Inferential procedures for partially observed functional data[*]

David Kraus[†]

May 27, 2019

Abstract: In functional data analysis it is usually assumed that all functions are completely, densely or sparsely observed on the same domain. Recent applications have brought attention to situations where each functional variable may be observed only on a subset of the domain while no information about the function is available on the complement. Various advanced methods for such partially observed functional data have already been developed but, interestingly, some essential methods, such as K-sample tests of equal means or covariances and confidence intervals for eigenvalues and eigenfunctions, are lacking. Without requiring any complete curves in the data, we derive asymptotic distributions of estimators of the mean function, covariance operator and eigenelements and construct hypothesis tests and confidence intervals. To overcome practical difficulties with storing large objects in computer memory, which arise due to partial observation, we use the nonparametric bootstrap approach. The proposed methods are investigated theoretically, in simulations and on a fragmentary functional data set from medical research.

Key words and phrases: Bootstrap; covariance operator; functional data; K-sample test; partial observation; principal components.

1 Introduction

Functional data analysis is an established field (Ramsay and Silverman, 2005; Ferraty and Romain, 2011; Horváth and Kokoszka, 2012; Kokoszka and Reimherr, 2017) with well-developed methodologies for common types of observation of random curves, i.e., full (or dense) and sparse observation regimes. Due to new applications recent years have seen the emergence of a new type of observation of functional data, called functional fragments or partially observed functional data. For various examples see Bugni (2012), Delaigle and Hall (2013), Liebl (2013), Gellar et al. (2014), Goldberg et al. (2014), Kraus (2015), Delaigle and Hall (2016), Gromenko et al. (2017), Kneip and Liebl (2017), Dawson and Müller (2018), Mojirsheibani and Shaw (2018), Stefanucci et al. (2018), Descary and Panaretos (2019), Kraus and Stefanucci (2019) or Liebl and Rameseder (2019).

Functional data are collections of observations of random elements of a function space, such as curves, images, surfaces, spatio-temporal fields. We consider random functions in a separable Hilbert space. Without loss of generality we work with the space $L^2([0,1])$ of square-integrable functions on $[0,1]$ equipped with inner product $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$ and norm $\|f\| = \langle f, f \rangle^{1/2}$ but our results are applicable to more general spaces. Partially observed functional data consist of realizations of random functions that are not observed on the entire domain. Each function in the sample may be observed on a different subset of the domain and no information is available on the function values at arguments in the complement of this subset. For the ith functional variable $X_i \in L^2([0,1])$ there is a subset $O_i \subseteq [0,1]$ such that $X_i(t)$ is observed for $t \in O_i$ and not observed for $t \in [0,1] \setminus O_i$. The observation sets may be random, corresponding to data that are missing by happenstance, or non-random for designed experiments. We assume that the observation sets are mutually independent and independent of the curves. We refer to Liebl and Rameseder (2019) for a study of the case of dependent missingness.

Although some advanced procedures, such as goodness-of-fit tests, regression, classification and reconstruction methods, have been developed for functional fragments, basic methods of inference about the

---

[†]Department of Mathematics and Statistics, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic; *david.kraus@mail.muni.cz*.

fundamental characteristics of functional variables are still missing. In particular, the asymptotic distribution of estimators of the mean function and covariance operator, K-sample tests of equal means or covariances, and confidence intervals for eigenvalues and eigenfunctions have not been studied yet in the setting of incomplete functions. Users who wish to perform these basic tasks currently have the only option: to omit the partially observed functions and apply existing procedures to the complete data only. This approach is not only clearly sub-optimal due to a possibly large loss of information and resulting decay of power and accuracy, but also hardly or totally inapplicable in situations where the data contain few or no complete curves.

In this paper, we address this deficiency of existing methodology and develop essential methods of inference about the mean and covariance structure of incomplete functional data. Random functions are characterized by the mean function $\mu = \mathrm{E}\, X$ and the covariance operator $\mathscr{R} : L^2([0,1]) \to L^2([0,1])$ defined as

$$(\mathscr{R}f)(\cdot) = \int_0^1 \rho(\cdot, t)f(t)dt, \quad f \in L^2([0,1]),$$

where $\rho(s,t) = \mathrm{cov}\{X(s), X(t)\}$ is the covariance function, assuming it exists. The covariance structure is best understood via principal component analysis or eigendecomposition of $\mathscr{R}$ in the form

$$\mathscr{R} = \sum_{m=1}^{\infty} \lambda_m \varphi_m \otimes \varphi_m,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues, φ_m are the corresponding orthornormal eigenfunctions, and $(a \otimes b)f = \langle b, f \rangle a$ for $a, b, f \in L^2([0,1])$. For a theoretical background see, e.g., Bosq (2000).

We find appropriate assumptions on the observation pattern that enable us to establish the asymptotic distribution of estimators of μ and $\mathscr{R}$. We develop tests for comparing the mean functions in K populations of functional data based on samples of fragments. Next, we propose several tests of equal covariance operators in K samples. We also construct confidence intervals for the eigenvalues and eigenfunctions estimated from incomplete data.

The practical implementation of methods for functional fragments is more complicated than for complete curves. The main difficulty is that temporal averaging (e.g., in inner products for dimension reduction) is impossible due to missing values. This leads to asymptotic distributions whose parameters follow rather complicated formulas. More importantly, since dimension reduction is not possible, the asymptotic distributions are, upon discretization, characterized by large objects (matrices or arrays) that are difficult or even impossible to store and manipulate in computer memory. The bootstrap turns out to be a solution to this problem. We provide specific algorithms for resampling functional fragments for mean and covariance testing and for confidence intervals for eigenelements.

In a simulation study we investigate the performance of the proposed tests, focusing in particular on the impact of missingness on the different tests and on the effect of the interplay between missingness and the form of differences between groups. The study shows that the proposed methods are superior to the currently only available approach based on omitting incomplete curves.

The proposed methodology is applied to a data set of temporal profiles of heart rate. The data consist of several hundred curves recorded by an automatic device during several hours in the evening during the transition from the day to night regime of heart activity. The profiles are not observed always available on the entire domain of interest because either the device did not measure or record measurements, or the person switched off the device. These fragmentary data were previously analyzed in Kraus (2015), where further details can be found.

Section 2 develops methods of inference about means in one and K samples. Section 3 deals with tests about covariance operators and with inference about principal components. Section 4 presents bootstrap approximations. Results of the simulation study and the data example are reported in Sections 5 and 6. In the Appendix we provide a central limit theorem for non-identically distributed functional variables needed in the asymptotic analysis of fragments, and proofs of all theorems. Additional simulation results and further results of the data analysis.

2 Mean inference from incomplete curves

2.1 Estimation of the mean function

In this section we focus on inference about the mean of functional data. Let us first consider estimation of the mean function μ of a homogeneous population. Let there be n independent functional observations. Each curve X_i, $i \in \{1, \ldots, n\}$ may be observed incompletely, with values known only for arguments in a subset $O_i \subseteq [0,1]$, with no information on the complement of O_i. The observation sets may be non-random or random. They are assumed to be mutually independent and independent of the curves and to consist of a finite union of intervals. We denote by $O_i(t)$ the indicator that the value of $X_i(t)$ is observed.

The mean function $\mu(t)$ can be estimated by the cross-sectional average of available observations

$$\hat{\mu}(t) = \frac{J(t)}{N(t)} \sum_{i=1}^{n} O_i(t) X_i(t),$$

where $N(t) = \sum_{i=1}^{n} O_i(t)$ is the number of available observations at time t and $J(t) = 1_{[N(t)>0]}$. The estimator is defined to be zero when $N(t) = 0$. In Kraus (2015, Proposition 1) it was shown that under non-restrictive assumptions on the observation pattern the estimator $\hat{\mu}$ is consistent for the mean function μ, namely, it was proven that $\mathrm{E}\,\|\hat{\mu} - \mu\|^2 = O(n^{-1})$ as $n \to \infty$. We now aim to provide the asymptotic distribution of the estimator. The result will be essential in the derivation of the limiting distribution of the test statistics that we construct afterwards.

We denote $\pi_i(t) = \mathrm{E}\,O_i(t) = \mathrm{Pr}\{O_i(t) = 1\}$ and $\bar{\pi}(t) = n^{-1}\sum_{i=1}^{n}\pi_i(t)$. Furthermore, we denote by $U_i(s,t) = O_i(s)O_i(t)$ the indicator of observing the function values at the pair of arguments s and t, and define $\nu_i(s,t) = \mathrm{E}\,U_i(s,t)$, $\bar{\nu}(s,t) = n^{-1}\sum_{i=1}^{n}\nu_i(s,t)$ and $M(s,t) = \sum_{i=1}^{n}U_i(s,t)$. We need to introduce conditions on the observation pattern as follows.

Condition 1.

(a) Let there be a function $\pi(t)$ such that $\pi_0 = \inf_{t\in[0,1]}\pi(t) > 0$ and $\sup_{t\in[0,1]}|\bar{\pi}(t) - \pi(t)| \to 0$ for $n \to \infty$.

(b) Let there be a function $\nu(s,t)$ such that $\bar{\nu}(s,t) \to \nu(s,t)$ for all $s,t \in [0,1]$.

(c) Let there be a value $\nu_0 > 0$ such that for each $(s,t) \in [0,1]^2$ either $\nu(s,t) \geq \nu_0$ or $\nu(s,t) = 0$, and let the convergence $\sup_{(s,t)\in[0,1]^2}|\bar{\nu}(s,t) - \nu(s,t)| \to 0$ for $n \to \infty$ hold.

Condition (a) guarantees the consistency of the estimator $\hat{\mu}$, see Kraus (2015). Condition (b) is needed for the weak convergence of the estimator. Condition (c) is needed for consistent estimation of the covariance operator of the limiting distribution. We emphasize that no complete curves are required since these conditions may be satisfied even when the sample contains only fragments. We illustrate this attractive property in the simulation study in Section 5.

When the observation indicators $O_1, \ldots, O_n$ are identically distributed, then Condition (a) is stafisfied if $\pi(t) = P\{O_i(t) = 1\}$ is bounded away from zero, Condition (b) is satisfied automatically and Condition (c) is satisfied if for each $(s,t) \in [0,1]^2$, $\nu(s,t) = P\{O_i(s) = 1, O_i(t) = 1\}$ is either bounded away from zero or equal to zero. The case of non-identically distributed observation indicators may be relevant, for example, for designed experiments in which non-random, designed observation sets may vary across subjects.

By $\|\cdot\|_2$ below we denote the Hilbert–Schmidt norm of an operator.

Theorem 1. *Assume that* $\mathrm{E}(\|X_1\|^2) < \infty$. *Let Conditions 1(a) and 1(b) hold. Then*

$$n^{1/2}\{\hat{\mu}(\cdot) - \mu(\cdot)\}, \quad N(\cdot)^{1/2}\{\hat{\mu}(\cdot) - \mu(\cdot)\}$$

are asymptotically distributed as mean zero Gaussian processes with covariance operators $\mathscr{K}'$, $\mathscr{K}$ *with kernels*

$$\kappa'(s,t) = \pi(s)^{-1}\pi(t)^{-1}\nu(s,t)\rho(s,t), \quad \kappa(s,t) = \pi(s)^{-1/2}\pi(t)^{-1/2}\nu(s,t)\rho(s,t),$$

respectively.

If, moreover, Condition 1(c) is satisfied, then $\mathscr{K}'$ *and* $\mathscr{K}$ *can be consistently estimated by the operators* $\hat{\mathscr{K}}'$ *and* $\hat{\mathscr{K}}$ *with kernels* $\hat{\kappa}'(s,t) = \hat{\pi}(s)^{-1}\hat{\pi}(t)^{-1}\hat{\nu}(s,t)\hat{\rho}(s,t)$ *and* $\hat{\kappa}(s,t) = \hat{\pi}(s)^{-1/2}\hat{\pi}(t)^{-1/2}\hat{\nu}(s,t)\hat{\rho}(s,t)$,

respectively, i.e., $\mathrm{E}\,\|\hat{\mathscr{K}}' - \mathscr{K}'\|_2^2 \to 0$ and $\mathrm{E}\,\|\hat{\mathscr{K}} - \mathscr{K}\|_2^2 \to 0$, where $\hat{\pi}(t) = N(t)/n$, $\hat{\nu}(s,t) = M(s,t)/n$, $\hat{\rho}(s,t)$ is the empirical covariance based on all complete pairs of function values at s, t, and the value of the kernels is set to 0 whenever $\hat{\pi}(s)$ or $\hat{\pi}(t)$ is 0.

The proof of this and other theorems is provided in the Appendix. Since the observable functional variables may be non-identically distributed due to possibly non-identically distributed observation indicators, the proof uses a central limit theorem for non-identically distributed functional random variables given in the Appendix.

Notice that the covariance kernels $\kappa'(s,t)$ and $\kappa(s,t)$ of the limiting distributions are zero when $\nu(s,t) = 0$ regardless of the value of $\rho(s,t)$. Therefore, it is not necessary to estimate $\rho(s,t)$ at such points. This is why Condition 1(c) does not require the function $\nu(s,t)$ to be bounded away from zero on the entire domain $[0,1]^2$ which is needed for the estimation of $\mathscr{R}$, as will be seen in Section 3, Condition 2(a). This means that the theorem applies also in the context of short fragments of curves considered, e.g., by Delaigle and Hall (2016) or Descary and Panaretos (2019), where each curve in the sample is observed on a short interval and no completely observed curves are available.

2.2 Tests of equality of means in several populations

Let us now consider K independent samples of functional data. Let the jth sample ($j \in \{1,\ldots,K\}$) consist of independent curves $X_{j1},\ldots,X_{jn_j}$ coming from the same distribution with mean μ_j and covariance operator $\mathscr{R}_j$. The functions may not be observed completely. It is assumed that for each function X_{ji} its values are available on a subset O_{ij}. Let the observation subsets be mutually independent and independent of the curves. Our aim is to test the null hypothesis that $\mu_1 = \cdots = \mu_K$ against the general alternative that the null does not hold. The literature on hypothesis testing for means of functional data is rich. See, for example, Aue et al. (2009); Benko et al. (2009); Cao et al. (2012); Cuevas et al. (2004); Fogarty and Small (2014); Horváth and Kokoszka (2012); Liebl (2019); Mas (2007); Paparoditis and Sapatinas (2016b); Pini et al. (2018); Pini and Vantini (2016); Vsevolozhskaya et al. (2014); Zhang et al. (2010); Zhang (2013).

In the literature on complete functional samples there exist two main approaches to comparing mean functions. One is based on the L^2 distance between the means and one uses projections on finite dimensional subspaces.

The assessment of the hypothesis will be based on the contrasts of the group means and a null estimate of the common mean, i.e., on the differences $\hat{\mu}_j - \hat{\mu}$, $j \in \{1,\ldots,K\}$. Here we use $\hat{\mu}_j(t) = J_j(t)N_j(t)^{-1}\sum_{i=1}^{n_j} O_{ji}(t)X_{ji}(t)$, $j \in \{1,\ldots,K\}$, with $N_j(t) = \sum_{i=1}^{n_j} O_{ji}(t)$ and $J_j(t) = 1_{[N_j(t)>0]}$. The estimator $\hat{\mu}$ is obtained as a weighted average of the group means in the form $\hat{\mu}(t) = \sum_{j=1}^{K} \hat{w}_j(t)\hat{\mu}_j(t)$ with weights

$$\hat{w}_j(t) = \frac{N_j(t)/\hat{r}_j^2}{\sum_{k=1}^{K} N_k(t)/\hat{r}_k^2},$$

where $\hat{r}_j^2 = \mathrm{tr}\,\hat{\mathscr{R}}_j$ is the trace of the estimated covariance operator in the jth sample (the estimators $\hat{\mathscr{R}}_j$ are discussed later). The role of the scaling by $\hat{r}_j^2$ is to account for possibly different covariance structures in the samples. This way of combining estimated means of heteroscedastic samples is inspired by the univariate case and its standard multivariate extensions. If the covariance structures are known to be the same in all samples, the factors $\hat{r}_j^2$ can be replaced by the trace of an estimator of the common covariance operator, which leads to the estimated mean based on the pooled sample of curves.

The first test we propose is inspired by the method of Cuevas et al. (2004) who in the context of fully observed functional data developed an ANOVA test based on the L^2 norms of the contrasts of the group means and the pooled sample mean. A two-sample version of the test using the nonparametric bootstrap was proposed by Benko et al. (2009). Horváth et al. (2013) studied a two-sample test based on the L^2 norm in the context of functional time series. The standardized contrast processes $N_j(\cdot)^{1/2}\{\hat{\mu}_j(\cdot) - \hat{\mu}(\cdot)\}/\hat{r}_j$, $j \in \{1,\ldots,K\}$ can be collected into a K-dimensional vector that is a random element of the product space $\{L^2([0,1])\}^K$ with inner product $\langle f,g\rangle = \sum_{j=1}^{K}\langle f_j,g_j\rangle$ for $f = (f_1,\ldots,f_K)^\top$, $g = (g_1,\ldots,g_K)^\top$. We use its L^2 norm as the test statistic, i.e., base the test on

$$T_{L^2} = \sum_{j=1}^{K} \|N_j(\cdot)^{1/2}\{\hat{\mu}_j(\cdot) - \hat{\mu}(\cdot)\}/\hat{r}_j\|^2 = \sum_{j=1}^{K} \int_0^1 N_j(t)\{\hat{\mu}_j(t) - \hat{\mu}(t)\}^2/\hat{r}_j^2\,dt \tag{1}$$

4

and reject when the value of the statistic is significantly large.

Another main approach to curve mean testing uses dimension reduction. See, e.g., Aue et al. (2009), Horváth and Kokoszka (2012) or Horváth et al. (2013). The idea is to focus on a finite number of important features of the infinite-dimensional data. The functional observations are projected on a finite-dimensional subspace and multivariate ANOVA or a similar multivariate procedure is applied to the resulting vectors of Fourier scores. This strategy is not directly applicable in the situation of incompletely observed curves because, unlike in the fully observed case, Fourier scores of functional fragments cannot be computed by numerical integration as inner products of the functional variable and the basis function since the functional variable is not available on the entire domain.

Let $\hat{\psi}_1, \ldots, \hat{\psi}_d$ be some linearly independent functions in $L^2([0,1])$. Without loss of generality we assume that they are orthonormal. These functions may be either deterministic or random (estimated from the data). In the construction of our projection tests we use Fourier scores of the standardized contrast processes with respect to the basis functions $\hat{\psi}_l$. We denote these scores $Q_{jl} = \langle N_j(\cdot)\{\hat{\mu}_j(\cdot) - \hat{\mu}(\cdot)\}, \hat{\psi}_l \rangle / (\hat{r}_j n_j^{1/2})$, $j \in \{1, \ldots, K\}$, $l \in \{1, \ldots, d\}$ and collect them in the score vector $Q = (Q_{11}, \ldots, Q_{1d}, \ldots, Q_{K1}, \ldots, Q_{Kd})^\top$. The score statistic is the quadratic form

$$T_d = Q^\top \hat{V}^- Q, \tag{2}$$

where $\hat{V}^-$ is the Moore–Penrose pseudoinverse of the estimated $(Kd) \times (Kd)$ covariance matrix of Q whose entry on the position with index (jl, km) is

$$\hat{V}_{jl,km} = \langle \hat{\pi}_j^{1/2}\hat{\psi}_l, \hat{\mathscr{V}}_{jk}(\hat{\pi}_k^{1/2}\hat{\psi}_m) \rangle = \int_{[0,1]^2} \hat{\pi}_j(s)^{1/2}\hat{\psi}_l(s)\hat{v}_{jk}(s,t)\hat{\psi}_m(t)\hat{\pi}_k(t)^{1/2}dsdt$$

for $j, k \in \{1, \ldots, K\}$, $l, m \in \{1, \ldots, d\}$. Here $\hat{\mathscr{V}}_{jk}$ is the covariance operator with kernel

$$\hat{v}_{jk}(s,t) = \sum_{l=1}^{K} \hat{r}_j^{-1}\{\delta_{jl} - N_j(s)^{1/2}\hat{w}_l(s)N_l(s)^{-1/2}\}\hat{\kappa}_l(s,t)\{\delta_{kl} - N_k(t)^{1/2}\hat{w}_l(t)N_l(t)^{-1/2}\}\hat{r}_k^{-1}, \tag{3}$$

where δ_{jk} is the Kronecker delta. The test rejects for large values of T_d.

Analogously to the case of one group considered in Subsection 2.1, we denote for $j \in \{1, \ldots, K\}$, $i \in \{1, \ldots, n_j\}$ the following quantities characterizing the observation patterns in each group, $\pi_{ji}(t) = \mathrm{E}\,O_{ji}(t) = \Pr(O_{ji}(t) = 1)$, $\bar{\pi}_j(t) = n_j^{-1}\sum_{i=1}^{n_j}\pi_{ji}(t)$, $U_{ji}(s,t) = O_{ji}(s)O_{ji}(t)$, $\nu_{ji}(s,t) = \mathrm{E}\,U_{ji}(s,t)$, $\bar{\nu}_j(s,t) = n_j^{-1}\sum_{i=1}^{n_j}\nu_{ij}(s,t)$ and $M_j(s,t) = \sum_{i=1}^{n_j}U_{ji}(s,t)$. Under mild assumptions we obtain the asymptotic distribution of both test statistics.

Theorem 2. *For $j \in \{1, \ldots, K\}$ assume that $n_j \to \infty$, $n_j/(n_1 + \cdots + n_K) \to a_j > 0$ and $\mathrm{E}\,\|X_{j1}\|^2 < \infty$. Let the observation patterns in each group satisfy Condition 1. Then under the null hypothesis of equal means we obtain the following results:*

(i) *The test statistic T_{L^2} is asymptotically distributed as $\sum_{k=1}^{\infty}\gamma_k C_k$, where C_k are independent chi-square distributed variables with one degree of freedom and γ_k can be consistently estimated by the eigenvalues of the operator $\hat{\mathscr{V}}$ given in (3).*

(ii) *Assume that there exist linearly independent non-random functions $\psi_1, \ldots, \psi_d$ such that $\|\hat{\psi}_l - \psi_l\| \xrightarrow{P} 0$ for $l \in \{1, \ldots, d\}$. Then the test statistic T_d is asymptotically chi-square distributed with $(K-1)d$ degrees of freedom.*

The test statistic based on the L^2 norm is not distribution-free but the critical values can be obtained straightforwardly by simulation, provided that the eigenvalues of $\hat{\mathscr{V}}$ consistently estimate γ_k. Similarly, the consistency of $\hat{\mathscr{V}}$ (and hence of $\hat{V}$) is needed for the score statistic. The consistency of $\hat{\mathscr{V}}$ is guaranteed by Condition 1(c). It may sometimes happen that $M_j(s,t)$ is low for some s, t, making the estimator $\hat{\mathscr{V}}$ less reliable. For this reason, and also for computational reasons, to avoid the estimation of the limiting covariance one can use the bootstrap method, as we describe in Section 4.

In the literature on complete functional data, the most common choice of the basis functions for the projection test is derived from principal component analysis (see Horváth and Kokoszka (2012) and references therein, or Fremdt et al. (2014)). The approach uses several leading eigenfunctions of the pooled sample

covariance operator. The motivation for this choice is the property that the first eigenfunctions capture the principal modes of variation, the most important features of random deviations of the functional variables from the mean. Another approach is to use a fixed set of basis functions, such as several elements of the Fourier basis of sines and cosines or several orthornomial Legendre polynomials.

For several reasons we prefer deterministic bases to the basis of eigenfunctions. One drawback of the latter approach is that the principal components of variability may be only weakly related or entirely unrelated (orthogonal) to the differences between the mean functions, resulting in a test that is weak or inconsistent against this alternative. It may of course happen that the deterministic functions we choose are orthogonal to the alternative too, or that the leading eigenfunctions capture the mean differences well. However, with fixed functions it is at least possible to say before the analysis which alternatives can be detected. With principal components it is not known beforehand which departures from the null can be captured because the eigenfunctions are usually unknown. Moreover, their property of capturing the largest portion of variability, which is typically the main argument for using them, is not exactly what one wishes in mean testing. In fact, one would rather wish to maximize the signal-to-noise ratio or non-centrality, which, for example, in the case of components with equal magnitude of means would mean to minimize variability. In reality, the true interplay between the magnitude of components of the mean difference and their variability is not known, and we, therefore, prefer fixed functions.

The choice of the number of basis functions is important with projection methods. For the approach using eigenfunctions, we follow the recommendation of Horváth et al. (2013) to use the smallest number of components needed to explain at least 85 % of the total variability. For the method using fixed functions, in light of the above discussion of the relation of the power and variability we do not base the choice of d on the explained variability. Instead, we can specify what shape differences we wish to detect and use the corresponding basis functions. For example, using just $d = 3$ Legendre polynomials describing constant, monotonic as well as convex or concave non-monotonic differences seems to be a good choice in many applications.

3   Covariance inference under partial observation

3.1   Asymptotics for the estimated covariance operator and principal components

Given a collection of independent realizations of curves $X_1, \ldots, X_n$ with mean function μ and covariance operator $\mathscr{R}$ observed on subsets $O_1, \ldots, O_n$, the covariance function $\rho(s,t)$ can be estimated by the empirical covariance using pairwise complete observations, that is, by

$$\hat{\rho}(s,t) = \frac{I(s,t)}{M(s,t)} \sum_{i=1}^{n} U_i(s,t)\{X_i(s) - \hat{\mu}_{st}(s)\}\{X_i(t) - \hat{\mu}_{st}(t)\},$$

where $I(s,t) = 1_{[M(s,t)>0]}$ and

$$\hat{\mu}_{st}(s) = \frac{1_{[M(s,t)>0]}}{M(s,t)} \sum_{i=1}^{n} U_i(s,t)X_i(s).$$

If $M(s,t) = 0$, we define $\hat{\rho}(s,t) = 0$ and $\hat{\mu}_{st}(s) = 0$. Under certain assumptions on the observation pattern, the operator $\hat{\mathscr{R}}$ with kernel $\hat{\rho}(s,t)$ was shown to be a consistent estimator of $\mathscr{R}$ in Kraus (2015, Proposition 1).

In the theorem below we give the asymptotic distribution under a set of conditions for which we denote $E_i(s,t,u,v) = O_i(s)O_i(t)O_i(u)O_i(v)$, the indicator that the observation of X_i at points s,t,u,v is available, and set $\theta_i(s,t,u,v) = \Pr\{E_i(s,t,u,v) = 1\}$, $\bar{\theta}(s,t,u,v) = \sum_{i=1}^{n} \theta_i(s,t,u,v)/n$ and $L(s,t,u,v) = \sum_{i=1}^{n} E_i(s,t,u,v)$.

Condition 2.

(a) Let there be a function $\nu(s,t)$ such that $\nu_0 = \inf_{(s,t)\in[0,1]^2} \nu(s,t) > 0$ and $\sup_{(s,t)\in[0,1]^2} |\bar{\nu}(s,t) - \nu(s,t)| \to 0$ for $n \to \infty$.

(b) Let there be a function $\theta(s,t,u,v)$ such that $\bar{\theta}(s,t,u,v) \to \theta(s,t,u,v)$ for all $s,t,u,v \in [0,1]$.

(c) Let there be a value $\theta_0 > 0$ such that for each $(s, t, u, v) \in [0,1]^4$ either $\theta(s, t, u, v) \geq \theta_0$ or $\theta(s, t, u, v) = 0$, and let the convergence $\sup_{(s,t,u,v) \in [0,1]^4} |\bar{\theta}(s, t, u, v) - \theta(s, t, u, v)| \to 0$ for $n \to \infty$ hold.

Condition (a) means that there are enough observations at all pairs of arguments. The condition is needed for the consistency of $\hat{\mathscr{R}}$, see Kraus (2015) for a proof under an essentially equivalent condition. Condition (b) guarantees the weak convergence in the theorem below, and the additional condition (c) guarantees that the covariance of the asymptotic distribution can be estimated. We stress that these conditions do not require that the data contain any complete curves. They may be satisfied even in situations, where all functional observations are fragmentary. When the observation indicators $O_1, \ldots, O_n$ are identically distributed, then Condition (a) is stafisfied if $\nu(t) = P\{O_i(s) = 1, O_i(t) = 1\}$ is bounded away from zero, Condition (b) is satisfied automatically and Condition (c) is satisfied if for each $(s, t, u, v) \in [0,1]^4$, $\theta(s, t, u, v) = P\{O_i(s) = 1, O_i(t) = 1, O_i(u) = 1, O_i(v) = 1\}$ is either bounded away from zero or equal to zero.

Theorem 3. *Assume that $\mathrm{E}(\|X_1\|^4) < \infty$. Let Conditions 2(a) and 2(b) hold. Then $n^{1/2}(\hat{\mathscr{R}} - \mathscr{R})$ and the operator with kernel $M(\cdot, \cdot)^{1/2}\{\hat{\rho}(\cdot, \cdot) - \rho(\cdot, \cdot)\}$ are asymptotically distributed as mean zero Gaussian operators whose covariance operators $\mathfrak{H}', \mathfrak{H}$ have kernels*

$$\eta'(s, t, u, v) = \nu(s, t)^{-1}\nu(u, v)^{-1}\theta(s, t, u, v)\{\zeta(s, t, u, v) - \rho(s, t)\rho(u, v)\},$$

$$\eta(s, t, u, v) = \nu(s, t)^{-1/2}\nu(u, v)^{-1/2}\theta(s, t, u, v)\{\zeta(s, t, u, v) - \rho(s, t)\rho(u, v)\},$$

respectively, where $\zeta(s, t, u, v) = \mathrm{E}[\{X(s) - \mu(s)\}\{X(t) - \mu(t)\}\{X(u) - \mu(u)\}\{X(v) - \mu(v)\}]$.

If, in addition, Condition 2(c) is satisfied, then $\mathfrak{H}'$ and $\mathfrak{H}$ can be consistently estimated by the operators $\hat{\mathfrak{H}}'$ and $\hat{\mathfrak{H}}$ with kernels $\hat{\eta}'(s, t, u, v) = \hat{\nu}(s, t)^{-1}\hat{\nu}(u, v)^{-1}\hat{\theta}(s, t, u, v)\{\hat{\zeta}(s, t, u, v) - \hat{\rho}(s, t)\hat{\rho}(u, v)\}$ and $\hat{\eta}(s, t, u, v) = \hat{\nu}(s, t)^{-1/2}\hat{\nu}(u, v)^{-1/2}\hat{\theta}(s, t, u, v)\{\hat{\zeta}(s, t, u, v) - \hat{\rho}(s, t)\hat{\rho}(u, v)\}$, respectively, i.e., $\mathrm{E}\|\hat{\mathfrak{H}}' - \mathfrak{H}'\|_2^2 \to 0$ and $\mathrm{E}\|\hat{\mathfrak{H}} - \mathfrak{H}\|_2^2 \to 0$, where $\hat{\eta}'(s, t, u, v)$ and $\hat{\eta}(s, t, u, v)$ are set to 0 whenever $\hat{\nu}(s, t)$ or $\hat{\nu}(u, v)$ is 0, $\hat{\theta}(s, t, u, v) = L(s, t, u, v)/n$ and $\hat{\zeta}(s, t, u, v)$ is the empirical fourth central moment of the functional random variable computed using all complete quadruples of function values at arguments s, t, u, v.

The weak convergence in the theorem above is on the separable Hilbert space of Hilbert–Schmidt operators equipped with the Hilbert–Schmidt norm $\| \cdot \|_2$. The limiting covariance operator $\mathfrak{H}$ is an operator that maps a Hilbert–Schmidt operator $\mathscr{F}$ with kernel $f(u, v)$ to an operator with kernel $\int_0^1 \int_0^1 \eta(s, t, u, v)f(u, v)dudv$, similarly for other objects in the theorem.

Next, we study the estimators $\hat{\lambda}_m$ and $\hat{\varphi}_m$ of the eigenvalues and eigenfunctions of $\mathscr{R}$. The estimators are obtained by the eigendecomposition of $\hat{\mathscr{R}}$. Their root-n consistency was established by Kraus (2015, Proposition 2). Here we find the approximate distribution of the fluctuation of the estimators around their true counterparts (with appropriate sign for the eigenfunctions as usual).

Theorem 4. *Assume that $\mathrm{E}(\|X_1\|^4) < \infty$ and $\mathscr{R}$ has eigenvalues with multiplicity 1. Let Conditions 2(a) and 2(b) hold. Denote by $\mathscr{H}'^\infty$ a random operator following the limiting Gaussian distribution of $n^{1/2}(\hat{\mathscr{R}} - \mathscr{R})$ with mean zero and covariance $\mathfrak{H}'$ given in Theorem 3. Then, for $n \to \infty$, we obtain the following results:*

(i) $n^{1/2}(\hat{\lambda}_m - \lambda_m)$ is asymptotically distributed as $\langle \mathscr{H}'^\infty \varphi_m, \varphi_m \rangle$, which is a normal variable with mean zero and variance

$$\int_{[0,1]^4} \varphi_m(s)\varphi_m(t)\eta'(s, t, u, v)\varphi_m(u)\varphi_m(v)dsdtdudv.$$

(ii) $n^{1/2}(\hat{\varphi}_m - \hat{s}_m\varphi_m)$, where $\hat{s}_m = \mathrm{sign}\langle \hat{\varphi}_m, \varphi_m \rangle$, is asymptotically distributed as the Gaussian random function $\mathscr{Q}_m\mathscr{H}'^\infty\varphi_m$, where

$$\mathscr{Q}_m = \sum_{\substack{k=1 \\ k \neq m}}^{\infty} \frac{\varphi_k \otimes \varphi_k}{\lambda_m - \lambda_k}.$$

The limiting covariance operator of $n^{1/2}(\hat{\varphi}_m - \hat{s}_m\varphi_m)$ is

$$\sum_{\substack{k=1 \\ k \neq m}}^{\infty} \sum_{\substack{l=1 \\ l \neq m}}^{\infty} \frac{\varphi_k \otimes \varphi_l}{(\lambda_m - \lambda_k)(\lambda_m - \lambda_l)} \int_{[0,1]^4} \varphi_k(s)\varphi_m(t)\eta'(s, t, u, v)\varphi_m(u)\varphi_l(v)dsdtdudv.$$

7

If, additionally, Condition 2(c) is satisfied, then the limiting variance and covariance above can be consistently estimated by plugging-in estimates from Theorem 3.

The theorem is proved in the Appendix with the help of perturbation theory. The theorem generalizes the classic results of Dauxois et al. (1982) who considered completely observed functions. See Kokoszka and Reimherr (2013) for related results for functional time series. In the case of complete Gaussian curves Dauxois et al. (1982) showed that the limiting covariance structure of the empirical covariance operator simplifies (see also Panaretos et al., 2010) which eventually leads to a simpler form of the limiting variance of the empirical eigenvalue, namely to $2\lambda_m^2$. No such simplification is in general possible in the case of incomplete curves, even if they are Gaussian. Therefore, to make inference about eigenvalues or eigenfunctions, e.g., to construct confidence intervals, one possibility is to estimate the function $\eta'(s,t,u,v)$ and use the complicated expressions above for the limiting covariance structure. In Section 4 we provide an alternative approach based on the bootstrap which enables to avoid the possibly unstable estimation of η' and computer memory demanding storage and manipulation with the estimate.

3.2 Testing the equality of covariance operators

We now study tests for equality of covariance operators of several populations. Let there be K independent samples of partially observed functions with mean μ_j and covariance $\mathscr{R}_j$ in the jth sample, as described in Subsection 2.2. We aim to test the null hypothesis that $\mathscr{R}_1 = \cdots = \mathscr{R}_K$ against the general alternative. The general problem of hypothesis testing for covariance operators was previously studied in various contexts by various methods. See, e.g., Benko et al. (2009); Boente et al. (2017); Cabassi et al. (2017); Fremdt et al. (2013); Guo et al. (2018a,b); Jarušková (2013); Kashlak et al. (2018); Kraus and Panaretos (2012); Masarotto (2019); Panaretos et al. (2010); Paparoditis and Sapatinas (2016b); Pigoli et al. (2014); Pini et al. (2019); Zhang (2013); Zhang and Liang (2014).

Tests of the null hypothesis of equal covariance operators can be based on the differences between the estimators $\hat{\mathscr{R}}_j$ and the null estimator $\hat{\mathscr{R}}$ which is the pooled covariance operator with kernel

$$\hat{\rho}(s,t) = \sum_{j=1}^{K} \hat{w}_j(s,t)\hat{\rho}_j(s,t),$$

where

$$\hat{w}_j(s,t) = \frac{M_j(s,t)}{\sum_{k=1}^{K} M_k(s,t)}.$$

The differences are expressed by the constrast operators with kernels $M_j(\cdot,\cdot)^{1/2}\{\hat{\rho}_j(\cdot,\cdot) - \hat{\rho}(\cdot,\cdot)\}$. We propose two types of tests measuring the importance of the contrasts: one approach is based on the Hilbert–Schmidt norm of the contrasts and one is based on their projections on a subspace.

The first approach is inspired by methods that were previously considered in the case of fully observed functions, e.g., by Boente et al. (2017). The importance of the contrasts is expressed by the Hilbert–Schmidt norm. The test statistic takes the form

$$S_{\mathrm{HS}} = \sum_{j=1}^{K} \|M_j(\cdot,\cdot)^{1/2}\{\hat{\rho}_j(\cdot,\cdot) - \hat{\rho}(\cdot,\cdot)\}\|_2^2 = \sum_{j=1}^{K} \int_{[0,1]^2} M_j(s,t)\{\hat{\rho}_j(s,t) - \hat{\rho}(s,t)\}^2 ds\,dt \tag{4}$$

(in this notation we identify kernels and the corresponding operators).

The second approach uses projections of the contrasts onto a finite-dimensional subspace of the space of Hilbert–Schmidt operators. This type of tests was used for complete functions in various settings, e.g., by Horváth et al. (2010), Panaretos et al. (2010), Panaretos et al. (2011), Kraus and Panaretos (2012), Fremdt et al. (2013), and Jarušková (2013). It is natural to project on the subspace generated by the leading eigenfunctions of $\hat{\mathscr{R}}$ because they carry information about the object of interest, the covariance operator (unlike in the case of mean functions where we prefer to use a fixed basis for the projection test). Let $\hat{\varphi}_1, \ldots, \hat{\varphi}_d$ be the first d eigenfunctions of $\hat{\mathscr{R}}$. Then the operators

$$\hat{\mathscr{U}}_{lm} = \begin{cases} \hat{\varphi}_l \otimes \hat{\varphi}_l, & l = m, \\ (\hat{\varphi}_l \otimes \hat{\varphi}_m + \hat{\varphi}_m \otimes \hat{\varphi}_l)/2^{1/2}, & l < m \end{cases}$$

8

with kernels $\hat{u}_{ll}(s,t) = \hat{\varphi}_l(s)\hat{\varphi}_l(t)$ and $\hat{u}_{lm}(s,t) = \{\hat{\varphi}_l(s)\hat{\varphi}_m(t) + \hat{\varphi}_m(s)\hat{\varphi}_l(t)\}/2^{1/2}$, $l < m$ form an orthonormal basis of a $d(d+1)/2$-dimensional subspace of $\mathrm{HS}(L^2([0,1]))$. The Fourier coefficients of the projection of the jth standardized contrast on this subspace are

$$R_{jlm} = \langle M_j(\cdot,\cdot)\{\hat{\rho}_j(\cdot,\cdot) - \hat{\rho}(\cdot,\cdot)\}/n_j^{1/2}, \hat{\mathscr{U}}_{lm}\rangle = \int_{[0,1]^2} M_j(s,t)\{\hat{\rho}_j(s,t) - \hat{\rho}(s,t)\}\hat{u}_{lm}(s,t)dsdt/n_j^{1/2}. \quad (5)$$

Denote by R the $Kd(d+1)/2$-dimensional score vector with components R_{jlm}, $j \in \{1,\ldots,K\}$, $1 \le l \le m \le d$. The test statistic measures the size of the projection of the contrast operators on the subspace. It takes the form

$$S_d = R\hat{W}^- R, \quad (6)$$

where $\hat{W}^-$ is the Moore–Penrose pseudoinverse of the estimator of the asymptotic covariance matrix whose entry with indices (jlm, kpq) is

$$\begin{aligned}
\hat{W}_{jlm,kpq} &= \langle \hat{\nu}_j(\cdot,\cdot)^{1/2}\hat{u}_{lm}(\cdot,\cdot), \hat{\mathfrak{B}}_{jk}\{\hat{\nu}_k(\cdot,\cdot)^{1/2}\hat{u}_{pq}(\cdot,\cdot)\}\rangle \\
&= \int_{[0,1]^4} \hat{\nu}_j(s,t)^{1/2}\hat{u}_{lm}(s,t)\hat{\beta}_{jk}(s,t,u,v)\hat{u}_{pq}(u,v)\hat{\nu}_k(u,v)^{1/2}dsdtdudv,
\end{aligned} \quad (7)$$

$j,k = 1,\ldots,K$, $1 \le l \le m \le d$, $1 \le p \le q \le d$. The kernel of $\hat{\mathfrak{B}}_{jk}$ is

$$\begin{aligned}
\hat{\beta}_{jk}(s,t,u,v) &= \sum_{l=1}^K \{\delta_{jl} - M_j(s,t)^{1/2}\hat{w}_l(s,t)M_l(s,t)^{-1/2}\}\hat{\eta}_l(s,t,u,v) \\
&\quad \times \{\delta_{kl} - M_k(u,v)^{1/2}\hat{w}_l(u,v)M_l(u,v)^{-1/2}\}.
\end{aligned} \quad (8)$$

We now give the asymptotic distribution of the Hilbert–Schmidt and projection statistics.

Theorem 5. *For $j \in \{1,\ldots,K\}$ assume that $n_j \to \infty$, $n_j/(n_1 + \cdots + n_K) \to a_j > 0$, $\mathrm{E}\|X_{j1}\|^4 < \infty$ and all eigenvalues of $\mathscr{R}_j$ have multiplicity 1. Let the observation patterns in each group satisfy Condition 2. Then under the null hypothesis of equal covariance operators we obtain the following results:*

(i) *The test statistic S_{HS} is asymptotically distributed as $\sum_{k=1}^\infty \delta_k C_k$, where C_k are independent chi-square distributed variables with one degree of freedom and δ_k can be consistently estimated by the eigenvalues of the operator $\hat{\mathfrak{B}}$ given in (8).*

(ii) *The test statistic S_d is asymptotically chi-square distributed with $(K-1)d(d+1)/2$ degrees of freedom.*

The asymptotic distribution of S_{HS} can be approximated by simulation like in Boente et al. (2017). Section 4 presents a practical bootstrap implementation of these tests in which it is not necessary to compute the operator $\hat{\mathfrak{B}}$.

Tests based directly on covariance operators are not the only option. As an alternative we explore the approach of Pigoli et al. (2014) who argue that although covariance operators are contained in the Hilbert space of Hilbert–Schmidt operators, they do not form a linear subspace, and propose other distances than those based on the difference of covariances, such as the Procrustes distance and the square root distance. This direction of research was further investigated by Cabassi et al. (2017) and Masarotto (2019). One of the proposals of Pigoli et al. (2014) was to use the Hilbert–Schmidt distance between square root covariance operators $d_{\mathrm{sqrt}}(\mathscr{R}_1, \mathscr{R}_2) = \|\mathscr{R}_1^{1/2} - \mathscr{R}_2^{1/2}\|_2$. They report good power results for a two-sample test of equal covariances in the setting of complete functions based on this distance between estimated operators, $d_{\mathrm{sqrt}}(\hat{\mathscr{R}}_1, \hat{\mathscr{R}}_2)$. We extend this approach to K samples consisting of partially observed functions.

Since the data may contain incomplete functions, the empirical covariance operators $\hat{\mathscr{R}}_j$ used before may have negative eigenvalues. To be able to work with empirical square root covariance operators, we need to modify the covariance estimators to ensure they are non-negative definite. We use

$$\hat{\mathscr{R}}_{j+} = \sum_{l=1}^{n_j} (\hat{\lambda}_{jl})_+ \hat{\varphi}_{jl} \otimes \hat{\varphi}_{jl},$$

9

where $(\hat{\lambda}_{jl})_+ = \max(\hat{\lambda}_{jl}, 0)$ is the positive part of the eigenvalue $\hat{\lambda}_{jl}$ of $\hat{\mathscr{R}}_j$ and $\hat{\varphi}_{jl}$ is the corresponding eigenfunction. As discussed in Kraus (2015), negative eigenvalues are typically of small magnitude in comparison with leading eigenvalues and, therefore, are negligible in practice. For a test statistic, we need to use the distance d_{sqrt} to define a null estimator of $\mathscr{R}$ and contrasts between the group estimators $\hat{\mathscr{R}}_{j+}$ and the null estimator. The common covariance operator can be estimated by

$$\hat{\mathscr{R}}_{\text{sqrt}} = \left(\frac{\sum_{j=1}^{K} n_j \hat{\mathscr{R}}_{j+}^{1/2}}{\sum_{j=1}^{K} n_j} \right)^2,$$

which is the weighted Fréchet mean of the group-specific operators, i.e., the minimizer with respect to $\mathscr{R}$ of $\sum_{j=1}^{K} n_j d_{\text{sqrt}}(\hat{\mathscr{R}}_{j+}, \mathscr{R})^2$.

The attained minimum of this objective function,

$$S_{\text{sqrt}} = \sum_{j=1}^{K} n_j d_{\text{sqrt}}(\hat{\mathscr{R}}_{j+}, \hat{\mathscr{R}}_{\text{sqrt}})^2 = \sum_{j=1}^{K} \| n_j^{1/2} (\hat{\mathscr{R}}_{j+}^{1/2} - \hat{\mathscr{R}}_{\text{sqrt}}^{1/2}) \|_2^2, \tag{9}$$

can serve as a test statistic for comparing covariance operators in K samples. The statistic summarizes the size of the contrasts between the group and null estimators of the square root covariance operator. Following Pigoli et al. (2014) we use resampling to approximate the null distribution of the statistic.

Notice that the contrasts between the group and null estimators in S_{sqrt} and S_{HS} are weighted differently. In S_{HS} we weight the contrast kernels by $M_j(s,t)^{1/2}$ which in the fragmentary setting reflects the accuracy of the estimation of the covariance kernel at each point of $[0,1]^2$ due to the number of observations available at that point. In S_{sqrt} this would not be meaningful because the square root covariance operator is a function of the entire covariance operator and thus the accuracy of the estimation of the square root covariance kernel at one point depends also on the numbers of available observations at all other points. We therefore simply weight by $n_j^{1/2}$ reflecting the overall accuracy of the square root covariance estimator. Both S_{HS} and S_{sqrt} are the attained minimum of the corresponding objective functional that defines the null estimator.

4 Practical implementation and bootstrap approximations

Functional data procedures are practically implemented by discretization. Functional observations are evaluated at q points of a grid in the domain. Functions then correspond to q-vectors (possibly with missing values), operators on the function space correspond to $(q \times q)$-matrices and operators on operators correspond to four-way arrays with all dimensions q.

To make inference (tests and confidence intervals), one can use the asymptotic distributions found in the previous section. However, the implementation of such procedures would be excessively demanding in terms of computer memory, especially in the case of covariance inference. For example, when the evaluation grid consists of $q = 100$ points, arrays such as the one corresponding to the fourth moment kernel $\zeta(s,t,u,v)$ contain $q^4 = 10^8$ entries. To compare covariances, e.g., in $K = 3$ samples, one would have to work with an array with $K^2 q^4 = 9 \times 10^8$ entries whose size already approaches the memory limits of usual computers, even if symmetry is exploited. In the case of multivariate, spatial or image data the number of evaluation points q is typically much larger than for functions of a one-dimensional argument. Aston et al. (2017) give an example of acoustic phonetic data with bivariate, time-frequency argument with $q = 8100$. In conclusion, the size of objects representing the asymptotic covariance structure for tests or confidence intervals may be far beyond memory limits.

Projection covariance tests for complete functions can avoid the computation, storage and manipulation with such large arrays by computing principal scores of each function with respect to the required low number d of eigenfunctions (Horváth et al., 2010; Panaretos et al., 2010, 2011; Fremdt et al., 2013). The covariance matrix of the score then depends on easy-to-handle d-dimensional four-way arrays instead of large q-dimensional four-way arrays. This dimension reduction approach is not applicable in the case of incomplete functions because the principal scores $\langle X_{ji} - \hat{\mu}_j, \hat{\varphi}_m \rangle$ cannot be computed when X_{ji} is available only on a subset of its domain (they can only be predicted, see Kraus, 2015). Therefore, even the computation of the projection test statistic (6) is difficult due the large arrays the matrix $\hat{W}$ depends on.

The computation of the Hilbert–Schmidt statistic (4) and the square root covariance statistic (9) does not involve large four-way arrays. However, to use the asymptotic distribution of S_{HS} (see Theorem 5) one needs to estimate the eigenvalues of an operator on operators. Upon discretization and vectorization, this leads to a large eigenproblem of dimension $(Kq^2) \times (Kq^2)$, e.g., $30\,000 \times 30\,000$ for $K = 3$, $q = 100$. Again, dimension reduction cannot be used due to incomplete functions.

To overcome these difficulties we use the bootstrap. For completely observed functional data bootstrap tests of equal mean functions or covariance operators were studied by Benko et al. (2009) and Paparoditis and Sapatinas (2016b,a). In our missing data setting, all bootstrap procedures consist of appropriate resampling of fragmentary curves, which means that each bootstrap sample is again a collection of partially observed functions. The proposed procedures enable to completely avoid the computation of each entry of the large four-way covariance array and the storage and decomposition of the whole array.

The implementation of the tests of equal means is described in Algorithm 1. To correctly reproduce the limiting distribution of the group mean estimators under the null, the resampling is done separately in each group of groupwise centred fragmentary observations. The stratification guarantees that neither the missingness patterns nor distributional characteristics of the functions beyond the means need to be equal in all groups. The L^2 statistic is computed directly for each bootstrap sample and the observed value is then compared with the resampled values. The direct computation of the projection test statistic from observed or resampled data would require the estimation of the covariance functions $\hat{v}_{jk}$ in (3), which may be memory demanding and possibly unstable in regions with few complete pairs. We avoid it by estimating the covariance matrix of the score vector from the resampled score vectors, calculating the quadratic form statistic using the observed score vector and the bootstrap estimate of its covariance matrix, and comparing it with its asymptotic chi-square distribution.

Algorithm 1 Bootstrap approximation for tests of equal mean functions
---
1: Calculate $\hat{\mu}_j$ from observed samples of fragments $X_{j1}, \ldots, X_{jn_j}$, $j = 1, \ldots, K$, and $\hat{\mu}$
2: Calculate the test statistic T_{L^2} and the score vector Q
3: Set $X_{ji0} = X_{ji} - \hat{\mu}_j + \hat{\mu}$
4: For $b = 1, \ldots, B$
5:     For each $j = 1, \ldots, K$, sample with replacement from fragments $X_{j10}, \ldots, X_{jn_j0}$
       to get fragments $X_{j10}^*, \ldots, X_{jn_j0}^*$
6:     Calculate the statistic $T_{L^2}^{*(b)}$ and score vector $Q^{*(b)}$ from $X_{j10}^*, \ldots, X_{jn_j0}^*$, $j = 1, \ldots, K$
7: Approximate the p-value of the L^2-test using T_{L^2} and $T_{L^2}^{*(1)}, \ldots, T_{L^2}^{*(B)}$
8: Calculate the empirical covariance matrix $\hat{V}^*$ of $Q^{*(1)}, \ldots, Q^{*(B)}$ and the statistic $T_d = Q^\top \hat{V}^{*-} Q$
9: Approximate the p-value of the projection test using T_d and the $\chi^2_{(K-1)d}$ distribution

---

Algorithm 2 describes the bootstrap implementation of confidence intervals for eigenelements. Resampling is applied to fragments and eigenelements are computed. The resampled eigenfunction is possibly reflected about zero so that its sign agrees with that of the observed data empirical eigenfunction. Standard methods of construction of confidence intervals can then be used. Since we again wish to avoid the calculation of variance estimates of eigenelements (see Theorem 4), we use the normal or basic bootstrap method (Davison and Hinkley, 1997, Chapter 5). Intervals for eigenvalues are constructed on the logarithmic scale and untransformed. This is appropriate in general because in the case of completely observed Gaussian curves the asymptotic variance of $n^{1/2}(\hat{\lambda}_m - \lambda_m)$ is $2\lambda_m^2$ and thus the log-transformation approximately stabilizes variance.

Bootstrap covariance testing is described in Algorithm 3. Unlike in the case of mean testing, it is not possible to transform the data to the common null covariance structure and use stratified resampling. Bootstrap samples are instead drawn from the pooled sample of groupwise centred fragments, similarly to Paparoditis and Sapatinas (2016b, Subsection 2.2) for complete curves. Then, under the null hypothesis, if characteristics of observation patterns (θ_j) and fourth order moments (ζ_j) are the same in all groups, the pooled resampling asymptotically replicates the limiting distributions of interest. The Hilbert–Schmidt norm and square root covariance statistics are computed directly and the significance is decided upon by comparing the observed statistics with the resampled ones. Like in the case of mean testing, dimension reduction is impossible due to partial observation, and thus the computation of the covariance matrix of

Algorithm 2 Bootstrap confidence intervals for eigenvalues and eigenfunctions

1: Calculate $\hat{\hat{\mathscr{R}}}$ from the observed fragmentary functional data $X_1, \ldots, X_n$
2: Calculate the eigenvalues $\hat{\lambda}_m$ and eigenfunctions $\hat{\varphi}_m$ of $\hat{\hat{\mathscr{R}}}$
3: For $b = 1, \ldots, B$
4:    Sample with replacement from fragments $X_1, \ldots, X_n$ to get fragments $X_1^*, \ldots, X_n^*$
5:    Calculate $\hat{\hat{\mathscr{R}}}^*$ from $X_1^*, \ldots, X_n^*$ and its eigenvalues $\hat{\lambda}_m^{*(b)}$ and eigenfunctions $\hat{\varphi}_m^{*(b)}$
6:    Replace $\hat{\varphi}_m^{*(b)}$ by $\text{sign}\langle \hat{\varphi}_m^{*(b)}, \hat{\varphi}_m \rangle \hat{\varphi}_m^{*(b)}$
7: Based on $\hat{\lambda}_m^{*(b)}$, $\hat{\varphi}_m^{*(b)}$, $b = 1, \ldots, B$, calculate bootstrap confidence intervals for λ_m using log-transformation and pointwise bootstrap confidence intervals for $\varphi_m(t)$

the score vector would require to compute large four-way arrays. Instead, the boostrap is used to estimate the covariance matrix of the score and the quadratic statistic with this matrix is used.

Algorithm 3 Bootstrap approximation for tests of equal covariance operators

1: Calculate $\hat{\mu}_j$ and $\hat{\hat{\mathscr{R}}}_j$ from observed samples of fragments $X_{j1}, \ldots, X_{jn_j}$, $j = 1, \ldots, K$, and $\hat{\hat{\mathscr{R}}}$
2: Perform eigendecomposition of $\hat{\hat{\mathscr{R}}}$, determine d and calculate $\hat{\mathscr{U}}_{lm}$, $1 \leq l \leq m \leq d$
3: Calculate the test statistics S_{HS} and S_{sqrt} and the score vector R with respect to $\hat{\mathscr{U}}_{jm}$
4: Set $X_{ji0} = X_{ji} - \hat{\mu}_j$
5: For $b = 1, \ldots, B$
6:    For each $j = 1, \ldots, K$, sample with replacement from the pooled collection of fragments X_{ji0}, $j = 1, \ldots, K$, $i = 1, \ldots, n_j$ to get fragments $X_{j10}^*, \ldots, X_{jn_j0}^*$
7:    Calculate the statistics $S_{\text{HS}}^{*(b)}$ and $S_{\text{sqrt}}^{*(b)}$ and the score vector $R^{*(b)}$ with respect to $\hat{\mathscr{U}}_{jm}$ from $X_{j10}^*, \ldots, X_{jn_j0}^*$, $j = 1, \ldots, K$
8: Approximate the p-value of the Hilbert–Schmidt norm test using S_{HS} and $S_{\text{HS}}^{*(1)}, \ldots, S_{\text{HS}}^{*(B)}$ and the p-value of the square root covariance test using S_{sqrt} and $S_{\text{sqrt}}^{*(1)}, \ldots, S_{\text{sqrt}}^{*(B)}$
9: Calculate the empirical covariance matrix $\hat{W}^*$ of $R^{*(1)}, \ldots, R^{*(B)}$ and the statistic $S_d = R^\top \hat{W}^{*-} R$
10: Approximate the p-value of the projection test using S_d and the $\chi^2_{(K-1)d(d+1)/2}$ distribution

While we do not provide formal proofs of the validity of the bootstrap approximations, these could be obtained along the lines of the proofs in Paparoditis and Sapatinas (2016a) and Paparodits and Sapatinas (2016b) using our asymptotic results (Theorems 1–5). Note that in our setting the observation sets might be non-identically distributed (e.g., in the case of designed experiments), and hence the bootstrap is applied to possibly non-identically distributed observed fragments. Their average characteristics, however, converge under Conditions 1 and 2. It is possible to use the bootstrap even with mildly non-identically distributed data, as discussed in the general context by Liu (1988) who shows that if average moment characteristics of possibly non-identically distributed variables converge, the bootstrap is still applicable.

The use of the bootstrap for the square root covariance test is based on empirical evidence from simulation studies (Section 5 and the Supplementary Material). Its theoretical justification would require to first establish the asymptotic distribution of the estimated square root covariance operator, which is not available even in the case of completely observed curves (Pigoli et al., 2014).

5 Simulation results

The main goal of the study is to investigate the impact of partial observation on the performance of the different mean and covariance tests and compare the proposed tests using complete and incomplete curves with the simple approach using complete curves only.

We repeatedly generate three samples of curves of sizes $n_1 = 80$, $n_2 = 100$, $n_3 = 120$. Curves in the jth sample take the form

$$X(t) = \mu_j(t) + \lambda_{j0}^{1/2} \beta_{j0} h_j(t) + \sum_{k=1}^{20} \lambda_{jk}^{1/2} \beta_{jk} 2^{1/2} \cos(k\pi t), \quad t \in [0, 1],$$

where β_{jk}, $j \in \{1, 2, 3\}$, $k \in \{0, \dots, 20\}$ are mutually independent standard normal variables. Additional simulations with t_5 distributed coefficients are reported in the supplementary material. In all simulations we use 1000 repetitions of the test procedures, each based on 500 bootstrap samples. All tests are performed on the nominal level of 5 %. All results have been computed in R 3.4.

The tests are applied to complete trajectories, observation pattern (1), and to fragments obtained by deleting missing periods following several random or nonrandom patterns. Observation patterns (2) and (3) are nonrandom: under pattern (2), the period $[0, 0.5]$ is removed from 50 % of the curves in the first sample, 50 % in the second sample and 60 % in the third sample; pattern (3) is symmetric about 0.5, i.e., the period $[0.5, 1]$ instead of $[0, 0.5]$ is missing in the same subset of curves. Under patterns (4)–(7), a random missing period is generated independently for each curve and removed from the trajectory. First, we consider random missing periods taking the form $M = [C - E, C + E] \cap [0, 1]$ with $C = dU_1^{1/2}$ and $E = fU_2$, where U_1, U_2 are independent variables uniformly distributed on $[0, 1]$ and d, f are parameters. For missingness pattern (4) we set $d = 1.4$ and $f = 0.2$; this gives 39 % of completely observed curves and the cross-sectional percentage of observed values decreases from 99 % at time 0 to 79 % at time 1. Pattern (5) is symmetric about 0.5. For pattern (6) we use the same model as for (4) and set $d = 1.2$ and $f = 0.5$; this leads to 7 % of complete curves and the cross-sectional probability of observation is 94 % at 0 and decreases to about 45 % near 1. Pattern (7) is again obtained by reflecting pattern (6) about 0.5. Pattern (8) consists of observation periods generated independently for each curve in the form $O = [U_1, U_2] \cap [0, 1]$, where U_1, U_1 are independent variables uniformly distributed on $[a, C]$, $[C, 1 - a]$, respectively, $a = -0.3$ and C is uniformly distributed on $[0, 1]$; the percentage of complete curves in this case is 16 % and the cross-sectional observation probability at 0.5 is 77 % and decreases to 44 % towards both endpoints of the domain. Finally, for pattern (9) curves are observed on random intervals generated as $[C - 0.2, C + 0.2] \cap [0, 1]$, where C is uniformly distributed in $[0, 1]$. This corresponds to fragments of curves of length at most 0.4, hence the datasets contain no complete curves, the median length of observed fragments is 0.3 and the cross-sectional probability of observation is 0.3 in the middle of the domain and decreases towards the endpoints, where it is 0.15.

In the study of mean tests four configurations of the mean functions are considered. Under configuration A the null hypothesis is satisfied: all mean functions are zero. Under configuration B the mean functions differ by a constant vertical shift: $\mu_1(t) = 0$, $\mu_2(t) = 0.18$, $\mu_3(t) = -0.1$. Under configuration C there are monotonic differences between the means: $\mu_1(t) = 0$, $\mu_2(t) = 0.35 \exp(-4t)$, $\mu_3(t) = -0.25 \exp(-3t)$. Under configuration D the means differ in a more complex, nonmonotonic way and they cross: $\mu_1(t) = 0$, $\mu_2(t) = 2t \exp(-3t)$, $\mu_3(t) = 0.1 - 8t^2 \exp(-5t)$. We set $\lambda_{j0} = 0.5$, $\lambda_{jk} = 3^{-k}$ and $h_j(t) = 1$, that is, the covariance structure is the same in all three groups. Additional simulations with unequal covariance structures lead to similar results and are included in the Supplementary Material. We report in the first part of Table 1 the size and power of the L^2 test based on T_{L^2} given in (1) and of the projection test based on T_d given in (2) using $d = 3$ Legendre polynomials of order zero, one and two. Blank entries in the table correspond to situations where the true rejection probability is the same as in the entry above; such situations arise when the observation pattern is obtained by reflecting the preceding pattern and the processes $\{X(t) : t \in [0, 1]\}$ and the time-reversed processes $\{X(1 - t) : t \in [0, 1]\}$ have the same distribution.

We see in the first part of Table 1 that under the null hypothesis, configuration A, the rejection probability of the L^2 tests is close to the nominal level. The size of the projection test seems to be somewhat above the nominal level due to the sample size, especially under observation pattern (9), where the missingness rate is the highest. Our simulation study of power provides raw rejection probabilities in Table 1 and size-adjusted powers (using the method from Subsection 3.2 of Lloyd (2005)) in Table S2 in the Supplementary Material. The possibility of size issues should be kept in mind in applications: especially in marginal cases, users should not simply compare p-values with a single threshold but rather carefully report them.

Under scenario B the L^2 test is more powerful than the projection method. The reason is that the projection method uses in addition to the constant basis function two other terms (linear and quadratic) that do not contribute to the detection of the constant difference between the means but on the other hand they increase the degrees of freedom and hence decrease the power. The L^2 method uses infinitely many directions in the space of alternatives but these redundant features are downweighted by the decreasing eigenvalues (the constant difference of means agrees with the constant leading eigenfunction which receives the highest weight in the L^2 statistic). Most partial observation patterns lead to a relatively small decrease of power because under this scenario the mean functions differ by a constant vertical shift which is a very

Table 1
Empirical rejection probability (in %) of the L^2 test, T_{L^2}, and projection test, T_d, of equal means. A dash indicates the same value as on the preceding row. The observation patterns (1)–(9) and mean configurations A–D are described in the text.

Observation pattern	Mean configuration							
	A		B		C		D	
	T_{L^2}	T_d	T_{L^2}	T_d	T_{L^2}	T_d	T_{L^2}	T_d
	Tests using complete and incomplete curves (proposed approach)							
(1)	5.6	6.2	69	60	49	56	52	63
(2)	5.4	6.7	59	52	28	29	38	50
(3)	—	—	—	—	50	56	44	62
(4)	4.4	6.5	66	58	51	57	51	62
(5)	—	—	—	—	44	49	50	58
(6)	5.4	7.1	58	51	50	55	42	49
(7)	—	—	—	—	28	34	37	42
(8)	5.4	5.8	55	47	34	37	42	48
(9)	5.4	7.8	37	40	20	23	26	34
	Tests using complete curves only (simple approach)							
(2), (3)	5.7	7.4	40	34	26	32	27	35
(4), (5)	3.6	7.4	28	27	18	26	19	28
(6), (7)	4.9	26.8	7	31	6	29	6	31
(8)	4.0	11.5	13	22	8	20	10	21

simple, global feature that is easily detected even with reduced, fragmented data. The loss of power is largest under pattern (9), where also the reduction of observed data is considerably larger than under the other patterns.

Both tests have comparable power under scenario C. Both tests lose power under observation pattern (2) because a large portion of data is missing on the interval $[0, 0.5]$, where the difference between the means is the largest; on the other hand, the reflected pattern (3) does not lead to a loss of power because curves are missing only in $[0.5, 1]$, where the means do not differ much. A similar effect is seen under observation patterns (6) and (7).

Under scenario D the projection test seems to be slightly more powerful than the L^2 (even after the size adjustment in Table S2 in the Supplementary Material) because the nonmonotonic differences between the mean functions are well captured by both the first three Legendre polynomials and the first three eigenfunctions but the contribution of the latter is downweighted in the L^2 statistic whereas the projection statistic treats all three components equally.

The second part of Table 1 shows for each missingness pattern and mean configuration the performance of the tests applied to the subset of complete curves only. The complete curve approach would be the only possibility if the tests developed in this paper were not available. Results for the pairs of patterns (2) and (3), (4) and (5), (6) and (7) are presented on the same rows of the second part of the table because the subsets of complete curves are the same under both patterns in each pair. Pattern (9) is omitted because it contains no complete curves and hence inference is impossible without our methods. Under patterns (2) (or (3)) and (4) (or (5)), the use of complete curves only, which form 46% and 39%, respectively, of the whole sample, leads to a considerable loss of power in most situations. Configuration C under pattern (2) is an exception. Here removing incomplete curves does not decrease the power because they are observed on the subdomain $[0.5, 1]$, where the means do not differ much. Under patterns (6) (or (7)) and (8) there are only 7% and 16% complete curves, respectively. With such small sample sizes the projection test becomes unreliable in terms of level and the L^2 test loses almost all power.

Next, we study the behaviour of the tests for comparing covariance operators. Under all scenarios we generate mean zero trajectories. Configuration A satisfies the null hypothesis with $\lambda_{j0} = 0.5$, $\lambda_{jk} = 3^{-k}$ and $h_j(t) = 1$, $j \in \{1, 2, 3\}$. Under configuration B the same parameters are used except for the third sample where the overall scale is larger, namely $\lambda_{3,0} = 1.5 \times 0.5$ and $\lambda_{3,k} = 1.5 \times 3^{-k}$. Under scenario C the first two eigenvalues in the third sample are interchanged, i.e., $\lambda_{3,0} = 3^{-1}$, $\lambda_{3,1} = 0.5$ and $\lambda_{3,k} = 3^{-k}$, $k \in \{2, \ldots, 20\}$, otherwise the parameters are the same as in A. Scenario D differs from A in that we set $h_3(t) = 1$ for $t \in [0, 0.5]$ and $h_3(t) = 2.2^{1/2}$ for $t \in (0.5, 1]$. Table 2 shows the size and power of the

14

Table 2
Empirical rejection probability (in %) of the Hilbert–Schmidt norm test, S_{HS}, projection test, S_d, and square root covariance test, S_{sqrt}, of equal covariance operators. A dash indicates the same value as on the preceding row. The observation patterns (1)–(5) and covariance configurations A–D are described in the text.

Observation pattern	Covariance configuration											
	A			B			C			D		
	S_{HS}	S_d	S_{sqrt}	S_{HS}	S_d	S_{sqrt}	S_{HS}	S_d	S_{sqrt}	S_{HS}	S_d	S_{sqrt}
	Tests using complete and incomplete curves (proposed approach)											
(1)	5.4	5.8	4.8	69	82	80	69	58	69	78	62	81
(2)	4.6	6.4	4.9	54	63	41	37	32	38	76	64	54
(3)	—	—	—	—	—	—	—	—	—	46	30	48
(4)	5.0	5.1	5.8	64	74	72	61	53	62	72	56	73
(5)	—	—	—	—	—	—	—	—	—	77	60	77
	Tests using complete curves only (simple approach)											
(2), (3)	4.1	7.3	4.6	32	38	41	33	28	34	45	30	47
(4), (5)	4.3	5.5	4.2	26	32	33	25	24	28	34	23	36

Hilbert–Schmidt norm test based on S_{HS} in (4), projection test based on S_d in (6) with d selected to explain at least 85 % of the total variability of the null covariance estimate, and square root covariance test based on S_{sqrt} in (9). Like before, entries where the true rejection probability equals the one above are left blank. We use only observation patterns (1)–(5). Under the other patterns the amount of missing information is too large for second order inference.

Under the null hypothesis, configuration A, the first part of Table 2 shows that the rejection probability of all tests is close to the nominal level under all missingness patterns, with the projection test being slightly above the level in some cases.

It is interesting to notice the different impact of missingness on the power in different situations. We report raw power in Table 2 and size-adjusted power in Table S4 in the Supplementary Material. While in many situations the loss of power due to missingness is similar for all three tests, in some situations the square root test appears to be more sensitive to missingness. For example under scenario B and missingness pattern (2), the square root covariance test loses almost half of its power relative to no missingness, much more than the other two tests. This can be explained by the fact that the square root covariance estimator depends on the estimator of the covariance kernel at all arguments which means that uncertainty due to missingness localized in a certain region in the domain, like under pattern (2), propagates. Similarly, under scenario D and pattern (2) the Hilbert–Schmidt and projection tests do not lose much power and the square root test does because the difference between the covariances is due to the differences of $h_j(t)$ for $t \in [0.5, 1]$ while missingness occurs for $t \in [0, 0.5]$. For these reasons, under the same scenario, pattern (3) leads to a larger loss of power than pattern (2) for the Hibert–Schmidt and projection tests, whereas the loss of the square root covariance is not much higher than under pattern (2), where it was already high.

The second part of Table 2 shows results for tests applied to the subset of complete curves only. Like before, patterns (3) and (5) are shown on the same rows as patterns (2) and (4), respectively, because the subsets of complete curves are the same. We observe a large decrease of power in comparison with the power of the proposed tests in cases, where the neglected incomplete curves carry information on the difference between covariance operators. When the difference is mostly in the frequently missing region (e.g., configuration D, pattern (3)), removing incomplete curves affects the power much less.

These results highlight the usefulness of the proposed methods as an efficient, and often the only viable approach to testing with incomplete functions. In no situation the proposed methods behaved worse than the simple approach using complete curves only, and in many cases it behaved dramatically better. Additional results for non-Gaussian curves can be found in the Supplementary Material.

6 Application to partially observed heart rate temporal profiles

We illustrate our methods on curves describing the evolution of heart rate in 427 male participants in the period from 8 PM to 2 AM corresponding to the domain $[20, 26]$. The data come from the Swiss Kidney
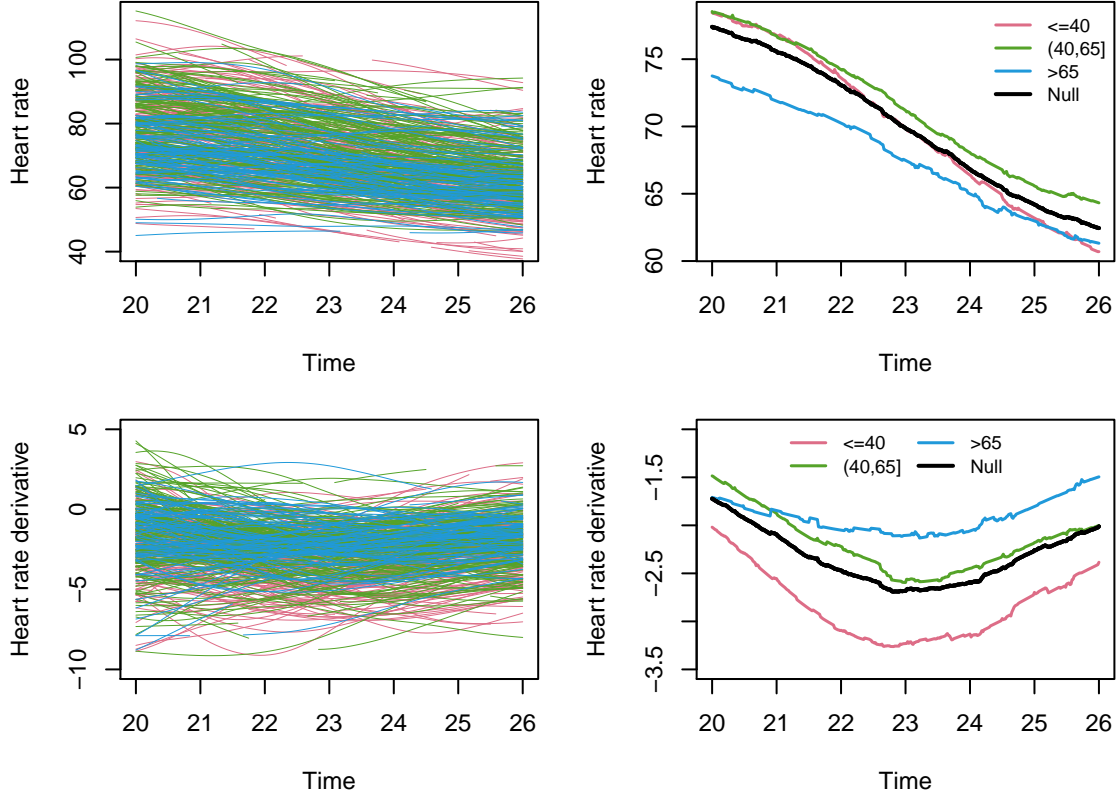
Fig. 1. Individual heart rate profiles and their first derivative (left panels) and the corresponding group-specific and null estimates of the mean (right panels).

Project on Genes in Hypertension. There are three groups of persons according to their age: younger than 40 years (164 persons), between 40 and 65 (180), and older than 65 (83). The curves and their first derivative are plotted in Fig. 1. Although the percentage of observed values at each time or at each pair of time points is relatively high (Fig. 2), only 58 % of the curves are complete.

Plots of the estimated mean functions in Fig. 1 indicate differences between the age groups both in terms the temporal profiles and their first derivative. We first compare the group means of heart rate profiles. The p-values of the L^2 test and projection test using three Legendre polynomials are 0.006 and less than 0.001, respectively, confirming the clearly visible differences. To compare the dynamics of heart rate during the transition between day and night we test whether the means of the first derivative differ. The L^2 and projection test have nearly zero p-values, meaning that the mean heart rate profiles differ between age groups more than by a vertical shift. The plots suggest it may be interesting to compare some pairs of groups. E.g., while the mean profiles of the middle and oldest group significantly differ ($p < 0.01$ for both tests), they appear to be approximately parallel. The difference between the derivatives is indeed insignificant ($p = 0.07$ for the L^2 test, $p = 0.09$ for the projection test).

Without the methods developed in this paper one would have to use complete curves only. There are 249 complete functions (43, 110 and 96 in the three age groups). The projection test still detects the differences between the three groups ($p = 0.008$) but the L^2 test loses significance ($p = 0.066$). When comparing the second and third group, the projection test now fails to detect the difference ($p = 0.13$) and the L^2 test gives a marginally significant result ($p = 0.048$). This can be explained by a loss of power seen in simulations because the removed incomplete curves are more often observed at earlier times, where also the difference between the two mean curves is more pronounced.

Estimates of the covariance function of heart rate profiles and of their derivatives for each age group are plotted in Fig. 3. Further plots can be found in the supplementary document. The plots suggest some differences between the groups. The variance and covariance appears to be higher in younger participants, especially earlier in the time interval (during the day). We assess the significance of these differences using
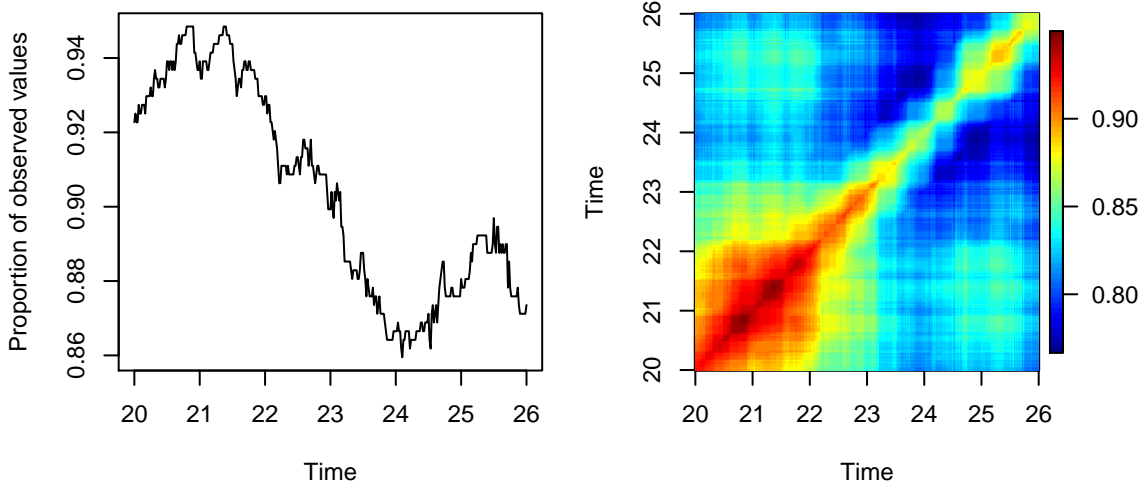
16

Fig. 2. Cross-sectional percentage of observed values (left) and percentage of pairwise complete observations (right).

Table 3

p-values of the Hilbert–Schmidt norm test, S_{HS}, the square root covariance test, S_{sqrt}, and the projection tests, S_d, with $d = 1, 2, 3$, for comparing covariance structures of heart rate profiles and of their first derivative in three age groups. The fraction of variance explained by the first d principal components of the null covariance estimate is indicated in parentheses.

	S_{HS}	S_{sqrt}	S_1	S_2	S_3
Curves	0.338	0.118	0.317 (88.2 %)	0.439 (97.3 %)	0.275 (99.1 %)
First derivative	0.226	0.114	0.322 (62.6 %)	0.131 (94.4 %)	0.094 (98.7 %)

the proposed tests. For the projection test we consider up to three principal components (plotted in the supplementary document), which corresponds to the projection on a subspace of dimension six in the space covariance operators. Table 3 reports the p-values. None of the tests rejects the null hypothesis on usual significance levels. Similarly, pairwise comparisons provided no overwhelming evidence of differences. It is of course possible that there are differences between groups that may be detected with larger samples. To gain further insight into the structure of possible differences one can inspect the values of the standardized score components $R_{jlm}/\hat{W}_{jlm,jlm}^{1/2}$ (see (5) and (7)) whose graphical representation is provided in the supplement.

Supplementary material

The supplementary document available online contains further simulation results and additional graphs for the data application. R code is available online.
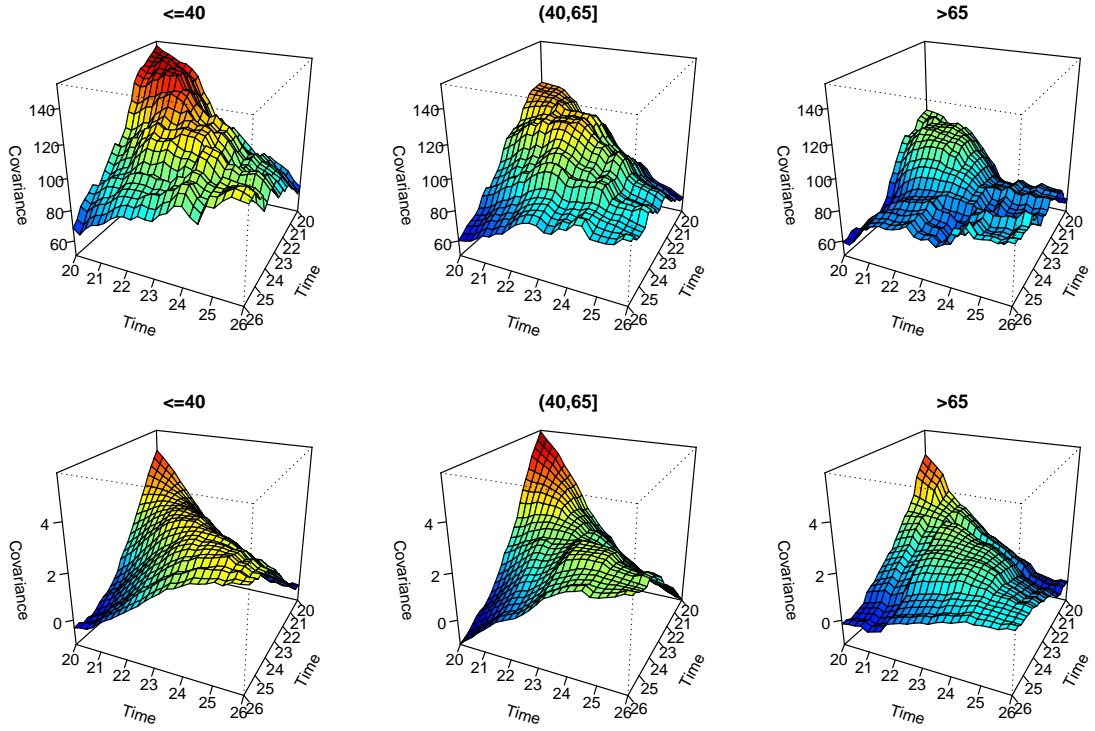
Acknowledgements

Fig. 3. Estimated covariance functions of heart rate profiles (top row) and of their derivatives (bottom row) in age groups.
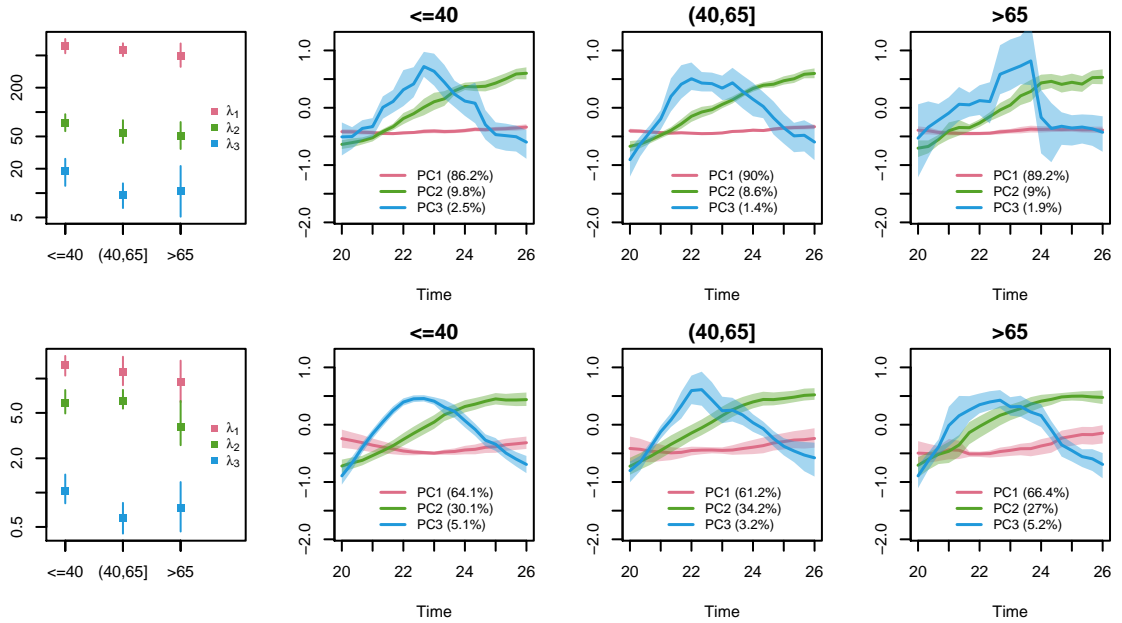

Fig. 4. Estimated eigenvalues and eigenfunctions of heart rate profiles (top row) and of their derivatives (bottom row) in age groups with pointwise 95 % bootstrap confidence intervals.

A   A central limit theorem

We provide a general central limit theorem for independent but not necessarily identically distributed random elements of a separable Hilbert space. It is needed in the proofs, where non-identical distributions arise due to partial observation, but is of more general interest. It extends the standard result for independent identically distributed functional variables (Bosq, 2000, Theorem 2.7) by relaxing the assumption of identical distributions and by considering triangular arrays. The notation $\|\cdot\|_\infty$ below means the operator norm.

Theorem 6. *Let Y_{ni}, $n \in \{1, 2, \dots\}$, $i \in \{1, \dots, n\}$ be random elements of a separable Hilbert space $\mathcal{H}$ with mean zero, $\mathrm{E}\,\|Y_{ni}\|^2 < \infty$ and covariance operators $\mathscr{C}_{ni}$. Let $Y_{n1}, \dots, Y_{nn}$ be mutually independent for each $n \in \{1, 2, \dots\}$. Denote $S_n = n^{-1/2} \sum_{i=1}^n Y_{ni}$ and $\mathscr{G}_n = n^{-1} \sum_{i=1}^n \mathscr{C}_{ni}$. Assume that*

(i) $\|\mathscr{G}_n - \mathscr{G}\|_\infty \to 0$ as $n \to \infty$ for some covariance operator $\mathscr{G}$,

(ii) for all $\varepsilon > 0$,

$$n^{-1} \sum_{i=1}^n \mathrm{E}(\|Y_{ni}\|^2 1_{[\|Y_{ni}\| > n^{1/2} \|\mathscr{G}_n\|_\infty \varepsilon]}) \to 0$$

as $n \to \infty$,

(iii) $\operatorname{tr} \mathscr{G}_n \to \operatorname{tr} \mathscr{G}$ as $n \to \infty$.

Then S_n converges in distribution to a Gaussian random element with mean zero and covariance operator $\mathscr{G}$.

B   Proofs

Proof of Theorem 1

We rewrite $N^{1/2}(\hat{\mu} - \mu) = \hat{\pi}^{1/2} n^{1/2}(\hat{\mu} - \mu)$. The main task is to establish the weak convergence of the process

$$n^{1/2}(\hat{\mu} - \mu) = \frac{1}{\pi} S_n + \left(\frac{J}{\hat{\pi}} - \frac{1}{\pi}\right) S_n + n^{1/2}(J - 1)\mu, \tag{10}$$

where $S_n = n^{-1/2} \sum_{i=1}^n O_i(X_i - \mu)$. We show that the first term on the right side of (10) converges in distribution to a mean zero Gaussian process with covariance operator with kernel $\pi(s)^{-1}\pi(t)^{-1}\nu(s,t)\rho(s,t)$ that can be consistently estimated by $\hat{\pi}(s)^{-1}\hat{\pi}(t)^{-1}\hat{\nu}(s,t)\hat{\rho}(s,t)$, and that the norms of the other two terms converge in probability to 0. The proof of the weak convergence of $N^{1/2}(\hat{\mu} - \mu)$ then follows from the convergence of $\hat{\pi}$ to π, the consistency of the estimator of its covariance kernel can be shown analogously.

The weak convergence of S_n is shown with the help of Theorem 6, a central limit theorem for independent non-identically distributed Hilbert space variables given in the Appendix. We apply the theorem with $Y_{ni} = O_i(X_i - \mu)$. The covariance operator $\mathscr{G}_n$ of S_n is given by the kernel $\bar{\nu}(s,t)\rho(s,t)$. Denote by $\mathscr{G}$ the covariance operator with kernel $\nu(s,t)\rho(s,t)$. Conditions of the central limit theorem Theorem 6 can be shown using Condition 1(b) as follows. Condition (i) of Theorem 6 is satisfied because

$$\|\mathscr{G}_n - \mathscr{G}\|_\infty^2 \le \|\mathscr{G}_n - \mathscr{G}\|_2^2 = \int_{[0,1]^2} \{\bar{\nu}(s,t) - \nu(s,t)\}^2 \rho(s,t)^2 ds\, dt \to 0$$

as $n \to \infty$ by the dominated convergence theorem. Condition (ii) of Theorem 6 holds because

$$n^{-1} \sum_{i=1}^n \mathrm{E}(\|Y_{ni}\|^2 1_{[\|Y_{ni}\| > n^{1/2}\|\mathscr{G}_n\|_\infty \varepsilon]}) \le n^{-1} \sum_{i=1}^n \mathrm{E}(\|X_i - \mu\|^2 1_{[\|X_i - \mu\| > n^{1/2}\|\mathscr{G}_n\|_\infty \varepsilon]})$$

$$= \mathrm{E}(\|X_1 - \mu\|^2 1_{[\|X_1 - \mu\| > n^{1/2}\|\mathscr{G}_n\|_\infty \varepsilon]}),$$

which converges to 0 by the dominated convergence theorem. Finally, $\int_0^1 \bar{\nu}(t,t)\rho(t,t)dt \to \int_0^1 \nu(t,t)\rho(t,t)dt$ by the dominated convergence theorem again, and thus condition (iii) of Theorem 6 is satisfied. Hence the process S_n is asymptotically Gaussian with covariance kernel $\nu(s,t)\rho(s,t)$.

The expectation of the squared norm of the second term on the right side of (10) can be rewritten as

$$\int_0^1 \mathrm{E}\left[\left\{\frac{J(t)}{\hat{\pi}(t)} - \frac{1}{\pi(t)}\right\}^2 S_n(t)^2 1_{[\hat{\pi}(t) \geq \pi_0/2]}\right]dt + \int_0^1 \mathrm{E}\left[\left\{\frac{J(t)}{\hat{\pi}(t)} - \frac{1}{\pi(t)}\right\}^2 S_n(t)^2 1_{[\hat{\pi}(t) < \pi_0/2]}\right]dt. \quad (11)$$

The first summand above is dominated by

$$\int_0^1 \mathrm{E}\left[\frac{\{\pi(t) - \hat{\pi}(t)\}^2}{\pi_0^4/4} S_n(t)^2\right]dt \leq \int_0^1 \mathrm{E}\left[\frac{\{\pi(t) - \hat{\pi}(t)\}^2}{\pi_0^4/4}\right]\rho(t,t)dt$$

which converges to zero by the dominated convergence theorem since $\mathrm{E}[\{\pi(t) - \hat{\pi}(t)\}^2] = \{\pi(t) - \bar{\pi}(t)\}^2 + n^{-2}\sum_{i=1}^n \pi_i(t)\{1 - \pi_i(t)\} \to 0$ for $n \to \infty$. Next, we first compute

$$\left\{\frac{J(t)}{\hat{\pi}(t)} - \frac{1}{\pi(t)}\right\}^2 1_{[\hat{\pi}(t) < \pi_0/2]} = \left[J(t)\left\{\frac{\pi(t) - \hat{\pi}(t)}{\hat{\pi}(t)\pi(t)}\right\}^2 + \{1 - J(t)\}\frac{1}{\pi(t)^2}\right]1_{[\hat{\pi}(t) < \pi_0/2]}$$

$$\leq [J(t)n^2/\pi_0^2 + \{1 - J(t)\}/\pi_0^2]1_{[\hat{\pi}(t) < \pi_0/2]} \leq n^2/\pi_0^2 1_{[\hat{\pi}(t) < \pi_0/2]}.$$

Then the second summand in (11) is smaller than or equal to

$$\int_0^1 \mathrm{E}\{n^2/\pi_0^2 1_{[\hat{\pi}(t) < \pi_0/2]} S_n(t)^2\}dt \leq \int_0^1 n^2/\pi_0^2 \Pr\{\hat{\pi}(t) < \pi_0/2\}\rho(t,t)dt \leq n^2 \sup_{t \in [0,1]} \Pr\{\hat{\pi}(t) < \pi_0/2\}/\pi_0^2 \operatorname{tr}\mathscr{R},$$

which converges to 0 because, in light of Hoeffding's inequality and Condition 1(a), for all $t \in [0,1]$,

$$\Pr\{\hat{\pi}(t) < \pi_0/2\} \leq \exp[-2n\{\bar{\pi}(t) - \pi_0/2\}^2] \leq \exp\left[-2n\left\{\pi_0/2 - \sup_{t \in [0,1]}|\bar{\pi}(t) - \pi(t)|\right\}^2\right] \to 0.$$

This completes the proof of the convergence in probability of the norm of the second term on the right hand side of (10) to zero. The last term in (10) can be shown to converge to zero using similar arguments based on Hoeffding's inequality.

We now turn to the proof of the consistency of the estimator of the covariance kernel. To show that

$$\mathrm{E}\int_{[0,1]^2}\left\{\frac{\hat{\nu}(s,t)\hat{\rho}(s,t)}{\hat{\pi}(s)\hat{\pi}(t)} - \frac{\nu(s,t)\rho(s,t)}{\pi(s)\pi(t)}\right\}^2 dsdt \to 0,$$

we can split the integral into the integrals over $A_0 = \{(s,t) \in [0,1]^2 : \nu(s,t) = 0\}$ and $A_1 = \{(s,t) \in [0,1]^2 : \nu(s,t) \geq \nu_0\}$ because Condition 1(c) implies that $A_0 \cup A_1 = [0,1]^2$. On A_0 we obtain

$$\mathrm{E}\int_{A_0}\left\{\frac{\hat{\nu}(s,t)\hat{\rho}(s,t)}{\hat{\pi}(s)\hat{\pi}(t)}\right\}^2 \{1_{[\min(\hat{\pi}(s),\hat{\pi}(t)) \geq \pi_0/2]} + 1_{[\min\{\hat{\pi}(s),\hat{\pi}(t)\} < \pi_0/2]}\}dsdt$$

$$\leq \int_{A_0}\mathrm{E}\{\hat{\nu}(s,t)^2\}\mathrm{E}\{\hat{\rho}(s,t)^2\}dsdt\left((\pi_0/2)^{-4} + n^4 \sup_{(s,t) \in [0,1]^2}\Pr[\min\{\hat{\pi}(s),\hat{\pi}(t)\} < \pi_0/2]\right).$$

Here the integral converges to zero by the dominated convergence theorem as the integrand can be shown to go to 0 and the second term in the brackets asymptotically vanishes due to an exponential rate of decrease of the supremum that can be established with the help of Hoeffding's inequality as before, hence the whole quantity above converges to 0. We now focus on A_1. We rewrite

$$\frac{\hat{\nu}(s,t)\hat{\rho}(s,t)}{\hat{\pi}(s)\hat{\pi}(t)} - \frac{\nu(s,t)\rho(s,t)}{\pi(s)\pi(t)} = \frac{\hat{\nu}(s,t)}{\hat{\pi}(s)\hat{\pi}(t)}\{\hat{\rho}(s,t) - \rho(s,t)\} + \left\{\frac{\hat{\nu}(s,t)}{\hat{\pi}(s)\hat{\pi}(t)} - \frac{\nu(s,t)}{\pi(s)\pi(t)}\right\}\rho(s,t) \quad (12)$$

and show that the integral over A_1 of the expectation of the square of each summand converges to zero. For the first summand we compute

$$\int_{A_1}\mathrm{E}\left(\left[\frac{\hat{\nu}(s,t)}{\hat{\pi}(s)\hat{\pi}(t)}\{\hat{\rho}(s,t) - \rho(s,t)\}\right]^2 \{1_{[\min(\hat{\pi}(s),\hat{\pi}(t)) \geq \pi_0/2]} + 1_{[\min\{\hat{\pi}(s),\hat{\pi}(t)\} < \pi_0/2]}\}\right)dsdt$$

$$\leq \mathrm{E}\int_{A_1}\{\hat{\rho}(s,t) - \rho(s,t)\}^2 dsdt\left[(\pi_0/2)^{-4} + n^4 \sup_{(s,t) \in [0,1]^2}\Pr(\min\{\hat{\pi}(s),\hat{\pi}(t)\} < \pi_0/2)\right],$$

20

where the integral term converges to 0 by similar arguments to those in the proof of Proposition 1 in Kraus (2015) with the help of Condition 1(c) and the second term goes to 0 by Hoeffding's inequality again. For the second summand on the right in (12) we can write

$$
\int_{A_1} \mathrm{E}\left[I(s,t) \left\{ \frac{\pi(s)\pi(t)\hat{\nu}(s,t) - \hat{\pi}(s)\hat{\pi}(t)\nu(s,t)}{\hat{\pi}(s)\hat{\pi}(t)\pi(s)\pi(t)} \right\}^2 \right] \rho(s,t)^2 ds dt
$$
$$
+ \int_{A_1} \mathrm{E}\left[\{1 - I(s,t)\} \left\{ \frac{\nu(s,t)}{\pi(s)\pi(t)} \right\}^2 \right] \rho(s,t)^2 ds dt. \tag{13}
$$

Like before, we split the first term in (13) into two summands by writing

$$
\int_{A_1} \mathrm{E}\left[I(s,t) \left\{ \frac{\pi(s)\pi(t)\hat{\nu}(s,t) - \hat{\pi}(s)\hat{\pi}(t)\nu(s,t)}{\hat{\pi}(s)\hat{\pi}(t)\pi(s)\pi(t)} \right\}^2 \left\{ 1_{[\min(\hat{\pi}(s),\hat{\pi}(t)) \geq \pi_0/2]} + 1_{[\min\{\hat{\pi}(s),\hat{\pi}(t)\} < \pi_0/2]} \right\} \right] \rho(s,t)^2 ds dt.
$$

The first summand is bounded by $16\pi_0^{-8} \int_{A_1} \mathrm{E}[\{\pi(s)\pi(t)\hat{\nu}(s,t) - \hat{\pi}(s)\hat{\pi}(t)\nu(s,t)\}^2]\rho(s,t)^2 ds dt$, which converges to 0 by the dominated convergence theorem since the expectation in the integrand can be shown to converge to 0; the second summand in the displayed expression above is dominated by $n^4 \pi_0^{-4} \|\mathscr{R}\|_2^2 \sup_{(s,t)\in[0,1]^2} \Pr(\min\{\hat{\pi}(s),\hat{\pi}(t)\} < \pi_0/2)$, which converges to 0 by Hoeffding's inequality. Finally, the second term in (13) is dominated by $\sup_{(s,t)\in A_1} \Pr(\hat{\nu}(s,t) < \nu_0/2)\pi_0^{-4}\|\mathscr{R}\|_2^2$, which converges to 0 again by Hoeffding's inequality.

Proof of Theorem 2

Denote $Z_j(\cdot) = N_j(\cdot)^{1/2}\{\hat{\mu}_j(\cdot) - \hat{\mu}(\cdot)\}/\hat{r}_j$ and $Z = (Z_1,\ldots,Z_K)^\top$. Under the null hypothesis we can write $Z = \hat{\mathscr{D}}H$, where $H = (H_1,\ldots,H_K)^\top$ with $H_j = N_j^{1/2}(\hat{\mu}_j - \mu)$ and $\hat{\mathscr{D}}$ is a bounded linear operator from $\{L^2([0,1])\}^K$ to $\{L^2([0,1])\}^K$ that maps an element f to an element g whose jth component is given by $g_j(t) = \sum_{l=1}^K (\hat{\mathscr{D}}_{jl}f_l)(t) = \sum_{l=1}^K \hat{r}_j^{-1}\{\delta_{jl} - N_j(t)^{1/2}\hat{w}_l(t)J_l(t)N_l(t)^{-1/2}\}f_l(t)$ (here δ_{jl} is the Kronecker delta and $J_l(t)N_l(t)^{-1/2}$ is zero if $J_l(t) = 1_{[N_l(t)>0]}$ is zero). From Theorem 1 we see that H converges in distribution to the random element $H^\infty = (H_1^\infty,\ldots,H_K^\infty)^\top$ whose components are mutually independent Gaussian processes with mean zero and covariance operators $\mathscr{K}_j, j = 1 \ldots, K$ analogous to the operator $\mathscr{K}$ in Theorem 1. The operator $\hat{\mathscr{D}}$ converges in probability to the operator $\mathscr{D}$ whose elements are defined by $(\mathscr{D}_{jl}f_l)(t) = r_j^{-1}\{\delta_{jl} - \pi_j(t)^{1/2}a_j^{1/2}w_l(t)\pi_l(t)^{-1/2}a_l^{-1/2}\}f_l(t)$ with $w_l(t) = a_l \pi_l(t)/r_l^2/(\sum_{k=1}^K a_k\pi_k(t)/r_k^2)$ (the convergence is in the operator norm, i.e., $\|\hat{\mathscr{D}} - \mathscr{D}\|_\infty \xrightarrow{P} 0$). Therefore, it follows from Slutsky's and continuous mapping theorem that $Z = \hat{\mathscr{D}}H$ converges weakly to $Z^\infty = \mathscr{D}H^\infty$. This is a K-dimensional mean zero Gaussian random process with cross-covariance operator between Z_j^∞ and Z_k^∞ equal to $\mathscr{V}_{jk} = \sum_{l=1}^K \mathscr{D}_{jl}\mathscr{K}_l\mathscr{D}_{kl}^*, j = 1,\ldots,K, k = 1,\ldots,K$. These can be consistently estimated by plugging-in the estimators $\hat{\mathscr{D}}_{jl}$ and $\hat{\mathscr{K}}_l$. The kernel of the estimator $\hat{\mathscr{V}}_{jk}$ takes the form $\hat{v}_{jk}(s,t) = \sum_{l=1}^K \hat{r}_j^{-1}\{\delta_{jl} - N_j(s)^{1/2}\hat{w}_l(s)N_l(s)^{-1/2}\}\hat{\kappa}_l(s,t)\{\delta_{kl} - N_k(t)^{1/2}\hat{w}_l(t)N_l(t)^{-1/2}\}\hat{r}_k^{-1}$.

For (i), the continuous mapping theorem gives that the statistic $T_{L^2} = \|Z\|^2$ converges weakly to the random variable $\|Z^\infty\|^2$. The process Z^∞ is a Gaussian random element of the separable Hilbert space $\{L^2([0,1])\}^K$. Therefore, it can be expanded in a Karhunen–Loève series with Gaussian coefficients. Consequently, the distribution of its squared norm is that of the series given in the theorem. The consistency of $\hat{\mathscr{V}}$ implies the consistency of the estimated eigenvalues.

To prove (ii), notice that the components of the score vector satisfy $Q_{jl} = \langle \hat{\pi}_j^{1/2}Z_j, \hat{\psi}_l \rangle$. The continuous mapping theorem and Slutsky's theorem in conjunction with the convergence of $\hat{\psi}_l$ imply that Q is asymptotically distributed as a Gaussian vector with mean zero and covariance matrix with entries $V_{jl,km} = \langle \pi_j^{1/2}\psi_l, \mathscr{V}_{jk}(\pi_k^{1/2}\psi_m) \rangle$. The consistency of $\hat{V}_{jl,km}$ follows from the consistency of $\hat{\mathscr{V}}_{jk}$ and $\hat{\pi}_j$ and convergence of $\hat{\psi}_l$. The process $(\hat{\pi}_1^{1/2}Z_1,\ldots,\hat{\pi}_K^{1/2}Z_K)$ lies in a $(K-1)$-dimensional subspace of the K-dimensional product space $\{L^2([0,1])\}^K$ and the same holds for its limit. Therefore, the score vector lies in a $(K-1)d$-dimensional subspace of $\mathbb{R}^{Kd}$, leading to $(K-1)d$ degrees of freedom of the chi-square distribution.

Proof of Theorem 3

The kernel of $n^{1/2}(\hat{\mathscr{R}} - \mathscr{R})$ is

$$n^{1/2}\{\hat{\rho}(s,t) - \rho(s,t)\} = n^{1/2}\{\hat{\rho}(s,t) - \check{\rho}(s,t)\} + \frac{1}{\nu(s,t)}\sigma(s,t) + \left\{\frac{I(s,t)}{\hat{\nu}(s,t)} - \frac{1}{\nu(s,t)}\right\}\sigma(s,t)$$
$$+ n^{1/2}\{I(s,t) - 1\}\rho(s,t), \tag{14}$$

where $\check{\rho}$ is defined like $\hat{\rho}$ with the true mean in place of the estimated mean and $\sigma(s,t) = n^{-1/2}\sum_{i=1}^{n}U_i(s,t)[\{X_i(s) - \mu(s)\}\{X_i(t) - \mu(t)\} - \rho(s,t)]$. Let us focus on the second summand on the right side of (14). All the other terms are negligible in the appropriate sense as we explain later. The kernel $\sigma(s,t)$ corresponds to the operator $\mathscr{S}_n = n^{-1/2}\sum_{i=1}^{n}\mathscr{Y}_{ni}$, where $\mathscr{Y}_{ni}$ are the integral operators with kernels $y_{ni}(s,t) = U_i(s,t)[\{X_i(s) - \mu(s)\}\{X_i(t) - \mu(t)\} - \rho(s,t)]$. We will apply Theorem 6 to $\mathscr{Y}_{ni}$, which is a triangular array of row-wise independent non-identically distributed zero-mean random elements of the separable Hilbert space of the Hilbert–Schmidt operators on $L^2([0,1])$. The covariance operator of $\mathscr{Y}_{ni}$ is the Hilbert–Schmidt operator $\mathfrak{C}_{ni}$ on Hilbert–Schmidt operators given by

$$\langle \mathscr{A}_1, \mathfrak{C}_{ni}\mathscr{A}_2 \rangle = \text{cov}(\langle \mathscr{Y}_{ni}, \mathscr{A}_2 \rangle, \langle \mathscr{Y}_{ni}, \mathscr{A}_1 \rangle) = \int_{[0,1]^4} \alpha_1(s,t)\,\text{cov}\{y_{ni}(s,t), y_{ni}(u,v)\}\alpha_2(u,v)\,dsdtdudv,$$

where $\mathscr{A}_1$, $\mathscr{A}_2$ are Hilbert–Schmidt operators with kernels α_1, α_2, respectively. The kernel of $\mathfrak{C}_{ni}$ is $c_{ni}(s,t,u,v) = \text{cov}\{y_{ni}(s,t), y_{ni}(u,v)\} = \theta_i(s,t,u,v)\{\zeta(s,t,u,v) - \rho(s,t)\rho(u,v)\}$. The covariance operator of $\mathscr{S}_n$ is $\mathfrak{G}_n = n^{-1}\sum_{i=1}^{n}\mathfrak{C}_{ni}$ with kernel $\bar{\theta}(s,t,u,v)\{\zeta(s,t,u,v) - \rho(s,t)\rho(u,v)\}$. Like in the proof of Theorem 1, one can use the dominated convergence theorem to show that $\|\mathfrak{G}_n - \mathfrak{G}\|_2 \to 0$, where $\mathfrak{G}$ has kernel $\theta(s,t,u,v)\{\zeta(s,t,u,v) - \rho(s,t)\rho(u,v)\}$. Thus condition (i) of Theorem 6 is verified. Condition (ii) can be verified like in the proof of Theorem 1. Next, condition (iii) is satisfied because $\text{tr}\,\mathfrak{G}_n = \int_{[0,1]^2}\bar{\theta}(s,t,s,t)\{\zeta(s,t,s,t) - \rho(s,t)^2\}dsdt$ converges to $\text{tr}\,\mathfrak{G} = \int_{[0,1]^2}\theta(s,t,s,t)\{\zeta(s,t,s,t) - \rho(s,t)^2\}dsdt$. Therefore, $\mathscr{S}_n$ is asymptotically distributed as a Gaussian random operator with mean zero and covariance operator $\mathfrak{G}$ and, consequently, by the continuous mapping theorem the second term on the right-hand side of (14) weakly converges to the mean zero Gaussian operator with covariance operator $\mathfrak{H}'$ given in Theorem 3.

The operators corresponding to the first and fourth summand on the right side in (14) were shown to converge to zero in the proof of Proposition 1 in Kraus (2015) in the sense that the expectation of their squared Hilbert–Schmidt norm converges to zero. Also, the Hilbert–Schmidt norm of the third term on the right in (14) converge to zero in mean square which can be shown by arguments analogous to those used for the second term on the right in (10) in the proof of Theorem 1. Therefore, in view of Slutsky's lemma these terms are negligible for the weak convergence.

The weak convergence of the operator with kernel $M(s,t)^{1/2}\{\hat{\rho}(s,t) - \rho(s,t)\}$ follows from the convergence of $\hat{\nu}(s,t)$ to $\nu(s,t)$. The consistency of the estimators of $\mathfrak{H}'$ and $\mathfrak{H}$ can be proved along the lines of the proof for $\mathscr{K}'$ and $\mathscr{K}$ in Theorem 1.

Proof of Theorem 4

The proof uses perturbation theory in which $\hat{\mathscr{R}}$ is regarded as a perturbed version of $\mathscr{R}$, i.e., $\hat{\mathscr{R}} = \mathscr{R} + (\hat{\mathscr{R}} - \mathscr{R})$. Recall that the perturbation satisfies $E\|\hat{\mathscr{R}} - \mathscr{R}\|_2^2 = O(n^{-1})$ (Kraus, 2015, Proposition 1), and, therefore, $\|\hat{\mathscr{R}} - \mathscr{R}\|_\infty = O_P(n^{-1/2})$.

Similarly to the proof of Theorem 3.1 in Cupidon et al. (2007), we rewrite $n^{1/2}(\hat{\lambda}_m - \lambda_m) = n^{1/2}(\hat{\lambda}_m - \lambda_m)1_{\Omega_n} + n^{1/2}(\hat{\lambda}_m - \lambda_m)1_{\Omega_n^{\complement}}$, where $\Omega_n = \{\omega : \|\hat{\mathscr{R}} - \mathscr{R}\|_\infty < \varepsilon_n\}$ for a numerical sequence ε_n satisfying $n^{-1/2} \ll \varepsilon_n \ll n^{-1/4}$. Since $\text{Pr}(\Omega_n) \to 1$ as $n \to \infty$, the term $n^{1/2}(\hat{\lambda}_m - \lambda_m)1_{\Omega_n^{\complement}}$ converges to 0 in probability. For $\|\hat{\mathscr{R}} - \mathscr{R}\|_\infty$ sufficiently small, i.e., on Ω_n for n large enough, we have by Corollary 3.4 of Gilliam et al. (2009) that $n^{1/2}(\hat{\lambda}_m - \lambda_m)1_{\Omega_n} = n^{1/2}\langle(\hat{\mathscr{R}} - \mathscr{R})\varphi_m, \varphi_m\rangle1_{\Omega_n} + n^{1/2}O(\|\hat{\mathscr{R}} - \mathscr{R}\|_\infty^2)1_{\Omega_n}$. Here the last term converges to 0 in probability because $\varepsilon_n \ll n^{-1/4}$ and the first term on the right side converges in distribution to the limit given in part (i) of the theorem. Hence the result follows from Slutsky's theorem. The expression for the limiting variance is obtained by rewriting $\text{var}\langle\mathscr{H}'^\infty\varphi_m, \varphi_m\rangle = \text{var}\langle\mathscr{H}'^\infty, \varphi_m \otimes \varphi_m\rangle = \langle\varphi_m \otimes \varphi_m, \mathfrak{H}'(\varphi_m \otimes \varphi_m)\rangle$.

Next, we can write $n^{1/2}(\hat{s}_m\hat{\varphi}_m - \varphi_m) = n^{1/2}(\hat{s}_m\hat{\varphi}_m - \varphi_m)1_{\Omega_n} + n^{1/2}(\hat{s}_m\hat{\varphi}_m - \varphi_m)1_{\Omega_n^\complement}$. For n sufficiently large, Corollary 3.3 of Gilliam et al. (2009) gives $n^{1/2}(\hat{s}_m\hat{\varphi}_m - \varphi_m)1_{\Omega_n} = n^{1/2}\mathscr{Q}_m(\hat{\mathscr{R}} - \mathscr{R})\varphi_m 1_{\Omega_n} + n^{1/2}O(\|\hat{\mathscr{R}} - \mathscr{R}\|_\infty^2)1_{\Omega_n}$. The first term on the right converges in distribution to the limiting distribution as claimed in part (ii) and the other terms converge in probability to 0. The limiting covariance operator is obtained by inspecting the cross-covariance operator for each pair of summands in the series $\mathscr{Q}_m\mathscr{H}'^\infty\varphi_m$. The cross-covariance between $(\varphi_k \otimes \varphi_k)\mathscr{H}'^\infty\varphi_m = \langle\varphi_k, \mathscr{H}'^\infty\varphi_m\rangle\varphi_k$ and $(\varphi_l \otimes \varphi_l)\mathscr{H}'^\infty\varphi_m = \langle\varphi_l, \mathscr{H}'^\infty\varphi_m\rangle\varphi_l$ is

$$\mathrm{cov}(\langle\varphi_k, \mathscr{H}'^\infty\varphi_m\rangle, \langle\varphi_l, \mathscr{H}'^\infty\varphi_m\rangle)(\varphi_k \otimes \varphi_l) = \mathrm{cov}\{\langle(\varphi_m \otimes \varphi_k), \mathscr{H}'^\infty\rangle, \langle(\varphi_m \otimes \varphi_l), \mathscr{H}'^\infty\rangle\}(\varphi_k \otimes \varphi_l)$$
$$= \langle(\varphi_m \otimes \varphi_k), \mathfrak{H}'(\varphi_m \otimes \varphi_l)\rangle(\varphi_k \otimes \varphi_l).$$

The inner product in the last expression above equals the integral in part (ii) of the theorem.

Proof of Theorem 5

Let $\hat{\mathfrak{D}}$ be the linear operator on the product space $\mathrm{HS}(L^2([0,1]))^K$ that maps $\mathscr{F} = (\mathscr{F}_1, \ldots, \mathscr{F}_K)^\top$, where $\mathscr{F}_j$ are Hilbert–Schmidt operators on $L^2([0,1])$ with kernels $f_j(s,t)$, to $\mathscr{G} = (\mathscr{G}_1, \ldots, \mathscr{G}_K)^\top$ where $\mathscr{G}_j$ has kernel $g_j(s,t) = \sum_{l=1}^{K}\{\delta_{jl} - M_j(s,t)^{1/2}\hat{w}_l(s,t)I_l(s,t)M_l(s,t)^{-1/2}\}f_l(s,t)$. The mapping $\hat{\mathfrak{D}}$ is a random linear operator on $\mathrm{HS}(L^2([0,1]))^K$ that acts by pointwise multiplication and linear combination of integral kernels; $\hat{\mathfrak{D}}$ itself is not an integral operator but it is bounded because the functions in the braces above are bounded. It converges in probability to the non-random bounded linear operator $\mathfrak{D}$ that maps $\mathscr{F}$ to $\mathscr{G}$ with $\mathscr{G}_j$ with kernel $\sum_{l=1}^{K}\{\delta_{jl} - \nu_j(s,t)^{1/2}a_j^{1/2}w_l(s,t)\nu_l(s,t)^{-1/2}a_l^{-1/2}\}f_l(s,t)$. The convergence is in the sense of the operator norm on linear operators on $\mathrm{HS}(L^2([0,1]))^K$, that is, $\|\hat{\mathfrak{D}} - \mathfrak{D}\|_\infty \xrightarrow{P} 0$, where $\|\mathfrak{D}\|_\infty = \sup\{\|\mathfrak{D}\mathscr{F}\|_2/\|\mathscr{F}\|_2 : \mathscr{F} \in \mathrm{HS}(L^2([0,1]))^K\}$ with $\|\cdot\|_2$ being the Hilbert–Schmidt norm on $\mathrm{HS}(L^2([0,1]))^K$.

Now consider the standardized contrasts $\mathscr{Z} = (\mathscr{Z}_1, \ldots, \mathscr{Z}_K)^\top$ with kernels $z_j(s,t) = M_j(s,t)^{1/2}\{\hat{\rho}_j(s,t) - \hat{\rho}(s,t)\}$. They are obtained as $\mathscr{Z} = \hat{\mathfrak{D}}\mathscr{H}$, where $\mathscr{H} = (\mathscr{H}_1, \ldots, \mathscr{H}_K)^\top$ with $\mathscr{H}_j$ with kernel $h_j(s,t) = M_j(s,t)^{1/2}\{\hat{\rho}(s,t) - \rho(s,t)\}$. Under the null hypothesis Theorem 3 yields that $\mathscr{H}$ converges in distribution to $\mathscr{H}^\infty$, a vector of K independent mean zero Gaussian random operators with covariance operators $\mathfrak{H}_j$. Therefore, $\mathscr{Z} = \hat{\mathfrak{D}}\mathscr{H}$ converges in distribution to $\mathscr{Z}^\infty = \mathfrak{D}\mathscr{H}^\infty$ by Slutsky's and continuous mapping theorem.

The covariance operator $\mathfrak{B}$ of $\mathscr{Z}^\infty$ is given by the cross-covariance operators $\mathfrak{B}_{jk}$ between the components $\mathscr{Z}_j$ and $\mathscr{Z}_k$ whose estimator $\hat{\mathfrak{B}}_{jk}$ has kernel

$$\hat{\beta}_{jk}(s,t,u,v) = \sum_{l=1}^{K}\{\delta_{jl} - M_j(s,t)^{1/2}\hat{w}_l(s,t)M_l(s,t)^{-1/2}\}\hat{\eta}_l(s,t,u,v)\{\delta_{kl} - M_k(u,v)^{1/2}\hat{w}_l(u,v)M_l(u,v)^{-1/2}\}.$$

The test statistic $S_{\mathrm{HS}} = \|\mathscr{Z}\|_2^2$ is asymptotically distributed as $\|\mathscr{Z}^\infty\|_2^2$. The random variable $\mathscr{Z}^\infty$ is a Gaussian element of the separable Hilbert space $\mathrm{HS}(L^2([0,1]))^K$, therefore it can be expanded in a Karhunen–Loève series with independent Gaussian coefficients. Therefore, its squared norm is distributed as the series of independent chi-square variables weighted by the eigenvalues of the covariance operator and part (i) of the theorem follows.

The components of the score vector satisfy $R_{jlm} = \langle\hat{\nu}_j(\cdot,\cdot)^{1/2}z_j(\cdot,\cdot), \hat{\mathscr{U}}_{lm}\rangle$. Due to the consistency of the estimated eigenfunctions (Kraus, 2015, Proposition 2), the operator $\hat{\mathscr{U}}_{lm}$ (up to the sign ambiguity for $l \neq m$) converges to $\mathscr{U}_{lm}$ defined by the true eigenfunctions, with kernel $u_{lm}(s,t)$. Therefore, the score vector weakly converges to the mean zero Gaussian vector with components $R_{jlm}^\infty = \langle\nu_j(\cdot,\cdot)^{1/2}z_j^\infty(\cdot,\cdot), \mathscr{U}_{lm}\rangle = \langle z_j^\infty(\cdot,\cdot), \nu_j(\cdot,\cdot)^{1/2}u_{lm}(\cdot,\cdot)\rangle$ whose covariance matrix has entries $W_{jlm,kpq} = \langle\nu_j(\cdot,\cdot)^{1/2}u_{lm}(\cdot,\cdot), \mathfrak{B}_{jk}\{\nu_k(\cdot,\cdot)^{1/2}u_{pq}(\cdot,\cdot)\}\rangle$, $j,k \in \{1,\ldots,K\}$, $1 \leq l \leq m \leq d$, $1 \leq p \leq q \leq d$. The vector of operators with kernels $\nu_j(s,t)^{1/2}z_j^\infty(s,t)$ lies in a hyperplane in $\mathrm{HS}(L^2([0,1]))^K$, thus the matrix W has rank $(K-1)d(d+1)/2$. The consistency of $\hat{W}$ follows from the convergence of all quantities involved. Hence the limiting distribution is the chi-square distribution as claimed in part (ii).

Proof of Theorem 6

First, we prove the convergence in distribution of one-dimensional projections using Lindeberg's central limit theorem. It follows from assumption (i) that for $f \in \mathcal{H}$ such that $\mathscr{G}f \neq 0$, $\mathrm{var}\langle S_n, f\rangle = \langle f, \mathscr{G}_n f\rangle \to \langle f, \mathscr{G}f\rangle$

as $n \to \infty$. To verify Lindeberg's condition, we compute

$$n^{-1} \sum_{i=1}^{n} \mathrm{E}(\langle Y_{ni}, f \rangle^2 1_{[|\langle Y_{ni}, f \rangle| > n^{1/2} \langle f, \mathscr{G}_n f \rangle^{1/2} \varepsilon]}) \leq n^{-1} \sum_{i=1}^{n} \mathrm{E}(\|Y_{ni}\|^2 \|f\|^2 1_{[\|Y_{ni}\| > n^{1/2} \langle f, \mathscr{G}_n f \rangle^{1/2} \|f\|^{-1} \varepsilon]}).$$

Now in light of assumption (i), there is a positive constant c such that for sufficiently large n, $\langle f, \mathscr{G}_n f \rangle^{1/2} / \|\mathscr{G}_n\|_\infty > c$, and the above expression is further dominated by $n^{-1} \sum_{i=1}^{n} \mathrm{E}(\|Y_{ni}\|^2 \|f\|^2 1_{[\|Y_{ni}\| > n^{1/2} \|\mathscr{G}_n\|_\infty c \|f\|^{-1} \varepsilon]})$, which converges to 0 by assumption (ii). Hence one-dimensional projections converge, and due to Theorem 2.3 of Bosq (2000), all finite-dimensional projections converge.

To complete the proof, let us prove the tightness of the sequence S_n, $n = 1, 2, \ldots$ The idea of the proof is similar to that of Bosq (2000, Theorem 2.7) but in the present situation the variables $Y_{n1}, \ldots, Y_{nn}$ are possibly non-identically distributed. Let v_j and δ_j, $j = 1, 2, \ldots$ be the eigenfunctions and eigenvalues of the limiting operator $\mathscr{G}$. Consider a sequence l_k, $k = 1, 2, \ldots$ such that $l_k \to \infty$ for $k \to \infty$. For $\varepsilon > 0$, let N_k, $k = 1, 2, \ldots$ be an increasing sequence of integers such that $\sum_{k=1}^{\infty} l_k r_{N_k}^2 < \varepsilon$, where $r_N^2 = \sum_{j=N}^{\infty} \delta_j$. Define $B_k = \{x \in \mathcal{H} : \sum_{j=N_k}^{\infty} \langle x, v_j \rangle^2 \leq l_k^{-1}\}$. It follows from assumptions (i) and (iii) that

$$\Pr(S_n \in B_k^{\complement}) = P\left(\sum_{j=N_k}^{\infty} \langle S_n, v_j \rangle^2 > l_k^{-1} \right) \leq l_k \mathrm{E}\left(\sum_{j=N_k}^{\infty} \langle S_n, v_j \rangle^2 \right) = l_k \mathrm{E}\left(\|S_n\|^2 - \sum_{j=1}^{N_k-1} \langle S_n, v_j \rangle^2 \right)$$

$$= l_k \left(\mathrm{tr}\, \mathscr{G}_n - \sum_{j=1}^{N_k-1} \langle v_j, \mathscr{G}_n v_j \rangle \right) \to l_k \left(\mathrm{tr}\, \mathscr{G} - \sum_{j=1}^{N_k-1} \langle v_j, \mathscr{G} v_j \rangle \right) = l_k \sum_{j=N_k}^{\infty} \langle v_j, \mathscr{G} v_j \rangle = l_k r_{N_k}^2.$$

Consider the compact set $K_\varepsilon = \cap_{k=1}^{\infty} B_k$ and compute

$$\limsup_{n \to \infty} \Pr(S_n \in K_\varepsilon^{\complement}) \leq \limsup_{n \to \infty} \sum_{k=1}^{\infty} \Pr(S_n \in B_k^{\complement}) \leq \sum_{k=1}^{\infty} \limsup_{n \to \infty} \Pr(S_n \in B_k^{\complement}) \leq \sum_{k=1}^{\infty} l_k r_{N_k}^2 < \varepsilon,$$

where the second inequality is due to Fatou's lemma. This proves the tightness.

References

Aston, J. A. D., Pigoli, D., and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics*, 45(4):1431–1461.

Aue, A., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269.

Benko, M., Härdle, W., and Kneip, A. (2009). Common functional principal components. *Annals of Statistics*, 37(1):1–34.

Boente, G., Rodriguez, D., and Sued, M. (2017). Testing equality between several populations covariance operators. *Annals of the Institute of Statistical Mathematics*, pages 1–32.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.

Bugni, F. A. (2012). Specification test for missing functional data. *Econometric Theory*, 28(5):959–1002.

Cabassi, A., Pigoli, D., Secchi, P., and Carter, P. A. (2017). Permutation tests for the equality of covariance operators of functional data with applications to evolutionary biology. *Electronic Journal of Statistics*, 11(2):3815–3840.

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

Cuevas, A., Febrero, M., and Fraiman, R. (2004). An anova test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122.

Cupidon, J., Gilliam, D., Eubank, R., and Ruymgaart, F. (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli*, 13(4):1179–1194.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154.

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

Dawson, M. and Müller, H.-G. (2018). Dynamic modeling of conditional quantile trajectories, with application to longitudinal snippet data. *Journal of the American Statistical Association*, 113(524):1612–1624.

Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *Journal of the American Statististical Association*, 108(504):1269–1283.

Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika*, 103(4):779–799.

Descary, M.-H. and Panaretos, V. M. (2019). Recovering covariance from functional fragments. *Biometrika*, 106(1):145–160.

Ferraty, F. and Romain, Y., editors (2011). *The Oxford Handbook of Functional Data Analysis.* Oxford University Press, Oxford.

Fogarty, C. B. and Small, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *The Annals of Applied Statistics*, 8(4):2002–2026.

Fremdt, S., Horváth, L., Kokoszka, P., and Steinebach, J. G. (2014). Functional data analysis with increasing number of projections. *Journal of Multivariate Analysis*, 124:313–332.

Fremdt, S., Steinebach, J. G., Horváth, L., and Kokoszka, P. (2013). Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 40(1):138–152.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2014). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association*, 109(508):1425–1439.

Gilliam, D. S., Hohage, T., Ji, X., and Ruymgaart, F. (2009). The Fréchet derivative of an analytic function of a bounded operator with some applications. *International Journal of Mathematics and Mathematical Sciences*, 2009.

Goldberg, Y., Ritov, Y., and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference*, 147:53–65.

Gromenko, O., Kokoszka, P., and Sojka, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics*, 11(2):898–918.

Guo, J., Zhou, B., and Zhang, J.-T. (2018a). New tests for equality of several covariance functions for functional data. *Journal of the American Statistical Association*. To appear.

Guo, J., Zhou, B., and Zhang, J.-T. (2018b). Testing the equality of several covariance functions for functional data: a supremum-norm based test. *Computational Statistics & Data Analysis*, 124:15–26.

Horváth, L., Hušková, M., and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*, 101(2):352–367.

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

Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 75(1):103–122.

Jarušková, D. (2013). Testing for a change in covariance operator. *Journal of Statistical Planning and Inference*, 143(9):1500–1511.

Kashlak, A., Aston, J., and Nickl, R. (2018). Inference on covariance operators via concentration inequalities: k-sample tests, classification, and clustering via Rademacher complexities. *Sankhya A*.

Kneip, A. and Liebl, D. (2017). On the optimal reconstruction of partially observed functional data. arXiv:1710.10099.

Kokoszka, P. and Reimherr, M. (2013). Asymptotic normality of the principal components of functional time series. *Stochastic Processes and their Applications*, 123(5):1546–1562.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 77(4):777–801.

Kraus, D. and Panaretos, V. M. (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika*, 99(4):813–832.

Kraus, D. and Stefanucci, M. (2019). Classification of functional fragments by regularized linear classifiers with domain selection. *Biometrika*, 106(1):161–180.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. *Annals of Applied Statistics*, 7(3):1562–1592.

Liebl, D. (2019). Nonparametric testing for differences in electricity prices: The case of the Fukushima nuclear accident. *The Annals of Applied Statistics*. To appear.

Liebl, D. and Rameseder, S. (2019). Partially observed functional data: The case of systematically missing parts. *Computational Statistics & Data Analysis*, 131:104 – 115.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, 16(4):1696–1708.

Lloyd, C. J. (2005). Estimating test power adjusted for size. *Journal of Statistical Computation and Simulation*, 75(11):921–933.

Mas, A. (2007). Testing for the mean of random curves: a penalization approach. *Stat. Inference Stoch. Process.*, 10(2):147–163.

Masarotto, V. (2019). *Procrustes Metric and Optimal Transport for Covariance Operators*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne.

Mojirsheibani, M. and Shaw, C. (2018). Classification with incomplete functional covariates. *Statistics & Probability Letters*, 139:40–46.

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association*, 105(490):670–682.

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2011). Second-order inference for functional data with application to DNA minicircles. In *Recent advances in functional data analysis and related topics*, pages 245–250. Springer.

Paparoditis, E. and Sapatinas, T. (2016a). Bootstrap-based *k*-sample testing for functional data. arXiv:1409.4317v4.

Paparoditis, E. and Sapatinas, T. (2016b). Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika*, 103(3):727–733.

Pigoli, D., Aston, J. A., Dryden, I. L., and Secchi, P. (2014). Distances and inference for covariance operators. *Biometrika*, 101(2):409–422.

Pini, A., Spreafico, L., Vantini, S., and Vietti, A. (2019). Multi-aspect local inference for functional data: Analysis of ultrasound tongue profiles. *Journal of Multivariate Analysis*, 170:162–185.

Pini, A., Stamm, A., and Vantini, S. (2018). Hotelling's T^2 in separable Hilbert spaces. *Journal of Multivariate Analysis*, 167:284–305.

Pini, A. and Vantini, S. (2016). The interval testing procedure: a general framework for inference in functional data analysis. *Biometrics*, 72(3):835–845.

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

Stefanucci, M., Sangalli, L. M., and Brutti, P. (2018). PCA-based discrimination of partially observed functional data, with an application to AneuRisk65 data set. *Statistica Neerlandica*. To appear.

Vsevolozhskaya, O., Greenwood, M., and Holodov, D. (2014). Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis. *The Annals of Applied Statistics*, 8(2):905–925.

Zhang, C., Peng, H., and Zhang, J.-T. (2010). Two samples tests for functional data. *Communications in Statistics – Theory and Methods*, 39(4):559–578.

Zhang, J.-T. (2013). *Analysis of Variance for Functional Data*. Chapman and Hall/CRC.

Zhang, J.-T. and Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics*, 41(1):51–71.