

Functional Models for Time-Varying Random Objects

Paromita Dubey

University of California, Davis

E-mail: pdubey@ucdavis.edu

Hans-Georg Müller

University of California, Davis

Summary. Functional data analysis provides a popular toolbox of functional models for the analysis of samples of random functions that are real-valued. In recent years, samples of time-varying object data such as time-varying networks that are not in a vector space have been increasingly collected. These data can be viewed as elements of a general metric space that lacks local or global linear structure and therefore common approaches that have been used with great success for the analysis of functional data, such as functional principal component analysis, cannot be applied. In this paper we propose *metric covariance*, a novel association measure for paired object data lying in a metric space (Ω, d) that we use to define a metric auto-covariance function for a sample of random Ω -valued curves, where Ω generally will not have a vector space or manifold structure. The proposed metric auto-covariance function is non-negative definite when the squared semimetric d^2 is of negative type. Then the eigenfunctions of the linear operator with the auto-covariance function as kernel can be used as building blocks for an *object functional principal component analysis* for Ω -valued functional data, including time-varying probability distributions, covariance matrices and time-dynamic networks. Analogues of functional principal components for time-varying objects are obtained by applying Fréchet means and projections of distance functions of the random object trajectories in the directions of the eigenfunctions, leading to real-valued *Fréchet scores*. Using the notion of *generalized Fréchet integrals*, we construct *object functional principal components* that lie in the metric space Ω . We establish asymptotic consistency of the sample based estimators for the corresponding population targets under mild metric entropy conditions on Ω and continuity of the Ω -valued random curves. These concepts are illustrated with samples of time-varying probability distributions for human mortality, time-varying covariance matrices derived from trading patterns, and time-varying networks that arise from New York taxi trips.

KEY WORDS: Dynamic Networks, Fréchet Integral, Functional Data Analysis, Metric Covariance, Object Data, Principal Component Analysis, Stochastic Processes.

1. Introduction

Time-varying data where one collects an i.i.d. sample of random functions, which take values in a general object space that does not have a linear structure, are increasingly common, while the statistical methodology for the analysis of such data has been lagging behind. We aim to introduce techniques that will help to fill this gap. For the case where observations consist of samples of random trajectories that take values in \mathbb{R}^p , the methodology of choice is often Functional Data Analysis (FDA) (Ramsay and Silverman, 2005; Horvath and Kokoszka, 2012; Wang et al., 2016), where methodology for one-dimensional ($p = 1$) functional data is readily available. Models for functional data that consist of vector-valued processes ($p > 1$) have been studied more recently (Zhou et al., 2008; Berrendero et al., 2011; Chiou et al., 2014; Claeskens et al., 2014; Verbeke et al., 2014; Chiou et al., 2016) as well as the case where at each time point one records a random function, i.e. function-valued stochastic processes (Park and Staicu, 2015; Chen and Müller, 2012; Chen et al., 2017). In these models, the responses are situated in a linear space, either the Euclidean space \mathbb{R}^p or the Hilbert space L^2 . Functions of objects in spaces that can be locally approximated by linear spaces such as Riemannian manifolds such as spheres have also been considered more recently (Lin et al., 2017; Dai and Müller, 2018). The major objective of this paper is to overcome the global or local linearity assumptions inherent in these previous approaches. The challenge is that existing FDA methodology relies on vector operations and inner products, which are no longer available.

Functional Principal Component Analysis (FPCA) (Kleffe, 1973; Dauxois et al., 1982) has emerged as the method of choice to represent and interpret samples of random functions that take values in linear spaces. It also provides dimension reduction by expanding an underlying random process into the basis functions given by the eigenfunctions of the auto-covariance operator and then truncating this expansion at a finite number of expansion terms. A related tool are the modes of variation, which enable exploration of the effects of single eigendirections (Castro et al., 1986; Jones and Rice, 1992) and are useful in practical applications (Dong et al., 2018). FPCA also provides a starting point for many theoretical investigations and FDA techniques such as functional clustering (Chiou and Li, 2007; Jacques and Preda, 2014; Suarez and Ghosal, 2016) or regression and classification (Yao et al., 2005a; Dai et al., 2017).

As we enter the era of big data, it has become increasingly common to observe more complex, often non-Euclidean, data on a time grid. Technological advances have made it possible to record and efficiently store time courses of image, network, sensor or other com-

plex data. For example, neuroscientists are interested in dynamic functional connectivity, where one essentially observes samples of time-varying covariance or correlation matrices obtained from functional Magnetic Resonance Imaging (fMRI) data for each subject in a sample. Time-varying network data arise in various forms, e.g. road or internet traffic networks or time-evolving social networks, and it is of interest to extract structure and patterns from such data.

To obtain efficient and interpretable summaries of the information contained in samples of complex observations is a major task in modern statistics that has led for example to the development of methods such as Geodesic Principal Component Analysis (GPCA) in the space of probability distributions on \mathbb{R} ([Bigot et al., 2017](#)) and on more general Hilbert spaces ([Seguy and Cuturi, 2015](#)) that utilize optimal transport geometry and geodesic curves under the Wasserstein metric. These approaches utilize geodesics to connect the random distributions with the Wasserstein barycenters. We aim here at identifying dominant directions of variation in a sample of time-varying random object trajectories, where the random objects are indexed by time and are situated in a general metric space. The time-varying aspect provides for a novel and little explored setting, and to develop tools that are supported by theory and are useful for the further exploration and analysis of such data is the main motivation for this paper.

While FPCA for samples of functions taking values in smooth Riemannian manifolds has been studied both practically and theoretically ([Anirudh et al., 2017; Dai and Müller, 2018](#)), these approaches critically depend on the local Euclidean property of Riemannian manifolds and cannot be extended to functional data objects that take values in more general metric spaces that do not have a tractable and relatively simple Riemannian geometry. FPCA for doubly functional data, where the observations at each time point are functions rather than scalars ([Chen et al., 2017](#)), is based on a tensor product representation of the underlying function-valued stochastic process. The functions need to be Hilbert-space valued, so that this approach cannot be applied to non-Hilbertian data. Due to the lack of linear structure, developing a form of FPCA for random functions taking values in a metric space, which we refer to as *functional random objects*, is a major challenge.

Consider a totally bounded metric space (Ω, d) and a random sample of fully observed Ω -valued functional data. Aiming to extend key tools of FDA to cover such data, we first revisit the well-established FPCA for the case of real valued functional data. The essence of FPCA is contained in the auto-covariance structure of the underlying random

functions, i.e. their covariance at different time points. This leads to the question how to quantify correlation between random objects in general metric spaces that correspond to the values of the random function at different time points. An example for such an extension of Pearson correlation to the case of multivariate data is the RV coefficient (Robert and Escoufier, 1976), which is zero if all of the vector components are uncorrelated and strictly positive otherwise.

In this article we introduce *metric covariance*, which is a novel association measure for paired data in general metric spaces. Metric covariance differs in key aspects from distance correlation, another measure of dependence between metric space data (Lyons, 2013; Székely and Rizzo, 2017), the latter being primarily suited to measure probabilistic independence rather than for quantifying the strength of ‘positive’ or ‘negative’ association, which is the primary goal of the former. Unlike distance covariance, the magnitude of metric covariance quantifies the degree of association between paired data objects. The key component of FPCA is to decompose the variation in a sample of trajectories into orthogonal directions. An important difference between metric covariance and distance covariance, which is specifically relevant in this context, arises when considering the associated notion of variance. In contrast to distance correlation, metric covariance of a random object with itself leads to an interpretable notion of variance for data objects, as we will demonstrate below. We also show that metric covariance is symmetric and non-negative definite whenever the squared distance d^2 is a semimetric of negative type (Sejdinovic et al., 2013; Lyons, 2013; Schoenberg, 1938). The notion of *metric correlation* can then be easily derived from metric covariance and random objects will be considered to be uncorrelated if they have a metric correlation of size 0.

In FPCA for \mathbb{R} -valued functional data, once the auto-covariance function has been determined, one defines a linear Hilbert-Schmidt operator whose eigenfunctions represent the orthonormal directions of variance for the functional data in the Hilbert space L^2 . The corresponding eigenvalues represent the fraction of variance explained by the respective functional principal components (FPCs), which are the lengths of the projections of the functional data in the direction of each eigenfunction. How can one extend these ideas to object-valued functional data, where one does not have a linear structure or inner product? We proceed by constructing a linear Hilbert-Schmidt operator using the proposed metric covariance as its kernel and utilize its eigenfunctions and eigenvalues. For real valued functional data, one obtains the FPCs by the Karhunen-Loëve expansion of centered functional data in the eigenbasis, where the FPCs are the inner products of the centered

functional data with respect to the eigenfunctions. Unfortunately it is not possible to ‘center’ object functional data living in general metric spaces and one also does not have an inner product. In the case of FDA in the Hilbert space L^2 , the inner products can be expressed as integrals. While due to the lack of linear structure there is no integral for functional random objects, the interpretation of inner products as integrals nevertheless provides a way forward that we develop in this paper. We propose two approaches for obtaining FPCs for object functional data, one in which the FPCs are scalar irrespective of the nature of the metric space in which the random objects live, and an alternative approach in which the FPCs themselves are random objects, i.e. Ω valued.

To obtain FPCs in object space, we introduce the notion of a *generalized Fréchet integral* of an Ω -valued curve with respect to a real valued function, where this integral resides in Ω . Generalized Fréchet integrals depend on the underlying metric d in Ω and are defined under the constraint that the real valued function in the integrand integrates to one. We draw inspiration from the covariance integral for multivariate functional data that was previously introduced as a Fréchet integral (Petersen and Müller, 2016). This previous integral is a special case of the generalized Fréchet integral introduced here; it corresponds to the special case where Ω is the space of covariance matrices and the real valued function in the integrand is the constant function one. We demonstrate that the resulting Object Functional Principal Components (Object FPCs), which reside in Ω , provide useful insights about the structure of the underlying functional random objects.

For an alternative scalar approach, we extract relatively simple characteristics from the object functional data. A first step is to define a ‘mean’ function using the notion of Fréchet means (Fréchet, 1948). This mean function resides in the object function space and serves as a ‘central’ trajectory for the object functional data. To obtain a representative scalar FPC for a specific random object trajectory and eigenfunction, we utilize projections of the distance function between the specific random object trajectory and the Fréchet mean trajectory on each of the eigenfunctions. The resulting *Fréchet scores* encapsulate variation in the departures of functional random objects from the Fréchet mean trajectory. As we illustrate in simulations and data analysis, plotting these Fréchet scores against each other often illustrates meaningful patterns in the sample of object functional data that are generally hard to capture visually, due to their complexity and non-linearity. For example, such plots can aid in detecting the presence of clusters or outliers in functional random objects.

In this paper, we have three major objectives. First, we lay out a framework for

extending FPCA to general metric space valued functional data. The population target parameters are the *metric auto-covariance* operator, its eigenvalues and eigenfunctions and the population Fréchet mean function, which are introduced in section 2; additionally, the object FPCs, which are generalized Fréchet integrals and the Fréchet scores (section 3). Second, we provide sample based estimators of these population targets and establish their asymptotic properties under mild restrictions on the metric entropy of the metric space Ω and the continuity of the object functional data (section 4). Proofs of all results are in section S1 of the online supplement. Third, we illustrate our results through simulations (section 5) and various data examples (section 6), which include samples of time-varying probability distributions of age at death obtained from human mortality data of 32 countries, time-varying yellow taxi trip networks of different regions in Manhattan observed daily during the year 2016, and of changing trade patterns between countries that can be represented as time-varying covariance matrices, followed by a brief discussion (section 7).

2. Metric Covariance

2.1. Covariance and correlation for random objects

We consider a totally bounded metric space (Ω, d) where d is a metric and an Ω -valued stochastic process $X = \{X(t)\}_{t \in [0,1]}$ on the interval $[0, 1]$. With P denoting the probability measure of the random process X , we are given a sample $\{X_i = (X_i(t))_{t \in [0,1]} : i = 1, 2, \dots, n\}$ of random Ω -valued functions on $[0, 1]$ generated by P . The simplest case is $\Omega = \mathbb{R}$ with the intrinsic Euclidean metric, where $\{X_1, X_2, \dots, X_n\}$ is a sample of real valued functional data. For general metric spaces Ω , we refer to $\{X_1, X_2, \dots, X_n\}$ as a sample of functional random objects. Inspired by the approach to obtain FPCA for real valued functional data, our first goal is to quantify the association between random objects $X(s)$ and $X(t)$ in Ω , where s and t are two arbitrary points in the domain $[0, 1]$.

For motivation, consider first the case of real random variables (U, V) with finite covariance. Imagine for a moment that we cannot add, subtract or multiply these r.v.s and are restricted to compute their distances $d_E(U, V) = |U - V|$. As is well known, one then can write the variance of U using an i.i.d. copy U' of U by $\text{Var}(U) = \frac{1}{2}E d_E^2(U, U')$.

Interestingly, this non-algebraic construction can be extended to the covariance of U, V : Let (U', V') be an i.i.d. copy of (U, V) . We then obtain an alternate formulation of

$\text{Cov}(U, V)$ in terms of pairwise distances as follows,

$$\begin{aligned}\text{Cov}(U, V) &= E((U - E(U))(V - E(V))) \\ &= \frac{1}{4}E(d_E^2(U, V') + d_E^2(U', V) - 2d_E^2(U, V)).\end{aligned}$$

If (U, V) are \mathbb{R}^d -valued random variables with $d_E(\cdot, \cdot)$ denoting the Euclidean distance in \mathbb{R}^d , a simple calculation shows that in this case,

$$\begin{aligned}&\frac{1}{4}E(d_E^2(U, V') + d_E^2(U', V) - 2d_E^2(U, V)) \\ &= E((U - E(U))^T(V - E(V))),\end{aligned}$$

which is the inner product in the Hilbert space of \mathbb{R}^d -valued random variables with finite $E(U^T U)$. Next consider the case where (U, V) are \mathcal{H} -valued random variables, where \mathcal{H} is a Hilbert space and $d_E(\cdot, \cdot)$ is replaced by $d_{\mathcal{H}}(U, V) = \|U - V\|_{\mathcal{H}}$, the metric that arises from the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ of the Hilbert space. If the metric $d_{\mathcal{H}}(\cdot, \cdot)$ is bounded then $E(\|U\|_{\mathcal{H}}^2) < \infty$. One can show with some simple algebra and utilizing the Riesz representation theorem that

$$\frac{1}{4}E(d_{\mathcal{H}}^2(U, V') + d_{\mathcal{H}}^2(U', V) - 2d_{\mathcal{H}}^2(U, V)) = E(\langle U - E(U), V - E(V) \rangle_{\mathcal{H}}), \quad (1)$$

which is the inner product in $L^2(\mathcal{H})$, the Hilbert space of \mathcal{H} -valued random variables U such that $E(\|U\|_{\mathcal{H}}^2) < \infty$.

What happens if (U, V) are Ω -valued random variables and we replace $d_{\mathcal{H}}$ by d where (Ω, d) is a general metric space with no vector space structure to rely on? Or, for which spaces does the function $\frac{1}{4}E(d^2(U, V') + d^2(U', V) - 2d^2(U, V))$ retain desirable properties? Proposition 3 of [Sejdinovic et al. \(2013\)](#) implies that whenever d is a semi-metric of negative type, there exists a Hilbert space \mathcal{H} and an injective map, say $f : \Omega \rightarrow \mathcal{H}$, with

$$d^2(U, V) = \|f(U) - f(V)\|_{\mathcal{H}}^2, \quad (2)$$

and therefore it follows from (1) that for some ‘remote’ Hilbert space \mathcal{H} and the unknown map $f(\cdot)$,

$$\begin{aligned}&\frac{1}{4}E(d^2(U, V') + d^2(U', V) - 2d^2(U, V)) \\ &= E(\langle f(U) - E(f(U)), f(V) - E(f(V)) \rangle_{\mathcal{H}}).\end{aligned} \quad (3)$$

Here a space (\mathcal{Z}, ρ) with a semi-metric ρ is of negative type if for all $n \geq 2$, $z_1, z_2, \dots, z_n \in \mathcal{Z}$ and $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$ with $\sum_{i=1}^n \alpha_i = 0$ one has

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho(z_i, z_j) \leq 0.$$

These considerations motivate the following definition of a generalized version of covariance $\text{Cov}_\Omega(U, V)$ for paired random objects (U, V) that take values in $\Omega \times \Omega$, where (Ω, d) is a metric space,

$$\text{Cov}_\Omega(U, V) = \frac{1}{4}E(d^2(U, V') + d^2(U', V) - 2d^2(U, V)), \quad (4)$$

where as above (U', V') is an i.i.d. copy of (U, V) . We refer to $\text{Cov}_\Omega(U, V)$ as *metric covariance* of U and V . Metric covariance is always finite if the underlying metric space is bounded and coincides with the usual notion of covariance in Euclidean spaces.

We also define *metric correlation* between two Ω -valued random variables as follows,

$$\rho_\Omega(U, V) = \frac{\text{Cov}_\Omega(U, V)}{\sqrt{\text{Cov}_\Omega(U, U)\text{Cov}_\Omega(V, V)}}.$$

By the Cauchy-Schwarz inequality one has $-1 \leq \rho_\Omega(U, V) \leq 1$. Metric covariance/correlation depends on the choice of the metric d and different choices of d might reveal different aspects of association between random objects, depending on the underlying geometry of the metric.

2.2. Metric auto-covariance operators

As in the real valued Euclidean case, we define the metric auto-covariance function $C(s, t)$ for functional random objects $\{X_1, X_2, \dots, X_n\} \in \Omega$ as

$$C(s, t) = \text{Cov}_\Omega(X(s), X(t)),$$

for all $(s, t) \in [0, 1] \times [0, 1]$. Obviously, $C(s, t)$ is a symmetric kernel and therefore has real eigenvalues when used as the kernel of a linear Hilbert-Schmidt operator. The following result shows that for metric spaces (Ω, d) for which the squared distance function d^2 is of negative type, the metric auto-covariance operator is positive semidefinite.

PROPOSITION 1. *If d^2 is of negative type, then $C(s, t)$ is a nonnegative definite kernel.*

By Proposition 3 in [Sejdinovic et al. \(2013\)](#) and equation (3), $\text{Cov}_\Omega(U, V) = 0$ implies that there exists an abstract Hilbert space \mathcal{H} and an injective map $f : \Omega \rightarrow \mathcal{H}$ such that $f(U)$ and $f(V)$ are orthogonal in $L^2(\mathcal{H})$. Note that $\text{Var}_\Omega(U) = \text{Cov}_\Omega(U, U) = \frac{1}{2}E(d^2(U, U'))$, which for real valued random variables equals $\text{Var}(U)$.

Formally, one can define the *metric auto-covariance operator* as a linear Hilbert-Schmidt integral operator T_C that operates on functions $g \in L^2([0, 1])$ and utilizes the metric auto-covariance kernel,

$$(T_C g)(s) = \int_0^1 C(s, t)g(t)dt.$$

We note that for example Theorem 4.6.4 of [Hsing and Eubank \(2015\)](#) implies the non-negative definiteness of the kernel $C(s, t)$, in the sense that $\langle T_C f, f \rangle \geq 0$ for all f .

By Mercer's theorem there is an orthonormal basis $\{\phi_i\}_{i=1}^{\infty}$ of $L^2([0, 1])$ consisting of eigenfunctions of T_C such that the corresponding sequence of eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$, which are ordered in declining order, is nonnegative, since $C(s, t)$ is positive semidefinite. The eigenfunctions corresponding to non-zero eigenvalues are continuous on $[0, 1]$ and C has the representation

$$C(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t),$$

where the convergence is absolute and uniform; see, e.g. Lemma 4.6.1 and Theorems 4.5.2, 4.6.2, 4.6.5 and 4.6.7 of [Hsing and Eubank \(2015\)](#).

We thus accomplished the first step of extending FPCA from Euclidean valued functional data to general metric space valued functional data. The eigenfunctions $\{\phi_j\}_{j=1}^{\infty}$ can be interpreted as principal directions of variation of the functional object process and will be ordered according to the size of the associated eigenvalues. We can view the eigenvalues as representing a metric version of ‘fraction of variance explained’, which is their common interpretation in the real-valued case. The only requirement for this extension is that the squared metric d^2 is of negative type but this is not a severe restriction and in light of Proposition 3 of [Sejdinovic et al. \(2013\)](#) is true for the following examples:

1. (Ω, d) where Ω is the space of univariate probability distributions on a common compact support in $T \subset \mathbb{R}$. Choices of d include the popular 2-Wasserstein metric or the L^2 metric.
2. (Ω, d) where Ω is the space of correlation matrices of a fixed dimension r , where the choice of metrics includes the Frobenius metric, log Frobenius metric, power Frobenius metric and Procrustes metric ([Dryden et al., 2009](#); [Pigoli et al., 2014](#); [Tavakoli et al., 2016](#)).
3. (Ω, d) where Ω is the space of networks with a fixed number, say r , of nodes. One can view networks as adjacency matrices or graph Laplacians equipped with the Frobenius metric ([Ginestet et al., 2017](#)) or as resistance matrices equipped with the resistance perturbation metric ([Monnig and Meyer, 2018](#)).

We conclude that in most cases of interest the auto-covariance operator and its eigenfunctions will be well defined.

2.3. Interpretation of metric covariance

When X and Y are real valued, classical Pearson correlation captures the strength and sign of linear (also monotone) associations between X and Y . From a geometrical perspective, Pearson correlation can be interpreted as the cosine of the angle between X and Y . In \mathbb{R}^d , angles between vectors are defined using inner products, which can also be used for data in Hilbert space to characterize dependency. Specifically, for random functions in the metric space L^2 this idea leads to the notion of “dynamic correlation” in functional data analysis (Dubin and Müller, 2005), which was found to be useful for data analysis in genetics (Opgen-Rhein and Strimmer, 2006) and psychology (Liu et al., 2016). Dynamic correlation turns out to be equivalent to metric covariance when the random objects are in the Hilbert space $L^2([0, 1])$, equipped with the usual L^2 metric. Metric covariance then provides a generalization beyond Hilbert spaces.

For general metric spaces, under the weak assumption that the squared metric is of negative type, the map f from object to Hilbert space in (2) implies that metric covariance can be derived from the inner product in an abstract Hilbert space, while metric correlation is obtained by standardizing metric covariance, and is thus tied to the notion of an angle in an abstract space. Hence its magnitude can be interpreted as strength of association between random objects. While we use the existence of a map f and an associated abstract Hilbert space, we do not require knowledge about f . Metric covariance is thus a natural extension of Pearson covariance to general metric spaces.

In recent work (Petersen and Müller, 2019b), Wasserstein covariance for pairs of univariate probability distributions was introduced, and was shown to have an appealing interpretation as an expected value of an inner product of optimal transport maps. More specifically, if f_1 and f_2 are the components of a random bivariate density process and $F_1^{-1}(\cdot)$ and $F_2^{-1}(\cdot)$ the corresponding random quantile functions, the squared Wasserstein distance between f_1 and f_2 is given by

$$d_W^2(f_1, f_2) = \int_0^1 \{Q_1(t) - Q_2(t)\}^2 dt$$

and Wasserstein covariance between f_1 and f_2 was introduced as

$$\text{Cov}_W(f_1, f_2) = E \left[\int_0^1 \{Q_1(t) - E(Q_1(t))\} \{Q_2(t) - E(Q_2(t))\} dt \right].$$

Wasserstein covariance is then easily seen to be a special case of metric covariance when the metric space-valued random objects are probability distributions and the Wasserstein metric is used. This Wasserstein version of metric covariance was found to quantify the degree of synchronization of the movement of probability mass from the marginal Fréchet

means of the probability distributions to the random components of a multivariate density process. In applications to fMRI data, this Wasserstein version led to new findings and insights about differences in brain connectivity of normal versus Alzheimer disease patients, a topic of special interest in neuroimaging (Petersen and Müller, 2019b). The examples of dynamic correlation for Hilbert space valued random variables in functional data analysis and of Wasserstein covariance/correlation demonstrate the utility of metric covariance/correlation in non-standard spaces and its interpretability in applications. This provides evidence that metric covariance and metric correlation are indeed useful tools for data analysis in general metric spaces.

A word of caution is in order. While metric covariance covariance can be universally applied, and in the space of distributions with the Wasserstein metric has an interpretation as an inner product of transport maps, such interpretations hinge on the specific metric space in which the random objects are located and may not be available for all spaces. In practice, interpretations for specific scenarios can be important. The choice of the metric also matters and should be considered carefully, as it will affect the interpretation of metric covariance.

Apart from the interpretation of covariance as the expectation of an inner product, the diagonal elements of the metric auto-covariance surface reflect a natural notion of variance of metric-space valued objects, as

$$\text{Var}_{\Omega} = \frac{1}{2} E d^2(U, U'), \quad (5)$$

where U' is an independent copy of U . This provides a variation measure that is tied to the average squared distance of objects that are independently sampled from the underlying population, which is a natural and interpretable measure of spread that is well known to coincide with conventional variance in the Euclidean case.

Since it is sensible to define variance for metric-space valued random objects as

$$\frac{1}{2} E d^2(U, U') = \frac{1}{2} E[d^2(U, U') - d^2(U, U)],$$

it is then natural to extend this to a covariance measure between random objects (U, V) that reflects the difference between squared distances when sampling independently from the marginal distributions of U and V and when sampling from the joint distribution of (U, V) . This simple idea provides another avenue to suggest

$$\widetilde{\text{Cov}}_{\Omega}(U, V) = E[d^2(U, V') - d^2(U, V)].$$

Symmetrizing this expression and adding the factor .25 to match the usual definition of covariance in the Euclidean case then leads to formula (4). The above arguments

also lead to an interpretation of the total variance that corresponds to the trace of the proposed metric covariance operator $C(s, t)$, as an integrated squared distance between the functional random objects X and an independent copy X' ,

$$\sum_{j=1}^{\infty} \lambda_j = \int_0^1 \text{Cov}_{\Omega}(X(t), X(t)) dt = \frac{1}{2} \int_0^1 E d^2(X(t), X'(t)) dt, \quad (6)$$

We find in our examples and applications that the eigenfunctions derived from metric covariance lead to useful and often well-interpretable modes of variation of the time-varying metric random objects in the sense of [Jones and Rice \(1992\)](#), adding to the practical appeal of metric covariance for the analysis of functional random objects.

To conclude this discussion, we note that metric covariance differs substantially from distance correlation ([Székely et al., 2007](#); [Lyons, 2013](#)). A distinguishing feature of distance correlation is that it is equivalent to probabilistic independence between the distributions of U and V when it is 0, but we find that it is not suitable as a covariance or correlation measure for random objects in the situations that we study here. Specifically, the auto-covariance operator it generates is not useful for our purposes. For further details on this, see section [S2](#) in the supplement.

3. Functional Principal Components: Generalized Fréchet Integrals and Fréchet Scores

3.1. Generalized Fréchet Integrals and Object Functional Principal Components

FPCs in the case of real valued functional data are projections of the centered process onto the directions of the eigenfunctions and therefore summarize how a function differs from the mean function along orthonormal eigenfunction directions. Formally the FPC of the i^{th} process $X_i(t)$ and the k^{th} eigenfunction $\phi_k(t)$ is

$$\xi_{ik} = \int_0^1 (X_i(t) - \mu(t)) \phi_k(t) dt,$$

where $\mu(\cdot)$ is the mean process. The part of the score contributing to the variability of the functional data is $\int_0^1 X_i(t) \phi_k(t) dt$, which is just a horizontal shift of the actual scores, so centering is not needed when our goal is to decompose the variability of the random processes X , which is fortuitous as one cannot ‘center’ object data to obtain an analogue of $X(t) - \mu(t)$, as algebraic operations such as subtraction are not feasible in metric spaces.

In the Euclidean case for any function $\phi(\cdot)$ on $[0, 1]$, whenever $\int_0^1 \phi(t) dt \neq 0$, one can obtain a scaled version of the integral of $X(t)$ with respect to $\phi(t)$ as follows,

$$\int_0^1 X(t) \frac{\phi(t)}{\int_0^1 \phi(t) dt} dt = \operatorname{arginf}_{\omega \in \mathbb{R}} \int_0^1 d_E^2(\omega, X(t)) \frac{\phi(t)}{\int_0^1 \phi(t) dt} dt.$$

This suggests to define an integral of an Ω -valued function $S(\cdot)$ with respect to a real valued function $\phi(\cdot)$ which integrates to 1. For any real valued function $\phi(\cdot)$ with $\int_0^1 \phi(t)dt = 1$, we define the *generalized Fréchet integral* of $S(\cdot)$ with respect to $\phi(\cdot)$ as

$$\int_{\oplus} S(t)\phi(t)dt = \operatorname{arginf}_{\omega \in \Omega} \int_0^1 d^2(\omega, S(t))\phi(t)dt, \quad (7)$$

provided that the integral $\int_0^1 d^2(\omega, S(t))\phi(t)dt$ exists as a limit of Riemann sums for all $\omega \in \Omega$ and the minimizer of the integrals over $\omega \in \Omega$ exists and is unique. A special case of the integral in (7) was introduced as Fréchet integral in Petersen and Müller (2016), where an integral for the space of covariance matrices was constructed for $\phi(t)1$.

The Fréchet integrals defined here are far more general. Generalized Fréchet integrals can be interpreted as an extension of weighted Fréchet means (Fréchet, 1948). We omit the additional term “generalized” in the following and note that Fréchet integrals can be interpreted as projections of functional random objects onto functions ϕ , by weighing the elements $S(t)$ according to the value of $\phi(t)$, in direct analogy to projections in the linear function space L^2 . This feature motivates to employ Fréchet integrals to obtain object FPCs.

For fixed $\omega \in \Omega$ consider the Fréchet integral function

$$I(\omega) = \int_0^1 d^2(\omega, S(t))\phi(t)dt,$$

which, if it exists, is the limit of Riemann sums. A sufficient condition for its existence is that $d^2(\omega, S(t))\phi(t)$ is a continuous function of $t \in [0, 1]$. If the metric is bounded and $S(\cdot)$ and $\phi(\cdot)$ are continuous for $t \in [0, 1]$, the function $d^2(\omega, S(t))\phi(t)$ is a continuous function of $t \in [0, 1]$ and the integral $I(\omega)$ exists for all ω . Note that for any $\omega \in \Omega$, $I(\omega)$ is finite by the Cauchy Schwarz inequality whenever the metric space is bounded and the L^2 norm of the function $\phi(\cdot)$ is finite.

If the integrals $I(\omega)$ exist as limits of Riemann sums, the question arises under which conditions the minimizers of the Riemann sums converge and whether the limit of the minimizers coincides with the Fréchet integral $\int_{\oplus} S\phi$. Proposition 2 addresses this question. Let $0 = x_0 < x_1 < x_2 < \dots < x_k = 1$ be a partition \mathcal{P} of $[0, 1]$, where the $[x_j, x_{j+1}]$ are the subintervals of the partition and the length of the j^{th} subinterval is $\Delta_j = x_{j+1} - x_j$. The mesh size $\epsilon_{\mathcal{P}}$ of the partition is given by $\epsilon_{\mathcal{P}} = \max_j \Delta_j$. We select t_0, t_1, \dots, t_{k-1} such that for each j , $t_j \in [x_j, x_{j+1}]$. For each $\omega \in \Omega$, the Riemann sum $I_{\mathcal{P}}(\omega)$ corresponding to the partition \mathcal{P} and t_0, t_1, \dots, t_{k-1} is given by

$$I_{\mathcal{P}}(\omega) = \sum_{j=0}^{k-1} d^2(\omega, S(t_j))\phi(t_j)\Delta_j$$

and the Riemann integral $I(\omega)$ is obtained as a limit of Riemann sums as the partition gets finer. Formally, $I(\omega) = \lim_{\epsilon_P \rightarrow 0} I_P(\omega)$.

We will invoke the following assumptions for the integral function $I(\omega)$. For ease of notation, we suppress t in $\int_{\oplus} S(t)\phi(t)dt$, writing $\int_{\oplus} S\phi$ in the following.

- (I1) The integrand function $H(\omega, t) = d^2(\omega, S(t))\phi(t)$ is uniformly equicontinuous in $t \in [0, 1]$ and $\omega \in \Omega$.
- (I2) $\int_{\oplus} S\phi = \operatorname{argmin}_{\omega \in \Omega} I(\omega)$ exists and is unique, and $\inf_{d(\omega, \int_{\oplus} S\phi) > \delta} I(\omega) > I(\int_{\oplus} S\phi)$ for all $\delta > 0$.
- (I3) There exist constants $\beta > 0$, $\nu > 0$ and $C > 0$ such that

$$I(\omega) - I(\int_{\oplus} S\phi) \geq C d^{\beta}(\omega, \int_{\oplus} S\phi)$$

whenever $d(\omega, \int_{\oplus} S\phi) < \nu$.

Define $\sum_{P,\oplus} S\phi = \operatorname{argmin}_{\omega \in \Omega} I_P(\omega)$.

PROPOSITION 2. (a) Under assumption (I1), $I_P(\omega)$ converges to $I(\omega)$ uniformly in ω as $\epsilon_P \rightarrow 0$.

(b) Under assumptions (I1) and (I2), $\lim_{\epsilon_P \rightarrow 0} d(\sum_{P,\oplus} S\phi, \int_{\oplus} S\phi) = 0$.

(c) If $\lim_{\epsilon_P \rightarrow 0} h(\epsilon_P) \sup_{\omega \in \Omega} |I_P(\omega) - I(\omega)| = 0$ for a function h with $h(\delta) \rightarrow \infty$ as $\delta \rightarrow 0$, then under (I3), $\lim_{\epsilon_P \rightarrow 0} h^{\frac{1}{\beta}}(\epsilon_P) d(\sum_{P,\oplus} S\phi, \int_{\oplus} S\phi) = 0$.

As a continuous function on a compact interval is uniformly continuous, whenever $S(\cdot)$ is continuous and $\phi(\cdot)$ is bounded and continuous, (I1) holds since for $D = \operatorname{diam}(\Omega)$,

$$\begin{aligned} & |H(\omega, t_1) - H(\omega, t_2)| \\ &= |d^2(\omega, S(t_1))\phi(t_1) - d^2(\omega, S(t_2))\phi(t_1) + d^2(\omega, S(t_2))\phi(t_1) - d^2(\omega, S(t_2))\phi(t_2)| \\ &\leq 2D d(S(t_1), S(t_2)) |\phi(t_1)| + D^2 |\phi(t_1) - \phi(t_2)|. \end{aligned}$$

Assumption (I1) is sufficient to guarantee that the Fréchet integrals are well defined, while Assumption (I3) is a restriction on the curvature of the function $I(\omega)$ near its minimizer, implying convergence rates of the approximations of the Fréchet integrals. A few examples of spaces that satisfy assumptions (I2) and (I3) are as follows.

1. Let (Ω, d_W) be the space of univariate probability distributions on a common support $T \subset R$. For any $\omega \in \Omega$, denote the corresponding random distribution and quantile

functions by $Q(\omega)$. The squared 2-Wasserstein metric between distributions ω_1 and ω_2 is

$$d_W^2(\omega_1, \omega_2) = d_{L^2}^2(Q(\omega_1), Q(\omega_2)) = \int_0^1 (Q(\omega_1)(u) - Q(\omega_2)(u))^2 du.$$

For any $S(t)$ taking values in Ω , where we view $Q(S(t))$ as the quantile function of the distribution at time $t \in [0, 1]$, writing $Q(S(t))(u)$ for the u^{th} quantile of the distribution at time t , define $Q^*(u) = \int_0^1 Q(S(t))(u)\phi(t)dt$. Since $\int_0^1 \phi(t)dt = 1$, a simple calculation shows that for any $\omega \in \Omega$,

$$\operatorname{arginf}_{\omega \in \Omega} I(\omega) = \operatorname{arginf}_{\omega \in \Omega} d_{L^2}^2(Q(\omega), Q^*),$$

therefore the minimizer exists and is unique by the convexity of the space of univariate quantile functions. By the orthogonal projection theorem the minimizer $\tilde{\omega}$ is uniquely characterized by

$$\langle Q(\tilde{\omega}) - Q^*, \omega \rangle_{L^2} = 0,$$

for all $\omega \in \Omega$ and therefore it is enough to choose $\nu = C = 1$ and $\beta = 2$ in (I3).

2. Consider the space of graph Laplacians or graph adjacency matrices of connected, undirected and simple graphs with a fixed number r of nodes (Ω, d_F) , equipped with the Frobenius metric d_F . For any $\omega \in \Omega$,

$$d_F^2(\omega_1, \omega_2) = \sum_{j=1}^r \sum_{k=1}^r (\omega_{1,jk} - \omega_{2,jk})^2.$$

For any $S(t)$ taking values in Ω , let $S_{jk}(t)$ be the $(j, k)^{th}$ entry of the graph Laplacian or the graph adjacency matrix. Define $S_{jk}^* = \int_0^1 S_{jk}(t)\phi(t)dt$. Since $\int_0^1 \phi(t)dt = 1$, it can be easily seen that for any $\omega \in \Omega$,

$$\operatorname{arginf}_{\omega \in \Omega} I(\omega) = \operatorname{arginf}_{\omega \in \Omega} d_F^2(\omega, S^*),$$

and so the minimizer exists and is unique by the convexity of the space of graph Laplacians (Ginestet et al., 2017) and the space of graph adjacency matrices. Again, by the orthogonal projection theorem, the minimizer $\tilde{\omega}$ is uniquely characterized by

$$\sum_{j=1}^r \sum_{k=1}^r (\tilde{\omega}_{jk} - S_{jk}^*)\omega_{jk} = 0,$$

for all $\omega \in \Omega$ and therefore it is enough to choose $\nu = C = 1$ and $\beta = 2$ in (I3).

3. The same arguments also imply that (Ω, d_F) satisfies assumptions (I2) and (I3) when Ω is the space of correlation matrices of a fixed dimension r .

As we have seen, for general metric spaces Ω , under mild assumptions on the boundedness of the metric and continuity of the functions $S(\cdot)$ and $\phi(\cdot)$, the Fréchet integral has nice properties if it exists and is unique. Moreover, when Ω is bounded and $\int_0^1 |\phi(t)|dt < \infty$,

$$|I(\omega_1) - I(\omega_2)| \leq 2D d(\omega_1, \omega_2) \int_0^1 |\phi(t)|dt,$$

and therefore $I(\omega)$ is a continuous function of $\omega \in \Omega$. This ensures that the Fréchet integral always exists when Ω is compact.

We now define the FPCs corresponding to the bounded continuous eigenfunctions ϕ_k of the metric auto-covariance operator in the object space using Fréchet integrals. For this, we assume that all trajectories $\{X_i(t)\}_{t \in [0,1]}$ have continuous sample paths almost surely and the metric space Ω is bounded, and furthermore that

$$(A1) \int_0^1 \phi_k(t)dt \neq 0$$

$$(A2) \int_{\oplus} X_i \phi_k^* \text{ exists and is unique almost surely for all } i = 1, \dots, n, \text{ where } \phi_k^*(t) = \phi(t)/\int_0^1 \phi(t)dt.$$

Then *object functional principal components* (object FPCs). for X_i and ϕ_k are defined as the Fréchet integrals

$$\psi_{\oplus}^{ik} = \int_{\oplus} X_i \phi_k^*, \quad (8)$$

which are random objects in Ω . Similar to ordinary FPCA one can choose a number of basis functions aiming to explain a desired percentage of variation in the data utilizing the eigenvalues of the metric auto-covariance operator. If $\Omega = \mathbb{R}$, the object FPCs correspond to a location and scale shifted version of the ordinary FPCs.

3.2. Fréchet Scores

Exploratory data analysis such as checking for clusters or outliers often benefits from plotting the FPCs against each other for the case of real valued functional data. FPCs defined using Fréchet integral live in the object space Ω and therefore visualizing them is non-trivial. One approach is to obtain their projections to a lower-dimensional real space using multi-dimensional scaling or its variants (Kruskal, 1964; Belkin and Niyogi, 2002) and then visualizing the projections. Here we propose another approach for obtaining a scalar version of object FPCs. The resulting scalar FPCs are interpretable and can be plotted against each other and are thus useful for exploratory data analysis.

In the real valued case, one obtains projections of the deviations of the observed random curves from the mean curve onto dominant eigenfunctions. While the concept of a mean

function can be generalized to object functional data using Fréchet means (Fréchet, 1948), one cannot ‘center’ object data and does not have directional information. Nevertheless, it is possible to study how distances of sample curves from the mean curve project onto a few dominant eigenfunctions, in analogy to the real valued case. Formally, given a random object process $\{X(t)\}_{t \in [0,1]}$, the population Fréchet mean function is

$$\mu_{\oplus}(t) = \operatorname{argmin}_{\omega \in \Omega} E(d^2(\omega, X(t))),$$

where we assume existence and uniqueness of the minimizer. For real valued functional data under the Euclidean metric the Fréchet mean function coincides with the usual pointwise mean function. Defining distance functions

$$D_i(t) = d(X_i(t), \mu_{\oplus}(t))$$

for sample trajectories X_i , we represent the scalar functions D_i in the eigenbasis of the metric auto-covariance operator, obtaining the coefficients

$$\beta_{ik} = \int_0^1 D_i(t) \phi_k(t) dt = \int_0^1 d(X_i(t), \mu_{\oplus}(t)) \phi_k(t) dt. \quad (9)$$

We refer to the scalars β_{ik} as the *Fréchet scores*.

The Fréchet scores can be interpreted as decomposition of the departures of the sample elements from the ‘central’ Fréchet mean curve in predominant directions of variation. They can be plotted against each other and have the potential to provide interesting insights, as we will illustrate in the data applications. Considering the existence of the Fréchet scores, with D denoting as before the diameter of the totally bounded metric space Ω , continuity of the Fréchet mean function implies that for any $t_1, t_2 \in [0, 1]$,

$$\begin{aligned} & |d^2(X_i(t_1), \mu_{\oplus}(t_1)) - d^2(X_i(t_2), \mu_{\oplus}(t_2))| \\ &= |d^2(X_i(t_1), \mu_{\oplus}(t_1)) - d^2(X_i(t_1), \mu_{\oplus}(t_2)) + d^2(X_i(t_1), \mu_{\oplus}(t_2)) - d^2(X_i(t_2), \mu_{\oplus}(t_2))| \\ &\leq 2D \{d(\mu_{\oplus}(t_1), \mu_{\oplus}(t_2)) + d^2(X_i(t_1), X_i(t_2))\}. \end{aligned}$$

Thus for continuous eigenfunction $\phi_k(\cdot)$, the function $d^2(X_i(t), \mu_{\oplus}(t))\phi_k(t)$ is a continuous function of $t \in [0, 1]$ almost surely and therefore the Fréchet scores will exist. Proposition 3 shows that under assumption (A3) the Fréchet mean function is indeed continuous.

(A3) For each $t \in [0, 1]$, the pointwise Fréchet mean $\mu_{\oplus}(t)$ exists and is unique, and

$$\inf_{d(\omega, \mu_{\oplus}(t)) > \gamma} E(d^2(\omega, X(t))) > E(d^2(\mu_{\oplus}(t), X(t)))$$

for any $\gamma > 0$.

PROPOSITION 3. *If the random object process $\{X(t)\}_{t \in [0,1]}$ has almost surely continuous paths, then $\mu_{\oplus}(\cdot)$ is continuous under (A3).*

Assumption (A3) is satisfied for the space (Ω, d_W) of univariate probability distributions with the 2-Wasserstein metric and also for the space (Ω, d_F) , where Ω is the space of covariance matrices or alternatively graph Laplacians of fixed dimension with the Frobenius metric d_F (Dubey and Müller, 2017; Petersen and Müller, 2019a).

4. Estimation and Theory

Having defined suitable population targets, our goal now is to construct appropriate estimators, starting with a sample of functional random objects. An empirical estimator of the metric auto-covariance operator $C(s, t)$ as defined in section 2 is given by

$$\hat{C}(s, t) = \frac{1}{4n(n-1)} \sum_{i \neq j} f_{s,t}(X_i, X_j), \quad (10)$$

where

$$f_{s,t}(X_i, X_j) = d^2(X_i(s), X_j(t)) + d^2(X_j(s), X_i(t)) - d^2(X_i(s), X_i(t)) - d^2(X_j(s), X_j(t)).$$

Observe that for each $s, t \in [0, 1]$, $\hat{C}(s, t)$ is a U -statistic and the class $\{\hat{C}(s, t) : s, t \in [0, 1]\}$ is a family of U -statistics.

Noting that $\hat{C}(s, t)$ can be viewed as a stochastic process indexed by the function class $\mathcal{F} = \{f_{s,t}(\cdot, \cdot) : s, t \in [0, 1]\}$, where $f_{s,t}(x, y) = d^2(x(s), y(t)) + d^2(y(s), x(t)) - d^2(x(s), x(t)) - d^2(y(s), y(t))$ allows us to apply the theory of U -processes (Nolan and Pollard, 1987, 1988; Arcones and Giné, 1993) for weak convergence (Billingsley, 1968; Van der Vaart and Wellner, 1996). For the uniform convergence of $\{\hat{C}(s, t) : s, t \in [0, 1]\}$, we need an assumption on the rate of continuity of the functional random objects.

- (A4) The process $X(\cdot)$ is almost surely α -Hölder continuous for some $0 < \alpha \leq 1$, where the Hölder constant has a finite second moment, i.e. for some non negative function $G(X)$ one has

$$d(X(s), X(t)) \leq G(X)|s - t|^\alpha,$$

where $E(G(X))^2 < \infty$.

THEOREM 1. *Under assumption (A4), the sequence of stochastic processes*

$$U_n(s, t) = \sqrt{n}(\hat{C}(s, t) - C(s, t))$$

converges weakly to a Gaussian process with mean 0 and covariance function

$$R_{(s,t),(u,v)} = \text{Cov}(f_{s,t}(X, Y), f_{u,v}(X, Y)).$$

Writing $\hat{\lambda}_j$ and $\hat{\phi}_j$ for the eigenvalues and eigenfunctions of $\hat{C}(s, t)$, uniform convergence and rates of convergence of these estimates of the eigenvalues and eigenfunctions of the metric auto-covariance operator to their targets are obtained as a direct consequence of Proposition 1 under the following assumption on the spacings of the eigenvalues.

- (A5) For each $j \geq 1$, the eigenvalue λ_j has multiplicity 1, i.e it holds that $\delta_j > 0$, where

$$\delta_j = \min_{1 \leq l \leq j} (\lambda_l - \lambda_{l+1}).$$

COROLLARY 1. (*Bosq (2000)*) Under assumptions (A4) and (A5),

$$|\hat{\lambda}_j - \lambda_j| = O_P(1/\sqrt{n}).$$

$$\sup_{s \in [0, 1]} |\hat{\phi}_j(s) - \phi_j(s)| = O_P(1/\delta_j \sqrt{n}).$$

As in classical FDA, the eigenfunctions ϕ_j are uniquely identifiable only up to a sign change. For theoretical considerations such as the convergence in Corollary 1, we may always assume that true and estimated eigenfunctions are aligned in the sense that $\langle \hat{\phi}_j, \phi_j \rangle \geq 0$. Our next objective is to obtain sample estimators for the object FPCs (8) defined in section 3.1. For each j , consider the following estimators of $\phi_j^*(t)$,

$$\hat{\phi}_j^*(t) = \frac{\hat{\phi}_j(t)}{\int_0^1 \hat{\phi}_j(t) dt}.$$

A natural estimator for the Fréchet integral ψ_{\oplus}^{ik} is then

$$\hat{\psi}_{\oplus}^{ik} = \int_{\oplus} X_i \hat{\phi}_j^* = \operatorname{argmin}_{\omega \in \Omega} \int_0^1 d^2(\omega, X_i(t)) \hat{\phi}_j^*(t) dt. \quad (11)$$

To obtain convergence of $\hat{\psi}_{\oplus}^{ik}$ to its population target, we make the following assumptions.

- (A6) For every i and k , ψ_{\oplus}^{ik} and $\hat{\psi}_{\oplus}^{ik}$ exist and are unique almost surely. Moreover for any $\epsilon > 0$, $c_{\epsilon} = \inf_{d(\omega, \psi_{\oplus}^{ik}) > \epsilon} \left(\int_0^1 d^2(\omega, X_i(t)) \phi^*(t) dt - \int_0^1 d^2(\psi_{\oplus}^{ik}, X_i(t)) \phi^*(t) dt \right) > 0$ almost surely.
- (A7) There exist constants $\beta_1 > 1$, $\nu' > 0$ and $C' > 0$ such that almost surely,

$$\left(\int_0^1 d^2(\omega, X_i(t)) \phi^*(t) dt - \int_0^1 d^2(\psi_{\oplus}^{ik}, X_i(t)) \phi^*(t) dt \right) \geq C' d^{\beta_1}(\omega, \psi_{\oplus}^{ik}),$$

whenever $d(\omega, \psi_{\oplus}^{ik}) < \nu'$.

Assumption (A6) on the existence and uniqueness of the Fréchet integrals is used to establish consistency. Assumption (A7) is a restriction on the local behavior of the integrals around the minimizer and determines the rate of convergence.

THEOREM 2. *Under assumptions (A1), (A2), (A4) and (A6),*

$$d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) = o_P(1).$$

If additionally (A7) holds, then

$$d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) = O_P(n^{-1/(2\beta_1)}).$$

Here we choose $\hat{\phi}_j$ to be such that $\langle \hat{\phi}_j, \phi_j \rangle \geq 0$ which ensures matching signs for the true and estimated eigenfunctions in the computation of $\hat{\psi}_\oplus^{ik}$ and ψ_\oplus^{ik} .

Next we provide estimates of the Fréchet scores and study their asymptotics. The starting point is the following estimator of the population Fréchet mean function,

$$\hat{\mu}_\oplus(t) = \operatorname{argmin}_{\omega \in \Omega} \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \omega). \quad (12)$$

We need the following assumptions:

- (A8) The Fréchet mean function estimate $\hat{\mu}_\oplus(t)$ exists and is unique almost surely for all $t \in [0, 1]$. Additionally, for every $\varepsilon > 0$, there exists $\tau(\varepsilon) > 0$ such that

$$\lim_{n \rightarrow \infty} P \left(\inf_{s \in [0,1]} \inf_{d(\omega, \hat{\mu}_\oplus(s)) > \epsilon} \frac{1}{n} \sum_{l=1}^n \{d^2(X_l(s), \omega) - d^2(X_l(s), \hat{\mu}_\oplus(s))\} \geq \tau(\varepsilon) \right) = 1.$$

- (A9) There exists a sufficiently small $\delta > 0$ and constants $0 < \nu_\delta \leq 1$ and $H_\delta > 0$, such that for all Ω -valued functions $\omega(\cdot)$ with $d_\infty(\omega, \mu_\oplus) < \delta$, where $d_\infty(\omega, \mu_\oplus) = \sup_{s \in [0,1]} d_\infty(\omega(s), \mu_\oplus(s))$, the functions $\omega(\cdot)$ are ν_δ -Hölder continuous with Hölder constant bounded above by H_δ , i.e.

$$d(\omega(s), \omega(t)) \leq H_\delta |s - t|^{\nu_\delta}.$$

- (A10) For $I(\delta) = \int_0^1 \sup_{s \in [0,1]} \sqrt{\log N(A\varepsilon\delta, B_\delta(\mu_\oplus(s)), d)} d\varepsilon$, it holds that $I(\delta) = O(1)$ as $\delta \rightarrow 0$ for all sufficiently small $\delta > 0$ and for any constant $A > 0$. Here $B_\delta(\mu_\oplus(s)) = \{\omega \in \Omega : d(\omega, \mu_\oplus(s)) < \delta\}$ is the δ -ball around $\mu_\oplus(s)$ and $N(\varepsilon, B_\delta(\mu_\oplus(s)), d)$ is the covering number, i.e. the minimum number of balls of radius ε required to cover $B_\delta(\mu_\oplus(s))$ ([Van der Vaart and Wellner, 1996](#)).

- (A11) There exist $\alpha > 0$, $D > 0$ and $\beta_2 > 1$ such that

$$\inf_{s \in [0,1]} \inf_{d(\omega, \mu_\oplus(s)) < \alpha} \{E(d^2(X(s), \omega)) - E(d^2(X(s), \mu_\oplus(s))) - Dd^{\beta_2}(\omega, \mu_\oplus(s))\} \geq 0.$$

PROPOSITION 4. *Under assumptions (A3) and (A8),*

$$\sup_{t \in [0,1]} d(\hat{\mu}_\oplus(t), \mu_\oplus(t)) = o_P(1).$$

Assumptions (A4) and (A9)-(A11) are required to obtain an entropy condition for the space of functional random objects (Lemma 1 below), which is used to establish the rate of convergence of the sample Fréchet mean function. We note that (A9), where we assume that in a sufficiently close neighborhood of the true Fréchet mean function $\mu_{\oplus}(t)$ all object functions have a common rate of Hölder continuity and a common Hölder constant, is weaker than assumptions that have been required in classical FDA (see e.g. Müller et al., 2006), where one deals with real-valued random functions. Assumption (A10) is a bound on the covering number of the object metric space and is satisfied by common instances for random objects that include the examples discussed at the end of section 3.2.

We write $\omega(\cdot)$ for Ω -valued functions $[0, 1] \rightarrow \Omega$ and define

$$V_n(\omega, s) = \frac{1}{n} \sum_{i=1}^n \{d^2(X_i(s), \omega(s)) - d^2(X_i(s), \mu_{\oplus}(s))\},$$

$$V(\omega, s) = E\{d^2(X(s), \omega(s)) - d^2(X(s), \mu_{\oplus}(s))\}.$$

Here $\hat{\mu}_{\oplus}(\cdot)$ is the minimizer of $V_n(\omega, s)$ and $\mu_{\oplus}(\cdot)$ is the minimizer of $V(\omega, s)$. We refer to $\hat{\mu}_{\oplus}(\cdot)$, $\mu_{\oplus}(\cdot)$ and $\omega(\cdot)$ as $\hat{\mu}_{\oplus}$, μ_{\oplus} and ω in the following. To derive the rate of convergence of $\hat{\mu}_{\oplus}$, we first obtain a bound for $E\left(\sup_{s \in [0, 1]} \sup_{d_{\infty}(\omega, \mu_{\oplus}) < \delta} |V_n(\omega, s) - V(\omega, s)|\right)$ for small $\delta > 0$, where $d_{\infty}(\omega, \mu_{\oplus}) = \sup_{s \in [0, 1]} d(\omega(s), \mu_{\oplus}(s))$. For this, we define function classes

$$\mathcal{F}_{\delta} = \{f_{\omega, s}(x) = d^2(x(s), \omega(s)) - d^2(x(s), \mu_{\oplus}(s)) : s \in [0, 1], d_{\infty}(\omega, \mu_{\oplus}) < \delta\}. \quad (13)$$

It is easy to see that an envelope function for this class is the constant function $F(x) = 2M\delta$, where M is the diameter of Ω . The L^2 -norm of this envelope function is $\|F\|_2 = 2M\delta$. By Theorem 2.14.2 of Van der Vaart and Wellner (1996) we have

$$E\left(\sup_{s \in [0, 1]} \sup_{d_{\infty}(\omega, \mu_{\oplus}) < \delta} |V_n(\omega, s) - V(\omega, s)|\right) \leq \frac{2M\delta J_{[]} (1, \mathcal{F}_{\delta}, L^2(P))}{\sqrt{n}}, \quad (14)$$

where $J_{[]} (1, \mathcal{F}_{\delta}, L^2(P))$ is the bracketing integral of the function class \mathcal{F}_{δ} ,

$$J_{[]} (1, \mathcal{F}_{\delta}, L^2(P)) = \int_0^1 \sqrt{1 + \log N(\varepsilon \|F\|_2, \mathcal{F}_{\delta}, L^2(P))} d\varepsilon.$$

Here $N(\varepsilon \|F\|_2, \mathcal{F}_{\delta}, L^2(P))$ is the minimum number of balls of radius $\varepsilon \|F\|_2$ required to cover the function class \mathcal{F}_{δ} under the $L^2(P)$ norm. Lemma 1 provides the behavior of the bracketing integral of the function class \mathcal{F}_{δ} , a key step for the proof of Theorem 3.

LEMMA 1. *Under assumptions (A4), (A9) and (A10), it holds for the function class \mathcal{F}_{δ} as defined in (13) that $J_{[]} (1, \mathcal{F}_{\delta}, L^2(P)) = O(\sqrt{\log 1/\delta})$ as $\delta \rightarrow 0$.*

THEOREM 3. *Under assumptions (A3)-(A4) and (A8)-(A11),*

$$\sup_{s \in [0,1]} d(\hat{\mu}_\oplus(s), \mu_\oplus(s)) = O_P \left(\left(\frac{\sqrt{\log n}}{n} \right)^{1/\beta_2} \right).$$

Setting $\hat{D}_i(t) = d(X_i(t), \hat{\mu}_\oplus(t))$, an application is the convergence of the estimated Fréchet scores

$$\hat{\beta}_{ik} = \int_0^1 \hat{D}_i(t) \hat{\phi}_k(t) dt. \quad (15)$$

COROLLARY 2. *Under assumptions (A3)-(A5) and (A8)-(A11),*

$$|\hat{\beta}_{ik} - \beta_{ik}| = O_P \left(n^{-1/2} + \left(\frac{\sqrt{\log n}}{n} \right)^{1/\beta_2} \right).$$

Following widely adopted convention, we assume throughout that true and estimated eigenfunctions are aligned in the sense that $\langle \hat{\phi}_j, \phi_j \rangle \geq 0$, as the scores are identifiable only up to a sign change.

5. Simulations

We illustrate the utility of the proposed methods through simulations for two settings: In the first setting, the space Ω consists of univariate probability distributions equipped with the 2-Wasserstein metric and in the second setting, Ω consists of networks with fixed number of nodes, represented as graph adjacency matrices and equipped with the Frobenius metric.

5.1. Time-varying probability distributions

We generated random samples of sizes $n = 25, 50$ and 100 of ‘distribution’-valued curves on the domain $[0, 1]$, where for each $t \in [0, 1]$, $X_i(t)$ is a normal distribution with mean $\mu_i(t)$ and variance $\sigma_i^2(t)$ with

$$\begin{aligned} \mu_i(t) &= 1 + U_i \phi_1(t) + V_i \phi_3(t), & U_i \sim N(0, 12), & V_i \sim N(0, 1), \\ \sigma_i^2(t) &= 3 + W_i \phi_2(t) + Z_i \phi_3(t), & W_i \sim \sqrt{72} U(0, 1), & Z_i \sim \sqrt{9} U(0, 1), \end{aligned}$$

with $\phi_1(t) = (t^2 - 0.5)/0.3416$, $\phi_2(t) = \sqrt{3}t$, $\phi_3(t) = (t^3 - 0.3571t^2 - 0.6t + 0.1786)/0.0895$ where ϕ_1, ϕ_2 and ϕ_3 are orthonormal on $[0, 1]$. We use the 2-Wasserstein metric for the distribution space Ω . For these specifications, the metric auto-covariance function is

$$C(s, t) = 12\phi_1(s)\phi_1(t) + 6\phi_2(s)\phi_2(t) + 1.75\phi_3(s)\phi_3(t), \quad (16)$$

and $\phi_1(\cdot), \phi_2(\cdot)$ and $\phi_3(\cdot)$ are the first 3 eigenfunctions.

We applied the proposed method to estimate the metric auto-covariance operator to the simulated data and obtained its eigenvalues and eigenfunctions. Denoting the estimated metric auto-covariance surface and the estimated j -th eigenvalue and eigenfunction obtained at the k^{th} simulation run by $\hat{C}_k(s, t)$, respectively $\hat{\lambda}_{j,k}$ and $\hat{\phi}_{j,k}$, we computed mean integrated squared errors (MISE)

$$\begin{aligned} \text{MISE}(C) &= \frac{1}{100} \sum_{k=1}^{100} \int_0^1 \int_0^1 \left(\hat{C}_k(s, t) - C(s, t) \right)^2 ds dt, \\ \text{MISE}(\phi_j) &= \frac{1}{100} \sum_{k=1}^{100} \int_0^1 \left(\hat{\phi}_{j,k}(s) - \phi_j(s) \right)^2 ds, \quad \text{MISE}(\lambda_j) = \frac{1}{100} \sum_{k=1}^{100} \left(\hat{\lambda}_{j,k} - \lambda_j \right)^2. \end{aligned}$$

Figure 1 shows the true and estimated metric auto-covariance surfaces and their eigenfunctions for one randomly chosen simulation run for $n = 25$ and $n = 100$. We find that the proposed method has negligible bias as sample size increases. The MISEs are reported in Table 7 and seen to decrease with increasing sample sizes.

Table 1: Mean Squared Errors for the estimators of the metric auto-covariance kernel C and the eigenfunctions ϕ_1, ϕ_2 in dependence on sample size when the functional random objects are distributions.

n	C	ϕ_1	ϕ_2	ϕ_3	λ_1	λ_2	λ_3
25	12.2709	1.2841	1.5351	1.5202	10.8634	7.8471	3.5798
50	8.6598	0.0504	0.0201	0.0030	0.9748	0.6482	0.3680
100	4.0697	0.0158	0.0084	0.0047	0.1239	0.0607	0.0314

To illustrate the nature of the simulated random density trajectories, four density-valued random functions that are part of a sample of density-valued random functions as generated in one Monte Carlo run are displayed in Figure 2, reflecting variation in means and variances of the Gaussian distributions as a function of time for the four selected subjects. The estimated object FPCs, i.e. the Fréchet integrals of the object curves along the first two eigenfunctions, from one Monte Carlo run are in Figure 19 for sample size 50. Here the first object FPCs reflect variation in location of the distributions and the second object FPCs variation in the variance of the distributions, which is what we expect in view of how these data were generated. The object FPCs are found to be useful for discovering the underlying modes of variation for distributions as functional random objects.

5.2. Time-varying networks

In each iteration, we generated random samples of sizes $n = 25, 50$ and 100 of time varying random networks with 10 nodes each in the time interval $[0, 1]$. For generating the edge

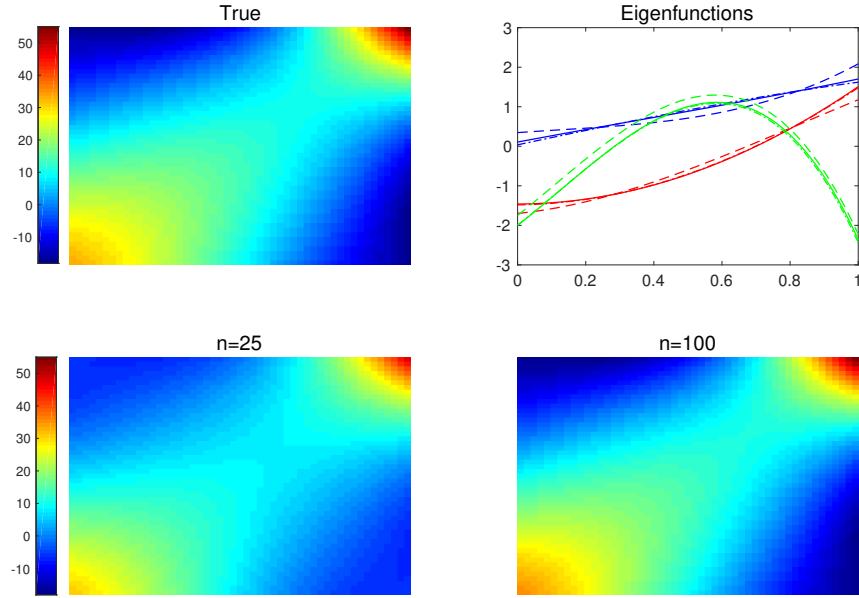


Fig. 1: True (top left) and estimated (bottom panels) metric auto-covariance contour plots (10) for simulation with distributions as functional random objects. The red curves in the top right panel correspond to the first, blue curves to the second and green curves to the third eigenfunction, depicting true (solid) and estimated (dashed for $n = 25$ and dot-dashed for $n = 100$) eigenfunctions.

weights, we followed the model described below. We assumed that the network has two communities, the first five nodes belonging to one community and the second five nodes to the other one. For each fixed time t , the edge weights within each community and also those between the communities are the same, where the latter are smaller than the within community edge weights. Formally, if $p_{1,i}(t)$, $p_{2,i}(t)$ and $p_{12,i}(t)$ denote the edge weight at time $t \in [0, 1]$ for the first community, the second community and between communities, for the i^{th} network valued curve we generated

$$p_{1,i}(t) = 0.5 + U_i\phi_1(t) + V_i\phi_3(t), \quad p_{2,i}(t) = 0.5 + W_i\phi_2(t) + Z_i\phi_3(t), \quad p_{12,i}(t) = 0.1. \quad (17)$$

Here the U_i , V_i , W_i and Z_i were generated from the uniform distributions $U(0, 0.4)$, $U(0, 0.1)$, $U(0, 0.3)$ and $U(0, 0.1)$, respectively. The functions $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ are orthonormal polynomials derived from Jacobi polynomials $P_n^{(\alpha, \beta)}(x)$ (Totik, 2005), which are classical orthogonal polynomials for $\alpha, \beta > 1$. They are orthogonal with respect to the basis $(1+x)^\beta(1-x)^\alpha$ on $[-1, 1]$. With a suitable change of basis, one can obtain a version of the Jacobi polynomials on $[0, 1]$ which are orthonormal with respect to the

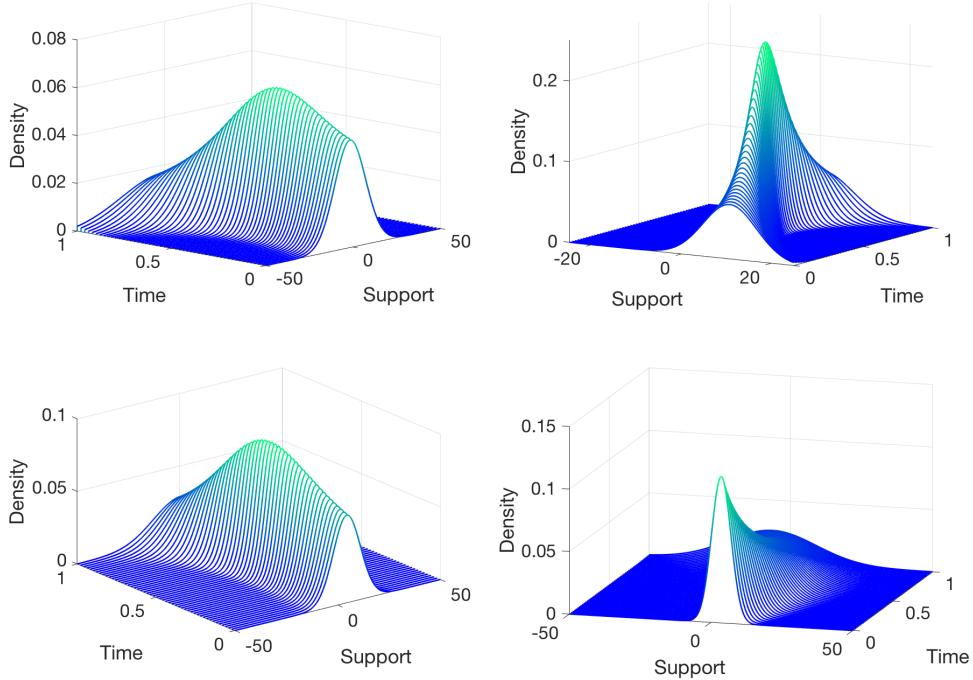


Fig. 2: Four randomly chosen observations of density-valued trajectories, selected from the sample of distributions generated by one of the Monte Carlo runs. The densities are plotted as a function of time.

weight function $x^\beta(1-x)^\alpha$ on $[0, 1]$. We selected $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ as

$$\phi_j(t) = \frac{(P_{2j}^{(4,3)}(2t-1))t^{1.5}(1-t)^2}{[\int_0^1 (P_{2j}^{(4,3)}(2t-1))^2 t^{1.5}(1-t)^2 dt]^{1/2}} \quad \text{for } j = 1, 2, 3.$$

The weighted networks are represented as graph adjacency matrices with the Frobenius metric. Here the true metric auto-covariance function is

$$C(s, t) = 0.266\phi_1(s)\phi_1(t) + 0.15\phi_2(s)\phi_2(t) + 0.0417\phi_3(s)\phi_3(t), \quad (18)$$

and $\phi_1(\cdot)$, $\phi_2(\cdot)$ and $\phi_3(\cdot)$ are the first three eigenfunctions.

We estimated the metric auto-covariance operator from the simulated data and obtained its eigenfunctions for different sample sizes. Figure 4 displays the true and estimated metric auto-covariance surfaces and corresponding eigenfunctions for one randomly chosen simulation run for $n = 25$ and $n = 100$. The MISEs were computed as described for the previous simulation setting and are reported in Table 2. They decrease with increasing sample sizes. The proposed method is seen to work very well.

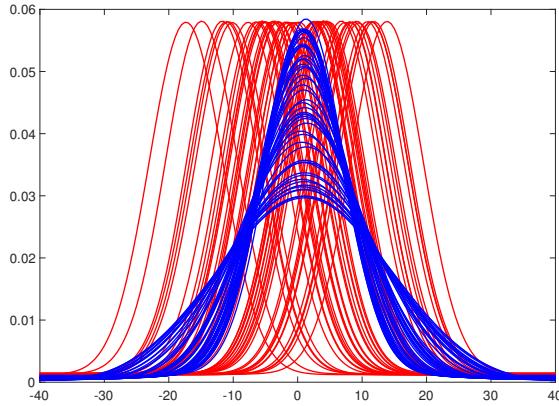


Fig. 3: Estimated Fréchet integrals (11) for the first (red) and second (blue) eigenfunction for the sample elements, for simulated time-varying probability distributions.

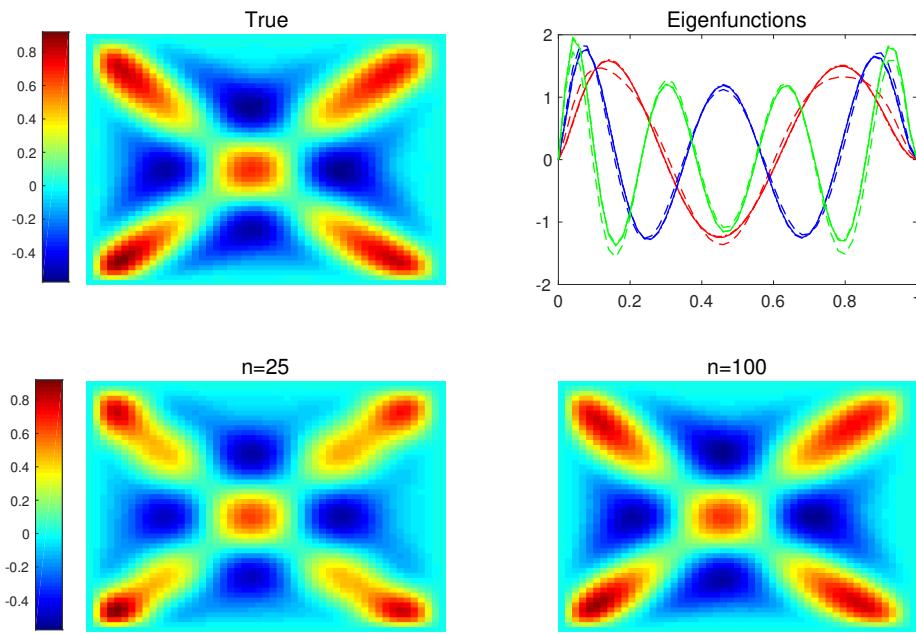


Fig. 4: True (top left) and estimated (bottom panels) metric auto-covariance contour plots (10) for simulation with networks as functional random objects. The red curves in the top right panel correspond to the first, blue curves to the second and green curves to the third eigenfunction, depicting true (solid) and estimated (dashed for $n = 25$ and dot-dashed for $n = 100$) eigenfunctions.

Table 2: Mean Integrated Squared Errors for the estimators of the metric auto-covariance kernel C and eigenfunctions/eigenvalues $\phi_j, \lambda_j, j = 1, 2, 3$, in dependence on sample size for samples of functional random objects that correspond to time-varying networks.

n	C	ϕ_1	ϕ_2	ϕ_3	λ_1	λ_2	λ_3
25	0.0039	0.0039	0.0017	0.0007	0.0025	0.0010	0.0007
50	0.0017	0.0093	0.0046	0.0021	0.0007	0.0003	0.0001
100	0.0010	0.0130	0.0063	0.0028	0.0001	0.0001	0.0001

The object FPCs were obtained using Fréchet integrals (11). For visualization they are presented as a “networks.mov” in the supplementary materials. In the movie the leftmost plot corresponds to Fréchet integrals for the first eigenfunction which, as expected due to the true model, shows variation only in the edge weights of the first community. The middle plot corresponds to Fréchet integrals for the second eigenfunction and indicates variation only in the edge weights of the second community. The rightmost plot corresponds to Fréchet integrals for the third eigenfunction and where variation in both the first and the second community edge weights can be discerned.

6. Data Applications

6.1. Mortality data

The Human Mortality Database provides life table data differentiated by gender and is available at www.mortality.org. Currently the mortality database contains life table data for 37 countries spanning over 5 decades. One can obtain histograms from life tables and smooth these with local least squares to obtain estimated probability density functions for age at death. We carried this out for the age interval [0, 80]. The mortality data can then be viewed as samples of time varying univariate probability distributions, for a sample of 32 countries, where the time axis corresponds to calendar years between 1960 and 2009 and the observation made at each calendar year for each country corresponds to the age at death distribution for that year. We included the 32 countries which had complete records over the entire calendar period. For each country and year, we used the Hades package available at <https://stat.ucdavis.edu/hades/> for smoothing the histograms and used bandwidth= 2 to obtain the age-at-death densities. For illustration, the time-varying age at death distributions represented as density functions for the age interval [0, 80] and indexed by calendar year are displayed for four selected countries, USA, Ukraine, Russia and Portugal, for males in Figure 5 and for females in Figure 6.

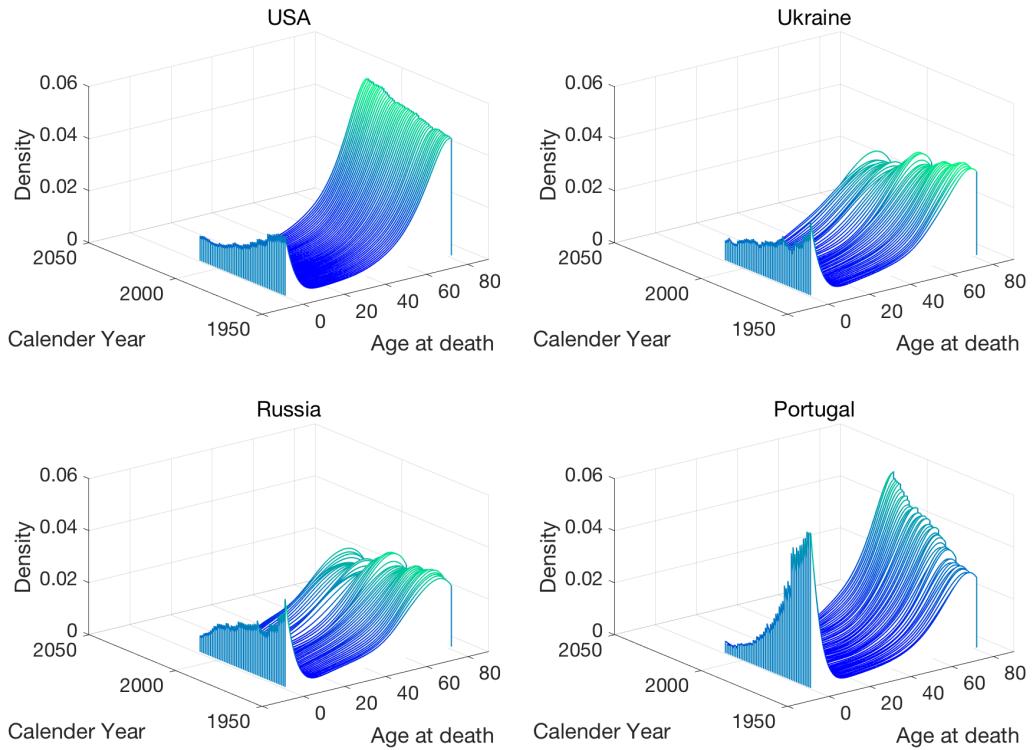


Fig. 5: Time-varying age at death density functions for the age interval $[0, 80]$ for males in USA, Ukraine, Russia and Portugal.

Choosing the 2-Wasserstein metric for the probability distributions space, the estimated metric auto-covariance surfaces for males and females can be inspected in Figure 7 and the eigenfunctions of the corresponding auto-covariance operators in Figure 8. The auto-covariance functions and eigenfunctions indicate that there are systematic differences between males and females.

The resulting object FPCs, i.e. the Fréchet integrals, are illustrated in Figure 9 for the first two eigenfunctions. The object FPCs are distributions that are represented as densities for males and females. The Eastern European countries included in the data base, namely, Belarus, Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Poland, Slovakia, Ukraine, Russia and Estonia underwent major political upheaval due to the end of Communist rule in these regions during the period between the late 1980s and early 1990. This is reflected in clear distinctions between the Eastern European countries (red) and the rest (blue) in the Fréchet integrals for the males but much less so for the females, which indicates that particularly male mortality was affected by the political upheavals.

The sample Fréchet mean function at a particular calendar year corresponds to the

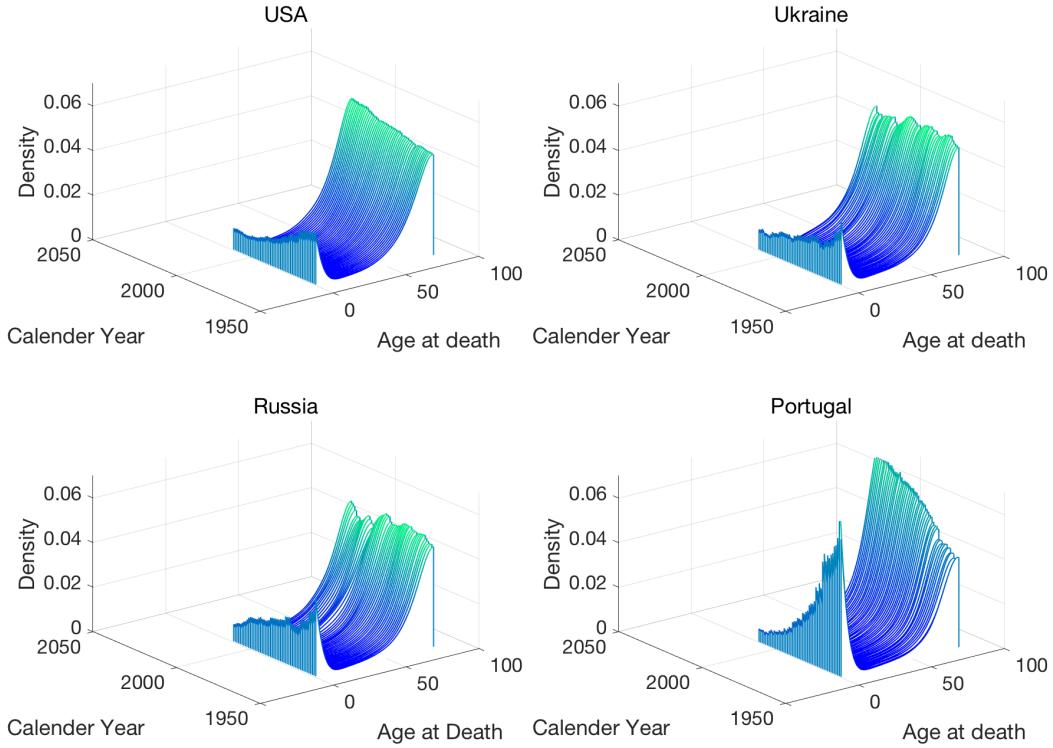


Fig. 6: Time-varying age at death density functions for the age interval $[0, 80]$ for females in USA, Ukraine, Russia and Portugal.

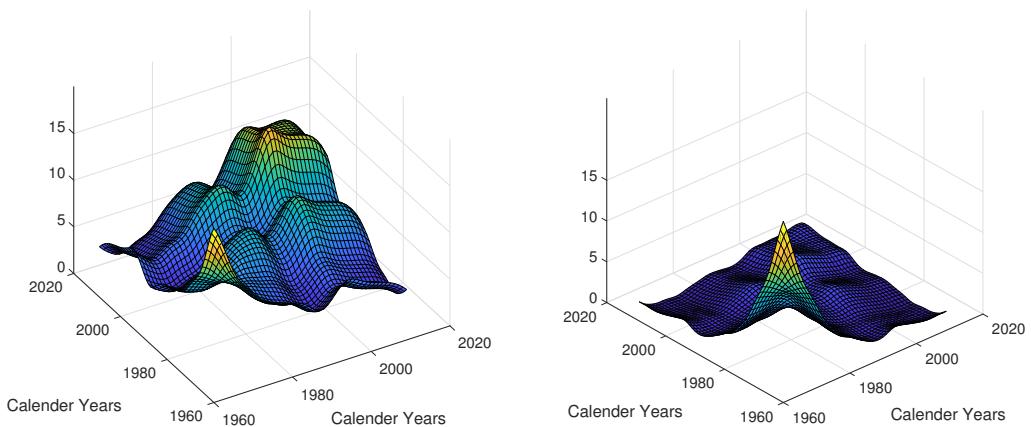


Fig. 7: Estimated metric auto-covariance surfaces (10) for males (left) and females (right) for densities as functional random objects, as obtained for the mortality data.

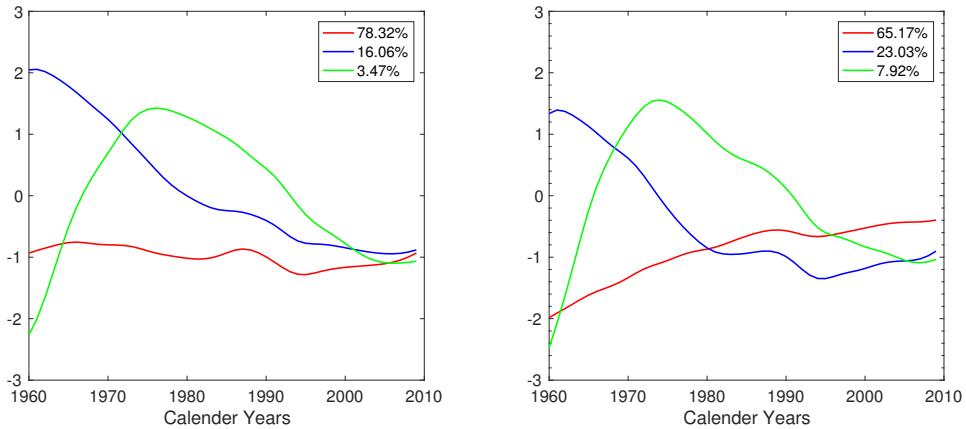


Fig. 8: The eigenfunctions of the estimated metric auto-covariance surface for males (left) and for females (right) for the mortality data.

sample average of the quantile functions of the different countries at that calendar year and is illustrated in the movies “mean_males.mov” and “mean_females.mov” in the supplementary materials. Figure 10 illustrates the scalar FPCs, i.e. the Fréchet scores for the second eigenfunction plotted against the Fréchet scores for the first eigenfunction for males and females. Russia is an outlier for the first eigenfunction for males and Portugal is an outlier for the second eigenfunction, even though it does not belong to the above list of Eastern European countries. One could speculate that this might be related to the fact Portugal in 1974 moved to a democratic government after four decades of authoritarian dictatorship. Figures 5 and 6 suggest higher infant mortality for both males and females in Portugal during the earlier era. Another interesting observation is that the order of outliers is reversed for females, as Russia turns out to be an outlier for females for the second eigenfunction and Portugal for the first. The plots of the Fréchet scores against each indicate that there are clear distinctions between the two groups of countries and Portugal.

6.2. Time-varying networks for New York taxi data

New York City Taxi and Limousine Commission (NYC TLC) provides records on pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts for yellow and green taxis available at http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml. The time resolution of these data is in the order of seconds. Of interest are networks which represent how many people traveled between places of interest and the

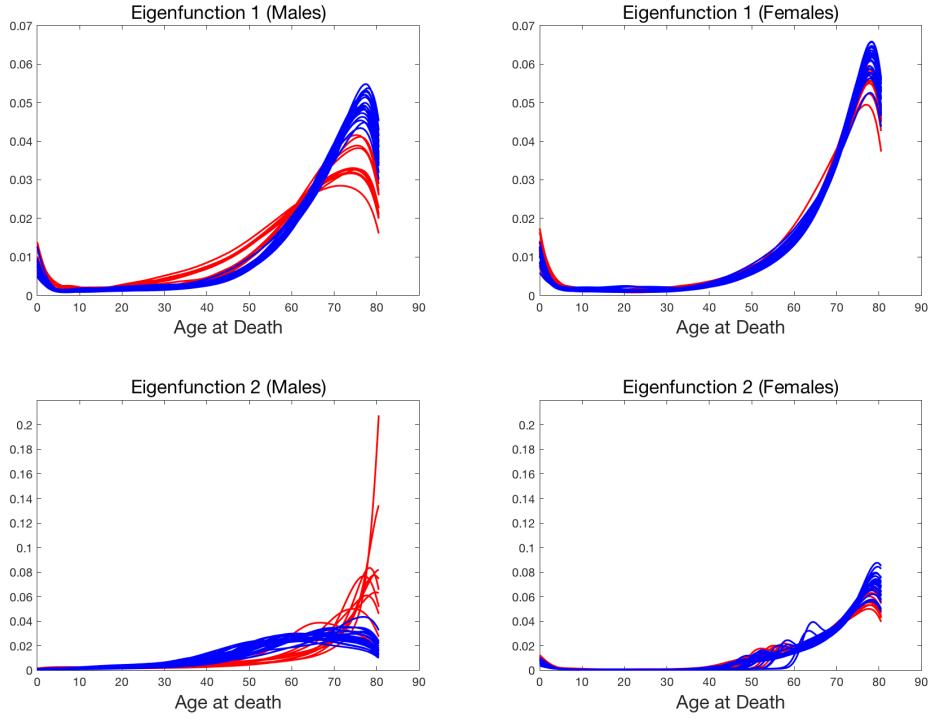


Fig. 9: Fréchet integrals (11) for the first eigenfunction for males (top left) and females (top right) and for the second eigenfunction for males (bottom left) and females (bottom right) for the mortality data. The red curves correspond to the Eastern European countries, namely, Belarus, Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Poland, Slovakia, Ukraine, Russia and Estonia and the blue curves to the other countries.

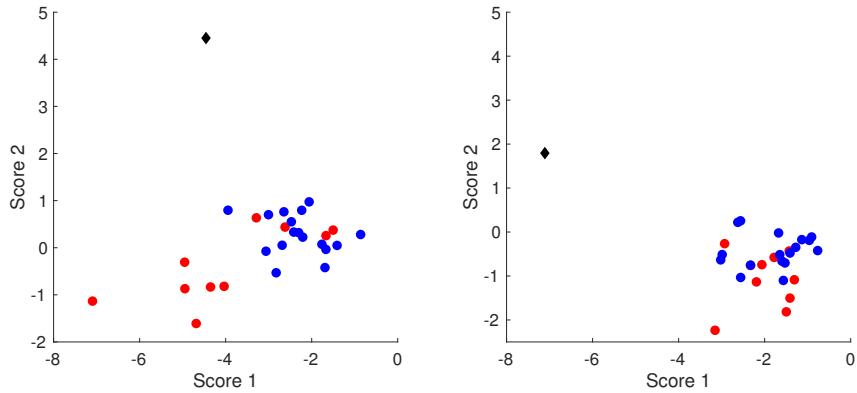


Fig. 10: Fréchet scores (15) for the first and the second eigenfunctions plotted against each other for males (left) and for females (right) for the mortality data. The black diamond corresponds to Portugal, the red dots to the Eastern European countries, namely, Belarus, Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Poland, Slovakia, Ukraine, Russia and Estonia and the blue dots to the other countries.

evolution of these networks during a typical day. To study this, we constructed samples of time-varying networks where the sample elements are the recordings for each day in the year 2016. Three days (23 and 24 Jan and 13 March) were excluded from the study due to incomplete records.

We focus on the Manhattan area, which has the highest traffic and split the area according to the provided location shape files into 10 zones, which form the regions of interest. Details about the zones are in section S6.1 of the online supplement. Yellow taxis provide the predominant taxi service in Manhattan. We divided each day into 20 minute intervals, and for each interval constructed a network with nodes corresponding to the 10 selected zones and edge weights representing the number of people who traveled between the zones connecting the edges within the 20 minute interval. The edge weights were normalized by the maximum edge weight for each day so that they lie in $[0, 1]$. We thus have a time-varying network for each of the 363 days in 2016 for which complete records are available, where the time points where the network-valued functions are evaluated correspond to the 20 minute intervals of a 24 hour day. The observations at each time point correspond to a 10 dimensional graph adjacency matrix which characterizes the network between the 10 zones of Manhattan for the particular 20 minute interval.

We choose the Frobenius metric as metric between the graph adjacency matrices. The sample Fréchet mean function at a particular time point therefore corresponds to the sample average of the graph adjacency matrices of 363 networks corresponding to different days for that time point. It is illustrated in the movie “mean_NY.mov” in the supplementary materials. Figure 11 illustrates the estimated auto-covariance function and associated eigenfunctions. The plots of the Fréchet scores for second, third and fourth eigenfunction against the scores for the first eigenfunction can be found in Figure 12, where the blue dots correspond to Mondays to Thursdays, the green dots to Fridays and the red dots to Saturdays and Sundays. Several interesting patterns emerge: Weekdays and weekends form clearly distinguishable clusters. Special holidays show similar patterns as weekends. Several outliers can be identified using the projection scores for the eigenfunctions, which turn out to be special days: For the first eigenfunction, the outliers correspond to New Year day and November 6, 2016, which is the day when daylight saving ends. March 13, 2016 is the day the daylight saving begins but was excluded as it did not have complete records. For the second eigenfunction, an outlying point is Independence Day, 4 July 2016, and for the third eigenfunction, 14 February 2016 which is Valentine’s day. Another day that stands out is 18 September 2016. On

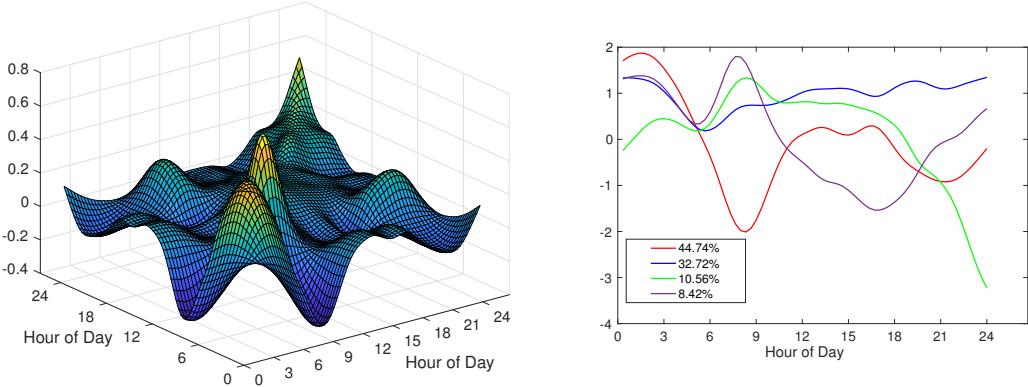


Fig. 11: Estimated metric auto-covariance surface (10) (left) and the corresponding eigenfunctions (right) for the New York taxi data, viewed as time-varying networks.

further investigating it was found that between 17 to 19 September 2016 three bombs exploded and several unexploded ones were found in the New York metropolitan area (https://en.wikipedia.org/wiki/2016_New_York_and_New_Jersey_bombings).

We then repeated the analysis separately for three groups of days, namely the weekdays Monday-Thursday (group 1), Fridays and weekends (group 2) and holidays (group 3). We present the results in Figure 17 (Supplement) and in several movies whose descriptions can be found in sections S3 and S5 of the online supplement.

6.3. World trade data

The Center for International Data at University of California, Davis (<http://cid.econ.ucdavis.edu/nberus.html>) provides detailed documentation of United Nations trade data for the years 1962-2000. The dataset, publicly available at www.nber.org, contains bilateral trade data during this time period for several commodities and countries. We studied the time period 1970 to 1999 for 46 actively trading countries and the 26 most common types of commodities. The list of chosen countries and commodities can be found in the section S6.2 of the online supplement. For each country, commodity and year, we represent current trade as the ratio of the amount of total trade, i.e. import export value (in thousands of US dollars), to the amount of total trade recorded for the same commodity and country in the year 2000, yielding a 26-dimensional vector of trade ratios.

Viewing the countries as sampling units, we obtain for each country and calendar

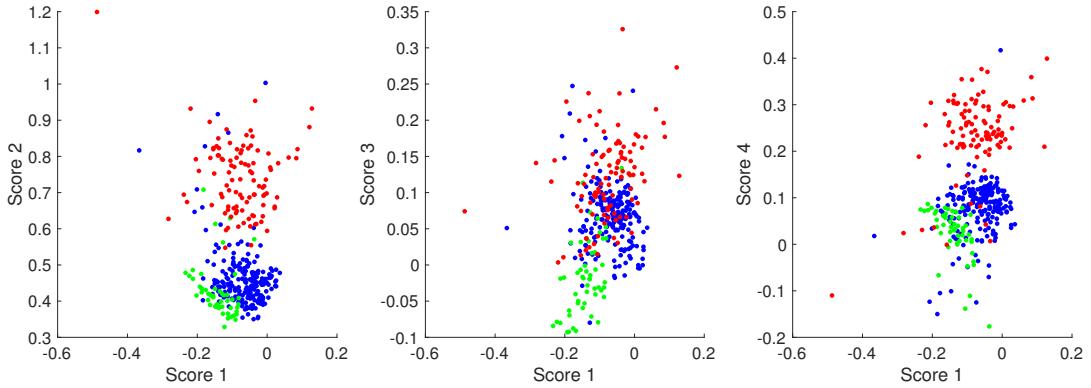


Fig. 12: Fréchet scores (15) for the second (left), third (middle) and fourth (right) eigenfunctions in the y -axis plotted against Fréchet scores for the first eigenfunction in the x -axis, for the New York taxi data. The blue dots correspond to Mondays to Thursdays, the green dots to Fridays and the red dots to Saturdays and Sundays.

year t a 26×26 -dimensional raw covariance matrix of commodities trade ratios as $\tilde{\Sigma}(t) = (Q(t) - \bar{Q}(t))(Q(t) - \bar{Q}(t))^T$, where $Q(t)$ is the country-specific 26-dimensional vector of commodities trade ratios for year t and the mean vector \bar{Q} is obtained as cross-sectional average over all 46 countries. These raw time-varying raw covariances were then smoothed using local Fréchet regression with a Gaussian kernel (Petersen and Müller, 2019a; Petersen et al., 2019) to obtain samples of smooth time-varying 26-dimensional covariance matrices between the components of commodities trade for each of the 46 countries over the time period 1970 to 2000, yielding time-varying covariance matrices over the time period 1970 to 2000 as functional random objects.

When adopting the Frobenius metric, the sample Fréchet mean function at calendar year t corresponds to the sample average of the smoothed covariance across 46 countries for year t . Figure 13a illustrates the estimated metric auto-covariance function and Figure 13b its eigenfunctions. The metric auto-covariance and its eigenfunctions provide insights about world trade patterns over the time period 1970 to 1999. The first eigenfunction represents increased variability due to overall expansion in world trade over the years from 1970 to 1999. The slope of the first eigenfunction is more gradual before 1985 but increases sharply starting 1985, stagnates a little around 1990 and then again picks up. This can be connected to the boom in world trade towards the last decade of the new millennium. The second eigenfunction corresponds to a contrast before 1990 and after 1990. The peak in the second eigenfunction between 1980-1985 could be related to a major economic downturn caused by recession affecting several countries in the dataset during

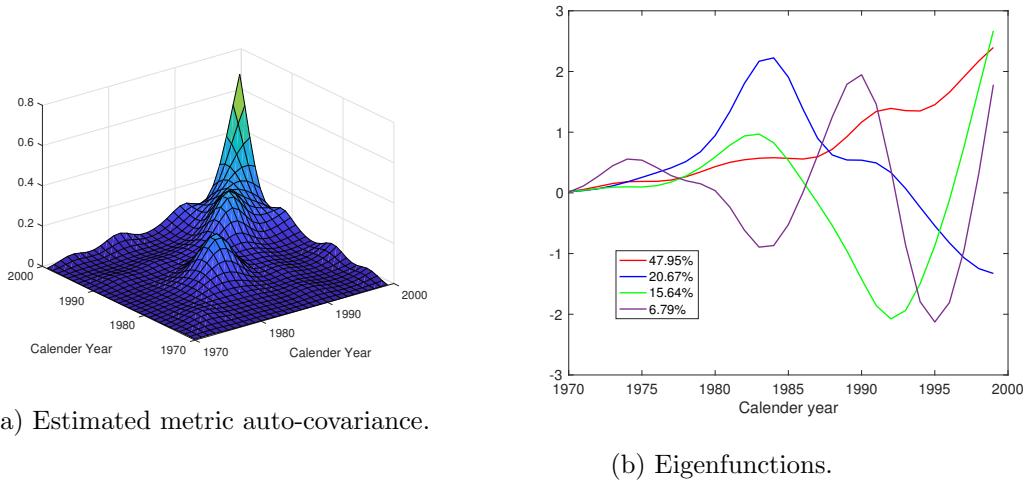


Fig. 13: Estimated metric auto-covariance surface (10) (left) and corresponding first four eigenfunctions (right) for the trade data.

the early 1980s. The recession began in USA in 1981 and continued till 1982 and affected many of the developed Western countries. The third eigenfunction captures effects of the early 1990s recession, which compared to the 1980s recession was much milder.

In Figure 14, the Fréchet scores for the first four eigenfunctions are plotted against each other. Thailand and Egypt have high Fréchet scores for the first eigenfunction and Saudi Arabia ranks the highest for the second eigenfunction. Chile, Israel, Hong Kong and Bulgaria turn out to figure prominently in the third eigenfunction. Further visualization can be found in section S4 of the online supplement, including a movie described in section S5 of the online supplement that demonstrates the object FPCs.

7. Discussion

We propose an extension of functional data methods to the case of functional random objects. The basis of our approach is metric covariance, a novel covariance measure for paired metric space valued data. Eigenfunctions of the metric covariance operator for time-varying object data aid in creating a version of object FPCA, where the object FPCs in the metric space Ω are obtained as Fréchet integrals, a general and versatile concept. Alternatively, components of variation can be quantified by Fréchet scores, which are real numbers. For the precursor problem, where one has non-functional time-varying object data, i.e. one has observations for just one random object function over time, methods for metric-space valued regression have been considered previously (Steinke et al., 2010; Faraway, 2014; Petersen and Müller, 2019a), often under the special assumption that the

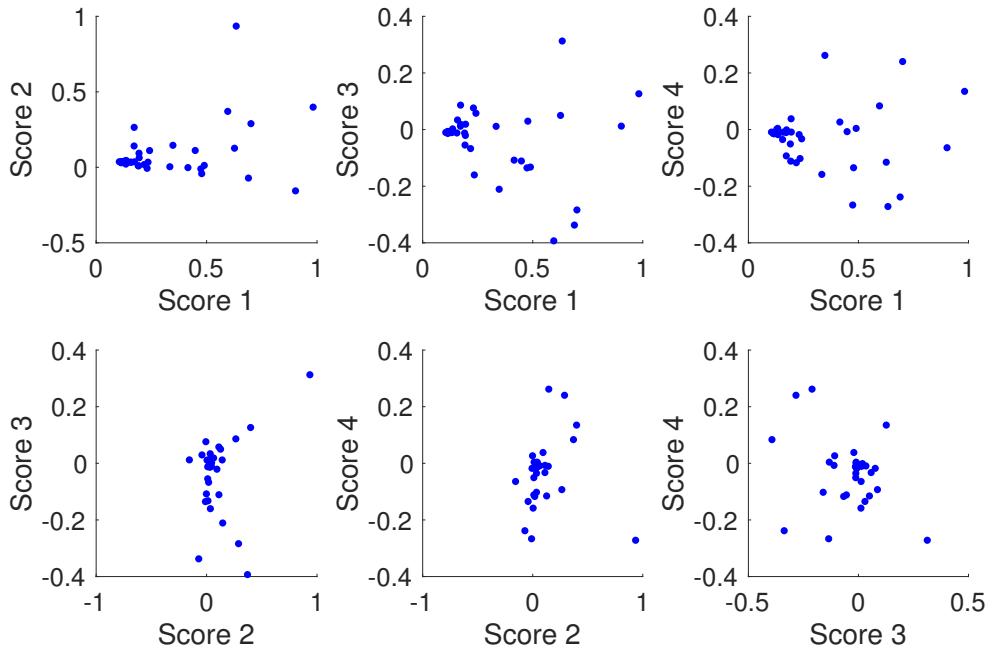


Fig. 14: Fréchet scores for different eigenfunctions plotted against each other for the trade data.

regression responses are on a Riemannian manifold (Shi et al., 2009; Fletcher, 2013; Hinkle et al., 2012; Su et al., 2012; Yuan et al., 2012; Cornea et al., 2017). However the more general object function case, which is characterized by samples of random functions that are object-valued, is considerably more challenging, as the absence of a linear structure in the object space both globally and locally imposes serious limitations on the methods that can be applied.

The tools we propose here for functional random objects, namely metric covariance, the metric auto-covariance operator and its eigenfunctions, the Fréchet integrals and the Fréchet scores make it possible to obtain compact summaries, visualizations and interpretations of the observed samples of time-varying object data that in themselves are highly complex and difficult to quantify. These tools can provide insights into the patterns of variability of the object trajectories, as we demonstrated in the simulations and data examples. The quantification of functional random objects can also be used for other tasks. For example the object FPCs that we introduce reside in the object space and can serve as responses for a regression model, where predictors are Euclidean vectors and responses are random object trajectories, which are summarized by these object FPCs. Implementing such a regression approach is analogous to the principal component

approach for function-to-function regression (Yao et al., 2005b). Various regression models can then be implemented through Fréchet regression (Petersen and Müller, 2019a). For the case where functional random objects feature as predictors in a regression setting, one can employ the vector of Fréchet scores that summarize each random object trajectory as predictors. The ensuing regression, classification and clustering models will be interesting topics for future research.

A core challenge that one faces when modeling and analyzing samples of random object trajectories is that in contrast to the situation for real-valued processes, one cannot expect to represent object-valued processes in terms of an analogue to the Karhunen-Loève expansion, due to the lack of a linear structure in the object space Ω . In some special cases such expansions are possible, for example through a transformation method, whenever random objects can be transformed to a linear space, as exemplified for the case of objects that are probability distributions (Petersen and Müller, 2016) or for the case of Riemannian manifold-valued objects (Dai and Müller, 2018). Apart from such special cases, it is an open problem whether more general useful representations of functional random objects can be found. Another open problem is inference for such data, e.g., comparing two groups or testing for structural features of auto-covariance. Here the metric auto-covariance operator that we introduce in this paper and also the Fréchet mean function could prove useful for the extension of tests that have been considered for real-valued functional data (for some recent examples, see Aston et al., 2017; Constantinou et al., 2017; Chen and Lynch, 2018; Choi and Reimherr, 2018). These and many other open problems in this area indicate that there is ample potential for future research.

Acknowledgements

We wish to thank the reviewers for their comprehensive and helpful comments that led to numerous improvements and new perspectives. This research was supported by NSF grant DMS-1712864 and NIH grant 5UH3OD023313.

References

- Anirudh, R., Turaga, P., Su, J. and Srivastava, A. (2017) Elastic functional coding of Riemannian trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 922–936.
- Arcones, M. A. and Giné, E. (1993) Limit theorems for U-processes. *The Annals of Probability*, 1494–1542.

- Aston, J. A., Pigoli, D. and Tavakoli, S. (2017) Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics*, **45**, 1431–1461.
- Belkin, M. and Niyogi, P. (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, 585–591.
- Berrendero, J., Justel, A. and Svarc, M. (2011) Principal components for multivariate functional data. *Computational Statistics and Data Analysis*, **55**, 2619–2634.
- Bigot, J., Gouet, R., Klein, T. and López, A. (2013) Geodesic PCA in the Wasserstein space. *arXiv preprint arXiv:1307.7721*.
- (2017) Geodesic PCA in the Wasserstein space by convex PCA. *Annales de l’Institut Henri Poincaré B: Probability and Statistics*, **53**, 1–26.
- Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley, New York.
- Bosq, D. (2000) *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer-Verlag.
- Castro, P. E., Lawton, W. H. and Sylvestre, E. A. (1986) Principal modes of variation for processes with continuous sample curves. *Technometrics*, **28**, 329–337.
- Chen, K., Delicado, P. and Müller, H.-G. (2017) Modeling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society, Series B (Theory and Methodology)*, **79**, 177–196.
- Chen, K. and Lynch, B. (2018) A test of weak separability for multi-way functional data, with application to brain connectivity studies. *Biometrika*, **105**, 815–831.
- Chen, K. and Müller, H.-G. (2012) Modeling repeated functional observations. *Journal of the American Statistical Association*, **107**, 1599–1609.
- Chiou, J.-M., Chen, Y.-T. and Yang, Y.-F. (2014) Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, **24**, 1571–1596.
- Chiou, J.-M. and Li, P.-L. (2007) Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B*, **69**, 679–699.
- Chiou, J.-M., Yang, Y.-F. and Chen, Y.-T. (2016) Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, **146**, 301–312.

- Choi, H. and Reimherr, M. (2018) A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**, 239–260.
- Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014) Multivariate functional half-space depth. *Journal of the American Statistical Association*, **109**, 411–423.
- Constantinou, P., Kokoszka, P. and Reimherr, M. (2017) Testing separability of space-time functional processes. *Biometrika*, **104**, 425–437.
- Cornea, E., Zhu, H., Kim, P. and Ibrahim, J. G. (2017) Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 463–482.
- Dai, X. and Müller, H.-G. (2018) Principal component analysis for functional data on Riemannian manifolds and spheres. *Annals of Statistics*, **46**, 3334–3361.
- Dai, X., Müller, H.-G. and Yao, F. (2017) Optimal Bayes classifiers for functional data and density ratios. *Biometrika*, **79**, 545–560.
- 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**, 136–154.
- Dong, J. J., Wang, L., Gill, J. and Cao, J. (2018) Functional principal component analysis of glomerular filtration rate curves after kidney transplant. *Statistical Methods in Medical Research*, **27**, 3785–3796.
- Dryden, I. L., Koloydenko, A. and Zhou, D. (2009) Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, **3**, 1102–1123.
- Dubey, P. and Müller, H.-G. (2017) Fréchet analysis of variance for random objects. *arXiv preprint arXiv:1710.02761*.
- Dubin, J. A. and Müller, H.-G. (2005) Dynamical correlation for multivariate longitudinal data. *Journal of the American Statistical Association*, **100**, 872–881.
- Faraway, J. J. (2014) Regression for non-Euclidean data using distance matrices. *Journal of Applied Statistics*, **41**, 2342–2357.

- Fletcher, P. T. (2013) Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, **105**, 171–185.
- Fréchet, M. (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, **10**, 215–310.
- Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. and Kolaczyk, E. D. (2017) Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, **11**, 725–750.
- Hinkle, J., Muralidharan, P., Fletcher, P. T. and Joshi, S. (2012) Polynomial regression on Riemannian manifolds. In *Computer Vision–ECCV 2012*, 1–14. Springer.
- Horvath, L. and Kokoszka, P. (2012) *Inference for Functional Data with Applications*. New York: Springer.
- Hsing, T. and Eubank, R. (2015) *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.
- Jacques, J. and Preda, C. (2014) Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, **71**, 92–106.
- Jakobsen, M. E. (2017) Distance covariance in metric spaces: Non-parametric independence testing in metric spaces. *arXiv preprint arXiv:1706.03490*.
- Jones, M. C. and Rice, J. A. (1992) Displaying the important features of large collections of similar curves. *The American Statistician*, **46**, 140–145.
- Kleffe, J. (1973) Principal components of random variables with values in a separable Hilbert space. *Statistics: A Journal of Theoretical and Applied Statistics*, **4**, 391–406.
- Kruskal, J. (1964) Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115–129.
- Lin, L., St. Thomas, B., Zhu, H. and Dunson, D. B. (2017) Extrinsic local regression on manifold-valued data. *Journal of the American Statistical Association*, **112**, 1261–1273.
- Liu, S., Zhou, Y., Palumbo, R. and Wang, J.-L. (2016) Dynamical correlation: A new method for quantifying synchrony with multivariate intensive longitudinal data. *Psychological Methods*, **21**, 291–308.
- Lyons, R. (2013) Distance covariance in metric spaces. *Annals of Probability*, **41**, 3284–3305.

- Monnig, N. D. and Meyer, F. G. (2018) The resistance perturbation distance: A metric for the analysis of dynamic networks. *Discrete Applied Mathematics*, **236**, 347–386.
- Müller, H.-G., Stadtmüller, U. and Yao, F. (2006) Functional variance processes. *Journal of the American Statistical Association*, **101**, 1007–1018.
- Nolan, D. and Pollard, D. (1987) U-processes: rates of convergence. *The Annals of Statistics*, 780–799.
- (1988) Functional limit theorems for U-processes. *The Annals of Probability*, 1291–1298.
- Opgen-Rhein, R. and Strimmer, K. (2006) Inferring gene dependency networks from genomic longitudinal data: A functional data approach. *REVSTAT - Statistical Journal*, **4**, 53–65.
- Park, S. Y. and Staicu, A.-M. (2015) Longitudinal functional data analysis. *Stat*, **4**, 212–226.
- Petersen, A., Deoni, S. and Müller, H.-G. (2019) Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *arXiv preprint arXiv:1806.09690 Annals of Applied Statistics, in press*.
- Petersen, A. and Müller, H.-G. (2016) Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika*, **103**, 103–120.
- Petersen, A. and Müller, H.-G. (2016) Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, **44**, 183–218.
- (2019a) Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics*, **47**, 691–719.
- (2019b) Wasserstein covariance for multiple random densities. *arXiv:1812.07694 Biometrika, in press*.
- Pigoli, D., Aston, J. A., Dryden, I. L. and Secchi, P. (2014) Distances and inference for covariance operators. *Biometrika*, **101**, 409–422.
- Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. Springer Series in Statistics. New York: Springer, second edn.

- Robert, P. and Escoufier, Y. (1976) A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **25**, 257–265.
- Schoenberg, I. J. (1938) Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, **44**, 522–536.
- Seguy, V. and Cuturi, M. (2015) Principal geodesic analysis for probability measures under the optimal transport metric. In *Advances in Neural Information Processing Systems*, 3312–3320.
- Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013) Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics*, **41**, 2263–2291.
- Shi, X., Styner, M., Lieberman, J., Ibrahim, J. G., Lin, W. and Zhu, H. (2009) Intrinsic regression models for manifold-valued data. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, 192–199. Springer.
- Steinke, F., Hein, M. and Schölkopf, B. (2010) Nonparametric regression between general Riemannian manifolds. *SIAM Journal on Imaging Sciences*, **3**, 527–563.
- Su, J., Dryden, I. L., Klassen, E., Le, H. and Srivastava, A. (2012) Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. *Image and Vision Computing*, **30**, 428–442.
- Suarez, A. J. and Ghosal, S. (2016) Bayesian clustering of functional data using local features. *Bayesian Analysis*, **11**, 71–98.
- Székely, G. J. and Rizzo, M. L. (2017) The energy of data. *Annual Review of Statistics and Its Application*, **4**, 447–479.
- Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007) Measuring and testing dependence by correlation of distances. *Annals of Statistics*, **35**, 2769–2794.
- Tavakoli, S., Pigoli, D., Aston, J. A. and Coleman, J. (2016) A spatial modeling approach for linguistic object data: Analysing dialect sound variations across Great Britain. *arXiv preprint arXiv:1610.10040*.
- Totik, V. (2005) Orthogonal polynomials. *Surveys in Approximation Theory*, **1**, 70–125.

- Van der Vaart, A. and Wellner, J. (1996) *Weak Convergence and Empirical Processes*. Springer, New York.
- Verbeke, G., Fieuws, S., Molenberghs, G. and Davidian, M. (2014) The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, **23**, 42–59.
- Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016) Functional data analysis. *Annual Review of Statistics and its Application*, **3**, 257–295.
- Yao, F., Müller, H.-G. and Wang, J.-L. (2005a) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590.
- (2005b) Functional linear regression analysis for longitudinal data. *Annals of Statistics*, **33**, 2873–2903.
- Yuan, Y., Zhu, H., Lin, W. and Marron, J. (2012) Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74**, 697–719.
- Zhou, L., Huang, J. and Carroll, R. (2008) Joint modelling of paired sparse functional data using principal components. *Biometrika*, **95**, 601–619.

Online Supplement

S1. Proofs

Proof of Proposition 1

It is clear that $C(s, t)$ is a symmetric function. To prove that it is nonnegative definite we need to show that for any positive integer k ,

$$\sum_{i=1}^k \sum_{j=1}^k a_i a_j C(s_i, s_j) \geq 0,$$

for any a_1, a_2, \dots, a_k in \mathbb{R} and s_1, s_2, \dots, s_k in $[0, 1]$. Since d^2 is a semimetric of negative type, by Proposition 3 in Sejdinovic et al. (2013) there exists a Hilbert space \mathcal{H} and an injective map $f : \Omega \rightarrow H$ such that $d^2(\omega_1, \omega_2) = \|f(\omega_1) - f(\omega_2)\|_{\mathcal{H}}^2$. We therefore have that for $x, x', y, y' \in \Omega$,

$$\begin{aligned} & d^2(x, y') + d^2(x', y) - 2d^2(x, y) \\ &= \|f(x) - f(y')\|_{\mathcal{H}}^2 + \|f(x') - f(y)\|_{\mathcal{H}}^2 - 2\|f(x) - f(y)\|_{\mathcal{H}}^2 \\ &= \|f(y')\|_{\mathcal{H}}^2 + \|f(x')\|_{\mathcal{H}}^2 - \|f(y)\|_{\mathcal{H}}^2 - \|f(x)\|_{\mathcal{H}}^2 + 4\langle f(x), f(y) \rangle_{\mathcal{H}} \\ &\quad - 2\langle f(x), f(y') \rangle_{\mathcal{H}} - 2\langle f(x'), f(y) \rangle_{\mathcal{H}}, \end{aligned}$$

which implies that for i.i.d copies (X, Y) and (X', Y')

$$C(X, Y) = \frac{1}{4}E(4\langle f(X), f(Y) \rangle_{\mathcal{H}} - 2\langle f(X), f(Y') \rangle_{\mathcal{H}} - 2\langle f(X'), f(Y) \rangle_{\mathcal{H}}).$$

Let $H_i = f(X(s_i))$ and $H'_i = f(X'(s_i))$ for $i = 1, 2, \dots, k$ where X' is an i.i.d copy of X .

Then

$$C(s_i, s_j) = \frac{1}{4}E(4\langle H_i, H_j \rangle_{\mathcal{H}} - 2\langle H_i, H'_j \rangle_{\mathcal{H}} - 2\langle H'_i, H_j \rangle_{\mathcal{H}}),$$

which leads to

$$\begin{aligned} & \sum_{i=1}^k \sum_{j=1}^k a_i a_j C(s_i, s_j) \\ &= \frac{1}{4}E\left(4\left\langle \sum_{i=1}^k a_i H_i, \sum_{j=1}^k a_j H_j \right\rangle_{\mathcal{H}} - 2\left\langle \sum_{i=1}^k a_i H_i, \sum_{j=1}^k a_j H'_j \right\rangle_{\mathcal{H}} - 2\left\langle \sum_{i=1}^k a_i H'_i, \sum_{j=1}^k a_j H_j \right\rangle_{\mathcal{H}}\right) \\ &\geq \frac{1}{4}E\left(4\left\|\sum_{i=1}^k a_i H_i\right\|_{\mathcal{H}}^2\right) - 2E\left(\left\|\sum_{i=1}^k a_i H_i\right\|_{\mathcal{H}}\right)E\left(\left\|\sum_{j=1}^k a_j H'_j\right\|_{\mathcal{H}}\right) \\ &\quad - 2E\left(\left\|\sum_{i=1}^k a_i H'_i\right\|_{\mathcal{H}}\right)E\left(\left\|\sum_{j=1}^k a_j H_j\right\|_{\mathcal{H}}\right) \\ &= \text{Var}\left(\left\|\sum_{i=1}^k a_i H_i\right\|_{\mathcal{H}}\right). \end{aligned}$$

The last step follows from the Cauchy-Schwarz inequality. This completes the proof.

Proof of Proposition 2

- (a) For any $\gamma > 0$, by (I1) there exists $\delta > 0$ such that, whenever $|t_1 - t_2| < \delta$,

$$\sup_{\omega \in \Omega} |H(\omega, t_1) - H(\omega, t_2)| < \gamma.$$

For any partition \mathcal{P} as described above such that $\epsilon_{\mathcal{P}} < \delta$ we find

$$\begin{aligned} & \sup_{\omega \in \Omega} |I_{\mathcal{P}}(\omega) - I(\omega)| \\ &= \sup_{\omega \in \Omega} \left| \sum_{j=0}^{k-1} H(\omega, t_j) \Delta_j - \int_0^1 H(\omega, t) dt \right| \\ &= \sup_{\omega \in \Omega} \left| \sum_{j=0}^{k-1} \int_{x_j}^{x_{j+1}} \{H(\omega, t_j) - H(\omega, t)\} dt \right| \\ &= \sum_{j=0}^{k-1} \int_{x_j}^{x_{j+1}} \sup_{\omega \in \Omega} |H(\omega, t_j) - H(\omega, t)| dt < \gamma, \end{aligned}$$

which completes the proof for part (a).

- (b) Observe that

$$\begin{aligned} & |I(\sum_{\mathcal{P}, \oplus} S\phi) - I(\int_{\oplus} S\phi)| \\ &\leq |I(\sum_{\mathcal{P}, \oplus} S\phi) - I_{\mathcal{P}}(\sum_{\mathcal{P}, \oplus} S\phi) + I_{\mathcal{P}}(\sum_{\mathcal{P}, \oplus} S\phi) - I(\int_{\oplus} S\phi)| \\ &\leq \sup_{\omega \in \Omega} |I(\omega) - I_{\mathcal{P}}(\omega)| + |\min_{\omega \in \Omega} I_{\mathcal{P}}(\omega) - \min_{\omega \in \Omega} I(\omega)| \\ &\leq 2 \sup_{\omega \in \Omega} |I(\omega) - I_{\mathcal{P}}(\omega)|, \end{aligned}$$

and therefore by part (a), $|I(\sum_{\mathcal{P}, \oplus} S\phi) - I(\int_{\oplus} S\phi)| \rightarrow 0$ as $\epsilon_{\mathcal{P}} \rightarrow 0$.

Now assume that $\lim_{\epsilon_{\mathcal{P}} \rightarrow 0} d(\sum_{\mathcal{P}, \oplus} S\phi, \int_{\oplus} S\phi) \neq 0$. Then there must exist a sequence of partitions $\{\mathcal{P}_n\}$ and a $\gamma > 0$ such that $\epsilon_{\mathcal{P}_n} \rightarrow 0$ but $d(\sum_{\mathcal{P}_n, \oplus} S\phi, \int_{\oplus} S\phi) \geq \gamma$. For this sequence of partitions we observe that,

$$|I(\sum_{\mathcal{P}_n, \oplus} S\phi) - I(\int_{\oplus} S\phi)| \geq \inf_{d(\omega, \int_{\oplus} S\phi) > \gamma} |I(\omega) - I(\int_{\oplus} S\phi)| > 0$$

and therefore $\lim_{\epsilon_{\mathcal{P}_n} \rightarrow 0} |I(\sum_{\mathcal{P}_n, \oplus} S\phi) - I(\int_{\oplus} S\phi)| \geq \inf_{d(\omega, \int_{\oplus} S\phi) > \gamma} |I(\omega) - I(\int_{\oplus} S\phi)| > 0$ by (I2), which is a contradiction to part (a). Therefore the assumption that $\lim_{\epsilon_{\mathcal{P}} \rightarrow 0} d(\sum_{\mathcal{P}, \oplus} S\phi, \int_{\oplus} S\phi) \neq 0$ cannot be true, which completes the proof for part(b).

- (c) Let $\delta > 0$ be such that whenever $\epsilon_{\mathcal{P}} < \delta$, it holds that $d(\sum_{\mathcal{P}, \oplus} S\phi, \int_{\oplus} S\phi) < \nu$. Assume that $\lim_{\epsilon_{\mathcal{P}} \rightarrow 0} h^{1/\beta}(\epsilon_{\mathcal{P}}) d(\sum_{\mathcal{P}, \oplus} S\phi, \int_{\oplus} S\phi) \neq 0$. Then there exists a sequence

of partitions $\{\mathcal{P}_n\}$ and a $\gamma > 0$ such that $\epsilon_{\mathcal{P}_n} < \delta$, while $d(\sum_{\mathcal{P}_n, \oplus} S\phi, \int_{\oplus} S\phi) \geq \frac{\gamma}{h^{1/\beta}(\epsilon_{\mathcal{P}})}$. For this sequence of partitions we observe that

$$\begin{aligned} h(\epsilon_{\mathcal{P}_n})|I(\sum_{\mathcal{P}_n, \oplus} S\phi) - I(\int_{\oplus} S\phi)| \\ \geq h(\epsilon_{\mathcal{P}_n})|\inf_{\frac{\gamma}{h^{1/\beta}(\epsilon_{\mathcal{P}_n})} \leq d(\omega, \int_{\oplus} S\phi) < \nu} I(\omega) - I(\int_{\oplus} S\phi)| \\ \geq \frac{Ch(\epsilon_{\mathcal{P}_n})\gamma^{\beta}}{h(\epsilon_{\mathcal{P}_n})} \end{aligned}$$

by (I3). Therefore $\lim_{\epsilon_{\mathcal{P}_n} \rightarrow 0} h(\epsilon_{\mathcal{P}_n})|I(\sum_{\mathcal{P}_n, \oplus} S\phi) - I(\int_{\oplus} S\phi)| \geq C\gamma^{\beta} > 0$, which results in a contradiction and completes the proof for part (c).

Proof of Lemma 3

Consider $t \in [0, 1]$ and a sequence $\{t_n\} \in [0, 1]$ such that $t_n \rightarrow t$. We aim to show that $d(\mu_{\oplus}(t_n), \mu_{\oplus}(t)) \rightarrow 0$.

Observe that almost surely continuous sample curves on the compact interval $[0, 1]$ are uniformly continuous and since Ω is bounded, by bounded convergence, for all $t \in [0, 1]$ and sequences $\{t_n\} \in [0, 1]$ such that $t_n \rightarrow t$, there exists a $\delta > 0$ for every $\epsilon > 0$ such that whenever $|t_n - t| < \delta$, it holds that $E(d(X(t_n), X(t))) < \epsilon$ for all but finitely many n . For processes $E(d^2(\omega, X(t)))$,

$$|E(d^2(\omega, X(t_n))) - E(d^2(\omega, X(t)))| \leq 2D E(d(X(t_n), X(t))),$$

and therefore, given any $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $|t_n - t| < \delta$, one has $\sup_{\omega \in \Omega} |E(d^2(\omega, X(t_n))) - E(d^2(\omega, X(t)))| < \epsilon$. This implies that

$$\begin{aligned} & |E(d^2(\mu_{\oplus}(t_n), X(t))) - E(d^2(\mu_{\oplus}(t), X(t)))| \\ & \leq |E(d^2(\mu_{\oplus}(t_n), X(t))) - E(d^2(\mu_{\oplus}(t_n), X(t_n)))| \\ & \quad + |E(d^2(\mu_{\oplus}(t_n), X(t_n))) - E(d^2(\mu_{\oplus}(t), X(t)))| \\ & \leq \sup_{\omega \in \Omega} |E(d^2(\omega, X(t_n))) - E(d^2(\omega, X(t)))| + |\min_{\omega \in \Omega} E(d^2(\omega, X(t_n))) - \min_{\omega \in \Omega} E(d^2(\omega, X(t)))| \\ & \leq 2 \sup_{\omega \in \Omega} |E(d^2(\omega, X(t_n))) - E(d^2(\omega, X(t)))| < 2\epsilon. \end{aligned}$$

Assume $d(\mu_{\oplus}(t_n), \mu_{\oplus}(t)) \not\rightarrow 0$. Then one can find an $\eta > 0$ such that for any $\delta > 0$, there exists a subsequence $\{t_{n_k}\}$ of $\{t_n\}$ for which $|t_{n_k} - t| < \delta$ but $d(\mu_{\oplus}(t_{n_k}), \mu_{\oplus}(t)) \geq \eta$. Then

by (A3),

$$\begin{aligned} & |E(d^2(\mu_{\oplus}(t_{n_k}), X(t))) - E(d^2(\mu_{\oplus}(t), X(t)))| \\ & \geq \inf_{d(\omega, \mu_{\oplus}(t)) > \eta} |E(d^2(\omega, X(t))) - E(d^2(\mu_{\oplus}(t), X(t)))| \\ & > 0. \end{aligned}$$

This leads to a contradiction for $\epsilon = |\inf_{d(\omega, \mu_{\oplus}(t)) > \eta} E(d^2(\omega, X(t))) - E(d^2(\mu_{\oplus}(t), X(t)))|/2$, thus completing the proof.

Proof of Theorem 1

Denoting the usual L^2 norm by $\|\cdot\|_2$, observe that for $s_1, t_1, s_2, t_2 \in [0, 1]$, one has

$$\begin{aligned} & |f_{s_1, t_1}(x, y) - f_{s_2, t_2}(x, y)| \\ & \leq 4M\{d(x(s_1), x(s_2)) + d(x(t_1), x(t_2)) + d(y(s_1), y(s_2)) + d(y(t_1), y(t_2))\} \\ & \leq 4M(G(x) + G(y))(|s_1 - s_2|^{\alpha} + |t_1 - t_2|^{\alpha}), \end{aligned}$$

implying

$$\|f_{s_1, t_1} - f_{s_2, t_2}\|_2 \leq 8M\|G\|_2(|s_1 - s_2|^{\alpha} + |t_1 - t_2|^{\alpha}).$$

Observe that for any $0 < u < 1$, if we take $|s_1 - s_2| < (\frac{u}{16})^{\frac{1}{\alpha}}$ and $|t_1 - t_2| < (\frac{u}{16})^{\frac{1}{\alpha}}$, then $\|f_{s_1, t_1}(X) - f_{s_2, t_2}(Y)\|_2 \leq Mu\|G\|_2$. Therefore, with s_1, s_2, \dots, s_K and t_1, t_2, \dots, t_L forming $(\frac{u}{4})^{\frac{1}{\alpha}}$ -nets for $[0, 1]$ with metric $|\cdot|$, the brackets $[f_{s_i, t_j} \pm Mu\|G\|_2]$ cover the function class \mathcal{F} (Van der Vaart and Wellner, 1996) and are of length $2Mu\|G\|_2$. This implies

$$N_{[]} (2Mu\|G\|_2, \mathcal{F}, L^2(P \otimes P)) \leq N \left(\left(\frac{u}{4} \right)^{\frac{1}{\alpha}}, [0, 1], |\cdot| \right)^2,$$

where $N_{[]}(\varepsilon, \mathcal{F}, L^2(P))$ is the bracketing number, which is the minimum number of ε -brackets needed to cover \mathcal{F} , where an ε -bracket is composed of pairs of functions $[l, u]$ such that $\|l - u\|_2 < \varepsilon$, and N is the covering number. Hence for any $\epsilon > 0$, for some constant $K > 0$,

$$N_{[]}(\epsilon, \mathcal{F}, L^2(P \otimes P)) \leq K\epsilon^{-2/\alpha} < \infty.$$

The result now follows from Theorem 4.10 of Arcones and Giné (1993), observing

$$\begin{aligned} & \int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L^2(P \otimes P))} d\varepsilon \\ & \leq \varepsilon \sqrt{\log K} + \int_0^1 \sqrt{-\frac{2}{\alpha} \log \varepsilon} d\varepsilon \\ & = \varepsilon \sqrt{\log K} + \sqrt{\frac{2}{\alpha}} \Gamma\left(\frac{3}{2}\right) < \infty. \end{aligned}$$

Proof of Corollary 1

For any fixed j , Lemma 4.3 in Bosq (2000) gives $|\hat{\lambda}_j - \lambda_j| \leq \sup_{s,t \in [0,1]} |\hat{C}(s,t) - C(s,t)|$.

Uniform mapping then implies

$$|\hat{\lambda}_j - \lambda_j| = O_P(1/\sqrt{n}).$$

Under assumption (A5), $\sup_{s \in [0,1]} |\hat{\phi}_j(s) - \phi_j(s)| \leq 2\sqrt{2}\delta_j^{-1} \sup_{s,t \in [0,1]} |\hat{C}(s,t) - C(s,t)|$,

and therefore

$$\sup_{s \in [0,1]} |\hat{\phi}_j(s) - \phi_j(s)| = O_P(1/\sqrt{n}),$$

which completes the proof.

Proof of Theorem 2

Since

$$\left| \int_0^1 \hat{\phi}(t)dt - \int_0^1 \phi(t)dt \right| \leq \sup_{s \in [0,1]} |\hat{\phi}(s) - \phi(s)| = O_P(1/\sqrt{n}), \quad (19)$$

for all sufficiently large n , $\left| \int_0^1 \hat{\phi}(t)dt \right| \geq \left| \int_0^1 \phi(t)dt \right|/2$, and since $\left| \int_0^1 \phi(t)dt \right| \neq 0$,

$$\begin{aligned} & \sup_{s \in [0,1]} |\hat{\phi}^*(s) - \phi^*(s)| \\ & \leq 2 \frac{\sup_{s \in [0,1]} |\hat{\phi}(s)| \left| \int_0^1 \hat{\phi}(t)dt - \int_0^1 \phi(t)dt \right| + \left| \int_0^1 \hat{\phi}(t)dt \right| \sup_{s \in [0,1]} |\hat{\phi}(s) - \phi(s)|}{\left| \int_0^1 \phi(t)dt \right|^2} \\ & = O_P(1/\sqrt{n}). \end{aligned}$$

As a direct consequence,

$$\begin{aligned} & \sup_{\omega \in \Omega} \left| \int_0^1 d^2(\omega, X_i(t)) \hat{\phi}^*(t)dt - \int_0^1 d^2(\omega, X_i(t)) \phi^*(t)dt \right| \\ & \leq M^2 \sup_{s \in [0,1]} |\hat{\phi}^*(s) - \phi^*(s)| = O_P(1/\sqrt{n}), \end{aligned} \quad (20)$$

whence

$$\begin{aligned} & P \left(d(\hat{\psi}_{\oplus}^{ik}, \psi_{\oplus}^{ik}) > \epsilon \right) \\ & \leq P \left(\left| \int_0^1 d^2(\hat{\psi}_{\oplus}^{ik}, X_i(t)) \phi^*(t)dt - \int_0^1 d^2(\psi_{\oplus}^{ik}, X_i(t)) \phi^*(t)dt \right| > c_{\epsilon} \right) \end{aligned} \quad (21)$$

$$\leq P \left(\sup_{\omega \in \Omega} \left| \int_0^1 d^2(\omega, X_i(t)) \hat{\phi}^*(t)dt - \int_0^1 d^2(\omega, X_i(t)) \phi^*(t)dt \right| > \frac{c_{\epsilon}}{2} \right), \quad (22)$$

which implies that $d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) = o_P(1)$ by equation (20). Here (22) follows from (21) using the fact that

$$\begin{aligned} & \left| \int_0^1 d^2(\hat{\psi}_\oplus^{ik}, X_i(t)) \phi^*(t) dt - \int_0^1 d^2(\psi_\oplus^{ik}, X_i(t)) \phi^*(t) dt \right| \\ & \leq \left| \int_0^1 d^2(\hat{\psi}_\oplus^{ik}, X_i(t)) (\phi^*(t) - \hat{\phi}^*(t)) dt \right| \\ & + \left| \inf_{\omega \in \Omega} \int_0^1 d^2(\omega, X_i(t)) \hat{\phi}^*(t) dt - \inf_{\omega \in \Omega} \int_0^1 d^2(\omega, X_i(t)) \phi^*(t) dt \right| \\ & \leq 2 \sup_{\omega \in \Omega} \left| \int_0^1 d^2(\omega, X_i(t)) \hat{\phi}^*(t) dt - \int_0^1 d^2(\omega, X_i(t)) \phi^*(t) dt \right|. \end{aligned}$$

From assumption (A7),

$$\begin{aligned} & P \left(n^{1/(2\beta_1)} d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) > 2^L \right) \\ & \leq P \left(\frac{2^L}{n^{1/(2\beta_1)}} < d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) < \nu' \right) + P \left(d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) \geq \nu' \right) \\ & = P \left(\sup_{\omega \in \Omega} \left| \int_0^1 d^2(\omega, X_i(t)) \hat{\phi}^*(t) dt - \int_0^1 d^2(\omega, X_i(t)) \phi^*(t) dt \right| > \frac{2^{\beta_1 L}}{n^{1/2}} \right) + P \left(d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) \geq \nu' \right) \\ & \leq P \left(\sqrt{n} \sup_{s \in [0,1]} |\hat{\phi}^*(s) - \phi^*(s)| > \frac{2^{\beta_1 L}}{M^2} \right) + P \left(d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) \geq \nu' \right). \end{aligned}$$

Therefore $P \left(n^{1/(2\beta_1)} d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) > 2^L \right)$ can be made as small as possible by choosing sufficiently large n and L , using equation (20) and the fact that $d(\hat{\psi}_\oplus^{ik}, \psi_\oplus^{ik}) = o_P(1)$, thus completing the proof.

Proof of Proposition 4

By Theorem 1.5.4 in [Van der Vaart and Wellner \(1996\)](#), it suffices to show asymptotic equicontinuity of the processes $Z_n(s) = d(\hat{\mu}_\oplus(s), \mu_\oplus(s))$, i.e. for any $\theta > 0$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P \left(\sup_{|s-t| < \delta} |Z_n(s) - Z_n(t)| > 2\theta \right) = 0, \quad (23)$$

in addition to the pointwise convergence of $Z_n(s)$, i.e. for all $s \in [0, 1]$ it holds that

$$Z_n(s) = o_P(1). \quad (24)$$

We observe that equation (24) follows from Theorem 1 in [Petersen and Müller \(2019a\)](#). To establish equation (23), by Lemma 3 and as $|Z_n(s) - Z_n(t)| \leq d(\mu_\oplus(s), \mu_\oplus(t)) + d(\hat{\mu}_\oplus(s), \hat{\mu}_\oplus(t))$, it suffices to show that

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P \left(\sup_{|s-t| < \delta} d(\hat{\mu}_\oplus(s), \hat{\mu}_\oplus(t)) > \theta \right) = 0. \quad (25)$$

To show equation (25), suppose that $d(\hat{\mu}_\oplus(s), \hat{\mu}_\oplus(t)) > \theta$ with $s, t \in [0, 1]$.

Step 1. Since the functions are continuous and the domain is compact, it holds that almost surely $\sup_{|s-t|<\delta} d(X(t), X(s)) \rightarrow 0$ as $\delta \rightarrow 0$. By the boundedness of the metric and dominated convergence,

$$\lim_{\delta \rightarrow 0} E(\sup_{|s-t|<\delta} d(X(t), X(s))) \rightarrow 0. \quad (26)$$

Now (26) implies that for any $a > 0$,

$$\begin{aligned} P & \left(\sup_{|s-t|<\delta} \sup_{\omega \in \Omega} \left| \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \omega) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(s), \omega) \right| > a \right) \\ & \leq \frac{2M E \left(\sup_{|s-t|<\delta} \frac{1}{n} \sum_{i=1}^n d(X_i(t), X_i(s)) \right)}{a} \\ & \leq \frac{2M E(\sup_{|s-t|<\delta} d(X(t), X(s)))}{a} \rightarrow 0 \quad \text{as } \delta \rightarrow 0, \end{aligned}$$

where $M = \text{diam}(\Omega)$.

Step 2. We observe that

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \hat{\mu}_\oplus(s)) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \hat{\mu}_\oplus(t)) \right| \\ & \leq \left| \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \hat{\mu}_\oplus(s)) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(s), \hat{\mu}_\oplus(s)) \right| \\ & \quad + \left| \frac{1}{n} \sum_{i=1}^n d^2(X_i(s), \hat{\mu}_\oplus(s)) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \hat{\mu}_\oplus(t)) \right| \\ & \leq 2 \sup_{\omega \in \Omega} \left| \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \omega) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(s), \omega) \right|. \end{aligned}$$

Step 3. Now we find

$$\begin{aligned} & P \left(\sup_{|s-t|<\delta} d(\hat{\mu}_\oplus(s), \hat{\mu}_\oplus(t)) > \theta \right) \\ & \leq P(B \cap A_n) + P(B \cap A_n^C) \\ & \leq P \left(\sup_{|s-t|<\delta} 2 \sup_{\omega \in \Omega} \left| \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \omega) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(s), \omega) \right| \geq \tau(S) \right) + P(B \cap A_n^C). \end{aligned} \quad (27)$$

where $A_n = \{\sup_{|s-t|<\delta} |\frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \hat{\mu}_\oplus(s)) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \hat{\mu}_\oplus(t))| \geq \tau(S)\}$, $B = \{\sup_{|s-t|<\delta} d(\hat{\mu}_\oplus(s), \hat{\mu}_\oplus(t)) > S\}$ in (27) with $\tau(S)$ as defined in (A8). The last step follows from (27) using Step 2. From Step 1, choosing $a = \frac{\tau(S)}{2}$,

$$\lim_{\delta \rightarrow 0} P \left(\sup_{|s-t|<\delta} 2 \sup_{\omega \in \Omega} \left| \frac{1}{n} \sum_{i=1}^n d^2(X_i(t), \omega) - \frac{1}{n} \sum_{i=1}^n d^2(X_i(s), \omega) \right| \geq \tau(S) \right) = 0$$

for any n , and from (A1), $\lim_{n \rightarrow \infty} P(B \cap A_n^C) = 0$. This completes the proof.

Proof of Lemma 1

Define functions $f_{\omega,s}(x) = d^2(x(s), \omega(s)) - d^2(x(s), \mu_{\oplus}(s))$. We find

$$\begin{aligned} & |f_{\omega_1,s_1}(x) - f_{\omega_2,s_2}(x)| \\ & \leq |f_{\omega_1,s_1}(x) - f_{\omega_1,s_2}(x)| + |f_{\omega_1,s_2}(x) - f_{\omega_2,s_2}(x)| \\ & \leq 2M(2d(x(s_1), x(s_2)) + d(\mu_{\oplus}(s_1), \mu_{\oplus}(s_2)) + d(\omega_1(s_1), \omega_1(s_2)) + d(\omega_1(s_2), \omega_2(s_2))). \end{aligned}$$

Note that $d(\omega_1(s_2), \omega_2(s_2)) \leq d(\omega_1(s_2), \mu_2(s_2)) + d(\omega_1(s_2), \mu_2(s_2))$. By assumptions (A4) and (A9), it holds that almost surely,

$$\begin{aligned} & |f_{\omega_1,s_1}(X) - f_{\omega_2,s_2}(X)| \leq 4M[G(X)|s_1 - s_2|^{\alpha} + H_{\delta}|s_1 - s_2|^{\nu_{\delta}}] \\ & + 4M[d(\omega_1(s_2), \mu_2(s_2)) + d(\omega_1(s_2), \mu_2(s_2))], \end{aligned}$$

which implies that

$$\begin{aligned} & \|f_{\omega_1,s_1} - f_{\omega_2,s_2}\|_2 \\ & \leq 4M[||G||_2|s_1 - s_2|^{\alpha} + H_{\delta}|s_1 - s_2|^{\nu_{\delta}} + d(\omega_1(s_2), \mu_2(s_2)) + d(\omega_1(s_2), \mu_2(s_2))]. \end{aligned}$$

It follows that for some $0 < u < 1$, if we take $|s_1 - s_2| < (\frac{u}{4})^{\frac{1}{V}}$ with $V = \min(\alpha, \nu_{\delta})$ and ω_1, ω_2 such that $d_{\infty}(\omega_1, \mu) < \frac{u}{8}$ and $d_{\infty}(\omega_2, \mu) < \frac{u}{8}$, then $\|f_{\omega_1,s_1} - f_{\omega_2,s_2}\|_2 < Mu(||G||_2 + H_{\delta} + 1)$. Therefore if s_1, s_2, \dots, s_K is a $(\frac{u}{4})^{\frac{1}{V}}$ -net for $[0, 1]$ with metric $|\cdot|$ and $\omega_1, \omega_2, \dots, \omega_L$ is a $\frac{u}{8}$ -net for $B_{\delta}(\mu(\cdot))$ with metric d_{∞} , the brackets $[f_{s_i, \omega_j} \pm Mu(||G||_2 + H_{\delta} + 1)]$ cover the function class \mathcal{F}_{δ} and are of length $2Mu(||G||_2 + H_{\delta} + 1)$. We conclude that

$$N_{[]} (2Mu(||G||_2 + H_{\delta} + 1), \mathcal{F}_{\delta}, L^2(P)) \leq N\left(\left(\frac{u}{4}\right)^{\frac{1}{V}}, [0, 1], |\cdot|\right) N\left(\frac{u}{8}, B_{\delta}(\mu_{\oplus}(\cdot)), d_{\infty}\right).$$

Applying [Van der Vaart and Wellner \(1996\)](#) (page 84), for any function class \mathcal{F} and for any r ,

$$N(\varepsilon, \mathcal{F}, L^r(P)) \leq N_{[]} (2\varepsilon, \mathcal{F}, L^r(P)),$$

so that for appropriate constants $K_1, K_2, C > 0$,

$$\begin{aligned} & \log N(2M\delta\varepsilon, \mathcal{F}_{\delta}, L^2(P)) \\ & \leq \log N\left(K_1(\varepsilon\delta)^{1/V}, [0, 1], |\cdot|\right) + \log N(K_2\varepsilon\delta, B_{\delta}(\mu_{\oplus}(\cdot)), d_{\infty}) \\ & \leq \log\left(C\left(\frac{1}{\varepsilon\delta}\right)^{1/V}\right) + \log N(K_2\varepsilon\delta, B_{\delta}(\mu_{\oplus}(\cdot)), d_{\infty}). \end{aligned}$$

Observe that $\log N(K_2\varepsilon\delta, B_{\delta}(\mu_{\oplus}(\cdot)), d_{\infty}) \leq \sup_{s \in [0, 1]} \log N(K_2\varepsilon\delta, B_{\delta}(\mu_{\oplus}(s)), d)$ because $d_{\infty}(\omega_1, \omega_2) = \sup_{s \in [0, 1]} d(\omega_1(s), \omega_2(s))$ and $d(\omega_1(s), \omega_2(s))$ is a uniformly continuous function in s so that the supremum is attained. Therefore, $d_{\infty}(\omega_1, \omega_2) = d(\omega_1(s^*), \omega_2(s^*))$ for

some $s^* \in [0, 1]$. Finally we observe that

$$\begin{aligned} & \int_0^1 \sqrt{1 + \log N(\varepsilon ||F||_2, \mathcal{F}_\delta, L^2(P))} d\varepsilon \\ &= \int_0^1 \sqrt{1 + \log N(2M\delta\varepsilon, \mathcal{F}_\delta, L^2(P))} d\varepsilon \\ &\leq \sqrt{\log(C)} + \int_0^1 \sqrt{-\frac{1}{V} \log(\varepsilon\delta)} d\varepsilon + \int_0^1 \sup_{s \in [0,1]} \sqrt{\log N(K_2\varepsilon\delta, B_\delta(\mu_\oplus(s)), d)} d\varepsilon \\ &\leq \sqrt{\log(C)} + \frac{1}{\sqrt{V}} \int_0^1 \sqrt{-\log(\varepsilon\delta)} d\varepsilon + \int_0^1 \sup_{s \in [0,1]} \sqrt{\log N(K_2\varepsilon\delta, B_\delta(\mu_\oplus(s)), d)} d\varepsilon. \end{aligned}$$

Assumption (A10) then implies $J_{[]} (1, \mathcal{F}_\delta, L^2(P)) = O(\sqrt{-\log \delta})$ as $\delta \rightarrow 0$, which completes the proof.

Proof of Theorem 3

For a sequence $\{q_n\}$ define the sets

$$S_{j,n}(x) = \{\omega(\cdot) : 2^{j-1} < q_n d_\infty^{\beta_2/2}(\omega, \mu_\oplus) \leq 2^j\}.$$

Choose $\alpha > 0$ to satisfy (A11) and also small enough such that (A3) and (A4) hold for all $\delta < \alpha$ and choose $\tilde{\alpha} = \alpha^{\beta_2/2}$. For any integer L ,

$$\begin{aligned} & P(q_n d_\infty^{\beta_2/2}(\hat{\mu}_\oplus, \mu_\oplus) > 2^L) \\ &\leq P(d_\infty(\hat{\mu}_\oplus, \mu_\oplus) \geq \alpha) + \sum_{j \geq L, 2^j \leq q_n \tilde{\alpha}} P(\hat{\mu}_\oplus \in S_{j,n}) \\ &\leq P(d_\infty(\hat{\mu}_\oplus, \mu_\oplus) \geq \alpha) + \sum_{j \geq L, 2^j \leq q_n \tilde{\alpha}} P\left(\sup_{\omega \in S_{j,n}} |V_n(\omega, s) - V(\omega, s)| \geq D \frac{2^{2(j-1)}}{q_n^2}, \right) \quad (28) \end{aligned}$$

where (28) follows by observing

$$\sup_{\omega \in S_{j,n}} |V_n(\omega, s) - V(\omega, s)| \geq \left| \inf_{\omega \in S_{j,n}} V_n(\omega, s) - \inf_{\omega \in S_{j,n}} V(\omega, s) \right| \geq D \frac{2^{2(j-1)}}{q_n^2}.$$

The first term in (28) goes to zero by Proposition 4 and for each j in the second term it holds that $d_\infty(\omega, \mu_\oplus) \leq \alpha$. By Lemma 1, $J_{[]} (1, \mathcal{F}_\delta, L^2(P)) = O(\sqrt{\log 1/\delta})$, and therefore is bounded above by $J \sqrt{\log 1/\delta}$ for all small enough $\delta > 0$, where $J > 0$ is a constant. Using equation (14), Lemma 1 and the Markov inequality, the second term is upper bounded up to a constant by

$$\sum_{j \geq L, 2^j \leq q_n \tilde{\alpha}} \frac{2M J 2^j}{n} \sqrt{\log n / 2^{j+1}} \frac{q_n^2}{D 2^{2(j-1)}}. \quad (29)$$

Since $\sqrt{\log n / 2^{j+1}}$ is dominated by $\sqrt{\log n / 2}$, setting $q_n = \frac{\sqrt{n}}{(\log n)^{1/4}}$, the series in (29) is upper bounded by $\frac{8MJ}{D} \sum_{j \geq L, 2^j \leq q_n \tilde{\alpha}} \frac{1}{2^j}$, which converges and can be made sufficiently

small by choosing L and n large. This proves the desired result that $d_\infty(\hat{\mu}_\oplus, \mu_\oplus) = O_P(q_n^{-2/\beta_2}) = O_P\left(\left(\frac{n}{\sqrt{\log n}}\right)^{-1/\beta_2}\right)$.

Proof of Corollary 2

Observing that

$$\begin{aligned} & |\hat{\beta}_{ik} - \beta_{ik}| \\ & \leq \left| \int_0^1 d(X_i(t), \hat{\mu}_\oplus(t)) (\hat{\phi}_k(t) - \phi_k(t)) dt \right| + \left| \int_0^1 \phi(t) (d(X_i(t), \hat{\mu}_\oplus(t)) - d(X_i(t), \mu_\oplus(t))) dt \right| \\ & \leq M \sup_{s \in [0,1]} |\hat{\phi}_k(s) - \phi_k(s)| + \int_0^1 |\phi(t)| dt \sup_{s \in [0,1]} d(\hat{\mu}_\oplus(s), \mu_\oplus(s)), \end{aligned}$$

the result follows from Corollary 1 and Theorem 3.

S2. Comparison of Metric Covariance with Distance Covariance

For two random variables X and Y with marginal probability measures P_X and P_Y and joint probability measure P_{XY} , testing for probabilistic dependence corresponds to testing

$$H_0 : P_{XY} = P_X P_Y \quad \text{versus} \quad H_1 : P_{XY} \neq P_X P_Y. \quad (30)$$

Implementation of these tests is usually based on a metric in the space of probability measures. As shown in [Lyons \(2013\)](#), distance correlation ([Székely et al., 2007](#); [Székely and Rizzo, 2017](#)) provides a suitable metric for this purpose, provided X and Y take values in metric spaces which are of “strong negative type” ([Lyons, 2013](#)) and include Euclidean spaces and separable Hilbert spaces. Then independence of X and Y is equivalent to the distance correlation being 0. While it is often of interest to determine whether distance correlation is zero, which then implies independence of X and Y , the magnitude of distance correlation if not zero is hard to interpret. This fact is emphasized for example in [Jakobsen \(2017\)](#) (page 61) where distance covariance is characterized to be useful to test for independence (30) but much less so to measure degree of dependence between random variables X and Y in general metric spaces.

As a concrete example, we compare distance correlation/covariance with metric correlation/covariance for the case of distribution spaces with the Wasserstein metric. Writing Q_1, Q_2 for the quantile function functions corresponding to distributions F_1, F_2 , the dis-

tance covariance $\text{dCov}(F_1, F_2)$ between F_1 and F_2 is found to correspond to

$$\begin{aligned} \text{dCov}(F_1, F_2) &= E \left[\int_0^1 \{Q_1(t) - Q'_1(t)\}^2 dt \int_0^1 \{Q_2(t) - Q'_2(t)\}^2 dt \right]^{1/2} \\ &\quad + E \left[\int_0^1 \{Q_1(t) - Q'_1(t)\}^2 dt \right]^{1/2} E \left[\int_0^1 \{Q_2(t) - Q'_2(t)\}^2 dt \right]^{1/2} \\ &\quad - 2E \left[\int_0^1 \{Q_1(t) - Q'_1(t)\}^2 dt \int_0^1 \{Q_2(t) - \tilde{Q}_2(t)\}^2 dt \right]^{1/2}, \end{aligned}$$

where F'_1 is an independent copy of F_1 and F'_2 , \tilde{F}_2 are two independent copies of F_2 . This expression for distance covariance is rather unintuitive, and it is hard to interpret as a measure for the strength of covariation between F_1 and F_2 .

A numerical comparison provides further illumination. We implemented distance covariance as an alternative covariance/correlation for functional random objects in a simulation study, where we compared the utility of the proposed metric covariance with that of distance covariance (Székely and Rizzo, 2017; Lyons, 2013) for carrying out FPCA of regular scalar-valued functional data. We consider a simple setting as in classical FDA where the time-varying random objects $X(s)$ are real valued for $s \in [0, 1]$. In the following, we take U_i , V_i and Y_i to be distributed as $N(0, 3)$, $N(0, 1)$ and $N(0, 0.25)$, respectively. As described in the simulation setting for time varying networks in Section 5.2, we take $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ to be orthonormal polynomials derived from the Jacobi polynomials $P_n^{(\alpha, \beta)}(x)$ (Totik, 2005), which are classical orthogonal polynomials for $\alpha, \beta > 1$. The expressions for $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ are

$$\begin{aligned} \phi_1(t) &= \frac{(P_3^{(1,0.5)}(2t-1))t^{0.25}(1-t)^{0.5}}{[\int_0^1 (P_3^{(1,0.5)}(2t-1))^2 t^{0.5}(1-t)dt]^{1/2}} \\ \phi_2(t) &= \frac{(P_4^{(1,0.5)}(2t-1))t^{0.25}(1-t)^{0.5}}{[\int_0^1 (P_4^{(1,0.5)}(2t-1))^2 t^{0.5}(1-t)dt]^{1/2}} \\ \phi_3(t) &= \frac{(P_5^{(1,0.5)}(2t-1))t^{0.25}(1-t)^{0.5}}{[\int_0^1 (P_5^{(1,0.5)}(2t-1))^2 t^{0.5}(1-t)dt]^{1/2}}. \end{aligned}$$

We generated 1000 i.i.d. realizations $X_i(s)$ as follows on a fine grid of $[0, 1]$,

$$X_i(s) = U_i\phi_1(t) + V_i\phi_2(t) + Y_i\phi_3(t). \quad (31)$$

It is clear from the construction of X_i that in model (31), the first, second and third eigenfunctions are given by $\phi_1(t)$, $\phi_2(t)$ and $\phi_3(t)$ respectively. We evaluated the estimated metric covariance and distance covariance surfaces on a fine grid, which led to the surfaces depicted in Figure 15, and then obtained the first three eigenfunctions using these surfaces as covariance kernels. The resulting eigenfunctions are presented in Figure

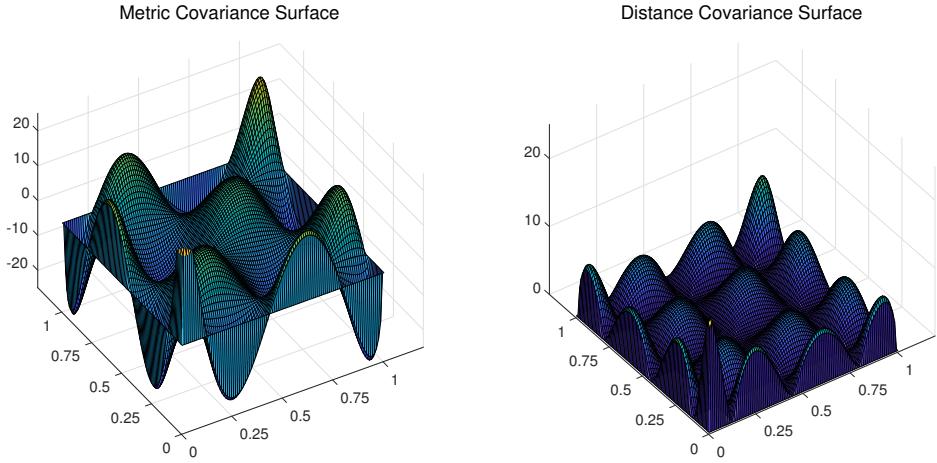


Fig. 15: Metric covariance surface (left) and distance covariance surface (right) for functional data generated according to model in (31).

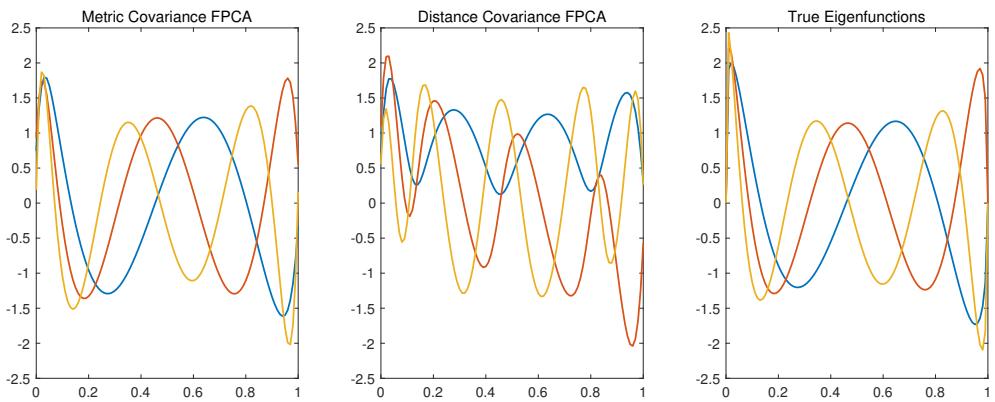


Fig. 16: Eigenfunctions obtained from using metric covariance based (left panel) and distance covariance based (middle panel) kernel for simulated functional data, generated according to model (31). Also shown are the true underlying eigenfunctions (right panel). The blue curves correspond to the first, the red curves to the second and the yellow curves to the third eigenfunction.

16. As can be seen from Figures 15 and 16, metric covariance delivers the eigenfunctions that one would expect for functional principal component analysis, while distance covariance as an alternate notion of covariance leads to seemingly arbitrary and uninterpretable eigenfunctions, so is clearly not suitable in this context.

In classical FPCA one aims to identify dominant modes of variation of functional data that are derived from the eigenfunctions of the auto-covariance operator. The interpretation of these modes of variation provides valuable insights in many applications, and this is why interpretability of the eigenfunctions is important. Another major goal is to decompose the variation of functional data in a parsimonious and interpretable way into orthogonal directions. As illustrated in the simulation above, the lack of clear interpretation of the eigenfunctions associated with the distance covariance operator is a big hurdle for this program. When using distance correlation, the corresponding distance variance and also the total variation have an unintuitive and complex form that makes distance covariance rather unsuitable for our purposes. In contrast, the proposed metric covariance (5) works well for quantifying the variation and co-variation of random objects. It gives rise to the total variation measure (6) for functional random objects and emerges as a bona fide extension of the proven and successful FPCA for scalar-valued functional data.

S3. Metric auto-covariance surfaces and eigenfunctions for New York taxi data

We repeated the analysis of the time-varying networks generated by the New York taxi data separately for three groups of days, namely the weekdays Monday-Thursday (group 1), Fridays and weekends (group 2) and holidays (group 3). The results are visualized in Figure 17. This figure clearly indicates that the metric covariance structure and the eigenfunctions differ across the groups. The Fréchet integrals for the dominant eigenfunctions reveal different aspects of variation, both within and between daily networks in three groups and are presented in the movies “week.mov”, “friday.mov” and “Hol.mov” which are included in the supplementary materials. In the movie frames, the top left panels correspond to the FPCs for the first eigenfunction, the top right panels to those of the second, the bottom left panels to those of the third and the bottom right panels to those of the fourth eigenfunction. The edge weights in the graphs are proportional to their line widths.

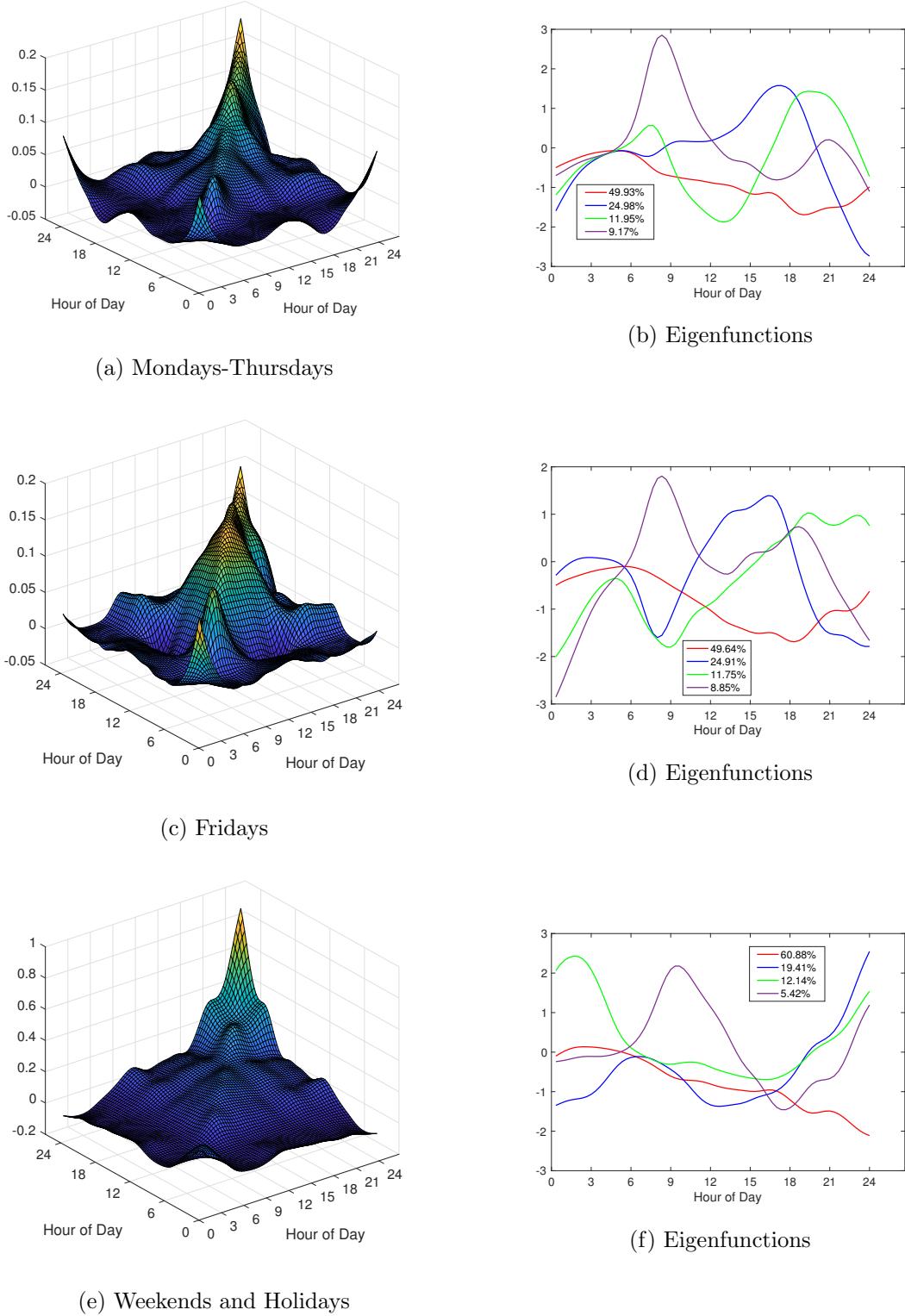


Fig. 17: Estimated metric auto-covariance surface (10) (left) and associated eigenfunctions (right), for the New York taxi data, viewed as time-varying networks, separated by groups of days.

S4. Additional Visualization for the World Trade Data analysis

We selected USA, Saudi Arabia, Hong Kong and Thailand and display their time evolving inter-commodity trade correlations as obtained from the fitted model for the years 1970, 1982, 1992, 1999 (bottom to top) in Figure 18.

We also computed the Fréchet integral covariance matrices for the first four eigenfunctions. For visualization of commodities trade similarities these Fréchet integral covariance matrices were converted to correlation matrices that can be viewed in the movie “trade.mov”, available in the online supplement. In the movie frames, the FPCs for the first, second, third and fourth eigenfunctions are at the top left, top right, bottom left and bottom right, respectively.

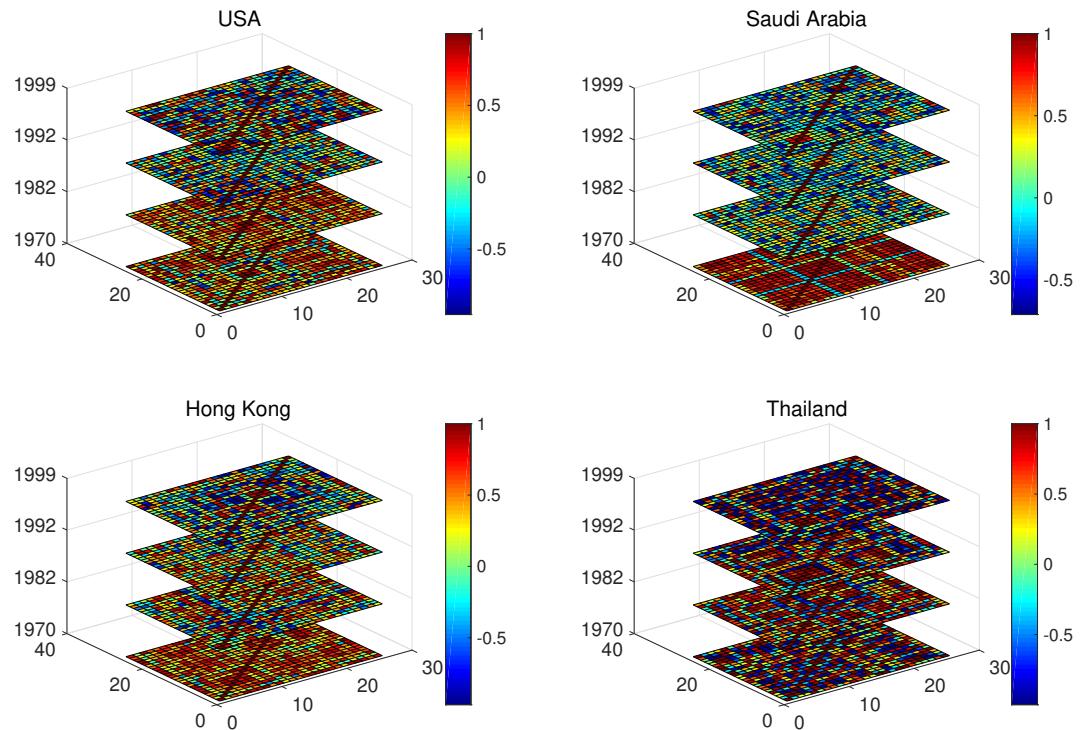


Fig. 18: International commodity trade correlation matrices for 1970 (first slice from bottom), 1982 (second slice from bottom), 1992 (third slice from bottom) and 1999 (top slice) for four countries.

S5. Description of Movies

The following is a list of the movies that have been included as supplementary materials and a brief discussion of their content.

Filename	Content Description
networks.mov	Object FPCs obtained as Fréchet integrals (11) for one randomly chosen simulation run for $n = 50$, using the model in (17) for time-varying networks in section 5.2.
mean_males.mov	Estimated Fréchet mean function for males represented as density functions indexed by year, derived from the yearly sample average of the quantile functions of the countries included in the mortality data in section 6.1 for each calendar year.
mean_females.mov	Estimated Fréchet mean function for females. The description is the same as for males.
mean_NY.mov	Estimated Fréchet mean function represented as time varying network adjacency matrices, obtained for each 20 minute time interval by averaging the network adjacency matrices over 363 daily networks for the New York taxi data as described in section 6.2.
week.mov	Fréchet integrals represented as network adjacency matrices for the dominant eigenfunctions, obtained from the analysis of the New York taxi data as described in section S3 of the supplement, for weekdays.
friday.mov	Same as previous, for Fridays.
Hol.mov	Same as previous, for weekends and special holidays.
trade.mov	Fréchet integrals represented as covariance matrices for the dominant eigenfunctions for the trade dataset as described in section 6.3.

S6. Data Descriptions

S6.1 Zones in Manhattan, New York

New York City Taxi and Limousine Commission (NYC TLC) provides records on pick-up and drop-off dates/times, pick-up and drop-off latitudes and longitudes, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts for yellow and green taxis. The data are available at http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml. The polygon shape files available at this website represent the boundaries zones for taxi pickups as delimited by the New York City Taxi and Limousine Commission (TLC). The latitudes and longitudes in New York are split into 6 boroughs: Bronx, Brooklyn, Newark Liberty International Airport, Manhattan, Queens and Staten Island. Since yellow taxis operate predominantly in Manhattan, we consider so-called towns in Manhattan which are further grouped into 10 zones as described in the following Table. We excluded the islands from our study. For a description of towns, we refer to Figure 19.

Zone	Towns
1	Inwood, Fort George, Washington Heights, Hamilton Heights, Harlem, East Harlem
2	Upper West Side, Morningside Heights, Central Park
3	Yorkville, Lenox Hill, Upper East Side
4	Lincoln Square, Clinton, Chelsea, Hell's Kitchen
5	Garment District, Theater District
6	Midtown
7	Midtown South
8	Turtle Bay, Murray Hill, Kips Bay, Gramercy Park, Sutton, Tudor, Medical City, Stuy Town
9	Meat packing district, Greenwich Village, West Village, Soho, Little Italy, Chinatown, Civic center, Noho
10	Lower East Side, East Village, ABC Park, Bowery, Two Bridges, Southern tip, White Hall, Tribeca, Wall Street

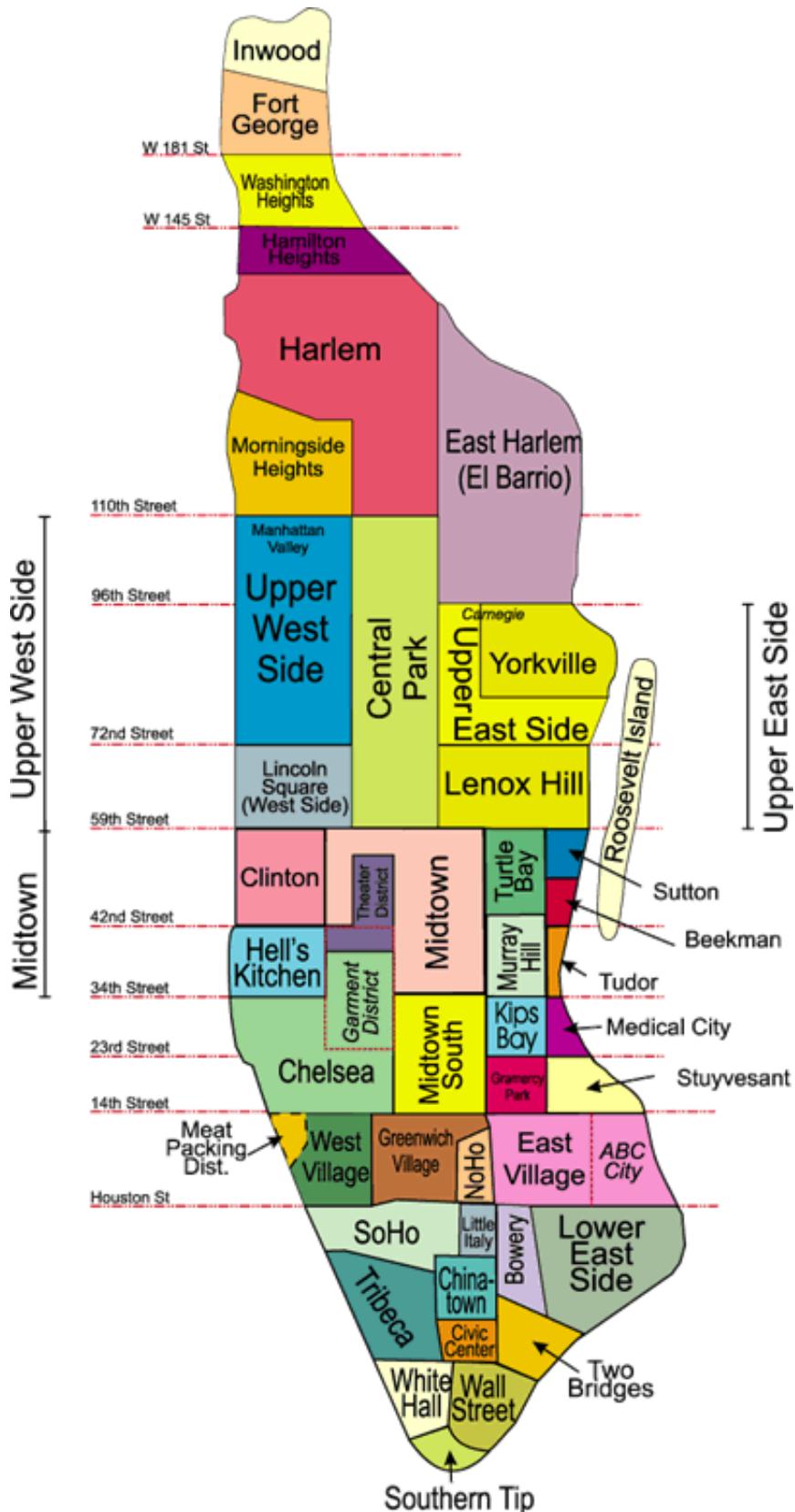


Fig. 19: Towns in Manhattan, New York.

S6.2 Trade Data

The countries chosen for the analysis were Morocco, Tunisia, Egypt, Canada, USA, Argentina, Brazil, Chile, Mexico, Venezuela, Dominican Republic, Israel, Japan, Cyprus, Lebanon, Saudi Arabia, United Arab Emirates, Turkey, Hong Kong, Indonesia, Korea Republic, Malaysia, Philippines, Singapore, Thailand, Taiwan, China, Belgium-Luxemburg, Denmark, France, Greece, Ireland, Italy, Netherlands, Portugal, Spain, UK, Austria, Finland, Norway, Sweden, Switzerland, Malta, Bulgaria, Australia and New Zealand. A list of the traded commodities can be found in the following table,

Number	Products
1	Sugar, Honey
2	Road Vehicles
3	Fruits and Vegetables
4	Non metallic Minerals Manufactures
5	Coffee, Tea, Cocoa, Spices
6	Tobacco and Tobacco Manufactures
7	Textiles, Yarns, Fabrics
8	Printed Books, Maps, Charts, Paper, Stationery
9	Beverages
10	Chemical Materials, Products
11	Machineries
12	Transport Equipments
13	Rubber
14	General Industrial Machinery
15	Dairy Products, Eggs
16	Fish and Seafood
17	Cereals
18	Petroleum
19	Dye
20	Medicines
21	Oil, Perfumes, Toilet, Cleansing
22	Paper, Paper Board, Articles
23	Iron and Steel
24	Manufacture of Metals
25	Power Generating Machinery
26	Telecommunications, Sound Recording, Reproducing Equipments