Journal of Machine Learning Research 26 (2026) 1-40 Submitted 11/25; Revised XXX; Published XXX

Adaptive Simultaneous Prediction Bands in Concurrent Functional Linear Regression

Michael L. Creutzinger XXXX TODO
Department of Statistics, Colorado State University
Fort Collins, CO 80523, USA

Dominik Liebl dliebl@uni-bonn.de
Institute of Finance and Statistics, University of Bonn
Bonn, Germany

Tim Mensinger mensingertim@gmail.com
Institute of Finance and Statistics, University of Bonn
Bonn, Germany

Julia L. Sharp julia.sharp@nist.gov
National Institute of Standards and Technology
Gaithersburg, MD 20899, USA

Editor: My editor XXXXX TODO

Abstract

We develop novel, fast simultaneous prediction and confidence bands for concurrent functional linear regression with time-adaptive critical values. The bands balance the local false positive rate to enable local interpretability, provide covariate-adjusted conditional guarantees, and remain valid under fat-tailed error processes. In simulations, our bands are more precise than conformal prediction while delivering conditional (not merely marginal) coverage. We illustrate the method on Sprint Start Kinetics data motivated by World Athletics Rule 6.3.4.

Keywords: functional data analysis, simultaneous inference, statistical fairness guarantees, false positive rate balance

1 Introduction

Functional data analysis has emerged as a powerful framework for studying data observed as functions over a continuum, such as time or space (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012; Kokoszka and Reimherr, 2017). Many classical methods from multivariate statistics have been extended to this setting, including a wide range of regression models (see, for instance, Cai and Hall, 2006; Hall and Horowitz, 2007; Crambes et al., 2009). Yet, methods and theoretical guarantees for simultaneous confidence bands (SCBs) and, in particular, simultaneous prediction bands (SPBs) in functional regression remain limited.

In this paper, we contribute novel simultaneous prediction and confidence bands for functional regression models. Inference in such models is challenging and often infeasible, since the typically ill-posed inverse problem induces a lack of tightness in the estimation, which in turn prevents asymptotic normality (Cardot et al., 2007). We focus on the concurrent functional linear regression model, where this tightness issue does not arise and which is among the most frequently applied functional regression models in the literature (see, for instance, Zhang et al., 2011; Li et al., 2016; Chu et al., 2016; Ghosal et al., 2020; Torti et al., 2021; Petrovich et al., 2024; Depaoli et al., 2024). Nonetheless, our methodology also applies to other functional linear regression models without tightness issues, such as the functional factor regression model of Otto and Winter (2025).

Our simultaneous prediction bands are a key contribution. To the best of our knowledge, there is no existing method for constructing SPBs for the concurrent functional linear regression model beyond the conformal-inference approaches of Diquigiovanni et al. (2022b) and Fontana et al. (2023). However, as shown in Lei and Wasserman (2014), conformal prediction regions provide only marginal (unconditional) coverage guarantees. While there are conformal methods that achieve asymptotic conditional coverage for real-valued data (Izbicki et al., 2020), they rely on density estimation which does not extend readily to functional data (Delaigle and Hall, 2010). This lack of conditional coverage is consequential in our case study (Section 4), which requires coverage conditional on given covariate values, and our simulation studies (Section 3) further show that conformal-based SPBs are overly wide.

While point-wise prediction intervals typically assume Gaussian errors, our simultaneous prediction bands accommodate heavy-tailed error processes, which is particularly valuable in settings with costly false positives, as in our application (Section 4). Our bands build upon the work of Liebl and Reimherr (2023) and generalize their results to predictive inference for the concurrent functional linear regression model. In addition, we extend their work by establishing uniform asymptotic normality results for the concurrent functional linear regression model.

Moreover, unlike existing methods, our bands are based on adaptive critical value functions, allowing their width to adjust to different objectives. In this work, we use this feature to construct bands that support inference under fairness constraints, such as maintaining a balanced false positive rate across the domain of the functions. We also explain how this framework can be used to derive bands with minimized width.

In our application, we address the long-standing methodological challenge of assessing whether amputee sprinters gain a competitive advantage from prostheses under Rule 6.3.4 of World Athletics, which requires evidence "on the balance of probabilities" (World Athletics, 2020a,b). Existing statistical methods have failed in this context, leading to inconclusive or even retracted studies. Our simultaneous prediction bands fill this gap by providing conditional, covariate-adjusted, and distributionally robust prediction bands for high-resolution functional data such as sprint start force curves. Applying them to force data from amputee and non-amputee sprinters reveals systematic differences in the sprint start, offering strong, statistically grounded evidence relevant to Rule 6.3.4.

None of the existing works on functional predictive inference address the problem of conditional predictive inference in the concurrent linear regression model. The early works of Olshen et al. (1989) and Lenhoff et al. (1999) are inspired by biomechanic curve data, but do not allow for covariate adjustments. More recently, Franco-Villoria and Ignaccolo (2017) and Paparoditis and Shang (2023) develop SPBs for spatial functional data and functional time series, respectively, but also these works do not take into account covariate adjustments. In related works, Rathnayake and Choudhary (2016) and de Silva and Choudhary (2024) propose bootstrap-based simultaneous tolerance bands for Gaussian and exponential family functional data; however, these also do not allow for covariate adjustments. Das et al. (2023) propose a linear mixed-effects model for predictive inference in a biomechanical functional data context, however, their method is valid only point-wise and thus does not allow simultaneous inference over the domain of the curves. Kraus (2025) also contributes simultaneous prediction bands, but with a focus on reconstructing curves from noisy discrete observations that are assumed to be normally distributed.

The literature on SCBs is substantially broader than the literature on SPBs, but the majority of contributions also do not allow for covariate adjustments (see, for instance, Bunea et al., 2011; Cao et al., 2012; Cao, 2014; Degras, 2011; Wang et al., 2020; Telschow and Schwartzman, 2022; Liebl and Reimherr, 2023). SCBs that take into account covariate adjustments in a function-on-scalar regression model are contributed by Chang et al. (2017) and Abramowicz et al. (2018). Belloni et al. (2018) develop a very general theory that allows constructing SCBs for functional parameters. Chang and McKeague (2024) contribute a SCB for the concurrent functional linear regression model and is probably the closest to our work on SCBs. None of the former works on SCBs, however, allow for local inference over sub-intervals, $[a_l, b_l] \subset [a, b]$. Most similar to our SCB is the SCB of Ecker et al. (2024), which, like ours, builds upon the work of Liebl and Reimherr (2023). However, Ecker et al. (2024) require the assumption of a Gaussian error process, while our approach only assumes that the error process has finite variances and two-times continuously differentiable sample paths.

The remainder of the paper is organized as follows. Section 2 presents the model, the estimators, and our simultaneous prediction and confidence bands, including the algorithm for computing fair (Algorithm 1) critical value functions. Section 3 reports a simulation study comparing our approach with conformal inference bands. Section 4 contains the case study on Sprint Start Kinetics data, demonstrating how our methods can be used to assess Rule 6.3.4 of World Athletics (World Athletics, 2020a,b). Section 5 concludes with a discussion of the results. Mathematical derivations are provided in Appendix A. Code and data to reproduce both the simulation study and the case study are available at `https://github.com/timmens/fspb`.

2 Bands for Concurrent Regression

Let $Y = \{Y(t) \in \mathbb{R},\, t \in [a, b]\}$ and $X = \{X(t) \in \mathbb{R}^K,\, t \in [a, b]\}$ be stochastic processes, where X is twice continuously differentiable almost surely. For simplicity, and without loss of generality, we standardize the interval $[a, b] \subset \mathbb{R}$ to $[0, 1]$, and focus on this case throughout. We study the concurrent function-on-function linear regression model (Hastie and Tibshirani, 1993; Ramsay and Silverman, 2005, Ch. 14)

$$Y(t) = X^{\mathsf{T}}(t)\beta(t) + \varepsilon(t), \quad \text{for} \quad t \in [0, 1], \tag{1}$$

where $\beta = \{\beta(t) \in \mathbb{R}^K,\, t \in [0, 1]\}$ is the twice continuously differentiable parameter function and $\varepsilon = \{\varepsilon(t) \in \mathbb{R},\, t \in [0, 1]\}$ denotes an almost surely twice continuously differentiable stochastic error process, with covariance function $\sigma_\varepsilon(s, t|X) = \mathbb{E}[\varepsilon(s)\varepsilon(t)|X]$. The error is assumed to be mean independent of X, such that $\mathbb{E}[\varepsilon(t)|X] = 0$ for all $t \in [0, 1]$. We consider the case of an independent and identically distributed (iid) random sample

$(Y_1, X_1), \dots, (Y_n, X_n) \sim_{\text{iid}} (Y, X)$. The squared cross-moments of $(X^\mathsf{T}(t), \varepsilon(t))$ with itself, as well as their first and second derivatives, are assumed to be finite in the supremum norm. We further require that $\Sigma_X(s,t) = \mathbb{E}[X(s)X^\mathsf{T}(t)]$ has full rank in an open neighborhood around the diagonal $s = t$, for all $t \in [0,1]$. In the special case when X contains only constant functions, Model (1) simplifies to the function-on-scalar model.

While the estimation of confidence bands does not require a distributional assumption on the error term, such an assumption is necessary in the case of prediction bands, as required by our application. A common assumption in the literature on prediction bands is that the error term is Gaussian. To deal with heavy-tailed phenomena, and to provide conservative inference in the case of thin-tailed phenomena, which is particularly useful in cases with costly false positives, as in our application (Section 4), we model a heavy-tailed stochastic error process. Specifically, for our prediction bands, we assume that the error process is a Student's t type of process with $\nu > 4$ degrees of freedom, defined as

$$\varepsilon = Z\sqrt{\nu/\chi^2_\nu}, \tag{2}$$

where χ^2_ν is a real-valued Chi-squared distributed random variable with ν degrees of freedom that is independent of X, $Z = \{Z(t), t \in [0,1]\}$ is a mean-zero Gaussian process with covariance kernel $\sigma_Z(s,t|X) = \mathbb{E}[Z(s)Z(t)|X]$, and where Z and χ^2_ν are independent. The covariance kernel of the error process is thus given by $\sigma_\varepsilon(s,t|X) = \sigma_Z(s,t|X)\nu/(\nu-2)$. Note that the error distribution is Gaussian in the case of $\nu \to \infty$. The above-described setup and further regularity assumptions are listed in Appendix A.

2.1 Pointwise Confidence and Prediction Bands

Let $(Y_{\mathsf{new}}, X_{\mathsf{new}}) \sim (Y, X)$ denote a new data pair, independent of the original sample, following Model (1), that is $Y_{\mathsf{new}}(t) = X^\mathsf{T}_{\mathsf{new}}(t)\beta(t) + \varepsilon_{\mathsf{new}}(t)$, where $\varepsilon_{\mathsf{new}}$ is the corresponding error term. Assume we have observed the predictor function $X_{\mathsf{new}} = x_{\mathsf{new}}$, but not Y_{new}. For confidence bands (CB), we construct bands around the conditional mean response $\mathbb{E}\left[Y_{\mathsf{new}}(t)|X_{\mathsf{new}}(t) = x_{\mathsf{new}}(t)\right] = x^\mathsf{T}_{\mathsf{new}}(t)\beta(t)$, while for prediction bands (PB), we aim to construct bands around the response itself $Y_{\mathsf{new}}(t) = x^\mathsf{T}_{\mathsf{new}}(t)\beta(t) + \varepsilon_{\mathsf{new}}(t)$, taking into account the uncertainty in the error term. We begin with the derivation of pointwise confidence and prediction bands for which the coverage probability of $1-\alpha$ (for example, $\alpha = 0.05$) is guaranteed to hold at each $t \in [0,1]$. In Section 2.2, we generalize these to simultaneous confidence and prediction bands, for which the coverage probability is guaranteed to hold simultaneously for all $t \in [0,1]$.

2.1.1 Pointwise Confidence Band

A natural estimator of the conditional mean response is $x^\mathsf{T}_{\mathsf{new}}(t)\hat{\beta}(t)$, where $\hat{\beta}(t)$ denotes the functional ordinary least squares (OLS) estimator

$$\hat{\beta}(t) = \left(\sum_{i=1}^n X_i(t)X_i^\mathsf{T}(t)\right)^{-1} \sum_{i=1}^n X_i(t)Y_i(t), \quad t \in [0,1]. \tag{3}$$

From classical asymptotic theory on the OLS estimator (for example, Hayashi, 2000, chap. 2.3) it follows that, pointwise for each $t \in [0,1]$

$$\sqrt{n}\,(\hat{c}_{\mathsf{CB}}(t,t))^{-1/2}\left(x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) - x_{\mathsf{new}}^{\mathsf{T}}(t)\beta(t)\right) \xrightarrow{d} \mathcal{N}(0,1), \tag{4}$$

as $n \to \infty$, where $\hat{c}_{\mathsf{CB}}(s,t)$ is a consistent estimator of the asymptotic variance

$$c_{\mathsf{CB}}(s,t) = x_{\mathsf{new}}^{\mathsf{T}}(s)\Sigma_X^{-1}(s,t)\Sigma_{X\varepsilon}(s,t)\Sigma_X^{-1}(s,t)x_{\mathsf{new}}(t), \tag{5}$$

with $\Sigma_X(s,t) = \mathbb{E}[X(s)X^{\mathsf{T}}(t)]$ and $\Sigma_{X\varepsilon}(s,t) = \mathbb{E}[X(s)\varepsilon(s)\varepsilon(t)X(t)]$. While it is not necessary in the pointwise case to define objects for time points (s,t) instead of just t, we do so to align our notation with the simultaneous case.

Under homoskedasticity, (5) simplifies to $c_{\mathsf{CB}}(t,t) = x_{\mathsf{new}}^{\mathsf{T}}(t)\sigma_\varepsilon(t,t)\Sigma_X^{-1}(t,t)x_{\mathsf{new}}(t)$, which can be estimated using a plug-in approach with $\hat{\Sigma}_X(s,t) = n^{-1}\sum_{i=1}^n X_i(s)X_i^{\mathsf{T}}(t)$ and $\hat{\sigma}_\varepsilon(s,t) = n^{-1}\sum_{i=1}^n e_i(s)e_i(t)$, with $e_i(t) = Y_i(t) - X_i^{\mathsf{T}}(t)\hat{\beta}(t)$ denoting the residual.

Under heteroskedasticity, $\Sigma_{X\varepsilon}$ can be estimated using a heteroskedasticity-consistent estimator $\hat{\Sigma}_{X\varepsilon}$, for example, the HC3 estimator

$$\hat{\Sigma}_{X\varepsilon}(s,t) = n^{-1}\sum_{i=1}^n e_i(s)e_i(t)(1-h_i)^{-2}X_i(s)X_i^{\mathsf{T}}(t),$$

where h_i denotes the leverage score for the i-th observation (see e.g, White, 1980; Long and Ervin, 2000; Cribari-Neto, 2004). Analogously, c_{CB} can then be estimated by a plug-in approach

$$\hat{c}_{\mathsf{CB}}(t,t) = x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\Sigma}_X^{-1}(t,t)\hat{\Sigma}_{X\varepsilon}(t,t)\hat{\Sigma}_X^{-1}(t,t)x_{\mathsf{new}}(t).$$

Inverting (4) leads to the asymptotic, pointwise $(1-\alpha)$ confidence band for $x_{\mathsf{new}}^{\mathsf{T}}(t)\beta(t)$,

$$\mathrm{CB}_{1-\alpha}(t) = \left[x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) \pm \frac{q_{\alpha,n}}{\sqrt{n}}\,\hat{c}_{\mathsf{CB}}(t,t)^{1/2}\right], \tag{6}$$

where $q_{\alpha,n}$ is the $(1-\alpha/2)$-quantile of any distribution that is asymptotically equivalent to the standard normal distribution. A popular choice is the Student's t distribution with $n-K$ degrees of freedom, which is more conservative in finite samples than the standard normal distribution, and can therefore act as a finite sample correction for the otherwise unconsidered estimation errors in $\hat{c}_{\mathsf{CB}}(t,t)$; see, for example, Hansen (2022, chap. 7.13.)

2.1.2 Pointwise Prediction Band

To construct a pointwise prediction band for $Y_{\mathsf{new}}(t)$, in addition to the estimation error of $\hat{\beta}(t)$, we need to take into account the distribution of the error term $\varepsilon_{\mathsf{new}}(t)$. From (4) we know that $Y_{\mathsf{new}}(t) - x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) = o_P(1) + \varepsilon_{\mathsf{new}}(t)$, pointwise for each $t \in [0,1]$. And hence, pointwise for each $t \in [0,1]$,

$$\hat{c}_{\mathsf{PB}}(t,t)^{-1/2}\left(x_{\mathsf{new}}^{\mathsf{T}}\hat{\beta}(t) - Y_{\mathsf{new}}(t)\right) \xrightarrow{d} \mathbf{t}_\nu, \tag{7}$$

as $n \to \infty$, where $\mathbf{t}_\nu$ is Student's t-distributed with ν degrees of freedom, and where $\hat{c}_{\mathsf{PB}}(t,t)$ is a consistent estimator of the conditional scaling covariance $\sigma_Z(t,t|x_{\mathsf{new}})$. Typically,

$\hat{c}_{\mathsf{PB}}(t,t)$ also contains a finite sample correction, such as $\hat{c}_{\mathsf{PB}}(t,t) = \hat{c}_{\mathsf{CB}}(t,t)/n + \hat{\sigma}_Z(t,t|x_{\mathsf{new}})$, which can be motivated by the fact that

$$\begin{aligned}
&\mathrm{Var}[x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) - Y_{\mathsf{new}}(t) \mid \{X_i\}_{i=1}^n, X_{\mathsf{new}} = x_{\mathsf{new}}] \\
&\quad = \mathrm{Var}[x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) - (x_{\mathsf{new}}^{\mathsf{T}}(t)\beta(t) + \varepsilon_{\mathsf{new}}(t)) \mid \{X_i\}_{i=1}^n, X_{\mathsf{new}} = x_{\mathsf{new}}] \\
&\quad = \mathrm{Var}[(x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) - x_{\mathsf{new}}^{\mathsf{T}}(t)\beta(t)) - \varepsilon_{\mathsf{new}}(t) \mid \{X_i\}_{i=1}^n, X_{\mathsf{new}} = x_{\mathsf{new}}] \\
&\quad \approx \hat{c}_{\mathsf{CB}}(t,t)/n + \sigma_\varepsilon(t,t|x_{\mathsf{new}}).
\end{aligned}$$

Notice that the second summand of $\hat{c}_{\mathsf{PB}}$ is an estimator of σ_Z and not σ_ε, in order to obtain the correct limiting distribution in (7). Following the methodology of Singh (1988), the degrees of freedom parameter ν can be estimated using

$$\hat{\nu} = \begin{cases} \min_{t\in[0,1]}\{\hat{\nu}(t) : \hat{a}(t) > 3+\delta\} & \text{if } \exists t \in [0,1] : \hat{a}(t) > 3+\delta, \\ 30 & \text{otherwise,} \end{cases}$$

where $\hat{\nu}(t) = 4 + 6/(\hat{a}(t) - 3)$ are the degrees of freedom estimates at time t, with $\hat{a}(t) = n^{-1}\sum_{i=1}^n e_i(t)^4 \left(n^{-1}\sum_{i=1}^n e_i(t)^2\right)^{-2}$ denoting the kurtosis estimates of the residuals, and δ being a small positive number (for instance, $\delta = 0.1$). If the kurtosis estimates are all close to 3, we interpret this as evidence for Gaussian tails and therefore set $\hat{\nu} = 30$. Given the distributional assumption on the error term, a natural estimator of $\sigma_Z(t,t|x_{\mathsf{new}})$ is then $\hat{\sigma}_Z(t,t|x_{\mathsf{new}}) = \hat{\sigma}_\varepsilon(t,t|x_{\mathsf{new}})(\hat{\nu}-2)/\hat{\nu}$, where the conditional variance estimator $\hat{\sigma}_\varepsilon(t,t|x_{\mathsf{new}})$ can be constructed relying on functional form assumptions or non-parametric methods (see, for example, Hansen, 2022, chap. 19.15; Liebl, 2019a,b). Inverting (7) leads to the asymptotic, pointwise $(1-\alpha)$ prediction band for $Y_{\mathsf{new}}(t) = x_{\mathsf{new}}^{\mathsf{T}}(t)\beta(t) + \varepsilon_{\mathsf{new}}(t)$,

$$\mathrm{PB}_{1-\alpha}(t) = \left[x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) \pm q_{\alpha,\hat{\nu}}\, \hat{c}_{\mathsf{PB}}(t,t)^{1/2}\right], \tag{8}$$

where $q_{\alpha,\hat{\nu}}$ denotes the $(1-\alpha/2)$-quantile of the univariate Student's t distribution with $\hat{\nu}$ degrees of freedom.

2.2 Simultaneous Confidence and Prediction Bands

The pointwise bands in (6) and (8) are valid only at each $t \in [0,1]$. In general, this means that the probability of the whole curve falling in the confidence (or prediction) band is not guaranteed to be at least $1-\alpha$. Building upon the framework of Liebl and Reimherr (2023), we solve this problem by constructing time-adaptive critical value functions that take into account the correlation structure of the asymptotic process. A key object in this construction is what Liebl and Reimherr (2023) call the *roughness function* τ, defined for any covariance kernel c as

$$\tau_c(t) = \left(\frac{\partial^2}{\partial s \partial t} \frac{c(s,t)}{\sqrt{c(s,s)c(t,t)}}\bigg|_{(s,t)=(t,t)}\right)^{1/2},$$

measuring the local variability of a process with covariance c, and thereby enabling the quantification of the multiple testing problem's extent. In the following, we will show that the pointwise limit statements (4) and (7) extend to the space of continuous functions, when considering convergence of the process as a whole, and how to use these results to construct simultaneous confidence and prediction bands guided by Liebl and Reimherr (2023).

2.2.1 Simultaneous Confidence Bands

We begin this section by extending the limit result from Equation (4) to the whole process.

Theorem 1 *Let $\hat{\beta}$ be the OLS estimator of β as defined by Equation* (3)*. Under our model setup and some additional regularity conditions (see Appendix A), we find that*

$$\left\{\sqrt{n}\,(\hat{c}_{\mathsf{CB}}(t,t))^{-1/2}\left(x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t)-x_{\mathsf{new}}^{\mathsf{T}}(t)\beta(t)\right), t\in[0,1]\right\}\xrightarrow{d}\mathcal{GP}(0,c_{\mathsf{SCB}}), \tag{9}$$

where the convergence happens in $C[0,1]$ and $\mathcal{GP}(0,c_{\mathsf{SCB}})$ denotes a mean-zero Gaussian process with covariance kernel $c_{\mathsf{SCB}}(s,t)=c_{\mathsf{CB}}(s,t)/\sqrt{c_{\mathsf{CB}}(s,s)c_{\mathsf{CB}}(t,t)}$.

A proof of Theorem 1 is provided in Appendix A. Inverting (9) at a point $t\in[0,1]$ leads to the same pointwise confidence band as stated in Equation (6). For simultaneous confidence bands we need to replace the pointwise, and time-independent, critical value $q_{\alpha,n}$ with a time-adaptive critical value $u(t)$ that allows for simultaneous inference across $t\in[0,1]$. The discussion of how to construct such critical value functions is provided in the next section. Given such a $u(t)$, the simultaneous confidence band around $x_{\mathsf{new}}^{\mathsf{T}}(t)\beta(t)$ is defined as

$$\mathrm{SCB}_{1-\alpha}(t)=\left[x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t)\pm\frac{u(t)}{\sqrt{n}}\,\hat{c}_{\mathsf{CB}}(t,t)^{1/2}\right]. \tag{10}$$

2.2.2 Constructing time-adaptive critical value functions

We adapt the approach of Liebl and Reimherr (2023) by focusing on piecewise constant critical value functions instead of piecewise linear. This simplifies the theoretical derivation and speeds up the practical computation, without losing significant flexibility. For a given finite partition $0=t_0<t_1<\cdots<t_{M-1}<t_M=1$ of the interval $[0,1]$, we therefore consider critical value functions of the form

$$u(t)=\sum_{m=1}^{M-1}u_m\mathbb{1}_{[t_{m-1},t_m)}(t)+u_M\mathbb{1}_{[t_{M-1},1]}(t),$$

where $u_m>0$ denotes the critical value in the m-th interval $[t_{m-1},t_m)$. The number of intervals M and their cutoffs can be chosen based on the desired trade-off between flexibility and computational complexity, or motivated by time points of interest.

For the following, let $\mathcal{Z}\sim\mathcal{GP}(0,c_{\mathsf{SCB}})$, and consider the random counting variable

$$N_{u,\mathcal{Z}}([a,b])=\#\{t\in[a,b]:\mathcal{Z}(t)=u(t),\mathcal{Z}'(t)>0\},$$

which counts the locations where $\mathcal{Z}$ crosses the critical value function u in an upward-sloping manner on the subdomain $[a,b]\subset[0,1]$. If $N_{u,\mathcal{Z}}([a,b])=0$, the only way $\mathcal{Z}$ could have exceeded u was if $\mathcal{Z}$ started above of u at $t=a$. This logic, and an application of Boole's and Markov's inequality, leads to the expected Euler characteristic inequality that allows us to bound the one-sided simultaneous non-coverage probability:

$$\begin{aligned}\mathbb{P}\left(\exists\, t\in[a,b]:\mathcal{Z}(t)\geq u(t)\right)&=\mathbb{P}\left(\mathcal{Z}(a)\geq u(a)\ \text{ or }\ N_{u,\mathcal{Z}}([a,b])\geq 1\right)\\&\leq\mathbb{P}\big(\mathcal{Z}(a)\geq u(a)\big)+\mathbb{P}\big(N_{u,\mathcal{Z}}([a,b])\geq 1\big)\\&\leq\mathbb{P}\big(\mathcal{Z}(a)\geq u(a)\big)+\mathbb{E}\left[N_{u,\mathcal{Z}}([a,b])\right]\\&=\mathbb{E}\left[\varphi_u([a,b])\right],\end{aligned} \tag{11}$$

where $\varphi_u([a,b]) = \mathbb{1}_{\{\mathcal{Z}(a)\geq u(a)\}} + N_{u,\mathcal{Z}}([a,b])$ denotes the Euler characteristic of the excursion set $\{t \in [a,b] : \mathcal{Z}(t) \geq u(t)\}$.

The sharpness of inequality (11) has been analyzed under various assumptions on the stochastic process $\mathcal{Z}$. Piterbarg (1982) and Azaïs et al. (2002) studied smooth Gaussian processes with stationary covariance functions (that is $\mathrm{Cov}(\mathcal{Z}(t), \mathcal{Z}(s)) = \mathrm{Cov}(\mathcal{Z}(0), \mathcal{Z}(|t - s|))$ for all $t, s \in [0,1]$) and without global singularities (that is $\mathrm{Corr}(\mathcal{Z}(t), \mathcal{Z}(s)) = 1$ only if $t = s$). Under additional technical conditions, they proved that for small significance levels ($\alpha \to 0$) and thus for large critical values $u = u_\alpha \to \infty$ the approximation error satisfies $|\mathbb{P}\big(\exists t \in [0,1] : \mathcal{Z}(t) \geq u(t)\big) - \mathbb{E}[\varphi_{u,X}(0)]| \leq c_1 \exp\big(-c_2 u^2/2\big)$ for some constants $c_1, c_2 > 0$. Extending this, Taylor et al. (2005) showed analogous results for non-stationary covariance functions. In our setting with non-constant critical value functions $u(t)$, the same conclusions follow by taking $\bar{u} = \min_{t\in[0,1]} u(t)$ as $\bar{u} = \bar{u}_\alpha \to \infty$ for small $\alpha \to 0$. Note that global singularities are permitted in (11), though they reduce approximation sharpness.

Explicit formulas for the expectation of an Euler characteristic are often called Kac-Rice formulas, acknowledging the works of Kac (1943) and Rice (1945); see also Adler and Taylor (2007). A function $u_{\alpha,n}(t)$ that solves

$$\mathbb{E}\left[\varphi_{u_{\alpha,n}}([0,1])\right] = \alpha/2$$

is a valid critical value function to the one-sided simultaneous inference problem, as it bounds the one-sided simultaneous non-coverage probability (11) from above by $\alpha/2$. For any symmetric limit distribution, such as the Gaussian process $\mathcal{Z}$, the critical value function $u_{\alpha,n}$ also bounds the two-sided simultaneous non-coverage probability

$$\mathbb{P}\left(\exists\, t \in [0,1] : \mathcal{Z}(t) \notin [-u_{\alpha,n}(t), u_{\alpha,n}(t)]\right) \leq \alpha.$$

Liebl and Reimherr (2023) derive Kac-Rice formulas for general elliptical processes. The following corollary provides a simplified version of their Kac-Rice formula under our setup.

Corollary 2 *(Based on Corollary 3.2 in Liebl and Reimherr, 2023.) Let τ_{SCB} denote the roughness function corresponding to the covariance kernel of $\mathcal{Z} \sim \mathcal{GP}(0, c_{\mathsf{SCB}})$. Additionally, let $[a,b] \subset [0,1]$ and assume that $u'(t) = 0$ almost everywhere. Then*

$$\mathbb{E}\left[\varphi_u([a,b])\right] = \Phi(-u(a)) + \int_a^b \frac{\tau_{\mathsf{SCB}}(t)}{2\pi} \exp\left(-\frac{1}{2}u(t)^2\right) \mathrm{d}t, \tag{12}$$

where Φ denotes the standard normal cumulative distribution function.

The only unknown parameter in (12) is the roughness function τ_{SCB}, which can be estimated consistently uniformly over $t \in [0,1]$ by

$$\hat{\tau}_{\mathsf{SCB}}(t) = \left(\partial_{s,t}\, \hat{c}_{\mathsf{SCB}}(s,t)\big|_{(s,t)=(t,t)}\right)^{1/2}, \tag{13}$$

where $\hat{c}_{\mathsf{SCB}}(s,t)$ denotes a uniformly consistent estimator of the covariance kernel, such as the plug-in estimator $\hat{c}_{\mathsf{SCB}}(s,t) = \hat{c}_{\mathsf{CB}}(s,t)/\sqrt{\hat{c}_{\mathsf{CB}}(s,s)\hat{c}_{\mathsf{CB}}(t,t)}$; see Section 2.1.2 for the definition of $\hat{c}_{\mathsf{CB}}$, and Appendix A for a proof of uniform consistency. This leads to the empirical Kac-Rice equation

$$\frac{\alpha}{2} = \Phi(-u_1) + \sum_{m=1}^{M} \exp(-u_m^2/2)\frac{1}{2\pi}\int_{t_{m-1}}^{t_m} \hat{\tau}_{\mathsf{SCB}}(t)\mathrm{d}t, \tag{14}$$

where we replaced the arbitrary critical value function $u(t)$ with the piecewise constant function $u(t) = u_m$ for $t \in [t_{m-1}, t_m)$. Equation (14) has infinitely many solutions, which means that we can choose $(u_1, \ldots, u_M)$ such that additional constraints are satisfied. An intuitive choice is to solve (14) such that the width of the resulting confidence band is small, for example, by minimizing the squared width of the band $\int u(t)^2/n\, \hat{c}_{\mathsf{CB}}(t,t)\mathrm{d}t = \sum_{m=1}^{M} u_m^2/n \int_{t_{m-1}}^{t_m} \hat{c}_{\mathsf{CB}}(t,t)\mathrm{d}t$, subject to (14). Another approach is to choose each u_m, $m = 1, \ldots, M$, such that the significance level is distributed according to the share of the interval length, that is

$$\mathbb{P}\left(\exists\, t \in [t_{m-1}, t_m] : \mathcal{Z}(t) \notin [-u_m, u_m]\right) \leq \alpha\,(t_m - t_{m-1}),$$

where each u_m is determined by the local Kac-Rice equation

$$\frac{\alpha}{2}(t_m - t_{m-1}) = \Phi(-u_m) + \exp(-u_m^2/2)\frac{1}{2\pi}\int_{t_{m-1}}^{t_m} \hat{\tau}_{\mathsf{SCB}}(t)\mathrm{d}t.$$

The latter approach is what Liebl and Reimherr (2023) call *fair*, as it ensures that the false positive rate is balanced across the intervals $[t_{m-1}, t_m]$, $m = 1, \ldots, M$. As in the case of pointwise confidence bands, we typically want to incorporate a finite sample correction to account for the asymptotically negligible, but practically relevant estimation errors in $\hat{c}_{\mathsf{SCB}}(t,t)$. For this case, we derive alternative formulas to (12) and (14) for the Student's t type process $\mathcal{Z}(t)((n-K)/\chi^2_{n-K})^{1/2}$, where χ^2_{n-K} denotes a Chi-square distributed random variable with $n - K$ degrees of freedom (see Corollary 3). The distribution of $\mathcal{Z}(t)((n-K)/\chi^2_{n-K})^{1/2}$ is heavier tailed than that of $\mathcal{Z}(t)$, for every finite $n > K$, and therefore leads to more conservative inference in finite sampling, which can act as a finite sample correction. In order to incorporate these different cases into one algorithm, we abstract from the specific distribution and scaling of the integral of the roughness function in (14), and propose Algorithm 1, an algorithm that computes fair critical values u_m that satisfy the global and local Kac-Rice equations. We regain (14) by setting $F = \Phi$ and $S(u) = \exp(-u^2/2)/2\pi$.

2.2.3 Simultaneous Prediction Bands

To construct a simultaneous prediction band for $Y_{\mathsf{new}}(t)$, we need to consider the distribution of the error term $\varepsilon_{\mathsf{new}}(t)$. Extending the limit result from Equation (7) to the whole process, we directly have that

$$\left\{\hat{c}_{\mathsf{PB}}(t,t)^{-1/2}\left(x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) - Y_{\mathsf{new}}(t)\right), t \in [0,1]\right\} \xrightarrow{d} \mathcal{T}_\nu,$$

as $n \to \infty$, where $\hat{c}_{\mathsf{PB}}(t,t)$ is a consistent estimator of the conditional scaling covariance $\sigma_Z(t,t|x_{\mathsf{new}})$, and where $\mathcal{T}_\nu$ is a Student's t-type distributed process with ν degrees of freedom, and correlation kernel $c_{\mathsf{SPB}}(s,t) = \sigma_Z(s,t|x_{\mathsf{new}})/\sqrt{\sigma_Z(s,s|x_{\mathsf{new}})\sigma_Z(t,t|x_{\mathsf{new}})}$. In line with the motivation in the section on pointwise prediction bands, we typically add a finite sample correction to $\hat{c}_{\mathsf{PB}}(t,t)$, such as $\hat{c}_{\mathsf{PB}}(t,t) = \hat{c}_{\mathsf{CB}}(t,t)/n + \hat{\sigma}_Z(t,t|x_{\mathsf{new}})$. In this case, we estimate c_{SPB} by the consistent plug-in estimator $\hat{c}_{\mathsf{SPB}}(s,t) = \hat{c}_{\mathsf{PB}}(s,t)/\sqrt{\hat{c}_{\mathsf{PB}}(s,s)\hat{c}_{\mathsf{PB}}(t,t)}$. For the construction of $\hat{\sigma}_Z$, see Section 2.1.2. Given a valid critical value function $u(t)$, the simultaneous prediction band around $Y_{\mathsf{new}}(t)$ is defined as

$$\mathrm{SPB}_{1-\alpha}(t) = \left[x_{\mathsf{new}}^{\mathsf{T}}(t)\hat{\beta}(t) \pm u(t)\, \hat{c}_{\mathsf{PB}}(t,t)^{1/2}\right]. \tag{15}$$

To construct $u(t)$, we can use a similar approach as for simultaneous confidence bands. The main difference is that we need to adapt the Kac-Rice formula (12) to the Student's t-type process $\mathcal{T}_\nu$. The following corollary provides a simplified version of the Kac-Rice formula under our setup.

Corollary 3 *(Based on Corollary 3.3 in Liebl and Reimherr, 2023.) Let τ_{SPB} denote the roughness function corresponding to the covariance kernel of the Student's t-type distributed process $\mathcal{T}_\nu$ with ν degrees of freedom. Additionally, let $[a,b] \subset [0,1]$ and assume that $u'(t) = 0$ almost everywhere. Then*

$$\mathbb{E}\left[\varphi_u([a,b])\right] = F_{\mathbf{t}}(-u(a);\nu) + \int_a^b \frac{\tau_{\mathsf{SPB}}(t)}{2\pi}\left(1 + \frac{u(t)^2}{\nu}\right)^{-\nu/2} \mathrm{d}t,$$

where $F_{\mathbf{t}}(-u(a);\nu)$ denotes the cumulative distribution function of a Student's t-distributed random variable with ν degrees of freedom.

This gives us the empirical Kac-Rice equation for the Student's t-distributed process

$$\frac{\alpha}{2} = F_{\mathbf{t}}(-u_1;\nu) + \sum_{m=1}^{M}\left(1 + \frac{u_m^2}{\nu}\right)^{-\nu/2} \frac{1}{2\pi}\int_{t_{m-1}}^{t_m} \hat{\tau}_{\mathsf{SPB}}(t)\mathrm{d}t.$$

We obtain fair critical values using Algorithm 1 (setting $F = F_{\mathbf{t}}(\cdot;\nu)$ and $S(u) = (1 + u^2/\nu)^{-\nu/2}/2\pi$). The uniformly consistent estimated roughness function $\hat{\tau}_{\mathsf{SPB}}$ is defined using the same approach as in (13), replacing $\hat{c}_{\mathsf{SCB}}$ with $\hat{c}_{\mathsf{SPB}}$.

Algorithm 1: Fair critical value function selection (two-sided)

Input:

Significance level $\alpha \in (0,1)$ and time points $t_0 = 0 < t_1 < \cdots < t_M = 1$
Cumulative distribution function F and integral scaling $S : \mathbb{R}_+ \to \mathbb{R}$
Estimated roughness function $\hat{\tau} : [0,1] \to \mathbb{R}$

for $m = 1$ **to** M **do**

Solve: $F(-u^\star_{\alpha,m}) + S(u^\star_{\alpha,m}) \int_{t_{m-1}}^{t_m} \hat{\tau}(t)\mathrm{d}t = \alpha/2\,(t_m - t_{m-1})$

return $(u^\star_{\alpha,1}, \ldots, u^\star_{\alpha,M})$

3 Simulation Study

We assess the finite-sample properties of the proposed simultaneous prediction bands (SPB) by means of a comprehensive simulation study. Throughout, we contrast the SPB with the conformal inference approach of Diquigiovanni et al. (2022b) and Fontana et al. (2023).

3.1 Data-generating Process

We simulate a random sample from a concurrent functional linear regression model

$$Y_i(t) = \beta_0(t) + X_i(t)\beta_1(t) + \varepsilon_i(t), \quad i = 1, \dots, n,$$

with coefficient functions $\beta_0(t) = \exp(-2t)/2$ and $\beta_1(t) = t \cdot \sin(2\pi t)$. The predictor $X_i(t)$ is a randomly scaled cosine function with a vertical shift that is positive with probability 0.5 and otherwise negative; specifically, we set

$$X_i(t) = B_i + S_i \cdot \cos(2\pi t),$$

with $B_i \sim_{iid} B$, where $B \in \{-1, 1\}$ with $\mathbb{P}(B = -1) = \mathbb{P}(B = 1) = 0.5$, and $S_i \sim_{iid} \mathcal{U}[0.75, 1.25]$. The error process $\varepsilon_i(t)$ is drawn independently of a Student's-t type of process (as in Equation 2) for which we consider two different covariance structures and fixed degrees of freedom ν. For both structures, the covariance is given by a Matérn kernel

$$C(s,t) = (1/3)^2 \left(2^{1-\gamma_{st}}/\Gamma(\gamma_{st})\right) \left(\sqrt{2\gamma_{st}}\,|t-s|\right)^{\gamma_{st}} K_{\gamma_{st}} \left(\sqrt{2\gamma_{st}}\,|t-s|\right),$$

where Γ is the gamma function, $K_{\gamma_{st}}$ is the modified Bessel function of the second kind, and $\gamma_{st} > 0$ controls the roughness of the sample paths. For the first structure, we set $\gamma_{st} = 3/2$ for all $s, t \in [0, 1]$, which leads to a stationary process and continuously differentiable sample paths. For the second structure, we set $\gamma_{st} = 2 + \sqrt{\max(s,t)} \cdot (1/4 - 2)$, which results in a non-stationary process and sample paths that begin smooth in the sampling domain and later become rough. A visualization of the outcomes $Y_i(t)$ induced by this setup is given in Figure 1. To mimic empirical functional data, all curves are discretized on an equidistant grid of 101 points.

3.2 Design

We examine eight simulation scenarios by varying three factors: the sample size (n), the degrees of freedom (ν), and the type of covariance structure (γ_{st}):

$$n \in \{30, 100\}, \quad \nu \in \{5, 15\}, \quad \gamma_{st} \in \{\text{Stationary}, \text{Non-Stationary}\},$$

with stationary meaning $\gamma_{st} = 3/2$, and non-stationary $\gamma_{st} = 2 + \sqrt{\max(s,t)} \cdot (1/4 - 2)$. For each of the eight scenarios, we conduct 1 000 simulation runs. During every iteration, we simulate a random sample $\{(Y_i(\cdot), X_i(\cdot)) : i = 1, \dots, n\}$ from the model, as well as an independent new observation $(Y_{\mathsf{new}}(\cdot), X_{\mathsf{new}}(\cdot))$. We then calculate the prediction and confidence bands for the new observation $Y_{\mathsf{new}}(\cdot)$ given $X_{\mathsf{new}}(\cdot)$, using the proposed method and the conformal inference method. The bands are then evaluated based on several criteria.

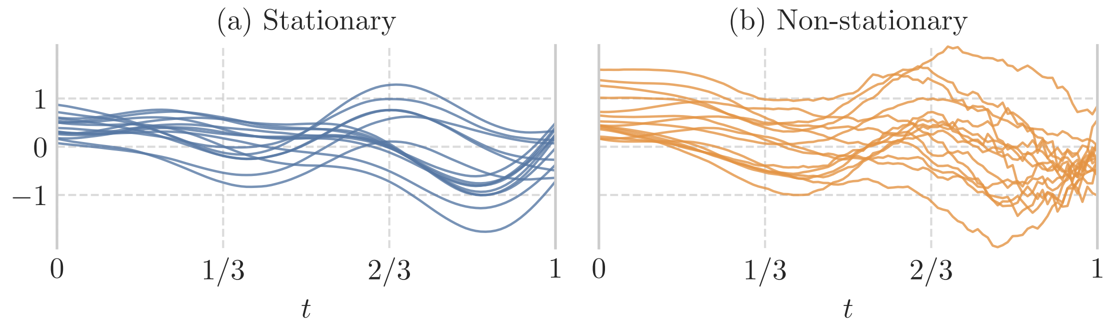

Figure 1: A random sample of $Y_i(t)$ from the simulation model, with $\nu = 5$ degrees of freedom. The left panel (a) shows data generated using the stationary (blue) Matérn covariance structure, while the right panel (b) shows data generated using the non-stationary (orange) Matérn covariance structure.

3.3 Evaluation Criteria

Let $L(t)$ and $U(t)$ denote the lower and upper bounds of a band at time $t \in [0, 1]$. We consider three criteria to evaluate the performance of such bands.

3.3.1 Coverage

For confidence bands, we assess the simultaneous coverage by checking that for all $t \in [0, 1]$: $L(t) \leq X_{\text{new}}(t)^\top \beta_n(t) \leq U(t)$. For prediction bands, we verify that for all $t \in [0, 1] : L(t) \leq Y_{\text{new}}(t) \leq U(t)$.

3.3.2 Maximum width

Since an infinitely-wide band would always cover the true function, we are interested in the smallest bands that achieve a given coverage. We evaluate the width of a band using the maximum width statistic $\max_{t\in[0,1]}|U(t) - L(t)|$.

3.3.3 Band score

To assess the trade-off between the width of the band and its ability to cover the new observation $Y_{\text{new}}(t)$, we use the *band score*, a functional version of the interval score (Gneiting and Raftery, 2007). The band score is defined as

$$\max_{t\in[0,1]}|U(t) - L(t)| + \frac{2}{\alpha} \max_{t\in[0,1]} (L(t) - Y_{\text{new}}(t))\, \mathbf{1}(Y_{\text{new}}(t) < L(t)) + \frac{2}{\alpha} \max_{t\in[0,1]} (Y_{\text{new}}(t) - U(t))\, \mathbf{1}(Y_{\text{new}}(t) > U(t)),$$

where α is the significance level. This metric penalizes bands with large maximum widths when they cover the new observation and penalizes the distance between the band and $Y_{\text{new}}(t)$ when the band fails to cover it.

3.4 Results

We compare the performance of our prediction bands with the conformal inference prediction band for functional data of Diquigiovanni et al. (2022b); Fontana et al. (2023). The conformal inference method was implemented using the R package `conformalInference.fd` (Diquigiovanni et al., 2022a).

In Table 1, we show the simulation results for our 90%-prediction bands. The critical value function is calculated using Algorithm 1 based on $F = F_{\mathtt{t}}(\cdot;\nu)$ and $S(u) = (1 + u^2/\nu)^{-\nu/2}/2\pi$. Our approach and the conformal inference method achieve roughly the same conservative empirical coverage in the small sample size case ($n = 30$) under both covariance structures. In the larger sample size case ($n = 100$), our approach remains conservative, while the conformal inference method approaches the nominal coverage more closely, with a slight undercoverage in two cases. Although, our method is weakly more conservative, the maximum width of our bands is substantially smaller in all scenarios but one. The same trend is observed for the band score, where we achieve a better score in all but two cases.

In Table 2, we present the simulation results for our 90%-confidence bands. The critical value function is calculated using Algorithm 1 based on $F = \Phi$ and $S(u) = \exp(-u^2/2)/2\pi$. We do not report results for the conformal inference method, as they are not applicable in the confidence bands case. While the empirical coverage is slightly conservative in most cases, it is less conservative than in the prediction bands case.

Figure 2 shows an example of the 90%-prediction bands generated using the conformal inference method and our method (Fair).

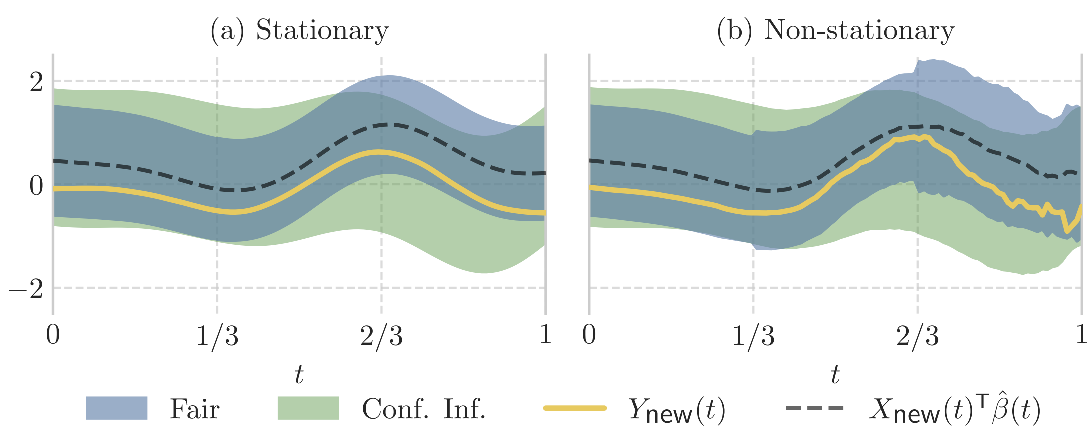

Figure 2: 90%-conditional prediction bands for the outcome Y given covariates $X_{\mathsf{new}}(t)$. The yellow line represents the corresponding observed outcomes $Y_{\mathsf{new}}(t)$. The bands are constructed using our Fair method (blue) and the Conformal inference method (green). The dashed black line represents the predicted outcome $X_{\mathsf{new}}(t)^\intercal \hat{\beta}(t)$. In the left panel (a) the error process is generated using the stationary Matérn covariance structure, while the right panel (b) shows data generated using the non-stationary Matérn covariance structure. In both cases we set $\nu = 15$, and estimate the bands on the same $n = 30$ samples.

Stationary

n	ν	Coverage		Maximum Width		Band Score	
		Fair	Conf. Inf.	Fair	Conf. Inf.	Fair	Conf. Inf.
30	5	0.93 (0.25)	0.93 (0.26)	2.42 (0.50)	3.46 (0.88)	2.78 (2.05)	3.74 (1.62)
	15	0.94 (0.23)	0.94 (0.24)	1.94 (0.24)	3.03 (0.52)	2.11 (1.00)	3.22 (0.97)
100	5	0.94 (0.24)	0.89 (0.31)	2.29 (0.27)	2.75 (0.26)	2.67 (2.10)	3.33 (2.45)
	15	0.94 (0.23)	0.90 (0.30)	1.83 (0.13)	2.57 (0.20)	2.02 (1.23)	2.97 (1.61)

Non-Stationary

n	ν	Coverage		Maximum Width		Band Score	
		Fair	Conf. Inf.	Fair	Conf. Inf.	Fair	Conf. Inf.
30	5	0.94 (0.24)	0.95 (0.23)	3.79 (0.81)	3.84 (1.02)	4.17 (2.24)	4.12 (1.79)
	15	0.95 (0.21)	0.94 (0.23)	2.88 (0.36)	3.25 (0.53)	3.03 (0.95)	3.42 (0.98)
100	5	0.94 (0.24)	0.90 (0.30)	3.52 (0.47)	3.03 (0.27)	3.94 (2.43)	3.68 (2.89)
	15	0.95 (0.21)	0.89 (0.31)	2.65 (0.21)	2.77 (0.20)	2.83 (1.19)	3.14 (1.48)

Table 1: Simulation results for the coverage, maximum width, and band score of our prediction bands for $\alpha = 10\%$ under a stationary (top table) and non-stationary (bottom table) Matérn covariance structure. The results are reported for varying sample sizes (n) and degrees of freedom (ν), and compared to the conformal inference method. Means and standard deviations (in parentheses) are calculated over 1 000 simulation runs, and rounded to two decimal places.

n	ν	Coverage		Maximum Width		Band Score	
		Stat.	Non-Stat.	Stat.	Non-Stat.	Stat.	Non-Stat.
30	5	0.93 (0.26)	0.92 (0.28)	0.59 (0.13)	0.78 (0.17)	0.66 (0.36)	0.85 (0.36)
	15	0.92 (0.27)	0.91 (0.29)	0.48 (0.08)	0.63 (0.09)	0.53 (0.26)	0.69 (0.27)
100	5	0.93 (0.26)	0.94 (0.24)	0.31 (0.04)	0.40 (0.05)	0.34 (0.14)	0.42 (0.14)
	15	0.92 (0.27)	0.92 (0.27)	0.25 (0.02)	0.32 (0.03)	0.29 (0.16)	0.36 (0.15)

Table 2: Simulation results for the coverage, maximum width, and band score of our Fair confidence bands for $\alpha = 10\%$ under a stationary (Stat.) and non-stationary (Non-Stat.) Matérn covariance structure. The results are reported for varying sample sizes (n) and degrees of freedom (ν). Means and standard deviations (in parentheses) are calculated over 1 000 simulation runs, and rounded to two decimal places.

4 Application: Sprint Start Kinetics

Determining whether amputee sprinters gain a competitive advantage from prostheses is a long-standing challenge in sports science. World Athletics Rule 6.3.4 requires evidence "on the balance of probabilities" that a mechanical aid does not confer an overall competitive advantage (World Athletics, 2020a,b). We illustrate our methodology on this question.

To judge "on the balance of probabilities," one must compare an amputee sprinter's movement/force pattern to a probabilistically justified *conditional* range from *analogous* non-amputee sprinters. We therefore apply our conditional simultaneous prediction bands to sprint-start force trajectories.

4.1 Data

We use the Sprint Start Kinetics data of Willwacher et al. (2016) comprising 154 non-amputee sprinters and 7 amputee sprinters. The functional response is the vertical force at the front starting block during the push-off phase, scaled to 0-100 % of push-off and discretized at 101 points. Scalar predictors are age, height, mass, sex, and push-time. The original data are misaligned due to inter-individual differences in the timing of peak force. To correct for this, we align the trajectories to the subject-specific peak, truncate post-peak, and then linearly inter- and extrapolate onto the original time grid (see Figure 3).

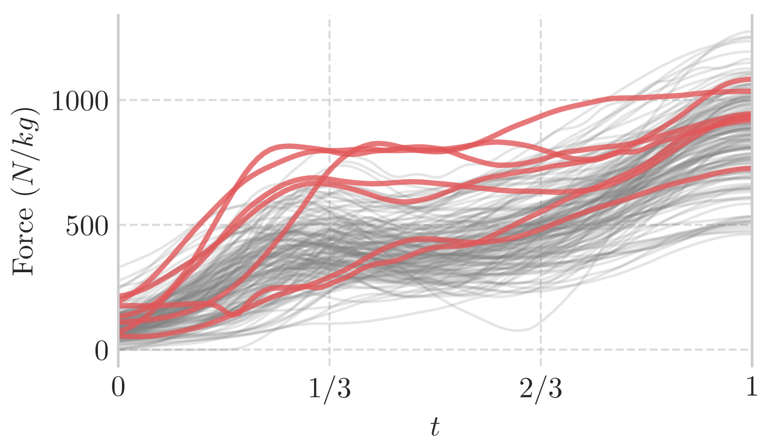

Figure 3: Front vertical force trajectories for amputee (red) and non-amputee (gray) sprinters, realigned at peak and truncated thereafter.

4.2 Analysis Setup

We fit a function-on-scalar concurrent regression using only non-amputee sprinters, and then construct 90%-conditional simultaneous prediction bands for the non-amputee sprinters with the same covariates as each amputee. As a sanity check, for each amputee we also select a nearest-neighbor non-amputee with similar covariates.

4.3 Findings

At the 90% level, 5 out of 7 amputee trajectories exhibit at least one exceedance of their conditional band, with exceedances concentrated in the middle third of push-off, whereas, with one exception, matched non-amputee counterparts lie entirely within their bands. For reference, conformal bands flag 3 out of 7 amputees. Figure 4 illustrates the analysis result for a single amputee. The trajectory of this individual exceeds the band in the middle third under our method, but not under the conformal method. These results provide statistically grounded and conditional evidence relevant to Rule 6.3.4 by deciding if and localizing where amputee and non-amputee force patterns differ during the sprint start.

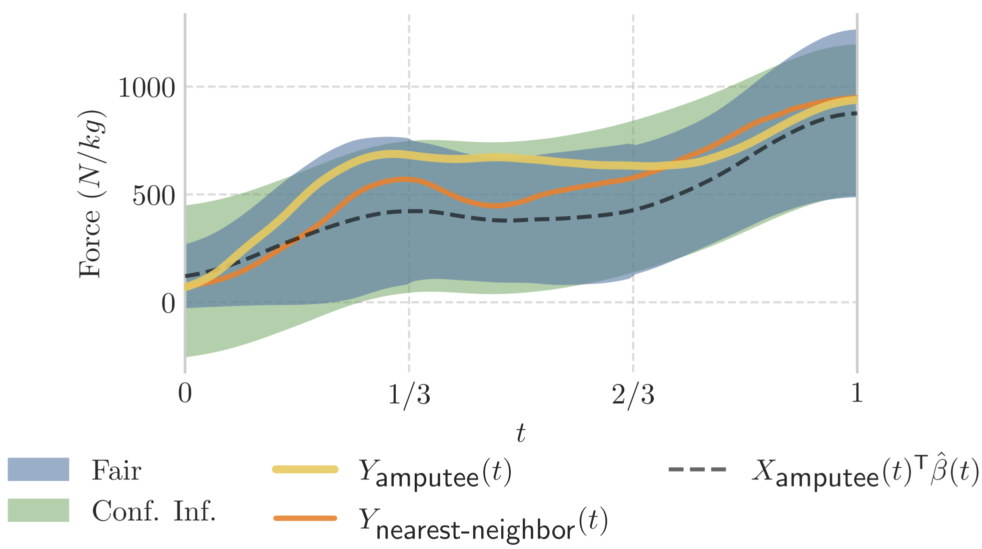

Figure 4: 90%-conditional prediction bands for the front vertical force given covariates $X_{\mathsf{amputee}}(t)$ of an amputee. The yellow line represents the corresponding observed outcomes $Y_{\mathsf{amputee}}(t)$ of the amputee. The bands are constructed using our Fair method (blue), and the Conformal Inference method (green). The dashed black line represents the predicted outcome $X_{\mathsf{amputee}}(t)^{\mathsf{T}}\hat{\beta}(t)$. The orange line is the trajectory $Y_{\mathsf{nearest\text{-}neighbor}}(t)$ of a matched non-amputee sprinter with similar covariates.

5 Conclusion

We developed simultaneous confidence and prediction bands (SCB/SPB) for concurrent functional linear regression that enable *conditional*, locally interpretable inference across the domain. On the theory side, we establish a functional central limit theorem for predicted conditional mean process (Theorem 1) and combine it with piecewise-constant critical values obtained via an empirical Kac-Rice formula. This delivers bands that adapt to local uncertainty and support conditional conclusions rather than purely marginal guarantees.

In our simulations, across stationary and non-stationary error structures and $n \in \{30, 100\}$, our proposed SPBs demonstrated improved performance compared to conformal prediction bands. Our SCBs empirical coverage is close to nominal levels and slightly less conservative than the SPBs, as expected. See Tables 1 and 1 for details.

The sprint-start case study illustrates practical value: at the 90% level, conditional SPBs reveal localized exceedances for several amputee trajectories concentrated in the middle third of push-off, whereas matched non-amputee trajectories largely remain within band. This provides statistically grounded and conditional evidence relevant to assessments under Rule 6.3.4.

Our analysis relies on a linear functional regression model with homoskedastic errors, random sampling, C^2 smoothness of underlying functions with finite moment conditions, a full-rank condition ensuring identification, and positive roughness preventing degenerate correlation; SPBs additionally assume a Student-t-type error process to accommodate heavy tails. Importantly, our results extend to heteroskedastic settings with appropriate covariance estimation.

Several extensions merit investigation. While our framework assumes i.i.d. sampling, the theoretical foundation supports extension to weakly dependent functional time series via mixing conditions. The method could accommodate sparsely observed functional data through pre-smoothing or basis expansions. Extensions to higher-dimensional domains (e.g., $[0, 1]^2$ for image data) represent natural generalizations. Finally, alternative critical value optimizations, including minimum-width strategies and smoother parametric forms beyond piecewise-constant designs, could further enhance local adaptation while maintaining computational tractability.

Acknowledgments and Disclosure of Funding

TODO NEEDS TO BE WRITTEN All acknowledgements go at the end of the paper before appendices and references. Moreover, you are required to declare funding (financial activities supporting the submitted work) and competing interests (related financial activities outside the submitted work). More information about this disclosure can be found on the JMLR website.

Appendix A. Proofs

In the following, we provide the proofs of the theoretical results presented in the main text. We start by introducing some definitions. In Section A.1, we list the assumptions under which we develop our theory. In Section A.2, we present some auxiliary lemmas that are used in the main proofs. Section A.3 contains the proofs of the main results.

Definition 4 (Roughness Parameter) *Let $c : [0, 1]^2 \to \mathbb{R}$ be a differentiable covariance kernel. The roughness parameter $\tau_c : [0, 1] \to \mathbb{R}$ associated with c, is defined as*

$$\tau_c(t) = \left(\frac{\partial^2}{\partial s \partial t} \frac{c(s,t)}{\sqrt{c(s,s)c(t,t)}} \bigg|_{(s,t)=(t,t)} \right)^{1/2}.$$

Definition 5 (Differential Operator) *Let $f(t) : [0,1] \to K$ and $g(s,t) : [0,1]^2 \to K$ be differentiable functions, where K is a Banach space (in this text, $\mathbb{R}$, $\mathbb{R}^K$ or $\mathbb{R}^{K\times K}$). The differential operator $D^{(d)}$ is defined as*

$$D^{(d)} f(t) = \frac{\partial^d}{(\partial t)^d} f(t), \quad d \in \{0,1,2\},$$

and we sometimes write $f^{(d)}(t) = D^{(d)} f(t)$. Note that $f^{(0)} = D^{(0)} f = f$. The partial differential operator $D^{(d,p)}$ is defined as

$$D^{(d,p)} g(s,t) = \frac{\partial^d}{(\partial s)^d} \frac{\partial^p}{(\partial t)^p} g(s,t), \quad d,p \in \{0,1\},$$

and we sometimes write $g^{(d,p)}(s,t) = D^{(d,p)} g(s,t)$. Note that $g^{(0,0)} = D^{(0,0)} g = g$.

A.1 Assumptions

The following list summarizes the assumptions under which we develop our uniform confidence and prediction bands. The confidence bands require Assumption 1 - 6, while the prediction bands require the additional Assumption 7. We discuss the assumptions in Section A.1.1.

1 Linear model:
$$Y(t) = X^{\mathsf{T}}(t)\beta(t) + \varepsilon(t), \quad t \in [0,1],$$
where $X = \{X(t) \in \mathbb{R}^K : t \in [0,1]\}$ is a K-dimensional vector-valued random function $X(t) = (X_1(t), \dots, X_K(t))^{\mathsf{T}}$ with intercept $X_1(t) = 1$, for all $t \in [0,1]$. Further, $\varepsilon = \{\varepsilon(t) : t \in [0,1]\}$ denotes the unobserved error function with $\mathbb{E}[\varepsilon(t)|X(t)] = 0$ for all $t \in [0,1]$.

2 Homoskedasticity: The covariance kernel of the error term ε is assumed to not depend on the predictor functions X:

$$\mathbb{E}[\varepsilon(s)\varepsilon(t)|X] = \mathbb{E}[\varepsilon(s)\varepsilon(t)] = \sigma_\varepsilon(s,t).$$

3 Random sampling: $\{(Y_i, X_i)|i = 1, \dots, n\}$ is an iid sample from (Y, X).

4 Smoothness and moments:

Define
$$(G_1(t), \dots, G_K(t), G_{K+1}(t))^{\mathsf{T}} = (X_1(t), \dots, X_K(t), \varepsilon(t))^{\mathsf{T}}.$$

For all $j,k = 1, \dots, K+1$,

(a) The random functions G_j and their expectations are twice continuously differentiable almost surely, that is $G_j \in C^2[0,1]$ and $\mathbb{E}[G_j] \in C^2[0,1]$ almost surely.

(b) There exists a finite constant $C < \infty$ such that

$$\mathbb{E}\left[\sup_{t\in[0,1]} \left| D^d \left(G_k(t)G_j(t)\right)\right|^2\right] \leq C < \infty$$

for all $d \in \{0,1,2\}$, where D^d is the differential operator from Definition 5.

(c) The regression coefficient functions are twice continuously differentiable almost surely, that is $\beta_j \in C^2[0,1]$ almost surely.

5 Rank condition: There exists a positive constant $\delta > 0$, such that for all $s, t \in [0,1]$ with $|s-t| < \delta$, $\Sigma_X(s,t) = \mathbb{E}[X(s)X^{\mathsf{T}}(t)]$ has full rank.

6 Roughness condition:

Let $x_{\mathsf{new}} : [0,1] \to \mathbb{R}^K \in C^1([0,1], \mathbb{R}^K)$ be a continuously differentiable deterministic function with $x_{\mathsf{new}}(t) \neq 0$ for all $t \in [0,1]$. Define $\sigma_{x_{\mathsf{new}}}(s,t) = x_{\mathsf{new}}^{\mathsf{T}}(s)\Sigma_X^{-1}(s,t)x_{\mathsf{new}}(t)$. Then, for all $t \in [0,1]$: $\tau_\sigma(t) > 0$ for $\sigma \in \{\sigma_\varepsilon, \sigma_{x_{\mathsf{new}}}\}$, where τ_σ is the roughness parameter as defined in Definition 4, and σ_ε is the covariance kernel of the error term ε, as defined in Assumption 2.

7 Error distribution (required only for prediction bands): The error ε is a Student's t type of process with $\nu_0 > 2$ degrees of freedom defined as $\varepsilon \overset{D}{=} Z\sqrt{\frac{\nu_0}{\chi^2_{\nu_0}}}$, where $\chi^2_{\nu_0}$ is a real-valued Chi-squared distributed random variable with ν_0 degrees of freedom, $Z = \{Z(t),\, t \in [0,1]\}$ is a mean zero Gaussian process with continuous covariance kernel $\mathbb{E}[Z(t)Z(s)] = \sigma_Z(s,t)$, and where Z and $\chi^2_{\nu_0}$ are independent. The kernel then has the following structure (relying on the homoskedasticity assumption from Assumption 2):

$$\sigma_\varepsilon(s,t) = \mathbb{E}[\varepsilon(s)\varepsilon(t)|X] = \mathbb{E}[\varepsilon(s)\varepsilon(t)] = \mathbb{E}[Z(s)Z(t)]\mathbb{E}\left[\frac{\nu_0}{\chi^2_{\nu_0}}\right] = \sigma_Z(s,t)\frac{\nu_0}{\nu_0 - 2}.$$

A.1.1 Discussion

Assumption 1 is a standard linear model assumption with mean-independent error terms. Assumption 2 supposes homoskedastic error terms, which makes it easier to present and derive formulas for all objects that depend on the covariance kernel of the error term. Importantly, our results also hold in the heteroskedastic case, in which only the estimation of the error kernel has to be adjusted. Assumption 3 requires an iid random sampling scheme. However, our results also hold for weakly dependent stationary functional time series. Assumption 4 (a) and (b) impose smoothness conditions on the sample paths, which are required for the uniform convergence results. The typical applications considered in functional data analysis are concerned with relatively smooth functions and thus fit into our theoretical framework. Assumption 4 (c) imposes a smoothness condition on the population coefficient function, which ensures point-wise identification and uniform consistency of its estimator. Assumption 5 is the functional analog of the "no multicollinearity" condition: it ensures identification of $\beta(t)$ and its derivatives via the inverse $\Sigma_X^{-1}(s,t)$ near the diagonal, which is exactly the region entering the variance and roughness calculations. Assumption 6 rules out flat spots (degenerate local correlation) in the covariance kernels σ_ε and $\sigma_{x_{\mathsf{new}}}$. Intuitively, it ensures that correlation decays quadratically away from $s = t$, which is exactly what controls the excursion probabilities underlying simultaneous bands. The condition is mild and holds for common kernels (for instance, squared-exponential with length ℓ gives $\tau = 1/\ell$; Matérn with smoothness $\nu > 1$ yields $\tau > 0$); it excludes non-twice-differentiable kernels such as Brownian motion. Assumption 7 allows for heavy-tailed error distributions in the prediction band case via a common random scale, while preserving the correlation structure of Z. Our bands are therefore able to adapt to heavy-tails (low ν_0) as well as Gaussian tails ($\nu_0 \to \infty$). To ensure finite second moments we require $\nu_0 > 2$.

A.2 Lemmas

In Assumption 4, we defined $(G_1(t), \dots, G_K(t), G_{K+1}(t))^\top = (X_1(t), \dots, X_K(t), \varepsilon(t))^\top$. Recall the differential operator from Definition 5, and let $G_{jk}^{(d)}(t) = D^d G_j(t) G_k(t)$. Additionally, define its "de-meaned" version:

$$\dot{G}_{jk}^{(d)}(t) = D^d\left(G_j(t)G_k(t)\right) - \mathbb{E}\left[D^d\left(G_j(t)G_k(t)\right)\right], \tag{16}$$

for $j, k = 1, \dots, K+1$ and $d \in \{0, 1, 2\}$.

Lemma 6 (Stochastic Lipschitz Continuity) *Under Assumption 4 (a) and (b), we have for all $j, k = 1, \dots, K+1$, and for all $d \in \{0, 1\}$*

$$\max\left\{\left|G_k^{(d)}(t) - G_k^{(d)}(s)\right|, \left|\dot{G}_{jk}^{(d)}(t) - \dot{G}_{jk}^{(d)}(s)\right|\right\} \leq A_{jkd}\, \phi_{jkd}(|t-s|),$$

for all $s, t \in [0,1]$, where ϕ_{jkd} is a deterministic, non-decreasing continuous function on $[0,1]$ with $\phi_{jkd}(0) = 0$; and A_{jkd} is a real-valued random variable with $\mathbb{E}[A_{jkd}^2] < \infty$.

Proof Without loss of generality, let $0 \leq s < t \leq 1$. Under Assumption 4 (a) and (b), $\dot{G}_{jk}^{(d)}$ is continuous over $[0,1]$ for all $d \in \{0,1\}$ and thus, by the Mean Value Theorem, there exists a $\xi \in (s,t)$ such that

$$\dot{G}_{jk}^{(d)}(t) - \dot{G}_{jk}^{(d)}(s) = \dot{G}_{jk}^{(d+1)}(\xi)(t-s).$$

This implies that

$$\begin{aligned}\left(\dot{G}_{jk}^{(d)}(t) - \dot{G}_{jk}^{(d)}(s)\right)^2 &\leq \sup_{\xi \in (s,t)} \left(\dot{G}_{jk}^{(d+1)}(\xi)\right)^2 (t-s)^2 \\ &\leq \sup_{\xi \in [0,1]} \left(\dot{G}_{jk}^{(d+1)}(\xi)\right)^2 (t-s)^2.\end{aligned}$$

Taking the expectation and applying Assumption 4 (b) we have that

$$\begin{aligned}\mathbb{E}\left[\left(\dot{G}_{jk}^{(d)}(t) - \dot{G}_{jk}^{(d)}(s)\right)^2\right] &\leq \mathbb{E}\left[\sup_{\xi \in [0,1]} \left|\dot{G}_{jk}^{(d+1)}(\xi)\right|^2\right] (t-s)^2 \\ &= \mathbb{E}\left[\sup_{\xi \in [0,1]} \left|D^{d+1}G_j(\xi)G_k(\xi) - \mathbb{E}\left[D^{d+1}G_j(\xi)G_k(\xi)\right]\right|^2\right] (t-s)^2 \\ &\leq 4\mathbb{E}\left[\sup_{\xi \in [0,1]} \left|D^{d+1}G_j(\xi)G_k(\xi)\right|^2\right] (t-s)^2 \\ &\leq 4C\,(t-s)^2 =: f(|t-s|)\end{aligned} \tag{17}$$

for all $s, t \in [0,1]$, where $C < \infty$ is a non-zero, finite constant. Note that f is a deterministic non-negative function on $[0,1]$ which is non-decreasing in a neighborhood of 0. Moreover, it is easy to show that

$$\int_0^1 x^{-3/2} f^{1/2}(x) dx = 2\sqrt{4C} < \infty. \tag{18}$$

The result for $\left|\dot{G}_{jk}^{(d)}(t) - \dot{G}_{jk}^{(d)}(s)\right|$ now follows directly from Equations (17) and (18) by applying Theorem 2.3 in Hahn (1977) for the case of $r = 2$ moments. For the case of $\left|G_k^{(d)}(t) - G_k^{(d)}(s)\right|$, notice that we can apply the same proof technique such that

$$\begin{aligned}\mathbb{E}\left[\left(G_k^{(d)}(t) - G_k^{(d)}(s)\right)^2\right] &\leq \mathbb{E}\left[\sup_{\xi \in [0,1]} \left|D^{d+1} G_k(\xi)\right|^2\right] (t-s)^2 \\ &\leq C\,(t-s)^2 \leq f(|t-s|),\end{aligned}$$

where the last inequality follows by Assumption 4 (b). ■

For the next Lemma, let $\{\dot{G}_{jki}^{(d)}(t) : i = 1, \ldots, n\}$ be an iid sample of $\dot{G}_{jk}^{(d)}(t)$, as defined in (16). Define the corresponding sample mean as

$$B_{jkn}^{(d)}(t) = \frac{1}{n} \sum_{i=1}^{n} \dot{G}_{jki}^{(d)}(t) = \frac{1}{n} \sum_{i=1}^{n} D^d \left(G_{ji}(t) G_{ki}(t)\right) - \mathbb{E}\left[D^d \left(G_j(t) G_k(t)\right)\right].$$

Lemma 7 (Uniform consistency, 1-dimensional) *Under Assumptions 4 (a) and (b), it holds that, for all $j, k = 1, \ldots, K+1$, and for all $d \in \{0, 1\}$,*

$$\sup_{t \in [0,1]} \left|B_{jkn}^{(d)}(t)\right| \overset{a.s.}{\to} 0, \quad as \quad n \to \infty.$$

Proof We first consider pointwise convergence for each $t \in [0,1]$ and then expand this to uniform convergence. As the $\dot{G}_{jki}^{(d)}(t)$ are iid, by Kolmogorov's strong law of large numbers (SLLN),

$$B_{jkn}^{(d)}(t) \overset{a.s.}{\to} 0, \quad \text{as} \quad n \to \infty, \tag{19}$$

pointwise for each $t \in [0,1]$, each $j, k = 1, \ldots, K+1$, and each $d \in \{0, 1\}$. To expand this to uniform convergence, we need to show that $B_{jkn}^{(d)}(t)$ is strongly stochastically equicontinuous. By Lemma 6, we have that for all $i = 1, \ldots, n$

$$|\dot{G}_{jki}^{(d)}(t) - \dot{G}_{jki}^{(d)}(s)| \leq A_{jkdi} \phi_{jkd}(|t-s|) \quad \text{for all} \quad s, t \in [0,1], \tag{20}$$

where the $\{A_{jkdi} : i = 1, \ldots, n\}$ are iid as A_{jkd}, with $\mathbb{E}[A_{jkd}^2] < \infty$. This allows us to derive the following approximation for all $s, t \in [0, 1]$:

$$\begin{aligned}\left|B_{jkn}^{(d)}(t) - B_{jkn}^{(d)}(s)\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left(\dot{G}_{jki}^{(d)}(t) - \dot{G}_{jki}^{(d)}(s)\right)\right| \\ &\le \frac{1}{n}\sum_{i=1}^{n}\left|\dot{G}_{jki}^{(d)}(t) - \dot{G}_{jki}^{(d)}(s)\right| \\ &\le \frac{1}{n}\sum_{i=1}^{n} A_{jkdi}\phi_{jkd}(|t-s|), \end{aligned} \tag{21}$$

where the last inequality follows by (20). By the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} A_{jkdi} \overset{a.s.}{\to} \mathbb{E}\left(A_{jkd}\right) < \infty \quad \text{as} \quad n \to \infty. \tag{22}$$

The results of (21) and (22) imply, by Theorem 22.10 in Davidson (2021), that $B_{jkn}^{(d)}$ is strongly stochastic equicontinuous for every $j, k = 1, \ldots, K+1$, and for all $d \in \{0, 1\}$. This, together with the pointwise consistency in (19), implies, by Theorem 22.8 in Davidson (2021), that

$$\sup_{t\in[0,1]}\left|B_{jkn}^{(d)}(t)\right| \overset{a.s.}{\to} 0, \quad \text{as} \quad n \to \infty$$

for each $j, k = 1, \ldots, K+1$, and for all $d \in \{0, 1\}$. ■

For the next lemma, we need to extend our notation further. Define

$$G_{jk}^{(d,p)}(s,t) = D^{(d,p)}\left(G_j(s)G_k(t)\right),$$

for $j, k = 1, \ldots, K+1$, and for $d, p \in \{0, 1\}$, employing the differential operator $D^{(d,p)}$ from Definition 5. This leads to the following four cases:

$$G_{jk}^{(d,p)}(s,t) = \begin{bmatrix} G_j(s)G_k(t) & G_j(s)\frac{\partial}{\partial t}G_k(t) \\ \frac{\partial}{\partial s}G_j(s)G_k(t) & \frac{\partial}{\partial s}G_j(s)\frac{\partial}{\partial t}G_k(t) \end{bmatrix}_{(d,p)}.$$

As before, we define the "de-meaned" version of $G_{jk}^{(d,p)}$ as

$$\begin{aligned}\dot{G}_{jk}^{(d,p)}(s,t) &= D^{(d,p)}\left(G_j(s)G_k(t)\right) - \mathbb{E}\left[D^{(d,p)}G_j(s)G_k(t)\right] \\ &= G_{jk}^{(d,p)}(s,t) - \mathbb{E}\left[G_{jk}^{(d,p)}(s,t)\right].\end{aligned}$$

Let $\dot{G}_{jki}^{(d,p)}(s,t)$, for $i = 1, \ldots, n$, denote iid copies of $\dot{G}_{jk}^{(d,p)}(s,t)$ and define the corresponding sample mean as

$$B_{jkn}^{(d,p)}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\dot{G}_{jki}^{(d,p)}(s,t) = \frac{1}{n}\sum_{i=1}^{n} D^{(d,p)}\left(G_{ij}(s)G_{ik}(t)\right) - \mathbb{E}\left[D^{(d,p)}\left(G_j(s)G_k(t)\right)\right].$$

Lemma 8 (Uniform consistency, 2-dimensional) *Under Assumptions 4 (a) and (b), it holds that, for all* $j,k=1,\dots,K+1$ *and for all* $d,p\in\{0,1\}$,

$$\sup_{s,t\in[0,1]}\left|B^{(d,p)}_{jkn}(s,t)\right|\overset{a.s.}{\to}0,\quad as\quad n\to\infty.$$

Proof Note that, by the SLLN,

$$B^{(d,p)}_{jkn}(s,t)\overset{a.s.}{\to}\mathbb{E}\left[\dot{G}^{(d,p)}_{jk}(s,t)\right]=0,\quad\text{as}\quad n\to\infty,\tag{23}$$

pointwise for each $s,t\in[0,1]$, for all $j,k=1,\dots,K+1$, and for all $d,p\in\{0,1\}$. Next, we need to show that $B^{(d,p)}_{jkn}(s,t)$ is strongly stochastically equicontinuous for every $j,k=1,\dots,K+1$ and for all $d,p\in\{0,1\}$. For this, let $s,t,u,v\in[0,1]$, such that (without loss of generality) $s<u$ and $t<v$. Then,

$$\begin{aligned}
&\left|B^{(d,p)}_{jkn}(s,t)-B^{(d,p)}_{jkn}(u,v)\right|=\left|\frac{1}{n}\sum_{i=1}^{n}\left(\dot{G}^{(d,p)}_{ijk}(s,t)-\dot{G}^{(d,p)}_{ijk}(u,v)\right)\right|\\
&\le\frac{1}{n}\sum_{i=1}^{n}\left|\dot{G}^{(d,p)}_{ijk}(s,t)-\dot{G}^{(d,p)}_{ijk}(u,v)\right|\\
&=\frac{1}{n}\sum_{i=1}^{n}\left|G^{(d,p)}_{ijk}(s,t)-\mathbb{E}\left[G^{(d,p)}_{jk}(s,t)\right]-\left(G^{(d,p)}_{ijk}(u,v)-\mathbb{E}\left[G^{(d,p)}_{jk}(u,v)\right]\right)\right|\\
&\le\frac{1}{n}\sum_{i=1}^{n}\left|G^{(d,p)}_{ijk}(s,t)-G^{(d,p)}_{ijk}(u,v)\right|+\frac{1}{n}\sum_{i=1}^{n}\left|\mathbb{E}\left[G^{(d,p)}_{jk}(s,t)\right]-\mathbb{E}\left[G^{(d,p)}_{jk}(u,v)\right]\right|\\
&\le\frac{1}{n}\sum_{i=1}^{n}\left|G^{(d,p)}_{ijk}(s,t)-G^{(d,p)}_{ijk}(u,v)\right|+\mathbb{E}\left|G^{(d,p)}_{jk}(s,t)-G^{(d,p)}_{jk}(u,v)\right|=:(*)
\end{aligned}$$

Now, notice that

$$G^{(d,p)}_{jk}(s,t)-G^{(d,p)}_{jk}(u,v)=(G^{(d)}_{j}(s)-G^{(d)}_{j}(u))G^{(p)}_{k}(t)+(G^{(p)}_{k}(t)-G^{(p)}_{k}(v))G^{(d)}_{j}(u),$$

and thus,

$$\begin{aligned}
(*)\le\;&\\
&\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left|G^{(d)}_{ji}(s)-G^{(d)}_{ji}(u)\right|\left|G^{(p)}_{ki}(t)\right|}_{I_{1,1}}+\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left|G^{(p)}_{ki}(t)-G^{(p)}_{ki}(v)\right|\left|G^{(d)}_{ji}(u)\right|}_{I_{1,2}}\\
&+\underbrace{\mathbb{E}\left(\left|G^{(d)}_{j}(s)-G^{(d)}_{j}(u)\right|\left|G^{(p)}_{k}(t)\right|\right)}_{I_{2,1}}+\underbrace{\mathbb{E}\left(\left|G^{(p)}_{k}(t)-G^{(p)}_{k}(v)\right|\left|G^{(d)}_{j}(u)\right|\right)}_{I_{2,2}}.
\end{aligned}$$

Let us focus on $I_{1,1}$ first. By the Cauchy-Schwarz inequality, and Lemma 6, we have

$$\begin{aligned}
I_{1,1} &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left|G_{ji}^{(d)}(s)-G_{ji}^{(d)}(u)\right|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left|G_{ki}^{(p)}(t)\right|^2\right)^{1/2} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left|G_{ji}^{(d)}(s)-G_{ji}^{(d)}(u)\right|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,1]}\left|G_{ki}^{(p)}(t)\right|^2\right)^{1/2} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n}A_{jdi}^2\phi_{jd}^2(|s-u|)\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,1]}\left|G_{ki}^{(p)}(t)\right|^2\right)^{1/2} \\
&= \phi_{jd}(|s-u|)\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}A_{jdi}^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,1]}\left|G_{ki}^{(p)}(t)\right|^2\right)^{1/2}}_{=:F_n^{(d,p)}}.
\end{aligned}$$

Notice that $F_n^{(d,p)}$ does not depend on the choice of s, u, t, v. Hence, doing the same for $I_{1,2}$, and using the $\ell_1 - \ell_2$ norm inequality, and that ϕ_{jd} is non-decreasing, we get

$$\begin{aligned}
I_{1,1} + I_{1,2} &\leq F_n^{(d,p)}\left(\phi_{jd}(|s-u|)+\phi_{kd}(|t-v|)\right) \\
&\leq F_n^{(d,p)}2\phi_{jd}(|s-u|+|t-v|) \\
&\leq F_n^{(d,p)}2\phi_{jd}(2\sqrt{|s-u|^2+|t-v|^2}).
\end{aligned}$$

Next, we focus on $I_{2,1}$ and $I_{2,2}$. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\mathbb{E}\left(\left|G_j^{(d)}(s)-G_j^{(d)}(u)\right|\left|G_k^{(p)}(t)\right|\right) \\
&\leq \mathbb{E}\left(\left|G_j^{(d)}(s)-G_j^{(d)}(u)\right|^2\right)^{1/2}\mathbb{E}\left(\left|G_k^{(p)}(t)\right|^2\right)^{1/2} \\
&\leq \mathbb{E}\left(\left|G_j^{(d)}(s)-G_j^{(d)}(u)\right|^2\right)^{1/2}\mathbb{E}\left(\left|\sup_{t\in[0,1]}G_k^{(p)}(t)\right|^2\right)^{1/2} \\
&\leq \mathbb{E}\left(A_{jd}^2\phi_{jd}(|s-u|)^2\right)^{1/2}\mathbb{E}\left(\left|\sup_{t\in[0,1]}G_k^{(p)}(t)\right|^2\right)^{1/2} \\
&= \phi_{jd}(|s-u|)\underbrace{\mathbb{E}\left(A_{jd}^2\right)^{1/2}\mathbb{E}\left(\left|\sup_{t\in[0,1]}G_k^{(p)}(t)\right|^2\right)^{1/2}}_{=:F^{(d,p)}}.
\end{aligned}$$

Using the same arguments as above, we see

$$I_{2,1} + I_{2,2} \leq F^{(d,p)}2\phi_{jd}(2\sqrt{|s-u|^2+|t-v|^2}).$$

Now let $h(x) = 2\phi_{jd}(2x)$, and notice that $h(x) \downarrow 0$ as $x \downarrow 0$, since ϕ_{jd} is continuous and $\phi_{jd}(0) = 0$. Thus, we have

$$I_{1,1} + I_{1,2} + I_{2,1} + I_{2,2} \leq (F_n^{(d,p)} + F^{(d,p)})h(\sqrt{|s-u|^2 + |t-v|^2}). \tag{24}$$

By the SLLN and the continuous mapping theorem, we know that

$$F_n^{(d,p)} + F^{(d,p)} \overset{a.s.}{\to} 2F^{(d,p)} = 2\mathbb{E}\left(A_{jd}^2\right)^{1/2} \mathbb{E}\left(\left|\sup_{t\in[0,1]} G_k^{(p)}(t)\right|^2\right)^{1/2} < \infty, \tag{25}$$

where finiteness holds by Lemma 6 and Assumption 4 (b). Results (24) and (25) imply, by Theorem 22.10 in Davidson (2021), that $B_{jkn}^{(d,p)}$ is strongly stochastic equicontinuous for every $j, k = 1, \ldots, K+1$, and for all $d, p \in \{0, 1\}$. This, together with the pointwise consistency in (23), implies, by Theorem 22.8 in Davidson (2021), that

$$\sup_{s,t\in[0,1]} \left|B_{jkn}^{(d,p)}(s,t)\right| \overset{a.s.}{\to} 0, \quad \text{as} \quad n \to \infty$$

for each $j, k = 1, \ldots, K+1$, and for all $d, p \in \{0, 1\}$. ■

Lemma 9 (Roughness Functions) *Under Assumptions 2, 6, and 7, we find that*

(a) $\tau_{\mathsf{SCB}} = \tau_c$*, with* $c(s,t) = \sigma_\varepsilon(s,t)\sigma_{x_{\mathsf{new}}}(s,t)$*, and*

(b) $\tau_{\mathsf{SPB}} = \tau_c$*, with* $c(s,t) = \sigma_\varepsilon(s,t)$*.*

Proof (a) τ_{SCB} is defined as the roughness function corresponding to the covariance kernel σ_{CB}. Under homoskedasticity (Assumption 2), we have that $\sigma_{\mathsf{CB}}(s,t) = \sigma_\varepsilon(s,t)\sigma_{x_{\mathsf{new}}}(s,t)$.

(b) τ_{SPB} is defined as the roughness function corresponding to the covariance kernel σ_{PB}. Under homoskedasticity (Assumption 2) and the assumption on the error process in the prediction case (Assumption 7), we have that $\sigma_{\mathsf{PB}}(s,t) = \sigma_\varepsilon(s,t)(\nu-2)/\nu$. However, since the correlation function is invariant to constant scalings, we can ignore the factor $(\nu-2)/\nu$. ■

A.3 Theorems

We first recall the notation of the paper. The sample counterpart of Σ_X is defined as

$$\hat{\Sigma}_X(s,t) = \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i^{\mathsf{T}}(t),$$

and the sample covariance of the error term is defined as

$$\hat{\sigma}_\varepsilon(s,t) = \frac{1}{n}\sum_{i=1}^{n} e_i(s)e_i(t),$$

where $e_i(t) = Y_i(t) - X_i^{\mathsf{T}}(t)\hat{\beta}(t)$ denotes the OLS-residual. The sample counterpart of $\sigma_{x_{\mathsf{new}}}$ is defined as

$$\hat{\sigma}_{x_{\mathsf{new}}}(s,t) = x_{\mathsf{new}}^{\mathsf{T}}(s)\hat{\Sigma}_X(s,t)^{-1}x_{\mathsf{new}}(t).$$

In subsequent theorems, we will follow Definition 5 and denote the first derivative of a function using the superscript d, with $d = 0$ indicating the function itself and $d = 1$ indicating the first derivative. For the partial derivative, we will use the superscript (d,p), with $(d,p) = (0,0)$ indicating the function itself, $(d,p) = (1,0)$ indicating the partial derivative in the first argument, $(d,p) = (0,1)$ indicating the partial derivative in the second argument, and $(d,p) = (1,1)$ indicating the partial derivative in both the first and second argument, respectively.

Theorem 10 (Uniform Consistency) *Under Assumptions 1-6, we have that as $n \to \infty$, for $d, p \in \{0,1\}$:*

(a) $\sup_{t\in[0,1]} \|\hat{\Sigma}_X^{(d)}(t,t) - \Sigma_X^{(d)}(t,t)\| \overset{a.s.}{\to} 0,$

(b) $\sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{(d,p)}(s,t) - \Sigma_X^{(d,p)}(s,t)\| \overset{a.s.}{\to} 0,$

(c) $\sup_{t\in[0,1]} \|\hat{\Sigma}_X^{-1^{(d)}}(t,t) - \Sigma_X^{-1^{(d)}}(t,t)\| \overset{a.s.}{\to} 0,$

(d) $\sup_{s,t\in[0,1],|s-t|<\delta} \|\hat{\Sigma}_X^{-1^{(d,p)}}(s,t) - \Sigma_X^{-1^{(d,p)}}(s,t)\| \overset{a.s.}{\to} 0,$

(e) $\sup_{t\in[0,1]} \|\hat{\beta}^{(d)}(t) - \beta^{(d)}(t)\| \overset{a.s.}{\to} 0,$

(f) $\sup_{t\in[0,1]} |\hat{\sigma}_{\varepsilon}^{(d,p)}(s,t) - \sigma_{\varepsilon}^{(d,p)}(s,t)| \overset{a.s.}{\to} 0,$

(g) $\sup_{t\in[0,1]} |\hat{\sigma}_{x_{\mathsf{new}}}^{(d,p)}(s,t) - \sigma_{x_{\mathsf{new}}}^{(d,p)}(s,t)| \overset{a.s.}{\to} 0.$

Proof (a) Recall, that component-wise univariate almost sure convergence implies almost sure convergence of the vectors and matrices built from these components. Pick any $d \in \{0,1\}$. Using this, together with the continuous mapping theorem for function spaces (see van der Vaart, 1998, p. 259, Theorem 18.11), the desired result follows directly from Lemma 7:

$$\begin{aligned}
&\sup_{t\in[0,1]} \left|\hat{\Sigma}_X^{(d)}(t,t) - \Sigma_X^{(d)}(t,t)\right| \\
&\qquad = \sup_{t\in[0,1]} \left|\frac{1}{n}\sum_{i=1}^{n} D^d X_i(t)X_i^{\mathsf{T}}(t) - \mathbb{E}\left[D^d X(t)X^{\mathsf{T}}(t)\right]\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}.
\end{aligned}$$

(b) Pick any $d, p \in \{0,1\}$. Using the same arguments as above, the desired result follows directly from Lemma 8:

$$\begin{aligned}
&\sup_{s,t\in[0,1]} \left|\hat{\Sigma}_X^{(d,p)}(s,t) - \Sigma_X^{(d,p)}(s,t)\right| \\
&\qquad = \sup_{s,t\in[0,1]} \left|\frac{1}{n}\sum_{i=1}^{n} X_i^{(d)}(s)X_i^{(p)}(t)^{\mathsf{T}} - \mathbb{E}\left[X^{(d)}(s)X^{(p)}(t)^{\mathsf{T}}\right]\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}.
\end{aligned}$$

(c) For the case $d = 0$, the claim follows directly from the continuous mapping theorem for function spaces (see van der Vaart, 1998, p. 259, Theorem 18.11), Theorem 10 (a), and the rank condition in Assumption 5:

$$\sup_{t\in[0,1]} \left|\hat{\Sigma}_X^{-1}(t,t) - \Sigma_X^{-1}(t,t)\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}. \tag{26}$$

For the first derivative ($d = 1$), matrix calculus yields

$$D^{(1)}\,\hat{\Sigma}_X^{-1}(t,t) = -\hat{\Sigma}_X^{-1}(t,t)\left(D^{(1)}\,\hat{\Sigma}_X(t,t)\right)\hat{\Sigma}_X^{-1}(t,t).$$

Thus, by the convergence results of Theorem 10 (a) and Equation (26), and an application of the continuous mapping theorem for function spaces, we have

$$\sup_{t\in[0,1]} \left|\hat{\Sigma}_X^{-1}(t,t)\left(D^{(1)}\,\hat{\Sigma}_X(t,t)\right)\hat{\Sigma}_X^{-1}(t,t) - \Sigma_X^{-1}(t,t)\left(D^{(1)}\,\Sigma_X(t,t)\right)\Sigma_X^{-1}(t,t)\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}.$$

And since $D^{(1)}\,\Sigma_X^{-1}(t,t) = -\Sigma_X^{-1}(t,t)\left(D^{(1)}\,\Sigma_X(t,t)\right)\Sigma_X^{-1}(t,t)$, we get the desired result:

$$\sup_{t\in[0,1]} \left|D^{(1)}\,\hat{\Sigma}_X^{-1}(t,t) - D^{(1)}\,\Sigma_X^{-1}(t,t)\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}.$$

(d) For the case $(d,p) = (0,0)$, the claim follows directly from the continuous mapping theorem for function spaces (see van der Vaart, 1998, p. 259, Theorem 18.11), Theorem 10 (b), and the rank condition in Assumption 5:

$$\sup_{\substack{s,t\in[0,1]\\|s-t|<\delta}} \left|\hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t)\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}. \tag{27}$$

Consider now the case $(d,p) = (0,1)$. Matrix calculus yields

$$D^{(0,1)}\,\hat{\Sigma}_X^{-1}(s,t) = -\hat{\Sigma}_X^{-1}(s,t)\left(D^{(0,1)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t).$$

Thus, by the convergence results of Theorem 10 (b) and Equation (27), and an application of the continuous mapping theorem for function spaces, we have

$$\sup_{\substack{s,t\in[0,1]\\|s-t|<\delta}} \left|\hat{\Sigma}_X^{-1}(s,t)\left(D^{(0,1)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t)\left(D^{(0,1)}\,\Sigma_X(s,t)\right)\Sigma_X^{-1}(s,t)\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}.$$

And since $D^{(0,1)}\,\Sigma_X^{-1}(s,t) = -\Sigma_X^{-1}(s,t)\left(D^{(0,1)}\,\Sigma_X(s,t)\right)\Sigma_X^{-1}(s,t)$, we get the desired result:

$$\sup_{\substack{s,t\in[0,1]\\|s-t|<\delta}} \left|D^{(0,1)}\,\hat{\Sigma}_X^{-1}(s,t) - D^{(0,1)}\,\Sigma_X^{-1}(s,t)\right| \overset{a.s.}{\to} \underset{(K\times K)}{0}.$$

Now, for the last case $(d,p)=(1,1)$, the product rule for matrix derivatives yields

$$\begin{aligned}
&D^{(1,1)}\,\hat{\Sigma}_X^{-1}(s,t)\\
&\quad= D^{(1,0)}\,D^{(0,1)}\,\hat{\Sigma}_X^{-1}(s,t)\\
&\quad= D^{(1,0)}\left(-\hat{\Sigma}_X^{-1}(s,t)\left(D^{(0,1)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t)\right)\\
&\quad= \left(\hat{\Sigma}_X^{-1}(s,t)\left(D^{(1,0)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t)\left(D^{(0,1)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t)\right.\\
&\qquad+\hat{\Sigma}_X^{-1}(s,t)\left(D^{(1,1)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t)\\
&\qquad\left.+\hat{\Sigma}_X^{-1}(s,t)\left(D^{(0,1)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t)\left(D^{(1,0)}\,\hat{\Sigma}_X(s,t)\right)\hat{\Sigma}_X^{-1}(s,t)\right).
\end{aligned}$$

Hence, employing the same arguments as above, using the convergence results of Theorem 10 (b), as well as Equation (27), and the continuous mapping theorem for function spaces, we get the desired result:

$$\sup_{\substack{s,t\in[0,1]\\|s-t|<\delta}}\left|D^{(1,1)}\,\hat{\Sigma}_X^{-1}(s,t)-D^{(1,1)}\,\Sigma_X^{-1}(s,t)\right|\overset{a.s.}{\to}\underset{(K\times K)}{0}.$$

(e) Under the linearity assumption (Assumption 1), we can express the estimation error of the OLS estimator as

$$\begin{aligned}
\left|\hat{\beta}(t)-\beta(t)\right| &= \left|\left(\frac{1}{n}\sum_{i=1}^{n}X_i(t)X_i^{\mathsf{T}}(t)\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}X_i(t)\varepsilon_i(t)\right|\\
&= \left|\hat{\Sigma}_X^{-1}(t,t)\frac{1}{n}\sum_{i=1}^{n}X_i(t)\varepsilon_i(t)\right|,
\end{aligned}$$

and of its first derivative (using the product rule) as

$$\begin{aligned}
&\left|\hat{\beta}^{(1)}(t)-\beta^{(1)}(t)\right|\\
&=\left|D^{(1)}\big(\hat{\Sigma}_X^{-1}(t,t)\big)\frac{1}{n}\sum_{i=1}^{n}D^{(0)}\big(X_i(t)\varepsilon_i(t)\big)+D^{(0)}\big(\hat{\Sigma}_X^{-1}(t,t)\big)\frac{1}{n}\sum_{i=1}^{n}D^{(1)}\big(X_i(t)\varepsilon_i(t)\big)\right|, \quad (28)
\end{aligned}$$

where $\left|\hat{\beta}(t)-\beta(t)\right|$ and $\left|\hat{\beta}^{(1)}(t)-\beta^{(1)}(t)\right|$ denote the absolute values of the K-dimensional vectors of estimation errors. Recall that derivatives of vector- or matrix-valued functions are simply computed separately for each vector- or matrix-component. By Lemma 8, we have that for $d\in\{0,1\}$

$$\sup_{s,t\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}D^{(d)}\big(X_i(s)\varepsilon_i(t)\big)-\mathbb{E}\left[D^{(d)}\big(X(s)\varepsilon(t)\big)\right]\right|\overset{a.s.}{\to}\underset{(K\times 1)}{0}. \quad (29)$$

And under the assumptions of Theorem 10 we find, for $d\in\{0,1\}$,

$$\mathbb{E}\left[D^{(d)}\big(X(s)\varepsilon(t)\big)\right]=D^{(d)}\left(\mathbb{E}\big(X(s)\varepsilon(t)\big)\right)=\underset{(K\times 1)}{0}. \quad (30)$$

The case of $d = 0$ follows directly by conditional independence. For $d = 1$, we can interchange differentiation and expectation by the dominated convergence theorem, and the fact that $|X(t)\varepsilon(t)|$ can be bounded by $Z = \sup_{t\in[0,1]} |X(t)\varepsilon(t)|^2 + 1$ almost surely, where $\mathbb{E}[Z] < \infty$ by Assumption 4 (b). Thus, by Equation (29) and (30), the result from Theorem 10 (c) and the uniform continuous mapping theorem, we find

$$\begin{aligned}\sup_{t\in[0,1]} \left|\hat{\beta}(t) - \beta(t)\right| &= \sup_{t\in[0,1]} \left|\hat{\Sigma}_X^{-1}(t,t)\frac{1}{n}\sum_{i=1}^{n} X_i(t)\varepsilon_i(t)\right| \\ &\overset{a.s.}{\to} \sup_{t\in[0,1]} \left|\Sigma_X^{-1}(t,t)\mathbb{E}\left(X(t)\varepsilon(t)\right)\right| = \underset{(K\times 1)}{0}.\end{aligned}$$

Following Equation (28), to show the claim for $\hat{\beta}^{(1)}$ we require uniform convergence results for the terms $n^{-1}\sum_{i=1}^{n} D^{(1)}\left(X_i(t)\varepsilon_i(t)\right)$ and $D^{(1)}\hat{\Sigma}_X^{-1}(t,t)$. For the first term, the claim holds directly by Lemma 7, and for the second term by Theorem 10 (c). Thus, by the uniform continuous mapping theorem, we then have

$$\begin{aligned}&\sup_{t\in[0,1]} |\hat{\beta}^{(1)}(t) - \beta^{(1)}(t)| \\ &\quad\leq \sup_{t\in[0,1]} |D^{(1)}\big(\hat{\Sigma}_X^{-1}(t,t)\big)| \sup_{t\in[0,1]} |\frac{1}{n}\sum_{i=1}^{n} D^{(0)}(X_i(t)\varepsilon_i(t))| \\ &\qquad + \sup_{t\in[0,1]} |D^{(0)}\big(\hat{\Sigma}_X^{-1}(t,t)\big)| \sup_{t\in[0,1]} |\frac{1}{n}\sum_{i=1}^{n} D^{(1)}(X_i(t)\varepsilon_i(t))| \\ &\quad\overset{a.s}{\to} \sup_{t\in[0,1]} |D^1(\mathbb{E}[X(t)X^{\mathsf{T}}(t)]^{-1})| \underbrace{\sup_{t\in[0,1]} |\mathbb{E}[X(t)\varepsilon(t)]|}_{=0} \\ &\qquad + \sup_{t\in[0,1]} |(\mathbb{E}[X(t)X^{\mathsf{T}}(t))]^{-1}| \underbrace{\sup_{t\in[0,1]} |\mathbb{E}[D^1(X(t)\varepsilon(t))]|}_{=\sup_{t\in[0,1]} |D^1\mathbb{E}[X(t)\varepsilon(t)]|=0} = \underset{(K\times 1)}{0}.\end{aligned}$$

(f) We will first show the case $(d,p) = (0,0)$. Recall that

$$\begin{aligned}e_i(t) &= Y_i(t) - X_i^{\mathsf{T}}(t)\hat{\beta}(t) \\ &= X_i^{\mathsf{T}}(t)\beta(t) + \varepsilon_i(t) - X_i^{\mathsf{T}}(t)\hat{\beta}(t) \\ &= \varepsilon_i(t) - X_i^{\mathsf{T}}(t)\left(\hat{\beta}(t) - \beta(t)\right).\end{aligned}$$

And thus, we can decompose the covariance estimator as follows:

$$
\begin{aligned}
\hat{\sigma}_\varepsilon(s,t) &= \frac{1}{n}\sum_{i=1}^{n} e_i(s)e_i(t) \\
&= \frac{1}{n}\sum_{i=1}^{n} \left(\varepsilon_i(s) - X_i^\intercal(s)\left(\hat{\beta}(s)-\beta(s)\right)\right)\left(\varepsilon_i(t) - X_i^\intercal(t)\left(\hat{\beta}(t)-\beta(t)\right)\right) \\
&= \Bigg(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(s)\varepsilon_i(t) \\
&\qquad - \left(\hat{\beta}(s)-\beta(s)\right)^\intercal \; \frac{1}{n}\sum_{i=1}^{n} X_i(s)\varepsilon_i(t) \\
&\qquad - \left(\hat{\beta}(t)-\beta(t)\right)^\intercal \; \frac{1}{n}\sum_{i=1}^{n} X_i(t)\varepsilon_i(s) \\
&\qquad + \left(\hat{\beta}(s)-\beta(s)\right)^\intercal \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i^\intercal(t)\left(\hat{\beta}(t)-\beta(t)\right)\Bigg) \\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(s)\varepsilon_i(t) + U_n(s,t) \\
&= \sigma_\varepsilon(s,t) + \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i(s)\varepsilon_i(t) - \mathbb{E}[\varepsilon(s)\varepsilon(t)]\right) + U_n(s,t),
\end{aligned}
$$

which implies that

$$
\sup_{s,t\in[0,1]} |\hat{\sigma}_\varepsilon(s,t) - \sigma_\varepsilon(s,t)| \le \sup_{s,t\in[0,1]} \left|\frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i(s)\varepsilon_i(t) - \mathbb{E}[\varepsilon(s)\varepsilon(t)]\right)\right| + \sup_{s,t\in[0,1]} |U_n(s,t)|\,.
$$

Now, the first term on the right-hand side converges to zero almost surely by Lemma 8, while the second term converges to zero almost surely by an application of the Cauchy-Schwarz inequality, Theorem 10 (b) and (e), Equation (29), and the uniform continuous mapping theorem. And hence,

$$
\sup_{s,t\in[0,1]} |\hat{\sigma}_\varepsilon(s,t) - \sigma_\varepsilon(s,t)| \overset{a.s.}{\to} 0.
$$

For the case $(d,p) = (0,1)$, remember from above that

$$
\hat{\sigma}_\varepsilon(s,t) - \sigma_\varepsilon(s,t) = \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i(s)\varepsilon_i(t) - \mathbb{E}[\varepsilon(s)\varepsilon(t)]\right) + U_n(s,t),
$$

and so,

$$
\begin{aligned}
&\sup_{s,t\in[0,1]} |\hat{\sigma}_\varepsilon^{(0,1)}(s,t) - \sigma_\varepsilon^{(0,1)}(s,t)| \\
&\qquad\qquad \le \sup_{s,t\in[0,1]} \left|\frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i(s)\varepsilon_i^{(1)}(t) - \mathbb{E}[\varepsilon(s)\varepsilon^{(1)}(t)]\right)\right| + \sup_{s,t\in[0,1]} |U_n^{(0,1)}(s,t)|.
\end{aligned}
$$

The first term on the right-hand side converges to zero almost surely by Lemma 8. For the second term, notice that $U_n^{(0,1)}(s,t)$ is a continuous composition of $\hat{\beta}^{(d)}(t) - \beta^{(d)}(t)$, and averages of $X_i^{(d)}(s)\varepsilon^{(p)}(t)$ and $X_i^{(d)}(s)X_i^{(p)}(t)^{\mathsf{T}}$. For the OLS estimator and its derivative, almost sure uniform convergence has been shown by Theorem 10 (e), and for the averages, it has been shown in Lemma 8. Thus, by the uniform continuous mapping theorem

$$\sup_{s,t\in[0,1]} |U_n^{(0,1)}(s,t)| \overset{a.s.}{\to} 0,$$

which proves the result. The case $(d,p) = (1,0)$ follows by symmetry. And for the case $(d,p) = (1,1)$, notice that

$$\begin{aligned}\sup_{s,t\in[0,1]} &|\hat{\sigma}_\varepsilon^{(1,1)}(s,t) - \sigma_\varepsilon^{(1,1)}(s,t)| \\ &\leq \sup_{s,t\in[0,1]} \left| \frac{1}{n} \sum_{i=1}^n \left(\varepsilon_i^{(1)}(s)\varepsilon_i^{(1)}(t) - \mathbb{E}[\varepsilon^{(1)}(s)\varepsilon^{(1)}(t)] \right) \right| + \sup_{s,t\in[0,1]} |U_n^{(1,1)}(s,t)|.\end{aligned}$$

Employing the same arguments as for the case $(d,p) = (0,1)$, the result follows by Theorem 10 (e), Lemma 8 and an application of the uniform continuous mapping theorem.

(g) Let $x_{\mathsf{new}} : [0,1] \to \mathbb{R}^K \in C^1([0,1], \mathbb{R}^K)$ be a deterministic function with $x_{\mathsf{new}}(t) \neq 0$ for all $t \in [0,1]$. Since x_{new} is deterministic and each path $x_{\mathsf{new}}(t)_j$ as well as its derivative, are continuous, there exists a finite upper bound $0 < M < \infty$, such that $\max\{|x_{\mathsf{new}}(t)_j|, |x_{\mathsf{new}}^{(1)}(t)_j|\} \leq M$ for all $j \in \{1,\ldots,K\}$ and $t \in [0,1]$. We further recall, that $\sigma_{x_{\mathsf{new}}}(s,t) = x_{\mathsf{new}}^{\mathsf{T}}(s)\Sigma_X^{-1}(s,t)x_{\mathsf{new}}(t)$ and $\hat{\sigma}_{x_{\mathsf{new}}}(s,t) = x_{\mathsf{new}}^{\mathsf{T}}(s)\hat{\Sigma}_X^{-1}(s,t)x_{\mathsf{new}}(t)$. In Theorem 10 (d) it has been shown that, for $d,p \in \{0,1\}$,

$$\sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1^{(d,p)}}(s,t) - \Sigma_X^{-1^{(d,p)}}(s,t)\| \overset{a.s.}{\to} 0.$$

Now, notice that we can write

$$\begin{aligned}\hat{\sigma}_{x_{\mathsf{new}}}(s,t) - \sigma_{x_{\mathsf{new}}}(s,t) &= x_{\mathsf{new}}^{\mathsf{T}}(s) \left(\hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t) \right) x_{\mathsf{new}}(t) \\ &= \sum_{l=1}^K \sum_{m=1}^K \underbrace{x_{\mathsf{new}}(s)_l \left\{ \hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t) \right\}_{lm} x_{\mathsf{new}}(t)_m}_{=:c_{lm}(s,t)}.\end{aligned}$$

And so,

$$\sup_{s,t\in[0,1]} |\hat{\sigma}_{x_{\mathsf{new}}}^{(d,p)}(s,t) - \sigma_{x_{\mathsf{new}}}^{(d,p)}(s,t)| \leq \sum_{l=1}^K \sum_{m=1}^K \sup_{s,t\in[0,1]} |c_{lm}^{(d,p)}(s,t)|$$

Pick any $l,m \in \{1,\ldots,K\}$ and focus on the case $d = p = 0$ first. We find

$$\begin{aligned}\sup_{s,t\in[0,1]} |c_{lm}^{(0,0)}(s,t)| &= \sup_{s,t\in[0,1]} |x_{\mathsf{new}}(s)_l \left\{ \hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t) \right\}_{lm} x_{\mathsf{new}}(t)_m| \\ &\leq M^2 \sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t)\| \\ &\overset{a.s.}{\to} 0.\end{aligned}$$

For the case $d = 1$ and $p = 0$, applying the product rule, we have

$$
\begin{aligned}
&c_{lm}^{(1,0)}(s,t) \\
&\quad = x_{\mathsf{new}}^{(1)}(s)_l \left\{ \hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t) \right\}_{lm} x_{\mathsf{new}}(t)_m \\
&\qquad + x_{\mathsf{new}}(s)_l \left\{ \hat{\Sigma}_X^{-1^{(1,0)}}(s,t) - \Sigma_X^{-1^{(1,0)}}(s,t) \right\}_{lm} x_{\mathsf{new}}(t)_m,
\end{aligned}
$$

and so,

$$
\begin{aligned}
\sup_{s,t\in[0,1]} |c_{lm}^{(1,0)}(s,t)| \le M^2 &\sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t)\| \\
&+ M^2 \sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1^{(1,0)}}(s,t) - \Sigma_X^{-1^{(1,0)}}(s,t)\| \\
&\xrightarrow{a.s.} 0.
\end{aligned}
$$

The case $d = 0$ and $p = 1$ is analogous by symmetry. For the last case, $d = p = 1$, we have

$$
\begin{aligned}
c_{lm}^{(1,1)}(s,t) = &\, x_{\mathsf{new}}^{(1)}(s)_l \left\{ \hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t) \right\}_{lm} x_{\mathsf{new}}^{(1)}(t)_m \\
&+ x_{\mathsf{new}}^{(1)}(s)_l \left\{ \hat{\Sigma}_X^{-1^{(0,1)}}(s,t) - \Sigma_X^{-1^{(0,1)}}(s,t) \right\}_{lm} x_{\mathsf{new}}(t)_m \\
&+ x_{\mathsf{new}}(s)_l \left\{ \hat{\Sigma}_X^{-1^{(1,0)}}(s,t) - \Sigma_X^{-1^{(1,0)}}(s,t) \right\}_{lm} x_{\mathsf{new}}^{(1)}(t)_m \\
&+ x_{\mathsf{new}}(s)_l \left\{ \hat{\Sigma}_X^{-1^{(1,1)}}(s,t) - \Sigma_X^{-1^{(1,1)}}(s,t) \right\}_{lm} x_{\mathsf{new}}(t)_m,
\end{aligned}
$$

and therefore,

$$
\begin{aligned}
\sup_{s,t\in[0,1]} |c_{lm}^{(1,1)}(s,t)| \le M^2 &\sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1}(s,t) - \Sigma_X^{-1}(s,t)\| \\
&+ M^2 \sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1^{(0,1)}}(s,t) - \Sigma_X^{-1^{(0,1)}}(s,t)\| \\
&+ M^2 \sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1^{(1,0)}}(s,t) - \Sigma_X^{-1^{(1,0)}}(s,t)\| \\
&+ M^2 \sup_{s,t\in[0,1]} \|\hat{\Sigma}_X^{-1^{(1,1)}}(s,t) - \Sigma_X^{-1^{(1,1)}}(s,t)\| \\
&\xrightarrow{a.s.} 0.
\end{aligned}
$$

Since K is finite, we get

$$
\sup_{s,t\in[0,1]} |\hat{\sigma}_{x_{\mathsf{new}}}^{(d,p)}(s,t) - \sigma_{x_{\mathsf{new}}}^{(d,p)}(s,t)| \le \sum_{l=1}^{K} \sum_{m=1}^{K} \sup_{s,t\in[0,1]} |c_{lm}^{(d,p)}(s,t)| \xrightarrow{a.s.} 0.
$$

■

Theorem 11 (CLT for Functional OLS Estimator) *Let $\hat{\beta} = \{\hat{\beta}(t), t \in [0,1]\}$ denote the ordinary least squares (OLS) estimator,*

$$\hat{\beta}(t) = \left(\sum_{i=1}^{n} X_i(t) X_i^{\mathsf{T}}(t) \right)^{-1} \sum_{i=1}^{n} X_i(t) Y_i(t)$$

of $\beta = \{\beta(t), t \in [0,1]\}$. Under Assumptions 1 - 5, it holds that

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to_D \mathcal{G}_K(0, c_\beta), \quad in \quad C[0,1], \quad as \quad n \to \infty,$$

where $\mathcal{G}_K(0, c_\beta)$ is a mean-zero K-dimensional Gaussian process with covariance function $c_\beta(s,t) = \sigma_\varepsilon(s,t)\Sigma_X^{-1}(s,t)$, where $\Sigma_X^{-1}(s,t) = (\mathbb{E}[X(s)X^{\mathsf{T}}(t)])^{-1}$.

Proof We can write

$$\sqrt{n}\left(\hat{\beta}(t) - \beta(t)\right) = \left(\frac{1}{n} \sum_{i=1}^{n} X_i(t) X_i^{\mathsf{T}}(t) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i(t) \varepsilon_i(t),$$

and in Theorem 10 (a) we have shown that

$$\sup_{t \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} X_i(t) X_i^{\mathsf{T}}(t) - \mathbb{E}\left[X(t) X^{\mathsf{T}}(t) \right] \right| \overset{a.s.}{\to} \underset{(K \times K)}{0}.$$

Let $X_i \cdot \varepsilon_i = \{X_i(t)\varepsilon_i(t) : t \in [0,1]\}$ denote the pointwise multiplication of the processes X_i and ε_i. We start by showing that $n^{-1/2} \sum_{i=1}^{n} (X_i \cdot \varepsilon_i)$ converges in $C[0,1]$ to a mean zero Gaussian process with covariance kernel

$$c_{X\varepsilon}(s,t) = \mathbb{E}[(X(s)\varepsilon(s))(X(t)\varepsilon(t))^{\mathsf{T}}].$$

To show convergence of the multivariate process, we employ the Cramer-Wold device. Let $\lambda \in \mathbb{R}^K$ be arbitrary and focus on the process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda^{\mathsf{T}} (X_i \cdot \varepsilon_i).$$

To show convergence of this real-valued process in $C[0,1]$, we need to show (1) its finite-dimensional distributions converge to those of the limit process, and (2) the process is stochastically equicontinuous. The result then follows by Theorem 7.5 in Billingsley (1999). Since the data is iid, the finite-dimensional distributions converge by a standard central limit theorem. Moreover, above we showed that $X_{i,k} \cdot \varepsilon_i$ is stochastically equicontinuous for each $k = 1, \ldots, K$; and therefore $\lambda^{\mathsf{T}}(X_i \cdot \varepsilon_i)$ is stochastically equicontinuous. Thus,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda^{\mathsf{T}} (X_i \cdot \varepsilon_i) \to_D \mathcal{G}\left(0, \lambda^{\mathsf{T}} c_{X\varepsilon} \lambda\right) \quad \text{in} \quad C[0,1],$$

As λ was arbitrarily chosen, by the Cramer-Wold device, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \cdot \varepsilon_i \to_D \mathcal{G}_K\left(0, c_{X\varepsilon}\right) \quad \text{in} \quad C[0,1].$$

A functional version of Slutsky's Theorem then gives

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \to_D \mathcal{G}_K\left(0, c_\beta\right) \quad \text{in} \quad C[0,1],$$

where $c_\beta(s,t) = \mathbb{E}\left[X(s)X(t)\right]^{-1} c_{X\varepsilon}(s,t)\mathbb{E}\left[X(s)X(t)\right]^{-1}$. Under Assumption 2, we have

$$\mathbb{E}[X(s)\varepsilon(s)\varepsilon(t)X^{\mathsf{T}}(t)] = \sigma_\varepsilon(s,t)\mathbb{E}[X(s)X(t)],$$

and so,

$$c_\beta(s,t) = \sigma_\varepsilon(s,t)\mathbb{E}\left[X(s)X(t)\right]^{-1},$$

which corresponds to the homoskedastic case. In the heteroskedastic case, the formula for the covariance kernel does not simplify and is identified by

$$c_\beta(s,t) = \mathbb{E}\left[X(s)X(t)\right]^{-1}\mathbb{E}[X(s)\varepsilon(s)\varepsilon(t)X^{\mathsf{T}}(t)]\mathbb{E}\left[X(s)X(t)\right]^{-1}.$$

■

Theorem 12 (Consistency of the roughness estimators) *Under Assumptions 1-6, it holds that, as $n \to \infty$,*

(a) $\sup_{t\in[0,1]} |\hat{\tau}_{\mathsf{SPB}}(t) - \tau_{\mathsf{SPB}}(t)| \overset{a.s.}{\to} 0$ *with* $\tau_{\mathsf{SPB}} > 0$, *and*

(b) $\sup_{t\in[0,1]} |\hat{\tau}_{\mathsf{SCB}}(t) - \tau_{\mathsf{SCB}}(t)| \overset{a.s.}{\to} 0$ *with* $\tau_{\mathsf{SCB}} > 0$.

Proof (a) By Lemma 9, we know that $\tau_{\mathsf{SPB}} = \tau_{\sigma_\varepsilon}$ which we estimate using $\hat{\tau}_{\mathsf{SPB}} = \tau_{\hat{\sigma}_\varepsilon}$. By Assumption 6, we therefore have $\tau_{\mathsf{SPB}} > 0$. In Theorem 10 (f), we have shown that

$$\sup_{s,t\in[0,1]} \left|\hat{\sigma}_\varepsilon^{(d,p)}(s,t) - \sigma_\varepsilon^{(d,p)}(s,t)\right| \overset{a.s.}{\to} 0. \tag{31}$$

for $d, p \in \{0,1\}$. The partial derivative

$$\frac{\partial^2}{\partial s \partial t}\frac{\hat{\sigma}_\varepsilon(s,t)}{\sqrt{\hat{\sigma}_\varepsilon(s,s)\hat{\sigma}_\varepsilon(t,t)}}$$

can be constructed by combining the results in Equation (31) using continuous functions (addition, multiplication, and division). Therefore, by the functional continuous mapping theorem,

$$\sup_{s,t\in[0,1]} \left|\frac{\partial^2}{\partial s \partial t}\frac{\hat{\sigma}_\varepsilon(s,t)}{\sqrt{\hat{\sigma}_\varepsilon(s,s)\hat{\sigma}_\varepsilon(t,t)}} - \frac{\partial^2}{\partial s \partial t}\frac{\sigma_\varepsilon(s,t)}{\sqrt{\sigma_\varepsilon(s,s)\sigma_\varepsilon(t,t)}}\right| \overset{a.s.}{\to} 0,$$

and so

$$\sup_{t\in[0,1]} \left|\frac{\partial^2}{\partial s \partial t}\frac{\hat{\sigma}_\varepsilon(s,t)}{\sqrt{\hat{\sigma}_\varepsilon(s,s)\hat{\sigma}_\varepsilon(t,t)}}\right|_{(s,t)=(t,t)} - \frac{\partial^2}{\partial s \partial t}\frac{\sigma_\varepsilon(s,t)}{\sqrt{\sigma_\varepsilon(s,s)\sigma_\varepsilon(t,t)}}\Bigg|_{(s,t)=(t,t)}\Bigg| \overset{a.s.}{\to} 0$$

$$\implies \sup_{t\in[0,1]} \left|\tau_{\hat{\sigma}_\varepsilon}(t)^2 - \tau_{\sigma_\varepsilon}(t)^2\right| \overset{a.s.}{\to} 0 \implies \sup_{t\in[0,1]} \left|\hat{\tau}_{\mathsf{SPB}}(t) - \tau_{\mathsf{SPB}}(t)\right| \overset{a.s.}{\to} 0.$$

(b) Let $x_{\mathsf{new}} : [0,1] \to \mathbb{R}^K \in C^1([0,1], \mathbb{R}^K)$ be a continuously differentiable deterministic function with $x_{\mathsf{new}}(t) \neq 0$ for all $t \in [0,1]$. Define $\sigma_{x_{\mathsf{new}}}(s,t) = x_{\mathsf{new}}^{\mathsf{T}}(s)\Sigma_X^{-1}(s,t)x_{\mathsf{new}}(t)$ as in Assumption 6, and $\hat{\sigma}_{x_{\mathsf{new}}}(s,t) = x_{\mathsf{new}}^{\mathsf{T}}(s)\hat{\Sigma}_X^{-1}(s,t)x_{\mathsf{new}}(t)$. By Lemma 9, we know that $\tau_{\mathsf{SPB}} = \tau_c$, with $c(s,t) = \sigma_\varepsilon(s,t)\sigma_{x_{\mathsf{new}}}(s,t)$. Consider the correlation function $\tilde{c}$ based on c:

$$\begin{aligned}\tilde{c}(s,t) &= \frac{c(s,t)}{\sqrt{c(s,s)c(t,t)}} = \left(\frac{\sigma_\varepsilon(s,t)}{\sqrt{\sigma_\varepsilon(s,s)\sigma_\varepsilon(t,t)}}\right)\left(\frac{\sigma_{x_{\mathsf{new}}}(s,t)}{\sqrt{\sigma_{x_{\mathsf{new}}}(s,s)\sigma_{x_{\mathsf{new}}}(t,t)}}\right) \\ &= \tilde{\sigma}_\varepsilon(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}(s,t).\end{aligned}$$

Taking partial derivatives yields

$$\begin{aligned}&\frac{\partial^2}{\partial s \partial t}\tilde{c}(s,t) = \frac{\partial^2}{\partial s \partial t}\left(\tilde{\sigma}_\varepsilon(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}(s,t)\right) \qquad (32)\\ &= \tilde{\sigma}_\varepsilon^{(1,1)}(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}(s,t) + \tilde{\sigma}_\varepsilon(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}^{(1,1)}(s,t) + \tilde{\sigma}_\varepsilon^{(1,0)}(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}^{(0,1)}(s,t) + \tilde{\sigma}_\varepsilon^{(0,1)}(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}^{(1,0)}(s,t),\end{aligned}$$

where $\sigma^{(d,p)}$ denotes the d-th partial derivative in the first variable and p-th partial derivative in the second variable of σ. This implies that

$$\begin{aligned}\tau_{\mathsf{SPB}}^2(t) &= \frac{\partial^2}{\partial s \partial t}\tilde{c}(s,t)|_{(s,t)=(t,t)} \\ &= \tau_{\sigma_\varepsilon}^2(t) + \tau_{\sigma_{x_{\mathsf{new}}}}^2(t) + \left(\tilde{\sigma}_\varepsilon^{(1,0)}(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}^{(0,1)}(s,t) + \tilde{\sigma}_\varepsilon^{(0,1)}(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}^{(1,0)}(s,t)\right)|_{(s,t)=(t,t)}.\end{aligned}$$

And since $\tau_{\sigma_\varepsilon}^2(t) > 0$ and $\tau_{\sigma_{x_{\mathsf{new}}}}^2(t) > 0$ by Assumption 6, and because

$$\tilde{\sigma}_\varepsilon^{(1,0)}(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}^{(0,1)}(s,t)|_{(s,t)=(t,t)} \geq 0 \quad \text{and} \quad \tilde{\sigma}_\varepsilon^{(0,1)}(s,t)\tilde{\sigma}_{x_{\mathsf{new}}}^{(1,0)}(s,t)|_{(s,t)=(t,t)} \geq 0,$$

we know that $\tau_{\mathsf{SPB}}(t) > 0$.

Let us now show uniform convergence. We have already shown uniform convergence of the partial derivatives of $\hat{\sigma}_\varepsilon$ (Equation 31). In Theorem 10 (g), we have further shown that

$$\sup_{s,t\in[0,1]} |\hat{\sigma}_{x_{\mathsf{new}}}^{(d,p)}(s,t) - \sigma_{x_{\mathsf{new}}}^{(d,p)}(s,t)| \overset{a.s.}{\to} 0,$$

for $d, p \in \{0,1\}$. Writing out Equation (32) for the empirical case (where we estimate $\sigma_{\mathsf{SCB}}(s,t)$ by $\hat{\sigma}_{\mathsf{SCB}}(s,t) = \hat{\sigma}_\varepsilon(s,t)\hat{\sigma}_{x_{\mathsf{new}}}(s,t)$), we see that the above statements, combined with the uniform continuous mapping theorem, imply that

$$\sup_{s,t\in[0,1]}\left|\frac{\partial^2}{\partial s\partial t}\frac{\hat{\sigma}_{\mathsf{SCB}}(s,t)}{\sqrt{\hat{\sigma}_{\mathsf{SCB}}(s,s)\hat{\sigma}_{\mathsf{SCB}}(t,t)}} - \frac{\partial^2}{\partial s\partial t}\frac{\sigma_{\mathsf{SCB}}(s,t)}{\sqrt{\sigma_{\mathsf{SCB}}(s,s)\sigma_{\mathsf{SCB}}(t,t)}}\right| \overset{a.s.}{\to} 0,$$

and so,

$$\sup_{s,t\in[0,1]}\left|\frac{\partial^2}{\partial s\partial t}\frac{\hat{\sigma}_{\mathsf{SCB}}(s,t)}{\sqrt{\hat{\sigma}_{\mathsf{SCB}}(s,s)\hat{\sigma}_{\mathsf{SCB}}(t,t)}}|_{(s,t)=(t,t)} - \frac{\partial^2}{\partial s\partial t}\frac{\sigma_{\mathsf{SCB}}(s,t)}{\sqrt{\sigma_{\mathsf{SCB}}(s,s)\sigma_{\mathsf{SCB}}(t,t)}}|_{(s,t)=(t,t)}\right| \overset{a.s.}{\to} 0$$

$$\implies \sup_{t\in[0,1]}\left|\tau_{\hat{\sigma}_{\mathsf{SCB}}}(t)^2 - \tau_{\sigma_{\mathsf{SCB}}}(t)^2\right| \overset{a.s.}{\to} 0 \implies \sup_{t\in[0,1]}\left|\hat{\tau}_{\mathsf{SCB}}(t) - \tau_{\mathsf{SCB}}(t)\right| \overset{a.s.}{\to} 0.$$

■

References

Konrad Abramowicz, Charlotte K. Häger, Alessia Pini, Lina Schelin, Sara Sjöstedt De Luna, and Simone Vantini. Nonparametric inference for functional-on-scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament. *Scandinavian Journal of Statistics*, 45(4):1036–1061, 2018.

Robert J Adler and Jonathan E Taylor. *Random Fields and Geometry*. Springer, 2007.

Jean-Marc Azaïs, Jean-Marc Bardet, and Mario Wschebor. On the tails of the distribution of the maximum of a smooth stationary gaussian process. *ESAIM: Probability and Statistics*, 6:177–184, 2002.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Ying Wei. Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *The Annals of Statistics*, 46(6):3643–3675, 2018.

Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 1999.

Florentina Bunea, Andrada E. Ivanescu, and Marten H. Wegkamp. Adaptive inference for the mean of a Gaussian process in functional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):531–558, 2011.

T. Tony Cai and Peter Hall. Prediction in functional linear regression. *The Annals of Statistics*, 34(5):2159 – 2179, 2006. doi: 10.1214/009053606000000830. URL `https://doi.org/10.1214/009053606000000830`.

Guanqun Cao. Simultaneous confidence bands for derivatives of dependent functional data. *Electronic Journal of Statistics*, 8(2):2639–2663, 2014.

Guanqun Cao, Lijian Yang, and David Todem. Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24(2):359–377, 2012.

Hervé Cardot, André Mas, and Pascal Sarda. Clt in functional linear regression models. *Probability Theory and Related Fields*, 138(3):325–361, 2007.

Chung Chang, Xuejing Lin, and R. Todd Ogden. Simultaneous confidence bands for functional regression models. *Journal of Statistical Planning and Inference*, 188:67–81, 2017.

Hsin-wen Chang and Ian W. McKeague. Concurrent functional linear regression via plug-in empirical likelihood. *Sankhya A*, 2024. ISSN 0976-8378. doi: 10.1007/s13171-024-00369-3. URL `http://dx.doi.org/10.1007/s13171-024-00369-3`.

Wanghuan Chu, Runze Li, and Matthew Reimherr. Feature screening for time-varying coefficient models with ultrahigh-dimensional longitudinal data. *The Annals of Applied Statistics*, 10(2):596–617, 2016. ISSN 1932-6157. doi: 10.1214/16-aoas912. URL `http://dx.doi.org/10.1214/16-AOAS912`.

Christophe Crambes, Alois Kneip, and Pascal Sarda. Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, 37(1):35 – 72, 2009. doi: 10.1214/07-AOS563. URL `https://doi.org/10.1214/07-AOS563`.

Francisco Cribari-Neto. Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, 45(2):215–233, 2004. ISSN 0167-9473. doi: https://doi.org/10.1016/S0167-9473(02)00366-3. URL `https://www.sciencedirect.com/science/article/pii/S0167947302003663`.

Kishor Das, Thiago De Paula Oliveira, and John Newell. Comparison of markerless and marker-based motion capture systems using 95% functional limits of agreement in a linear mixed-effects modelling framework. *Scientific Reports*, 13(1):22880, 2023.

James Davidson. *Stochastic Limit Theory: An Introduction for Econometricians*. Advanced Texts in Econometrics. Oxford University Press, second edition edition, 2021. ISBN 978-0-19-284450-7.

Galappaththige S. R. de Silva and Pankaj K. Choudhary. Tolerance bands for exponential family functional data. *Canadian Journal of Statistics*, 52(4):e11808, 2024. doi: https://doi.org/10.1002/cjs.11808.

David A Degras. Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4):1735–1765, 2011.

Aurore Delaigle and Peter Hall. Defining probability density for a distribution of random functions. *The Annals of Statistics*, 38(2), April 2010. ISSN 0090-5364. doi: 10.1214/09-aos741. URL `http://dx.doi.org/10.1214/09-AOS741`.

Emanuele Giovanni Depaoli, Marco Stefanucci, and Stefano Mazzuco. Functional concurrent regression with compositional covariates and its application to the time-varying effect of causes of death on human longevity. *The Annals of Applied Statistics*, 18(2):1668 – 1685, 2024. doi: 10.1214/23-AOAS1853. URL `https://doi.org/10.1214/23-AOAS1853`.

Jacopo Diquigiovanni, Matteo Fontana, Aldo Solari, Simone Vantini, and Paolo Vergottini. *conformalInference.fd: Tools for Conformal Inference for Regression in Multivariate Functional Setting*, 2022a. URL `https://CRAN.R-project.org/package=conformalInference.fd`. R package version 1.1.1.

Jacopo Diquigiovanni, Matteo Fontana, and Simone Vantini. Conformal prediction bands for multivariate functional data. *Journal of Multivariate Analysis*, 189:104879, 2022b. doi: https://doi.org/10.1016/j.jmva.2021.104879.

Kreske Ecker, Xavier de Luna, and Lina Schelin. Causal inference with a functional outcome. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 73(1):221–240, 2024.

Frédéric Ferraty and Philippe Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, 2006.

Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: A unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

Maria Franco-Villoria and Rosaria Ignaccolo. Bootstrap based uncertainty bands for prediction in functional kriging. *Spatial Statistics*, 21:130–148, 2017.

Rahul Ghosal, Arnab Maity, Timothy Clark, and Stefano B Longo. Variable selection in functional linear concurrent regression. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 69(3):565–587, 2020.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Marjorie G. Hahn. Conditions for sample-continuity and the central limit theorem. *The Annals of Probability*, 5(3), 1977.

Peter Hall and Joel L. Horowitz. Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70 – 91, 2007. doi: 10.1214/009053606000000957. URL https://doi.org/10.1214/009053606000000957.

Bruce Hansen. *Econometrics*. Princeton University Press, 2022.

Trevor Hastie and Robert Tibshirani. Varying coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779, 1993. ISSN 00359246. doi: 10.1111/j.2517-6161.1993.tb01939.x.

Fumio Hayashi. *Econometrics*. Princeton University Press, 2000.

Lajos Horváth and Piotr Kokoszka. *Inference for Functional Data with Applications*. Springer Science & Business Media, 2012.

Rafael Izbicki, Gilson Shimizu, and Rafael Stern. Flexible distribution-free conditional predictive bands using density estimators. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 3068–3077. PMLR, 26–28 Aug 2020. URL https://proceedings.mlr.press/v108/izbicki20a.html.

M. Kac. On the average number of real roots of a random algebraic equation. *Bulletin of the American Mathematical Society*, 49(4):314–320, 1943.

Piotr Kokoszka and Matthew Reimherr. *Introduction to Functional Data Analysis*. Chapman and Hall/CRC, 2017.

David Kraus. Prediction intervals and bands with improved coverage for functional data under noisy discrete observation. *Journal of Applied Statistics*, 52(6):1258–1277, 2025. doi: 10.1080/02664763.2024.2420223.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1): 71–96, 2014.

Mark W. Lenhoff, Thomas J. Santner, James C. Otis, Margaret G.E. Peterson, Brian J. Williams, and Sherry I. Backus. Bootstrap prediction and confidence bands: A superior statistical method for analysis of gait data. *Gait & Posture*, 9(1):10–17, 1999.

Kui-Wai Li, Xianbo Zhou, and Zhewen Pan. Cross-country output convergence and growth: Evidence from varying coefficient nonparametric method. *Economic Modelling*, 55:32–41, 2016.

Dominik Liebl. Inference for sparse and dense functional data with covariate adjustments. *Journal of Multivariate Analysis*, 170:315–335, 2019a. doi: https://doi.org/10.1016/j.jmva.2018.04.006.

Dominik Liebl. Nonparametric testing for differences in electricity prices: The case of the fukushima nuclear accident. *The Annals of Applied Statistics*, 13(2):1128–1146, 2019b. doi: 10.1214/18-AOAS1230.

Dominik Liebl and Matthew Reimherr. Fast and fair simultaneous confidence bands for functional parameters. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):842–868, 2023.

J. Scott Long and Laurie H. Ervin. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224, 2000. doi: 10.1080/00031305.2000.10474549. URL `https://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474549`.

Richard A. Olshen, Edmund N. Biden, Marilynn P. Wyatt, and David H. Sutherland. Gait analysis and the bootstrap. *The Annals of Statistics*, 17(4):1419–1440, 1989.

Sven Otto and Luis Winter. Functional factor regression with an application to electricity price curve modeling. *arXiv preprint 2503.12611*, 2025. URL `https://arxiv.org/abs/2503.12611`.

Efstathios Paparoditis and Han Lin Shang. Bootstrap prediction bands for functional time series. *Journal of the American Statistical Association*, 118(542):972–986, 2023.

Justin Petrovich, Bahaeddine Taoufik, and Zachary George Davis. Instrumental variable estimation for functional concurrent regression models. *Journal of Applied Statistics*, 51 (8):1570–1589, 2024. doi: 10.1080/02664763.2023.2229968.

V I Piterbarg. Comparison of distribution functions of maxima of gaussian processes. *Theory of Probability & Its Applications*, 26(4):687–705, 1982.

J. O. Ramsay and B. W. Silverman. *Functional data analysis.* Springer, 2nd edition, 2005. ISBN 978-0-387-40080-8.

Lasitha N. Rathnayake and Pankaj K. Choudhary. Tolerance bands for functional data. *Biometrics*, 72(2):503–512, 2016.

Stephen O Rice. Mathematical analysis of random noise. *Bell System Technical Journal*, 24:46–156, 1945.

Radhey S. Singh. Estimation of error variance in linear regression models with errors having multivariate student-t distribution with unknown degrees of freedom. *Economics Letters*, 27(1):47–53, 1988.

Jonathan Taylor, Akimichi Takemura, and Robert J Adler. Validity of the expected euler characteristic heuristic. *The Annals of Probability*, 33(4):1362–1396, 2005.

Fabian J.E. Telschow and Armin Schwartzman. Simultaneous confidence bands for functional data using the Gaussian kinematic formula. *Journal of Statistical Planning and Inference*, 216:70–94, 2022.

Agostino Torti, Alessia Pini, Simone Vantini, et al. Modelling time-varying mobility flows using function-on-function regression: analysis of a bike sharing system in the city of milan. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(1): 226–247, 2021.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

Yueying Wang, Guannan Wang, Li Wang, and R. Todd Ogden. Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics*, 76 (2):427–437, 2020.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980.

Steffen Willwacher, Volker Herrmann, Kai Heinrich, Johannes Funken, Gerda Strutzenberger, Jan-Peter Goldmann, Björn Braunstein, Adam Brazil, Gareth Irwin, Wolfgang Potthast, and Gert-Peter Brüggemann. Sprint start kinetics of amputee and non-amputee sprinters. *PLOS ONE*, 11(11):e0166219, 2016.

World Athletics. *Amendments to Rule 6.3.4 of the Technical Rules*. 2020a. URL `https://timesofsandiego.com/wp-content/uploads/2023/07/Technical-Rules-amendment-to-Rule-6.3.4.pdf`.

World Athletics. *Competition and Technical Rules*. 2020b. URL `https://www.worldathletics.org/download/download?filename=0400178d-ce56-43ff-aa9c-6b4433aafcea.pdf&urlslug=Competition%20%26%20Technical%20Rules%20%E2%80%93%202020%20Edition`.

Jun Zhang, Murray K Clayton, and Philip A Townsend. Functional concurrent linear regression model for spatial images. *Journal of Agricultural, Biological, and Environmental Statistics*, 16:105–130, 2011.