nu\{\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + W_\lambda(\beta_0)\} - \frac{1}{n}\sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i \tau_\lambda(X_i) \right\|_K^2 \\
&= O_p\big(n\, \lambda^{(2a+1)/(2D)}\, v_n^2\big) = O_p\Big\{ n\, \lambda^{(2a+1)/(2D)} \times n^{-1}\lambda^{-2\varsigma} \big(\lambda + n^{-1}\lambda^{-(2a+1)/(2D)}\big)(\log n) \Big\} \\
&= O_p\Big\{ \big(\lambda^{-2\varsigma+(2D+2a+1)/(2D)} + n^{-1}\lambda^{-2\varsigma}\big)(\log n) \Big\} = o_p(1)\,,
\end{aligned} \tag{S1.31}$$

where we used Assumption A5 in the last step. By Theorem 2, there exists a Gaussian process $\{\Gamma(s,\nu)\}_{s\in[0,1],\nu\in[\nu_0,1]}$ in $\mathcal{F}$ defined in the set (4.11) such that

$$\sup_{\nu\in[0,1]} \left\| n^{-1/2}\lambda^{(2a+1)/(2D)} \sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i \tau_\lambda(X_i) - \Gamma(\cdot,\nu) \right\|_{L^2}^2 = \sup_{\nu\in[0,1]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\nu\rfloor} U_i - \Gamma(\cdot,\nu) \right\|_{L^2}^2 = o_p(1)\,.$$

Combining the above finding with (S1.30) and (S1.31) yields

$$\begin{aligned}
&\sup_{\nu\in[\nu_0,1]} \left\| \sqrt{n}\lambda^{(2a+1)/(2D)} \nu\big\{\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0\big\} - \Gamma(\cdot,\nu) \right\|_{L^2}^2 \\
&\leq 3 \sup_{\nu\in[\nu_0,1]} \left\| \sqrt{n}\lambda^{(2a+1)/(2D)} \left[\nu\{\widehat{\beta}_{n,\lambda}(\cdot,\nu) - \beta_0 + W_\lambda(\beta_0)\} - \frac{1}{n}\sum_{i=1}^{\lfloor n\nu\rfloor} \varepsilon_i \tau_\lambda(X_i) \right] \right\|_{L^2}^2
\end{aligned} \tag{S1.32}$$

$$+ 3 \sup_{\nu\in[\nu_0,1]} \left\| \sqrt{n}\lambda^{(2a+1)/(2D)}\nu W_\lambda(\beta_0) \right\|_{L^2}^2 + 3 \sup_{\nu\in[0,1]} \left\| n^{-1/2}\lambda^{(2a+1)/(2D)} \sum_{i=1}^{\lfloor n\nu \rfloor} \varepsilon_i \tau_\lambda(X_i) - \Gamma(\cdot,\nu) \right\|_{L^2}^2 = o_p(1)\,.$$

Recall the definition of C_U in Assumption A4.4, and let $\{\kappa_j\}_{j=1}^\infty$ and $\{\psi_j\}_{j=1}^\infty$ denote the eigenvalues and eigenfunctions of C_U, respectively, that is,

$$C_U(s,t) = \sum_{j=1}^\infty \kappa_j\, \psi_j(s)\, \psi_j(t)\,, \qquad \int_0^1 C_U(s,t)\, \psi_j(s)\, ds = \kappa_j\, \psi_j(t)\,. \tag{S1.33}$$

By Theorem 1.1 in Berkes, Horváth and Rice (2013), we have $\Gamma(s,\nu) = \sum_{j=1}^\infty \sqrt{\kappa_j}\, \psi_j(s)\, W_j(\nu)$, for $s, \nu \in [0,1]$, where the W_j's are i.i.d. standard Brownian motions on $[0,1]$. Since $\mathrm{E}\big\{\sup_{\nu\in[\nu_0,1]} W_1^2(\nu)\big\}$ is finite, we obtain

$$\mathrm{E}\bigg\{ \sup_{\nu\in[0,1]} \|\Gamma(\cdot,\nu)\|_{L^2}^2 \bigg\} \le \sum_{j=1}^\infty \kappa_j\, \mathrm{E}\bigg\{ \sup_{\nu\in[0,1]} W_j^2(\nu) \bigg\} < \infty\,. \tag{S1.34}$$

Furthermore, observing (S1.32), we deduce from direct calculations that

$$\begin{aligned} \widehat{\mathbb{G}}_n(\nu) &= \sqrt{n}\lambda^{(2a+1)/(2D)}\nu^2 \int_0^1 \big\{ \widehat{\beta}_{n,\lambda}^2(s,\nu) - \beta_0^2(s) \big\}\, ds \\ &= 2\nu \int_0^1 \beta_0(s)\, \Gamma(s,\nu)\, ds + I_{1,n}(\nu) + I_{2,n}(\nu) + I_{3,n}(\nu) + I_{4,n}(\nu)\,, \end{aligned} \tag{S1.35}$$

where

$$\begin{aligned} I_{1,n}(\nu) &= \frac{1}{\sqrt{n}\lambda^{(2a+1)/(2D)}} \int_0^1 \Big[\sqrt{n}\lambda^{(2a+1)/(2D)}\nu \big\{ \widehat{\beta}_{n,\lambda}(s,\nu) - \beta_0(s) \big\} - \Gamma(s,\nu) \Big]^2 ds\,, \\ I_{2,n}(\nu) &= \frac{1}{\sqrt{n}\lambda^{(2a+1)/(2D)}} \int_0^1 \Gamma^2(s,\nu)\, ds\,, \\ I_{3,n}(\nu) &= \frac{2}{\sqrt{n}\lambda^{(2a+1)/(2D)}} \int_0^1 \Big[\sqrt{n}\lambda^{(2a+1)/(2D)}\nu \big\{ \widehat{\beta}_{n,\lambda}(s,\nu) - \beta_0(s) \big\} - \Gamma(s,\nu) \Big] \Gamma(s,\nu)\, ds\,, \\ I_{4,n}(\nu) &= 2\nu \int_0^1 \beta_0(s) \Big[\sqrt{n}\lambda^{(2a+1)/(2D)}\nu \big\{ \widehat{\beta}_{n,\lambda}(s,\nu) - \beta_0(s) \big\} - \Gamma(s,\nu) \Big]\, ds\,. \end{aligned}$$

Note that (S1.34) implies that $\sup_{\nu\in[0,1]} \|\Gamma(\cdot,\nu)\|_{L^2}^2 = O_p(1)$. Therefore, observing that $n\lambda^{(2a+1)/D} \to \infty$ as $n \to \infty$ (see Assumption A5), (S1.32) and the Cauchy-Schwarz inequality, it follows that $\sup_{\nu\in[\nu_0,1]}\big\{|I_{1,n}(\nu)|+|I_{2,n}(\nu)|+|I_{3,n}(\nu)|+|I_{4,n}(\nu)|\big\}=o_p(1)$ as $n\to\infty$. Consequently, in view of (S1.35), we have that,

$$\sup_{\nu\in[\nu_0,1]}\left|\widehat{\mathbb{G}}_n(\nu)-2\nu\int_0^1\beta_0(s)\,\Gamma(s,\nu)\,ds\right|=o_p(1)\,. \tag{S1.36}$$

This proves the finite-dimensional convergence, i.e., for any $k\in\mathbb{N}_+$ and $\nu_1,\dots,\nu_k\in[\nu_0,1]$,

$$\big(\widehat{\mathbb{G}}_n(\nu_1),\dots,\widehat{\mathbb{G}}_n(\nu_k)\big)\xrightarrow{d}\left(2\nu_1\int_0^1\beta_0(s)\,\Gamma(s,\nu_1)\,ds\,,\dots,\,2\nu_k\int_0^1\beta_0(s)\,\Gamma(s,\nu_k)\,ds\right). \tag{S1.37}$$

Next, we shall show the tightness of the process $\{\widehat{\mathbb{G}}_n(\nu)\}_{\nu\in[\nu_0,1]}$. To achieve this, we shall show that the process $\{\widehat{\mathbb{G}}_n(\nu)\}_{\nu\in[\nu_0,1]}$ is asymptotically uniformly equicontinuous in probability (see Lemma 1.5.7 in van der Vaart and Wellner, 1996). By the Cauchy-Schwarz inequality and (S1.36), we deduce that

$$\begin{aligned}
&\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[\nu_0,1]}}|\widehat{\mathbb{G}}_n(\nu_1)-\widehat{\mathbb{G}}_n(\nu_2)| \qquad\qquad (\text{S1.38})\\
&\le\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[\nu_0,1]}}\left|2\nu_1\int_0^1\beta_0(s)\Gamma(s,\nu_1)ds-2\nu_2\int_0^1\beta_0(s)\Gamma(s,\nu_2)ds\right|+2\sup_{\nu\in[\nu_0,1]}\left|\widehat{\mathbb{G}}_n(\nu)-2\nu\int_0^1\beta_0(s)\Gamma(s,\nu)ds\right|\\
&\le\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[\nu_0,1]}}\left\{2|\nu_1-\nu_2|\times\left|\int_0^1\beta_0(s)\Gamma(s,\nu_1)ds\right|\right\}+2\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[\nu_0,1]}}\left|\int_0^1\beta_0(s)\{\Gamma(s,\nu_1)-\Gamma(s,\nu_2)\}ds\right|+o_p(1)\\
&\le 2\delta\|\beta_0\|_{L^2}\sup_{\nu\in[\nu_0,1]}\left|\int_0^1\Gamma^2(s,\nu)ds\right|^{1/2}+2\|\beta_0\|_{L^2}\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[0,1]}}\left|\int_0^1\{\Gamma(s,\nu_1)-\Gamma(s,\nu_2)\}^2ds\right|^{1/2}+o_p(1).
\end{aligned}$$

By Lemma 2.1 in Berkes, Horváth and Rice (2013), we have $\sup_{\nu\in[0,1]}\int_0^1\Gamma^2(s,\nu)\,ds<\infty$ a.s. In addition, in view of (S1.33), note that the ψ_j's are orthogonal in $L^2([0,1])$, so that

$$\begin{aligned}
&\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[0,1]}}\int_0^1\{\Gamma(s,\nu_1)-\Gamma(s,\nu_2)\}^2\,ds=\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[0,1]}}\int_0^1\left[\sum_{j=1}^\infty\sqrt{\kappa_j}\,\psi_j(s)\,\{W_j(\nu_1)-W_j(\nu_2)\}\right]^2ds\\
&=\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[0,1]}}\sum_{j=1}^\infty\kappa_j\{W_j(\nu_1)-W_j(\nu_2)\}^2\le\sup_{\substack{|\nu_1-\nu_2|<\delta\\ \nu_1,\nu_2\in[0,1]}}\{W_1(\nu_1)-W_1(\nu_2)\}^2\sum_{j=1}^\infty\kappa_j=o_p(1)\,, \qquad (\text{S1.39})
\end{aligned}$$

where the last step is due to the modulus of continuity of Brownian motions and the fact that $\sum_{j=1}^{\infty} \kappa_j < \infty$. Therefore, combining (S1.38)–(S1.39), we deduce that, for any $e > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \mathrm{P}\left\{ \sup_{\substack{|\nu_1-\nu_2|<\delta \\ \nu_1,\nu_2\in[\nu_0,1]}} |\widehat{\mathbb{G}}_n(\nu_1) - \widehat{\mathbb{G}}_n(\nu_2)| > e \right\} = 0\,.$$

This proves the tightness of the process $\{\widehat{\mathbb{G}}_n(\nu)\}_{\nu\in[\nu_0,1]}$. Together with (S1.37), by Lemma 1.5.4 in van der Vaart and Wellner (1996), this implies that $\{\widehat{\mathbb{G}}_n(\nu)\}_{\nu\in[\nu_0,1]} \rightsquigarrow \left\{2\nu \int_0^1 \beta_0(s)\,\Gamma(s,\nu)\,ds\right\}_{\nu\in[\nu_0,1]}$ in $\ell^\infty([\nu_0,1])$. In addition, observing (S1.33), we have for the Gaussian process $\{\Gamma(s,\nu)\}_{s\in[0,1],\nu\in[\nu_0,1]}$,

$$\begin{aligned}
\mathrm{cov}\{\Gamma(s_1,\nu_1),\Gamma(s_2,\nu_2)\} &= \mathrm{cov}\left\{ \sum_{j_1=1}^{\infty} \sqrt{\kappa_{j_1}}\,\psi_{j_1}(s)\,W_{j_1}(\nu_1), \sum_{j_2=1}^{\infty} \sqrt{\kappa_{j_2}}\,\psi_{j_2}(s)\,W_{j_2}(\nu_2) \right\} \\
&= \mathrm{cov}\{W_j(\nu_1),W_j(\nu_2)\} \sum_{j=1}^{\infty} \kappa_j\,\psi_j(s_1)\,\psi_j(s_2) = (\nu_1\wedge\nu_2)\,C_U(s_1,s_2)\,,
\end{aligned}$$

where C_U is defined in Assumption A4.4 and $s_1, s_2 \in [0,1]$ and $\nu_1,\nu_2\in[\nu_0,1]$. Therefore, $\mathrm{cov}\left\{\int_0^1 \beta_0(s_1)\Gamma(s_1,\nu_1)ds, \int_0^1 \beta_0(s_2)\Gamma(s_2,\nu_2)ds_2\right\} = (\nu_1\wedge\nu_2)\int_0^1\int_0^1 C_U(s_1,s_2)\beta_0(s_1)\beta_0(s_2)ds_1ds_2 = (\nu_1\wedge\nu_2)\sigma_d^2$, where σ_d^2 is defined in (3.6). This implies that $\int_0^1 \beta_0(s)\,\Gamma(s,\nu)\,ds \overset{d.}{=} 2\sigma_d\,\mathbb{B}(\nu)$, where $\mathbb{B}$ denotes a standard Brownian motion. Hence, combining the above finding with (S1.36) yields $\{\widehat{\mathbb{G}}_n(\nu)\}_{\nu\in[\nu_0,1]} \rightsquigarrow \{2\sigma_d\,\nu\,\mathbb{B}(\nu)\}_{\nu\in[\nu_0,1]}$ in $\ell^\infty([\nu_0,1])$, which completes the proof.

S1.4 Proof of Theorem 4

Observing $\widehat{\mathbb{G}}_n(\nu)$ defined in (4.12), we have

$$\begin{aligned}
\sqrt{n}\lambda^{(2a+1)/(2D)}\widehat{\mathbb{V}}_n &= \sqrt{n}\lambda^{(2a+1)/(2D)}\left[\int_{\nu_0}^1 \left|\nu^2\int_0^1 \left\{\widehat{\beta}^2_{n,\lambda}(s,\nu) - \widehat{\beta}^2_{n,\lambda}(s,1)\right\}ds\right|^2 \omega(d\nu)\right]^{1/2} \\
&= \sqrt{n}\lambda^{(2a+1)/(2D)}\left[\int_{\nu_0}^1 \left|\nu^2\int_0^1 \left\{\widehat{\beta}^2_{n,\lambda}(s,\nu) - \beta_0^2(s)\right\}ds - \nu^2\int_0^1 \left\{\widehat{\beta}^2_{n,\lambda}(s,1) - \beta_0^2(s)\right\}ds\right|^2 \omega(d\nu)\right]^{1/2} \\
&= \left\{\int_{\nu_0}^1 \left|\widehat{\mathbb{G}}_n(\nu) - \nu^2\,\widehat{\mathbb{G}}_n(1)\right|^2 \omega(d\nu)\right\}^{1/2}.
\end{aligned}$$

In addition, for $\widehat{\mathbb{T}}_n$ defined in (3.4), $\sqrt{n}\lambda^{(2a+1)/(2D)}(\widehat{\mathbb{T}}_n - d_0) = \sqrt{n}\lambda^{(2a+1)/(2D)}\int_0^1 \left\{\widehat{\beta}^2_{n,\lambda}(s,1) - \beta_0^2(s)\right\}ds = \widehat{\mathbb{G}}_n(1)$. Therefore, we obtain $\sqrt{n}\lambda^{(2a+1)/(2D)}\left((\widehat{\mathbb{T}}_n - d_0), \widehat{\mathbb{V}}_n\right) = \Big(\widehat{\mathbb{G}}_n(1), \big\{\int_{\nu_0}^1 \big|\widehat{\mathbb{G}}_n(\nu) -$

$\nu^2\widehat{\mathbb{G}}_n(1)\big|^2\omega(d\nu)\big\}^{1/2}\Big)$. Since $\sigma_d^2 \neq 0$, by Theorem 3 and the continuous mapping theorem,

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

where $\mathbb{W}$ is defined in (4.14). This completes the proof.

S1.5 Proof of Theorem 5

When $d_0 = \int_0^1 |\beta_0(t)|^2 dt = 0$, we have $\widehat{\mathbb{T}}_n = \widehat{\mathbb{T}}_n - d_0 = o_p(1)$ and $\widehat{\mathbb{V}}_n = o_p(1)$, which implies that $\widehat{\mathbb{T}}_n - \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n = o_p(1)$, so that $\lim_{n\to\infty} \mathrm{P}\big\{\widehat{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta\big\} = \lim_{n\to\infty} \mathrm{P}\big\{\widehat{\mathbb{T}}_n - \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n > \Delta\big\} = 0$. When $0 < \int_0^1 |\beta_0(t)|^2 dt < \Delta$, we use

$$\mathrm{P}\big\{\widehat{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta\big\} = \mathrm{P}\left\{\frac{\widehat{\mathbb{T}}_n - d_0}{\widehat{\mathbb{V}}_n} > \mathcal{Q}_{1-\alpha}(\mathbb{W}) + \frac{\sqrt{n}\lambda^{(2a+1)/(2D)}(\Delta - d_0)}{\sqrt{n}\lambda^{(2a+1)/(2D)}\widehat{\mathbb{V}}_n}\right\}. \quad \text{(S1.40)}$$

Note that $\sqrt{n}\lambda^{(2a+1)/(2D)}\widehat{\mathbb{V}}_n = O_p(1)$ and $\sqrt{n}\lambda^{(2a+1)/(2D)} \to +\infty$ as $n \to \infty$ according to Assumption A5, so that $\sqrt{n}\lambda^{(2a+1)/(2D)}(\Delta - d_0) \to +\infty$. Hence, when $d_0 < \Delta$, it follows that $\mathrm{P}\big\{\widehat{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta\big\} \to 0$. When $d_0 = \int_0^1 |\beta_0(t)|^2 dt = \Delta$, we have

$$\lim_{n\to\infty} \mathrm{P}\big\{\widehat{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta\big\} = \lim_{n\to\infty} \mathrm{P}\left\{\frac{\widehat{\mathbb{T}}_n - d_0}{\widehat{\mathbb{V}}_n} > \mathcal{Q}_{1-\alpha}(\mathbb{W})\right\} = \alpha\,.$$

When $d_0 = \int_0^1 |\beta_0(t)|^2 dt > \Delta$, we have $\sqrt{n}\lambda^{(2a+1)/(2D)}(\Delta - d_0) \to -\infty$ as $n \to \infty$, so that we obtain from (S1.40) that $\mathrm{P}\big\{\widehat{\mathbb{T}}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\widehat{\mathbb{V}}_n + \Delta\big\} \to 0$, which completes the proof.

S2. Auxiliary lemmas

Lemma 4. *Under Assumptions A1 and A2, there exists a constant $c > 0$ such that, for any $x \in L^2([0,1])$, $\|\tau_\lambda(x)\|_K^2 \le c\,\lambda^{-(2a+1)/(2D)}\|x\|_{L^2}^2$ and $\mathrm{E}\|\tau_\lambda(x)\|_K^2 \le c\,\lambda^{-1/(2D)}$.*

Proof. Recalling (3.8) and using the orthogonality of the functions φ_k, we have

$$\|\tau_\lambda(x)\|_K^2 = \sum_{k=1}^\infty \frac{\langle x, \varphi_k\rangle_{L^2}^2}{1 + \lambda\rho_k} \le \|x\|_{L^2}^2 \sum_{k=1}^\infty \frac{\|\varphi_k\|_{L^2}^2}{1 + \lambda\rho_k} \le c\,\lambda^{-(2a+1)/(2D)}\|x\|_{L^2}^2\,.$$

By Assumption A2, $\mathrm{E}\big(\langle X,\varphi_k\rangle_{L^2}^2\big)=1$, so that $\mathrm{E}\|\tau_\lambda(X)\|_K^2=\sum_{k=1}^{\infty}(1+\lambda\rho_k)^{-1}\le c\,\lambda^{-1/(2D)}$.

Lemma 5 (Lemma 3.1 in Shang and Cheng, 2015)**.** *Under Assumptions A1 and A2, there exists a constant $c_K>0$ such that for any $\beta\in\mathcal{H}$, $\|\beta\|_{L^2}^2\le c_K\lambda^{-(2a+1)/(2D)}\|\beta\|_K^2$.*

Lemmas 6 and 7 below shows that the long-run covariance $C_{X\varepsilon}(s,t)=\sum_{\ell=-\infty}^{+\infty}\mathrm{cov}\{\varepsilon_0X_0(s),\varepsilon_\ell X_\ell(t)\}$ of $\{X_i\varepsilon_i\}_{i\in\mathbb{Z}}$ satisfies $C_{X\varepsilon}\in L^2([0,1]^2)$ and $\int_0^1 C_{X\varepsilon}(t,t)dt<\infty$.

Lemma 6. *Under Assumption A3, we have $C_{X\varepsilon}\in L^2([0,1]^2)$.*

Proof. Note that by Assumption A3, for $\ell\ge1$, $\varepsilon_{0,\ell}$ and $\varepsilon_{-\ell}$ are independent; $X_{0,\ell}$ and $X_{-\ell}$ are independent. Since $\mathrm{E}(X)\equiv0$, we find that, for $\ell\ge1$, $\mathrm{E}\{X_0(s)\,X_\ell(t)\}=\mathrm{E}\big[\big\{X_0(s)-X_{0,\ell}(s)\big\}\,X_{-\ell}(t)\big]$. Observing that $\mathrm{E}(\varepsilon_0\varepsilon_{-\ell})=\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}+\mathrm{E}(\varepsilon_{0,\ell}\varepsilon_{-\ell})=\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}$, for $\ell\ge1$, we deduce from the above equation that, for $\ell\ge1$,

$$\mathrm{cov}\big\{\varepsilon_0\,X_0(s)\,,\varepsilon_{-\ell}\,X_{-\ell}(t)\big\}=\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\,\mathrm{E}\Big[\big\{X_0(s)-X_{0,\ell}(s)\big\}\times X_{-\ell}(t)\Big]\,. \qquad \text{(S2.1)}$$

Therefore, we find from the above equation that $\int_0^1\int_0^1\{C_{X\varepsilon}(s,t)\}^2dsdt\le 2I_1+8I_2$, where

$$I_1=\{\mathrm{E}(\varepsilon_0^2)\}^2\int_0^1\int_0^1\Big[\mathrm{E}\big\{X_0(s)\times X_0(t)\big\}\Big]^2dsdt\,,$$
$$I_2=\int_0^1\int_0^1\left(\sum_{\ell=1}^{+\infty}\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\,\mathrm{E}\Big[\big\{X_0(s)-X_{0,\ell}(s)\big\}\times X_{-\ell}(t)\Big]\right)^2dsdt\,. \qquad \text{(S2.2)}$$

For the first term I_1, we have $I_1=\{\mathrm{E}(\varepsilon_0^2)\}^2\times\|C_X\|_{L^2}^2<\infty$. For the second term I_2, by the Cauchy-Schwarz inequality, we find

$$\left(\sum_{\ell=1}^{+\infty}\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\,\mathrm{E}\Big[\big\{X_0(s)-X_{0,\ell}(s)\big\}\times X_{-\ell}(t)\Big]\right)^2$$
$$\le\left[\sum_{\ell=1}^{+\infty}\mathrm{E}\big\{(\varepsilon_0-\varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}^2\right]\times\left[\sum_{\ell=1}^{+\infty}\mathrm{E}\big\{X_0(s)-X_{0,\ell}(s)\big\}^2\right]\times\mathrm{E}\{X_0(t)\}^2$$
$$=\mathrm{E}(\varepsilon_0^2)\times\left(\sum_{\ell=1}^{+\infty}\mathrm{E}|\varepsilon_0-\varepsilon_{0,\ell}|^2\right)\times\left[\sum_{\ell=1}^{+\infty}\mathrm{E}\big\{X_0(s)-X_{0,\ell}(s)\big\}^2\right]\times\mathrm{E}\{X_0(t)\}^2\,. \qquad \text{(S2.3)}$$

Therefore, by Assumption A3, we deduce from the above equation that $I_2 \leq \mathrm{E}(\varepsilon_0)^2 \times \big(\sum_{\ell=1}^{+\infty} \mathrm{E}|\varepsilon_0 - \varepsilon_{0,\ell}|^2\big) \times \big(\sum_{\ell=1}^{+\infty} \mathrm{E}\|X_0 - X_{0,\ell}\|_{L^2}^2\big) \times \mathrm{E}\|X_0\|_{L^2}^2 < \infty$. Finally, we deduce that $\int_0^1 \int_0^1 \{C_{X\varepsilon}(s,t)\}^2 dsdt < \infty$ and complete the proof. □

Lemma 7. *Under Assumption A3, we have* $\int_0^1 C_{X\varepsilon}(s,s)ds < \infty$.

Proof. By the arguments similar to the ones used to obtain (S2.1), it follows that, for $\ell \geq 1$, $\mathrm{cov}\big\{\varepsilon_0 X_0(s), \varepsilon_{-\ell} X_{-\ell}(s)\big\} = \mathrm{E}\big\{(\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\mathrm{E}\big[\big\{X_0(s) - X_{0,\ell}(s)\big\}X_{-\ell}(s)\big]$. Therefore, following the proof of Lemma 6, we deduce that

$$\int_0^1 C_{X\varepsilon}(s,s)ds = \int_0^1 \left[\mathrm{cov}\big\{\varepsilon_0\, X_0(s)\,,\varepsilon_0\, X_0(s)\big\} + 2\sum_{\ell=1}^{+\infty} \mathrm{cov}\big\{\varepsilon_0\, X_0(s)\,,\varepsilon_{-\ell}\, X_{-\ell}(s)\big\}\right]ds$$

$$= \mathrm{E}(\varepsilon_0^2)\,\mathrm{E}\|X_0\|_{L^2}^2 + 2\int_0^1 \left(\sum_{\ell=1}^{+\infty} \mathrm{E}\big\{(\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell}\big\}\,\mathrm{E}\Big[\big\{X_0(s) - X_{0,\ell}(s)\big\}X_{-\ell}(s)\Big]\right)ds$$

$$\leq \mathrm{E}(\varepsilon_0^2)\mathrm{E}\|X_0\|_{L^2}^2 + 2\int_0^1 \left\{\mathrm{E}(\varepsilon_0^2)\left(\sum_{\ell=1}^{+\infty} \mathrm{E}|\varepsilon_0 - \varepsilon_{0,\ell}|^2\right)\left[\sum_{\ell=1}^{+\infty} \mathrm{E}\big\{X_0(s) - X_{0,\ell}(s)\big\}^2\right]\mathrm{E}\{X_0(s)\}^2\right\}^{1/2} ds\,,$$

where in the last step we applied (S2.3) by taking $t = s$. Therefore, by the Cauchy-Schwarz inequality and Assumption A3, we conclude that $\int_0^1 C_{X\varepsilon}(s,s)ds \leq \mathrm{E}(\varepsilon_0^2)\mathrm{E}\|X_0\|_{L^2}^2 + 2\{\mathrm{E}(\varepsilon_0)^2\}^{1/2}\big(\sum_{\ell=1}^{+\infty}\mathrm{E}|\varepsilon_0 - \varepsilon_{0,\ell}|^2\big)^{1/2}\big(\mathrm{E}\|X_0\|_{L^2}^2\big)^{1/2}\big(\sum_{\ell=1}^{+\infty}\mathrm{E}\|X_0 - X_{0,\ell}\|_{L^2}^2\big)^{1/2} < \infty$. □

Lemma 8 below is used to prove Theorem 1. Recall the definition of $H_{n,k}$ in (S1.15).

Lemma 8. *For* $p_n \geq 1$, *let* $\mathcal{F}_{p_n} = \{\beta \in \mathcal{H} : \|\beta\|_{L^2} \leq 1, J(\beta,\beta) \leq p_n\}$. *Then, under Assumptions A1–A4, as* $n \to \infty$,

$$\max_{1\leq k\leq n}\sup_{\beta\in\mathcal{F}_{p_n}} \frac{\|H_{n,k}(\beta)\|_K}{p_n^{1/(2m)}\|\beta\|_{L^2}^{(m-1)/m} + n^{-1/2}} = O_p\big(\lambda^{-1/(2D)}\log n\big)^{1/2}.$$

Proof. The proof of Lemma 8 follows a modified argument of the proof of Lemma 3.4 in Shang and Cheng (2015). For any $x \in L^2([0,1])$, let $g(x,\beta) = \tau_\lambda(x)\int_0^1 \beta(s)x(s)ds$. We have

$$H_{n,k}(\beta_1) - H_{n,k}(\beta_2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{j}\Big(g(X_i,\beta_1-\beta_2)\times\mathbb{1}\{\mathcal{E}_n(c)\} - \mathrm{E}\big[g(X_i,\beta_1-\beta_2)\times\mathbb{1}\{\mathcal{E}_n(c)\}\big]\Big).$$

Note that, on the event $\mathcal{E}_n(c)$ defined below equation (S1.6), $\big|\langle \beta_1-\beta_2, X_i\rangle_{L^2}\,\mathbb{1}\{\mathcal{E}_n(c)\} - \mathrm{E}\big[\langle \beta_1-\beta_2, X_i\rangle_{L^2}\,\mathbb{1}\{\mathcal{E}_n(c)\}\big]\big| \le c\,\|\beta_1-\beta_2\|_{L^2}$. Hence, we deduce that

$$\Big\|g(X_i,\beta_1-\beta_2)\mathbb{1}\{\mathcal{E}_n(c)\} - \mathrm{E}\big[g(X_i,\beta_1-\beta_2)\mathbb{1}\{\mathcal{E}_n(c)\}\big]\Big\|_K^2 \le 2(\log n)^2\|\beta_1-\beta_2\|_{L^2}^2\|\tau_\lambda(X_i)\|_K^2\,.$$

For $1\le k\le n$, let $W_{n,k}^2 = n^{-1}\sum_{i=1}^k \|\tau_\lambda(X_i)\|_K^2$ and $\mathcal{X}_n = \{\|\tau_\lambda(X_i)\|_K\}_{i=1}^n$. By Lemma 4, $\mathrm{E}(W_{n,k}^2) \le (k/n)\mathrm{E}\|\tau_\lambda(X_i)\|_K^2 \le c\,\lambda^{-1/(2D)}$. By Theorem 3.5 in Pinelis (1994), for any $1\le j\le n$, for any $\beta_1,\beta_2\in\mathcal{H}$ and for $1\le j\le n$,

$$\begin{aligned}
&\mathrm{P}\left\{\|H_{n,k}(\beta_1)-H_{n,k}(\beta_2)\|_K \ge x\,\big|\mathcal{X}_n\right\}\\
&= \mathrm{P}\left\{\left\|\frac{1}{\sqrt{k}}\sum_{i=1}^k\Big[g(X_i,\beta_1-\beta_2)-\mathrm{E}\big\{g(X_i,\beta_1-\beta_2)\big\}\Big]\right\|_K \ge \sqrt{n/k}\,x\,\big|\mathcal{X}_n\right\}\\
&\le 2\exp\left(-\frac{nk^{-1}x^2}{2k^{-1}\sum_{i=1}^k\|\tau_\lambda(X_i)\|_K^2\,\|\beta_1-\beta_2\|_{L^2}^2}\right) \le 2\exp\left(-\frac{x^2}{2\,W_{n,k}^2\,\|\beta_1-\beta_2\|_{L^2}^2}\right)\,. \qquad \text{(S2.4)}
\end{aligned}$$

Following the proof of Lemma 3.4 in Shang and Cheng (2015) (see p. 13 of Shang and Cheng, 2015b), we deduce that, for any $1\le k\le n$,

$$\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le\delta}\|H_{n,k}(\beta)\|_K\ge x\,\big|\mathcal{X}_n\right\} \le 2\exp\big(-c_1^{-2}W_{n,k}^{-2}\,p_n^{-1/(2m)}\delta^{-2+1/m}x^2\big)\,.$$

Taking $\gamma = 1-1/(2m)$, $b_n=\sqrt{n}\,p_n^{1/(4m)}$, $\theta_n = b_n^{-1}$, $Q_n = \lfloor -\log_2\theta_n+\gamma-1\rfloor$ and $T_n = c_2(\lambda^{-1/(2D)}\log n)^{1/2}$, for some constant $c_2>0$ to be specified below, yields that

$$\begin{aligned}
&\mathrm{P}\left\{\max_{1\le k\le n}\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le 2}\frac{\sqrt{n}\|H_{n,k}(\beta)\|_K}{b_n\|\beta\|_{L^2}^\gamma+1}\ge T_n\,\big|\mathcal{X}_n\right\} \le \sum_{k=1}^n\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le\theta_n^{1/\gamma}}\sqrt{n}\|H_{n,k}(\beta)\|_K\ge T_n\,\big|\mathcal{X}_n\right\}\\
&\qquad+\sum_{k=1}^n\sum_{j=0}^{Q_n}\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,(\theta_n2^j)^{1/\gamma}\le\|\beta\|_{L^2}\le(\theta_n2^{j+1})^{1/\gamma}}\frac{\sqrt{n}\|H_{n,k}(\beta)\|_K}{b_n\|\beta\|_{L^2}^\gamma+1}\ge T_n\,\big|\mathcal{X}_n\right\}\\
&\le\sum_{k=1}^n\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le\theta_n^{1/\gamma}}\sqrt{n}\|H_{n,k}(\beta)\|_K\ge T_n\,\big|\mathcal{X}_n\right\}\\
&\qquad+\sum_{k=1}^n\sum_{j=0}^{Q_n}\mathrm{P}\left\{\sup_{\beta\in\mathcal{F}_{p_n},\,\|\beta\|_{L^2}\le(\theta_n2^{j+1})^{1/\gamma}}\sqrt{n}\|H_{n,k}(\beta)\|_K\ge(b_n\theta_n2^j+1)T_n\,\big|\mathcal{X}_n\right\}\\
&\le 2\sum_{k=1}^n\exp\big(-c_1^{-2}W_{n,k}^{-2}\,p_n^{-1/(2m)}\theta_n^{(-1+1/m)/\gamma}n^{-1}T_n^2\big)
\end{aligned}$$

$$+ 2\sum_{k=1}^{n}\sum_{j=0}^{Q_n}\exp\Big\{-c_1^{-2}W_{n,k}^{-2}p_n^{-1/m}(\theta_n 2^{j+1})^{(-1+1/m)/\gamma}(b_n\theta_n 2^j+1)^2 n^{-1}T_n^2\Big\}$$
$$\le 2\sum_{k=1}^{n}\exp\big(-c_1^{-2}W_{n,k}^{-2}T_n^2\big)+2(Q_n+1)\sum_{k=1}^{n}\exp\big(-c_1^{-2}W_{n,k}^{-2}T_n^2/4\big)$$
$$\le 2(Q_n+2)\sum_{k=1}^{n}\exp\big(-c_1^{-2}W_{n,k}^{-2}T_n^2/4\big)\le 2(Q_n+2)\exp\big(\log n - c_1^{-2}W_{n,n}^{-2}T_n^2/4\big).$$

Denote the event $\mathcal{A}_n=\{W_{n,n}^2\le c_3\lambda^{-1/(2D)}\}$ for some constant $c_3>0$. Since $\mathrm{E}(W_{n,n}^2)\le c\lambda^{-1/(2D)}$, we have that, for c_3 large enough, $\mathrm{P}(\mathcal{A}_n)\to 1$ as $n\to\infty$. On the event $\mathcal{A}_n$, by taking $c_2>2c_1c_3^{-1/2}$, as $n\to\infty$, $2(Q_n+2)\exp\big(\log n-c_1^{-2}W_{n,n}^{-2}T_n^2/4\big)\le 2(Q_n+2)\exp\big\{\log n - c_1^{-2}c_2^2c_3\log n/4\big\}=o(1)$, which together with (S2.4) completes the proof. □

References

Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.*, **22**, 1679–1706.

Shang, Z. and Cheng, G. (2015b) Supplement to "Nonparametric inference in generalized functional linear models". DOI:10.1214/15-AOS1322SUPP.

Van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes. *Springer*, New York.