Received: 4 July 2016 | Revised: 8 December 2017 | Accepted: 16 March 2018

DOI: 10.1111/sjos.12333

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Nonparametric inference for functional-on-scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament

Konrad Abramowicz[1] | Charlotte K. Häger[2] | Alessia Pini[3,4] | Lina Schelin[2,3] | Sara Sjöstedt de Luna[1] | Simone Vantini[5]

[1]Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden
[2]Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Sweden
[3]Department of Statistics, Umeå School of Business, Economics and Statistics, Umeå University, Umeå, Sweden
[4]Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy
[5]MOX - Modelling and Scientific Computing Laboratory, Department of Mathematics, Politecnico di Milano, Milan, Italy

Correspondence
Alessia Pini, Department of Statistical Sciences, Università Cattolica del Sacro Cuore, 20123 Milan, Italy.
Email: alessia.pini@unicatt.it; alessia.pini@polimi.it

Funding information
Swedish Scientific Research Council, Grant/Award Number: K2014-99X-21876-04-4, 340-2013-5203, and 2016-02763; Västerbotten County Council, Grant/Award Number: ALF VLL548501 and VLL-358901 Project No. 7002795

Abstract

Motivated by the analysis of the dependence of knee movement patterns during functional tasks on subject-specific covariates, we introduce a distribution-free procedure for testing a functional-on-scalar linear model with fixed effects. The procedure does not only test the global hypothesis on the entire domain but also selects the intervals where statistically significant effects are detected. We prove that the proposed tests are provided with an asymptotic control of the intervalwise error rate, that is, the probability of falsely rejecting any interval of true null hypotheses. The procedure is applied to one-leg hop data from a study on anterior cruciate ligament injury. We compare knee kinematics of three groups of individuals (two injured groups with different treatments and one group of healthy controls), taking individual-specific covariates into account.

KEYWORDS

analysis of covariance, functional data, human movement, intervalwise testing, permutation test

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2018 The Authors Scandinavian Journal of Statistics published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics

1 | INTRODUCTION

Functional data analysis is a dynamically developing research area within the field of statistics. In recent literature, linear models for functional data have been widely studied (see, e.g., Abramovich & Angelini, 2006; Abramowicz et al., 2014; Cardot, Prchal, & Sarda, 2007; Fan & Zhang, 2000; Fogarty & Small, 2014; Gertheiss, Goldsmith, Crainiceanu, & Greven, 2013; Reiss, Huang, & Mennes, 2010).

Motivated by the analysis of the dependence of knee movement patterns during functional tasks on subject-specific covariates, we consider in the present paper a functional-on-scalar linear model. Specifically, we model a functional response with a set of covariates multiplied by functional parameters. Such models find their applications in a wide range of research fields where modern techniques enable the collection of high-resolution time-continuous data. In this context, many of the empirically relevant questions address the effect of covariates on a functional response. They may also desire the identification of significant domain subsets, that is, time intervals characterized by significant effects of a specific covariate. The incitement for this work comes from a follow-up study after anterior cruciate ligament (ACL) injury. ACL injuries are common worldwide, especially in sports, and are typically treated conservatively either with physiotherapy or with reconstructive surgery in combination with physiotherapy. We analyze the knee-joint kinematics data of a sagittal plane, that is, knee flexion/extension, during a one-leg hop for distance for $n = 95$ individuals (see Figure 1). We compare individuals from the surgery and physiotherapy groups with healthy-knee controls matched with age and sex.

Previous studies indicated differences in the movement patterns between these groups (Hébert-Losier et al., 2015; Tengman, Grip, Stensdotter, & Häger, 2015). Tengman et al. (2015) had a limitation because it is only considering selected landmarks (events) of the curves. The previous analysis included group effects and covariates but did not take into account the information coming from the functional nature of the data. In the work of Tengman et al. (2015), the complete functional data were considered but without taking covariates into account. Both these studies indicate less knee flexion among the individuals treated with physiotherapy compared with the control group. Here, we overcome both the aforementioned limitations by introducing a statistical tool that both exploits the functional nature of the data and takes into account possible covariate effects. In this paper, we investigate if these differences are only due to group effects or if they can be explained by means of additional individual-specific covariates. Furthermore, the introduced methodology enables the detection of the intervals where the covariates have significant effects

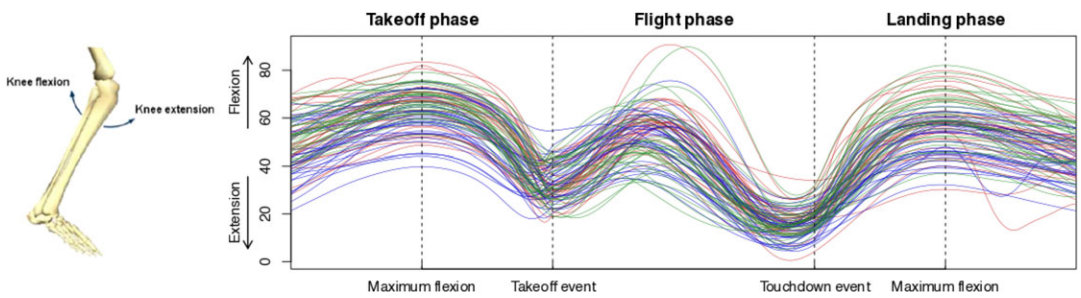

FIGURE 1 Knee flexion and extension in the (left) sagittal plane and (right) flexion/extension angle curves of the (blue) physiotherapy, (red) reconstructive surgery, and (green) control groups. Higher values indicate greater knee flexion [Colour figure can be viewed at wileyonlinelibrary.com]

(domain selection). Such information provides additional insight into the importance of different time segments of the movement. For instance, the active preparation phase with, for example, knee bending, prior to the actual takeoff of the jump determines much of the performance, but there is no obvious single event that would necessarily be the most representative for the comparison of movement control for individuals or groups. Likewise, the control of the knee in the landing phase is essential for maintaining balance during the task, but it is not known to a full extent how this is controlled or even how to best assess it. The present method enables comparisons of the whole movement pattern tailored also to individuals and may, in addition, help in identifying critical intervals within the larger phases of the movement execution.

Parameter estimation of the functional model is handled by least squares estimation, as suggested, for instance, by Ramsay and Silverman (2005). However, forming valid tests of various hypotheses about the functional regression parameters, with the control of the error rate, is not straightforward. One solution adopted in the literature is to develop global tests for the parameters of the model (Abramovich & Angelini, 2006; Antoniadis & Sapatinas, 2007; Cardot et al., 2007; Cuesta-Albertos & Febrero-Bande, 2010; Cuevas, Febrero, & Fraiman, 2004; Kayano, Matsui, Yamaguchi, Imoto, & Miyano, 2015; Schott, 2007; Staicu, Li, Crainiceanu, & Ruppert, 2014; Zhang & Liang, 2014). Such tests investigate if a covariate has a significant effect on the response but does not provide any domain selection. Another approach, proposed by Fan and Zhang (2000), Reiss et al. (2010), and Ramsay and Silverman (2005), is to provide pointwise confidence bands for the functional parameters. The results indicate which parts of the domain that the covariates have an effect, with only a pointwise control of the type I error rate. As clearly discussed in the work of Ramsay and Silverman (2005, pp. 243–244), pointwise limits are not equivalent to confidence regions for the entire estimated curves.

Assuming that data are expressed through a functional basis, inference can be based directly on the expansion coefficients, as proposed, for instance, by Spitzner, Marron, and Essick (2003) and Pini and Vantini (2016). In the latter work, single-component tests are performed and their p values adjusted with multiple testing methods. In this way, results are compensated for the many dependent tests performed on the same data set. A drawback with this type of procedures is that they rely on the basis expansion. The test conclusions thus depend on the type of functional basis selected to initialize the methods.

Vsevolozhskaya et al. (2013) and Vsevolozhskaya, Greenwood, and Holodov (2014) proposed an inferential procedure in the functional analysis of variance (ANOVA) framework based on partitioning the domain into preselected subintervals. The procedure, relying on permutation tests, results in a selection of subintervals where significant differences are observed between groups. However, the test conclusions depend on the initially chosen partition and do not take into account covariates.

In the framework of testing differences between two populations, Pini and Vantini (2016) proposed the intervalwise testing (IWT) procedure that tests functional hypotheses and selects intervals of the domain where the null hypothesis is rejected. The procedure relies on the definition of an adjusted p value function provided with a control of the intervalwise error rate (IWER), that is, control of the probability of wrongly rejecting any interval of the domain.

Assuming continuity of the functional data, it is possible to discretize the domain on a fine grid and perform inference separately for each point of the grid. In multivariate data analysis, several techniques have been proposed in the statistical literature to adjust such univariate p values controlling the familywise error rate (FWER; e.g., Holm, 1979; Holmes, Blair, Watson, & Ford, 1996; Marcus, Eric, & Gabriel, 1976). However, the well-known Bonferroni or Bonferroni–Holm adjustments are not suitable for the case of discretized functional data because their power tends

to zero as the number of components (points of the grid) increases. Another popular alternative, the closed testing procedure (Marcus et al., 1976), becomes computationally impractical as the number of points of the grid increases. One of the operational methods that can be straightforwardly applied to discrete pointwise evaluations of functional data in the t-test framework was proposed by Holmes et al. (1996). It was extended to the case of vector-on-scalar linear models with high-dimensional responses by Winkler, Ridgway, Webster, Smith, and Nichols (2014) and is based on the maximum of the F-test statistics for global tests, therefore called an $F_{\max}$ procedure. This procedure can be used for domain selection even if no information on a partition of the domain is available, and it does not require a basis expansion.

In our work, we extend the IWT framework proposed by Pini and Vantini (2017) to the functional-on-scalar linear model to test various hypotheses on the functional coefficients and prove that the proposed tests are exact or asymptotically exact. We utilize tests based on permutations, and hence, the procedure does not require normality of the residual functions and knowledge about the covariance structure of the data. We study the performance of our procedure on the kinematics data and in a small simulation study. The simulation study investigates the performance of the the IWT procedure. In addition, it includes a comparison of IWT with the $F_{\max}$ procedure originating from the multivariate setting, to shed light on the behavior of the latter procedure adapted to the functional data framework.

The paper is outlined as follows. In Section 2, we describe the functional-on-scalar linear model, discussing the methodology proposed for functional parameter estimation and inference. Section 3 reports the theoretical properties of the proposed methodology. The proofs of theorems in Section 3 are reported Section A of the Appendix. In Section 4, we provide the implementation details of the proposed procedure. In Section 5, we report the results of the analysis of the kinematics data, whereas in Section 6, we present the results of the simulation study. Finally, Section B of the Appendix reports some algorithmic details on the permutation scheme used to perform inference, and Sections C and D of the Appendix report supplementary results on the kinematics data. All computations and plots have been created using R (R Core Team, 2017).

2 | METHODOLOGY

2.1 | The functional-on-scalar linear model

Suppose we have observed a sample of n-continuous squared integrable random functions $y_i(t)$, $t \in [a,b]$, $i = 1, \dots, n$, that is, $y_i(\cdot) \in C^0[a,b]\,\forall i$. We want to study the following functional-on-scalar linear model:

$$y_i(t) = \beta_0(t) + \sum_{l=1}^{L} \beta_l(t) x_{li} + \varepsilon_i(t), \quad \forall i = 1, \dots, n, \, t \in [a,b], \tag{1}$$

where $x_{1i}, \dots, x_{Li} \in \mathbb{R}$ are known scalar covariates and $\beta_l(t)$, $l = 0, \dots, L$, are the unknown fixed functional regression parameters. The errors $\varepsilon_i(t)$, $t \in [a,b]$ are independent and identically distributed (with respect to units) zero-mean random functions (not necessarily Gaussian) with finite total variance, that is,

$$\int_a^b \mathbb{E}[\varepsilon_i(t)]^2 \mathrm{d}t < \infty, \quad \forall i = 1, \dots, n. \tag{2}$$

No further assumption is needed on the covariance structure of the error term.

2.2 | Model estimation

The ordinary least squares (OLS) estimators of the functional parameters $\beta_l(t), l = 0, \ldots, L$, can be found by minimizing the sum over units of the squared L^2 distances between the functional data $y_i(t)$ and the quantity $\beta_0(t) + \sum_{l=1}^{L} \beta_l(t)x_{li}$ with respect to $\beta_l(t), l = 0, \ldots, L$, (Ramsay & Silverman, 2005), hence minimizing

$$\sum_{i=1}^{n} \int_a^b \left(y_i(t) - \beta_0(t) - \sum_{l=1}^{L} \beta_l(t)x_{li} \right)^2 \mathrm{d}t. \tag{3}$$

From the interchangeability of summation and integration in (3), it follows that minimization can be done separately for each point of the domain, independently of the covariance structure of the errors $\varepsilon_i(t)$. The OLS estimate $\hat{\boldsymbol{\beta}}_n(t) = (\hat{\beta}_0(t), \ldots, \hat{\beta}_L(t))', t \in [a, b]$, minimizing (3), can thus be calculated pointwise for each given t as follows:

$$\hat{\boldsymbol{\beta}}_n(t) = \underset{\beta_0(t), \ldots, \beta_L(t)}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i(t) - \beta_0(t) - \sum_{l=1}^{L} \beta_l(t)x_{li} \right)^2. \tag{4}$$

For each t, $\hat{\boldsymbol{\beta}}_n(t)$ is thus the OLS estimator of the corresponding scalar-on-scalar linear model at point t.

Asymptotic properties for the OLS estimates at each point t can be immediately derived from classical results for scalar-on-scalar linear models. In detail, let X_n be the design matrix $X_n \in \mathbb{R}^{(n\times(L+1))}$ ($[X_n]_{i,1} = 1, \forall i = 1, \ldots, n$; $[X_n]_{i,j} = x_{j-1,i}, i = 1, \ldots, n, j = 2, \ldots, L+1$). Consider the following standard conditions.

- *C1* The matrix $X_n'X_n$ is nonsingular, and the inverse $V = (X_n'X_n)^{-1}$ is such that the elements $[V]_{ij} \to 0$ as $n \to \infty$, for all $i, j = 1, \ldots, L+1$.
- *C2* For each $t \in [a, b]$, the regression errors $\varepsilon_i(t)$ satisfy the following:

$$\mathbb{E}\left[\varepsilon_i(t)^2\right] < \infty.$$

Under conditions *C1* and *C2*, we have that, for each $t \in [a, b]$, the OLS estimate

$$\hat{\boldsymbol{\beta}}_n(t) = (X_n'X_n)^{-1}X_n'(y_1(t), \ldots, y_n(t))' \tag{5}$$

is a strongly consistent estimate of $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_L(t))'$ (Lai, Robbins, & Wei, 1979). Condition *C1* is a sufficient condition for finding an explicit expression of the OLS estimates and guarantees convergence in probability. Condition *C2* ensures almost sure convergence.

2.3 | Model inference

One of the main challenges with inference for functional linear models (1) is to perform valid tests of various hypotheses on the functional regression parameters and to select significant intervals of the domain. For instance, we are interested in the overall model test that none of the covariates have the functional version of the classical F-test hypotheses as follows:

$$\begin{cases} H_{0,F} : \beta_l(t) = 0, & \forall l \in 1, \ldots, L, \ \forall t \in [a, b] \\ H_{1,F} : \beta_l(t) \neq 0, & \text{for some } l \in \{1, \ldots, L\} \text{ and some } t \in [a, b], \end{cases} \tag{6}$$

together with tests on single functional parameters, that is, the functional version of the classical t-test hypotheses, as follows:

$$\begin{cases} H_{0,l} : \beta_l(t) = 0, & \forall t \in [a,b] \\ H_{1,l} : \beta_l(t) \neq 0, & \text{for some } t \in [a,b], \end{cases} \tag{7}$$

where $l \in \{0, \dots, L\}$.

It may also be of interest to test hypotheses on one or more linear combinations of the functional parameters of the regression, specified by a combination matrix C. In detail, let $C \in \mathbb{R}^{(q\times(L+1))}$ be any real-valued full rank matrix, where q denotes the number of hypotheses on the functional regression parameters to be jointly tested, with $1 \leq q \leq L+1$. Moreover, let $\mathbf{c}_0(t) = (c_{01}(t), \dots, c_{0q}(t))'$ be a vector of fixed functions in $C^0[a,b]$. The general testing problem can typically then be formulated as follows:

$$\begin{cases} H_{0,C} : C\boldsymbol{\beta}(t) = \mathbf{c}_0(t), & \forall t \in [a,b] \\ H_{1,C} : C\boldsymbol{\beta}(t) \neq \mathbf{c}_0(t), & \text{for some } t \in [a,b], \end{cases} \tag{8}$$

where the jth element of vector $C\boldsymbol{\beta}(t)$ is a function obtained by means of a linear combination of the functional regression parameters $\beta_l(t)$ with weights $[C]_{jl}$: $[C\boldsymbol{\beta}(\boldsymbol{t})]_j = \sum_{l=0}^{L} [C]_{jl}\beta_l(t)$, $j = 1, \dots, q$. There are two important special cases of the general functional linear hypotheses (8) as follows.

1. Let $q = L$, $C = C_F = (\mathbf{0}|I_L) \in \mathbb{R}^{L\times(L+1)}$, and $\mathbf{c_0}(\mathbf{t}) = \mathbf{0} \in \mathbb{R}^L$, where I_L is the $L \times L$ identity matrix. Then, the hypotheses of (8) correspond to the hypotheses in (6).
2. For a fixed l, let $q = 1$, $C = C_l \in \mathbb{R}^{1\times(L+1)}$ with $[C_l]_r = 1$ if $r = l$ and 0 otherwise, and $c(t) = 0$. Then, the hypotheses in (8) correspond to the hypotheses in (7).

For test (8), in case of rejection of the null hypothesis $H_{0,C}$, we want to select the intervals on $[a,b]$ where significant differences are detected. In theory, this problem can be addressed by performing an infinite family of tests along the domain $[a,b]$, of the form as follows:

$$\begin{cases} H^t_{0,C} : C\boldsymbol{\beta}(t) = \mathbf{c}_0(t) \\ H^t_{1,C} : C\boldsymbol{\beta}(t) \neq \mathbf{c}_0(t). \end{cases} \tag{9}$$

Based on classical results for scalar-on-scalar linear models, it is rather straightforward to test hypotheses (9) for each t. The challenge is to control the FWER arising from the uncountable infinite number of (dependent) hypotheses tests. In this paper, we extend the domain selection IWT procedure by Pini and Vantini (2017) to functional-on-scalar linear models to control the probability of wrongly rejecting each interval characterized by only true hypotheses, that is, the control of the IWER. The result of the procedure is an adjusted p value function $\tilde{p}_C(t)$ for testing (9) provided with a control of the IWER. The adjusted p value function can be thresholded at level α to select the portions of the domain imputable for the rejection of the null hypothesis $H_{0,C}$. This type of control guarantees that, for every interval of the domain where $H_{0,C}$ is true (i.e., where $C\boldsymbol{\beta}(t) = \mathbf{c}_0(t)$), the probability that $H_{0,C}$ is rejected at least at one point of the interval is less or equal to α. The domain selection procedure that we propose is based on three steps presented in detail in the following paragraphs.

2.3.1 | First step: IWT

Given any closed interval $\mathcal{I} \subseteq [a, b]$, we want to test the following:

$$\begin{cases} H_{0,C}^{\mathcal{I}} : C\boldsymbol{\beta}(t) = \mathbf{c}_0(t), \quad \forall t \in \mathcal{I} \\ H_{1,C}^{\mathcal{I}} : C\boldsymbol{\beta}(t) \neq \mathbf{c}_0(t), \quad \text{for some } \ t \in \mathcal{I}. \end{cases} \tag{10}$$

To perform tests of linear hypotheses (10), we propose to use the statistics as follows:

$$T_C^{\mathcal{I}} = \int_{\mathcal{I}} T_C(t)\mathrm{d}t, \tag{11}$$

where

$$T_C(t) = \left(C\hat{\boldsymbol{\beta}}(t) - \mathbf{c}_0(t)\right)' \left(C\hat{\boldsymbol{\beta}}(t) - \mathbf{c}_0(t)\right), \tag{12}$$

and $\hat{\boldsymbol{\beta}}(t)$ is the OLS estimate (5). In particular, for the overall model hypotheses on $\mathcal{I}$

$$\begin{cases} H_{0,F}^{\mathcal{I}} : \beta_l(t) = 0, \quad \forall l \in 1, \ldots, L, \forall t \in \mathcal{I} \\ H_{1,F}^{\mathcal{I}} : \beta_l(t) \neq 0, \quad \text{for some } \ l \in \{1, \ldots, L\} \ \text{ and } \ t \in \mathcal{I}, \end{cases} \tag{13}$$

we use the following statistics:

$$T_F^{\mathcal{I}} = \int_{\mathcal{I}} \sum_{l=1}^{L} (\hat{\beta}_{l_n}(t))^2 \mathrm{d}t. \tag{14}$$

For the hypotheses of the lth functional regression parameter on $\mathcal{I}$

$$\begin{cases} H_{0,l}^{\mathcal{I}} : \beta_l(t) = 0, \ \forall t \in \mathcal{I} \\ H_{1,l}^{\mathcal{I}} : \beta_l(t) \neq 0, \ \text{for some } \ t \in \mathcal{I}, \end{cases} \tag{15}$$

we use the test statistics as follows:

$$T_l^{\mathcal{I}} = \int_{\mathcal{I}} \left(\hat{\beta}_{l_n}(t)\right)^2 \mathrm{d}t. \tag{16}$$

Note that we have chosen $T_F^{\mathcal{I}}$ and $T_l^{\mathcal{I}}$ to be the special cases of $T_C^{\mathcal{I}}$.

2.3.2 | Second step: Adjusted *p* value functions

Let $p_C^{\mathcal{I}}$ denote the p value of test (10) obtained using functional permutation tests based on the Freedman and Lane permutation scheme (Freedman & Lane, 1983), as defined by Pesarin and Salmaso (2010). It is based on the permutation of the estimated residuals under the reduced model, that is, the linear model constrained to the null hypothesis of the test. The p value of the permutation test is then obtained by calculating the proportion of permutations leading to a larger value of the test statistics than the test statistics on the observed data. This scheme is the most commonly used scheme for linear models and presents many advantages compared with other permutation techniques (Anderson & Legendre, 1999; Anderson & Robinson, 2001; Davison & Hinkley, 1997; Winkler et al., 2014; Zeng, Pan, MaWhinney, Barón, & Zerbe, 2011). In particular, it can be shown empirically that its power is typically higher than the power of tests based on other permutation schemes (Anderson & Legendre, 1999; Winkler et al., 2014). For more details, see Appendix B.

In order to identify significant subdomains, we make use of adjusted p value functions. The adjusted p value function $\tilde{p}_C(t)$ at point t for testing general linear hypotheses with contrast C is defined as the supremum p value of all intervalwise tests on intervals containing t, as follows:

$$\tilde{p}_C(t) = \sup_{\mathcal{I}: t \in \mathcal{I}} p_C^{\mathcal{I}}, \quad t \in [a, b]. \tag{17}$$

Analogously, denoting by $p_F^{\mathcal{I}}$, and $p_l^{\mathcal{I}}$ the p values from testing Equations 13 and 15, respectively, the adjusted p value functions for testing hypotheses on the overall model and on the lth functional parameter are defined as

$$\tilde{p}_F(t) = \sup_{\mathcal{I}: t \in \mathcal{I}} p_F^{\mathcal{I}}; \quad \tilde{p}_l(t) = \sup_{\mathcal{I}: t \in \mathcal{I}} p_l^{\mathcal{I}}, \quad t \in [a, b], \tag{18}$$

respectively.

Due to the nature of permutation tests used here, the adjusted p value functions are quantized continuous functions with the step size decreasing as the sample size n tends to infinity. The continuity of the limiting function is guaranteed by the continuity of the pointwise test statistics and of the observed functions $y_i(t)$.

2.3.3 | Third step: Domain selection

The intervals of the domain presenting a rejection of any of the null hypotheses are obtained by thresholding the corresponding adjusted p value functions at level α. For example, we select the intervals presenting at least one significant effect by thresholding $\tilde{p}_F(t)$ and intervals presenting a significant effect of the lth covariate by thresholding $\tilde{p}_l(t)$.

The introduced domain selection procedure is provided with a (asymptotic) control of the IWER. This type of control implies that the probability of detecting false positive intervals is (asymptotically) controlled at level α. The supporting theoretical results are presented in the following section.

3 | THEORETICAL RESULTS

Here, we present the theoretical properties of the permutation-based inference on functional-on-scalar linear models performed along the line depicted in Section 2. All proofs are reported in Appendix A.

First, we prove that the domain selection IWT procedure for testing general linear hypothesis is provided with an asymptotic control of the IWER. Pini and Vantini (2016) proved that, if all intervalwise tests used to build IWT are exact, the IWT is provided with a control of the IWER. This result can be applied directly in the case of the overall functional model test on the regression model but has to be extended in the more general case of tests on linear hypotheses (including tests on single functional parameters) because, in the latter case, the exactness of all tests is only asymptotical.

Theorem 1. *Under assumptions C1 and C2, the domain selection procedure for testing general functional linear hypotheses is provided with an asymptotic control of the IWER.*

Formally, the adjusted p value function $\tilde{p}_C(t)$ is such that, $\forall \alpha \in (0, 1]$,

$$\forall \mathcal{I} \subseteq [a, b] \,:\, H_{0,C}^{\mathcal{I}} \text{ is true } \Rightarrow \limsup_{n \to \infty} \mathbb{P}\left[\forall t \in \mathcal{I}, \tilde{p}_C(t) \leq \alpha\right] \leq \alpha.$$

Because tests on single functional parameters are specific cases of linear hypotheses, we obtain directly the following corollary.

Corollary 1. *Under assumptions C1 and C2, the domain selection procedure for testing hypotheses for single functional regression parameter is provided with an asymptotic control of the IWER.*

Furthermore, the following proposition provides exact results for IWT-based overall functional model tests.

Proposition 1. *The domain selection procedure for testing the overall functional model hypotheses is provided with a control of the IWER. Formally, the adjusted p value function $\tilde{p}_F(t)$ is such that, $\forall \alpha \in (0,1]$,*

$$\forall \mathcal{I} \subseteq [a,b] : H_{0,F}^{\mathcal{I}} \text{ is true } \Rightarrow \mathbb{P}[\forall t \in \mathcal{I}, \tilde{p}_F(t) \leq \alpha] \leq \alpha.$$

Next, we focus on the property of consistency of the proposed tests. The following theorem states that the probability of detecting every interval $\mathcal{I}$, such that $\forall\, t \in \mathcal{I}\ H_{0,C}^t$ is false, converges to 1 as the sample size increases. This property implies that the probability of truly detecting any point where the null hypothesis is false converges to 1.

Theorem 2. *The domain selection procedure for testing general functional linear hypotheses is consistent. Formally, $\forall \alpha \in (0,1]$ the adjusted p-value $\tilde{p}_C(t)$ is as follows:*

$$\forall \mathcal{I} \subseteq [a,b] \text{ such that } H_{0,C}^t \text{ is false } \forall t \in \mathcal{I} \Rightarrow \lim_{n\to\infty} \mathbb{P}[\forall t \in \mathcal{I}, \tilde{p}_C(t) \leq \alpha] = 1.$$

As a consequence, we also obtain consistency results for the IWT-based tests of the overall functional model and on single functional parameters.

Corollary 2. *The domain selection procedure for the overall functional model hypotheses is consistent.*

Corollary 3. *The domain selection procedure for single functional parameter hypotheses is consistent.*

4 | DETAILS ON THE IMPLEMENTATION

In practice, evaluations of the adjusted p value functions, even at one fixed point $t \in [a,b]$, are practically infeasible because it involves taking supremum over uncountably many subintervals covering point t. We therefore propose a numerical procedure resulting in pointwise constant approximation of the p value functions and the significance regions chosen accordingly. The approximation error of the procedure is dependent on two parameters. The first parameter, $K \in \mathbb{N}$, defines the size of the initial partition of interval $[a,b]$. The second parameter, $B \in \mathbb{N}$, determines the number of random permutations used in the conditional Monte Carlo algorithm (Pesarin & Salmaso, 2010) to approximate the intervalwise permutation tests p values.

The following steps describe the procedure for linear hypotheses of the functional parameters with combination matrix C.

Step 1 Partition the domain $[a,b]$ into K equally sized atomic subintervals of length $\Delta t = (b-a)/K$, that is,

$$\mathcal{P}_i = [a + (i-1)\Delta t, a + i\Delta t] \qquad i = 1, \ldots, K.$$

Let $\mathcal{S}$ be the family (of size $K(K+1)/2$) of all possible *intervals* constructed from the K atomic subintervals.

Step 2 For $i = 1, \ldots, K$, approximate the value of the integrated test statistics $T_C^{\mathcal{P}_i}$ by

$$\widehat{T_C^{\mathcal{P}_i}} = T_C\left(a + (i-1)\Delta t\right)\Delta t.$$

This is equivalent to approximating the pointwise test statistics $T_C(t)$ in atomic subintervals P_i with a constant equal to $T_C(a + (i-1)\Delta t)$, that is, the value of the statistics in the left endpoint of the interval P_i.

Step 3 For all intervals $\mathcal{J} \in \mathcal{S}$, approximate the value of the integrated test statistics by

$$\widehat{T_C^{\mathcal{J}}} = \sum_{\mathcal{P}_i \subset \mathcal{J}} \widehat{T_C^{\mathcal{P}_i}}.$$

The approximations of the integrated test statistics correspond to a rectangle quadrature with step Δt applied to the pointwise test statistics $T_C(\cdot)$.

Step 4 Estimate the p values $\widehat{p_C^{\mathcal{J}}}$ of the tests on each $\mathcal{J} \in \mathcal{S}$ using the Freedman and Lane permutation scheme (see Appendix B) with B randomly chosen permutations.

Step 5 The adjusted p value function at point t is estimated by the maximum of the corresponding p values of all intervals $\mathcal{J}$ in $\mathcal{S}$ containing t, that is,

$$\widehat{\tilde{p}_C}(t) = \max\left\{ \widehat{\tilde{p}_C^{\mathcal{J}}} \ : \ \mathcal{J} \in \mathcal{S} \text{ such that } t \in \mathcal{J} \right\}.$$

Furthermore, by construction, $\widehat{\tilde{p}_C}(t)$ is constant at each atomic subinterval, and therefore, the estimate of $\tilde{p}_C(t)$ for all $t \in [a, b]$ is obtained by evaluating the above maximum K times (one time per $\mathcal{P}_1, \ldots, \mathcal{P}_K$).

Step 6 Select the significant domain subset by thresholding the obtained approximated adjusted p value function.

Observe that, as K tends to infinity, the error arising due to the discretization of the domain becomes negligible. This is guaranteed by the integrability of the pointwise test statistics $T_C(t)$ and resulting continuity of the integrated test statistics with respect to the integration limits. Furthermore, as the number of simulated permutations B increases, the Monte Carlo error tends to zero. It is worth to notice that in Step 4, we use the same B permutations for each subinterval, which decreases the computational complexity of the algorithm.

5 | ANALYSIS OF KNEE KINEMATICS

We will now report the results of the analysis of the knee kinematics data briefly presented in the introduction. To investigate potential long-term differences in movement patterns following ACL injury, individuals treated with either reconstructive surgery in combination with physiotherapy (ACL_R) or physiotherapy alone (ACL_PT) were tested in a motion laboratory. The resulting functional data consist of the joint angle motion in different movement planes during a one-leg hop, sampled at 240 Hz.

The functional data set that we analyze here corresponds to the knee movement in the sagittal plane, that is, knee flexion/extension (see Figure 1) during a one-leg hop for distance. Our primary focus is to compare (ACL_R) and (ACL_PT) individuals with healthy-knee controls (CTRL) matched with age and sex, taking individual-specific covariates into account. The three groups are only matched on group level (Hébert-Losier et al., 2015), and hence, age and sex are included as possible covariates together with available data about jump length and body mass index (BMI). The knee movement in the sagittal plane for the hop of maximal length (out of three attempts)

on the injured leg for the two ACL groups was compared with the nondominant leg for controls. In the case of group differences, it is of clinical interest to find out which time intervals of the jump these differences occur. Functional data were recorded at 240 Hz using a calibrated 8-camera 3D motion analysis system. Kinematic parameters were calculated using rigid-body analysis and Euler angles. Knee-joint angles (°) were computed using an *x*-*y*-*z* Cardan sequence equivalent to the knee-joint coordinate system. The knee-joint kinematic curves in the sagittal plane of motion were extracted and smoothed through a functional B-spline basis with 150 equally spaced knots. In order to take into account the different time lengths, data were aligned by means of four landmarks: the event of maximal knee flexion before takeoff, the takeoff event (the instant at which the foot leaves the ground), the touchdown event (the time instant at which the foot touches the ground), and the event of maximal knee flexion after landing. The aligned data and the landmarks are displayed in the right panel of Figure 1. For more details on data collection and preprocessing, we refer to the work of Hébert-Losier et al. (2015). The hop consists of three phases: time preceding the takeoff event (takeoff phase), time between the takeoff event and the landing event (flight phase), and time succeeding the landing event (landing phase).

The initial model we use to describe the knee-joint angle motion in the sagittal plane, $y_i(t)$, is the following:

$$\begin{aligned} y_i(t) = \beta_0(t) + \beta_{\text{Jump}}(t)x_{\text{Jump},i} + \beta_{\text{BMI}}(t)x_{\text{BMI},i} + \beta_{\text{Age}}(t)x_{\text{Age},i} \\ + \beta_{\text{Sex}}(t)x_{\text{Sex},i} + \beta_{\text{CTRL}}(t)x_{\text{CTRL},i} + \beta_{\text{R}}(t)x_{\text{R},i} + \epsilon_i(t), \quad i = 1, \ldots, 95, \end{aligned}$$

where the covariates $x_{\text{Jump},i}$, $x_{\text{BMI},i}$, and $x_{\text{Age},i}$ indicate the jump length, BMI, and age, respectively, of individual i. Furthermore, $x_{\text{Sex},i}$, $x_{\text{CTRL},i}$, and $x_{\text{R},i}$ are indicator functions attaining the value of 1 if individual i is a man in the control group and in the reconstructive surgery group, respectively, and otherwise 0.

As in classical multiple regression, we start with the initial model using all available covariates and then use backward elimination to reduce it to a final model that includes only the significant covariates. We start by performing a global overall significance test that simultaneously tests if at least one of the coefficients is significant somewhere on the domain; compare (6). Separate global tests are then performed for each coefficient as in (7). Table 1 presents the p values obtained by performing permutation tests with $B = 1000$ and test statistics Equations 14 and 16, where the interval $\mathcal{I}$ corresponds to the whole domain. Starting from these results, we apply the backward elimination procedure and stepwise remove the covariates with the largest p value (one at a time, and re-estimate the reduced model) until only the significant coefficients remain in the model. We use a 5% significance level at all steps of the procedure. The final model includes the continuous covariate Jump and the group indicators. The corresponding p values are presented in Table 2.

After assessing the overall significance of the covariates on the whole domain, we apply the introduced IWT-based procedure to the study in which parts of the domain, the coefficients, are significant. To illuminate all between-group differences, we use suitable combination matrices and introduce an additional hypothesis comparing the performance of the control and

TABLE 1 Global p values of the *initial model* for the overall test (6) and for each single functional parameter test (7) obtained by performing permutation tests with $B = 1{,}000$ based on test statistics in Equations 14 and 16, respectively

Test	Overall	β_0	β_{Jump}	β_{BMI}	β_{Sex}	β_{Age}	β_{CTRL}	β_{R}
p value	0.000	0.000	0.000	0.302	0.277	0.327	0.000	0.037

TABLE 2 Global p values of the reduced *final model* for the overall test (6) and for each single functional parameter test (7) obtained by performing permutation tests with $B = 1{,}000$ based on test statistics in Equations 14 and 16, respectively

Test	Overall	β_0	β_{Jump}	β_{CTRL}	β_{R}
p value	0.000	0.000	0.000	0.000	0.029

reconstructive surgery groups (see below). The adjusted p value functions are computed according to Section 4 with $K = 300$ and $B = 1000$.

Functional parameter estimates together with estimated adjusted p value functions for the overall functional model and single functional parameter tests are given for the final model in Figure 2. For illustrative purposes, we also present the results of the ITW-based procedure for the full model in Appendix C.

The first row in Figure 2 shows the individual knee-joint angle kinematics curves (left) and, in black solid line, the estimated adjusted p value functions for the overall functional tests (right), indicating the presence of at least one significant effect in the majority of the jump. The gray-shaded parts of the domain (left) correspond to significant effects at the 5% level (i.e., the points t with associated adjusted p values ≤ 0.05, indicated with a dashed red line in the right panel).

As expected, the jump length has a significant effect throughout all three jump phases, in the majority of the domain (Figure 2, row 3). The functional parameter estimate $\hat{\beta}_{\text{Jump}_n}(t)$ is positive in intervals containing the maximal flexion during both takeoff and landing and negative directly after takeoff and just before touchdown, confirming that the movement is more pronounced for longer jumps. The parts of the domain where the effect is nonsignificant are expected because $\beta_{\text{Jump}}(\cdot)$ is anticipated to change sign.

The last three rows of Figure 2 present results of group comparisons, testing respectively

$$H_{0,\text{CTRL}-\text{PT}} : \beta_{\text{CTRL}}(t) = 0,\ H_{1,\text{CTRL}-\text{PT}} : \beta_{\text{CTRL}}(t) \neq 0,$$
$$H_{0,\text{R}-\text{PT}} : \beta_{\text{R}}(t) = 0,\ H_{1,\text{R}-\text{PT}} : \beta_{\text{R}}(t) \neq 0,$$
$$H_{0,\text{CTRL}-\text{R}} : \beta_{\text{CTRL}}(t) - \beta_{\text{R}}(t) = 0,\ H_{1,\text{CTRL}-\text{R}} : \beta_{\text{CTRL}}(t) - \beta_{\text{R}}(t) \neq 0,$$

with the parameter estimates to the left and corresponding adjusted p value functions to the right. The ACL_{PT} group is significantly different with respect to the control group during both takeoff and landing (Figure 2, row 4). The functional parameter estimate $\hat{\beta}_{\text{CTRL}_n}(\cdot)$ associated to the differences indicates less flexion in the ACL_{PT} group during these two phases compared with individuals in the control group because there, $\hat{\beta}_{\text{CTRL}_n}(t)$ is positive. These results are in line with previously reported results (Hébert-Losier et al., 2015; Tengman et al., 2015), indicating significant differences between physiotherapy-treated individuals and controls. We do not detect any significant differences between the reconstructive surgery group and the controls (Figure 2, row 6). The overall test of $\beta_{\text{R}}(\cdot)$ indicates that the reconstructive surgery group is significantly different from the ACL_{PT} group (Table 2). However, there is no sufficient evidence to identify which parts of the domain the significant differences between the two groups occur (the corresponding adjusted p value function never goes below the 5% level); compare Figure 2, row 5.

We observe significant group differences between 0% and 56% of the takeoff phase and between 36% and 100% of the landing phase. We thus validate the clinical hypotheses that the preparation of the jump in the takeoff phase and the stabilization in the landing phase are of particular interest in relation to movement control after injury. Our analysis confirms that the

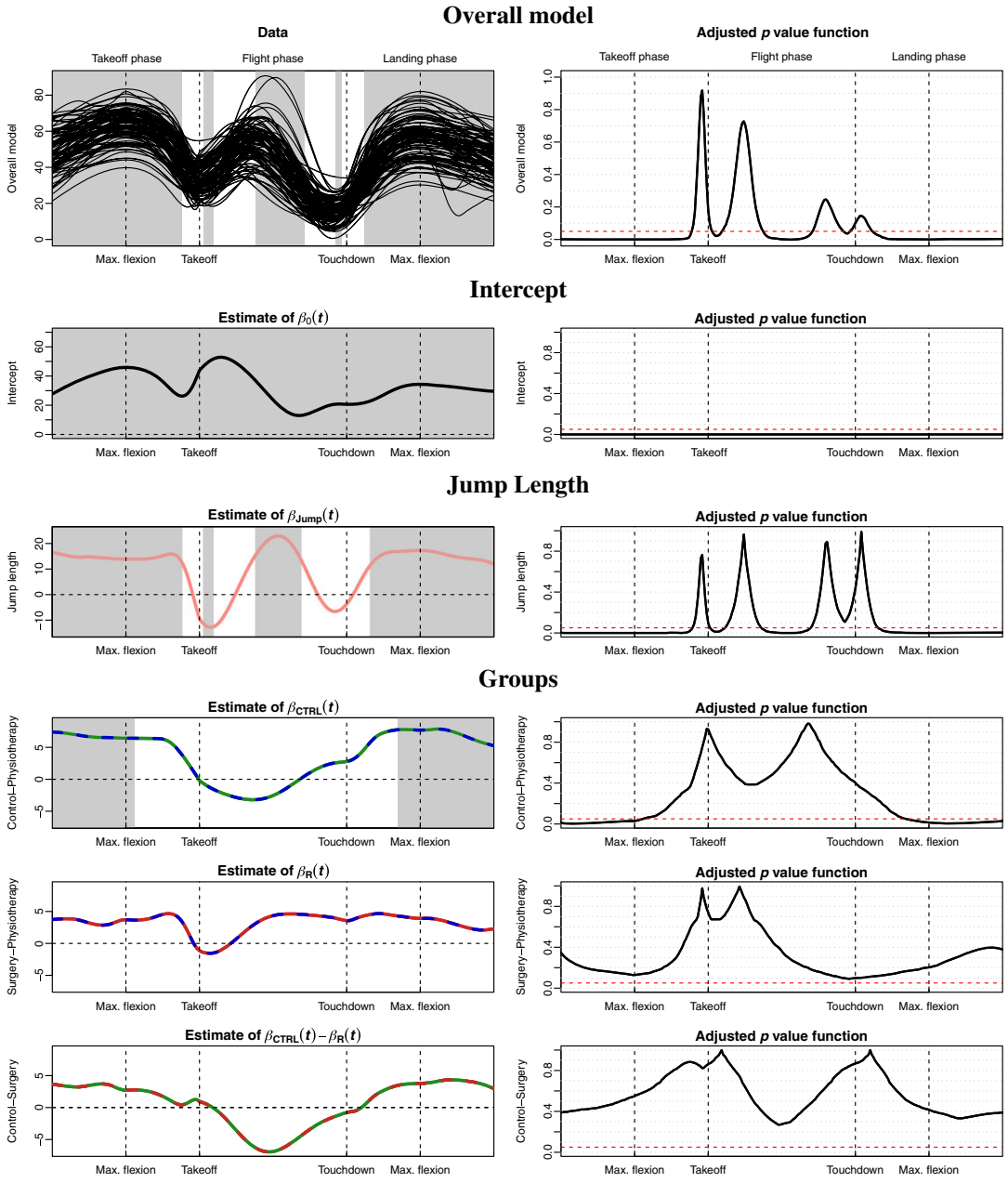

FIGURE 2 (Left) functional parameter estimates and (right) estimated adjusted *p* value functions for the final (reduced) functional-on-scalar model. Top panel: (left) knee-joint angle kinematics data (functional response variable) for the 95 individuals and (right) estimated adjusted *p* value function for the overall functional model tests. Bottom panels: (left) functional parameter estimates and (right) corresponding estimated adjusted *p* value functions for the intercept, jump length, and group effects, respectively. (Left) Gray-shaded parts correspond to significant effects at the 5% level, that is, the points *t* with associated adjusted *p* values less or equal to 0.05. (Right) Black solid lines correspond to intervalwise-testing-adjusted *p* value; (right) red dashed horizontal line corresponds to 5% threshold [Colour figure can be viewed at wileyonlinelibrary.com]

events of maximal flexion, analyzed for example in the work of Tengman et al. (2015), provides some insight into how the groups may differ. However, basing only on analyses on comparisons of movement data taken at one particular point in time provides limited information, and with the present procedure, we are able to detect significant differences in large parts of the takeoff and landing phases.

As a complement to the analysis of the kinematics data within the IWT framework, in Appendix D, we report the results with the $F_{\max}$ procedure applied (see Section 6 for details about the implementation of $F_{\max}$). The $F_{\max}$ procedure provides very similar results to the ones obtained with our IWT-based procedure, corroborating the results discussed in this section.

6 | SIMULATION STUDY

In this section, we report the results of a simulation study carried out to investigate the behavior of the IWT procedure and also to make a numerical comparison between the IWT procedure and the $F_{\max}$ procedure (Holmes et al., 1996; Winkler et al., 2014). Both are nonparametric methods for performing inference on linear models, and they can be used to perform domain selection. However, it is important to note that, methodologically, the two procedures differ because the $F_{\max}$ procedure is developed for vector-on-scalar linear models, that is, models where the response is a vector of finite dimension and the covariates are scalar.

6.1 | $F_{\max}$ procedure

We start by providing a brief description of the $F_{\max}$ procedure. When applied to vector-on-scalar linear models in a multivariate setting, the procedure of computing the adjusted p value function can be summarized in the following steps.

Step 1 The overall F-test statistics is computed for every component of the response vector.
Step 2 Freedman and Lane permutations (see Appendix B) are applied jointly to all components of the response vector.
Step 3 For each permutation, the F-test statistics is computed again for each component, and the maximum of the permuted statistics over all components is computed.
Step 4 For each component, the F-test statistics based on the observed data is compared with the permutational distribution of its maximum. The adjusted p value associated to each component is the proportion of permutations where the maximum test statistics is higher or equal to the observed test statistics.

Further details on construction, properties, and implementation of the $F_{\max}$ procedure can be found in the work of Winkler et al. (2014). One of the eminent properties making the comparison of our procedure and $F_{\max}$ especially interesting is the strong control of the FWER.

The procedure extends to functional-on-scalar linear models by applying the test to discrete pointwise evaluations of the functional data. In detail, the functional response is discretized on a grid, and a vector-on-scalar linear model is fitted for the discretized response. The pointwise test statistics is compared with the permutational distribution of the maximum of such statistics over the discretized domain.

6.2 | Simulation study setting

We simulate functional data according to the following intercept-free linear model:

$$y_i(t) = \beta_1(t)x_{1i} + \varepsilon_i(t) \quad i = 1, \ldots, n, t \in [0, 1],$$

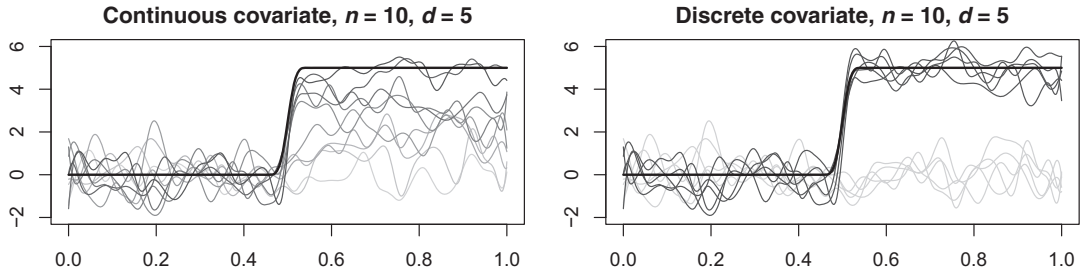

FIGURE 3 One instance of simulated data with parameters $n = 10$ and $d = 5$ and (left) continuous or (right) discrete covariate. Gray levels of the curves are associated to the covariate values (dark: high value; light: low value). The black bold line indicates the functional term $d \cdot f(t)$

where $\beta_1(t) = d \cdot f(t)$, $d \in \mathbb{R}$. The function $f(t)$ is a fixed curve obtained with a cubic B-spline expansion with 40 basis functions. The first 20 coefficients of the expansion are set to zero, and the last 20 coefficients are set to one. The resulting function is equal to zero in the first part of the domain ($t \in [0, 17/37]$) and to one in the second part of the domain ($t \in [20/37, 1]$), with a continuous and differentiable transition between the two parts. With such a definition of $f(t)$, the parameter d is a magnitude parameter that accentuates the effect of the covariate. The error functions $\varepsilon_i(t)$ are simulated with the same cubic B-spline basis. The coefficients of the basis expansion are sampled independently from a standard normal distribution. We investigate the following two cases for the covariate type.

1. Continuous covariate:

$$x_{1i} = \frac{i-1}{n-1}, \qquad i = 1, \ldots, n.$$

2. Discrete covariate:

$$x_{1i} = \begin{cases} 0 & i = 1, \ldots, n/2 \\ 1 & i = n/2+1, \ldots, n. \end{cases}$$

When evaluating performance, we consider the raster of values $d = 0, 0.5, 1, \ldots, 5$ together with three different sample sizes $n = 10, 20, 40$. We perform the overall tests of $H_{0,F}$ against $H_{1,F}$ (that, in this case, coincides with the test on $\beta_1(t)$) with the procedure proposed in this paper and with the $F_{\max}$ procedure. In the following, we denote by D_0 the interval of the domain where $H_{0,F}$ is true and with $D_1 = D \backslash D_0$ the interval of the domain where $H_{0,F}$ is false. For performing the comparison, the same discretization of functional data based on 50 pointwise evaluations is used for both procedures. An instance of the simulated functional data with $n = 10$, $d = 5$, and a continuous or discrete covariate are displayed in Figure 3. The figure displays the functional data in gray with the intensity of the gray color being proportional to the value of the covariate. The term $d \times f(t)$ is displayed in both panels with a black bold line.

6.3 | Performance measures

Let $\tilde{p}(t)$ denote the adjusted p value function for the F test resulting from the inferential procedure. We investigate the behavior of the two procedures in terms of the following quantities (as functions of d).

1. The estimated FWER, that is, the probability of rejecting $H_{0,F}$ for at least one point of D_0 (where the null hypothesis is true):

$$\text{FWER} = \mathbb{P}[\exists t \in D_0 \ : \ \tilde{p}(t) \leq \alpha].$$

2. The estimated power, that is, the probability of rejecting $H_{0,F}$ for at least one point of D_1 (where the null hypothesis is false):

$$\text{Power} = \mathbb{P}[\exists t \in D_1 \ : \ \tilde{p}(t) \leq \alpha].$$

3. The estimated sensitivity, that is, the average measure of the correctly detected domain over the measure of the domain where the null hypothesis is false:

$$\text{Sensitivity} = \mathbb{E}\left[\frac{|t \in D_1 \ : \ \tilde{p}(t) \leq \alpha|}{|D_1|}\right],$$

where $|\cdot|$ denotes the Lebesgue measure.

All three quantities are estimated for both procedures with a Monte Carlo simulation based on 1,000 randomly generated data sets.

6.4 | Results

The results of the simulations for continuous and discrete covariate types are shown in Figure 4 (first and second row, respectively). First of all, note that, as supported by theory, both procedures control the FWER for all values of d and for all sample sizes. Furthermore, as expected, increasing d results in higher power and sensitivity for both procedures. The results in terms of power and

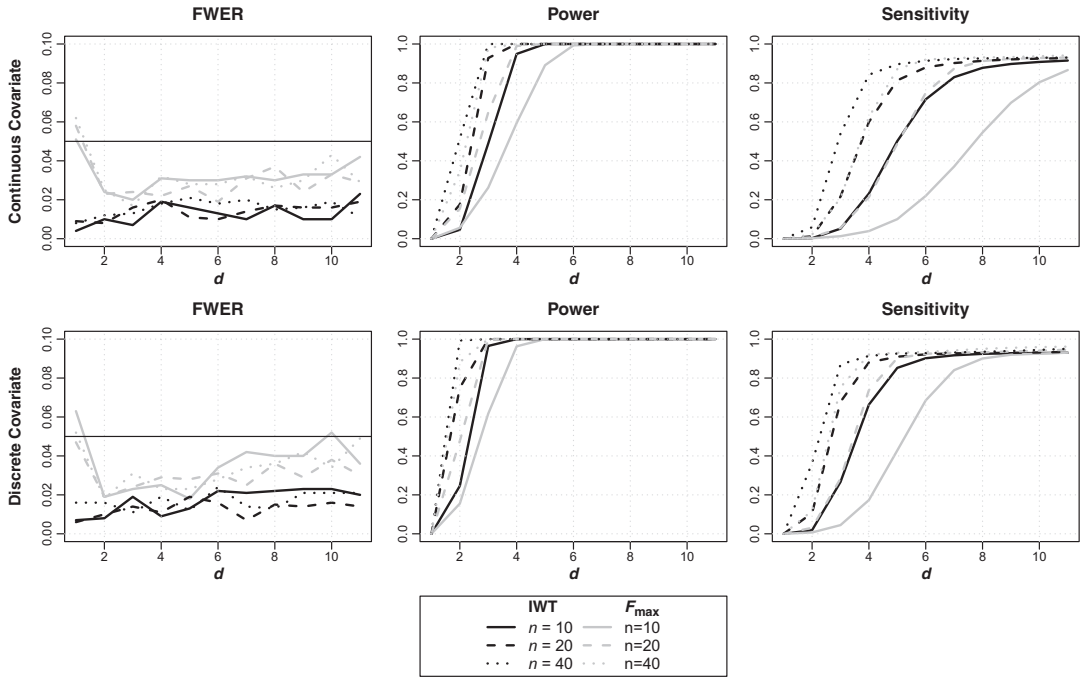

FIGURE 4 Results of numerical comparison between intervalwise testing (IWT) and $F_{\max}$ based on 1,000 randomly simulated data sets for different values of d and n. Top: continuous covariate; bottom: discrete covariate. Left: estimated familywise error rate (FWER); center: estimated power; right: estimated sensitivity

sensitivity indicate that (in the case of our data) the IWT procedure appears to be more powerful and more precise with respect to the identification of the interval where the null hypothesis is violated. This is in line with the fact that the $F_{\max}$ is provided with a strong control of the FWER, whereas the IWT control is only intervalwise. When investigating the figure, we also see that, as the sample size increases, the performance of the two procedures in terms of both power and sensitivity becomes similar.

In addition to presented results, we carried out several additional simulations by changing some of the parameters of the generative model and of the simulation itself. For brevity, we do not report the full results here but provide a short summary. When varying the signal-to-noise ratio or increasing the number of coefficients of the basis expansion, we have found that the IWT generally appears to be more powerful than $F_{\max}$ in noisy situations. Conversely, in less noisy situations, the performance of the two procedures becomes more similar. Furthermore, we have seen that introducing a positive or negative correlation between the coefficients in the error term does not alter the results significantly.

7 | DISCUSSION AND CONCLUSIONS

In this work, we have introduced a nonparametric methodology to test the functional parameters of a functional-on-scalar linear model with fixed effects. We provide IWT procedures based on permutations to test hypotheses on the functional regression parameters, including domain selection. We show that our proposed IWT-based procedure is asymptotically exact and consistent.

Due to the nonparametric nature of the testing procedures that we propose, the test statistics in Equations 11, 14, and 16 can be replaced by the integrated versions of other feasible pointwise test statistics, given that they are continuous functions on $[a, b]$. The continuity of the pointwise test statistics with respect to t is required to guarantee that the numerical procedure described in Section 4 provides a proper estimate of the adjusted p value functions. If integrable pointwise test statistics are to be used, more sophisticated numerical algorithms that can deal with improper integration need to be used to estimate the adjusted p value functions.

The IWT procedure used in this paper provides control of the IWER. As shown by Pini and Vantini (2016), the control can also be extended to the complementary sets of all intervals. In our simulation study, we report a numerical comparison with the $F_{\max}$ procedure described in the work of Winkler et al. (2014), which is provided with a strong control of the FWER. The results indicate that the IWT and $F_{\max}$ procedures perform similarly in terms of power and sensitivity with the IWT being superior for smaller sample sizes and more noisy data.

The analysis of the knee kinematics data set showed that the effect of jump length on knee kinematics is significantly different from zero, whereas the effects of BMI, sex, and age are not. In line with previous findings, even after having been discounted for the jump length, the physiotherapy group remains significantly different with respect to the control group during takeoff and landing. Our detected significant domain segments confirm the importance of the landmarks analyzed earlier by Tengman et al. (2015) in the problem of identifying group differences, simultaneously indicating statistical differences on wider segments of the time domain.

When performing backward or forward selection of significant covariates in a scalar-on-scalar linear model, the order in which covariates are removed or added to the model can influence the final result. In the case of functional-on-scalar linear models discussed in this paper, the situation could be even more difficult because the adjusted p value for each covariate changes along the domain. Nevertheless, note that backward selection is not the only possibility for selecting significant covariates. For instance, it is straightforward to define a functional coefficient of

determination (R^2) and use it as an index to perform model selection. Also, in this case, selection of covariates would not be straightforward because a functional R^2 will also change along the domain. In the case study presented in the current paper, this was not an issue because the results of the full model (reported in the Appendix) and the ones of the reduced model lead to similar conclusions. In general, a careful study of the results of all steps of backward or forward selection is advised.

Estimation and IWT of functional-on-scalar linear models are of interest in many recent applications. For instance, it could be applied for comparing pulmonary volume over time of different individuals, such as the data analyzed by Fogarty and Small (2014), for comparing hemolysis curves—the percent hemolysis as a function of time—at various treatment levels (Vsevolozhskaya et al., 2014), or for modeling the functional connectivity between brain regions as a function of distance between the regions, as in the work of Reiss et al. (2010). In all the mentioned cases, the methodology proposed in this work would additionally allow the selection of the intervals of the domain (i.e., time or space intervals) presenting significant effects of the covariates.

Even though the technique presented in this paper is meant to deal with functional data without requiring any basis expansion, rarely functional data are observed continuously. Usually, they are instead given as discrete observations of some variables. If the approximation described in Section 4 requires another resolution than the one of the raw data, or if the data are irregularly sampled, preprocessing of functional data through a basis expansion is needed. In such a case and in the presence of noisy observations, smoothing has to be performed carefully because both oversmoothing and undersmoothing the data would have as consequence a loss of power of the inferential procedure.

Because IWT aims at selecting the parts of the domain that are responsible for a rejection of a null hypothesis, the registration of functional data will also impact the results. Being a local technique, the IWT essentially compares functional data in terms of amplitude along the domain, implicitly assuming that the functions are without phase variation. In the presence of phase variation, registration is a crucial step of preprocessing because it makes the comparison in terms of amplitude between the different curves possible. An extensive study on the impact of the preprocessing (including both smoothing and registration) of functional data on the IWT is beyond the scope of this paper, but it would be of clear interest for future research.

Finally, it would be of interest to extend the domain-selective inference described here to functional-on-functional linear models. In such a framework, the functional regression coefficients $\beta_l(t)$ can be replaced by functional linear operators, and the concept of intervalwise inference has to be extended accordingly. Such a model would allow us to introduce the effects of time-varying covariates on the functional responses and could be of interest in applications where the covariates also change in time. Another possible extension of the methodology proposed in this paper would be to incorporate random effects in the model.

ACKNOWLEDGEMENTS

Part of this work was funded by the Swedish Scientific Research Council (Häger grant K2014-99X-21876-04-4, Sjöstedt de Luna grant 340-2013-5203, and Schelin grant 2016-02763) and by the Västerbotten County Council (Häger grant ALF VLL548501 and Strategic funding VLL-358901 through Project 7002795).

ORCID

Alessia Pini http://orcid.org/0000-0001-9235-3062

REFERENCES

Abramovich, F., & Angelini, C. (2006). Testing in mixed-effects FANOVA models. *Journal of Statistical Planning and Inference*, *136*, 4326–4348.

Abramowicz, K., Häger, C., Hébert-Losier, K., Pini, A., Schelin, L., Strandberg, J., & Vantini, S. (2014). An inferential framework for domain selection in functional ANOVA. In E. G. Bongiorno, E. Salinelli, A. Goia, & P. Vieu (Eds.), *Contributions in infinite-dimensional statistics and related topics* (pp. 13–18). Bologna, Italy: Società Editrice Esculapio.

Anderson, M. J., & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, *62*, 271–303.

Anderson, M. J., & Robinson, J. (2001). Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*, *43*, 75–88.

Antoniadis, A., & Sapatinas, T. (2007). Estimation and inference in functional mixed-effects models. *Computational Statistics & Data Analysis*, *51*, 4793–4813.

Cardot, H., Prchal, L., & Sarda, P. (2007). No effect and lack-of-fit permutation tests for functional regression. *Computational Statistics*, *22*, 371–390.

Cuesta-Albertos, J. A., & Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data. *TEST*, *19*, 537–557.

Cuevas, A., Febrero, M., & Fraiman, R. (2004). An ANOVA test for functional data. *Computational Statistics & Data Analysis*, *47*, 111–122.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

Fan, J., & Zhang, J. T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *62*, 303–322.

Fogarty, C. B., & Small, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Annals of Applied Statistics*, *8*, 2002–2026.

Freedman, D., & Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics*, *1*, 292–298.

Gertheiss, J., Goldsmith, J., Crainiceanu, C., & Greven, S. (2013). Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics*, *14*, 447–461.

Hébert-Losier, K., Pini, A., Vantini, S., Strandberg, J., Abramowicz, K., Schelin, L., & Häger, C. K. (2015). One-leg hop kinematics 20 years following anterior cruciate ligament ruptures: Data revisited using functional data analysis. *Clinical Biomechanics*, *30*, 1153–1161.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1996). Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow and Metabolism*, *16*, 7–22.

Kayano, M., Matsui, H., Yamaguchi, R., Imoto, S., & Miyano, S. (2015). Gene set differential analysis of time course expression profiles via sparse estimation in functional logistic model with application to time-dependent biomarker detection. *Biostatistics*, *17*, 235–248.

Lai, T. L., Robbins, H., & Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression. *Journal of Multivariate Analysis*, *9*, 343–361.

Marcus, R., Eric, P., & Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, *63*, 655–660.

Pesarin, F., & Salmaso, L. (2010). *Permutation tests for complete data: Theory, applications and software*. Chichester, UK: John Wiley & Sons Inc.

Pini, A., & Vantini, S. (2016). The interval testing procedure: A general framework for inference in functional data analysis. *Biometrics*, *72*, 835–845.

Pini, A., & Vantini, S. (2017). Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, *29*, 407–424.

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

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

Reiss, P. T., Huang, L., & Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *International Journal of Biostatistics*, *6*, 1–28.

Schott, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. *Journal of Multivariate Analysis*, *98*, 1825–1839.

Spitzner, D. J., Marron, J. S., & Essick, G. K. (2003). Mixed-model functional ANOVA for studying human tactile perception. *Journal of the American Statistical Association*, *98*, 263–272.

Staicu, A.-M., Li, Y., Crainiceanu, C. M., & Ruppert, D. (2014). Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian Journal of Statistics*, *41*, 932–949.

Tengman, E., Grip, H., Stensdotter, A. K., & Häger, C. K. (2015). Anterior cruciate ligament injury about 20 years post-treatment: A kinematic analysis of one-leg hop. *Scandinavian Journal of Medicine & Science in Sports*, *25*, 818–827.

Vsevolozhskaya, O. A., Greenwood, M. C., Bellante, G. J., Powell, S. L., Lawrence, R. L., & Repasky, K. S. (2013). Combining functions and the closure principle for performing follow-up tests in functional analysis of variance. *Computational Statistics & Data Analysis*, *67*, 175–184.

Vsevolozhskaya, O., Greenwood, M., & Holodov, D. (2014). Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis. *Annals of Applied Statistics*, *8*, 905–925.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, *92*, 381–397.

Zeng, C., Pan, Z., MaWhinney, S., Barón, A. E., & Zerbe, G. O. (2011). Permutation and *F* distribution of tests in the multivariate general linear model. *The American Statistician*, *65*, 31–36.

Zhang, J., & Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise *F*-test. *Scandinavian Journal of Statistics*, *41*, 51–71.

How to cite this article: Abramowicz K, Häger CK, Pini A, Schelin L, Sjöstedt de Luna S, Vantini S. Nonparametric inference for functional-on-scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament. *Scand J Statist*. 2018;1–26. https://doi.org/10.1111/sjos.12333

APPENDIX A: PROOFS

We here provide the proofs of the theorems stated in Section 3. We first report the theoretical properties of the functional intervalwise tests in Equations 10, 13, and 15 based on the Freedman and Lane scheme and on integrated test statistics in Equations 12, 14, and 16. Then, we prove that the IWT-based tests of linear hypotheses on the functional-on-scalar linear model are provided with an asymptotic control of the IWER and that they are consistent. Additionally, we show that the IWT-based *F* test on the regression model is provided with an exact control of the IWER.

A.1 | Intervalwise tests

We start by showing asymptotic exactness of functional tests on linear hypotheses for a given interval $\mathcal{I}$.

Lemma 1. *Under assumptions C1 and C2 and for each interval* $\mathcal{I} \subseteq [a, b]$, *the functional test of linear hypotheses on the regression parameters (10) based on statistics* $T_C^{\mathcal{I}}$ *(11) is asymptotically exact.*

Proof. Let $H_{0,C}^{\mathcal{I}}$ hold, that is, $C\boldsymbol{\beta}(t) = \mathbf{c}_0(t), \forall t \in \mathcal{I}$. Under the null hypothesis and for any $t \in \mathcal{I}$, the model can be reduced by solving the linear system $C\boldsymbol{\beta}(t) = \mathbf{c}_0(t)$. In particular, because C has full rank, $q \leq L + 1$ regression parameters can be removed from the model. Let $\mathcal{Q}$ denote

the set of indexes removed. The reduced model is then $y_i(t) = \sum_{r\notin Q}\beta_r(t)a_r(t)x_{ri} + \varepsilon_i(t)$, where $x_{0i} = 1$, $a_r(t)$ is a fixed known function (depending only on the solution of linear systems $C\boldsymbol{\beta}(t) = \mathbf{c}_0(t)$) and $\varepsilon_i(t)$ are i.i.d. and zero-mean random functions.

The generalization to the functional case of the Freedman and Lane permutation scheme is based on the joint permutations (the same for each $t \in \mathcal{I}$) of the residuals $\hat{\varepsilon}_{i,C}(t) = y_i(t) - \sum_{r\notin Q}\hat{\beta}_{r,C_n}(t)a_r(t)x_{ri}$, where $\hat{\beta}_{r,C_n}(t)$, $r \notin \mathcal{Q}$ are the OLS estimates of the parameters $\beta_r(t)$, $r \notin \mathcal{Q}$ under the reduced model. Under conditions *C1* and *C2*, we have strong consistency of the OLS parameters estimates, that is, in our case, $\hat{\beta}_{r,C_n}(t)\xrightarrow{a.s.}\beta_r(t)$, $\forall r \notin \mathcal{Q}$, and $\forall t \in \mathcal{I}$. Hence, we also have the strong convergence of the residuals, that is, $\hat{\varepsilon}_{i,C}(t)\xrightarrow{a.s.}\varepsilon_i(t)$, $\forall i = 1, \ldots, n$ and $\forall t \in \mathcal{I}$.

The errors $\varepsilon_i(t)$ of the linear model are jointly exchangeable. Hence, the likelihood of every joint permutation is invariant and equal to $1/n!$. Therefore, the test based on the joint permutations of the errors $\varepsilon_i(t)$ is exact. As $\hat{\varepsilon}_{i,C}(t)\xrightarrow{a.s.}\varepsilon_i(t)$, $\forall t \in \mathcal{I}$, the residuals are jointly asymptotically exchangeable, that is, the likelihood of every joint permutation is asymptotically invariant and converges to $1/n!$. Hence, the test based on joint permutations of the residuals is asymptotically exact. □

Asymptotical exactness for the test of hypothesis (15) for the lth functional parameter, based on statistics $T_l^{\mathcal{I}}$ (16), is a direct consequence of the aforementioned lemma. In addition to asymptotic results for functional tests on any linear hypothesis, we prove exactness for the overall functional model test of hypothesis (13):

Lemma 2. *For each interval $\mathcal{I} \subseteq [a, b]$, the functional overall model test of hypothesis (13) based on statistics $T_F^{\mathcal{I}}$ (14) is exact.*

Proof. Under $H_{0,F}^{\mathcal{I}}$, we have $y_i(t) = \beta_0(t) + \varepsilon_i(t)$, $\forall t \in \mathcal{I}$. The estimated residuals of this model are $\hat{\varepsilon}_{i,0}(t) = \beta_0(t) + \varepsilon_i(t) - \hat{\beta}_{0_n}(t)$, where $\hat{\beta}_{0_n}(t) = \bar{y}(t)$ is the sample mean of the responses $y_i(t)$. Note that the quantity $\beta_0(t) + \varepsilon_i(t) - \hat{\beta}_{0_n}(t)$ is permutationally invariant. Hence, the independence between the random functions $\varepsilon_i(t)$ implies the exchangeability with respect to the units of the residual functions $\hat{\varepsilon}_{i,0}(t)$ under $H_{0,F}^{\mathcal{I}}$. Thus, the test is exact, as it is based on the permutation of exchangeable quantities (Pesarin & Salmaso, 2010). □

In the next step, we verify the consistency of functional tests on linear hypotheses.

Lemma 3. *For each interval $\mathcal{I} \subseteq [a, b]$, the functional test of linear hypotheses on the regression parameters (10) based on the test statistics $T_C^{\mathcal{I}}$ (11) is consistent.*

Proof. The statement follows directly from the fact that the test statistics $T_C^{\mathcal{I}}$ is stochastically greater under $H_{1,C}^{\mathcal{I}}$ than under $H_{0,C}^{\mathcal{I}}$ (Pesarin & Salmaso, 2010). □

As a direct implication of Lemma 3, we get consistency for the overall functional model test of hypothesis (13) and the lth functional parameter test of hypothesis (15) based on test statistics $T_F^{\mathcal{I}}$ and $T_l^{\mathcal{I}}$, respectively.

A.2 | Properties of domain selection IWT procedures

We start by proving Theorem 1, establishing asymptotic intervalwise control of the domain selection IWT procedure for tests of linear hypotheses.

Proof of Theorem 1. Let $\mathcal{I} \subseteq [a, b]$ be an interval associated to a true null hypothesis, that is, let $H^{\mathcal{I}}_{0,C}$ hold. Consider a point $t \in \mathcal{I}$ and let $\mathcal{S}_t$ denote the set of all intervals containing the point t. The IWT-adjusted p value associated to point t is $\tilde{p}_C(t) = \sup_{\mathcal{J} \in \mathcal{I}_t} p^{\mathcal{J}}_C$, where $p^{\mathcal{J}}_C$ is the p value of the permutation test on interval $\mathcal{J}$. In particular, because $\mathcal{I} \in \mathcal{S}_t$, we have that $\forall t \in \mathcal{I}$, $\tilde{p}_C(t) \geq p^{\mathcal{I}}_C$. Therefore, $\mathbb{P}_{H^{\mathcal{I}}_{0,C}}[\forall t \in \mathcal{I} : \tilde{p}_C(t) \leq \alpha] \leq \mathbb{P}_{H^{\mathcal{I}}_{0,C}}[p^{\mathcal{I}}_C \leq \alpha]$. Because all tests are asymptotically exact (Lemma 1), we have

$$\lim_{n \to \infty} \mathbb{P}_{H^{\mathcal{I}}_{0,C}} \left[p^{\mathcal{I}}_C \leq \alpha \right] = \alpha,$$

and therefore,

$$\limsup_{n \to \infty} \mathbb{P}_{H^{\mathcal{I}}_{0,C}} [\forall t \in \mathcal{I} : \tilde{p}_C(t) \leq \alpha] \leq \alpha.$$

□

The assertion of Proposition 1 follows directly from the results of the work of Pini and Vantini (2016) and the fact that intervalwise tests used to build the procedure are exact (Lemma 2).

Furthermore, because all functional intervalwise tests are consistent (Lemma 3), the IWT procedure is also consistent (Theorem 2), due to the result of the work of Pini and Vantini (2017; theorem 3). Theorem 2 and Corollaries 2 and 3 follow immediately as special cases of Theorem 2.

APPENDIX B: THE FREEDMAN AND LANE PERMUTATION SCHEME

In this section, we give some details of the implementation of the Freedman and Lane permutation scheme for testing linear hypotheses on the regression model for each interval $\mathcal{I} \subseteq [a, b]$. In detail, we start from the restriction of the functional linear model (1) to interval $\mathcal{I}$, as follows:

$$y_i(t) = \sum_{l=0}^{L} \beta_l(t) x_{li} + \varepsilon_i(t), \quad \forall i = 1, \ldots, n; t \in \mathcal{I} \tag{B1}$$

with $x_{0i} = 1, \forall i$, and we describe the extension of the Freedman and Lane permutation scheme to the functional case, for implementing the functional test (10).

The Freedman and Lane permutations are based on the following steps:

- The residuals of the reduced model (that is the linear model under the null hypothesis) are estimated.
- The residuals of the reduced model are permuted.
- The permuted responses are computed through the reduced model and permuted residuals.

For more details about this method, we refer to the works of Freedman and Lane (1983) and Anderson and Legendre (1999). In the following subsections, we present the details of the scheme for the hypotheses introduced in our paper.

B.1 | Tests on general linear hypotheses

Under the null hypothesis (10), the model (B1) can be reduced by solving the linear system $C\boldsymbol{\beta}(t) = \mathbf{c}_0(t)$. In particular, because C has full rank, q regression parameters can be removed from the model by expressing them in terms of the others. Let Q denote the set of indexes of the removed

regression parameters. The reduced model is then

$$y_i(t) = \sum_{r \notin Q} \beta_r(t)\tilde{x}_{ri} + \varepsilon_i(t), \tag{B2}$$

that is, responses can be written in terms of a linear combination of modified covariates $\tilde{x}_{ri} = a_r(t)x_{ri}$, where $a_r(t)$ are fixed known coefficients, depending only on the solution of the linear system $C\boldsymbol{\beta}(t) = \mathbf{c}_0(t)$, and $\varepsilon_i(t)$ are i.i.d and zero-mean errors.

The residuals of the reduced model can then be estimated as $\hat{\varepsilon}_{i,C}(t) = y_i(t) - \sum_{r \notin Q} \hat{\beta}_{r,C_n}(t)\tilde{x}_{ri}$, where $\hat{\beta}_{r,C_n}(t)$ are the OLS estimates of parameters $\beta_r(t)$, $r \notin Q$, of model (B2). Then, the residuals $\hat{\varepsilon}_{i,C}(t)$ are permuted, and the permuted responses are computed using the permuted residuals $\hat{\varepsilon}^*_{i,C}(t)$ in the reduced model (B2), as follows:

$$y^*_i(t) = \sum_{r \notin Q} \hat{\beta}_{r,C_n}(t)\tilde{x}_{ri} + \hat{\varepsilon}^*_{i,C}(t). \tag{B3}$$

B.2 | Overall model tests

In the case of tests on the overall functional model hypothesis (13), under the null hypothesis $H^{\mathcal{I}}_{0,F}$, all regression parameters except the intercept are null. Therefore, the reduced model is

$$y_i(t) = \beta_0(t) + \varepsilon_i(t).$$

The estimated residuals of the reduced model are $\hat{\varepsilon}_{i,F}(t) = y_i(t) - \bar{y}(t)$, where $\bar{y}(t)$ is the sample mean of the responses $y_i(t)$. Therefore, using the permuted residuals $\hat{\varepsilon}^*_{i,F}(t)$, we get

$$y_i{}^*(t) = \bar{y}(t) + \hat{\varepsilon}^*_{i,F}(t).$$

Note that, in this case, permuting the estimated residuals $\hat{\varepsilon}_{i,F}(t)$ is equivalent to permuting the responses $y_i(t)$, $i = 1, \ldots, n$.

B.3 | Single-parameter tests

In the case of tests of hypothesis on the lth functional parameter (15), the model under the null hypothesis $H^{\mathcal{I}}_{0,l}$ reduces to

$$y_i(t) = \beta_0(t) + \sum_{r \neq l} \beta_r(t)x_{ri} + \varepsilon_i(t).$$

The estimated residuals of such model are $\hat{\varepsilon}_{i,l}(t) = y_i(t) - \hat{\beta}_{0,l_n}(t) - \sum_{r \neq l} \hat{\beta}_{r,l_n}(t)x_{li}$, where $\hat{\beta}_{r,l_n}(t)$ are the OLS estimates of the parameters of the reduced model. Then, the permuted responses are

$$y^*_i(t) = \hat{\beta}_{0,l_n}(t) + \sum_{r \neq l} \hat{\beta}_{r,l_n}(t) + \hat{\varepsilon}^*_{i,l}(t), \tag{B4}$$

where $\hat{\varepsilon}^*_{i,l}(t)$ are the permuted residuals.

APPENDIX C: FULL MODEL RESULTS

Figure C1 presents the results for the full model initially estimated for the knee data.

APPENDIX D: $F_{\mathbf{MAX}}$ RESULTS ON KNEE KINEMATICS

The backward elimination based on global p values using the maximum F-statistics selects the same reduced model that is presented in Section 5. Test results obtained with the $F_{\max}$ procedure are reported in Figures D1 and D2 for the reduced and full models, respectively.

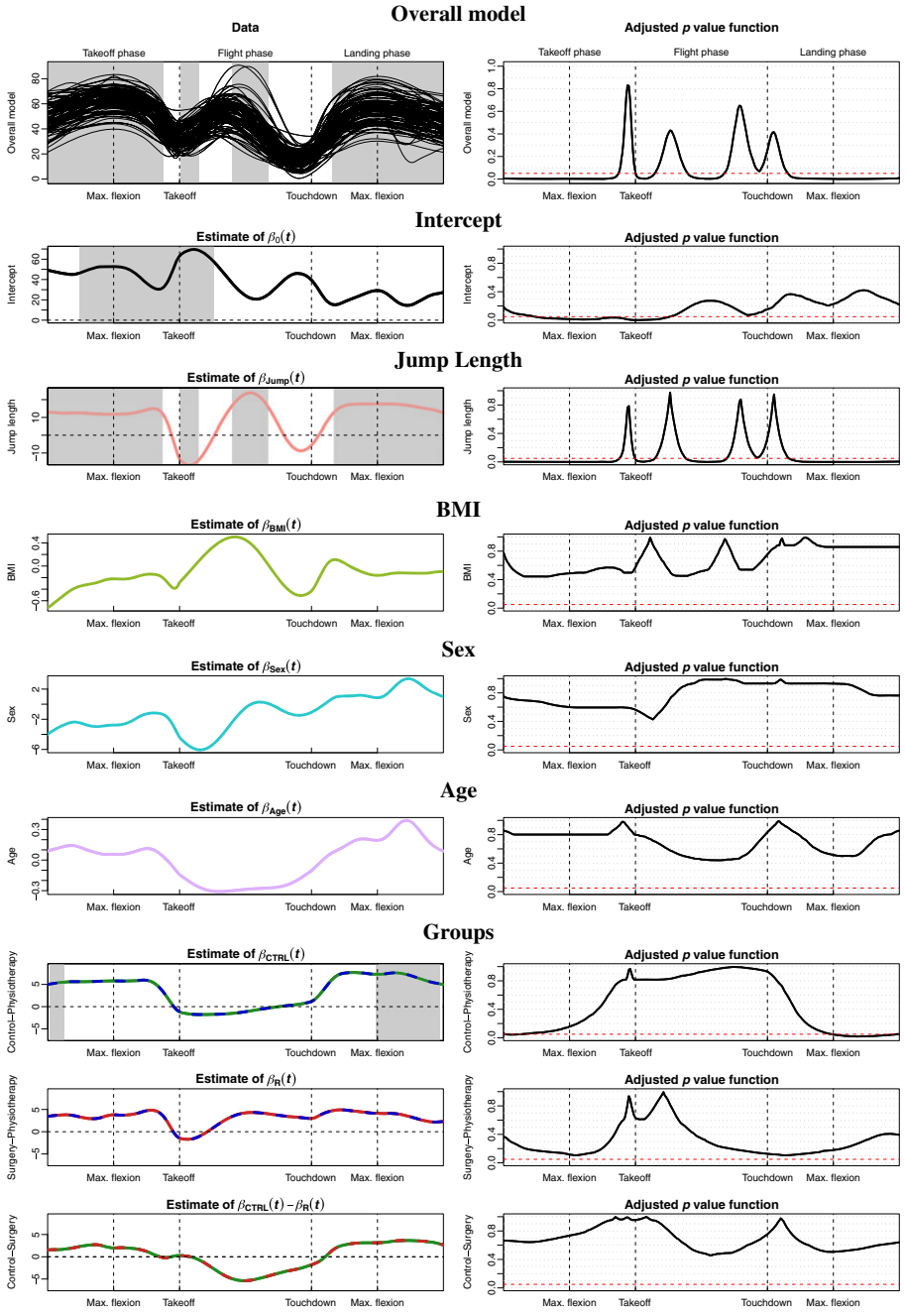

FIGURE C1 Functional parameter estimates and adjusted *p* value functions for the initial (full) functional-on-scalar model. Top panel: (left) knee-joint angle kinematics data (functional response variable) for the 95 individuals and (right) estimated adjusted *p* value function for the overall functional model tests. Bottom panels: (left) functional parameter estimates and (right) corresponding estimated adjusted *p* value functions for the intercept, jump length, body mass index (BMI), sex, age, and group effects, respectively. (Left) Gray-shaded parts correspond to significant effects at the 5% level, that is, the points *t* with associated adjusted *p* values less or equal to 0.05. (Right) Black solid lines correspond to intervalwise-testing-adjusted *p* value; (right) red dashed horizontal lines correspond to 5% threshold [Colour figure can be viewed at wileyonlinelibrary.com]

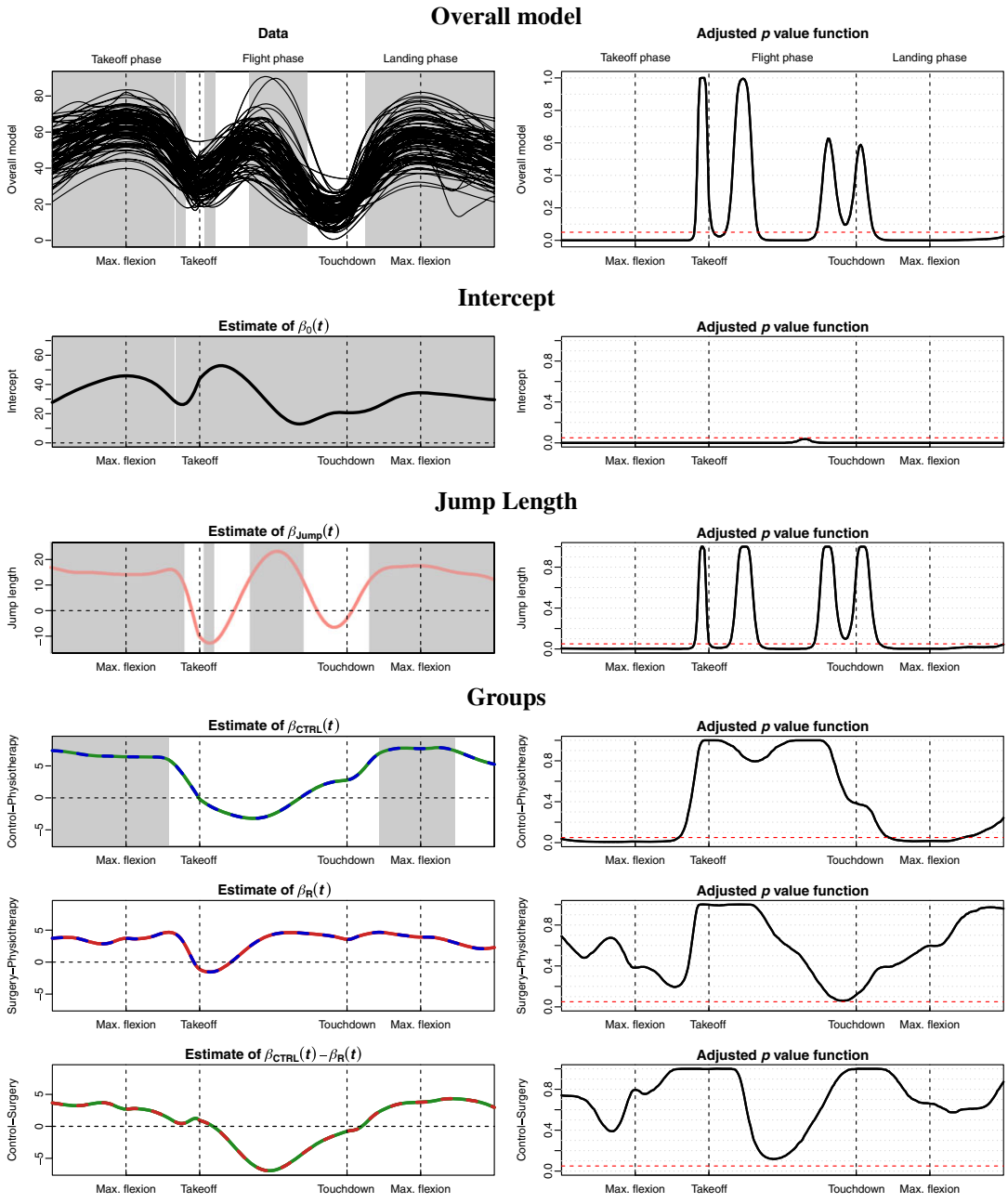

FIGURE D1 (Left) Functional parameter estimates and (right) F_{max}-adjusted p value functions for the final (reduced) functional-on-scalar model. Top panel: (left) knee-joint angle kinematics data (functional response variable) for the 95 individuals and (right) estimated adjusted p value function for the overall functional model tests. Bottom panels: (left) functional parameter estimates and (right) corresponding estimated adjusted p value functions for the intercept, jump length, and group effects, respectively. (Left) Gray-shaded parts correspond to significant effects at the 5% level, that is, the points t with associated F_{max}-adjusted p values less than or equal to 0.05. (Right) Black solid lines correspond to intervalwise-testing-adjusted p value; (Right) red dashed horizontal lines correspond to 5% threshold. BMI = body mass index [Colour figure can be viewed at wileyonlinelibrary.com]

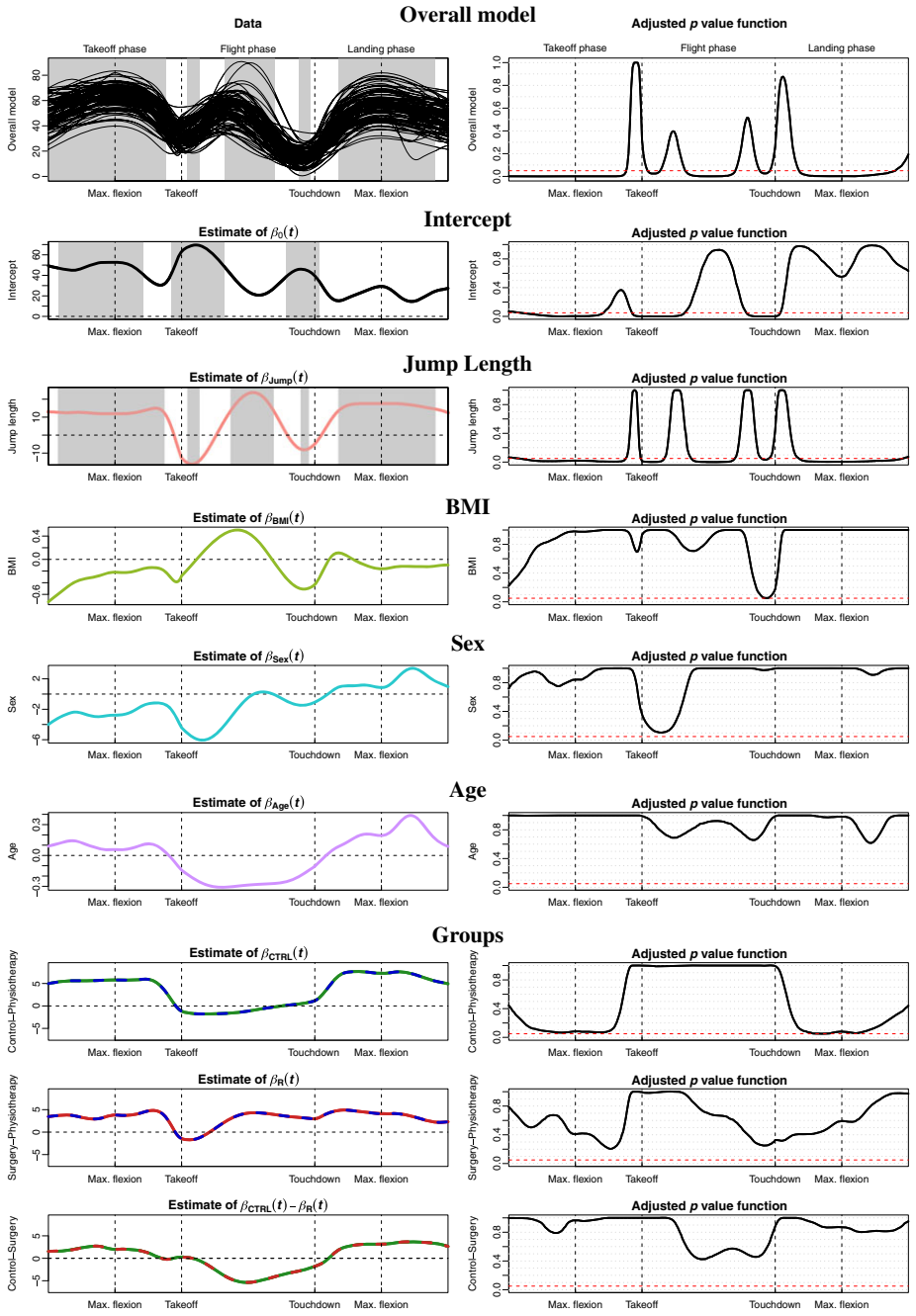

FIGURE D2 Functional parameter estimates and $F_{\max}$-adjusted p value functions for the initial (full) functional-on-scalar model. Top panel: (left) knee-joint angle kinematics data (functional response variable) for the 95 individuals and (right) estimated adjusted p value function for the overall functional model tests. Bottom panels: (left) functional parameter estimates and (right) corresponding estimated adjusted p value functions for the intercept, jump length, body mass index (BMI), sex, age, and group effects, respectively. (Left) Gray-shaded parts correspond to significant effects at the 5% level, that is, the points t with associated adjusted p values less or equal to 0.05. (Right) Black solid lines correspond to intervalwise-testing-adjusted p value; (right) red dashed horizontal line corresponds to 5% threshold [Colour figure can be viewed at wileyonlinelibrary.com]