DIRECTED LOCAL TESTING IN THE FUNCTIONAL LINEAR MODEL

By Christoph Rust

University of Regensburg

Abstract In this paper we develop a directed testing procedure for the coefficient function of the Functional Linear Model (FLM) with a scalar response. The suggested procedure provides local interpretability of the regression function in the sense that it is possible to infer from the data a subinterval of the random function's domain where the relationship between random curves and the scalar outcome variable differs from a null. The test is motivated from a situation where it can be assumed that the relationship is strongest at the beginning of the domain and it is of interest to find a subset of the domain where the relationship is statistically significant. Our real data application in the field of biomechanics is an example for such a situation. We show theoretical validity of our proposed procedure and evaluate the method by means of a simulation study.

1. Introduction. With the increasing availability of data collected on a dense grid, *functional data analysis* (FDA) has become an important field in statistics in recent years. The corresponding literature comprises numerous theoretical contributions and various interesting applications. In particular, the functional linear regression model with a scalar dependent variable is quite popular and has seen many applications in different areas (see Ullah and Finch, 2013, for an overview of various fields of application).

It is well known, however, that regression models with a functional predictor belong to the class of ill-posed inversion problems and that every estimate has to use some sort of regularisation. As a result, local inference in the functional linear model is a difficult issue. More specifically, Cardot et al. (2007a) show that it is impossible to derive a CLT for the estimated functional coefficient function. As a consequence, it is also not possible to draw pointwise inference about the functional coefficient, even asymptotically.

This does not mean that statistical tests for the parameter function are not possible in general. Indeed, there are several approaches to test whether the coefficient function deviates globally from a given null hypothesis. For instance, Cardot et al. (2003); Cardot, Goia and Sarda (2004) use properties of the cross-covariance operator, Swihart, Goldsmith and Crainiceanu (2014);

Keywords and phrases: functional data analysis, functional linear regression, testing

Kong, Staicu and Maity (2016) adapt classical methods such as likelihood-ratio and F tests to the functional case, and González-Manteiga et al. (2012) build on bootstrap techniques to construct such a global test. With these tests, however, it is only possible to find out whether the relationship between the functional predictors and the scalar outcome variable deviates from a prespecified null. The most important specific case, for instance, is the test of association in the functional linear model, where the null hypothesis corresponds to the zero function as functional coefficient.

Though, if the global test leads to a rejection, it is still not clear, whether the functional coefficient deviates from the null on the whole domain or is only different on a subpart of the domain. Therefore, a rejection of the global test does not uncover parts of the domain where the functional coefficient differs almost everywhere from a null coefficient. Evidently, this is an important shortcoming of the scalar-on-function regression model for practitioners that are interested in identifying subdomains where a functional covariate is related significantly to the scalar response.

An important contribution into this direction is the paper of Hall and Hooker (2016). Using their method, it is possible to consistently estimate the support of the coefficient function (that is, the the part of the domain where the coefficient function is different from zero). Very briefly, they also discuss the possibility of constructing a confidence interval for the boundaries of the support, based on the bootstrap. However, in their discussion, they also raise the issue that using the bootstrap is not appropriate in the context of smoothing. Again, this leaves the practitioner only with a point estimate.

We present an approach that shows one possibility to overcome these limitations under some additional assumptions about the structure of the coefficient function. Our testing procedure then is able to uncover a subinterval of the functional domain where the coefficient function is statistically different from zero almost everywhere. The underlying idea is to apply the global test sequentially on a family of subsets of the domain such that the family-wise error rate is controlled. The performance of available approaches for the global test is compared in Tekbudak et al. (2019), showing favorable properties of the F test suggested by Kong, Staicu and Maity (2016). Therefore, our method builds on this F test, although it might be replaced by other versions of the global test.

Inference in the functional regression model with scalar response has also been addressed by several other authors. In a generalised model framework, Müller and Stadtmüller (2005) propose (simultaneous) confidence bands for the coefficient function. These confidence bands though also reflect the global test and cannot be interpreted pointwise. Reiss and Ogden (2007) compute confidence intervals to illustrate the pointwise variability of the estimated coefficients in a simulation study which, however, cannot be used for inference. Imaizumi and Kato (2019) also propose simultaneous confidence bands which are based on weakened requirements for the coverage probability. It is, however, in general difficult to construct confidence regions for random elements in infinite dimensional Hilbert spaces (Choi and Reimherr, 2018).

The literature related to the scalar-on-function regression model is extensive and we refer to Reiss et al. (2017) for an overview of different methods. For readers with a general interest in FDA, the books by Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), and Hsing and Eubank (2015) provide a broad overview of available methods.

The remainder of this article is organized as follows. In Section 2, we introduce the model framework. The main results are given in Section 3, where we extend the global F test to a spline basis and introduce the sequential testing procedure. In that section, we also present our theoretical results. In Section 4 we illustrate the properties of the suggested method by means of a simulation study, and present a real data application in Section 5. Section 6 concludes.

2. Model framework.

We consider the functional linear model (FLM) with a scalar response variable

$$Y_i = \beta_0 + \int_{\mathcal{D}} \beta(t) X_i(t)\,\mathrm{d}t + \varepsilon_i, \tag{1}$$

reflecting the dependency between observations of a scalar variable, $Y_1, \dots, Y_n$, and a functional covariate $X_1, \dots, X_n$, the latter taking values in the Hilbert space $L^2(\mathcal{D})$. Without loss of generality, we set $\mathcal{D} = [0,1]$ and assume for notational simplicity that $E(Y_i) = 0$ and $E(X_i) = 0 \in L^2$, omitting the intercept, β_0 hereafter. For the error term we assume that the ε_i's are i.i.d. centered ($E(\varepsilon_i) = 0$) random variables with variance $Var(\varepsilon_i) = \sigma^2 < \infty$, and are independent of the X_i. The function-valued slope parameter $\beta \in L^2$ quantifies the effect of the functional predictors on the scalar outcome variable, and, very often is of central interest. Due to the reasons lined out above, available methods only leave a practitioner with a point estimate and the information whether the relationship between X_i and Y_i is significant at a given confidence level. While the method of Hall and Hooker (2016) makes it also possible to estimate a subinterval $[0, \theta]$ of the original domain where $\beta(t) \neq 0$, it cannot tell whether the function over the estimated support is almost everywhere statistically different from zero at a given significance level.

This, however, is of practical importance for the structural interpretation of β.

In some applications, it can be reasonable to assume that if there is a linear dependence between the scalar response and the functional covariate, it must be strongest at one boundary of the domain. For instance, if one can assume that $|\beta|$ is a monotonically decreasing function, it is then possible to construct a sequential test procedure that is capable to identify a subset of the domain, $[0, \tau^*] \in [0, 1]$ where at a prespecified level α, it holds that the function $\beta(t) \neq 0$ for almost every $t \in [0, \tau^*]$. Although in our view the assumption of a monotonic $|\beta|$ is the most important case in practice, one can also get qualitatively the same results under the weaker assumption 6.1 of Hall and Hooker (2016).

Before we introduce the test approach, let us define the basic notation. To this end, it is assumed that all realizations of the functional covariate are observed on the same grid values $t_1, \ldots, t_p$, with $t_1 = \frac{1}{2p},\ t_i - t_{i-1} = \frac{1}{p}$. In the $n \times p$ matrix $\mathbf{X}$, we collect the functional observations and the n-vector $\mathbf{Y}$ holds the observations of the dependent variable.

3. Testing procedure and theoretical results. To formulate our test procedure, we make the following assumption about the coefficient function β.

(2) The functional coefficient β is continuous on $[0, 1]$, the absolute value,$|\beta|$, is monotonically decreasing, and there exists a $\tau \in [0, 1]$ such that $\beta(t) \neq 0$ for almost all $t \in [0, \tau]$.

Under this assumption, the following procedure is able to find the largest number $\tau^* \in [0, \tau]$ such that β is statistically different from zero on the interval $[0, \tau^*]$ at a predefined significance level α. The core idea of the test procedure is to split the domain into two parts, $[0, t_l]$ and $[t_l, 1]$, and test whether $\beta(t) = 0$ for almost all $t \in [t_l, 1]$ sequentially for all split points t_l with $l = 1, \ldots, p - 1$.

A consistent estimator for β also requires some additional regularity assumptions regarding the process X and the functional coefficient β. These requirements depend on the estimation context, see Hall and Horowitz (2007) or Hall and Hosseini-Nasab (2006) for the classical approach based on functional principal components and Crambes, Kneip and Sarda (2009); Cardot et al. (2007b) for an estimator based on smoothing splines. Moreover, we do not discuss the consequences arising from the discretization of the curves. Hence, we assume that the number of observations points p is sufficiently large compared to the number of observations n, such that the approximation error can be neglected. Formal arguments can also be found, for example, in Crambes, Kneip and Sarda (2009). Furthermore, since this paper as currently written is primarily practically oriented, we assume that the function β can be represented using B-splines. A more rigorous development of the corresponding theory is therefore left for future work.

3.1. *Local test.* For each $l = 1, \dots, p-1$ the null and alternative hypothesis of the local test are

$$\begin{aligned} H_0 &: \beta(t) = 0 \text{ for almost all } t \in [t_l, 1], \\ H_1 &: \beta(t) \neq 0 \text{ for some } t \in [t_l, 1], \end{aligned} \tag{3}$$

where, for the alternative, the set $\{t \in [t_l, 1] \,|\, \beta(t) \neq 0\}$ must not be a zero set with respect to the Lebesgue measure.

The test can also be formulated differently by partitioning the integral in Model (1) into two parts. To this end, let $\beta_1(t) = \mathbb{1}_{[0,t_l]}(t)\beta(t)$ and $\beta_2(t) = \mathbb{1}_{[t_l,1]}(t)\beta(t)$ denote the first and second part of $\beta(t)$ splitted at t_l. Model (1), using that notation, is then equivalent to the splitted model

$$Y_i = \int \beta_1(t) X_i(t)\,\mathrm{d}t + \int \beta_2(t) X_i(t)\,\mathrm{d}t + \varepsilon_i, \tag{4}$$

and the test (3) is then equivalent to test (globally) for $\beta_2 = 0$ a.e. for which several methodologies have been suggested in the literature. Here, we build on the F test suggested by Kong, Staicu and Maity (2016) which builds on the classical Karhunen-Loève decomposition. However, using their approach straight away, it would be necessary to solve a relatively costly eigenvalue problem twice for every local test. To avoid this, we instead use a fixed spline basis which we have to assume to be appropriate for β. In the following section, we briefly explain how model (4) is fitted based on a B-spline expansion of the splitted β.

3.1.1. *Spline estimator for splitted model.* Let the integer k denote the dimension of the spline basis for the splitted model which should be chosen such that the corresponding basis is flexible enough to expand β. A function space corresponding to a cubic spline basis has at least four dimensions, hence, the function space for the splitted β has at least eight dimensions ($k \geq 8$). However, when $t_l < 4$ (or $t_l > p-3$), the matrix of a cubic spline basis evaluated at $0, \dots, t_l$ (or $t_l, \dots, 1$, respectively) are not regular, why these cases have to be treated separately.

Starting with the case $l \in \{4, \dots, p-3\}$, the corresponding knot sequence

$\tau_1, \ldots, \tau_{k+4}$, spanning $[0, 1]$, for the spline basis is given by

boundary knots:

$$\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$
$$\tau_{k+1} = \tau_{k+2} = \tau_{k+3} = \tau_{k+4} = 1$$

interior knots:

$$\tau_j = \begin{cases} t_l & \text{if } t_l \in (\tau_{j-1} + 0.5\,\delta, \tau_{j-1} + 1.5\,\delta] \wedge \tau_{j-3} \neq t_l \\ \tau_{j-5} + 2\delta & \text{if } \tau_{j-1} = t_l \\ \tau_{j-1} + \delta & \text{otherwise,} \end{cases}$$

where $\delta = \frac{1}{k-6}$ is the distance between interior knots. Intuitively, the above knot sequence has four boundary knots at 0 and 1, respectively, $k-7$ equidistant interior knots, where the interior knot which is closest to t_l is replaced by four knots at t_l. The corresponding B-spline basis, by design, spans a function space of twice continuously differentiable functions on the intervals $(0, t_l)$ and $(t_l, 1)$ and a jump point (due to repeating the knot t_l four times) at t_l. We denote by $\mathbf{b}(t) = (b_1(t), \ldots, b_k(t))^\top$ and $BS(k, t_l) = \langle b_1, \ldots, b_k \rangle$ the corresponding basis functions and the linear space. Every function $f \in BS(k, t_l)$ can be represented by a k-dimensional parameter θ, such that under the assumption $\beta \in BS(k, t_l)$, model (1) can be rewritten as

$$Y_i = \int X_i(t)\mathbf{b}(t)^\top \theta \,\mathrm{d}t + \varepsilon_i.$$

If we approximate the integral by the corresponding Riemann sum (midpoint rule), we get

$$Y_i = \frac{1}{p} \sum_{j=1}^{p} X_i(t_j)\mathbf{b}(t_j)^\top \theta + \varepsilon_i.$$

Let $\mathbf{B}$ be the $p \times k$ matrix of the k spline functions evaluated at the p grid values, that is $\mathbf{B}_{ij} = b_j(t_i)$. The matrix $\mathbf{B}$ has a block structure, where the upper-left $l \times m$ block $\mathbf{B_1}$ (m is the number of knots preceding the 4 knots for the split point) is the evaluated basis for the interval $[0, t_l]$ and the lower-right $(p-l) \times (k-m)$ block $\mathbf{B_2}$ is the basis for $(t_l, 1]$. The remaining upper-right and lower-left blocks are zero matrices.

In the other case, where $l < 4$ or $l > p-3$, the cubic spline basis evaluated at $0, \ldots, t_l$ (or $t_l, \ldots, 1$) is no longer invertible and, therefore, this singular block is replaced by the identity matrix of size l (or $p-l$, respectively).

A least-squares estimate of the spline coefficients $\widehat{\theta}$ can be obtained as usual via

$$\widehat{\theta} = p\left(\mathbf{B}^\top \mathbf{X}^\top \mathbf{X}\mathbf{B}\right)^{-1}\mathbf{B}^\top \mathbf{X}^\top \mathbf{y}.$$

Under the null, the last $k-m$ entries of θ are zero and the m remaining non-zero entries of $\widehat{\theta}$ in the restricted model become

$$\widehat{\theta}_1 = p\left(\mathbf{B_1}^\top \mathbf{X}^\top \mathbf{X}\mathbf{B_1}\right)^{-1}\mathbf{B_1}^\top \mathbf{X}^\top \mathbf{y}.$$

Analogous to the finite dimensional regression model, this results in the projection matrices

$$\mathbf{P_{XB}} = \mathbf{XB}\big(\mathbf{B}^\top \mathbf{X}^\top \mathbf{XB}\big)^{-1}\mathbf{B}^\top \mathbf{X}^\top$$
$$\mathbf{P_{XB_1}} = \mathbf{XB_1}\big(\mathbf{B_1}^\top \mathbf{X}^\top \mathbf{XB_1}\big)^{-1}\mathbf{B_1}^\top \mathbf{X}^\top$$

for full and null model.

3.1.2. *Test statistic.* The F test in Kong, Staicu and Maity (2016) is defined in terms of the residual sum of squares under the full and null models:

$$RSS_{\text{full}} = \mathbf{y}^\top(\mathbf{I}_n - \mathbf{P_{XB}})\mathbf{y}$$
$$RSS_{\text{null}} = \mathbf{y}^\top(\mathbf{I}_n - \mathbf{P_{XB_1}})\mathbf{y}$$

and the test statistic is

$$(5) \quad T_F = \frac{(RSS_{\text{null}} - RSS_{\text{full}})/(k-m)}{RSS_{\text{full}}/(n-k)} = \frac{\mathbf{y}^\top(\mathbf{P_{XB}} - \mathbf{P_{XB_1}})\mathbf{y}/(k-m)}{\mathbf{y}^\top(\mathbf{I}_n - \mathbf{P_{XB}})\mathbf{y}/(n-k)}.$$

Assuming that the discretization error is negligible, we can formulate the following result.

LEMMA 1. *If for fixed k and l, $\beta \in BS(k, t_l)$ (the B-spline space with k dimensions and split point at t_l) and the errors are centered i.i.d. normal, then, under the null, T_F is F-distributed with $k-m$ and $n-k$ degrees of freedom.*

By requiring that the true parameter function is an element of $BS(k, t_l)$, the above result is limited as is does not consider the more general case that $\beta \in L^2$. For many applications, this is a reasonable assumption, in particular when the true parameter function is sufficiently smooth. The general case requires that the model complexity (k in our notation) grows at a suitable rate with the number of observations n. As mentioned above, developing formal arguments for the general case, is left for future research.

Algorithm 1 Directed sequential test

Let α denote global global significance level and H_0^i denote the i-th local null hypothesis
Initialize $i = 1$
while $i < p \ \wedge H_0^{i-1}$ was rejected **do**
    test H_0^i at level α
    increment i
end while
H_0^{i-1} then is the last rejected hypothesis and therefore conclude that β is statistically different from zero over the interval $[0, t_{i-1}]$ at level α.

3.2. *Sequential directed test.* Based on the above result, we can formulate the testing procedure for detecting the interval $[0, \tau^*]$, on which the curves significantly influence the dependent variable Y. The central idea is to perform the above local test consecutively for every $l \in \{1, \ldots, p-1\}$ and stop once the test does not reject for the first time. The iterative procedure is given in Algorithm 1. The following result shows that our directed testing procedure controls the family-wise error rate (FWER) in the strong sense by design.[1]

THEOREM 1. *The sequential test described in Algorithm 1 controls FWER in the strong sense, that is, for any $\tau \in [0, 1]$ for which $\beta \mathbb{1}_{t>\tau} = 0$ a.e., the probability of the event that $\tau^* > \tau$ is bounded at global level α.*

As shown in the proof, our test is an example for a test meeting the requirements for the closed testing principle (Marcus, Peritz and Ruben, 1976). It, therefore, by design controls FWER without any further corrections. For the above result, it is even not necessary that β belongs to all $BS(k, t_l)$, $l = 1, \ldots, p-1$. It is sufficient, if the requirements of Lemma 1 are fulfilled for the first local test in the test sequence for which the null hypothesis is true. This local test then takes the role of a gatekeeper, such that the size of the overall procedure is maintained at level α.

4. Simulation study.

In the following simulation study, we investigate the performance of the sequential testing procedure. Apart from a numerical validation of the above theoretical results, it is of interest to analyse power properties of the test under different scenarios, in particular with respect

[1]A test procedure is said to control the FWER *in the strong sense*, if the probability of making at least one type I error is controlled at level α for every combination of true or non-true individual null hypotheses, whereas the *weak sense* control refers to the case where all individual null hypothesis have to be true.

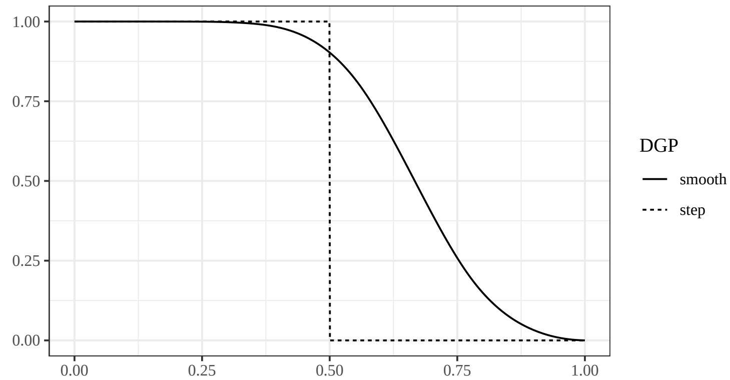

FIGURE 1. *Coefficient functions used in the simulation exercise.*

to the distance to τ. The simulation exercise considers two different Data Generating Processes (DGPs) which can be seen as extreme scenarios. While the first DGP is a step function that is useful to demonstrate how closely the procedure can approach the point where the true β becomes zero, the second DGP is a monotonically decreasing function which approaches zero but will never be exactly zero (see Figure 1 for an illustration of both DGP). The actual specification for the first DGP is $\beta_{\mathtt{step}} = \mathbb{1}_{[0,0.5]}$, and for the second DGP, $\beta_{\mathtt{smooth}}$ is a continuously decreasing function based on B-splines with 8 equidistant interior knots. While $\beta_{\mathtt{smooth}}$ by construction fulfills the assumptions of Lemma 1 for all individual tests, this is not true for $\beta_{\mathtt{step}}$ due to the discontinuity at $t = 0.5$. It can be easily seen, however, that $\beta_{\mathtt{step}}$ fulfills the requirements of Lemma 1 for one individual test, namely with split point $t_l = 0.5$, which is the first individual test for which the null hypothesis is true. Thus, both DGPs meet the necessary requirements for the procedure.

For each DGP we consider three different sample sizes ($n = 250, 500, 1000$) and six different signal-to-noise ratios ($\gamma = 0.1, 0.5, 1, 2, 5, 10$). For every scenario, we generate 10 000 replications of n random tuples (X_i, Y_i) via Model (1), where the random curves X_i are i.i.d. realizations of the functional random variable $\sum_{1<j<p} Z_j\, j^{-2} \phi_j$, where $\phi_j(t) = \sqrt{2}\cos((j-1)\pi t)$ and independent standard normal Z_j. In the appendix, we also consider random curves with a different covariance function that is more concentrated around the diagonal. Although the actual number of discretization points p is of minor importance, the simulation study also considers two cases ($p = 100, 300$). The error term ε_i is a centered normally distributed random variable with

TABLE 1
Type-I error rates for DGP with $\beta = \beta_{\text{step}}$ and at global level $\alpha = 0.05$,

	$p = 100$				$p = 300$			
n	100	250	500	1000	100	250	500	1000
γ								
0.1	0.010	0.009	0.011	0.013	0.010	0.009	0.010	0.015
0.5	0.012	0.015	0.015	0.020	0.014	0.016	0.017	0.019
1	0.018	0.018	0.018	0.021	0.018	0.018	0.018	0.018
2	0.021	0.021	0.023	0.021	0.020	0.019	0.022	0.024
5	0.025	0.024	0.024	0.026	0.024	0.026	0.026	0.028
10	0.026	0.024	0.026	0.028	0.028	0.028	0.029	0.028

variance σ^2. The signal-to-noise ratio is specified as the quotient of the variance of $\int X_i(t)\beta(t)\,\mathrm{d}t$ and the error variance σ^2. For the test, one also has to specify the number of spline basis functions for the expansion of $\widehat{\beta}$ which we set to $k = 12$. The simulation is implemented in GNU R (R Core Team, 2020) and, together with an R-package implementing the testing procedure, the code is part of the online supplement to this article.

Simulation results for the type-I error rate are summarized in Table 1 for the DGP involving $\beta_{\texttt{step}}$.[2] In case of the first DGP, a type-I error occurs if the test does not stop rejecting at or before $t = 0.5$. While every row of the table corresponds to a different signal-to-noise ratio γ, combinations of sample size n and number of discretization points p are found in the respective columns of the table. (Column 1-4 correspond to $p = 100$ whereas Column 5-8 correspond to $p = 300$). The global significance level here is set to $\alpha = 0.05$ and we can see that the procedure maintains size in all constellations, as expected.

Another interesting investigation of the simulation exercise is power analysis. Figure 2 summarizes the results for different scenarios. Left and right column correspond to $\beta_{\texttt{step}}$ and $\beta_{\texttt{smooth}}$, each row refers to a specific sample size, and different line types indicate signal-to-noise ratios (as outlined in the legend). The line of the graph depicts the probability of a rejection up to the respective value $t \in [0, 1]$ on the horizontal axis. It can be seen that sample size and signal-to-noise ratio have the expected influence on the rejection probability: the larger the sample size and the larger the signal-to-noise ratio, the larger also is the power.

[2]Note that there can be no false positive for any of the tests $H_0 : \beta(t) = 0 \forall t \in [t_l, 1]$ for the DGP involving $\beta_{\texttt{smooth}}$ since $\beta_{\texttt{smooth}}(t) \neq 0\, \forall t \in [0, 1]$.

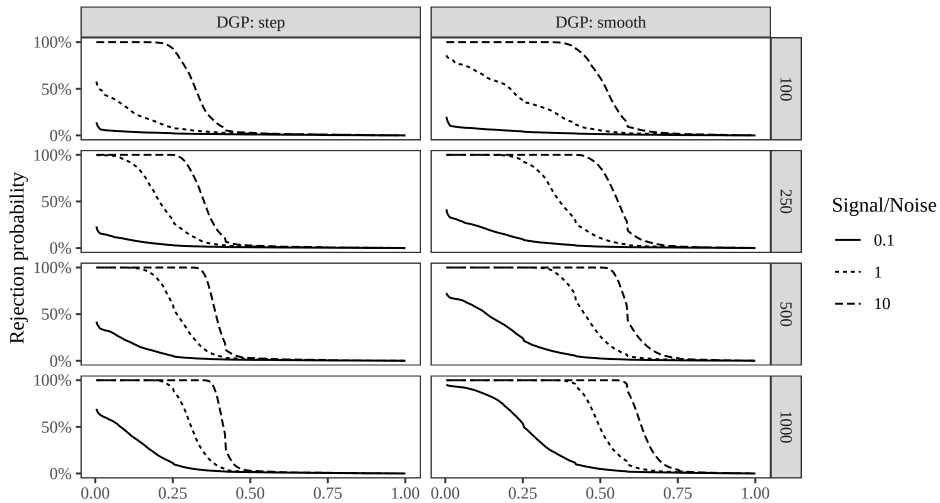

FIGURE 2. *Rejection probabilities of the directed testing procedure for the two DGPs with* $p = 300$.

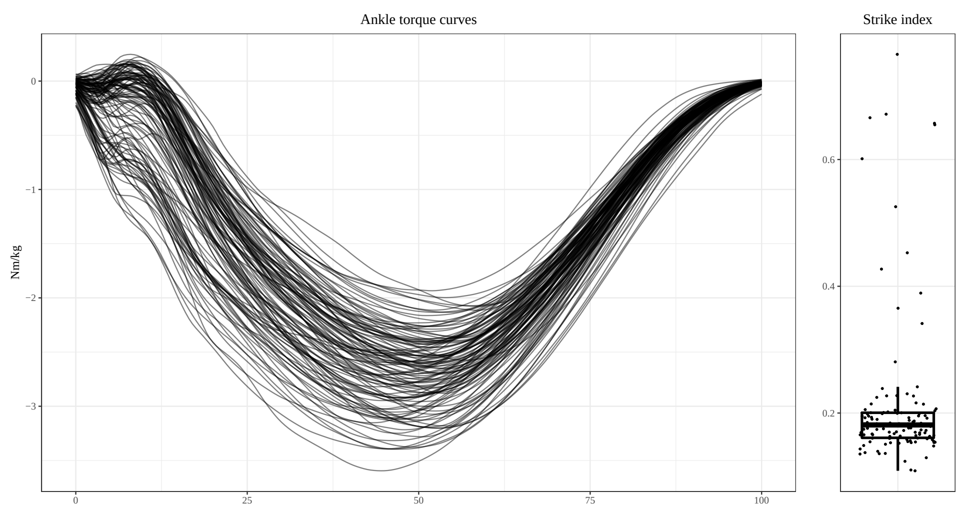

FIGURE 3. *Sample of torque curves measured at the ankle joint and boxplot of the strike index of recreational runners.*

5. Application. With the following real data exercise, we want to demonstrate the applicability of the sequential testing method in practice. We are drawing on data from a sports biomechanics experiment. Using the functional data approach to model such measurements is quite natural and frequently adopted by empirical researchers in biomechanics (Vanrenterghem et al., 2012; Liebl et al., 2014; Hamacher, Hollander and Zech, 2016; Warmenhoven et al., 2019).

The data shown in Figure 3 result from a sports biomechanics experiment conducted at the biomechanics lab of the German Sport University, Cologne, Germany. The sample comprises measurements of torque curves at the ankle joint as well the strike index of $n = 119$ recreational runners, obtained while they were running without shoes. Each curve describes the torque measured at the right ankle joint in the sagittal pane during the *stance phase*, a standardized interval $[0\%, 100\%]$ during which the foot has contact to the ground. Using an affine transformation, the individual stance phases were standardized such that the initial ground contact takes place at $t = 0\%$ while the foot leaves the ground at $t = 100\%$. Moreover, the torque measures are standardized by the participants' body weights to make the curves comparable. The right panel of Figure 3 shows the distribution of the so-called *strike index*, describing the center of pressure (CoP) at initial ground contact with respect to the long axis of the foot (Cavanagh and Lafortune, 1980; Altman and Davis, 2012). The strike index is used to classify runners into the groups rear foot, mid foot or fore food strikers, the majority of runners belonging to the group of rear foot strikers.

The torque curves as a whole show a similar pattern over all runners with a negative value throughout the largest part of the stance phase, representing a torque in the plentarflexion direction which is associated with the acceleration into the moving direction. Many curves, however, at the very beginning of the stance phase are positive which indicates a torque in dorsiflexion direction, associated with a controlled lowering of the fore food immediately after ground contact.

It is therefore obvious that the shape of the torque curves depend on the center of pressure at initial ground contact. Runners with a very low strike index (center of pressure is very close to the heel) are likely to have a torque curve which is positive in the beginning of the stance phase, while runners with a higher strike index (center of pressure is located more in the center of the foot's long axis) need a negative torque at the ankle joint in order to accelerate the body mass upwards as well as into the moving direction. Since it is not clear at which part of the stance phase the torque curves are related to the strike index, our sequential testing procedure can help to address this

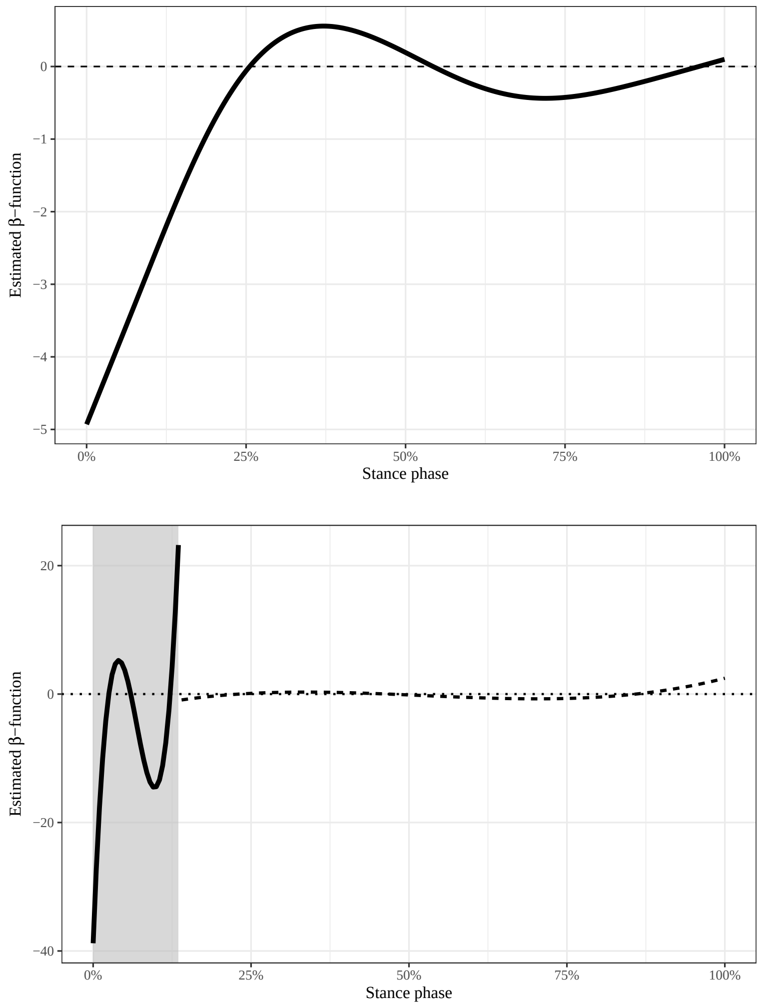

FIGURE 4. *Estimation result for FLM where the scalar variable strike index is explained by the torque curves of barefoot runners (upper panel), and result of the sequential testing procedure (lower panel), where the rejection region is shaded in grey.*

question.

With torque curves forming the functional predictor and the strike index as the scalar dependent variable, we applied the sequential test to find out for which part of the stance phase the dependence between torque curves and strike index is statistically different from zero. At first, however, we present the result of an application of a smoothing splines estimation (Crambes, Kneip and Sarda, 2009) of the corresponding functional linear model. Since the data is not standardized, we also include an intercept in the model. Figure 4 shows in the upper panel an estimated functional coefficient based on a smoothing parameter for which we set the effective degrees of freedom to 5. As expected, the estimate shows a negative relationship between torque curves and strike index and from 25% of the stance phase, the estimated functional coefficient approaches zero. As lined out in the introduction, such a point estimate cannot say anything about statistical significance of the result.

The result of applying our sequential testing procedure at global level $\alpha = 0.05$ is given in the lower panel of Figure 4. The grey area indicates the region where the test rejects the null of no association which in this application turns out to be the first 14% of the stance phase. The solid and dashed black lines depict the estimation result of the splitted model. This result makes sense if we once more take a look at the torque curves in Figure 3 where we can see that in particular at the first 13% the shape of the curves vary the most. In the later stance phase, the curves have a rather similar pattern across individuals. Consequently, it is not very likely that the center of pressure during the first ground contact is related to the torque in later stance phase.

6. Conclusion. In this paper, we propose a new testing method for the functional linear model in order to facilitate interpretability of model estimates. The proposed methodology can be helpful to detect a subregion of the functional predictors' domain where the relationship to a scalar outcome variable is statistically significant. The testing method builds on the (global) test of association in the functional linear model and performs this test sequentially on a sequence of decreasing domains.

The key results of our approach follow from the closed testing principle which ensures that the family-wise error rate (FWER) is maintained in the strong sense without any further corrections. For the theoretical analysis, we make some limiting assumptions in particular about the functional coefficient and the discretization error. Very likely, these assumptions can be relaxed to make the procedure valid in a more general setting. However, this requires a more thorough theoretical analysis that can be a future direction for this paper. In addition, the method can also be generalized to test against a general $\beta_0 \in L^2$ and, in principle, it is also not necessary to start the test at the left boundary of the domain. The procedure's numerical properties are also evaluated in a simulation exercise. We show that the global type-I error is maintained and explore the method's power properties under different scenarios with respect to sample size and signal-to-noise ratio.

We demonstrate the practical use of our method by applying it to data from a sports biomechanics experiment with recreational runners. In this example, we use a functional linear model to measure the dependency between the strike index as scalar outcome variable and torque curves measured at the ankle join as functional predictor. Using the proposed sequential test, we can show that the dependency between strike index and torque curves is statistically significant only at the very beginning of the stance phase (0%-14%) which corresponds to the first time after the foot's ground contact.

References.

ALTMAN, A. R. and DAVIS, I. S. (2012). A kinematic method for footstrike pattern detection in barefoot and shod runners. *Gait & Posture* **35** 298 - 300.

CARDOT, H., GOIA, A. and SARDA, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Communications in Statistics-Simulation and Computation* **33** 179–199.

CARDOT, H., FERRATY, F., MAS, A. and SARDA, P. (2003). Testing Hypotheses in the Functional Linear Model. *Scandinavian Journal of Statistics* **30** 241–255.

CARDOT, H., , MAS, A. and SARDA, P. (2007a). CLT in Functional Linear Regression Models. *Probab. Theory Relat. Fields* **138** 325–361.

CARDOT, H., CRAMBES, C., KNEIP, A. and SARDA, P. (2007b). Smoothing Splines Estimators in Functional Linear Regression with Errors-in-Variables. *Computational Statistics & Data Analysis* **51** 4832–4848.

CAVANAGH, P. R. and LAFORTUNE, M. A. (1980). Ground reaction forces in distance running. *Journal of Biomechanics* **13** 397 - 406.

CHOI, H. and REIMHERR, M. (2018). A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 239–260.

CRAMBES, C., KNEIP, A. and SARDA, P. (2009). Smoothing Splines Estimators for Functional Linear Regression. *The Annals of Statistics* **37** 35–72.

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis - Theory and Practice*. Spinger.

GONZÁLEZ-MANTEIGA, W., GONZÁLEZ-RODRÍGUEZ, G., MARTÍNEZ-CALVO, A. and GARCÍA-PORTUGUÉS, E. (2012). Bootstrap independence test for functional linear models. *arXiv preprint arXiv:1210.1072*.

HALL, P. and HOOKER, G. (2016). Truncated Linear Models for Functional Data. *Journal of the Royal Statistical Society Series B* **78** 637–653.

HALL, P. and HOROWITZ, J. (2007). Methodology and Convergence Rates for Functional Linear Regression. *Annals of Statistics* **35** 70–91.

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components. *Journal of The Royak Statistical Society Series B* **68** 106–126.

HAMACHER, D., HOLLANDER, K. and ZECH, A. (2016). Effects of ankle instability on running gait ankle angles and its variability in young adults. *Clinical Biomechanics* **33** 73 - 78.

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*. John Wiley & Sons.

IMAIZUMI, M. and KATO, K. (2019). A Simple Method to Construct Confidence Bands in Funtional Linear Regression. *Statistica Sinica* **29** 2055–2081.

KONG, D., STAICU, A.-M. and MAITY, A. (2016). Classical Testing in Functional Linear Models. *Journal of Nonparametric Statistics* **28** 813–838.

LIEBL, D., WILLWACHER, S., HAMILL, J. and BRÜGGEMANN, G.-P. (2014). Ankle plantarflexion strength in rearfoot and forefoot runners: A novel clusteranalytic approach. *Human Movement Science* **35** 104 - 120.

MARCUS, R., PERITZ, E. and RUBEN, G. K. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660.

MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized Functional Linear Models. *The Annals of Statistics* **33** 774–805.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2. ed. Springer.

Reiss, P. T. and Ogden, R. T. (2007). Functional Principal Component Regression and Functional Partial Least Squares. *Journal of the American Statistical Association* **102** 984–996.

Reiss, P. T., Goldsmith, J., Shang, H. L. and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review* **85** 228–249.

Swihart, B. J., Goldsmith, J. and Crainiceanu, C. M. (2014). Restricted Likelihood Ratio Tests for Functional Effects in the Functional Linear Model. *Technometrics* **56** 483–493.

Tekbudak, M. Y., Alfaro-Córdoba, M., Maity, A. and Staicu, A.-M. (2019). A Comparison of Testing Methods in Scalar-on-Function Regression. *AStA Advances in Statistical Analysis* **103** 411–436.

R Core Team (2020). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology* **13** 43.

Vanrenterghem, J., Venables, E., Pataky, T. and Robinson, M. A. (2012). The effect of running speed on knee mechanical loading in females during side cutting. *Journal of Biomechanics* **45** 2444 - 2449.

Warmenhoven, J., Cobley, S., Draper, C., Harrison, A., Bargary, N. and Smith, R. (2019). Considerations for the use of functional principal components analysis in sports biomechanics: examples from on-water rowing. *Sports Biomechanics* **18** 317-341. PMID: 29141500.

Appendix.

Proof of Lemma 1. Under the null and the prerequisites of the lemma, we have that $\beta \in BS(k, t_l)$ with $\beta(t) = 0$ for $t \in (t_l, 1]$. The model then can then be written as

$$Y_i = \int_0^{t_l} \beta(t) X_i(t)\,\mathrm{d}t + \varepsilon_i,$$

which in matrix notation (note that we assume the dense sampling design) for the sample is

$$\mathbf{y} = \frac{1}{p}\mathbf{X}\mathbf{B_1}\theta_1 + \boldsymbol{\varepsilon}.$$

Now, observe that the kernel of the projection matrix $\mathbf{P_{XB}} - \mathbf{P_{XB_1}}$ is given by the space spanned by the columns of $\mathbf{B_1}$ and, similarly, the kernel of $\mathbf{I}_n - \mathbf{P_{XB}}$ is the column space of $\mathbf{B}$.

It follows that

$$\mathbf{y}^\top (\mathbf{P_{XB}} - \mathbf{P_{XB_1}})\,\mathbf{y} = \boldsymbol{\varepsilon}^\top (\mathbf{P_{XB}} - \mathbf{P_{XB_1}})\,\boldsymbol{\varepsilon}$$

and

$$\mathbf{y}^\top (\mathbf{I}_n - \mathbf{P_{XB}})\,\mathbf{y} = \boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P_{XB}})\,\boldsymbol{\varepsilon}$$

which are scaled (by the variance of ε) versions of χ^2-distributed variables with $k - m$ and $n - k$ degrees of freedom, since the matrices $(\mathbf{P_{XB}} - \mathbf{P_{XB_1}})$ and $(\mathbf{I}_n - \mathbf{P_{XB}})$ have rank $k - m$ and $n - k$. It is easy to see that the matrices $(\mathbf{P_{XB}} - \mathbf{P_{XB_1}})$ and $(\mathbf{I}_n - \mathbf{P_{XB}})$ are orthogonal projections, completing the proof. □

Proof of Theorem 1. We will show that our sequential testing procedure is an example for a closed test, such that the FWER is controlled in the strong sense by design.

The family of hypotheses in our testing procedures is $\{H_i\}_{i=1,\dots,p}$. Now, consider H_i for some $i \in \{1, \dots, p\}$ that is rejected at local level α. It is easy to see that all possible intersection hypotheses involving H_i are the hypotheses $H_1, \dots, H_i$ which, by the design of our testing procedure, must have been rejected already at local level α. With the closed testing principle (Marcus, Peritz and Ruben, 1976), it follows that H_i can be rejected at global level α. □

Additional Simulation Results. To complement the simulation results of the main text, we present in this appendix additional simulation results with the same setup as in Section 4, but with different signal-to-noise ratios

TABLE 2
Type-I error rates for DGP with $\beta = \beta_{\text{step}}$ and at global level $\alpha = 0.05$,

	$p = 100$				$p = 300$			
n	100	250	500	1000	100	250	500	1000
γ								
0.1	0.021	0.029	0.026	0.032	0.020	0.027	0.030	0.032
0.2	0.027	0.031	0.032	0.037	0.022	0.030	0.033	0.033
0.4	0.033	0.034	0.036	0.032	0.033	0.031	0.037	0.035
0.6	0.037	0.028	0.035	0.032	0.028	0.038	0.034	0.034
0.8	0.034	0.034	0.034	0.033	0.040	0.034	0.037	0.038
0.9	0.040	0.037	0.035	0.036	0.040	0.037	0.035	0.034

$\gamma = 0.1, 0.2, 0.4, 0.6, 0.8, 0.9$ and with functions X_i that are generated from a p-dimensional spline basis with independent standard normal coefficients. The covariance function of these curves is only different from zero close to the diagonal and it holds that $\mathbb{E}[X_i(t)X_i(s)] = 0$ for $|t - s| > 4\delta$, where δ is the distance between adjacent knots. It can be observed that the power in this setting is much higher, which is a consequence of the curves' covariance structure. Even in the case where the signal-to-noise ratio is 10 and the number of observations is 1000, the power in the setting of Section 4 is lower than for any of the examples shown in Figure 5.

CHRISTOPH RUST
INSTITUTE OF ECONOMETRICS
UNIVERSITY OF REGENSBURG
UNIVERSITÄTSSTRASSE 31
93053 REGENSBURG, GERMANY
E-MAIL: christoph.rust@ur.de

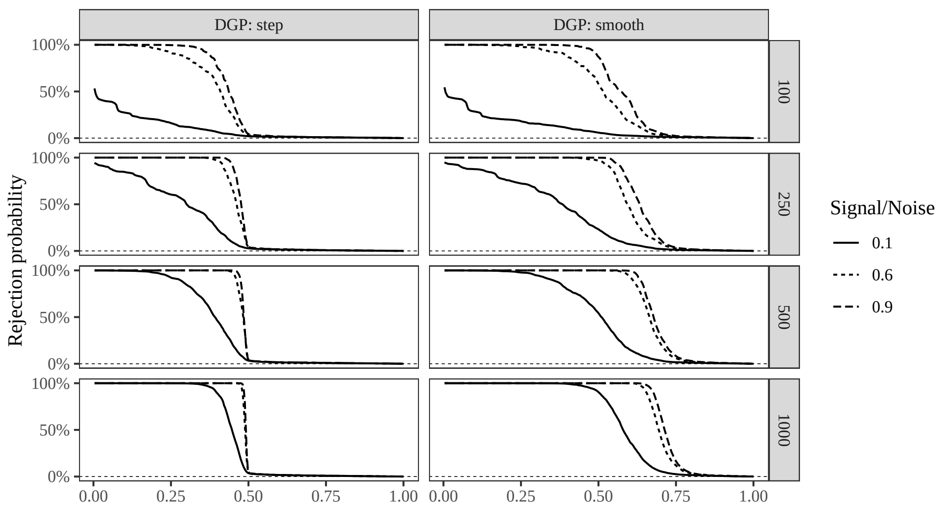

FIGURE 5. *Rejection probabilities of the directed testing procedure for the two DGPs with* $p = 300$.