DISSERTATION

METHODS FOR EXTREMES OF FUNCTIONAL DATA

Submitted by

Qian Xiong

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2018

Doctoral Committee:

Advisor: Piotr S. Kokoszka

Daniel Cooley
Olivier Pinaud
Haonan Wang

ABSTRACT

METHODS FOR EXTREMES OF FUNCTIONAL DATA

Motivated by the problem of extreme behavior of functional data, we develop statistical theory at the nexus of functional data analysis (FDA) and extreme value theory (EVT). A fundamental technique of functional data analysis is to replace infinite dimensional curves with finite dimensional representations in terms of functional principal components (FPCs). The coefficients of these projections, called the scores, encode the shapes of the curves. Therefore, the study of the extreme behavior of functional time series can be transformed to the study on functional principal component scores. We first derive two tests of significance of the slope function using functional principal components and their empirical counterparts (EFPC's). Applied to tropical storm data, these tests show a significant trend in the annual pattern of upper wind speed levels of hurricanes. Then we establish sufficient conditions under which the asymptotic extreme behavior of the multivariate estimated scores is the same as that of the population scores. We clarify these issues, including the rate of convergence, for Gaussian functions and for more general functional time series whose projections are in the Gumbel domain of attraction. Finally, we derive the asymptotic distribution of the sample covariance operator and of the sample functional principal components for functions which are regularly varying and whose fourth moment does not exist. The new theory is applied to establish the consistency of the regression operator in a functional linear model, with such errors.

ACKNOWLEDGEMENTS

First, and most of all, I wish to express my sincere gratitude to my advisor Professor Piotr Kokoszka for his guidance not only for academic research but also on career planning. He has shared his broad knowledge with me and inspired me in many ways. Started from basic concepts and methods, he helped me grow into a confident researcher who could break large tasks down into smaller, more manageable pieces. It has been an enjoyable and memorable experience discussing research related problems with him. He demonstrated huge enthusiasm for research and also encouraged me to pursue my own passion and dreams.

I would like to thank my committee members: Professor Daniel Cooley, Professor Olivier Pinaud and Professor Haonan Wang. Their valuable feedback and advice were essential and helpful throughout the research process. I express special thanks to Professor Haonan Wang. His brilliant insights are beneficial to my research consistently and substantially.

I also want to thank Petra Burdejova and Professor Wolfgang Hardle for their expertise in climate sciences and practical contributions to our collaborative project. I would thank Professor Stilian Stoev for his mathematical expertise and theoretical contributions to our research.

I want to thank my friends in the Department of statistics at Colorado State University for all the great times that we have shared. I would thank the students, faculty and staff for providing me such a positive and vibrant learning environment.

Lastly, I wish to thank my family for their endless love, encouragement and support. They give me the courage and strength to face challenges and obstacles in life. I cannot step further without their trust and support for all my decisions. The last word of acknowledgment I have saved for my boyfriend, Wenxiang Chen, for his patience and encouragement in helping me pursue my dream career.

DEDICATION

This dissertation is dedicated to my family and all the people who inspire and encourage me all along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Functional data analysis (FDA) is a dynamic branch of statistics that provides powerful tools and techniques to study information contained in a collection of curves or surfaces. It is concerned with observations, called functional data, that are viewed as smooth functions defined over some domain. Examples include daily stock returns over a specific period of time and annual patterns of temperature in some region. Methodological foundations of FDA are addressed in [1], its mathematical foundations in [2]. An important feature of FDA is its ability to take into account the temporal dependence between the observations. Functional time series arise in economics, finance, and demography to better analyze, model and forecast time series data. Compared to traditional methods studying scalar and vector time series, functional time series methods are often able to approximate the periodic components with irregular sampling patterns [3], and effectively reduce data noise through curve smoothing [4]. In spite of a huge amount of literature on scalar or vector time series, methodologies for functional time series are in strong demand.

Functional data are intrinsically infinite dimensional. Even though they are measured discretely over a finite subset of some interval, the dimensionality is still very high. The high dimensionality of these data presents big challenges for both theory and computation. Dimension reduction, therefore, is a central issue. Functional data analysis offers a way to approach high-dimensional or infinite-dimensional problems. The fundamental technique of FDA is functional principal component analysis (FPCA), see [1, 2, 5, 6]. It is derived from principal component analysis (PCA), extending the finite dimensional setting to the infinite dimensional one. Both settings thus have the same underlying concepts and objectives. The individual curves are represented by a linear combination of basis functions. One choice of basis functions is functional principal components (FPC's), the orthonormal eigenfunctions of the covariance operator of the process. By truncating the basis representation at a finite depth and estimating FPC's using empirical functional principal components (EFPC's), a lower dimensional representation is obtained to approximate the infinite dimensional curves. Asymptotic properties of EFPC's were established by [7] for iid cases and extended to weakly dependent functional time series by [8].

The study of extreme values is another important topic in applied sciences. Extreme value theory (EVT) provides theoretical foundations for quantifying the stochastic behavior of a process at unusually large or small levels. It is widely used in many disciplines, such as finance, insurance, geology and climatology. There are many excellent accounts of EVT, including [9–12]. One of the key issues is to estimate the distribution of the maximum or minimum values, which are called extremes. For a single process, the limit laws for the distribution of maximum value, called maxima, can be expressed by three extreme value distributions: Gumbel, Fréchet and Weibull, first derived by [13]. The behavior of extremes for functional data is also of interest, but it has not been studied extensively. It is therefore hoped that we provide useful contribution that merges the ideas of functional data analysis and extreme value theory. The major contributions consist of the study of the extreme behavior of functional time series and of functional principal components and scores.

The dissertation is organized as follows. Chapter 2 derives two tests of significance of the slope function using functional principal components and their empirical counterparts. One of the tests relies on a Monte Carlo distribution to compute the critical values, the other is pivotal with the chi-square limit distribution. These tests are applied to tropical storm data to detect the annual trend pattern of the upper wind speed levels of hurricanes. Chapter 3 establishes sufficient conditions under which the asymptotic extreme behavior of the multivariate estimated scores is the same as that of the population scores. We clarify these issues, including the rate of convergence, for Gaussian functions and for more general functional time series whose projections are in the Gumbel domain of attraction. Chapter 4 derives the asymptotic distribution of the sample covariance operator and of the sample functional principal components for functions, which are regularly varying and whose fourth moment does not exist. The new theory is applied to establish the consistency of the regression operator in a functional linear model, with such errors.

The following chapters of this dissertation are based on the papers during my Ph.D. program.

- Chapter 2 is based on the paper: P. Burdejova, W. Hardle, P. Kokoszka and Q. Xiong, Change point and trend analyses of annual expectile curves of tropical storms, Econometrics and Statistics , 1 , 101-117, 2017.
- Chapter 3 is based on the paper: P. Kokoszka and Q. Xiong, Extremes of projections of functional time series on data-driven basis systems, Extremes, 21, 177-204, 2018.
- Chapter 4 is based on the paper: P. Kokoszka, S. Stoev, Q. Xiong, Principal components analysis of regularly varying functions. Under review of Bernoulli.

Before moving to individual research topics, in the remainder of this chapter we will introduce some fundamental concepts that will be used in subsequent chapters. In particular, we will give an introduction to functional data, functional principal components, functional linear regression and review the classical change point tests. Then we look at some basic concepts of extreme value theory, extreme properties for stationary sequences, and introduce the definition of the regularly varying random element.

1.1 Functional Data

We introduce basic concepts of functional data in a separable Hilbert space H with inner product $\langle \cdot, \cdot \rangle$. Denote by $\mathcal{L}$ the space of bounded (continuous) linear operators on H with the norm

$$\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\| : \|x\| \le 1\}.$$

A linear operator $\Psi : H \to H$ is Hilbert–Schmidt if $\sum_{j=1}^{\infty} \|\Psi(e_j)\|^2 < \infty$, where $\{e_j\}$ is any orthonormal basis of H. The space $\mathcal{S}$ of Hilbert-Schmidt operators is a separable Hilbert space with the scalar product

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} = \sum_{j=1}^{\infty} \langle \Psi_1(e_j), \Psi_2(e_j) \rangle .$$

Suppose $X_1, X_2, \ldots, X_N$ are a realization of independent and identically distributed (iid) random function $X = \{X(t), t \in \mathcal{T}\}$ in $L^2 = L^2(\mathcal{T})$, the space of integrable functions on a compact interval $\mathcal{T}$. For simplicity, we work on $\mathcal{T} = [0,1]$ since any compact interval can be normalized to unit interval. The space $L^2 = L^2([0,1])$ is the set of measurable real-valued functions x defined on $[0,1]$ satisfying $\int_0^1 x^2(t)dt < \infty$. The space L^2 is a separable Hilbert space with the inner product

$$\langle x, y \rangle = \int_0^1 x(t)y(t)dt.$$

An important class of operators in L^2 are the integral operators defined by

$$\Psi(x)(t) = \int_0^1 \psi(t,s)x(s)ds, \quad x \in L^2, \tag{1.1.1}$$

with the real kernel $\psi(\cdot,\cdot)$. Then $\|\Psi\|_{\mathcal{S}}^2 = \iint \psi^2(t,s)dtds$ and $\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{S}}$. Such operators are Hilbert -Schmidt if and only if $\iint \psi^2(t,s)dtds < \infty$.

A random function X is said to be square integrable if $E\|X\|^2 = E\int_0^1 X^2(t)dt < \infty$. We then define the mean and covariance functions by

$$\mu(t) = E[X(t)],$$

$$c(t,s) = E[(X(t) - \mu(t))(X(s) - \mu(s))].$$

Similar to (1.1.1), the covariance operator of X can be defined by

$$C(x) = E[\langle (X - \mu), x \rangle (X - \mu)] = \int_0^1 c(t,s)x(s)ds, \quad x \in L^2. \tag{1.1.2}$$

If $X_1, X_2, \ldots, X_N$ are iid in L^2, all the functional parameters can be estimated by the sample equivalents. The sample mean function, sample covariance function and sample covariance operator are defined by

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

$$\hat{c}(t,s) = \frac{1}{N}\sum_{i=1}^{N}(X_i(t) - \hat{\mu}(t))(X_i(s) - \hat{\mu}(s)),$$

and

$$\widehat{C}(x) = \frac{1}{N}\sum_{i=1}^{N}\langle X_i - \hat{\mu}, x\rangle (X_i - \hat{\mu}), \quad x \in L^2. \tag{1.1.3}$$

For a more comprehensive introduction of functional data, see [6].

1.1.1 Functional principal components

The fundamental technique of functional data analysis (FDA) is functional principal component analysis (FPCA). Each curve X_i admits the expansion

$$X_i(t) = \sum_{j=1}^{\infty}\sqrt{\lambda_j}Z_{ij}v_j(t), \tag{1.1.4}$$

where the λ_j and v_j are, respectively, the eigenvalues and the eigenfunctions of the covariance operator C defined in (1.1.2), i.e. $C(v_j) = \lambda_j v_j,\ j \geq 1$. The eigenvalues must be identifiable, so we assume that $\lambda_1 > \lambda_2 > \cdots$. In practice, we are concerned about the p largest eigenvalues, and assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_p > \lambda_{p+1}$, which implies that the first p eigenvalues are nonzero. The random variables Z_{ij}, called the jth score of X_i, are defined by $Z_{ij} = \lambda_j^{-1/2}\langle X_i, v_j\rangle = \lambda_j^{-1/2}\int_0^1 X_i(t)v_j(t)dt$. They encode the shapes of the functions X_i with unit variance. In applications, the infinite expansion is replaced by truncating at a finite depth which involves estimated counterparts of the quantities in (1.1.4). That is to approximate the infinite dimensional curve using a p-dimensional representation with the form

$$X_i(t) \approx \sum_{j=1}^{p}\sqrt{\hat{\lambda}_j}\hat{Z}_{ij}\hat{v}_j(t), \tag{1.1.5}$$

where the $\hat{\lambda}_j$ and $\hat{v}_j$ are the eigenvalues and the eigenfunctions of the sample covariance operator $\widehat{C}$ defined in (1.1.3), i.e. $\widehat{C}(v_j) = \hat{\lambda}_j\hat{v}_j,\ j \geq 1$. The estimated scores $\hat{Z}_{ij}$ are defined by $\hat{Z}_{ij} = \hat{\lambda}_j^{-1/2}\langle X_i, \hat{v}_j\rangle = \hat{\lambda}_j^{-1/2}\int_0^1 X_i(t)\hat{v}_j(t)dt$. The eigenfunctions of the covariance operator C

are called the functional principal components (FPC's) and the eigenfunctions of the sample covariance operator $\widehat{C}$ are called the empirical functional principal components (EFPC's).

Suppose $E\|X\|^4 < \infty$ and $\lambda_1 > \lambda_2 > \cdots > \lambda_p > \lambda_{p+1}$. Large sample justifications of the applications of expansion (1.1.5) rely on the following bounds:

$$\limsup_{N\to\infty} NE\|\hat{c}_j\hat{v}_j - v_j\|^2 < \infty, \quad \limsup_{N\to\infty} NE|\hat{\lambda}_j - \lambda_j|^2 < \infty, \tag{1.1.6}$$

where $\hat{c}_j = \text{sign}(\langle \hat{v}_j, v_j\rangle)$. Relations (1.1.6) were established by [7] for iid functions, and extended to weakly dependent functional time series by [8]. The weak dependence is quantified by the condition known as $L^p - m-$approximability. For $p \geq 1$, we denote by $L^p_H = L^p(\Omega, \mathcal{A}, \mathcal{P})$ the space of $H = L^2$ valued random variables X such that

$$v_p(X) = (E\|X\|^p)^{1/p} = \left(E\left\{\int X^2(t)dt\right\}^{p/2}\right)^{1/p} < \infty.$$

DEFINITION 1.1.1. *A sequence $\{X_i\} \in L^p_H$ is called $L^p - m-$approximable if each X_i admits the representation*

$$X_i = f(\epsilon_i, \epsilon_{i-1}, \cdots),$$

where the ϵ_i are iid elements taking values in a measurable space S and f is a measurable function $f : S^\infty \to H$. Moreover we assume that if $\{\epsilon'_i\}$ is an independent copy of $\{\epsilon_i\}$ defined on the same probability space, then letting

$$X_i^{(j)} = f(\epsilon_i, \epsilon_{i-1}, \cdots, \epsilon_{i-j+1}, \epsilon'_{i-j}, \epsilon'_{i-j-1}, \cdots)$$

we have

$$\sum_{j=1}^{\infty} v_p(X_i - X_i^{(j)}) < \infty.$$

Definition 1.1.1 implies that $\{X_i\}$ is strictly stationary. We choose $p = 4$ for our applications of FDA. Theorem 16.2 in [6] shows that relations (1.1.6) still hold if $\{X_i\}$ is an L^4-m-approximable sequence and $\lambda_1 > \lambda_2 > \cdots > \lambda_p > \lambda_{p+1}$.

Furthermore, [14] derived the asymptotic normality of the random variables $N^{1/2}(\hat{\lambda}_j - \lambda_j)$ using complex arguments related to those developed by [7]. Suppose the random function X_i can be written as a infinite dimensional stochastic process, i.e. $X_i = \sum_{j=-\infty}^{+\infty} a_j(\epsilon_{i-j})$. The random variables ϵ_j are iid with mean 0 satisfying $E\left\|\epsilon\right\|^4 < \infty$. The bounded linear operators a_j map L^2 onto L^2 and satisfy $\sum_{j=-\infty}^{+\infty} \|a_j\|_{\mathcal{S}} < \infty$. $\{X_i\}$ is thus a strictly stationary sequence of random functions in L^2. [14] showed that there exists some Gaussian random operator $Z \in \mathcal{S}$, such that

$$Z_N := N^{1/2}(\widehat{C} - C) \xrightarrow{d} Z, \quad N \to \infty.$$

Denoted by C_ϵ the covariance operator of $\{\epsilon_j\}$ and e_j the jth largest eigenvector of $C\epsilon$. Then

$$N^{1/2}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, \sigma_j^2),$$

with the variance $\sigma_j^2 = E\left\langle Z(e_j), e_j \right\rangle^2$. In the course of his proof he also established the asymptotic normality of the random functions $N^{1/2}(\hat{c}_j\hat{v}_j - v_j)$. [15] then presented simpler arguments under which the asymptotic properties still hold. Let $y \otimes z$ be the integral operator in Hilbert space H defined by $(y \otimes z)(x) = \langle y, x \rangle z$ for $x \in H$. Set

$$T_j = \sum_{k \neq j} (\lambda_j - \lambda_k)^{-1} \left\langle Z, v_j \otimes v_k \right\rangle v_k.$$

Suppose $\{X_i\}$ is an $L^4 - m-$approximable sequence and $\lambda_1 > \lambda_2 > \cdots > \lambda_p > \lambda_{p+1}$. Then

$$N^{1/2}\left\{\hat{v}_j - v_j,\ 1 \leq j \leq p\right\} \xrightarrow{d} \left\{T_j,\ 1 \leq j \leq p\right\},\ \text{ in } (L^2)^p.$$

The $\{T_j,\ 1 \le j \le p\}$ are jointly Gaussian distributed with cross covariance operators (for $j_1 = 1, \cdots, p$ and $j_2 = 1, \cdots, p$)

$$\sum_{k_1 \neq j_1, k_2 \neq j_2} \frac{\langle \Gamma, (v_{j_1} \otimes v_{k_1}) \otimes (v_{j_2} \otimes v_{k_2}) \rangle}{(\lambda_{j_1} - \lambda_{k_1})(\lambda_{j_2} - \lambda_{k_2})} v_{k_1} \otimes v_{k_2}.$$

See [15] for more details.

1.1.2 Functional linear regression model

Functional regression model is one of the most powerful tools of functional data analysis. Recall that the simplest linear regression model is

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \cdots, N,$$

in which all random variables are scalars, and the regressors x_i are typically assumed to be known scalars. In a functional linear regression model, some of these quantities are curves, and analogs of the coefficients β_0 and β_1 must be then appropriately defined. Suppose the explanatory functions $X_1, \cdots, X_N$ are iid with mean zero in L^2. The fully functional model is

$$Y_i(t) = \int_0^1 \psi(t,s) X_i(s) ds + \epsilon_i(t), \quad i = 1, 2, \cdot, N,$$

where $\psi(\cdot,\cdot)$ is the kernel of $\Psi \in \mathcal{S}$ i.e. $\int\int \psi^2(t,s) dtds < \infty$. It reflects the effect of the explanatory functions X_i at time s on the response functions Y_i at time t. The error functions ϵ_i are assumed to be iid with mean zero in L^2 and independent of X_i. The responses Y_i are thus iid in L^2. Consider the expansions of explanatory and response functions

$$X(s) = \sum_{i=1}^{\infty} \xi_i v_i(s), \quad Y(t) = \sum_{j=1}^{\infty} \zeta_j u_j(t),$$

where the v_j are the FPC's of X and the u_j the FPC's of Y, and

$$\xi_i = \langle X, v_j \rangle, \quad \zeta_j = \langle Y, u_j \rangle.$$

An estimator of the kernel $\psi(\cdot,\cdot)$ is

$$\psi(t,s) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{E[\xi_l \zeta_k]}{\lambda_l} u_k(t) v_l(s) \tag{1.1.7}$$

with $\lambda_l = E[\xi_l^2]$, the eigenvalue corresponding to v_l. The series converging in $L^2([0,1] \times [0,1])$, equivalently in $\mathcal{S}$, see Lemma 8.1 in [6]. Another approach to estimate $\psi(\cdot,\cdot)$ is to consider estimating the form (1.1.7) using EFPC's, i.e.

$$\hat{\psi}_{KL}(t,s) = \sum_{k=1}^{K} \sum_{l=1}^{L} \hat{\lambda}_l^{-1} \hat{\sigma}_{lk} \hat{u}_k(t) \hat{v}_l(s),$$

where $\hat{\sigma}_{lk}$ is an estimator of $E[\xi_l \zeta_k]$. The simplest estimator is

$$\hat{\sigma}_{lk} = \frac{1}{N} \sum_{i=1}^{N} \langle X_i, \hat{v}_l \rangle \langle Y_i, \hat{u}_k \rangle.$$

1.1.3 Change point tests of the mean function

Recall that the empirical eigenvalues and eigenfunctions are estimated using the sample covariance operator defined in (1.1.3). This approach is however not valid if the observations X_i do not have the same mean. Furthermore, the inference based on the FPC's will no longer be valid. A simple type of change is that the mean function changes abruptly from one deterministic curve to another. For scalar observations, the model for an abrupt change is $X_i = \mu_1 + Y_i, 1 \le i \le k^\star$, $X_i = \mu_2 + Y_i, k^\star < i \le N$, where $k^\star$ is an unknown change point. Assuming $k^\star/N \to \theta$, then $\widehat{C}$ is close to $C_Y + \theta(1-\theta) \langle \Delta, \cdot \rangle \Delta$, where $\Delta = \mu_1 - \mu_2$. Therefore, the eigenfunctions of $\widehat{C}$ will then no longer estimate the eigenfunctions of C_Y, the covariance operator of the Y_i. Change point methodology is often applied to time series, for example, to detect changes in the average annual temperature at a specific location.

We assume that the observations $X_i \in L^2$ are independent. We want to test if their mean remains constant in i, i.e. test the null hypothesis

$$H_0 : EX_1 = EX_2 = \cdots = EX_N.$$

The specific value of the mean is not part of the null hypothesis. The alternative is that there is at least one unknown change point such that the equality under H_0 fails. Under the alternative we can also locate the change points, see Chapter 6 in [6].

Consider the estimated scores corresponding to the largest d eigenvalues

$$\hat{\xi}_{l,i} = \int_0^1 [X_i(t) - \bar{X}_N(t)]\hat{v}_l(t)dt, \;\; i = 1, 2, \cdots, N, \;\; l = 1, 2, \cdots, d,$$

where $\bar{X}_N(t) = N^{-1} \sum_{i=1}^{N} X_i(t)$. The statistic used to derive the test of the constancy of the mean function is

$$T_N(x) = \frac{1}{N} \sum_{l=1}^{d} \hat{\lambda}_l^{-1} \left(\sum_{1 \le i \le Nx} \hat{\xi}_{l,i} - x \sum_{1 \le i \le N} \hat{\xi}_{l,i} \right)^2.$$

The $\hat{\lambda}_l$ and $\hat{v}_l$ are the empirical functional principal eigenvalues and eigenfunctions. Under the null hypothesis, we can represent each functional observation as

$$X_i(t) = \mu(t) + Y_i(t), \quad EY_i(t) = 0. \tag{1.1.8}$$

Assume that $Y_i(t)$ in (1.1.8) are iid mean zero random elements of L^2 which satisfying $E\|Y_i\|^4 = \int_0^1 EY_i^4(t)dt < \infty$, and satisfy $\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1}$, for some $d > 0$. Then Theorem 6.1 in [6] showed that, under H_0,

$$T_N(x) \xrightarrow{d} \sum_{1 \le l \le d} B_l^2(x) \;\; (0 \le x \le 1),$$

in the Skorokhod topology on the space $D[0,1]$ of right-continuous functions on $[0,1]$ having limits to the left at each $t \in (0,1]$. Here $B_l(\cdot)$ are independent standard Brownian bridges. Therefore, we can produce effective tests by the Cramér-von-Mises functional. The convergence of $\int_0^1 T_N(x)dx \xrightarrow{d} \int_0^1 \sum_{1\le l\le d} B_l^2(x)dx$ can be rewritten as

$$S_{N,d} = \frac{1}{N^2}\sum_{l=1}^{d}\frac{1}{\hat{\lambda}_l}\sum_{k=1}^{N}\left(\sum_{1\le i\le k}\hat{\xi}_{l,i} - \frac{k}{N}\sum_{1\le i\le k}\hat{\xi}_{l,i}\right)^2 \xrightarrow{d} \int_0^1 \sum_{1\le l\le d} B_l^2(x)dx := K_d.$$

The limit distribution was derived by [16]. Denoting by $c_d(\alpha)$ its $(1-\alpha)$th quantile, the test rejects H_0 if $S_{N,d} > c_d(\alpha)$. The critical values $c_d(\alpha)$ are given in Table 6.1 in [6].

1.2 Extreme value theory

The asymptotic theory of sample extremes has been developed in parallel with the central limit theory. Let $X_1, X_2, \cdots$ be independent and identically distributed random variables. The central limit theory is concerned with the limit behavior of the partial sums $X_1 + X_2 + \cdots + X_N$ as $N \to \infty$, while the theory of sample extremes is concerned with the limit behavior of the sample extremes $\max(X_1, X_2, \cdots, X_N)$ or $\min(X_1, X_2, \cdots, X_N)$ as $N \to \infty$. Since

$$\min(X_1, X_2, \cdots, X_N) = -\max(-X_1, -X_2, \cdots, -X_N),$$

we will focus on sample maxima.

Suppose $X_1, X_2, \cdots$ are iid random variables with common cumulative distribution function (cdf) F. The maximum of the first N random variables is denoted by $M_N = \max(X_1, X_2, \cdots, X_N)$. Observe that

$$P(M_N \le x) = P(X_1 \le x, \cdots, X_N \le x) = F^N(x).$$

DEFINITION 1.2.1. *Suppose there exists a sequence of constants $a_N > 0$ and b_N such that*

$$\lim_{N\to\infty} P\left(\frac{M_N - b_N}{a_N} \le x\right) = \lim_{N\to\infty} F^N(a_N x + b_N) = G(x), \tag{1.2.9}$$

for every continuity point x of G and G a nondegenerate distribution function. The class of distributions F satisfying (1.2.9) *is called in the domain of attraction of G.*

The class of distributions that can occur as a limit in the relation (1.2.9) is called the class of extreme value distributions. It has the form of $G_\gamma(ax+b)$ with $a>0, b$ real, where

$$G_\gamma(x) = \exp\Big(-(1+\gamma x)^{-1/\gamma}\Big), \quad 1+\gamma x > 0,$$

with extreme value index γ real and where for $\gamma = 0$ the right-hand side is interpreted as $\exp(-e^{-x})$.

Besides Definition 1.2.1, there are several equivalent definitions of belonging to a domain of attraction. von Mises established a sufficient condition for F to be in the domain of attraction of G_γ.

DEFINITION 1.2.2. *The distribution function F with its right endpoint $x^\star = \sup\{x : F(x) < 1\}$ is said to satisfy von Mises' condition, if $F''(x)$ exists, $F'(x)$ is positive for all x in some left neighborhood of $x^\star$, and*

$$\lim_{t \uparrow x^\star} \left(\frac{1-F}{F'} \right)'(t) = \gamma,$$

where $\gamma \in \mathbb{R}$ is some constant.

The distribution function F satisfying von Mise's condition with some γ is in the domain of attraction of G_γ, see Theorem 1.1.8 in [12].

There are three classes for the extreme value distributions G_γ.

- Type I (Gumbel distribution): with $\gamma = 0$,

$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}.$$

- Type II (Fréchet distribution): for $\gamma > 0$, use $G_\gamma((x-1)/\gamma)$ and get $\alpha = 1/\gamma > 0$,

$$\Phi_\alpha(x) = \begin{cases} 0 & x \le 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases}$$

- Type III (Weibull distribution): for $\gamma < 0$, use $G_\gamma(-(1+x)/\gamma)$ and get $\alpha = -1/\gamma > 0$,

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{-\alpha}) & x < 0 \\ 1 & x \ge 0 \end{cases}.$$

Notice that the Gumbel distribution is light-tailed with $1 - \Lambda(x) \sim e^{-x}$, as $x \to \infty$. Fréchet distribution has a rather heavy right tail with $1 - \Phi_\alpha(x) \sim \alpha^\alpha x^{-\alpha}$, as $x \to \infty$. Weibull distribution has a short tail with the right endpoint α and $1 - \Psi_\alpha(\alpha - x) \sim (\alpha^{-1}x)^\alpha$, as $x \downarrow 0$.

1.2.1 Gumbel domain of attraction

The distribution function F is said to be in the Gumbel domain of attraction if (1.2.9) holds with $\gamma = 0$, i.e.

$$\lim_{N\to\infty} F^N(a_N x + b_N) = \exp(-e^{-x}), \quad x \in \mathbb{R}. \tag{1.2.10}$$

Gaussian distribution is a common example of a distribution in the Gumbel domain of attraction.

DEFINITION 1.2.3. *For any nondecreasing function f, $f^{\leftarrow}$ is the left-continuous inverse if*

$$f^{\leftarrow}(x) := \inf\Big\{y : f(y) \ge x\Big\}.$$

Let the function U be the left-continuous inverse of $1/(1-F)$. Define the function f by

$$f(t) := \frac{1 - F(t)}{F'(t)}.$$

If von Mises' condition in Definition 1.2.2 is satisfied for $\gamma = 0$, then the normalizing constants in (1.2.10) can be chosen as

$$a_N = f(U(N)), \quad b_N = U(N).$$

See Section 1.2 in [12] for more details.

1.2.2 Extremes of stationary sequence

Let $\tilde{X}_1, \tilde{X}_2, \cdots$ be a (strictly) stationary sequence of random variables with marginal distribution function F. The assumption entails that for integer $h \geq 0$ and $N \geq 1$, the distribution of the random vector $(\tilde{X}_{h+1}, \cdots, \tilde{X}_{h+N})$ does not depend on h. We also want to find the limiting distribution of the sample maximum $\tilde{M}_N = \max(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_N)$. However, the limit distribution need not be the same as for the maximum $M_N = \max(X_1, X_2, \cdots, X_N)$ of the associated, independent sequence $\{X_i\}$ with the same marginal distribution as $\{\tilde{X}_i\}$.

Let $M(I) = \max_{i \in I} \tilde{X}_i$ and

$$\mathcal{I}_{j,k}(u_n) = \{\{M(I) \leq u_n\} : I \subseteq \{j, \cdots, k\}\}$$

to be the set of all intersections of the events $\{\tilde{X}_i \leq u_n\}, j \leq i \leq k$. A mixing condition known as the $D(u_N)$ condition [17] is a sufficient condition that ensure that the $\tilde{M}_N$ and M_N have the similar limit distributions.

DEFINITION 1.2.4. *For all $A_1 \in \mathcal{I}_{1,l}(u_N), A_2 \in \mathcal{I}_{l+s,N}(u_N)$ and $1 \leq l \leq N - s$,*

$$|P(A_1 \cap A_2) - P(A_1)P(A_2)| \leq \alpha(N, s)$$

and $\alpha(N, s_N) \to 0$ as $N \to \infty$ for some positive integer sequence s_N such that $s_N = o(N)$. This is called $D(u_N)$ condition.

The $D(u_N)$ condition says that any two events of the form $\{M(I_1) \leq u_N\}$ and $\{M(I_2) \leq u_N\}$ can become approximately independent as N increases when the index sets $I_i \subset \{1, \cdots, N\}$

are separated by a relatively short distance $s_N = o(N)$. Hence the $D(u_N)$ condition limits the long-range dependence between such events.

Theorem 10.4 in [11] showed the relationship between the $\tilde{M}_N$ and M_N.

THEOREM 1.2.1. *Suppose there exist sequences of constants $a_N > 0$ and b_N and a non-degenerate distribution function $G(x)$ such that* (1.2.9) *is satisfied,i.e.*

$$P\Big(\frac{M_N - b_N}{a_N} \leq x\Big) \xrightarrow{d} G(x), \quad N \to \infty.$$

if $D(u_N)$ holds with $u_N = a_N x + b_N$ for each x such that $G(x) > 0$ and if $P((\tilde{M}_N - b_N)/a_N \leq x)$ converges for some x, then

$$P\Big(\frac{\tilde{M}_N - b_N}{a_N} \leq x\Big) \xrightarrow{d} \tilde{G}(x) := G^{\theta}(x), \quad N \to \infty,$$

for some constant $\theta \in [0, 1]$, which is called the extremal index.

For a stationary sequence $\tilde{X}_1, \tilde{X}_2, \cdots$ with the Gaussian marginal distribution, we can specify how close the distributions of $\tilde{M}_N$ and M_N will be. [18] showed that

$$P(\tilde{M}_N \leq u_N) - P(M_n \leq u_N) \sim e^{-C} R_N.$$

for some sequence u_N satisfying

$$N(1 - F(u_N)) \to C > 0, \quad N \to \infty.$$

The term of R_N has the order

$$N^{-(1-\rho)/(1+\rho)}(logN)^{-\rho/(1+\rho)}$$

for some $0 < \rho \leq 1$, and $N^{-1} \log N$ for $\rho = 0$.

1.2.3 Regular Variation

Recall that the Fréchet distribution is heavy-tailed. The fundamental relationship between regular variation and extreme value theory is that a random variable X lies in the domain of attraction of a Fréchet distribution with parameter $\alpha > 0$ if and only if $P(X > \cdot)$ is regularly varying with index $-\alpha$. Regularly varying random variables form thus the entire domain of attraction of the Fréchet extreme value distribution.

DEFINITION 1.2.5. *A measurable function $f : [a, \infty) \to \mathbb{R}_+$ is regularly varying with index $\alpha > 0$ if, for all $t > 0$,*

$$\frac{f(tu)}{f(u)} \to t^{\alpha} > 0, \quad u \to \infty.$$

We write $f \in RV_\alpha$. If $\alpha = 0$, f is said to be slowly varying (at infinity).

Slowly varying functions are generically denoted by $L(u)$. Functions of the form $R(u) = u^\alpha L(u)$ are exactly regularly varying with index α.

To define a regular varying random element in a separable Banach space $\mathbb{B}$, we define $B_\epsilon := \{z \in \mathbb{B} : \|z\| < \epsilon\}$ be the open ball of radius $\epsilon > 0$, centered at the origin. Hilbert space is an example of Banach space. A Borel measure μ defined on $\mathbb{B}_0 := \mathbb{B}\backslash\{\mathbf{0}\}$ is said to be *boundedly finite* if $\mu(A) < \infty$, for all Borel sets that are bounded away from $\mathbf{0}$, that is, such that $A \cap B_\epsilon = \emptyset$, for some $\epsilon > 0$. Let $\mathbb{M}_0$ be the collection of all such measures. For $\mu_n, \mu \in \mathbb{M}_0$, we say that the μ_n converge to μ in the M_0 topology, if $\mu_n(A) \to \mu(A)$, for all bounded away from $\mathbf{0}$, μ-continuity Borel sets A, i.e., such that $\mu(\partial A) = 0$, where $\partial A := \overline{A} \setminus A^\circ$ denotes the boundary of A. The M_0 convergence can be metrized such that $\mathbb{M}_0$ becomes a complete separable metric space (Theorem 2.3 in [19] and also Section 2.2. of [20]). Then we can give the definition of a regularly varying element in $\mathbb{B}$.

DEFINITION 1.2.6. *A random element X in $\mathbb{B}$ is regularly varying with index $\alpha > 0$ if there exists $V(u) \in RV_{-\alpha}$ and*

$$\frac{P(X \in u\cdot)}{V(u)} \xrightarrow{M_0} \mu(\cdot), \quad u \to \infty. \tag{1.2.11}$$

for some non-null meaure μ on the Borel σ-field $\mathcal{B}(\mathbb{B}_0)$ of $\mathbb{B}_0 = \mathbb{B}\setminus\{\mathbf{0}\}$.

$V(u) = P(\|X\| > u)$ is regularly varying with index $-\alpha$ hence this choice suits for (1.2.11). There are some equivalent definitions of regular variation, which we state in Proposition 1.2.1. See Section 2.2. of [20] and [21] for more details.

PROPOSITION 1.2.1. *Let X be a random element in a separable Banach space $\mathbb{B}$ and $\alpha > 0$. The following three statements are equivalent:*

(i) For some slowly varying function L,

$$P(\|X\| > u) = u^{-\alpha}L(u) \tag{1.2.12}$$

and

$$\frac{P(u^{-1}X \in \cdot)}{P(\|X\| > u)} \xrightarrow{M_0} \mu(\cdot), \quad u \to \infty,$$

where μ is a non-null measure on the Borel σ-field $\mathcal{B}(\mathbb{B}_0)$ of $\mathbb{B}_0 = \mathbb{B}\setminus\{\mathbf{0}\}$.

(ii) There exists a regularly varying sequence a_N with index $1/\alpha$ such that

$$NP\left(X \in a_N \cdot\right) \xrightarrow{M_0} \mu(\cdot), \quad N \to \infty,$$

for measure μ same as the one in (i).

(iii) There exists a probability measure Γ on the unit sphere $\mathbb{S}$ in $\mathbb{B}$ such that, for every $t > 0$,

$$\frac{P(\|X\| > tu, X/\|X\| \in \cdot)}{P(\|X\| > u)} \xrightarrow{w} t^{-\alpha}\Gamma(\cdot), \quad u \to \infty.$$

(iv) Relation (1.2.12) holds, and for the same spectral measure Γ in (iii),

$$P\left(X/\|X\| \in \cdot \mid \|X\| > u\right) \xrightarrow{w} \Gamma(\cdot), \quad u \to \infty.$$

If any one of the equivalent conditions in Proposition 1.2.1 hold, we shall say that X is regularly varying with index α. The measures μ and Γ will be referred to as exponent and angular measures of X, respectively. The exponent measure μ satisfies

$$\mu(tA) = t^{\alpha}\mu(A), \ \ \forall t > 0, \ \ A \in \mathbb{B}_0.$$

It admits the polar coordinate representation via the angular measure Γ. That is, if $x = r\theta$, where $r := \|x\|$ and $\theta = x/\|x\|$, for $x \neq \mathbf{0}$, we have

$$\mu(dx) = \alpha r^{-\alpha-1} dr \Gamma(d\theta).$$

This means that for every bounded measurable function f that vanishes on a neighborhood of $\mathbf{0}$, we have

$$\int_{\mathbb{B}} f(x)\mu(dx) = \int_{\mathbb{S}}\int_0^{\infty} f(r\theta)\alpha r^{-\alpha-1} dr \Gamma(d\theta).$$

Chapter 2

Change point and trend analyses of annual expectile curves of tropical storms

2.1 Introduction

A great deal of research in environmental and climate sciences has been dedicated to detecting change points and trends in various time series, including those related to temperature, precipitation and wind speed. In a typical setting, a scalar time series $X_1, X_2, \ldots, X_N$ is analyzed. Sometimes several correlated series are considered. Most environmental and climate series exhibit a pronounced annual periodicity which must be removed, or otherwise accounted for, before statements on change–points or trends can be inferred. Sometimes, it is difficult to approximate the periodic component by a Fourier expansion due to the irregular domain and amplitude of observations within a year. The data that motivate this work are tropical storm wind speed data, examples are shown in Figure 2.1 and Figure 2.2. By definition, only storms having the wind speed over 63 kilometers per hour are considered as tropical storms. The onset and end of typhoon and hurricane seasons, as well as their intensity, can change from year to year. We therefore propose to treat the data available for a whole year as a single high–dimensional data object and perform the change point and trend analyses on these objects rather than the scalar observations directly. Such an approach is now relatively well–established in the field of functional data analysis (FDA), the monographs of [6] or [22] contain many examples. Methodological foundations of FDA are addressed in [1], its mathematical foundations in [2]. While the amount of information available in the data is invariably reduced by various smoothing and dimension reduction methods, the most important and relevant features of the data come into focus. In the problems we study in this paper, we are interested in the evolution of the annual pattern of tropical storms strength over several decades, not in specific hourly measurements.

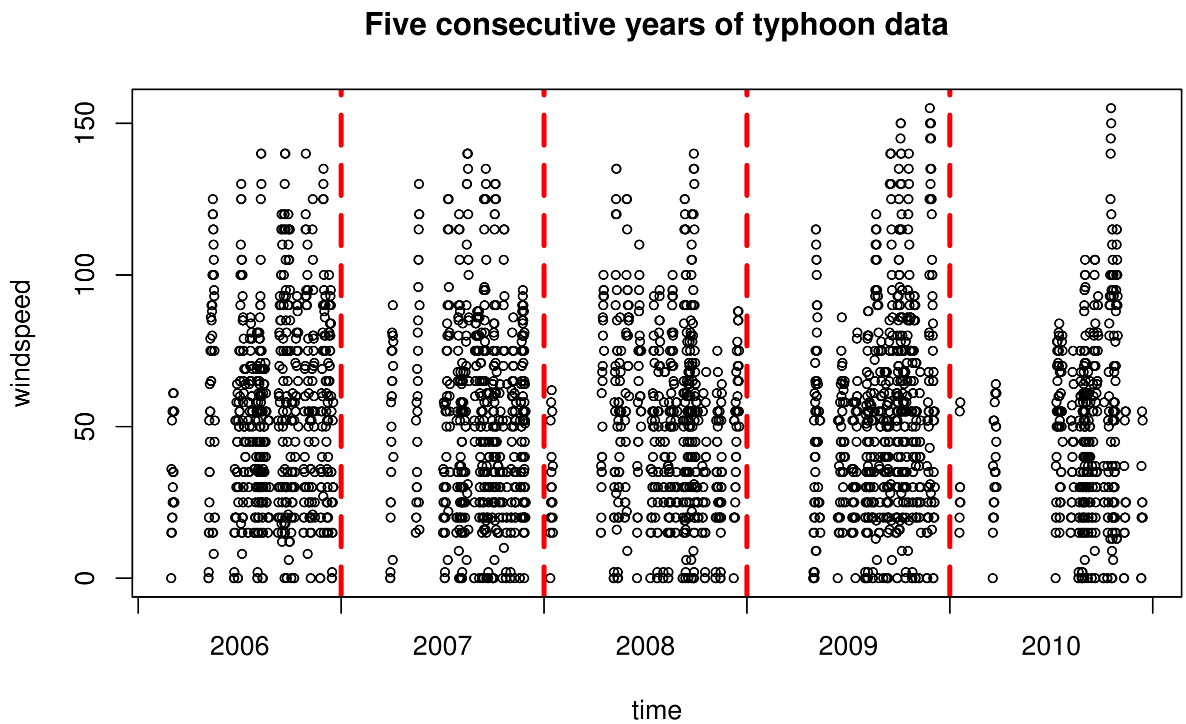

Figure 2.1: Five consecutive years (2006-2010) of typhoon data. The dots represent the wind speed measurements. Dashed vertical lines separate the years.

The data objects that this paper studies have the form $X_n(t)$, where n refers to year, and t to time within the year. In the framework of functional data analysis, t is viewed as a continuous argument. The data are observed at a regular or irregular grid, but are converted to functional objects by means of various basis expansions which are defined for every t. We consider a sequence of curves $X_n(t, \tau)$ for several expectile levels $\tau \in (0, 1)$; these are similar to quantile levels. Examples of expectile curves we study are given in Figure 2.2.

The index $\tau \in (0, 1)$ has the following interpretation. If $\tau = 0.5$, the curve $X_n(t, \tau)$ describes the median strength of tropical storms throughout the year. If τ is close to 1, the curve $X_n(t, \tau)$ captures the annual pattern of highest wind speeds. If τ is close to zero, it does the same for the lowest wind speeds. We are interested in detecting change points and trends in the functional time series $X_1(\cdot, \tau), X_2(\cdot, \tau), \ldots, X_N(\cdot, \tau)$. For this purpose, we use the existing change point test of [23] and develop two trend tests. No trend tests have presently been available for the data

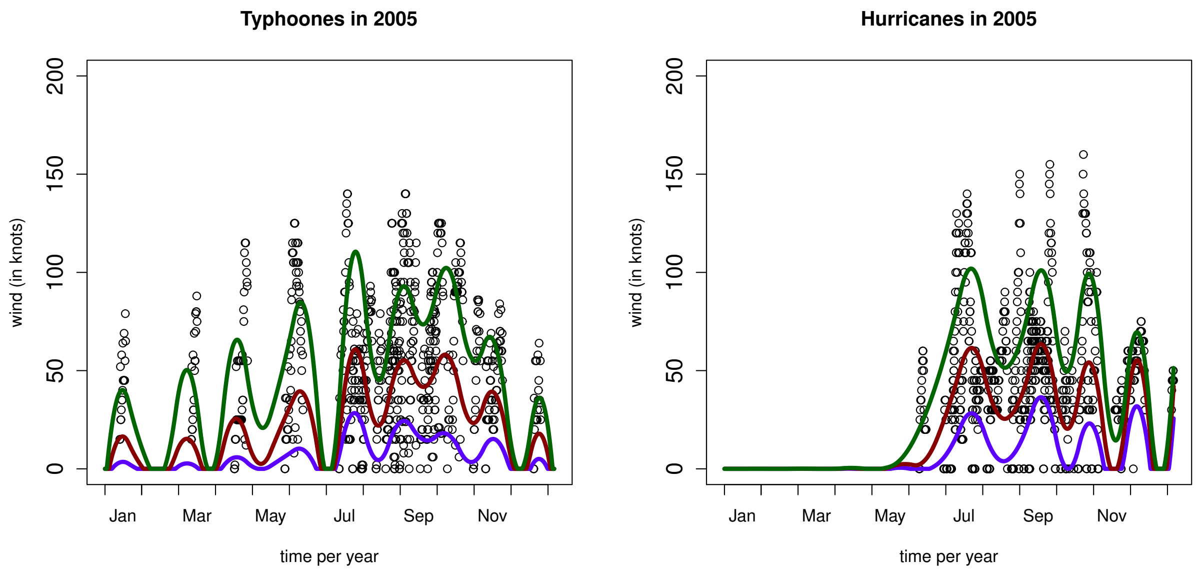

Figure 2.2: Typhoons (left) and hurricanes (right) data in 2005 with expectile curves for $\tau = 0.1, 0.5$ and 0.9. The dots represent the wind speed measurements. Generally, a vertical streak of dots represents one tropical storm event. The lines are the estimated expectile curves.

structure described above. These two tests form a methodological contribution to statistics, while the analysis of the expectile curves of tropical storms provides an insight to climate science.

We thus focus not only on the average pattern but on change points and trends in annual curves which describe the behaviour at various levels of wind speed. This is illustrated in Figure 2.2. The curves in the middle summarize the pattern of average wind speed. These curves will exhibit some evolution from year to year. The curves above them summarize the annual patterns of the highest speeds; they may exhibit a different evolution than the average curves. This issue is well–known in climate research; typically trends in the averages are contrasted with trends in extremes. In our application, no modeling of extreme behaviour is required, the expectile curves are within the range of the data points. They provide information of behaviour which lies between the typical behaviour and the unobservable extreme behaviour. Following the work of [24], evaluation of trends in extremes has attracted a great deal of attention, with respect to change point analysis of extremes, we are aware only of the work of [25].

The paper is organized as follows. After reviewing the notion of expectile curves in Section 2.2, we review in Section 2.3 the test of [23] and present the two trend tests. Section 2.4 presents the results of a simulation study. The tests are applied in Section 2.5 to the analysis of expectile curves. The last section contains the details of the asymptotic theory for the trend tests.

2.2 Expectile curves

In this section we provide some background needed to understand how the expectile curves studied in this paper are constructed. The underlying concept of expectiles was first discussed by [26] and further analyzed in several directions, for example [27] and [28] focused on time-varying expectiles. Most relevant to our setting is the paper by [29], which extended the work of [30]. It combined the LAWS (least average weighted squares) algorithm with P-splines in order to estimate expectile curves. Recent applications include [31–33] or more applicable one in finance by [34], where Value at risk (VaR) and Expected shortfall (ES) were estimated using expectiles. Expectiles have a similar interpretation as quantiles, but have some desirable properties outlined in the references cited above.

Consider a scatter plot of points $(t_i, x_i),\ 1 \leq i \leq I$. In our applications, the t_i correspond to times within a year at which wind speed is measured and x_i to the wind speed. Since the form of the dependence of the x_i on the t_i is unknown, a B-spline expansion is used. We thus assume that

$$x_i \approx g_a(t_i) = \sum_{j=1}^{J} a_j B_j(t_i),$$

and find coefficients $a = (a_1, a_2, \ldots, a_J)$ which minimize

$$S_\tau(a) = (1-\tau)S_-(a) + \tau S_+(a),$$

where

$$S_-(a) = \sum_{x_i \leq g_a(t_i)} \{x_i - g_a(t_i)\}^2 \quad \text{and} \quad S_+(a) = \sum_{x_i > g_a(t_i)} \{x_i - g_a(t_i)\}^2 .$$

If τ is close to 1, then $S_+(a)$ must be made small. This means that the curve g_a will be above most of the points (t_i, x_i).

Denote a matrix of B-splines differences as D. In order to control the smoothness of curves we can add penalization and minimize

$$S_\tau(a) + \lambda a^\top D^\top D a,$$

with λ as shrinkage parameter chosen by a desired criterion. We chose λ according to AIC criterion. After finding $\hat{a}_j$ using penalized spline estimation, the expectile curve is obtained as $\sum_{j=1}^{J} \hat{a}_j B_j(t_i)$. For our computation we set up J=20. The estimation algorithm is implemented in the `R` package `expectreg`, see [35]. Further details are presented in [36] or [37].

2.3 Change point and trend tests

This section presents the significance tests that will be applied to tropical storm data in Section 2.5. The change point test described in Section 2.3.1 was derived by [23], it is also described in Chapter 6 of [6]. Trend tests introduced in Section 2.3.2 are new; their full large sample justification is presented in the last section. In both inferential settings, we consider as sequence of curves $X_n(t), t \in [0,1],\ n = 1, 2, \ldots N$. The index n can be identified with year, the index t with time within the year normalized to unit interval. The exposition that follows uses now fairly standard concepts of functional data analysis, including functional principal components (FPC's) and their empirical counterparts (EFPC's), see, for example, Chapter 3 of [6].

2.3.1 Change point test

In change point tests, the null hypothesis is that the mean function does not change with year:

$$H_0 : \ \mathsf{E}X_1 = \mathsf{E}X_2 = \ldots = \mathsf{E}X_N.$$

The specific value of the mean is not part of the null hypothesis. The alternative is that there is at least one *unknown* change point k^* such that the equality under H_0 fails. The theory and practice

Table 2.1: Critical values of the distribution of K_d, which approximates the distribution of the statistic $\widehat{S}_d$ for large N.

d	5	6	7	8	9	10	11	12
10%	1.2797	1.4852	1.6908	1.8974	2.0966	2.2886	2.4966	2.6862
5%	1.4690	1.6847	1.8956	2.1242	2.3227	2.5268	2.7444	2.9490
1%	1.8667	2.1260	2.3423	2.5893	2.8098	3.0339	3.2680	3.4911

of change points tests have been described in many textbooks, for example, [38–40], so we do not dwell on the background and move on to the description of the test of [23].

The test is based on the normalized differences of estimated mean functions:

$$P_k(t) = \frac{k(N-k)}{N} \left\{ \hat{\mu}_k(t) - \tilde{\mu}_k(t) \right\},$$

where

$$\hat{\mu}_k(t) = k^{-1} \sum_{i=1}^{k} X_i(t), \quad \tilde{\mu}_k(t) = (N-k)^{-1} \sum_{i=k+1}^{N} X_i(t).$$

Next, we compute the estimated functional principal components $\hat{v}_\ell$ of the curves X_n and calculate the scores

$$\hat{\xi}_{j,n} = \int_0^1 \left\{ X_n(t) - \bar{X}_N(t) \right\} \hat{v}_j(t) dt, \quad \bar{X}_N(t) = N^{-1} \sum_{n=1}^{N} X_n(t). \tag{2.3.1}$$

We find the smallest d such that 85% of the variance is explained and calculate the test statistic

$$\widehat{S}_d = \frac{1}{N^2} \sum_{j=1}^{d} \frac{1}{\hat{\lambda}_j} \sum_{k=1}^{N} \left(\sum_{1 \le i \le k} \hat{\xi}_{j,i} - \frac{k}{N} \sum_{1 \le i \le k} \hat{\xi}_{j,i} \right).$$

As $N \to \infty$, the statistics $\widehat{S}_d$ converges in distribution to the random variable K_d whose critical values are given Table 2.1, see [6] for more details.

2.3.2 Trend tests

Suppose the functions $X_n(t)$ follow the trend model

$$X_n(t) = \alpha(t) + \beta(t)n + \varepsilon_n(t). \tag{2.3.2}$$

The testing problem in our setting is

$$H_0 : \beta = 0, \quad \text{vs.} \quad H_A : \beta \neq 0.$$

The paper thus focuses on a *linear* trend, which is the most common type of trend considered in atmospheric sciences. The review paper of [41] discusses research on linear trends in the context of tropical storms. The assumption of a linear trend makes the development of significance tests easier and leads to readily interpretable results if the null is rejected. More general nonlinear trends can often be displayed using various smoothing methods, but the assessment of their significance and interpretation are difficult due the lack of a simple parametrization. It is however possible to develop tests based on different approaches. [42] propose a permutation test based on the proportion of time t the curve $X_n(t)$ matches the record curve $r_n(t) = \max_{1\le k\le n} X_k(t)$. We are however not aware of other approaches to test the presence of an increasing trend in a sequence of functions. [43] consider curves $X(\mathbf{s}_k, t)$ defined at spatial locations $\mathbf{s}_k$ and test $H_0 : \beta = 0$ in the model $X(\mathbf{s}_k, t) = \alpha + \beta t + \varepsilon(\mathbf{s}_k, t)$.

Before proceeding with the description of our testing approach we state the assumptions on the objects appearing in (2.3.2).

ASSUMPTION 2.3.1. *The error curves ε_n are iid elements of the Hilbert space of square integrable functions with finite second moment: $E\int \varepsilon_n^2(t)dt < \infty$. The functions α and β are deterministic elements of that space: $\int \alpha^2(t)dt < \infty$, $\int \beta^2(t)dt < \infty$.*

Assumption 2.3.1 holds throughout the paper.

A natural approach to testing is based on an estimator of β. If this estimator is small for all $t \in [0, 1]$, there is not enough evidence to reject H_0.

Representing trend model 2.3.2 as the regression

$$\begin{bmatrix} X_1(t) \\ \vdots \\ X_N(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & N \end{bmatrix} \cdot \begin{bmatrix} \alpha(t) \\ \beta(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \vdots \\ \varepsilon_N(t) \end{bmatrix},$$

we obtain the least squares estimators

$$\hat{\alpha}(t) = \frac{2}{N(N-1)} \sum_{k=1}^{N} (2N + 1 - 3k) X_k(t) \tag{2.3.3}$$

and

$$\hat{\beta}(t) = \frac{6}{N(N+1)(N-1)} \sum_{k=1}^{N} (2k - N - 1) X_k(t). \tag{2.3.4}$$

Our first approach is based on the statistic $\int_0^1 \hat{\beta}^2(t)dt$. To describe its asymptotic distribution additional notation is needed. Introduce the covariance function of the errors $c_\varepsilon(t, s) = \mathsf{E}[\varepsilon_n(t)\varepsilon_n(s)]$. Denote by $\lambda_j, j = 1, 2, \ldots$ the eigenvalues of c_ε. Next, define the residuals

$$\hat{\varepsilon}_n(t) = X_n(t) - \hat{\alpha}_n(t) - \hat{\beta}_n(t)n \tag{2.3.5}$$

and denote by $\hat{\lambda}_j$ the eigenvalues of the empirical covariance function

$$\hat{c}_\varepsilon(t, s) = \frac{1}{N} \sum_{n=1}^{N} \hat{\varepsilon}_n(t)\hat{\varepsilon}_n(s). \tag{2.3.6}$$

Theorem 2.3.1 describes large sample properties of the suitably normalized statistic $\int_0^1 \hat{\beta}^2(t)dt$.

THEOREM 2.3.1. *(i) Under H_0,*

$$\widehat{\Lambda}_N = \frac{N^3}{12} \int_0^1 \left(\hat{\beta}(t)\right)^2 dt \xrightarrow{\mathcal{L}} \Lambda_\infty \stackrel{def}{=} \sum_{j=1}^{\infty} \lambda_j Z_j^2, \tag{2.3.7}$$

where $\{Z_j, j \geq 1\}$ are independent standard normal variables, and the λ_j are the eigenvalues of the covariance function c_ε.

(ii) Under H_A,

$$\mathrm{P}\left\{\widehat{\Lambda}_N > q_N(\alpha)\right\} \to 1, \quad \text{as } N \to \infty, \tag{2.3.8}$$

where $q_N(\alpha)$ is the $(1-\alpha)$th quantile of the distribution of $\Lambda_N = \sum_{j=1}^N \hat{\lambda}_j Z_j^2$.

Theorem 2.3.1 is proven in the last section.

The distribution of Λ_∞ can be approximated by the distribution of

$$\Lambda_N = \sum_{j=1}^{N} \hat{\lambda}_j Z_j^2. \tag{2.3.9}$$

This leads to the Monte Carlo test whose consistency is claimed in part (ii) of Theorem 2.3.1. To implement the test, we generate a large number, say $R = 10^4$, of independent replications of Λ_N (the $\hat{\lambda}_j$ are estimated only once, from the original sample). Denote these replications by $\Lambda_{N,r}, 1 \leq r \leq R$. The P–value of the test is computed as the fraction of the $\Lambda_{N,r}$ which are greater than $\widehat{\Lambda}_N$ (computed from the data).

It is also possible to develop a test similar to the test of [23] in the sense that a limit distribution is independent of the distribution of the data. In fact, in the trend model, the limit distribution is the usual chi–square distribution. This is stated in Theorem 2.3.2, in which we use the inner product notation $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$.

THEOREM 2.3.2. *Suppose $\mathsf{E}\,\|\varepsilon\|^4 < \infty$ and*

$$\lambda_1 > \lambda_2 > \ldots > \lambda_q > \lambda_{q+1} > 0. \tag{2.3.10}$$

i) Under H_0,

$$\widehat{T}_N = \frac{N^3}{12} \sum_{j=1}^{q} \hat{\lambda}_j^{-1} \left\langle \hat{\beta}, \hat{v}_j \right\rangle^2 \xrightarrow{\mathcal{L}} \chi_q^2. \tag{2.3.11}$$

ii) If for some $1 \leq j \leq q$, $\langle \beta, v_j \rangle \neq 0$, then the test is consistent, i.e.

$$\mathrm{P}\left\{\widehat{T}_N > q(\alpha)\right\} \to 1, \quad \text{as } N \to \infty, \tag{2.3.12}$$

where $q(\alpha)$ is the $(1-\alpha)$th quantile of the chi–square distribution with q degrees of freedom.

Theorem 2.3.2 is proven in the last section.

Observe that to establish the consistency, it is not enough to assume $\beta \neq 0$ in L^2. Since the statistic $\widehat{T}_N$ is based on projections on the first q EFPC's, we must assume that the slope function β is not orthogonal to the subspace spanned by the first q FPC's.

Under the assumption of iid error curves ε_n, cf. Assumption 2.3.1, the functional principal components used in this paper offer an optimal expansion. However, if the Assumption 2.3.1 is relaxed to allow some form of weak dependence, for example the approximability introduced in [8], then a different data–driven orthonormal system may offer some advantages. For example, the long–run FPC's of [44] or the dynamic FPC's of [45] could be used. These systems however require selections of kernel functions and other tuning parameters, whose selection and impact would need to be studied. We expect that the test statistics could be formulated in an analogous way and their asymptotic distribution would have a similar form to those we derived. Some work in relation to change point tests has been done by [46]. Theoretical and practical exploration of similar extensions of trend tests is an interesting topic for future research.

2.4 Finite sample performance of the trend tests

A simulation study validating the change point test of Section 2.3.1 is reported in [23]. In this section, we examine the finite sample performance of the trend tests introduced in Section 2.3.2.

We consider two models for the error functions $\varepsilon_n(t)$. The first is a generic Gaussian model in which we take the $\varepsilon_n(t)$ to be Brownian bridges $B_n(t)$. We represent Brownian bridge as a Fourier series with stochastic coefficients (the Karhunen–Loéve expansion, see [5]):

$$B_n(t) = \sqrt{2}\sum_{j=1}^{\infty} Z_{nj}\frac{\sin(j\pi t)}{j\pi} \approx \sqrt{2}\sum_{j=1}^{J} Z_{nj}\frac{\sin(j\pi t)}{j\pi},$$

where $\{Z_j, j \geq 1\}$ are independent standard normal random variables. We set $J = 100$ so the trajectories of the B_n have similar smoothness as the typhoon and hurricane expectile curves.

The second model for the ε_n is based more directly on the tropical storm data. We proceed as follows. We consider $\tau = 0.1, 0.5, 0.9$. For each level τ, we compute the sample mean function and the sample functional principal components $\hat{v}_j(t;\tau)$ of the expectile curves $X_n(t,\tau)$. Next we compute the scores $\xi_{jn}(\tau)$ according to (2.3.1). Denote by $\sigma_j(\tau)$ the standard deviation of the $\xi_{jn}(\tau), 1 \leq n \leq N, (N = 65)$. The ε_n are generated as independent realizations of the random function

$$\varepsilon(t;\tau) = \sum_{j=1}^{q} \sigma_j(\tau) Z_j \hat{v}_j(t;\tau), \quad Z_j \sim \text{ iid N}(0,1),$$

with q determined from the original expectile curves according to the 85% rule. We thus have four models for the error curves which we refer to as BB, E1, E5, E9. The errors E1, E5, E9 are different depending on whether hurricane or typhoon data are used. The empirical rejection rates are however very similar in both cases. We display the results for the errors based on the hurricane data.

We generate artificial data according to the specification

$$X_n(t) = b\beta(t)n + \varepsilon_n(t).$$

To find empirical size, we set $\beta(t) = \beta_0(t) = 0$. To find empirical power, we use the slope functions

$$\beta_1(t) = -\frac{\cos\left(\frac{t\pi 3}{2}\right)}{100}; \quad \beta_2(t) = \frac{\sin(t\pi 20)}{100},$$

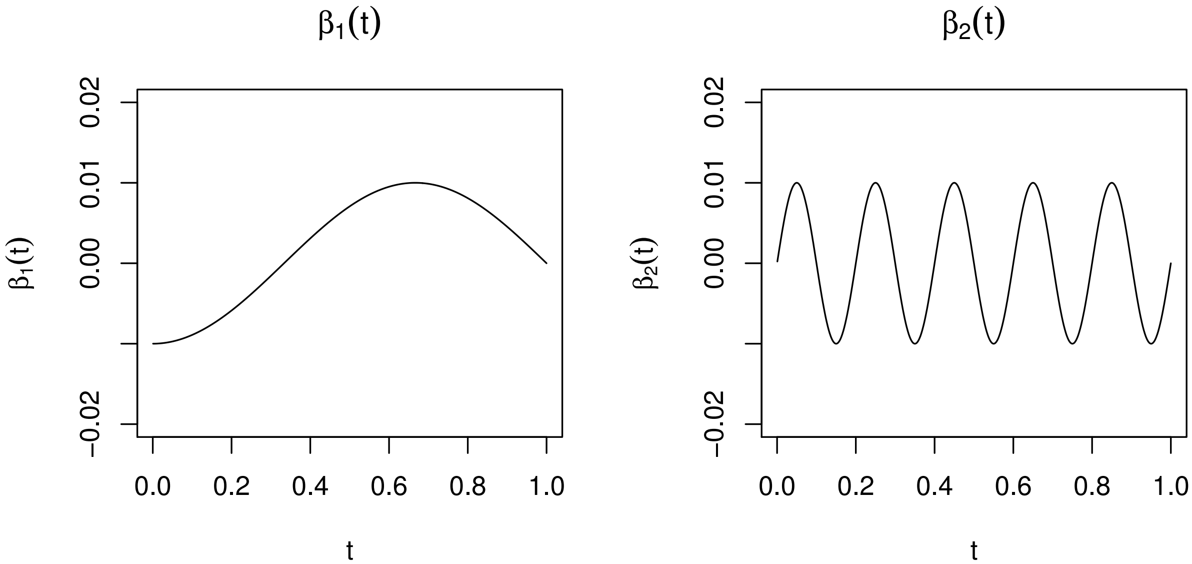

Figure 2.3: Slope functions $\beta_1(t)$ (left) and $\beta_2(t)$ (right) used to assess power.

which are graphed in Figure 2.3. The constant b is used to adjust the magnitude of the departure from the null hypothesis. For E1, E5 and E9 error curves we set $b = 20$, for BB errors we use $b = 1$. The different values are used to ensure similar signal to noise ration for both types of errors.

We consider sample sizes $N = 30, 60, 120$. Empirical rejection rates are shown in Tables 2.2 and 2.3. The Monte Carlo test, generally has slightly better size and power, but the pivotal chi–square test performs well too. The chi–square test tends to overreject under H_0 (for $N = 60$ and $N = 120$).

Table 2.2: Rejection rates of the **Monte Carlo test**. Columns corresponding to β_0 report empirical size, those to β_1 and β_2, empirical power.

BB	β_0	β_1	β_2
N=30	0.055	0.175	0.136
N=60	0.056	0.967	1.000
N=120	0.064	1.000	1.000

E1	β_0	β_1	β_2
N=30	0.060	0.082	0.078
N=60	0.045	0.438	0.440
N=120	0.042	1.000	1.000

E5	β_0	β_1	β_2
N=30	0.042	0.072	0.060
N=60	0.047	0.435	0.438
N=120	0.044	1.000	1.000

E9	β_0	β_1	β_2
N=30	0.069	0.081	0.091
N=60	0.058	0.435	0.404
N=120	0.042	1.000	1.000

Table 2.3: Rejection rates of the **Chi–square test**. Columns corresponding to β_0 report empirical size, those to β_1 and β_2, empirical power.

BB	β_0	β_1	β_2
N=30	0.064	0.344	0.053
N=60	0.058	0.995	0.085
N=120	0.069	1.000	0.238

E1	β_0	β_1	β_2
N=30	0.053	0.071	0.089
N=60	0.058	0.215	0.220
N=120	0.056	0.975	0.971

E5	β_0	β_1	β_2
N=30	0.047	0.065	0.044
N=60	0.064	0.249	0.193
N=120	0.049	0.982	0.898

E9	β_0	β_1	β_2
N=30	0.051	0.075	0.085
N=60	0.065	0.216	0.234
N=120	0.058	0.929	0.967

2.5 Application to typhoon and hurricane data

In this section we apply the tests of Section 2.3 to annual expectile curves of wind speed data. The data have the form $X_n(t_i)$, where the times t_i are separated by six hours, and the index n stands for year. The value $X_n(t_i)$ is the wind speed in knots (1 kn = 0.5144 m/s). We work with two data sets: typhoons in the West Pacific area over the period 1946–2010, and hurricanes across the North Atlantic basin over the period 1947-2011. Both datasets are accessible free of charge at the website of Unisys Weather Information, [47].

Since there are about 1,460 time points t_i per year, we treat time $0 \leq t \leq T$ within a year as continuous, and the observed curves as functional data. For each year n, we construct expectile curves $X_n(t, \tau)$, for $\tau = 0.1, 0.2, \ldots, 0.9$. Examples of expectile curves we study are given in Figure 2.2.

2.5.1 Change point analysis

The results of the application of the change–point test of Section 2.3.1 are shown in Table 2.4. For both data sets and at all levels τ, the test rejects the null hypothesis that the mean pattern does not change. As explained in Section 2.2, the construction of the expectile curves involves the selection of a smoothing parameter λ. Table 2.4 shows the results for λ selected by the AIC

Table 2.4: Results of the application of the change point test of Section 2.3.1 to typhoon (upper panel) and hurricane (lower panel) expectile curves. Usual significance codes are used: ** – significant at 5% level, *** - at 1% level.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d	10	11	12	12	12	12	12	12	12
$\widehat{S}_d$	3.3522	3.2291	3.4317	3.4978	3.6564	3.8554	4.0342	4.2317	4.5084
	***	**	**	***	***	***	***	***	***

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d	5	5	5	6	6	6	7	7	7
$\widehat{S}_d$	2.7440	3.3993	3.8759	4.4640	4.7141	4.8680	5.0366	4.9247	4.5740
	***	***	***	***	***	***	***	***	***

criterion. To validate our conclusions, we performed the same analysis using λ which is either twice or half of the λ selected by AIC. In both cases, all empirical significance levels remained under 5%.

The change point test shows that for all expectile levels τ, there are statistically significant changes in the annual pattern. It is instructive to complement the above inferential analysis by simple exploratory analysis that reveals some dependence on the level τ. Consider squared norms

$$P_k(\tau) = \int_0^T P_k^2(t,\tau)dt,$$

where the $P_k(t,\tau)$ are the normalized differences $P_k(t)$ introduced in Section 2.3.1 computed for the expectile level τ. The plot of $P_k(\tau)$ against the year index k shows the magnitude of change of the mean function. We display such plots in Figure 2.4. They suggest that the largest changes occur for the expectile levels τ close to one, but it must be kept in mind that they may just reflect the fact that the curves $X_n(t)$ are "larger" for larger τ. By contract, the statistic $\widehat{S}_d$ contains a normalization with the variances $\hat{\lambda}_j$, and is scale invariant.

The change point analysis above shows that the pattern of typhoon and hurricane wind speeds cannot be treated as stable over the sample periods we study. In the next section, we investigate if this instability can be attributed to systematic trends.

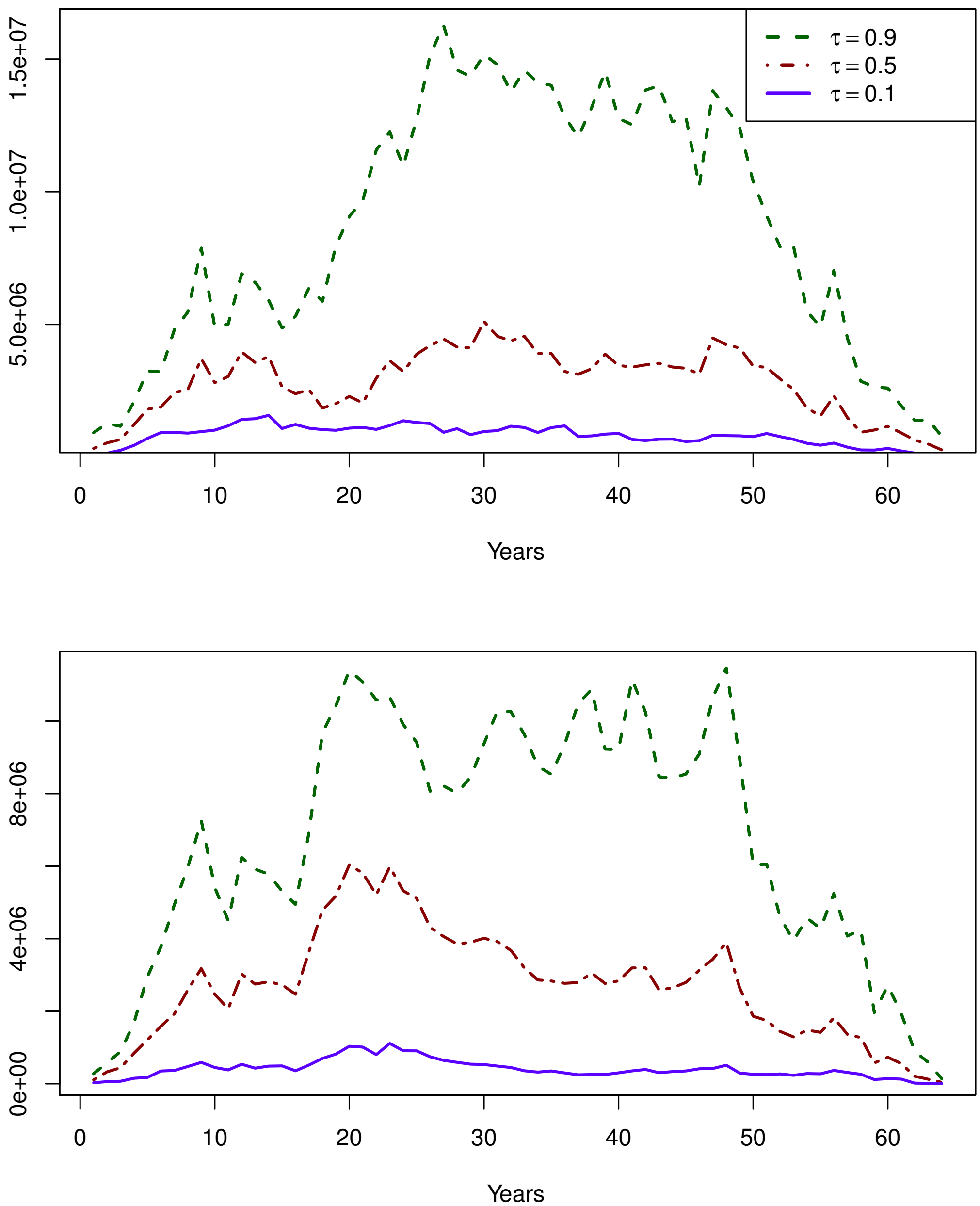

Figure 2.4: The squared norms $P_k(\tau)$ showing the magnitude of change in mean annual pattern for expectile curves of typhoons (upper panel) and hurricanes (lower panel). The largest changes occur in the expectile curves corresponding to $\tau = 0.9$.

Table 2.5: P–values for the Monte Carlo trend test based on Theorem 2.3.1.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
typhoons P–value	0.365	0.537	0.545	0.495	0.438	0.381	0.329	0.316	0.269
hurricanes P–value	0.439	0.239	0.133	0.081	0.062	0.047	0.038	0.040	0.055

Table 2.6: P–values for the chi–square trend test based on Theorem 2.3.2.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
q	10	11	12	12	12	12	12	12	12
typhoons P–value	0.534	0.705	0.722	0.688	0.587	0.466	0.382	0.371	0.453
q	5	5	5	6	6	6	7	7	7
hurricanes P–value	0.069	0.024	0.015	0.006	0.003	0.003	0.004	0.006	0.035

2.5.2 Trend analysis

We now apply the trend tests introduced in Section 2.3.2 to typhoon and hurricane expectile curves. In the Monte Carlo test based on Theorem 2.3.1, we use 10^4 replications of the random variable Λ_N defined by (2.3.9). In the chi–square test based on Theorem 2.3.2, we determine q as the smallest number which explains at least 85% of the variance of the residual curves $\hat{\varepsilon}_n$ defined by (2.3.5). The results of the tests are presented in Tables 2.5 and 2.6.

For the typhoon data, none of the two tests finds evidence of a trend. For the hurricane data, the Monte Carlo test based on Theorem 2.3.1 indicates the existence of a trend for expectile levels $\tau = 0.6 - 0.9$ while the chi–square test of Theorem 2.3.2 for all τ except $\tau = 0.1$. Simulations reported in Section 2.4 show that the chi–square test tends to overreject for data generating processes (DGP's) of length and error structure similar to the tropical storm expectile curves. We therefore conclude that there is evidence for the existence of a trend for upper expectile functions of hurricane data. The estimated slope functions $\hat{\beta}$ are plotted in Figure 2.5. While general shapes look similar, the curves are different for different values of τ, with difference of the order 0.05-0.10 on the same scale as in Figures 2.5 and 2.6.

We conclude the trend analysis by showing in Figure 2.7 the dependence on τ of the norm $\|\hat{\beta}\| = \sqrt{\int \hat{\beta}^2(t)dt}$ of the estimated slope function. Even though there is statistical evidence for nonzero slope function only for the upper expectiles of hurricane data, the exploratory analysis of the norms indicates that there is a very clear increasing dependence of the slope on τ. Again, the

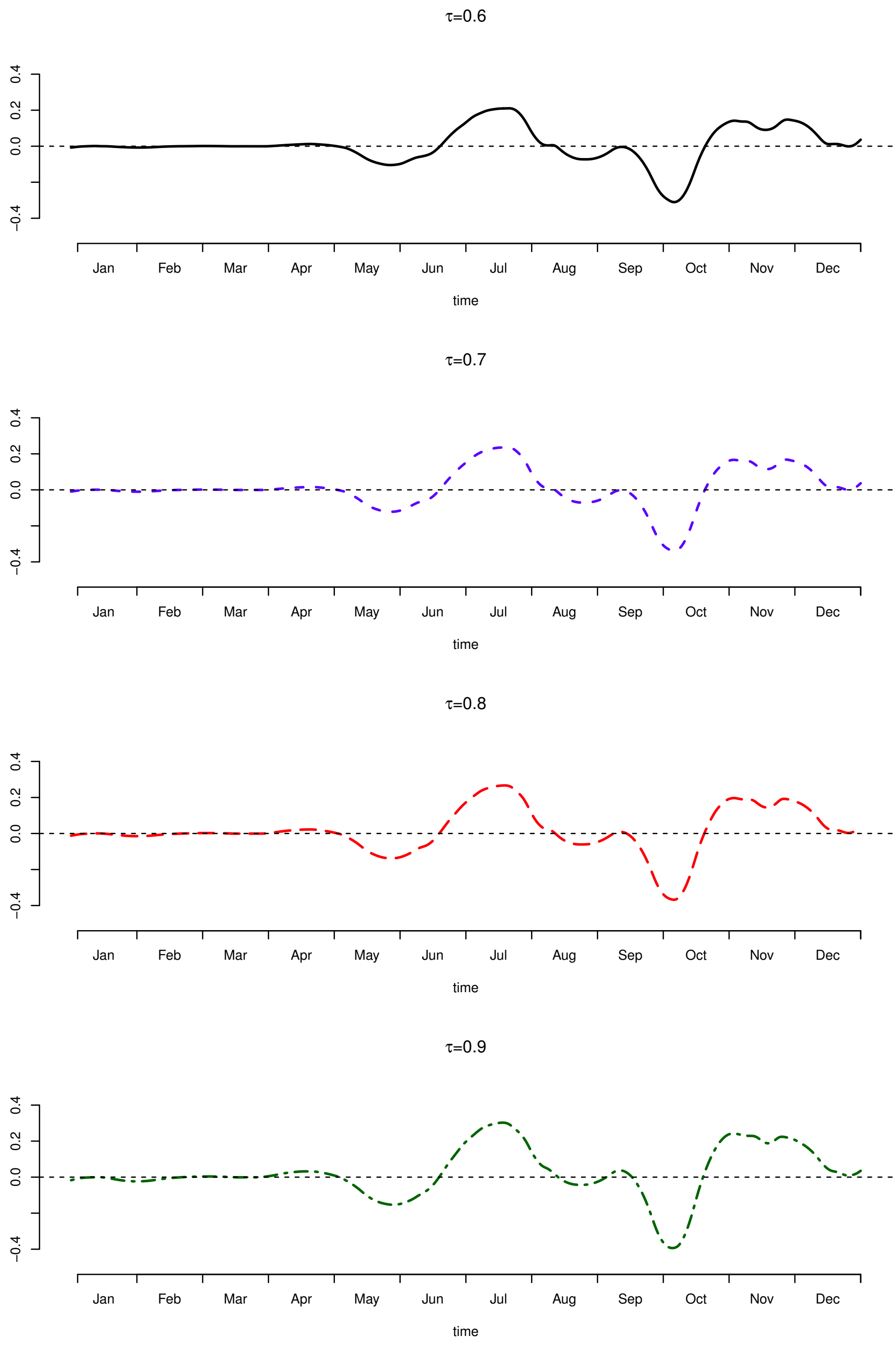

Figure 2.5: Estimated slope functions, $\hat{\beta}$, for upper expectile curves of **hurricane** data.

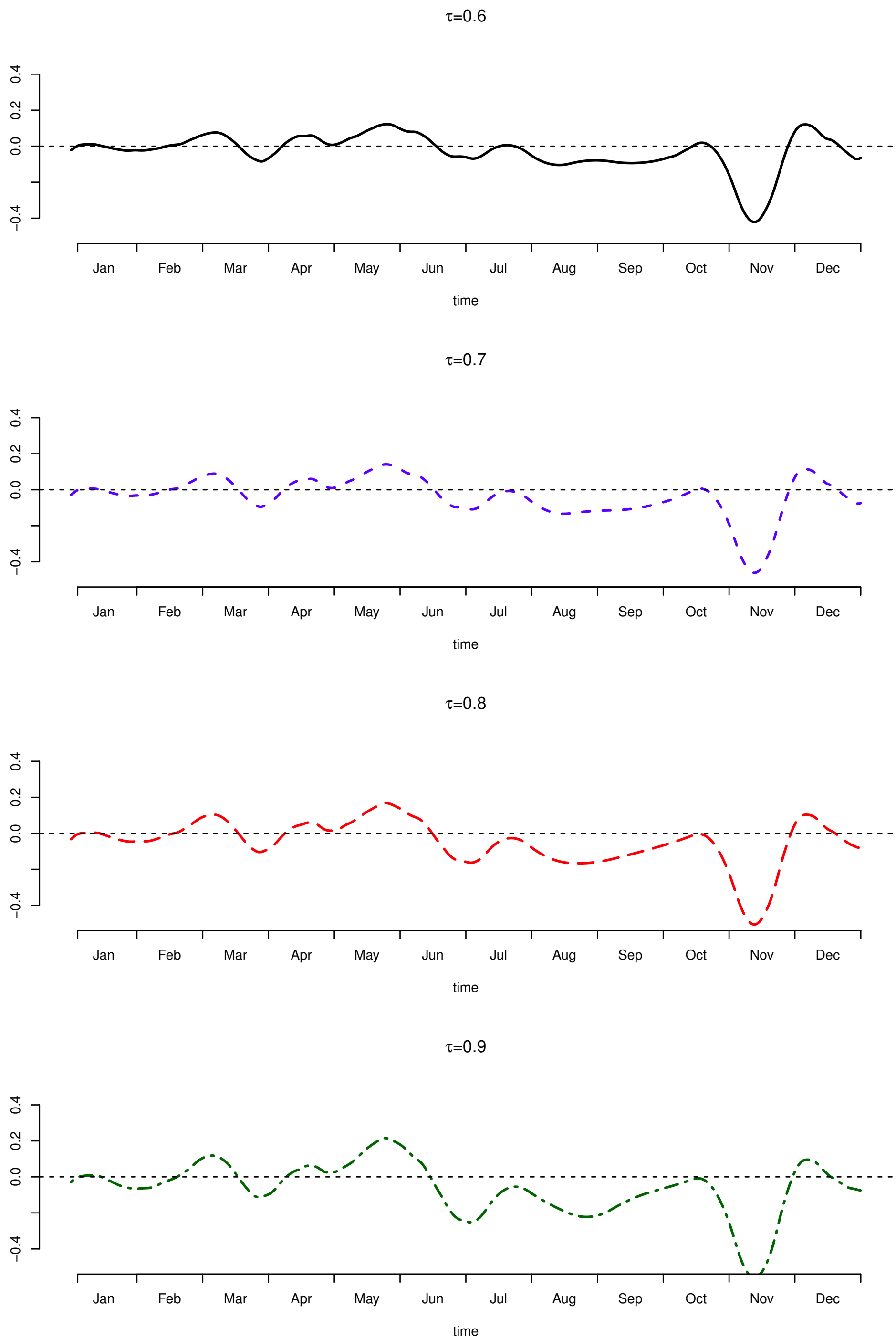

Figure 2.6: Estimated slope functions, $\hat{\beta}$, for upper expectile curves of **typhoon** data.

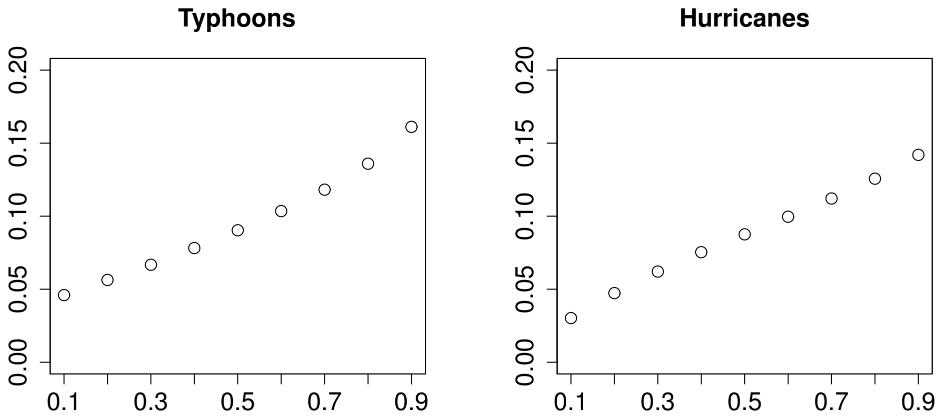

Figure 2.7: Norm of the slope function estimate, $\hat{\beta}$, as a function of the expectile level τ; typhoons (left), hurricanes (right).

increasing norms could be attributed to the increasing size of the curve $X_n(t)$, and the plots can be used only as an exploratory tool for comparing the hurricane and typhoon data.

There is not much difference between the size of $\hat{\beta}$, for typhoon and hurricane data, but the $\hat{\beta}$ for hurricanes show a clear pattern with positive mass around July and November, and negative mass in early autumn. For the typhoon curves the pattern of mass accumulation is spread more uniformly throughout the year, with a pronounced negative mass in November. The significance tests we developed provide a statistical justification for these fairly subtle visual differences.

2.5.3 Main conclusions of data analysis

The change point tests have shown that the annual pattern of wind speeds for both hurricanes and typhoons cannot be treated as constant, no matter what expectile level is considered. If there is one or two clear–cut change points, their location can be found as the years n for which the curves P_n shown in Figure 2.4 attain local maxima. For the tropical storm data, these plots show multiple local maxima indicating that either we must assume many change points or a continuous change, akin to trend. The application of the new trend tests has focused on a question which has however received a fair deal of attention: is there a trend in the intensity of tropical storms. A review of relevant research is not our aim, the paper of [41] provides background and references. There are two novel aspects to our approach: 1) focus on the annual curves, 2) separate analysis for each intensity level. Based on sixty years of data, our tests detect a trend in the upper wind speeds of Atlantic hurricanes. Exploratory analysis suggests a similar conclusion for Pacific typhoons, but it cannot be supported by low P–values with the amount of available data. These conclusions are similar to the findings of [41] who use different, custom–prepared, data sets. Their P–value for the existence of a trend in North Atlantic is less that 10^{-3}, but for the North–West pacific it is 0.03 (for South Pacific it is 0.09, 0.06 for the South Indian Ocean). Their analysis is concerned with the trend in the scalar data, not a trend in the annual pattern. They find all trends to be positive. In a sense, such trend coefficients can be viewed as averages of the annual curves like those displayed in Figures 2.5 and 2.6. The hurricane curves indeed have more positive mass, whereas for the typhoon curves the negative mass is larger (the typhoon curves are not statistically different from zero, according to our tests). The slope functions of the hurricanes indicate increasing intensity in summer and late fall, and decreasing intensity in early fall. For typhoons, these curves indicate decreasing intensity in November.

The conclusions of this paper which are supported by significance tests and do not contradict existing research are as follows:

1. The annual pattern of wind speeds of both hurricanes and typhoons has been changing at all wind speed levels over the last 60 years.

2. There is a significant trend in the shape of this pattern for upper wind speed levels of hurricanes.

2.6 Proofs of Theorems 2.3.1 and 2.3.2

Before proceeding to the proofs of Theorems 2.3.1 and 2.3.2, we observe that a direct verification shows that

$$c_\beta(t,s) \stackrel{\text{def}}{=} \text{Cov}\left\{\hat{\beta}(t), \hat{\beta}(s)\right\} = A_N c_\varepsilon(t,s),$$

where

$$A_N = \frac{12}{N(N+1)(N-1)}.$$

The constant A_N is repeatedly used in the proofs of Theorems 2.3.1 and 2.3.2.

2.6.1 Proof of Theorem 2.3.1

PROOF OF PART (I): Under H_0 ($\beta = 0$),

$$\hat{\beta}(t) = A_N \sum_{k=1}^{N} k\varepsilon_k(t) - \frac{1}{2} A_N (N+1) \sum_{k=1}^{N} \varepsilon_k(t).$$

Using the identity

$$\sum_{k=1}^{N} k\varepsilon_k = N \sum_{n=1}^{N} \varepsilon_n - \sum_{k=1}^{N-1} \sum_{n=1}^{k} \varepsilon_n, \tag{2.6.13}$$

we have

$$\begin{aligned} \hat{\beta}(t) &= A_N \sum_{k=1}^{N} k\varepsilon_k(t) - \frac{1}{2} A_N (N+1) \sum_{k=1}^{N} \varepsilon_k(t) \\ &= A_N \Big(N \sum_{n=1}^{N} \varepsilon_n(t) - \sum_{k=1}^{N-1} \sum_{n=1}^{k} \varepsilon_n(t) \Big) - \frac{1}{2} A_N (N+1) \sum_{n=1}^{N} \varepsilon_n(t). \end{aligned} \tag{2.6.14}$$

To determine the limit behaviour of $\hat{\beta}(t)$, we thus need an invariance principle for the partial sum process:

$$S_N(x,t) = \frac{1}{\sqrt{N}} \sum_{1 \le n \le [Nx]} \varepsilon_n(t), \quad 0 \le x, t \le 1.$$

A result of this type has recently been established by [48]. It states that

$$S_N(x,t) \xrightarrow{\mathcal{L}} \Gamma(x,t), \tag{2.6.15}$$

where $\Gamma(x,t)$ is the two parameter Gaussian process which admits the representation

$$\Gamma(x,t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} W_j(x) v_j(t), \tag{2.6.16}$$

where $\{W_j(x), 0 \le x \le 1\}$ are independent standard Wiener processes on $[0,1]$. The λ_j and the v_j are, respectively, the eigenvalues and the eigenfunctions of the covariance function $c_\varepsilon(t,s) = \mathsf{E}[\varepsilon_n(t)\varepsilon_n(s)]$. In (2.6.15), and whenever weak convergence of two parameter processes is concerned, $\xrightarrow{\mathcal{L}}$ denotes the convergence in the Skorokhod space $D([0,1], L^2)$.

Since $A_N \sim 12N^{-3}$, (2.6.14) implies

$$\begin{aligned}\hat\beta(t) &= A_N N^{\frac{3}{2}} S_N(1,t) - A_N N^{\frac{1}{2}} \sum_{k=1}^{N-1} S_N\left(\frac{k}{N}, t\right) - \frac{1}{2} A_N (N+1) N^{\frac{1}{2}} S_N(1,t) \\ &\sim 12 N^{-\frac{3}{2}} S_N(1,t) - 12 N^{-\frac{3}{2}} \left\{ \frac{1}{N} \sum_{k=1}^{N-1} S_N\left(\frac{k}{N}, t\right) \right\} - 6N^{-\frac{3}{2}} S_N(1,t) \\ &= 6N^{-\frac{3}{2}} S_N(1,t) - 12 N^{-\frac{3}{2}} \left\{ \frac{1}{N} \sum_{k=1}^{N-1} S_N\left(\frac{k}{N}, t\right) \right\}.\end{aligned}$$

By the continuous mapping theorem and (2.6.15)

$$\frac{1}{N} \sum_{k=1}^{N-1} S_N\left(\frac{k}{N}, t\right) \xrightarrow{\mathcal{L}} \int_0^1 \Gamma(x,t)dx.$$

Thus

$$\frac{N^{\frac{3}{2}}}{6} \hat\beta(t) \xrightarrow{\mathcal{L}} \Gamma(1,t) - 2\int_0^1 \Gamma(x,t)dx. \tag{2.6.17}$$

Using the continuous mapping theorem again, we obtain

$$\frac{N^3}{36} \int_0^1 \left\{\hat\beta(t)\right\}^2 dt \xrightarrow{\mathcal{L}} \int_0^1 \left\{ \Gamma(1,t) - 2\int_0^1 \Gamma(x,t)dx \right\}^2 dt.$$

Set

$$D_j = W_j(1) - 2\int_0^1 W_j(x)dx, \tag{2.6.18}$$

so that, by (2.6.16), we have

$$\Gamma(1,t) - 2\int_0^1 \Gamma(x,t)dx = \sum_{j=1}^{\infty} \sqrt{\lambda_j} D_j v_j(t).$$

Then, by Parseval's identity,

$$\int_0^1 \left\{\Gamma(1,t) - 2\int_0^1 \Gamma(x,t)dx\right\}^2 dt = \left\|\sum_{j=1}^{\infty} \sqrt{\lambda_j} D_j v_j\right\|^2 = \sum_{j=1}^{\infty} \lambda_j D_j^2. \tag{2.6.19}$$

The random variables D_j are independent normal with mean zero and variance

$$\begin{aligned}
\text{Var}[D_j] &= \mathsf{E}\left[W(1) - 2\int_0^1 W(x)dx\right]^2 \\
&= \mathsf{E}W^2(1) - 4\mathsf{E}\left[W(1)\int_0^1 W(x)dx\right] + 4\mathsf{E}\left[\int_0^1 W(x)dx\right]^2 \\
&= \frac{1}{3}.
\end{aligned}$$

We can write $D_j = \frac{1}{\sqrt{3}} Z_j$, where Z_j are standard normal variables. By $(2.6.19)$

$$\int_0^1 \left\{\Gamma(1,t) - 2\int_0^1 \Gamma(x,t)dx\right\}^2 dt = \frac{1}{3}\sum_{j=1}^{\infty} \lambda_j Z_j^2.$$

Thus $(2.3.7)$ is proven.

PROOF OF PART (II): The proof follows from several lemmas. It is assumed throughout that H_A holds, i.e $\|\beta\| > 0$. The argument relies on Lemma 2.6.1 whose proof follows from the relevant definitions, and so is omitted.

LEMMA 2.6.1. *Suppose* $\{X_n\}$ *and* $\{q_n\}$ *are sequences of random variables. Suppose further that* $\{X_n\}$ *diverges to infinity in probability and* $\{q_n\}$ *is bounded in probability , i.e. for each* $M, \lim_{n\to\infty} \mathrm{P}(X_n > M) = 1$ *and for each* $\varepsilon > 0$*, there are* M *and* n_0 *such that* $\mathrm{P}(q_n > M) < \varepsilon$*, if* $n > n_0$*. Then*

$$\lim_{n\to\infty} \mathrm{P}(X_n > q_n) = 1.$$

Relation (2.3.8) now follows from Lemmas 2.6.2 and 2.6.3.

LEMMA 2.6.2. *The statistic $\widehat{\Lambda}_N$ defined by* (2.3.7) *satisfies* $\widehat{\Lambda}_N \stackrel{P}{\to} \infty$.

PROOF: Decompose $\hat{\beta}(t)$ as

$$\hat{\beta}(t) = \beta(t) + G_N(t), \tag{2.6.20}$$

where

$$G_N(t) = \frac{1}{2} A_N \sum_{k=1}^{N} (2k - N - 1)\varepsilon_k(t).$$

Observe that $G_N(t)$ is equal to the estimator $\hat{\beta}(t)$ under H_0. Therefore, by (2.6.17),

$$N^{3/2} G_N(t) \xrightarrow{\mathcal{L}} 6 \left\{ \Gamma(1,t) - 2 \int_0^1 \Gamma(x,t)dx \right\} \stackrel{def}{=} U(t).$$

Consequently, as $N \to \infty$

$$N^3 \int \hat{\beta}^2(t)dt = \int \left\{ N^{3/2}\beta(t) + N^{3/2}G_N(t) \right\}^2 dt \sim \int \left\{ N^{3/2}\beta(t) + U(t) \right\}^2 dt \stackrel{P}{\to} \infty.$$

More precisely,

$$N^{-3}\widehat{\Lambda}_N \sim \frac{1}{12} \int \left\{ \beta(t) + N^{-3/2}U(t) \right\}^2 dt \stackrel{P}{\to} \frac{1}{12} \int \beta^2(t)dt.$$

LEMMA 2.6.3. *Under H_0, the sequence $\{\Lambda_N\}$ defined by* (2.3.9) *is bounded in probability.*

PROOF: Since the $\hat{\lambda}_j$ are fixed in the generation of the replications in the Monte Carlo test, the variables Z_j are independent of the $\hat{\lambda}_j$. Therefore, since $\mathsf{E}Z_j^2 = 1$,

$$\mathsf{E}\Lambda_N = \sum_{j=1}^{N} \mathsf{E}\hat{\lambda}_j.$$

The definition of the $\hat{\lambda}_j$ as the eigenvalues of the covariance operator with $\hat{c}_\varepsilon(\cdot,\cdot)$ defined by (2.3.5) and (2.3.6) implies that

$$\sum_{j=1}^{N} \hat{\lambda}_j = \frac{1}{N} \sum_{n=1}^{N} \left\| \hat{\varepsilon}_n \right\|^2.$$

This is the decomposition of functional sample variance, see details [6], p. 40. Therefore, if we can show that

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathsf{E}||\hat{\varepsilon}_n||^2 < \infty, \tag{2.6.21}$$

then we can conclude that $\limsup_{N\to\infty} \mathsf{E}\Lambda_N < \infty$, which in turn implies that the sequence $\{\Lambda_N\}$ is bounded in probability.

The decomposition

$$\hat{\varepsilon}_n(t) = \varepsilon_n(t) + \Big\{\alpha(t) - \hat{\alpha}(t)\Big\} + n\Big\{\beta(t) - \hat{\beta}(t)\Big\}, \tag{2.6.22}$$

implies that for some constant C,

$$||\hat{\varepsilon}_n||^2 \le C\Big(||\varepsilon_n||^2 + ||\hat{\alpha} - \alpha||^2 + ||n(\hat{\beta} - \beta)||^2\Big). \tag{2.6.23}$$

First note that

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} \mathsf{E}||\varepsilon_n||^2 &= \mathsf{E}\left[\int \Big\{\frac{1}{N}\sum_{n=1}^{N} \varepsilon_n^2(t)\Big\} dt\right] \\
&= \int \left\{\mathsf{E}\left[\frac{1}{N}\sum_{n=1}^{N} \varepsilon_n^2(t)\right]\right\} dt \\
&= \int \mathsf{E}\varepsilon_1^2(t) dt = \mathsf{E}||\varepsilon_1||^2 < \infty.
\end{aligned}$$

Next, observe that

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} \mathsf{E}||\hat{\alpha} - \alpha||^2 = \mathsf{E}||\hat{\alpha} - \alpha||^2& \\
&= \int \left\{\mathsf{E}\left[\hat{\alpha}(t) - \alpha(t)\right]^2\right\} dt \\
&= \int \mathsf{E}\left[\frac{2}{N(N-1)}\sum_{k=1}^{N}(2N+1-3k)\varepsilon_k(t)\right]^2 dt \\
&= \frac{2(2N+1)}{N(N-1)}\mathsf{E}||\varepsilon_1||^2 \to 0.
\end{aligned}$$

Similarly, with $H_N = \frac{(N+1)(2N+1)}{6}$,

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} \mathsf{E}||n(\hat{\beta}-\beta)||^2 &= H_N \mathsf{E}||\hat{\beta}-\beta||^2 \\
&= H_N \int \left\{ \mathsf{E}\left[\hat{\beta}(t)-\beta(t)\right]^2 \right\} dt \\
&= H_N \int \mathsf{E}\left[\frac{6}{N(N-1)(N+1)}\sum_{k=1}^{N}(2k-N-1)\varepsilon_k(t)\right]^2 dt \\
&= \frac{2(2N+1)}{N(N-1)}\mathsf{E}||\varepsilon_1||^2 \to 0.
\end{aligned}$$

Thus (2.6.21) holds. Therefore $\sup_N \mathsf{E}\Lambda_N =: C_\Lambda < \infty$, and so $\mathrm{P}(\Lambda_N > M) \leq M^{-1}C_\Lambda$ can be made arbitrarily small by choosing M sufficiently large. The conclusion follows.

2.6.2 Proof of Theorem 2.3.2

PROOF OF PART (I): Under H_0, by (2.6.14), (2.6.16) and consistency of estimated eigenfunctions $\hat{v}_j$, ($\hat{v}_j \xrightarrow{P} v_j$),

$$\begin{aligned}
\left\langle \frac{N^{\frac{3}{2}}}{6}\hat{\beta}, \hat{v}_j \right\rangle^2 &\xrightarrow{\mathcal{L}} \left\langle \Gamma(1,\cdot) - 2\int_0^1 \Gamma(x,\cdot)dx, v_j \right\rangle^2 \\
&= \left\langle \sum_{k=1}^{\infty}\sqrt{\lambda_k}W_k(1)v_k - 2\int_0^1 \sum_{k=1}^{\infty}\sqrt{\lambda_k}W_k(x)v_k, v_j \right\rangle^2 \\
&= \left[\sum_{k=1}^{\infty}\sqrt{\lambda_k}\left\{W_k(1) - 2\int_0^1 W_k(x)dx\right\}\langle v_k, v_j\rangle\right]^2 \\
&= \lambda_j\left\{W_j(1) - 2\int_0^1 W_j(x)dx\right\}^2 \\
&= \lambda_j D_j^2 = \frac{1}{3}\lambda_j Z_j^2,
\end{aligned}$$

with the random variables D_j defined in (2.6.18), and Z_j standard normal variables. It follows that

$$\widehat{T}_N = \frac{N^3}{12}\sum_{j=1}^{q}\hat{\lambda}_j^{-1}\left\langle \hat{\beta}, \hat{v}_j \right\rangle^2 = 3\sum_{j=1}^{q}\hat{\lambda}_j^{-1}\left\langle \frac{N^{\frac{3}{2}}}{6}\hat{\beta}, \hat{v}_j \right\rangle^2 \xrightarrow{\mathcal{L}} \sum_{j=1}^{q} Z_j^2 \stackrel{\mathcal{L}}{=} \chi_q^2.$$

PROOF OF PART (II): We must show that $\widehat{T}_N \xrightarrow{P} \infty$, if $\langle \beta, v_j \rangle \neq 0$ for some $1 \leq j \leq q$. It is enough to show that

$$\sum_{j=1}^{q} \hat{\lambda}_j^{-1} \left\langle \hat{\beta}, \hat{v}_j \right\rangle^2 \xrightarrow{P} \sum_{j=1}^{q} \lambda_j^{-1} \left\langle \beta, v_j \right\rangle^2,$$

because the right–hand side is positive. The verification of the above convergence reduces to

$$\left\| \hat{\beta} - \beta \right\| \xrightarrow{P} 0 \tag{2.6.24}$$

and, for $1 \leq j \leq q$,

$$\|\hat{v}_j - v_j\| \xrightarrow{P} 0, \quad \hat{\lambda}_j \xrightarrow{P} \lambda_j. \tag{2.6.25}$$

To prove relation (2.6.24), observe first that by decomposition (2.6.20),

$$\mathsf{E} \left\| \hat{\beta} - \beta \right\| = \mathsf{E} \|G_N\| \leq \left\{ \mathsf{E} \|G_N\|^2 \right\}^{\frac{1}{2}} = \left\{ \mathsf{E} \int G_N^2(t) dt \right\}^{\frac{1}{2}}.$$

To calculate the last expected value, we will use the identity

$$\frac{1}{4} A_N \sum_{k=1}^{N} (2k - N - 1)^2 = 1,$$

which follows from algebraic manipulations. The independence of the ε_k thus implies that

$$\mathsf{E} \int G_N^2(t) dt = \frac{1}{4} A_N^2 \sum_{k=1}^{N} (2k - N - 1)^2 \mathsf{E} \int \varepsilon_k^2(t) dt = A_N \mathsf{E} \|\varepsilon\|^2 = \mathcal{O}(N^{-3}).$$

By Lemmas 2.2. and 2.3 of [6], relations (2.6.25) will follow from $\|\hat{c}_\varepsilon - c_\varepsilon\|_{\mathcal{S}} \xrightarrow{P} 0$, where the subscript $\mathcal{S}$ denotes the Hilbert–Schmidt norm. Proposition 2.6.1 states that, in fact, $\mathsf{E} \|\hat{c}_\varepsilon - c_\varepsilon\|_{\mathcal{S}}^2 = \mathcal{O}(N^{-1})$. It thus extends a well–known result, e.g. Theorem 2.5. of [6], which states that

$$\mathsf{E} \int \left(\frac{1}{N} \sum_{i=1}^{N} \varepsilon_i(t) \varepsilon_i(s) - \mathsf{E} \left[\varepsilon(t) \varepsilon(s) \right] \right)^2 dt ds = \mathcal{O}(N^{-1}). \tag{2.6.26}$$

The covariance function $\hat{c}_\varepsilon$ is defined in terms of the residuals $\hat{\varepsilon}_n$, cf. (2.3.5), (2.3.6). Estimation of the intercept and slope functions introduces many additional terms which are, however, all asymptotically negligible. This is the content of the following proposition whose proof is very long as it requires the examination of 16 cross–terms. The proof is therefore not presented her, but is available upon request.

PROPOSITION 2.6.1. *Suppose model* (2.3.2) *holds and* $\mathsf{E}\,\|\varepsilon\|^4 < \infty$. *Then the sample covariance function* $\hat{c}_\varepsilon$, *defined by* (2.3.5) *and* (2.3.6), *satisfies* $\mathsf{E}\,\|\hat{c}_\varepsilon - c_\varepsilon\|_{\mathcal{S}}^2 = \mathcal{O}(N^{-1})$.

Chapter 3

Extremes of projections of functional time series on data–driven basis systems

3.1 Introduction

The work presented in this paper is motivated by a question that arises in the analysis of time series of functions, which came to be known as functional time series, many examples are studied in [6] and [49], and a large number of papers. Suppose $X_1, X_2, \ldots, X_N$ is a realization of a strictly stationary mean zero time series $\{X_i\}$ such that each X_i is a function in the space L^2 of square integrable functions on a compact interval. Many procedures of functional data analysis are based on the expansion

$$X_i(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_{ij} v_j(t), \tag{3.1.1}$$

in which the v_j are the eigenfunctions of the covariance operator and the λ_j are the corresponding eigenvalues. The precise definition will be given in Section 3.2. In applications, the infinite expansion is replaced by a finite sum which involves estimated counterparts of the quantities in (3.1.1), i.e. by

$$X_i(t) \approx \sum_{j=1}^{p} \hat{\lambda}_j^{1/2} \widehat{Z}_{ij} \hat{v}_j(t). \tag{3.1.2}$$

The idea is to replace the infinite dimensional objects, the functions X_i, by the vectors

$$\widehat{\mathbf{Z}}_i = [\widehat{Z}_{i1}, \widehat{Z}_{i2}, \ldots, \widehat{Z}_{ip}]^\top, \tag{3.1.3}$$

which can be stored in machine memory, and so are amenable to a variety of procedures. Over the last twenty years, a great deal of research in functional data analysis has focused on the examination of this approach in various settings. The central question has been how much is lost by reducing the functions X_i to the vectors $\widehat{\mathbf{Z}}_i$, and under what assumptions this loss is asymptotically negligible.

The random variables Z_{ij} in (3.1.1), called the scores, encode the shapes of the functions X_i. The $\widehat{Z}_{ij}$ are the *estimated* scores; they are computed from the whole sample $X_1, X_2, \ldots, X_N$, and are therefore dependent.

A question this paper seeks to answer is under what conditions the asymptotic extreme behavior, as $N \to \infty$, of the vectors (3.1.3) is the same as that of the unobservable vectors $\mathbf{Z}_i = [Z_{i1}, Z_{i2}, \ldots, Z_{ip}]^\top$. We focus on the case of Gaussian, or nearly Gaussian, functions X_n, as these occur in many climate research applications (as verified by QQ-plots of various projections). The simplest example is annual temperature records, $X_i(t)$, where i indexes the year, and t time, in days, within the year. These may be temperatures measured at a specific location, or temperature indexes, like Sea Surface Temperature indexes for various areas of the Pacific. There are many examples of non–Gaussian data of this type, like precipitation records. A corresponding theory for functional time series with heavy–tailed projections is not the subject of this paper. In either case, the scores encode the shapes of the annual curves, and their multivariate extremes describe the extreme shapes of these curves. This is illustrated in Figure 3.1 which shows the residual annual temperature curves obtained after subtracting the the sample mean function (in Fort Collins, Colorado) together with an "extreme curve" whose scores are the maxima of the scores of the observed curves. In accordance with the EVT paradigm, the black curve does look extreme, but it can be expected that a similar curve could be observed with a small positive probability. This functional, shape–centered, approach can be contrasted with the usual “annual maxima” approach which focuses on the series $\max_t X_i(t)$.

To illustrate mathematical difficulties arising when the scores are replaced by estimated scores, we consider only one projection, i.e. $j = 1$. In that case, $Z_{i1} = \lambda_1^{-1/2} Y_i$ and $\widehat{Z}_{i1} = \hat{\lambda}_1^{-1/2} Y_{i,N}$, where

$$Y_i = \langle X_i, v_1 \rangle \,, \quad Y_{i,N} = \langle X_i, \hat{v}_1 \rangle \,. \tag{3.1.4}$$

Assume further that the X_i are Gaussian and independent (hence identically distributed by the stationarity assumption). Then, the Y_i are iid normal, and the limit of their normalized maxima is the standard Gumbel distribution. However, even in this simplest case, the $Y_{i,N}$ are neither normal

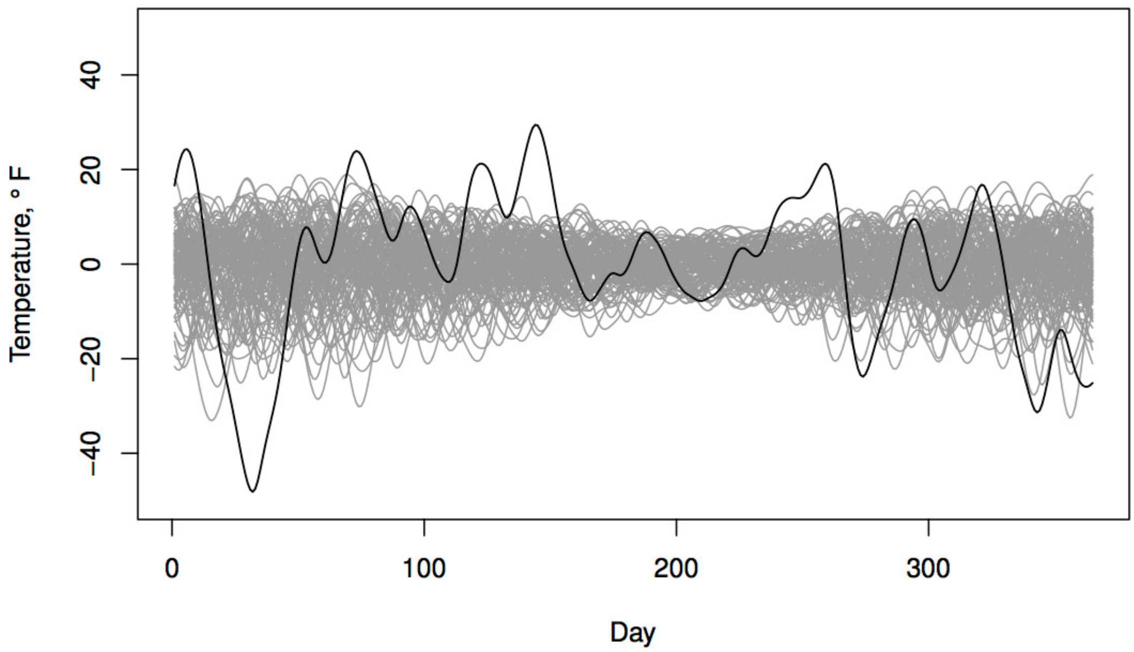

Figure 3.1: Residual temperature curves (grey) together with an "extreme curve" whose scores are the maxima of the scores of the observed curves.

nor independent (the $\hat{v}_1$ is estimated from the whole sample). They form a triangular array. [50] established conditions under which a triangular array of *normal* dependent random variables is in the Gumbel domain of attraction. Their approach is specific to Gaussian random variables, and does not apply to our problem because the $Y_{i,N}$ are not normal. Using a different technique, we will show that in the simplest case of iid Gaussian functions X_i, the extremal limiting behavior of the vectors $\widehat{\mathbf{Z}}_i$ is the same as that of the vectors $\mathbf{Z}_i$. Even this case is however not trivial, and requires a new approach.

In most applications, QQ-plots of projections and multivariate significance tests based on projections, can at best confirm that the observed functions $X_1, X_2, \dots, X_N$ are approximately normal. In our context, a relevant question is if the result mentioned above remains valid if the assumption of the exact Gaussian distribution is relaxed to the assumption that the projections are in the Gumbel domain of attraction. It turns out that for our technique of proof to be applicable in this more general context, we need to impose an additional assumption, which is however satisfied by the usual distributions in the Gumbel domain of attraction, including normal, exponential and gamma. Next, it is of practical importance to ask to what extend the assumption of independence can be relaxed. We show that the key property that is needed is a suitably formulated *Rootzen's condition*, roughly that the maxima of a dependent sequence are asymptotically equivalent to the maxima of its iid version. This condition has been shown to hold for univariate linearly dependent Gaussian sequences, including the ARMA time series. Finally, while expansions with respect to the functional principal components are most important and motivate our research, we show that the results hold in much greater generality. Essentially, all that is needed is that some population basis functions v_j can be estimated with an $O(N^{-1/2})$ rate.

This paper presents a new exploration at the nexus of Functional Data Analysis (FDA) and Extreme Value Theory (EVT). Mathematical foundations of FDA are presented in [2], a concise introduction to the subject is given in [49], the most widely read classic is [1]. There are many excellent accounts of EVT, including [9–12].

The remainder of the paper is organized as follows. Section 3.2 is concerned with convergence in distribution, while Section 3.3 focuses on the rate of convergence of the distribution functions. The proofs are presented in Sections 3.4 and 3.5.

3.2 Convergence in distribution

We assume throughout the paper that the functions $X_1, X_2, \ldots, X_N$ are a realization of a strictly stationary mean zero sequence of functions in L^2, the space of integrable functions on a compact interval $\mathcal{T}$. The assumption of stationarity could be relaxed and replaced by technical assumptions, but at the expense of making the exposition too technical. The space L^2 could be replaced by an abstract separable Hilbert space, again at the expense of more abstract formulations. The framework we consider is general enough to cover applications to functional time series. Writing each X_i as a function $X_i(t), t \in \mathcal{T}$, provides a ready connection to applications.

Each X_i admits expansion (3.1.1) in which the λ_j and v_i are, respectively, the eigenvalues and the eigenfunctions of the covariance operator defined by $C(x) = E[\langle X_i, x\rangle X_i],\ x \in L^2$, i.e. $C(v_j) = \lambda_j v_j,\ j \geq 1$, see e.g. Section 11.4 of [49]. A large number of applications of functional data analysis use the approximation (3.1.2) with $\hat{\lambda}_j$ and $\hat{v}_j$ being the eigenelements of the sample covariance operator defined by $\widehat{C}(x) = N^{-1}\sum_{i=1}^{N}\langle X_i, x\rangle X_i$. Large sample justifications of the applications of expansion (3.1.2) rely on the following bounds:

$$\limsup_{N\to\infty} NE\left\|\hat{c}_j\hat{v}_j - v_j\right\|^2 < \infty, \quad \limsup_{N\to\infty} NE|\hat{\lambda}_j - \lambda_j|^2 < \infty, \tag{3.2.5}$$

where $\hat{c}_j = \text{sign}(\langle \hat{v}_j, v_j\rangle)$. To lighten the notation, in the following we assume that $\hat{c}_j = 1$. For iid X_i with $E\left\|X_i\right\|^4 < \infty$, relations (3.2.5) were established by [7]. [8] showed that they continue to hold for weakly dependent stationary sequences. They used a specific, but very inclusive, concept of weak dependence, called L^4–m–approximability, which has been used in many other contexts, e.g. [15, 45, 51, 52]. Related concepts of dependence were used, e.g. by [53] and [54]. We refer to Chapter 16 of [6] for the definition of L^4–m–approximability. In our work, only conditions (3.2.5) are relevant, and they can be established under different quantifications of weak dependence, and for different data–driven basis systems, e.g. those based on the long–run covariance functions, see [44]. We therefore impose the following general assumption.

ASSUMPTION 3.2.1. *Assume that* $E\left\|X\right\|^4 < \infty$. *Let* v_j *be any deterministic vectors and set* $\lambda_j = \text{Var}(\langle X, v_j\rangle)$. *Assume that relations* (3.2.5) *hold for* $1 \le j \le p$, *where the* $\hat{v}_j$ *and* $\hat{\lambda}_j$ *are estimators of* v_j *and* λ_j, *and* $\hat{c}_j = \text{sign}(\langle \hat{v}_j, v_j\rangle)$.

Set

$$Y_i(j) = \langle X_i, v_j\rangle, \quad Y_{i,N}(j) = \langle X_i, \hat{v}_j\rangle,$$

and consider the maxima

$$M_N(j) = \max(Y_1(j), Y_2(j), \ldots, Y_N(j));$$
$$\widehat{M}_N(j) = \max(Y_{1,N}(j), Y_{2,N}(j), \ldots, Y_{N,N}(j)).$$

In this section, we want to specify sufficient conditions under which the asymptotic distribution, as $N \to \infty$, of the random vector

$$\widehat{\mathbf{M}}_N = \left[\widehat{M}_N(1), \widehat{M}_N(2), \ldots, \widehat{M}_N(p)\right]^T$$

is the same as the asymptotic distribution of the vector

$$\mathbf{M}_N = [M_N(1), M_N(2), \ldots, M_N(p)]^T .$$

From an applied perspective, the above equivalence means that the effect of the estimation of the v_j is negligible for the purpose of the study of multivariate extrema of projections. The assumptions we formulate in the following hold if the functions X_i are iid Gaussian. In that case, the asymptotic distribution of $\mathbf{M}_N$ is well-known:

$$P\left(\frac{M_N(1) - \sqrt{\lambda_1} b_N^{(G)}}{\sqrt{\lambda_1} a_N^{(G)}} \le x_1, \ldots, \frac{M_N(p) - \sqrt{\lambda_p} b_N^{(G)}}{\sqrt{\lambda_p} a_N^{(G)}} \le x_p\right) \to \exp\left\{-\sum_{j=1}^{p} e^{-x_j}\right\},$$

with

$$b_N^{(G)} = (2\log N - \log\log N - \log(4\pi))^{\frac{1}{2}}; \quad a_N^{(G)} = 1/b_N^{(G)}. \tag{3.2.6}$$

It will follow from our general results that in the iid Gaussian case,

$$P\left(\frac{\widehat{M}_N(1) - \sqrt{\hat{\lambda}_1} b_N^{(G)}}{\sqrt{\hat{\lambda}_1} a_N^{(G)}} \le x_1, \ldots, \frac{\widehat{M}_N(p) - \sqrt{\hat{\lambda}_p} b_N^{(G)}}{\sqrt{\hat{\lambda}_p} a_N^{(G)}} \le x_p\right) \to \exp\left\{-\sum_{j=1}^{p} e^{-x_j}\right\}. \tag{3.2.7}$$

We want to see how far the assumptions of independence and Gaussianity can be relaxed.

Since the X_i form a stationary sequence, so do the projections $\langle X_i, v_j\rangle$. Assuming these projections are in the Gumbel domain of attraction, under condition $D(u_n)$, e.g. [11] p. 373, their maxima converge to a Gumbel distribution with an extremal index θ_j. Rather than assuming condition $D(u_n)$, which is only a sufficient condition, we make the following assumption.

ASSUMPTION 3.2.2. *For each* $1 \le j \le p$,

$$\lim_{N\to\infty} P\left(\frac{M_N(j) - b_{j,N}}{a_{j,N}} \le x\right) = \exp\left\{-\theta_j e^{-x}\right\}, \quad x \in \mathbb{R},$$

for an extremal index $\theta_j \in [0,1]$*, and* $a_{j,N}$*,* $b_{j,N}$ *defined in* (3.2.9) *below.*

Assumption 3.2.2 is a finite–projection analog of the definition of the Gumbel domain of attraction implied by the theory in Chapter 9 of [12], who consider iid functions. Theorems 4.3.3 and 4.5.2 in [55], see also Theorem 4.4.8 in [10], show that practically all stationary Gaussian sequences have extremal index 1, so if the functions X_i are Gaussian, Assumption 3.2.2, will practically always hold with $\theta_j = 1$. However, for linear processes with subexponential innovations (which are in the Gumbel domain of attraction) the extremal index is generally smaller than 1, see Section 5.5.2 of [10].

The normalizing constants in Assumption 3.2.2 are defined as follows. Let F_j be the cdf of each $Y_i(j)$, and U_j the left–continuous inverse of $1/(1-F_j)$. Define the function f_j by

$$f_j(t) := \frac{1 - F_j(t)}{F_j'(t)}. \tag{3.2.8}$$

Then

$$a_{j,N} = f_j(U_j(N)); \quad b_{i,N} = U_j(N). \tag{3.2.9}$$

To prove the advertised asymptotic equivalence, we need to restrict the Gumbel domain of attraction. We will assume that for precisely specified values of κ

$$\lim_{N\to\infty} \frac{N^{-\kappa}}{a_{j,N}} = 0. \tag{3.2.10}$$

REMARK 3.2.1. *We show in Appendix A that condition* (3.2.10) *holds, for any* $\kappa > 0$*, for normal, exponential, and any gamma distribution. We could not find an example of a distribution in the Gumbel domain of attraction for which it would fail. If the distribution has a density* F'*,* (3.2.10) *is* *equivalent to* $N^{1-\kappa}F'(b_N) \to 0$*, where* b_N *is the* $(1-N^{-1})$*th quantile. For the Gumbel domain of attraction,* F' *decays in some approximately exponential fashion, but* b_N *can grow at a logarithmic rate, so it is not clear if* (3.2.10) *always holds. Outside the Gumbel domain of attraction, many examples of distributions which satisfy* (3.2.10) *exist, e.g. Pareto or Cauchy.*

The sequences $\{Y_i(j), i \geq 1\}$ are in general, not independent. However, to recover the same limit as in the independent case, it is enough to require asymptotic independence of the extremes of projections, as stated in the next assumption.

ASSUMPTION 3.2.3. *The maxima* $M_N(1), M_N(2), \ldots M_N(p)$ *are asymptotically independent, in the sense that*

$$P\left(M_N(1) \leq d_{1,N}(x_1), \ldots, M_N(p) \leq d_{p,N}(x_p)\right) \\ - P\left(M_N(1) \leq d_{1,N}(x_1)\right) \ldots P\left(M_N(p) \leq d_{p,N}(x_p)\right) = o(1),$$

where

$$d_{j,N}(x) = a_{j,N}x + b_{j,N}.$$

We can now state the results of this Section. All limits are taken as $N \to \infty$.

THEOREM 3.2.1. *Suppose Assumption 3.2.1 and condition* (3.2.10) *with* $\kappa = 1/4$ *hold. Then, for any* $1 \leq j \leq p$*, and the* $a_{j,N}$ *and* $b_{j,N}$ *defined in* (3.2.9),

$$\frac{\widehat{M}_N(j) - b_{j,N}}{a_{j,N}} - \frac{M_N(j) - b_{j,N}}{a_{j,N}} = o_P(1). \tag{3.2.11}$$

If, in addition, Assumptions 3.2.2 and 3.2.3 hold, then

$$P\left(\frac{\widehat{M}_N(1) - b_{1,N}}{a_{1,N}} \leq x_1, \ldots, \frac{\widehat{M}_N(p) - b_{p,N}}{a_{p,N}} \leq x_p\right) \to \exp\left\{-\sum_{j=1}^{p}\theta_j e^{-x_j}\right\}. \tag{3.2.12}$$

The next theorem focuses on the Gaussian case. The normalizing constants are then known explicitly up to a scale parameter, which is estimated. The conditions $E\|X\|^4$ holds automatically, so in the Gaussian case the only restrictions in Assumption 3.2.1 are conditions (3.2.5).

THEOREM 3.2.2. *Suppose the X_i form a stationary Gaussian sequence that satisfies condition $D(u_n)$ and Assumptions 3.2.1 and 3.2.3 hold. Then, for any $1 \le j \le p$, and the $a_N^{(G)}$ and $b_N^{(G)}$ defined in* (3.2.6)*,*

$$\frac{\widehat{M}_N(j) - \sqrt{\hat{\lambda}_j} b_N^{(G)}}{\sqrt{\hat{\lambda}_j} a_N^{(G)}} - \frac{M_N(j) - \sqrt{\lambda_j} b_N^{(G)}}{\sqrt{\lambda_j} a_N^{(G)}} = o_P(1), \tag{3.2.13}$$

and

$$P\left(\frac{\widehat{M}_N(1) - \sqrt{\hat{\lambda}_1} b_N^{(G)}}{\sqrt{\hat{\lambda}_1} a_N^{(G)}} \le x_1, \ldots, \frac{\widehat{M}_N(p) - \sqrt{\hat{\lambda}_p} b_N^{(G)}}{\sqrt{\hat{\lambda}_p} a_N^{(G)}} \le x_p \right) \to \exp\left\{ -\sum_{j=1}^{p} \theta_j e^{-x_j} \right\}. \tag{3.2.14}$$

COROLLARY 3.2.1. *Suppose the X_i are iid Gaussian and Assumption 3.2.1 holds. Then relation* (3.2.7) *holds.*

Theorems 3.2.1 and 3.2.2 are applicable to the special, but the most important, case when the v_j and λ_j are, respectively, the functional principal components and their eigenvalues, as we illustrate in the following example. The example also clarifies the structure of our assumptions.

EXAMPLE 3.2.1. *Suppose X is a random element of the Hilbert space $L^2 = L^2([0,1])$ with mean zero and*

$$E\|X\|^2 = E\int_0^1 X^2(t)dt < \infty.$$

Define the covariance function of X by $c(t,s) = E[X(t)X(s)]$. Define its eigenvalues λ_j and the eigenfunction v_j by

$$\int_0^1 c(t,s)v_j(s)ds = \lambda_j v_j(t), \quad j = 1, 2, \ldots.$$

The random function X admits the L^2–convergent expansion $X(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j v_j(t)$, with uncorrelated, mean zero and unit variance Z_j. All these facts are proven in Chapter 11 of [49] (and several other books). The functions v_j are called the functional principal components (FPCs).

Consider now the sample covariance function $\hat{c}(t,s) = N^{-1} \sum_{i=1}^{N} X_i(t) X_i(s)$, where the X_i are copies of X. If the $X_i, 1 \le i \le N$, are a realization of a strictly stationary weakly dependent time series, then $\hat{c}(t,s)$ is a consistent estimator of $c(t,s)$, and its eigenelements, $\hat{\lambda}_j, \hat{v}_j$, satisfy relations (3.2.5). *This has been proven in [8] assuming a weak dependence condition called L^4–m–approximability. We give a specific example rather than explain this condition.*

Suppose ε_i are independent copies of ε, which is a mean zero random element in L^2 satisfying $E\|\varepsilon\|^4 < \infty$. (If ε is Gaussian, then $E\|\varepsilon\|^p < \infty$, for every $p > 0$, see e.g. Corollary 1 on p. 338 of [56].) Suppose $\psi_k(\cdot,\cdot)$ are Hilbert–Schmidt kernels, and define the linear process

$$X_i(t) = \sum_{k=1}^{\infty} \int_0^1 \psi_k(t,s) \varepsilon_{i-k}(s) ds. \tag{3.2.15}$$

Proposition 16.1 and Theorem 16.2 of [6] imply that if

$$\sum_{m=1}^{\infty} \sum_{k=m}^{\infty} \left\{ \int_0^1 \int_0^1 \psi_k^2(t,s) dt ds \right\}^{1/2} < \infty,$$

then Assumption 3.2.1 holds with the $v_j, \lambda_j, \hat{v}_j, \hat{\lambda}_j$ defined above.

Projecting (3.2.15) *onto v_j, we obtain*

$$\langle X_i, v_j \rangle = \sum_{k=1}^{\infty} \int_0^1 \psi_{k,j}(s) \varepsilon_{i-k}(s) ds, \quad \psi_{k,j}(s) = \int_0^1 \psi_k(t,s) v_j(t) dt.$$

We see that, except the case when the kernels ψ_k do not depend s, the scalar sequence $\{\langle X_i, v_j \rangle, i \in \mathbb{Z}\}$ is not a linear process. With the current state of research, it is difficult to formulate assumptions on the error functions ε_i which would imply some specific extremal structure of this sequence. It is however a weekly dependent scalar sequence, so assuming that it satisfies condition $D(u_n)$ is not restrictive. Following the paradigm of Chapter 9 of [12], we could assume *that for some functions a_N and b_N,*

$$\max_{1\le i\le N} \frac{\varepsilon_i(s) - b_N(s)}{a_N(s)} \xrightarrow{d} G(s),$$

where G is a random function with marginal Gumbel distributions. But it is not clear, at present, how to infer that $\langle X_i, v_j\rangle$ would be in the Gumbel domain of attraction. For this reason, we make the direct Assumption 3.2.2.

If the ε_i are Gaussian random functions, then $\{\langle X_i, v_j\rangle, i \in \mathbb{Z}\}$ is a strictly stationary Gaussian sequence, and for practically all such sequences, Assumption 3.2.2 holds with $\theta_j = 1$, see the discussion following this Assumption.

3.3 Rates of convergence

Set

$$\begin{aligned} K_{j,N}(x) &= P\left(\frac{\widehat{M}_N(j) - b_{j,N}}{a_{j,N}} \le x\right) - P\left(\frac{M_N(j) - b_{j,N}}{a_{j,N}} \le x\right) \\ &= P\left(\widehat{M}_N(j) \le d_{j,N}(x)\right) - P\left(M_N(j) \le d_{j,N}(x)\right). \end{aligned}$$

It follows from the results of Section 3.2, that under weak assumptions, for each j and each x, $K_{j,N}(x) \to 0$, as $N \to \infty$. More precise information can be obtained by establishing a rate at which $K_{j,N}(x)$ tends to zero.

[18] and [57] showed that (subject to technical conditions) for a (scalar) stationary Gaussian linear time series,

$$P(M_N \le u_N) - F^N(u_N) = O(N^{-r}), \tag{3.3.16}$$

for some $r > 0$, and any sequence u_N which satisfies

$$N(1 - F(u_N)) \to C > 0, \quad N \to \infty.$$

The cdf F is the marginal distribution of each scalar observation, say Y_i, and $M_N = \max_{1 \le i \le N} Y_i$. Condition (3.3.16) quantifies the strength of dependence in a stationary time series in a manner that is relevant to the study of extremes. The dependence must be sufficiently weak, so that the maximum of the first N observations has asymptotically the same distribution as the maximum of their independent copies. The strength of the dependence is quantified by the exponent r. Rather than imposing specific assumptions on the structure of the projections of the functional time series, we assume in this section that condition (3.3.16) holds with some $r > 0$. The rationale for such an approach is similar as in Section 3.2; condition (3.3.16) may hold for different and more general classes of the process than those for which it has already been established. For these reasons, we impose the following assumption.

ASSUMPTION 3.3.1. *Denote by F_j the marginal cdf of each $Y_i(j) = \langle X_i, v_j \rangle$. Let $u_{j,N}$ be any sequence such that $\lim_{N\to\infty} N(1 - F_j(u_{j,N})) = C_j > 0$. Then, for $1 \le j \le p$, assume that*

$$P(M_N(j) \le u_{j,N}) - F_j^N(u_{j,N}) = O(N^{-r_j}), \tag{3.3.17}$$

for some $r_j > 0$.

Assumption 3.3.1 requires that $\theta_j = 1$ in Assumption 3.2.2. It could be satisfied by non–Gaussian processes. To lighten the notation, in the following we drop the subscript j, as the arguments apply to each j.

THEOREM 3.3.1. *Suppose the X_i form a stationary Gaussian sequence and Assumptions 3.2.1 and 3.3.1 hold. Then, for any real x, $K_N(x) = O(N^{-q})$, for any $0 < q \le \min(r, 1/8)$.*

If the assumption of Gaussianity is dropped, the membership in the Gumbel domain of attraction must be specifically assumed. In the context of projections of a dependent functional sequence, this is specified by Assumption 3.2.2. However, to establish the rate, we need to restrict the form of the univariate marginal distribution. We require the von Mises' condition and condition (3.2.10). While these conditions theoretically restrict the Gumbel domain of attraction, all distributions used in applications meet them.

DEFINITION 3.3.1. *The distribution function F with its right endpoint $x^\star = \sup\{x : F(x) < 1\}$ is said to satisfy von Mises' condition, if $F''(x)$ exists, $F'(x)$ is positive for all x in some left neighborhood of $x^\star$, and*

$$\lim_{t \uparrow x^\star} \left(\frac{1-F}{F'} \right)' (t) = \gamma,$$

where $\gamma \in \mathbb{R}$ is some constant.

Definition 3.3.1 is a sufficient condition for F to be in the domain of attraction of G_γ, see Theorem 1.1.8 in [12].

ASSUMPTION 3.3.2. *The distribution function F satisfies von Mises' condition with $\gamma = 0$.*

THEOREM 3.3.2. *Suppose Assumptions 3.2.1, 3.3.1, 3.3.2 hold. If condition* (3.2.10) *holds with some $0 < \kappa < 1/4$, then, for any real x,*

$$K_N(x) = O(N^{-q}),$$

for any $0 < q \leq \min(r, 1/8 - \kappa/2)$.

3.4 Proofs of the results of Section 3.2

The proof of Theorem 3.2.1 requires several lemmas.

LEMMA 3.4.1. *Suppose X_i are functions in L^2 with the same distribution. If $E\|X_i\|^4 < \infty$, then*

$$E \max_{1 \leq i \leq N} ||X_i||^2 = O(N^{1/2}).$$

Proof. Set $\xi_i = ||X_i||^2$, $1 \leq i \leq N$. Then $\xi_1, \xi_2, \ldots, \xi_N$ are nonnegative random variables with the same distribution. Observe that

$$\left\{ E \max_{1 \leq i \leq N} \xi_i \right\}^2 \leq E \left[\max_{1 \leq i \leq N} \xi_i^2 \right] \leq \sum_{i=1}^N E\xi_i^2 = O(N).$$

□

Set

$$\eta_{i,N}(j) = Y_{i,N}(j) - Y_i(j), \quad 1 \le i \le N,\ 1 \le j \le p. \tag{3.4.18}$$

LEMMA 3.4.2. *Under Assumption 3.2.1, for the $\eta_{i,N}(j)$ defined in* (3.4.18)*, and for any $1 \le j \le p$,*

$$E\Big(\max_{1\le i\le N} |\eta_{i,N}(j)|\Big) = O(N^{-1/4}).$$

Proof. Observe that for each i and j,

$$|\eta_{i,N}(j)| = |\,\langle X_i, \hat{v}_j - v_j\rangle\,| \le ||\hat{v}_j - v_j|| \cdot ||X_i||,$$

so

$$\max_{1\le i\le N} |\eta_{i,N}(j)| \le ||\hat{v}_j - v_j|| \cdot \max_{1\le i\le N} ||X_i||.$$

By the Cauchy-Schwarz inequality, for any $1 \le j \le p$,

$$E\Big(\max_{1\le i\le N} |\eta_{i,N}(j)|\Big) \le \Big(E||\hat{v}_j - v_j||^2\Big)^{\frac{1}{2}} \cdot \Big(E \max_{1\le i\le N} ||X_i||^2\Big)^{\frac{1}{2}}.$$

Under Assumptions 3.2.1, $\limsup_{N\to\infty} NE||\hat{v}_j - v_j||^2 < \infty$, so $E||\hat{v}_j - v_j||^2 = O(N^{-1})$. By Lemma 3.4.1, we then have

$$E\Big(\max_{1\le i\le N} |\eta_{i,N}(j)|\Big) \le \Big(O(N^{-1})\Big)^{1/2} \cdot \Big(O(N^{1/2})\Big)^{1/2} = O(N^{-1/4}).$$

□

LEMMA 3.4.3. *Under Assumption 3.2.1, for any $1 \le j \le p$,*

$$E\Big|\widehat{M}_N(j) - M_N(j)\Big| = O(N^{-1/4}).$$

Proof. Observe that

$$\begin{aligned}
\widehat{M}_N(j) - M_N(j) &= \max_{1\le i\le N} Y_{i,N}(j) - \max_{1\le i\le N} Y_i(j) \\
&= \max_{1\le i\le N} \Big(Y_i(j) + \eta_{i,N}(j)\Big) - \max_{1\le i\le N} Y_i(j) \\
&\le \max_{1\le i\le N} Y_i(j) + \max_{1\le i\le N} \eta_{i,N}(j) - \max_{1\le i\le N} Y_i(j) \\
&= \max_{1\le i\le N} \eta_{i,N}(j) \\
&\le \max_{1\le i\le N} |\eta_{i,N}(j)|.
\end{aligned}$$

A similar argument yields $M_N(j) - \widehat{M}_N(j) \le \max_{1\le i\le N} |\eta_{i,N}(j)|$. Therefore, the claim follows from Lemma 3.4.2. □

PROOF OF THEOREM 3.2.1 By Markov's inequality, for any $1 \le j \le p$, and any $\epsilon > 0$,

$$\begin{aligned}
&P\Big(\Big|\frac{\widehat{M}_N(j) - b_{j,N}}{a_{j,N}} - \frac{M_N(j) - b_{j,N}}{a_{j,N}}\Big| > \epsilon\Big) \\
&= P\Big(\Big|\widehat{M}_N(j) - M_N(j)\Big| > a_{j,N}\epsilon\Big) \\
&\le \frac{E\Big|\widehat{M}_N(j) - M_N(j)\Big|}{a_{j,N}\epsilon}.
\end{aligned}$$

Using Lemma 3.4.3, we have

$$\frac{E\Big|\widehat{M}_N(j) - M_N(j)\Big|}{a_{j,N}} = \frac{O(N^{-1/4})}{a_{j,N}} \to 0, \tag{3.4.19}$$

when condition (3.2.10) holds with $\kappa = 1/4$. Therefore, relation (3.2.11) follows.

Suppose Assumptions 3.2.2 and 3.2.3 hold. Then, as $N \to \infty$,

$$\begin{aligned}
&P\left(M_N(1) \le d_{1,N}(x_1), \ldots, M_N(p) \le d_{p,N}(x_p)\right) \\
&= \prod_{j=1}^{p} P(M_N(j) \le d_{j,N}(x_j)) + o(1) \\
&\to \exp\left\{-\sum_{j=1}^{p} \theta_j e^{-x_j}\right\}.
\end{aligned}$$

Equivalently,

$$\left(\frac{M_N(1) - b_{1,N}}{a_{1,N}}, \ldots, \frac{M_N(p) - b_{p,N}}{a_{p,N}}\right) \xrightarrow{d} (M_1, \ldots, M_p), \quad N \to \infty,$$

where $\mathbf{M} = (M_1, \ldots, M_p)$ has the cdf $F_{\mathbf{M}}(x_1, \ldots, x_p) = \exp\left\{-\sum_{j=1}^{p} \theta_j e^{-x_j}\right\}$. By the Cramér–Wold device, we therefore have

$$\sum_{j=1}^{p} c_j \frac{M_N(j) - b_{j,N}}{a_{j,N}} \xrightarrow{d} \sum_{j=1}^{p} c_j M_j, \quad N \to \infty, \tag{3.4.20}$$

for any constants c_j. Observe that

$$\begin{aligned}
\sum_{j=1}^{p} c_j \frac{\widehat{M}_N(j) - b_{j,N}}{a_{j,N}} &= \sum_{j=1}^{p} c_j \left(\frac{\widehat{M}_N(j) - b_{j,N}}{a_{j,N}} - \frac{M_N(j) - b_{j,N}}{a_{j,N}}\right) \\
&\quad + \sum_{j=1}^{p} c_j \frac{M_N(j) - b_{j,N}}{a_{j,N}}.
\end{aligned}$$

The first term is asymptotically negligible because of (3.2.11), so claim (3.2.12) follows from (3.4.20).

PROOF OF THEOREM 3.2.2 If the X_i form a stationary Gaussian sequence, then condition (3.2.10) holds for any $\kappa > 0$. Since condition $D(u_n)$ implies Assumption 3.2.2, relations (3.2.11) and (3.2.12) of Theorem 3.2.1 hold with

$$b_{j,N} = \sqrt{\lambda_j} b_N^{(G)}, \quad a_{j,N} = \sqrt{\lambda_j} a_N^{(G)},$$

where $a_N^{(G)}$ and $b_N^{(G)}$ are defined in (3.2.6). Therefore, relation (3.2.13) is equivalent to

$$\frac{\widehat{M}_N(j)}{a_N^{(G)}}\left(\frac{1}{\sqrt{\hat{\lambda}_j}}-\frac{1}{\sqrt{\lambda_j}}\right)=o_P(1). \tag{3.4.21}$$

By Theorem 3.2.1, we know that, for any $1 \le j \le p$, $(\widehat{M}_N(j)-\sqrt{\lambda_j}b_N^{(G)})/(\sqrt{\lambda_j}a_N^{(G)})=O(1)$. Since $b_N^{(G)} \sim (2\log N)^{\frac{1}{2}}$ and $a_N^{(G)} \sim (2\log N)^{-\frac{1}{2}}$, we have

$$\frac{\widehat{M}_N(j)}{a_N^{(G)}}=\left(O(1)+\frac{b_N^{(G)}}{a_N^{(G)}}\right)\sqrt{\lambda_j}=O(1)+O(\log N)=O(\log N).$$

Therefore, (3.4.21) will follow from

$$\log N\left(\frac{1}{\sqrt{\hat{\lambda}_j}}-\frac{1}{\sqrt{\lambda_j}}\right)=o_P(1). \tag{3.4.22}$$

By Assumptions 3.2.1 and Chebyshev's inequality, for any $0<\gamma<1$,

$$N^{\frac{1-\gamma}{2}}(\hat{\lambda}_j-\lambda_j)\xrightarrow{P}0, \quad N\to\infty. \tag{3.4.23}$$

Hence,

$$\log N\left(\frac{1}{\sqrt{\hat{\lambda}_j}}-\frac{1}{\sqrt{\lambda_j}}\right)=\frac{\log N}{N^{\frac{1-\gamma}{2}}}\frac{N^{\frac{1-\gamma}{2}}\left(\lambda_j-\hat{\lambda}_j\right)}{\sqrt{\hat{\lambda}_j}\sqrt{\lambda_j}\left(\sqrt{\hat{\lambda}_j}+\sqrt{\lambda_j}\right)}\xrightarrow{P}0, \quad N\to\infty,$$

establishing (3.4.22). Thus (3.2.13) follows.

Observe that

$$\sum_{j=1}^{p} c_j \frac{\widehat{M}_N(j) - \sqrt{\hat{\lambda}_j} b_N^{(G)}}{\sqrt{\hat{\lambda}_j} a_N^{(G)}} = \sum_{j=1}^{p} c_j \left(\frac{\widehat{M}_N(j) - \sqrt{\hat{\lambda}_j} b_N^{(G)}}{\sqrt{\hat{\lambda}_j} a_N^{(G)}} - \frac{\widehat{M}_N(j) - \sqrt{\lambda_j} b_N^{(G)}}{\sqrt{\lambda_j} a_N^{(G)}} \right) + \sum_{j=1}^{p} c_j \frac{\widehat{M}_N(j) - \sqrt{\lambda_j} b_N^{(G)}}{\sqrt{\lambda_j} a_N^{(G)}},$$

for any constants c_j. The first term is asymptotically negligible because of (3.2.13). Claim (3.2.14) thus follows from Theorem 3.2.1.

PROOF OF COROLLARY 3.2.1 For any $1 \le j \le p$, Assumption 3.2.2 holds with $\theta_j = 1$, and

$$b_{j,N} = \sqrt{\lambda_j} b_N^{(G)}, \quad a_{j,N} = \sqrt{\lambda_j} a_N^{(G)},$$

where $a_N^{(G)}$ and $b_N^{(G)}$ are defined in (3.2.6). Considering independent Gaussian sequence X_i, for any i, k, $\mathrm{Cov}(Y_i(j), Y_k(\ell)) = 0$, if $j \neq \ell$, so sequences $\{Y_i(j), i \ge 1\}$, $j = 1, 2, \ldots, p$, are independent. It follows that for each N, the random variables $M_N(1), M_N(2), \ldots, M_N(p)$ are independent. Therefore, Assumption 3.2.3 follows.

3.5 Proofs of the results of Section 3.3

If the X_i form a stationary Gaussian sequence, then

$$b_{j,N} = \sqrt{\lambda_j} b_N^{(G)}, \quad a_{j,N} = \sqrt{\lambda_j} a_N^{(G)},$$

where $a_N^{(G)}$ and $b_N^{(G)}$ are defined in (3.2.6). Therefore,

$$d_N(x) = \sqrt{\lambda} a_N^{(G)} x + \sqrt{\lambda} b_N^{(G)},$$

after dropping the subscript j. The proof of Theorem 3.3.1 requires several lemmas. The extremal behavior of iid sequence is described by the following well–known result:

THEOREM 3.5.1. *If F is the cdf of the $N(0,\lambda)$ distribution, then*

$$\lim_{N\to\infty} F^N(d_N(x)) = \exp\left\{-e^{-x}\right\}. \tag{3.5.24}$$

LEMMA 3.5.1. *Suppose F is the cdf of the $N(0,\lambda)$ distribution. For any $s > 0$, any real c, and any fixed x,*

$$F\left(d_N(x) + cN^{-s}\right) - F\left(d_N(x)\right) = O\left(N^{-1-s}\right).$$

Proof. Notice that $b_N^{(G)} \sim (2\log N)^{\frac{1}{2}}$ and $a_N^{(G)} \sim (2\log N)^{-\frac{1}{2}}$. Therefore $d_N(x) \sim \sqrt{\lambda} b_N^{(G)}$, so $d_N(x) + cN^{-s} > 0$, for sufficiently large N. Next observe that, for $c \geq 0$,

$$\begin{aligned}
&\int\limits_{d_N(x)}^{d_N(x)+cN^{-s}} \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{u^2}{2\lambda}\right\} du \\
&\leq cN^{-s} \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{1}{2\lambda} d_N^2(x)\right\} \\
&= O(N^{-s}) \exp\left\{-\frac{(a_N^{(G)}x + b_N^{(G)})^2}{2}\right\} \\
&= O(N^{-s}) O\left(N^{-1}\right) = O\left(N^{-1-s}\right).
\end{aligned}$$

For $c < 0$, since $N^{-s} d_N(x) \to 0$, we have

$$\begin{aligned}
&\int\limits_{d_N(x)+cN^{-s}}^{d_N(x)} \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{u^2}{2\lambda}\right\} du \\
&\leq cN^{-s} \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{1}{2\lambda}\left(d_N(x) + cN^{-s}\right)^2\right\} \\
&\leq O(N^{-s}) \exp\left\{-\frac{1}{2\lambda} d_N^2(x)\right\} \exp\left\{-\frac{1}{\lambda} cN^{-s} d_N(x)\right\} \\
&= O(N^{-s}) O\left(N^{-1}\right) = O\left(N^{-1-s}\right).
\end{aligned}$$

□

The following lemma is well-known if the a_N and b_N are normalizing constants for the extreme limit to whose domain of attraction F belongs, e.g. Theorem 1.1.6 of [12]. It, however, holds in greater generality, which we will need in the proofs that follow.

LEMMA 3.5.2. *Let F be any distribution function. For any real sequences $a_N > 0$ and b_N, the following statements are equivalent for any fixed $x \in \mathbb{R}$:*
(i) $\lim_{N\to\infty} F^N(a_N x + b_N) = \exp\{-e^{-x}\}$*;*
(ii) $\lim_{N\to\infty} N(1 - F(a_N x + b_N)) = e^{-x}$.

Proof. Taking logarithms of both sides of (i), we get an equivalent relation

$$\lim_{N\to\infty} N \log F(a_N x + b_N) = -e^{-x}. \tag{3.5.25}$$

Set $w_N = 1 - F(a_N x + b_N)$, so that $F(a_N x + b_N) = 1 - w_N$. Observe that either (3.5.25) or condition (ii) imply that $w_N \to 0$. By Taylor series expansion of the logarithm around 1,

$$\log F(a_N x + b_N) = -w_N - \frac{1}{2x_N^2} w_N^2, \quad |x_N - 1| \le w_N.$$

Since $1/x_N^2$ is bounded in a neighborhood of 1,

$$\lim_{N\to\infty} \frac{-N \log F(a_N x + b_N)}{N\Big(1 - F(a_N x + b_N)\Big)} = \lim_{N\to\infty} \left\{1 + \frac{1}{2x_N^2} w_N\right\} = 1,$$

completing the proof. □

LEMMA 3.5.3. *Suppose F is the cdf of the $N(0,\lambda)$ distribution. For any $s > 0$, any real c, and any fixed x,*

$$F^N(d_N(x) + cN^{-s}) \to \exp\left\{-e^{-x}\right\}, \quad N \to \infty. \tag{3.5.26}$$

Proof. Set $b'_N = b_N^{(G)} + cN^{-s}/\sqrt{\lambda}$, so that

$$F^N(d_N(x) + cN^{-s}) = F^N(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b'_N).$$

Hence relation (3.5.26) is equivalent to

$$\lim_{N\to\infty} F^N(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b'_N) = \exp\left\{-e^{-x}\right\}.$$

Then by Lemma 3.5.2, it is enough to prove the following relation

$$\lim_{N\to\infty} N(1 - F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b'_N)) = e^{-x}. \tag{3.5.27}$$

By Theorem 3.5.1 and Lemma 3.5.2, we have $\lim_{N\to\infty} N(1 - F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b_N^{(G)})) = e^{-x}$. Therefore, by Lemma 3.5.1,

$$\begin{aligned}
&N(1 - F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b'_N)) \\
&\quad = N\left[1 - F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b_N^{(G)})\right] + N\left[F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b_N^{(G)}) - F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b'_N)\right] \\
&\quad = N\left[1 - F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b_N^{(G)})\right] + N\left[F(d_N(x)) - F(d_N(x) + cN^{-s})\right] \\
&\quad = N\left[1 - F(\sqrt{\lambda}a_N^{(G)}x + \sqrt{\lambda}b_N^{(G)})\right] + O(N^{-s}) \\
&\quad \to e^{-x}.
\end{aligned}$$

□

LEMMA 3.5.4. *Suppose F is the cdf of the $N(0, \lambda)$ distribution. For any $s > 0$, any real c, and any fixed x,*

$$F^N(d_N(x) + cN^{-s}) - F^N(d_N(x)) = O(N^{-s}). \tag{3.5.28}$$

Proof. We will use the inequality

$$b^N - a^N \le (b-a) \cdot N b^{N-1}, \quad 0 \le a \le b. \tag{3.5.29}$$

By Lemma 3.5.1, we thus have, for $c \ge 0$,

$$\begin{aligned} & F^N(d_N(x) + cN^{-s}) - F^N(d_N(x)) \\ & \le \left\{F(d_N(x) + cN^{-s}) - F(d_N(x))\right\} \cdot N F^{N-1}(d_N(x) + cN^{-s}) \\ & = O\left(N^{-1-s}\right) N L_{1,N}^{N-1} = O\left(N^{-s}\right) L_{1,N}^{-1} L_{1,N}^{N}, \end{aligned}$$

where $L_{1,N} = F(d_N(x) + cN^{-s}) \to 1$. Moreover, by Lemma 3.5.3, $L_{1,N}^N \to \exp\{-e^{-x}\}$. Thus (3.5.28) follows.

Similarly, for $c < 0$,

$$\begin{aligned} & F^N(d_N(x)) - F^N(d_N(x) + cN^{-s}) \\ & \le \left\{F^N(d_N(x)) - F^N(d_N(x) + cN^{-s})\right\} \cdot N F^{N-1}(d_N(x)) \\ & = O\left(N^{-1-s}\right) N L_{2,N}^{N-1} = O\left(N^{-s}\right) L_{2,N}^{-1} L_{2,N}^{N}, \end{aligned}$$

where $L_{2,N} = F(d_N(x)) \to 1$, and by Theorem 3.5.1, $L_{2,N}^N \to \exp\{-e^{-x}\}$. Thus (3.5.28) follows. □

LEMMA 3.5.5. *Recall that $Y_i = Y_i(j) = \langle X_i, v_j \rangle$, for some fixed j. Suppose Assumption 3.3.1 holds. Then, for any $s > 0$, any real c, and any fixed x,*

$$P(Y_i \le d_N(x) + cN^{-s}, 1 \le i \le N) - P(Y_i \le d_N(x), 1 \le i \le N) = O(N^{-\min(r,s)}).$$

Proof. Denote by $\widetilde{Y}_i$ independent random variables with the same marginal distribution as the Y_i. By the triangle inequality,

$$
\begin{aligned}
&\left|P(Y_i \le d_N(x) + cN^{-s}, 1 \le i \le N) - P(Y_i \le d_N(x), 1 \le i \le N)\right| \\
&\quad \le \left|P(Y_i \le d_N(x) + cN^{-s}, 1 \le i \le N) - P(\widetilde{Y}_i \le d_N(x) + cN^{-s}, 1 \le i \le N)\right| \\
&\qquad + \left|P(\widetilde{Y}_i \le d_N(x) + cN^{-s}, 1 \le i \le N) - P(\widetilde{Y}_i \le d_N(x), 1 \le i \le N)\right| \\
&\qquad + \left|P(\widetilde{Y}_i \le d_N(x), 1 \le i \le N) - P(Y_i \le d_N(x), 1 \le i \le N)\right| \\
&\quad = T_1 + T_2 + T_3.
\end{aligned}
$$

Lemmas 3.5.3 and Theorem 3.5.1 imply that, $F^N(d_N(x)+cN^{-s}) \to \exp\{-e^{-x}\}$ and $F^N(d_N(x)) \to \exp\{-e^{-x}\}$. By Lemma 3.5.2, we have, for any fixed x,

$$
\begin{aligned}
&\lim_{N\to\infty} N(1 - F(d_N(x) + cN^{-s})) = e^{-x} > 0; \\
&\lim_{N\to\infty} N(1 - F(d_N(x))) = e^{-x} > 0.
\end{aligned}
$$

Then, by Assumption 3.3.1,

$$
P(M_N \le d_N(x) + cN^{-s}) - F^N(d_N(x) + cN^{-s}) = O(N^{-r})
$$

and

$$
P(M_N \le d_N(x)) - F^N(d_N(x)) = O(N^{-r}),
$$

for some $r > 0$. Therefore, $T_1 = O(N^{-r})$ and $T_3 = O(N^{-r})$, so the conclusion follows from Lemma 3.5.4. □

PROOF OF THEOREM 3.3.1 Recall the definition of $\eta_{i,N}$ in (3.4.18) and define the events

$$
B_N = \{Y_i \le d_N(x),\ 1 \le i \le N\}; \quad B_N^\star = \{Y_i + \eta_{i,N} \le d_N(x),\ 1 \le i \le N\}.
$$

Therefore, $K_N(x) = P(B_N^\star) - P(B_N)$.

Define the event

$$A_N = \left\{ \max_{1\le i\le N} |\eta_{i,N}| \le cN^{-s} \right\},$$

for any $0 < s < 1/4$ and $c > 0$. By Lemma 3.4.2, there is a positive constant C_0 such that

$$E\Big(\max_{1\le i\le N} |\eta_{i,N}| \Big) \le C_0 N^{-1/4}.$$

Therefore, by Markov's inequality,

$$P(A_N^c) \le c^{-1} N^s C_0 N^{-1/4} = O(N^{-(1/4-s)}).$$

Using the inequality $P(A\cap B) \ge P(A) + P(B) - 1$, we have

$$\begin{aligned}
&P(B_N^\star) - P(B_N) \\
&\quad = P(B_N^\star \cap A_N)P(A_N) + P(B_N^\star \cap A_N^c)P(A_N^c) \\
&\qquad - P(B_N \cap A_N)P(A_N) - P(B_N \cap A_N^c)P(A_N^c) \\
&\quad \le P(B_N^\star \cap A_N)P(A_N) + P(A_N^c)^2 - P(B_N \cap A_N)P(A_N) \\
&\quad \le P(B_N^\star \cap A_N)P(A_N) + P(A_N^c)^2 - \left[P(B_N) + P(A_N) - 1\right] P(A_N) \\
&\quad = \left[P(B_N^\star \cap A_N) - P(B_N)\right] P(A_N) + P(A_N^c).
\end{aligned}$$

The event $B_N^\star \cap A_N$ is contained in the event $\{Y_i \le d_N(x) + cN^{-s}, 1 \le i \le N\}$. Thus, by Lemma 3.5.5,

$$\begin{aligned}
K_N(x) &\le P(Y_i \le d_N(x) + cN^{-s}, 1 \le i \le N) - P(Y_i \le d_N(x),\ 1 \le i \le N) + P(A_N^c) \\
&= O(N^{-\min(r,s)}) + O(N^{-(1/4-s)}),
\end{aligned}$$

for any $0 < s < 1/4$. Similarly,

$$
\begin{aligned}
&P(B_N) - P(B_N^\star) \\
&\quad = P(B_N \cap A_N)P(A_N) + P(B_N \cap A_N^c)P(A_N^c) \\
&\qquad - P(B_N^\star \cap A_N)P(A_N) - P(B_N^\star \cap A_N^c)P(A_N^c) \\
&\quad \leq P(B_N)P(A_N) + P(A_N^c)^2 - P(B_N^\star \cap A_N)P(A_N) \\
&\quad = \left[P(B_N) - P(B_N^\star \cap A_N)\right] P(A_N) + P(A_N^c)^2.
\end{aligned}
$$

The event $B_N^\star \cap A_N$ contains the event $\{Y_i \leq d_N(x) - cN^{-s}, 1 \leq i \leq N\} \cap A_N$, so

$$
\begin{aligned}
-K_N(x) &\leq \big[P(B_N) - P\big(\big\{Y_i \leq d_N(x) - cN^{-s}, 1 \leq i \leq N\big\} \cap A_N\big)\big]\, P(A_N) + P(A_N^c)^2 \\
&\leq \big[P(B_N) - P(Y_i \leq d_N(x) - cN^{-s}, 1 \leq i \leq N) - P(A_N) + 1\big]\, P(A_N) + P(A_N^c)^2 \\
&\leq P(Y_i \leq d_N(x), 1 \leq i \leq N) - P\big(Y_i \leq d_N(x) - cN^{-s}, 1 \leq i \leq N\big) + P(A_N^c) \\
&= O(N^{-\min(r,s)}) + O(N^{-(1/4-s)}),
\end{aligned}
$$

for any $0 < s < 1/4$. Thus

$$
|K_N(x)| \leq O(N^{-\min(r,s)}) + O(N^{-(1/4-s)}).
$$

When $r > 1/8$, $K_N(x) = O(N^{-q})$ for any $0 < q \leq 1/8$; when $0 < r \leq 1/8$, $K_N(x) = O(N^{-q})$ for any $0 < q \leq r$. The conclusion then follows.

Next we prove Theorem 3.3.2. A key element of the proof is the uniform convergence in Lemma 3.5.6, which extends a result of [58]. The following definition, used in the proof of Lemma 3.5.6, was introduced by [58].

DEFINITION 3.5.1. *For a positive real number $g > 0$, define the distribution functions*

$$
F(g, x) = \begin{cases} 0 & \textit{if } x < -g^{-1} \\ 1 - (1 + gx)^{-g^{-1}} & \textit{if } x \geq -g^{-1} \end{cases}
$$

and

$$F(-g,x) = \begin{cases} 0 & \textit{if } x < 0 \\ 1-(1-gx)^{g^{-1}} & \textit{if } 0 \le x < g^{-1} \\ 1 & \textit{if } x \ge g^{-1}. \end{cases}$$

LEMMA 3.5.6. *Suppose Assumption 3.3.2 holds. For any fixed real number A, and the a_N and b_N defined in* (3.2.9)*,*

$$\sup_{x \ge A} |N(1 - F(a_N x + b_N)) - e^{-x}| \to 0, \quad N \to \infty. \tag{3.5.30}$$

Proof. Lemma 3.5.6, has been derived by [58] for $A = 0$ in (3.5.30). We extend it to an arbitrary real A. The nontrivial extension is only for $A < 0$, and this will be assumed throughout the proof. The general case involves two new factors, denoted $L_N(A)$ and $R_N(A)$. We show that their effect is asymptotically negligible.

Assumption 3.3.2 implies that F is in the Gumbel domain of attraction, and for all $t \in (t_0, x^\star)$, $t_0 < x^\star$,

$$1 - F(t) = C \exp\left\{ - \int_{t_0}^{t} \frac{1}{f(s)} ds \right\}, \quad C > 0, \tag{3.5.31}$$

where f is defined in (3.2.8). Consider a non-increasing function $g(x) \ge 0$, s.t. $\lim_{x\to\infty} g(x) = 0$ and $g(x) \ge |f'(x)|$. Such a function g always exists since we may take $g(x) = \sup_{t \ge x} |f'(t)|$.

Consider the change of variables

$$a_N x + b_N = a_N(x - A) + (Aa_N + b_N) := a_N y + b'_N. \tag{3.5.32}$$

Relation (3.5.30) then becomes

$$\sup_{y \ge 0} |N(1 - F(a_N y + b'_N)) - e^{-(y+A)}| \to 0, \quad N \to \infty. \tag{3.5.33}$$

For $t \geq b'_N$,

$$f(t) - f(b'_N) = \int_{b'_N}^{t} f'(s)ds \leq \int_{b'_N}^{t} g(s)ds \leq g(b'_N)(t - b'_N)$$

and

$$f(t) - f(b'_N) = \int_{b'_N}^{t} f'(s)ds \geq -\int_{b'_N}^{t} g(s)ds \geq -g(b'_N)(t - b'_N).$$

It follows that

$$\frac{1}{f(b'_N) + g(b'_N)(t - b'_N)} \leq \frac{1}{f(t)} \leq \frac{1}{f(b'_N) - g(b'_N)(t - b'_N)}.$$

From now on, we assume that $A < 0$. This implies that $b'_N = Aa_N + b_N < b_N$. Recall that $b_N = U(N)$, so

$$N = \frac{1}{1 - F(b_N)} = \frac{1}{\bar{F}(b_N)}.$$

Using (3.5.31), we therefore have

$$-\log\left\{N\bar{F}(a_N y + b'_N)\right\} = -\log\left\{\frac{\bar{F}(a_N y + b'_N)}{\bar{F}(b'_N)}\frac{\bar{F}(b'_N)}{\bar{F}(b_N)}\right\} = \int_{b'_N}^{a_N y + b'_N} \frac{1}{f(s)}ds - \int_{b'_N}^{b_N} \frac{1}{f(s)}ds.$$

Observe that, using $a_N = f(b_N)$,

$$\frac{1}{g(b'_N)} \log\left\{1 + \frac{g(b'_N)f(b_N)}{f(b'_N)}y\right\} \leq \int_{b'_N}^{a_N y + b'_N} \frac{1}{f(s)}ds \leq -\frac{1}{g(b'_N)} \log\left\{1 - \frac{g(b'_N)f(b_N)}{f(b'_N)}y\right\}.$$

The right inequality holds for $0 \leq y \leq \frac{f(b'_N)}{g(b'_N)f(b_N)}$ and the left for $y \geq 0$. Similarly, we have

$$\frac{1}{g(b'_N)} \log\left\{1 + \frac{g(b'_N)}{f(b'_N)}(b_N - b'_N)\right\} \leq \int_{b'_N}^{b_N} \frac{1}{f(s)}ds \leq -\frac{1}{g(b'_N)} \log\left\{1 - \frac{g(b'_N)}{f(b'_N)}(b_N - b'_N)\right\}.$$

Therefore, exponentiating, we find that for $y \geq 0$,

$$
\begin{aligned}
N\bar{F}(a_N y + b'_N) &\geq \left\{1 + \frac{g(b'_N)}{f(b'_N)}(b_N - b'_N)\right\}^{1/g(b'_N)} \left\{1 - \frac{g(b'_N)f(b_N)}{f(b'_N)} y\right\}^{1/g(b'_N)} \\
&= L_N(A)\bar{F}\left(-g(b'_N), \frac{f(b_N)}{f(b'_N)} y\right)
\end{aligned}
$$

and

$$
\begin{aligned}
N\bar{F}(a_N y + b'_N) &\leq \left\{1 - \frac{g(b'_N)}{f(b'_N)}(b_N - b'_N)\right\}^{-1/g(b'_N)} \left\{1 + \frac{g(b'_N)f(b_N)}{f(b'_N)} y\right\}^{-1/g(b'_N)} \\
&= R_N(A)\bar{F}\left(g(b'_N), \frac{f(b_N)}{f(b'_N)} y\right),
\end{aligned}
$$

where the functions $F(\pm g, x)$ are defined in Definition 3.5.1, and the coefficients $L_N(A)$ and $R_N(A)$ are given by

$$
L_N(A) = \left\{1 + \frac{g(b'_N)}{f(b'_N)}(b_N - b'_N)\right\}^{1/g(b'_N)}
$$

and

$$
R_N(A) = \left\{1 - \frac{g(b'_N)}{f(b'_N)}(b_N - b'_N)\right\}^{-1/g(b'_N)}.
$$

Then we have

$$
\begin{aligned}
&\sup_{y\geq 0} \left|N\left(1 - F(a_N y + b'_N)\right) - e^{-(y+A)}\right| \\
&\quad \leq \sup_{y\geq 0} \left|L_N(A)\bar{F}\left(-g(b'_N), \frac{f(b_N)}{f(b'_N)} y\right) - e^{-(y+A)}\right| \\
&\qquad \bigvee \sup_{y\geq 0} \left|R_N(A)\bar{F}\left(g(b'_N), \frac{f(b_N)}{f(b'_N)} y\right) - e^{-(y+A)}\right|.
\end{aligned}
$$

We will only show that

$$
\sup_{y\geq 0} \left|L_N(A)\bar{F}\left(-g(b'_N), \frac{f(b_N)}{f(b'_N)} y\right) - e^{-(y+A)}\right| = o(1). \tag{3.5.34}
$$

A similar argument applies to the second expression, and (3.5.33) follows. We represent the factor $L_N(A)$ as follows:

$$L_N(A) = \left\{1 + \frac{g(b'_N)}{f(b'_N)}(-A)f(b_N)\right\}^{1/g(b'_N)} = \left\{[1 + h_N(-A)]^{1/h_N}\right\}^{\frac{f(b_N)}{f(b'_N)}} = Q_N(A)^{\frac{f(b_N)}{f(b'_N)}},$$

where

$$h_N = \frac{g(b'_N)f(b_N)}{f(b'_N)}$$

and

$$Q_N(A) = [1 + h_N(-A)]^{1/h_N}.$$

Then

$$\begin{aligned}
&\sup_{y\geq 0}\left|L_N(A)\bar{F}\left(-g(b'_N), \frac{f(b_N)}{f(b'_N)}y\right) - e^{-(y+A)}\right| \\
&\leq \sup_{y\geq 0}\left|Q_N(A)^{\frac{f(b_N)}{f(b'_N)}}\bar{F}\left(-g(b'_N), \frac{f(b_N)}{f(b'_N)}y\right) - Q_N(A)^{\frac{f(b_N)}{f(b'_N)}}e^{-\frac{f(b_N)}{f(b'_N)}y}\right| \\
&\quad + \sup_{y\geq 0}\left|\left[Q_N(A)e^{-y}\right]^{\frac{f(b_N)}{f(b'_N)}} - Q_N(A)e^{-y}\right| + \sup_{y\geq 0}\left|Q_N(A)e^{-y} - e^{-(y+A)}\right| \\
&= L_{1,N}(A) + L_{2,N}(A) + L_{3,N}(A).
\end{aligned}$$

By Taylor series expansion, there exists $b^\star_N \in (b'_N, b_N)$ such that

$$f(b_N) = f(b'_N) + f'(b^\star_N)(b_N - b'_N) = f(b'_N) - f'(b^\star_N)Af(b_N).$$

Under Assumption 3.3.2, we then have

$$\frac{f(b_N)}{f(b'_N)} = \frac{1}{1 + f'(b^\star_N)A} \to 1, \quad N \to \infty.$$

Therefore, $0 \leq \frac{f(b_N)}{f(b'_N)} \leq 2$, for N large enough. Recall that $h_N = \frac{g(b'_N)f(b_N)}{f(b'_N)}$ and $g(x) \to 0$, as $x \to \infty$, so $h_N \to 0$, as $N \to \infty$. Thus

$$Q_N(A) = [1 + h_N(-A)]^{1/h_N} \to e^{-A}, \quad N \to \infty.$$

Since $A < 0$, for N large enough, $1 < Q_N(A) < e^{-A} + 1$, and

$$Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} \leq Q_N(A)^2 = O(1).$$

Proposition 3.1 in [58] shows that for $0 < g < 1$,

$$\sup_{x \geq 0} \left| \bar{F}(\pm g, x) - e^{-x} \right| \leq (2 + g) e^{-2} g.$$

Thus we have

$$\begin{aligned} L_{1,N}(A) &= Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} \sup_{y \geq 0} \left| \bar{F}\left(-g(b'_N), \frac{f(b_N)}{f(b'_N)} y\right) - e^{-\frac{f(b_N)}{f(b'_N)} y} \right| \\ &\leq O(1)\,(2 + g(b'_N))\, e^{-2} g(b'_N) \\ &= O(1)\,(2 + o(1))\, e^{-2} o(1) \\ &= o(1). \end{aligned}$$

For the second term,

$$\begin{aligned} L_{2,N}(A) &\leq \sup_{y \geq 0} \left| \left[Q_N(A) e^{-y}\right]^{\frac{f(b_N)}{f(b'_N)}} - Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} e^{-y} \right| + \sup_{y \geq 0} \left| Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} e^{-y} - Q_N(A) e^{-y} \right| \\ &=: L_{21,N}(A) + L_{22,N}(A). \end{aligned}$$

Proposition 3.2 in [58] shows that for $0 < g < 1$,

$$\sup_{x \geq 0} \left| e^{-x(1-g)} - e^{-x} \right| \leq e^{-1} \left(g/(1-g)\right)$$

and

$$\sup_{x \geq 0} \left| e^{-x(1+g)} - e^{-x} \right| \leq \left(g/(1+g)\right)(1+g)^{-1/g}.$$

Therefore,

$$
\begin{aligned}
L_{21,N}(A) =& Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} \sup_{y\geq 0} \left| e^{-\frac{f(b_N)}{f(b'_N)}y} - e^{-y} \right| \\
\leq & Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} e^{-1} \left(\frac{f(b'_N)}{f(b_N)} - 1 \right) I \left(0 \leq \frac{f(b_N)}{f(b'_N)} \leq 1 \right) \\
& \bigvee Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} \left(1 - \frac{f(b'_N)}{f(b_N)} \right) \left(\frac{f(b_N)}{f(b'_N)} \right)^{\frac{f(b'_N)}{f(b'_N)-f(b_N)}} I \left(1 \leq \frac{f(b_N)}{f(b'_N)} \leq 2 \right) \\
\leq & O(1)e^{-1}o(1) \bigvee O(1)o(1)O(1) \\
=& o(1).
\end{aligned}
$$

By Taylor series expansion, there exists f_0 between $\frac{f(b_N)}{f(b'_N)}$ and 1 such that

$$
\begin{aligned}
L_{22,N}(A) =& \left| Q_N(A)^{\frac{f(b_N)}{f(b'_N)}} - Q_N(A) \right| \\
=& Q_N(A)^{f_0} \left| \log Q_N(A) \right| \left| \frac{f(b_N)}{f(b'_N)} - 1 \right| \\
\leq & Q_N(A)^2 O(1) o(1) \\
=& o(1).
\end{aligned}
$$

We have thus verified that $L_{2,N}(A) = o(1)$. Finally, observe that $L_{3,N}(A) = \left| Q_N(A) - e^{-A} \right| = o(1)$. This completes the verification of (3.5.34) and the proof of Lemma 3.5.6. □

LEMMA 3.5.7. *Suppose Assumption 3.3.2 and condition* (3.2.10) *with some* $\kappa > 0$ *hold. Then, for any* $s > \kappa$*, any real* c*, and any fixed* x*,*

$$
F\left(d_N(x) + cN^{-s}\right) - F\left(d_N(x)\right) = o\left(N^{-1-(s-\kappa)}\right).
$$

Proof. Set $r_N(x) = N\left(1 - F(d_N(x))\right) - e^{-x}$. For any fixed x,

$$
\begin{aligned}
&F\left(d_N(x) + cN^{-s}\right) - F\left(d_N(x)\right) \\
&\quad = \left[1 - F\left(d_N(x)\right)\right] - \left[1 - F\left(d_N(x) + cN^{-s}\right)\right] \\
&\quad = N^{-1}\left[r_N(x) + e^{-x}\right] - N^{-1}\left[r_N\left(x + \frac{cN^{-s}}{a_N}\right) + e^{-\left(x + \frac{cN^{-s}}{a_N}\right)}\right] \\
&\quad = N^{-1}e^{-x}\left[1 - e^{-\frac{cN^{-s}}{a_N}}\right] + N^{-1}\left[r_N(x) - r_N\left(x + \frac{cN^{-s}}{a_N}\right)\right] \\
&\quad = D_{1,N}(x) + D_{2,N}(x).
\end{aligned}
$$

Observe that, for any $s > \kappa$,

$$
\frac{N^{-s}}{a_N} = \frac{N^{-\kappa}}{a_N} N^{-(s-\kappa)} = o\left(N^{-(s-\kappa)}\right). \tag{3.5.35}
$$

Then by Taylor series expansion, there exists $y^\star$ between 0 and $\frac{cN^{-s}}{a_N}$ such that

$$
1 - e^{-\frac{cN^{-s}}{a_N}} = e^{-y^\star}\frac{cN^{-s}}{a_N} = o\left(N^{-(s-\kappa)}\right).
$$

Therefore, for any fixed x,

$$
D_{1,N}(x) = N^{-1}e^{-x}o\left(N^{-(s-\kappa)}\right) = o\left(N^{-1-(s-\kappa)}\right).
$$

Moreover, there exists $z^\star$ between x and $x + \frac{cN^{-s}}{a_N}$ such that

$$
r_N(x) - r_N\left(x + \frac{cN^{-s}}{a_N}\right) = -r_N'(z^\star)\frac{cN^{-s}}{a_N}.
$$

We claim that

$$
r_N'(z^\star) = o(1). \tag{3.5.36}
$$

Then $D_{2,N}(x) = o\left(N^{-1-(s-\kappa)}\right)$ follows from (3.5.35). The verification of (3.5.36) relies on Lemma 3.5.6. Recall that $f(t) = \frac{1-F(t)}{F'(t)}$. Then

$$
\begin{aligned}
|r'_N(z^\star)| &= |a_N N F'(a_N z^\star + b_N) - e^{-z^\star}| \\
&= \Big| \frac{f(b_N)}{f(a_N z^\star + b_N)} N(1 - F(a_N z^\star + b_N)) - e^{-z^\star} \Big| \\
&= \Big| \frac{f(b_N)}{f(a_N z^\star + b_N)} N(1 - F(a_N z^\star + b_N)) - \frac{f(b_N)}{f(a_N z^\star + b_N)} e^{-z^\star} \Big| \\
&\quad + \Big| \frac{f(b_N)}{f(a_N z^\star + b_N)} e^{-z^\star} - e^{-z^\star} \Big| \\
&= T_{1,N} + T_{2,N}.
\end{aligned}
$$

By Taylor series expansion, there exists $t^\star$ between b_N and $a_N z^\star + b_N$ such that

$$
f(a_N z^\star + b_N) = f(b_N) + f'(t^\star) a_N z^\star = f(b_N) + f'(t^\star) f(b_N) z^\star,
$$

then

$$
\frac{f(b_N)}{f(a_N z^\star + b_N)} = \frac{1}{1 + f'(t^\star) z^\star}.
$$

(3.5.35) implies that there exists an integer $N_1 > 0$ such that for any $N > N_1$, $\left| \frac{N^{-s}}{a_N} \right| < 1$. Therefore, setting $A_1 = x - |c|$ and $A_2 = x + |c|$, we have, for $N > N_1$,

$$
x_N := x + \frac{cN^{-s}}{a_N} \in (x - |c|, x + |c|) = (A_1, A_2).
$$

Since x and c are fixed, A_1 and A_2 are also fixed. Therefore, $z^\star \in (A_1, A_2)$ and $t^\star \to \infty$, as $N \to \infty$. Under Assumption 3.3.2, we then have, as $N \to \infty$,

$$
\frac{f(b_N)}{f(a_N z^\star + b_N)} \to 1.
$$

Using Lemma 3.5.6, we then have

$$T_{1,N} = \Big|\frac{f(b_N)}{f(a_N z^\star + b_N)}\Big| |N(1 - F(a_N z^\star + b_N)) - e^{-z^\star}|$$
$$\leq O(1) \sup_{z \geq A_1} |N(1 - F(a_N z + b_N)) - e^{-z}| = o(1).$$

Thus

$$T_{2,N} = \Big|\frac{f(b_N)}{f(a_N z^\star + b_N)} - 1\Big| e^{-z^\star} = o(1),$$

and (3.5.36) follows. □

LEMMA 3.5.8. *Under the assumptions of Lemma 3.5.7, for any $s > \kappa$, any real c, and any fixed x,*

$$F^N(d_N(x) + cN^{-s}) \to \exp\left\{-e^{-x}\right\}, \quad N \to \infty. \tag{3.5.37}$$

Proof. Set $b_N^\star = b_N + cN^{-s}$, so that

$$F^N(d_N(x) + cN^{-s}) = F^N(a_N x + b_N^\star).$$

Hence, relation (3.5.37) is equivalent to

$$\lim_{N\to\infty} F^N(a_N x + b_N^\star) = \exp\left\{-e^{-x}\right\}.$$

Then by Lemma 3.5.2, it is enough to prove the following relation

$$\lim_{N\to\infty} N(1 - F(a_N x + b_N^\star)) = e^{-x}. \tag{3.5.38}$$

Assumption 3.3.2 implies that

$$F^N(d_N(x)) \to \exp\left\{-e^{-x}\right\}, \quad N \to \infty.$$

By Lemma 3.5.2, we then have for any fixed $x \in \mathbb{R}$,

$$\lim_{N\to\infty} N(1 - F(a_N x + b_N)) = e^{-x}.$$

Therefore, by Lemma 3.5.7,

$$\begin{aligned}
&N(1 - F(a_N x + b_N^\star)) \\
&\quad = N\left[1 - F(a_N x + b_N)\right] + N\left[F(a_N x + b_N) - F(a_N x + b_N^\star)\right] \\
&\quad = N\left[1 - F(a_N x + b_N)\right] + N\left[F(d_N(x)) - F(d_N(x) + cN^{-s})\right] \\
&\quad = N\left[1 - F(a_N x + b_N)\right] + o(N^{-(s-\kappa)}) \\
&\quad \to e^{-x}.
\end{aligned}$$

□

LEMMA 3.5.9. *Under the assumptions of Lemma 3.5.7, for any $s > \kappa$, any real c, and any fixed x,*

$$F^N(d_N(x) + cN^{-s}) - F^N(d_N(x)) = o(N^{-(s-\kappa)}). \tag{3.5.39}$$

Proof. Using the inequality shown as (3.5.29), and by Lemma 3.5.7, we have for $c \geq 0$,

$$\begin{aligned}
&F^N(d_N(x) + cN^{-s}) - F^N(d_N(x)) \\
&\quad \leq \left\{F(d_N(x) + cN^{-s}) - F(d_N(x))\right\} \cdot NF^{N-1}(d_N(x) + cN^{-s}) \\
&\quad = o\left(N^{-1-(s-\kappa)}\right) NL_{1,N}^{N-1} = o\left(N^{-(s-\kappa)}\right) L_{1,N}^{-1} L_{1,N}^{N},
\end{aligned}$$

where $L_{1,N} = F(d_N(x) + cN^{-s}) \to 1$. Moreover, by Lemma 3.5.8, $L_{1,N}^N \to \exp\{-e^{-x}\}$. Thus (3.5.39) follows.

Similarly, for $c < 0$,

$$\begin{aligned}
&F^N(d_N(x)) - F^N(d_N(x) + cN^{-s}) \\
&\quad \leq \left\{F^N(d_N(x)) - F^N(d_N(x) + cN^{-s})\right\} \cdot NF^{N-1}(d_N(x)) \\
&\quad = o\left(N^{-1-(s-\kappa)}\right) NL_{2,N}^{N-1} = o\left(N^{-(s-\kappa)}\right) L_{2,N}^{-1} L_{2,N}^{N},
\end{aligned}$$

where $L_{2,N} = F(d_N(x)) \to 1$, and Assumption 3.3.2 implies that, $L_{2,N}^N \to \exp\{-e^{-x}\}$. Thus (3.5.39) follows. □

LEMMA 3.5.10. *Recall that $Y_i = Y_i(j) = \langle X_i, v_j \rangle$, for some fixed j. Suppose Assumptions 3.3.1, 3.3.2 and condition* (3.2.10) *hold with some $\kappa > 0$. For any $s > \kappa$, any real c, and any fixed x,*

$$P(Y_i \leq d_N(x) + cN^{-s}, 1 \leq i \leq N) - P(Y_i \leq d_N(x), 1 \leq i \leq N) = O\left(N^{-\min(r, s-\kappa)}\right).$$

Proof. Denote by $\widetilde{Y}_i$ independent random variables with the same marginal distribution as the Y_i. Proceeding as in the proof of Lemma 3.5.5,

$$\begin{aligned}
&\left| P(Y_i \leq d_N(x) + cN^{-s}, 1 \leq i \leq N) - P(Y_i \leq d_N(x), 1 \leq i \leq N) \right| \\
&\quad \leq \left| P(Y_i \leq d_N(x) + cN^{-s}, 1 \leq i \leq N) - P(\widetilde{Y}_i \leq d_N(x) + cN^{-s}, 1 \leq i \leq N) \right| \\
&\qquad + \left| P(\widetilde{Y}_i \leq d_N(x) + cN^{-s}, 1 \leq i \leq N) - P(\widetilde{Y}_i \leq d_N(x), 1 \leq i \leq N) \right| \\
&\qquad + \left| P(\widetilde{Y}_i \leq d_N(x), 1 \leq i \leq N) - P(Y_i \leq d_N(x), 1 \leq i \leq N) \right| \\
&\quad = T_1 + T_2 + T_3.
\end{aligned}$$

Lemmas 3.5.8 and Assumption 3.3.2 imply that $F^N(d_N(x) + cN^{-s}) \to \exp\{-e^{-x}\}$ and $F^N(d_N(x)) \to \exp\{-e^{-x}\}$. Following the argument used in the proof of Lemma 3.5.5, we know that $T_1 = O(N^{-r})$ and $T_3 = O(N^{-r})$. Then the conclusion follows from Lemma 3.5.9. □

PROOF OF THEOREM 3.3.2 Recall the definitions of events $B_N, B_N^\star$ and A_N introduced at the beginning of the proof of Theorem 3.3.1 and the relation $P(A_N^c) = O(N^{-(1/4-s)})$.

Proceeding as in the proof of Theorem 3.3.1, we have

$$
\begin{aligned}
K_N(x) &= P(B_N^\star) - P(B_N) \\
&\le P(Y_i \le d_N(x) + cN^{-s}, 1 \le i \le N) - P(Y_i \le d_N(x),\ 1 \le i \le N) + P(A_N^c).
\end{aligned}
$$

Thus, by Lemma 3.5.10,

$$K_N(x) \le O(N^{-\min(r,s-\kappa)}) + O(N^{-(1/4-s)}),$$

for any $\kappa < s < 1/4$. Similarly,

$$
\begin{aligned}
-K_N(x) &= P(B_N) - P(B_N^\star) \\
&\le P(Y_i \le d_N(x), 1 \le i \le N) - P\left(Y_i \le d_N(x) - cN^{-s}, 1 \le i \le N\right) + P(A_N^c) \\
&= O(N^{-\min(r,s-\kappa)}) + O(N^{-(1/4-s)}),
\end{aligned}
$$

for any $\kappa < s < 1/4$. Therefore, $|K_N(x)| \le O(N^{-\min(r,s-\kappa)}) + O(N^{-(1/4-s)})$. When $r > 1/8 - \kappa/2$, $K_N(x) = O(N^{-q})$ for any $0 < q \le 1/8 - \kappa/2$; when $0 < r \le 1/8 - \kappa/2$, $K_N(x) = O(N^{-q})$ for any $0 < q \le r$. The conclusion then follows.

Chapter 4

Principal components analysis of regularly varying functions

4.1 Introduction

A fundamental technique of functional data analysis is to replace infinite dimensional curves by coefficients of their projections onto suitable, fixed or data–driven, systems, e.g. [1, 2, 5, 6]. A finite number of these coefficients encode the shape of the curves and are amenable to various statistical procedures. The best systems are those that lead to low dimensional representations, and so provide the most efficient dimension reduction. Of these, the functional principal components (FPCs) have been most extensively used, with hundreds of papers dedicated to the various aspects of their theory and applications.

If $X_1, X_2, \ldots, X_N$ are mean zero iid functions in L^2 with $E\|X_n\|^2 < \infty$, then

$$X_n(t) = \sum_{j=1}^{\infty} \xi_{nj} v_j(t), \quad E\xi_{nj}^2 = \lambda_j. \tag{4.1.1}$$

The FPCs v_j and the eigenvalues λ_j are estimated by $\hat{v}_j$ and $\hat{\lambda}_j$ defined by

$$\int \hat{c}(t,s)\hat{v}_j(s)ds = \hat{\lambda}_j \hat{v}_j(t), \tag{4.1.2}$$

where

$$\hat{c}(t,s) = \frac{1}{N}\sum_{n=1}^{N} X_n(t)X_n(s). \tag{4.1.3}$$

Under the existence of the fourth moment,

$$E\|X\|^4 = \left\{\int X^2(t)dt\right\}^2 < \infty, \tag{4.1.4}$$

it has been shown that for each $j \geq 1$,

$$\limsup_{N\to\infty} NE \|\hat{v}_j - v_j\|^2 < \infty, \quad \limsup_{N\to\infty} NE \left(\hat{\lambda}_j - \lambda_j\right)^2 < \infty, \tag{4.1.5}$$

$$N^{1/2}(\hat{\lambda}_j - \lambda_j) \xrightarrow{d} N(0, \sigma_j^2), \tag{4.1.6}$$

$$N^{1/2}(\hat{v}_j - v_j) \xrightarrow{d} N(0, C_j), \tag{4.1.7}$$

for a suitably defined variance σ_j^2 and a covariance operator C_j. The above relations, especially (4.1.5), have been used to derive large sample justifications of inferential procedures based on the estimated FPCs $\hat{v}_j$. In most scenarios, one can show that replacing the $\hat{v}_j$ by the v_j and the $\hat{\lambda}_j$ by the λ_j is asymptotically negligible. Relations (4.1.5) were established by [7] and extended to weakly dependent functional time series by [8]. Relations (4.1.6) and (4.1.7) follow from the results of [15]. In case of continuous functions satisfying regularity conditions, they follow from the results of [59]. Multivariate versions of (4.1.6) and (4.1.7) have been established in these papers.

A crucial assumption for the relations (4.1.5)–(4.1.7) to hold is the existence of the fourth moment, i.e. (4.1.4), the iid assumption can be relaxed in many ways. Nothing is at present known about the asymptotic properties of the FPCs and their eigenvalues if (4.1.4) does not hold. Our objective is to explore what can be said about the asymptotic behavior of $\widehat{C}$, $\hat{v}_j$ and $\hat{\lambda}_j$ if (4.1.4) fails. We would thus like to consider the case of $E\|X_n\|^2 < \infty$ and $E\|X_n\|^4 = \infty$. Such an assumption is however too general. From mid 1980s to mid 1990s similar questions were posed for scalar time series for which the fourth or even second moment does not exist. A number of results pertaining to the convergence of sample covariances and the periodogram have been derived under the assumption of regularly varying tails, e.g. [60–65]; many others are summarized in the monograph of [10]. The assumption of regular variation is natural because non–normal stable limits can be derived by establishing a connection to random variables in a stable domain of

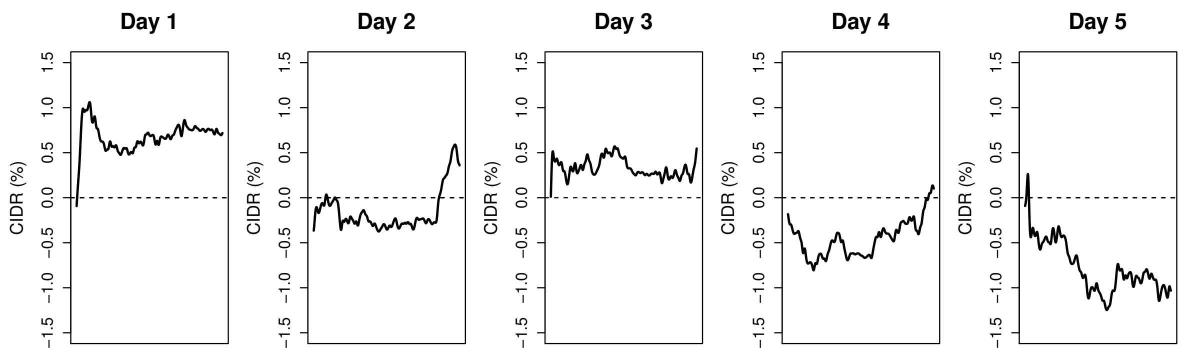

Figure 4.1: Five consecutive intraday return curves, Walmart stock.

attraction, which is characterized by regular variation. This is the approach we take. We assume that the functions X_n are regularly varying in the space L^2 with the index $\alpha \in (2, 4)$, which implies $E\|X_n\|^2 < \infty$ and $E\|X_n\|^4 = \infty$. Suitable definitions and assumptions are presented in Section 4.2.

The paper is organized as follows. The remainder of the introduction provides a practical motivation for the theory we develop. It is not necessary to understand the contribution of the paper, but, we think, it gives a good feel for what is being studied. The formal exposition begins in Section 4.2, in which notation and assumptions are specified. Section 4.3 is dedicated to the convergence of the sample covariance operator (the integral operator with kernel (4.1.3)). These results are then used in Section 4.4 to derive various convergence results for the sample FPCs and their eigenvalues. Section 4.5 shows how the results derived in previous sections can be used in a context of a functional regression model. Its objective is to illustrate the applicability of our theory in a well–known and extensively studied setting. It is hoped that it will motivate and guide applications to other problems of functional data analysis. All proofs which go beyond simple arguments are presented in Section 4.6.

We conclude this introduction by presenting a specific data context. Denote by $P_i(t)$ the price of an asset at time t of trading day i. For the assets we consider in our illustration, t is time in minutes between 9:30 and and 16:00 EST (NYSE opening times) rescaled to the unit interval $(0, 1)$. The intraday return curve on day i is defined by $X_i(t) = \log P_i(t) - \log P_i(0)$. In practice,

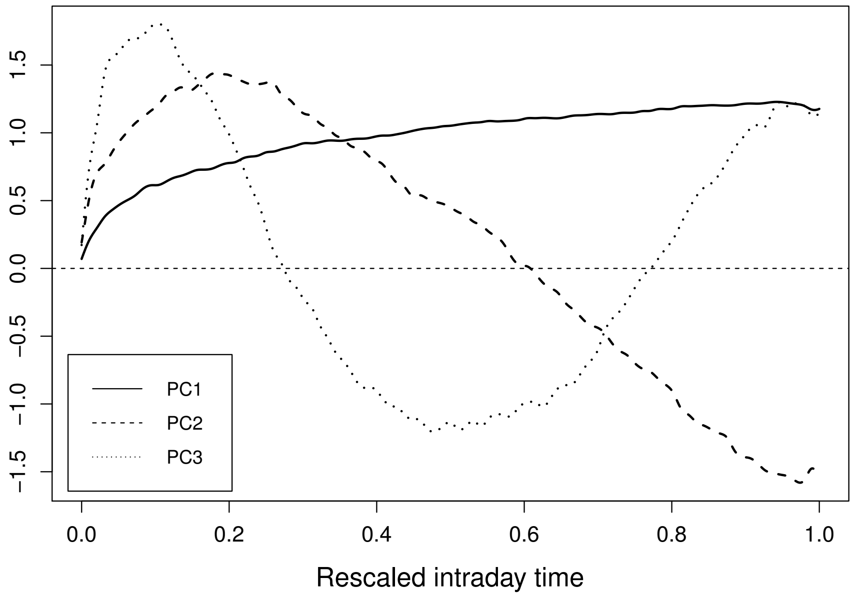

Figure 4.2: The first three sample FPCs of intraday returns on Walmart stock.

$P_i(0)$ is the price after the first minute of trading. The curves R_i show how the return accumulates over the trading day, see e.g. [66]; examples of are shown in Figure 4.1.

The first three sample FPCs, $\hat{v}_1, \hat{v}_2, \hat{v}_3$, are shown in Figure 4.2. They are computed, using (4.1.2), from minute-by-minute Walmart returns form July 05, 2006 to Dec 30, 2011, $N = 1,378$ trading days. (This time interval is used for the other assets we consider.) The curves $\widehat{X}_i(t) = \sum_{j=1}^{3} \hat{\xi}_{ij} \hat{v}_j$, with the scores $\hat{\xi}_{ij} = \int X_i(t)\hat{v}_j(t)dt$, visually approximate the curves X_i very well. One can thus expect that the $\hat{v}_j$ (with properly adjusted sign) are good estimators of the population FPCs v_j in (4.1.1). Relations (4.1.5) and (4.1.7) show that this is indeed the case, *if* $E\|X_1\|^4 < \infty$. (The curves X_i can be assumed to form a stationary time series in L^2, see [67].) We will now argue that the assumption of the finite fourth moment is not realistic, so, with the currently available theory, it is not clear if the $\hat{v}_j$ are good estimators of the v_j. If $E\|X_1\|^4 < \infty$, then $E\xi_{1j}^4 < \infty$ for every j. Figures 4.3 and 4.4 show the Hill plots of the sample score $\hat{\xi}_{ij}$ for four stocks and for $j = 1, 2, 3$. These plots illustrate several properties. 1) It is reasonable to assume that the scores have Pareto tails. 2) The tail index α is smaller than 4, implying that the fourth moment does not exist. 3) It is reasonable to assume that the tail index does not depend on j and is between 2 and 4. The evidence for the last assumption is stronger in Figure 4.3, but Figure 4.4 does not refute it. With such a motivation, we are now able to formalize in the next section the setting of this paper.

4.2 Preliminaries

The functions X_n are assumed to be independent and identically distributed in L^2, with the same distribution as X, which is regularly varying with index $\alpha \in (2, 4)$. By $L^2 := L^2(\mathcal{T})$, we denote the usual separable Hilbert space of square integrable functions on some compact subset $\mathcal{T}$ of an Euclidean space. In a typical FDA framework, $\mathcal{T} = [0, 1]$, e.g. Chapter 2 of [6]. Regular variation in finite–dimensional spaces has been a topic of extensive research for decades, see e.g. [9, 68, 69]. We shall need the concept of regular variation of measures on *infinitely-dimensional* function spaces. To this end, we start by recalling some terminology and fundamental facts about regularly varying functions.

A measurable function $L : (0, \infty) \to \mathbb{R}$ is said to be slowly varying (at infinity) if, for all $\lambda > 0$,

$$\frac{L(\lambda u)}{L(u)} \to 1, \qquad \text{as } u \to \infty.$$

Functions of the form $R(u) = u^\rho L(u)$ are said to be regularly varying with exponent $\rho \in \mathbb{R}$.

The notion of regular variation extends to measures and provides an elegant and powerful framework for establishing limit theorems. It was first introduced by [70] and has been since extended to Banach and even metric spaces using the notion of M_0 convergence (see e.g. [19]). Even though we will work only with Hilbert spaces, we review the theory in a more general context.

Consider a separable Banach space $\mathbb{B}$ and let $B_\epsilon := \{z \in \mathbb{B} : \|z\| < \epsilon\}$ be the open ball of radius $\epsilon > 0$, centered at the origin. A Borel measure μ defined on $\mathbb{B}_0 := \mathbb{B} \backslash \{\mathbf{0}\}$ is said to be *boundedly finite* if $\mu(A) < \infty$, for all Borel sets that are bounded away from $\mathbf{0}$, that is, such that $A \cap B_\epsilon = \emptyset$, for some $\epsilon > 0$. Let $\mathbb{M}_0$ be the collection of all such measures. For $\mu_n, \mu \in \mathbb{M}_0$, we say that the μ_n converge to μ in the M_0 topology, if $\mu_n(A) \to \mu(A)$, for all bounded away from $\mathbf{0}$, μ-continuity Borel sets A, i.e., such that $\mu(\partial A) = 0$, where $\partial A := \overline{A} \setminus A^\circ$ denotes the boundary

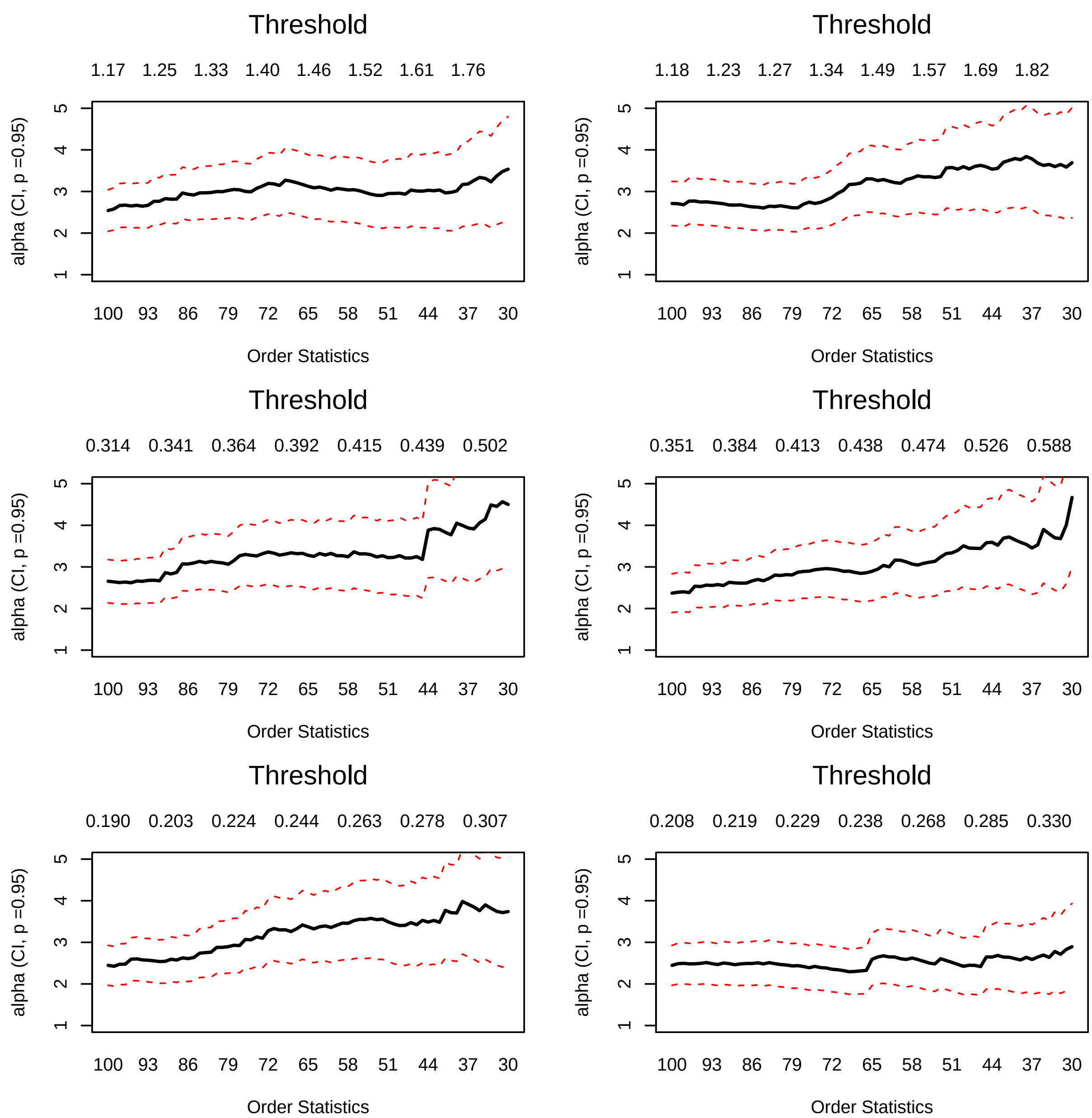

Figure 4.3: Hill plots (an estimate of α as a function of upper order statistics) for sample FPC scores for **Walmart** (left) and **IBM** (right). From top to bottom: levels $j = 1, 2, 3$.

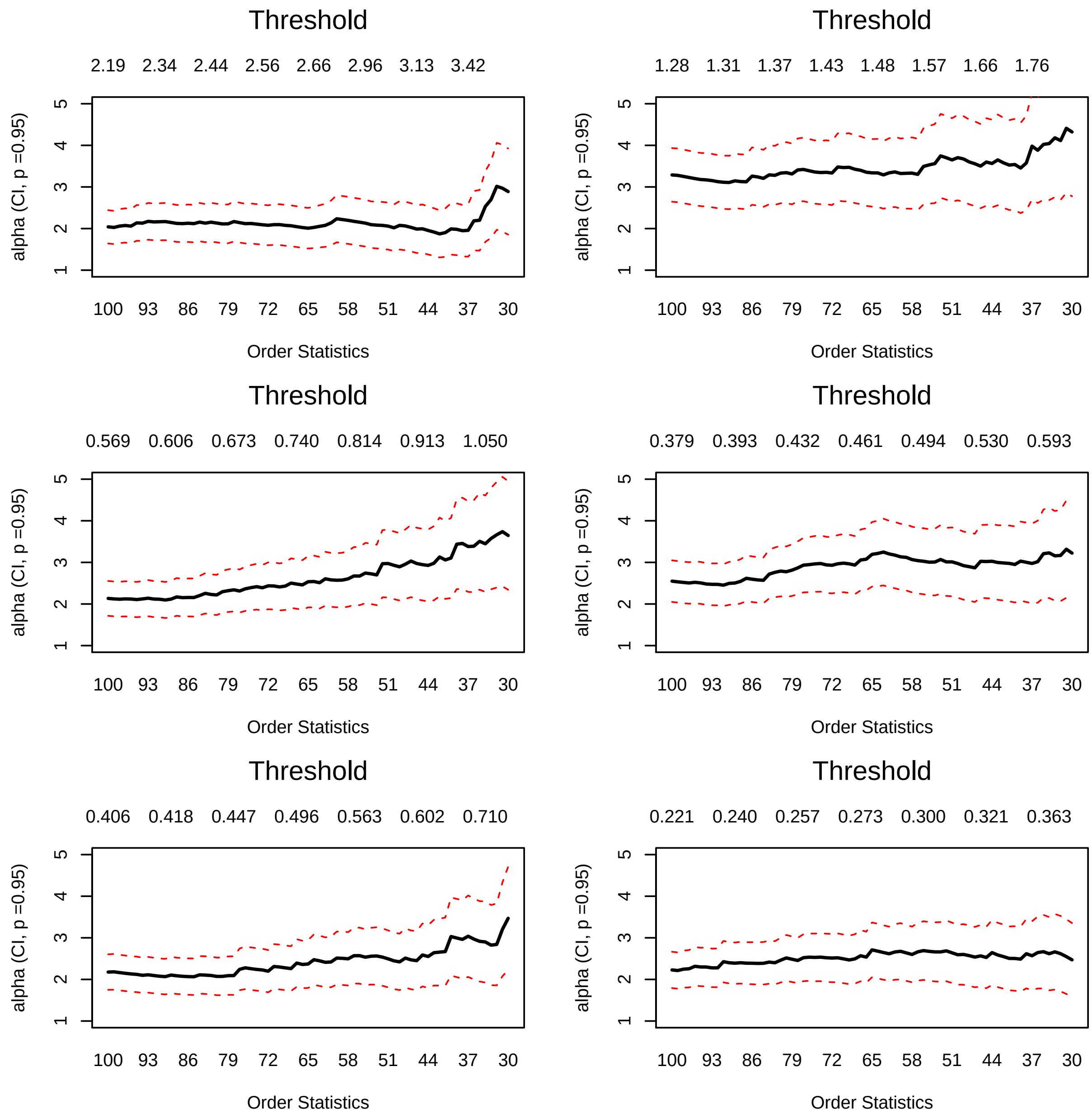

Figure 4.4: Hill plots as in Figure 4.3 for **J. P. Morgan–Chase** (left) and **Exxon–Mobil** (right).

of A. The M_0 convergence can be metrized such that $\mathbb{M}_0$ becomes a complete separable metric space (Theorem 2.3 in [19] and also Section 2.2. of [20]). The following result is known, see e.g. Chapter 2 of [20] and references therein.

PROPOSITION 4.2.1. *Let X be a random element in a separable Banach space $\mathbb{B}$ and $\alpha > 0$. The following three statements are equivalent:*

(i) For some slowly varying function L,

$$P(\|X\| > u) = u^{-\alpha} L(u) \tag{4.2.8}$$

and

$$\frac{P(u^{-1}X \in \cdot)}{P(\|X\| > u)} \xrightarrow{M_0} \mu(\cdot), \quad u \to \infty, \tag{4.2.9}$$

where μ is a non-null measure on the Borel σ-field $\mathcal{B}(\mathbb{B}_0)$ of $\mathbb{B}_0 = \mathbb{B} \backslash \{\mathbf{0}\}$.

(ii) There exists a probability measure Γ on the unit sphere $\mathbb{S}$ in $\mathbb{B}$ such that, for every $t > 0$,

$$\frac{P(\|X\| > tu, X/\|X\| \in \cdot)}{P(\|X\| > u)} \xrightarrow{w} t^{-\alpha}\Gamma(\cdot), \quad u \to \infty.$$

(iii) Relation (4.2.8) *holds, and for the same spectral measure Γ in (ii),*

$$P\left(X/\|X\| \in \cdot \mid \|X\| > u\right) \xrightarrow{w} \Gamma(\cdot), \quad u \to \infty.$$

DEFINITION 4.2.1. *If any one of the equivalent conditions in Proposition 4.2.1 hold, we shall say that X is regularly varying with index α. The measures μ and Γ will be referred to as exponent and angular measures of X, respectively.*

The measure Γ is sometimes called the spectral measure, but we will use the adjective "spectral" in the context of stable measures which appear in Section 4.3. It is important to distinguish the angular measure of a regularly varying random function and a spectral measure of a stable distribution, although they are related. We also note that we call α the tail index, and $-\alpha$ the tail exponent.

We will work under the following assumption.

ASSUMPTION 4.2.1. *The random element X in the separable Hilbert space $H = L^2$ is regularly varying with index $\alpha \in (2,4)$. The observations $X_1, X_2, \ldots$ are iid copies of X.*

Assumption 4.2.1 is a coordinate free condition not related in any way to functional principal components. The next assumption relates the asymptotic behavior of the FPC scores to the assumed regular variation. It implies, in particular, that the expansion $X(t) = \sum_{j=1}^{\infty} \xi_j v_j(t)$ contains infinitely many terms, so that we study infinite dimensional objects. We will see in the proofs of Proposition 4.3.1 and Theorem 4.3.2 that under Assumption 4.2.1 the limit

$$Q_{nm} = \lim_{u\to\infty} \frac{P\left(\left\{\sum_{j=n}^{\infty} \xi_j^2\right\}^{1/2} \left\{\sum_{j=m}^{\infty} \xi_j^2\right\}^{1/2} > u\right)}{P\left(\sum_{j=1}^{\infty} \xi_j^2 > u\right)}$$

exists and is finite. We impose the following assumption related to condition (4.2.9).

ASSUMPTION 4.2.2. *For every $n, m \geq 1$, $Q_{nm} > 0$.*

Assumption 4.2.2 postulates, intuitively, that the tail sums $\sum_{j=n}^{\infty} \xi_j^2$ must have extreme probability tails comparable to that of $\|X\|^2$.

We now collect several useful facts that will be used in the following. The exponent measure μ satisfies

$$\mu(tA) = t^{-\alpha}\mu(A), \quad \forall t > 0, \quad A \in \mathcal{B}(\mathbb{B}_0). \tag{4.2.10}$$

It admits the polar coordinate representation via the angular measure Γ. That is, if $x = r\theta$, where $r := \|x\|$ and $\theta = x/\|x\|$, for $x \neq \mathbf{0}$, we have

$$\mu(dx) = \alpha r^{-\alpha-1} dr \Gamma(d\theta). \tag{4.2.11}$$

This means that for every bounded measurable function f that vanishes on a neighborhood of $\mathbf{0}$, we have

$$\int_{\mathbb{B}} f(x)\mu(dx) = \int_{\mathbb{S}}\int_0^\infty f(r\theta)\alpha r^{-\alpha-1}dr\Gamma(d\theta).$$

There exists a sequence $\{a_N\}$ such that

$$NP\left(X \in a_N A\right) \to \mu(A), \tag{4.2.12}$$

for any set A in $\mathcal{B}(\mathbb{B}_0)$ with $\mu(\partial A) = 0$. One can take, for example,

$$a_N = N^{1/\alpha}L_0(N), \tag{4.2.13}$$

with a slowly varying function L_0 satisfying $L_0^{-\alpha}(N)L(N^{1/\alpha}L_0(N)) \to 1$.

We will work with Hilbert–Schmidt operators. A linear operator $\Psi : H \to H$ is Hilbert–Schmidt if $\sum_{j=1}^{\infty} \|\Psi(e_j)\|^2 < \infty$, where $\{e_j\}$ is any orthonormal basis of H. Every Hilbert–Schmidt operator is bounded. The space of Hilbert–Schmidt operators will be denoted by $\mathcal{S}$. It is itself a separable Hilbert space with the inner product

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} = \sum_{j=1}^{\infty} \langle \Psi_1(e_j), \Psi_2(e_j) \rangle .$$

If Ψ is an integral operator defined by $\Psi(x)(t) = \int \psi(t,s)x(s)ds,\ x \in L^2$, then $\|\Psi\|_{\mathcal{S}}^2 = \iint \psi^2(t,s)dtds$.

Relations (4.1.5) essentially follow from the bound

$$E\left\|\widehat{C} - C\right\|_{\mathcal{S}}^2 \le N^{-1}E\left\|X\right\|^4,$$

where the subscript $\mathcal{S}$ indicates the Hilbert–Schmidt norm. Under Assumption 4.2.1 such a bound is useless because, by (4.2.8), $E\left\|X\right\|^4 = \infty$. In fact, one can show that under Assumption 4.2.1, $E\|\widehat{C}\|_{\mathcal{S}}^2 = \infty$, so no other bound on $E\|\widehat{C} - C\|_{\mathcal{S}}^2$ can be expected. The following Proposition 4.2.2

implies however that under Assumption 4.2.1 the population covariance operator C is a Hilbert-Schmidt operator, and $\widehat{C} \in \mathcal{S}$ with probability 1. This means that the space $\mathcal{S}$ does provide a convenient framework.

PROPOSITION 4.2.2. *Suppose X is a random element of L^2 with $E\|X\|^2 < \infty$ and $\widehat{C}$ is the sample covariance operator based on n iid copies of X. Then $C \in \mathcal{S}$ and $\widehat{C} \in \mathcal{S}$ with probability 1.*

4.3 Limit distribution of $\widehat{C}$

We will show that $Nk_N^{-1}(\widehat{C} - C)$ converges to an $\alpha/2$–stable Hilbert–Schmidt operator, for an appropriately defined regularly varying sequence $\{k_N\}$. Unless stated otherwise, all limits in the following are taken as $N \to \infty$.

Observe that for any $x \in H$,

$$\begin{aligned} Nk_N^{-1}\left(\widehat{C} - C\right)(x) &= Nk_N^{-1}\left(N^{-1}\sum_{n=1}^{N}\langle X_n, x\rangle X_n - E[\langle X_1, x\rangle X_1]\right) \\ &= k_N^{-1}\left(\sum_{n=1}^{N}\langle X_n, x\rangle X_n - NE[\langle X_1, x\rangle X_1]\right) \\ &= k_N^{-1}\left(\sum_{n=1}^{N}(X_n \otimes X_n)(x) - NE[(X_1 \otimes X_1)](x)\right), \end{aligned} \tag{4.3.14}$$

where $(X_n \otimes X_n)(x) = \langle X_n, x\rangle X_n$. Since the $X_n \otimes X_n$ are Hilbert–Schmidt operators, the last expression shows a connection between the asymptotic distribution of $\widehat{C}$ and convergence to a stable limit in the Hilbert space $\mathcal{S}$ of Hilbert–Schmidt operators. We therefore restate below, as Theorem 4.3.1, Theorem 4.11 of [71] which provides conditions for the stable domain of attraction in a separable Hilbert space. The Hilbert space we will consider in the following will be $\mathcal{S}$ and the stability index will be $\alpha/2, \alpha \in (2, 4)$. However, when stating the result of Kuelbs and Mandrekar, we will use a generic Hilbert space H and the generic stability index $p \in (0, 2)$. Recall that for a stable random element $S \in H$ with index $p \in (0, 2)$, there exists a *spectral* measure σ_S defined on the unit sphere $\mathbb{S}_H = \{z \in H : \|z\| = 1\}$, such that the characteristic functional of S is given by

$$E \exp\{i \langle x, S \rangle\} = \exp\left\{ i \langle x, \beta_S \rangle - \int_{\mathbb{S}} | \langle x, s \rangle |^p \sigma_S(ds) + iC(p, x) \right\}, \quad x \in H, \qquad (4.3.15)$$

where

$$C(p, x) = \begin{cases} \tan \frac{\pi p}{2} \int_{\mathbb{S}} \langle x, s \rangle \, | \langle x, s \rangle |^{p-1} \sigma_S(ds) & \text{if } p \neq 1, \\ \frac{2}{\pi} \int_{\mathbb{S}} \langle x, s \rangle \log | \langle x, s \rangle | \sigma_S(ds) & \text{if } p = 1. \end{cases}$$

We denote the above representation by $S \sim [p, \sigma_S, \beta_S]$. The p-stable random element S is necessarily regularly varying with index $p \in (0, 2)$. In fact, its angular measure is precisely the normalized spectral measure appearing in (4.3.15), i.e.,

$$\Gamma_S(\cdot) = \frac{\sigma_S(\cdot)}{\sigma_S(\mathbb{S}_H)}.$$

[71] derived sufficient and necessary conditions on the distribution of Z under which

$$b_N^{-1} \left(\sum_{i=1}^{N} Z_i - \gamma_N \right) \xrightarrow{d} S, \qquad (4.3.16)$$

where the Z_i are iid copies of Z. They assume that the support of the distribution of S, equivalently of the distribution of Z, spans the whole Hilbert space H. In our context, we will need to work with Z whose distribution is not supported on the whole space. Denote by $L(Z)$ the smallest closed subspace which contains the support of the distribution of Z. Then $L(Z)$ is a Hilbert space itself with the inner product inherited from H. Denote by $\{e_j, \ j \in \mathbb{N}\}$ an orthonormal basis of $L(Z)$. We assume that this is an infinite basis because we consider infinite dimensional data. (The finite dimensional case has already been dealt with by [72].) Introduce the projections

$$\pi_m(z) = \sum_{j=m}^{\infty} \langle z, e_j \rangle \, e_j, \quad z \in H.$$

THEOREM 4.3.1. *Let Z_1, Z_2, ... be iid random elements in a separable Hilbert space H with the same distribution as Z. Let $\{e_j, \ j \in \mathbb{N}\}$ be an orthonormal basis of $L(Z)$. There exist normalizing constants b_N and γ_N such that* (4.3.16) *holds* **if and only if**

$$\frac{P\left(\|\pi_m(Z)\| > tu\right)}{P\left(\|Z\| > u\right)} \to \frac{c_m}{c_1} t^{-p}, \;\; u \to \infty, \tag{4.3.17}$$

where for each $m \geq 1$*,* $c_m > 0$*, and* $\lim_{m\to\infty} c_m = 0$*, and where*

$$\frac{P\left(\|Z\| > u, Z/\left\|Z\right\| \in A\right)}{P\left(\|Z\| > u, Z/\left\|Z\right\| \in A^\star\right)} \to \frac{\sigma_S(A)}{\sigma_S(A^\star)}, \;\; u \to \infty, \tag{4.3.18}$$

for all continuity sets A*,* $A^\star \in \mathcal{B}(\mathbb{S}_H)$ *with* $\sigma_S(A^\star) > 0$*.*

If (4.3.16) *holds, the sequence* b_N *must satisfy*

$$b_N \to \infty, \;\; \frac{b_N}{b_{N+1}} \to 1, \;\; N b_N^{-2} E\left(\|Z\|^2 I_{\{||Z|| \leq b_N\}}\right) \to \lambda_p \sigma_S(\mathbb{S}_H), \tag{4.3.19}$$

where

$$\lambda_p = \begin{cases} \frac{p(1-p)}{\Gamma(3-p)\cos(\pi p/2)} & , \text{ if } p \neq 1 \\ 2/\pi & , \text{ if } p = 1, \end{cases} \tag{4.3.20}$$

and $\Gamma(a) := \int_0^\infty e^{-x} x^{a-1} dx, a > 0$ *is the Euler gamma function. Furthermore, the* $\gamma_N \in H$ *may be chosen as*

$$\gamma_N = NE\left(Z I_{\{||Z|| \leq b_N\}}\right). \tag{4.3.21}$$

REMARK 4.3.1. *The origin of the constant* λ_p *appearing in* (4.3.19) *can be understood as follows. Consider the simple scalar case* $H = \mathbb{R}$*. Let* Z *be symmetric* α*-stable with* $E[e^{iZx}] = e^{-c|x|^\alpha}$*,* $x \in \mathbb{R}$*, where in this case,* $c = \sigma(\mathbb{S}_H) \equiv \sigma(\{-1, 1\}) > 0$*. Consider iid copies* Z_i*,* $i = 1, 2, \ldots$ *of* Z *and observe that by the* p*-stability property*

$$\frac{1}{N^{1/\alpha}} \sum_{j=1}^{N} Z_j \stackrel{d}{=} Z \equiv S,$$

and hence (4.3.16) *holds trivially with* $b_N := N^{1/\alpha}$ *and* $\gamma_N := 0$*.*

On the other hand, by Proposition 1.2.15 on page 16 in [73], we have

$$P(|Z| > x) \sim \frac{c(1-p)}{\Gamma(2-p)\cos(\pi p/2)} x^{-p}, \quad \textit{as } x \to \infty.$$

This along with an integration by parts and an application of Karamata's theorem yield $N b_N^{-2} E[Z^2 I_{\{|Z| \leq b_N\}}] \to \lambda_p \sigma_S(\mathbb{S}_H)$*, giving the constant in* (4.3.19).

PROPOSITION 4.3.1. *Conditions* (4.3.17) *and* (4.3.18) *in Theorem 4.3.1 hold if and only if* Z *is regularly varying in* H *with index* $p \in (0,2)$ *and for each* $m \geq 1$, $\mu_Z(A_m) > 0$*, where*

$$A_m = \left\{ z \in H : \|\pi_m(z)\| = \left\| \sum_{j=m}^{\infty} \langle z, e_j \rangle e_j \right\| > 1 \right\}. \tag{4.3.22}$$

Our next objective is to show that if X is a regularly varying element of a separable Hilbert space H whose index is $\alpha > 0$, then the operator $Y = X \otimes X$ is regularly varying with index $\alpha/2$, in the space of Hilbert–Schmidt operators. If $y, z \in H$, then $y \otimes z$ is an element of $\mathcal{S}$ defined by $(y \otimes z)(x) = \langle y, x \rangle z, \ x \in H$. It is easy to check that $\|y \otimes z\|_{\mathcal{S}} = \|y\| \, \|z\|$. If $B_1, B_2 \subset H$, we denote by $B_1 \otimes B_2$ the subset of $\mathcal{S}$ defined as the set of operators of the form $x_1 \otimes x_2$, with $x_1 \in B_1, x_2 \in B_2$. Denote by $\mathbb{S}_H$ the unit sphere in H centered at the origin, and by $\mathbb{S}_{\mathcal{S}}$ such a sphere in $\mathcal{S}$.

The next result is valid for all $\alpha > 0$.

PROPOSITION 4.3.2. *Suppose* X *is a regularly varying element with index* $\alpha > 0$ *of a separable Hilbert space* H*. Then the operator* $Y = X \otimes X$ *is a regularly varying element with index* $\alpha/2$ *of the space* $\mathcal{S}$ *of Hilbert-Schmidt operators.*

REMARK 4.3.2. *The proof of Proposition 4.3.2 shows that the angular measure of* $X \otimes X$ *is supported on the* diagonal (4.6.40) *and*

$$\Gamma_{X \otimes X}(B \otimes B) = \Gamma_X(B), \ \ \forall B \subset \mathcal{B}(\mathbb{S}_H).$$

The next result specifies the limit distribution of the sums of the $X_i \otimes X_i$ based on the results derived so far.

THEOREM 4.3.2. *Suppose Assumptions 4.2.1 and 4.2.2 hold. Then, there exist normalizing constants k_N and operators ψ_N such that*

$$k_N^{-1}\left(\sum_{i=1}^{N} X_i \otimes X_i - \psi_N\right) \xrightarrow{d} S, \tag{4.3.23}$$

where $S \in \mathcal{S}$ is a stable random operator, $S \sim [\alpha/2, \sigma_S, 0]$, where the spectral measure σ_S is defined on the unit sphere $\mathbb{S}_{\mathcal{S}} = \{y \in \mathcal{S} : \|y\|_{\mathcal{S}} = 1\}$. The normalizing constants may be chosen as follows

$$k_N = \left(\frac{\alpha}{4-\alpha}\right)^{2/\alpha} a_N^2, \quad \psi_N = NE\left[(X \otimes X)\, I_{\{||X||^2 \le k_N\}}\right]. \tag{4.3.24}$$

The final result of this section specifies the asymptotic distribution of $\widehat{C} - C$.

THEOREM 4.3.3. *Suppose Assumptions 4.2.1 and 4.2.2 hold. Then,*

$$Nk_N^{-1}(\widehat{C} - C) \xrightarrow{d} S - \frac{\alpha}{\alpha - 2}\int_{\mathbb{S}_H} (\theta \otimes \theta)\, \Gamma_X(d\theta), \tag{4.3.25}$$

where $S \in \mathcal{S}$ and $\{k_N\}$ are as in Theorem 4.3.2. ($k_N = N^{2/\alpha}L(N)$ for a slowly varying L.)

If the X_i are scalars, then the angular measure Γ_X is concentrated on $\mathbb{S}_H = \{-1, 1\}$, with $\Gamma_X(1) = p, \Gamma_X(-1) = 1 - p$, in the notation of [61]. Thus $\int_{\mathbb{S}_H} \theta^2 \Gamma_X(d\theta) = 1$, and we recover the centering $\alpha/(\alpha - 2)$ in Theorem 2.2 of [61]. Relation (4.3.25) explains the structure of this centering in a much more general context.

Theorem 4.3.3 readily leads to a strong law of large numbers which can be derived by an application of the following result, a consequence of Theorem 3.1 of [74].

THEOREM 4.3.4. *Suppose $Y_i, i \ge 1$, are iid mean zero elements of a separable Hilbert space with $E\|Y_i\|^\gamma < \infty$, for some $1 \le \gamma < 2$. Then,*

$$\frac{1}{N^{1/\gamma}}\sum_{i=1}^{N} Y_i \xrightarrow{P} 0 \quad \textit{if and only if} \quad \frac{1}{N^{1/\gamma}}\sum_{i=1}^{N} Y_i \xrightarrow{a.s.} 0.$$

Set $Y_i = X_i \otimes X_i - E[X \otimes X]$. Then the Y_i are iid mean zero elements of $\mathcal{S}$ which, by Proposition 4.3.2, satisfy $E\|Y_i\|_{\mathcal{S}}^{\gamma} < \infty$, for any $\gamma \in (0, \alpha/2)$. Theorem 4.3.3 implies that for any $\gamma \in (0, \alpha/2)$, $N^{-1/\gamma}\sum_{i=1}^{N} Y_i \xrightarrow{P} 0$. Thus Theorem 4.3.4 leads to the following corollary.

COROLLARY 4.3.1. *Suppose Assumptions 4.2.1 and 4.2.2 hold. Then, for any* $\gamma \in [1, \alpha/2)$, $N^{1-1/\gamma}\|\widehat{C} - C\|_{\mathcal{S}} \to 0$ *with probability 1.*

4.4 Convergence of eigenfunctions and eigenvalues

We first formulate and prove a general result which allows us to derive the asymptotic distributions of the eigenfunctions and eigenvalues of an estimator of the covariance operator from the asymptotic distribution of the operator itself. The proof of this results is implicit in the proofs of the results of Section 2 of [15], which pertain to the asymptotic normality of the sample covariance operator if $E\|X\|^4 < \infty$. The result and the technique of proof are however more general, and can be used in different contexts, so we state and prove it in detail.

ASSUMPTION 4.4.1. *Suppose C is the covariance operator of a random function X taking values in L^2 such that $E\|X\|^2 < \infty$. Suppose $\widehat{C}$ is an estimator of C which is a.s. symmetric, nonnegative–definite and Hilbert–Schmidt. Assume that for some random operator $Z \in \mathcal{S}$, and for some $r_N \to \infty$,*

$$Z_N := r_N(\widehat{C} - C) \xrightarrow{d} Z.$$

In our setting, $Z \in \mathcal{S}$ is specified in (4.3.25), and $r_N = N^{\beta}L(N)$ for some $0 < \beta < 1/2$. More precisely,

$$r_N = Na_N^{-2}, \; a_N = N^{1/\alpha}L_0(N), \; \alpha \in (2, 4).$$

We will work with the eigenfunctions and eigenvalues defined by

$$C(v_j) = \lambda_j v_j, \;\; \widehat{C}_j(\hat{v}_j) = \hat{\lambda}_j \hat{v}_j, \;\; j \geq 1.$$

Assumption 4.4.1 implies that $\hat{\lambda}_j \geq 0$ and the $\hat{v}_j$ are orthogonal with probability 1. We assume that that, like the v_j, the $\hat{v}_j$ have unit norms. To lighten the notation, we assume that sign$\langle \hat{v}_j, v_j \rangle$ = 1. This sign does not appear in any of our final results, it cancels in the proofs. We assume that both sets of eigenvalues are ordered in decreasing order. The next assumption is standard, it ensures that the population eigenspaces are one dimensional.

ASSUMPTION 4.4.2. $\lambda_1 > \lambda_2, \ldots, > \lambda_p > \lambda_{p+1}$.

Set

$$T_j = \sum_{k \neq j} (\lambda_j - \lambda_k)^{-1} \langle Z, v_j \otimes v_k \rangle \, v_k.$$

Lemma 4.6.2 shows that the series defining T_j converges a.s. in L^2.

THEOREM 4.4.1. *Suppose Assumptions 4.4.1 and 4.4.2 hold. Then,*

$$r_N \left\{ \hat{v}_j - v_j, \; 1 \leq j \leq p \right\} \xrightarrow{d} \left\{ T_j, \; 1 \leq j \leq p \right\}, \;\; \text{in } (L^2)^p,$$

and

$$r_N \left\{ \hat{\lambda}_j - \lambda_j, \; 1 \leq j \leq p \right\} \xrightarrow{d} \left\{ \langle Z(v_j), v_j \rangle, \; 1 \leq j \leq p \right\}, \;\; \text{in } \mathbb{R}^p.$$

If Z is an $(\alpha/2)$–stable random operator in $\mathcal{S}$, then the T_j are jointly $(\alpha/2)$–stable random functions in L^2, and $\langle Z(v_j), v_j \rangle$ are jointly $(\alpha/2)$–stable random variables. This follows directly from the definition of a stable distribution, e.g. Section 6.2 of [75]. Under Assumption 4.2.1, $r_N = N^{1-2/\alpha} L_0^{-2}(N)$. Theorem 4.4.1 thus leads to the following corollary.

COROLLARY 4.4.1. *Suppose Assumptions 4.2.1, 4.2.2 and 4.4.2 hold. Then,*

$$N^{1-2/\alpha} L_0^{-2}(N) \{\hat{v}_j - v_j,\ 1 \le j \le p\} \xrightarrow{d} \{T_j,\ 1 \le j \le p\},\ \text{ in } (L^2)^p,$$

where the T_j are jointly $(\alpha/2)$–stable in L^2, and

$$N^{1-2/\alpha} L_0^{-2}(N) \left\{\hat{\lambda}_j - \lambda_j,\ 1 \le j \le p\right\} \xrightarrow{d} \{S_j,\ 1 \le j \le p\},\ \text{ in } \mathbb{R}^p,$$

where the S_j are jointly $(\alpha/2)$–stable in $\mathbb{R}$.

Corollary 4.4.1 implies the rates in probability $\hat{v}_j - v_j = O_P(r_N^{-1})$ and $\hat{\lambda}_j - \lambda_j = O_P(r_N^{-1})$, with $r_N = N^{1-2/\alpha} L_0^{-2}(N)$. This means, that the distances between $\hat{v}_j$ and $\hat{\lambda}_j$ and the corresponding population parameters are approximately of the order $N^{2/\alpha-1}$, i.e. are asymptotically larger that these distances in the case of $E\|X\|^4 < \infty$, which are of the order $N^{-1/2}$. Note that $2/\alpha - 1 \to -1/2$, as $\alpha \to 4$.

It is often useful to have some bounds on moments, analogous to relations (4.1.5). Since the tails of $\|T_j\|$ and $|S_j|$ behave like $t^{-\alpha/2}$, e.g. Section 6.7 of [75], $E\|T_j\|^\gamma < \infty,\ 0 < \gamma < \alpha/2$, with an analogous relation for $|S_j|$. We can thus expect convergence of moments of order $\gamma \in (0, \alpha/2)$. The following theorem specifies the corresponding results.

THEOREM 4.4.2. *If Assumptions 4.2.1 and 4.2.2 hold, then for each $\gamma \in (0, \alpha/2)$, there is a slowly varying function L_γ such that*

$$\limsup_{N\to\infty} N^{\gamma(1-2/\alpha)} L_\gamma(N) E \left\| \widehat{C} - C \right\|_{\mathcal{S}}^{\gamma} < \infty$$

and for $j \ge 1$,

$$\limsup_{N\to\infty} N^{\gamma(1-2/\alpha)} L_\gamma(N) E |\hat{\lambda}_j - \lambda_j|^\gamma < \infty.$$

If, in addition, Assumption 4.4.2 holds, then for $1 \le j \le p$,

$$\limsup_{N\to\infty} N^{\gamma(1-2/\alpha)} L_\gamma(N) E \|\hat{v}_j - v_j\|^\gamma < \infty.$$

Several cruder bounds can be derived from Theorem 4.4.2. In applications, it is often convenient to take $\gamma = 1$. Then $E\|\widehat{C} - C\|_{\mathcal{S}} \le N^{2/\alpha-1} L_1(N)$. By Potter bounds, e.g. Proposition 2.6 (ii) in [68], for any $\epsilon > 0$ there is a constant C_ϵ such that for $x > x_\epsilon$ $L_1(x) \le C_\epsilon x^\epsilon$. For each $\alpha \in (2,4)$, we can choose ϵ so small that $-\delta(\alpha) := 2/\alpha - 1 + \epsilon < 0$. This leads to the following corollary.

COROLLARY 4.4.2. *If Assumptions 4.2.1 and 4.2.2 hold, then for each $\alpha \in (2,4)$, there are constant C_α and $\delta(\alpha) > 0$ such that*

$$E\|\widehat{C} - C\|_{\mathcal{S}} \le C_\alpha N^{-\delta(\alpha)} \quad \text{and} \quad E\|\hat{\lambda}_j - \lambda_j\| \le C_\alpha N^{-\delta(\alpha)}.$$

If, in addition, Assumption 4.4.2 holds, then for $1 \le j \le p$, $E\|\hat{v}_j - v_j\| \le C_\alpha(j) N^{-\delta(\alpha)}$.

Corollary 4.4.2 implies that $E\|\widehat{C} - C\|_{\mathcal{S}}$, $E\|\hat{\lambda}_j - \lambda_j\|$ and $E\|\hat{v}_j - v_j\|$ tend to zero, for any $\alpha \in (2,4)$.

4.5 An application: functional linear regression

One of the most widely used tools of functional data analysis is the functional regression model, e.g. [1, 6, 49]. Suppose $X_1, X_2, \ldots, X_N$ are explanatory functions, $Y_1, Y_2, \ldots, Y_N$ are response functions, and assume that

$$Y_i(t) = \int_0^1 \psi(t,s) X_i(s) ds + \varepsilon_i(t), \quad 1 \le i \le N, \tag{4.5.26}$$

where $\psi(\cdot,\cdot)$ is the kernel of $\Psi \in \mathcal{S}$. The X_i are mean zero iid functions in $L^2 = L^2([0,1])$, and so are the error functions ε_i. Consequently, the Y_i are iid in L^2. A question that has been investigated from many angles is how to consistently estimate the regression kernel $\psi(\cdot,\cdot)$. An estimator that has become popular following the work of [76] can be constructed as follows.

The population version of (4.5.26) is $Y(t) = \int \psi(t,s)X(s)ds + \varepsilon(t)$. Denote by v_i the FPCs of X and by u_j those of Y, so that

$$X(s) = \sum_{i=1}^{\infty} \xi_i v_i(s), \quad Y(t) = \sum_{j=1}^{\infty} \zeta_j u_j(t).$$

If ε is independent of X, then, with $\lambda_\ell = E[\xi_\ell^2]$,

$$\psi(t,s) = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{E[\xi_\ell \zeta_k]}{\lambda_\ell} u_k(t) v_\ell(s),$$

with the series converging in $L^2([0,1] \times [0,1])$, equivalently in $\mathcal{S}$, see Lemma 8.1 in [6]. This motivates the estimator

$$\hat{\psi}_{KL}(t,s) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \frac{\hat{\sigma}_{\ell k}}{\hat{\lambda}_\ell} \hat{u}_k(t) \hat{v}_\ell(s),$$

where $\hat{u}_k$ are the eigenfunctions of $\widehat{C}_Y$ and $\hat{\sigma}_{\ell k}$ is an estimator of $E[\xi_\ell \zeta_k]$. [76] study the above estimator under the assumption that data are observed sparsely and with measurement errors. This requires two-stage smoothing, so their assumptions focus on conditions on the various smoothing parameters and the random mechanism that generates the sparse observations. Like in all work of this type, they assume that the underlying functions have finite fourth moments: $E\|X\|^4 < \infty$, $E\|\varepsilon\|^4 < \infty$, and so $E\|Y\|^4 < \infty$. Our objective is to show that if the X_i satisfy the assumptions of Section 4.2, then

$$\left\| \widehat{\Psi}_{KL} - \Psi \right\|_{\mathcal{L}} \stackrel{a.s.}{\to} 0, \tag{4.5.27}$$

as $N \to \infty$, and $K, L \to \infty$ at suitable rates determined by the rate of decay of the eigenvalues. The integral operators Ψ and $\widehat{\Psi}_{KL}$ are defined by their kernels $\psi(\cdot,\cdot)$ and $\hat{\psi}_{KL}(\cdot,\cdot)$, respectively. We focus on moment conditions, so we assume that the functions X_i, Y_i are fully observed, and use the estimator

$$\hat{\sigma}_{\ell k} = \frac{1}{N} \sum_{i=1}^{N} \hat{\xi}_{i\ell} \hat{\zeta}_{ik}, \quad \hat{\xi}_{i\ell} = \langle X_i, \hat{v}_\ell \rangle, \; \hat{\zeta}_{ik} = \langle Y_i, \hat{u}_k \rangle.$$

Since the regression operator Ψ is infinitely dimensional, we strengthen Assumption 4.4.2 to the following assumption.

ASSUMPTION 4.5.1. *The eigenvalues* $\lambda_i = E\xi_i^2$ *and* $\gamma_j = E\zeta_j^2$ *satisfy*

$$\lambda_1 > \lambda_2 > \ldots > 0, \quad \gamma_1 > \gamma_2 > \ldots > 0.$$

Many issues related to the infinite dimension of the functional data in model (4.5.26) are already present when considering projections on the unobservable subspaces

$$\mathcal{V}_L = \operatorname{span}\{v_1, v_2, \ldots, v_L\}, \quad \mathcal{U}_K = \operatorname{span}\{u_1, u_2, \ldots, u_K\}.$$

Therefore we first consider the convergence of the operator with the kernel

$$\psi_{KL}(t,s) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \frac{\sigma_{\ell k}}{\lambda_\ell} u_k(t) v_\ell(s).$$

Set $\sigma_{\ell k} = E[\xi_\ell \zeta_k]$ and observe that

$$\psi_{KL}(t,s) - \psi(t,s) = -\sum_{k>K \text{ or } \ell>L} \frac{\sigma_{\ell k}}{\lambda_\ell} u_k(t) v_\ell(s).$$

Therefore

$$\|\Psi_{KL} - \Psi\|_{\mathcal{L}}^2 \le \|\Psi_{KL} - \Psi\|_{\mathcal{S}}^2 = \sum_{k>K \text{ or } \ell>L} \frac{\sigma_{\ell k}^2}{\lambda_\ell^2}. \tag{4.5.28}$$

The condition

$$\sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{\sigma_{\ell k}^2}{\lambda_\ell^2} < \infty, \tag{4.5.29}$$

which is Assumption (A1) of [76], implies that the remainder term is asymptotically negligible. It is instructive to rewrite condition (4.5.29) in a different form. Observe that

$$\sigma_{\ell k} = E[\xi_l \langle \Psi(X) + \varepsilon, u_k \rangle] = E[\xi_l \sum_{i=1}^{\infty} \xi_i \langle \Psi(v_i), u_k \rangle] = \lambda_\ell \langle \Psi(v_\ell), u_k \rangle. \tag{4.5.30}$$

Therefore

$$\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\frac{\sigma_{\ell k}^2}{\lambda_\ell^2}=\sum_{\ell=1}^{\infty}\frac{1}{\lambda_\ell^2}\sum_{k=1}^{\infty}\lambda_\ell^2\left\langle\Psi(v_\ell),u_k\right\rangle^2=\sum_{\ell=1}^{\infty}\left\|\Psi(v_\ell)\right\|^2=\left\|\Psi\right\|_{\mathcal{S}}^2. \tag{4.5.31}$$

We see that condition (4.5.29) simply means that Ψ is a Hilbert–Schmidt operator, and so it holds under our general assumptions on model (4.5.26).

The last assumption implicitly restricts the rates at which K and L tend to infinity with N. Under Assumption 4.5.1, the following quantities are well defined

$$\alpha_j=\min\left\{\lambda_j-\lambda_{j+1},\lambda_{j-1}-\lambda_j\right\},\ j\geq 2,\quad \alpha_1=\lambda_1-\lambda_2, \tag{4.5.32}$$

$$\beta_j=\min\left\{\gamma_j-\gamma_{j+1},\gamma_{j-1}-\gamma_j\right\},\ j\geq 2,\quad \beta_1=\gamma_1-\gamma_2. \tag{4.5.33}$$

ASSUMPTION 4.5.2. *The truncation levels K and L tend to infinity with N in such a way that for some $\gamma\in(1,\alpha/2)$,*

$$\limsup_{N\to\infty}\lambda_L^{-3/2}L^{1/2}N^{1/\gamma-1}<\infty, \tag{4.5.34}$$

$$\limsup_{N\to\infty}\lambda_L^{-1}\left(\sum_{j=1}^{L}\alpha_j^{-1}\right)N^{1/\gamma-1}<\infty, \tag{4.5.35}$$

$$\limsup_{N\to\infty}\lambda_L^{-1}K^{1/2}N^{1/\gamma-1}<\infty, \tag{4.5.36}$$

$$\limsup_{N\to\infty}\lambda_L^{-1}\left\{\left(\sum_{k=1}^{K}\beta_k^{-1}\right)+\left(\sum_{k=1}^{K}\beta_k^{-2}\right)^{1/2}\right\}N^{1/\gamma-1}<\infty. \tag{4.5.37}$$

The conditions in Assumption 4.5.2 could be restated or unified; and could be replaced by slightly different conditions by modifying the technique of proof. The essence of this assumption is that K and L must tend to infinity sufficiently slowly, and the rate is influenced by index α; the closer α is to 4, the larger γ can be taken, so K and L can be larger.

THEOREM 4.5.1. *Suppose model* (4.5.26) *holds with* $\Psi \in \mathcal{S}$*, the* X_i *and the* Y_i *satisfying Assumptions 4.2.1 and 4.2.2, and square integrable* ε_i*,* $E\|\varepsilon_i\|^2 < \infty$*. Then relation* (4.5.27) *holds under Assumptions 4.5.1 and 4.5.2.*

4.6 Proofs of the results of previous sections

Throughout the proofs, we will use relatively well–known properties of slowly varying functions, which we collect in Lemma 4.6.1 for ease of reference. For the proofs and many more details, see e.g., [9, 77].

LEMMA 4.6.1. *If* L *is a slowly varying function, then:*

(i) $L_1(u) = L(u^\rho),\ \rho > 0$ *and* $L_2(u) = |L(u)|^a,\ a \in \mathbb{R}$ *are slowly varying.*

(ii) (Potter bounds) *For all* $\delta > 0$*, we have* $L(u) = o(u^\delta)$*, as* $u \to \infty$*.*

(iii) (Karamata's Theorem) *For all* $\rho > -1$ *and* $\eta > 1$*, as* $u \to \infty$*, we have*

$$\int_0^u x^\rho L(x)dx \sim \frac{u^{\rho+1}L(u)}{(\rho+1)} \quad \textit{and} \quad \int_u^\infty x^{-\eta} L(x)dx \sim \frac{u^{-(\eta-1)}L(u)}{(\eta-1)},$$

where $a(u) \sim b(u)$ *means* $a(u)/b(u) \to 1$*, as* $u \to \infty$*.*

4.6.1 Proofs of Proposition 4.2.2 and of the results of Section 4.3

Proof of Proposition 4.2.2

Since C is a covariance operator, it is nuclear ($\sum_{j\geq 1} \lambda_j < \infty$), e.g. Theorem 11.2.2 of [49], and so it is Hilbert–Schmidt ($\sum_{j\geq 1} \lambda_j^2 < \infty$).

We now verify that $\widehat{C}$ is a.s. a Hilbert-Schmidt operator. Observe that

$$
\begin{aligned}
\|\widehat{C}\|_{\mathcal{S}}^2 &= \iint \hat{c}^2(t,s)dtds \\
&= \iint \left\{ \frac{1}{N} \sum_{n=1}^{N} X_n(t)X_n(s) \right\}^2 dtds \\
&= \frac{1}{N^2} \iint \left\{ \sum_{n=1}^{N} \sum_{j=1}^{\infty} \xi_{nj} v_j(t) \sum_{j'=1}^{\infty} \xi_{nj'} v_{j'}(s) \right\}^2 dtds.
\end{aligned}
$$

Set

$$
S_n(t,s) = \sum_{j=1}^{\infty} \xi_{nj} v_j(t) \sum_{j'=1}^{\infty} \xi_{nj'} v_{j'}(s)
$$

so that

$$
N^2 \|\widehat{C}\|_{\mathcal{S}}^2 = \sum_{n,m=1}^{N} \iint S_n(t,s) S_m(t,s) dtds.
$$

Next, observe that

$$
\iint S_n(t,s) S_m(t,s) dtds = \sum_{j,j'=1}^{\infty} \sum_{i,i'=1}^{\infty} \xi_{nj} \xi_{nj'} \xi_{mi} \xi_{mi'} \int v_j(t) v_i(t) dt \int v_{j'}(s) v_{i'}(s) ds.
$$

Therefore, by the orthonormality by the v_j,

$$
\iint S_n(t,s) S_m(t,s) dtds = \sum_{j,j'=1}^{\infty} \xi_{nj} \xi_{nj'} \xi_{mj} \xi_{mj'}
$$

and

$$
N^2 \|\widehat{C}\|_{\mathcal{S}}^2 = \sum_{n,m=1}^{N} \sum_{j,j'=1}^{\infty} \xi_{nj} \xi_{nj'} \xi_{mj} \xi_{mj'}. \tag{4.6.38}
$$

By (4.6.38), it suffices to verify that for each fixed m and n, the series $\sum_{j,j'=1}^{\infty} \xi_{nj} \xi_{nj'} \xi_{mj} \xi_{mj'}$ converges almost surely. Set

$$
W(J) = \sum_{j=1}^{J} \xi_{nj} \xi_{mj},
$$

and observe that

$$\sum_{j=1}^{J}\sum_{j'=1}^{J'}\xi_{nj}\xi_{nj'}\xi_{mj}\xi_{mj'} = W(J)W(J').$$

We see that it is enough to show that the sequence $W(J)$, $J \geq 1$, converges a.s. We will verify the absolute convergence of the series defining it. It follows from the bounds

$$\begin{aligned}\sum_{j=1}^{J}|\xi_{nj}\xi_{mj}| &\leq \frac{1}{2}\sum_{j=1}^{J}\xi_{nj}^2 + \frac{1}{2}\sum_{j=1}^{J}\xi_{mj}^2 \\ &\leq \frac{1}{2}\sum_{j=1}^{\infty}\xi_{nj}^2 + \frac{1}{2}\sum_{j=1}^{\infty}\xi_{mj}^2 \\ &= \frac{1}{2}\int_0^1 X_n^2(t)dt + \frac{1}{2}\int_0^1 X_m^2(t)dt.\end{aligned}$$

The right-hand side is finite a.s. because each X_n is a random element of L^2.

Proof of Proposition 4.3.1

Set

$$\Gamma(\cdot) = \frac{\sigma_S(\cdot)}{\sigma_S(\mathbb{S}_H)} \tag{4.6.39}$$

Recall that (4.6.39) specifies the relationship between the stable spectral measure σ_S and the angular measure Γ of a regularly varying distribution appearing in Proposition 4.2.1.

First we assume (4.3.17) and (4.3.18) hold. Take $m = 1$ in (4.3.17) and $A^\star = \mathbb{S}_H$ in (4.3.18), we then have for every $t > 0$,

$$\begin{aligned}\frac{P(\|Z\| > tu, Z/\|Z\| \in A)}{P(\|Z\| > u)} &= \frac{P(\|Z\| > tu, Z/\|Z\| \in A)}{P(\|Z\| > tu, Z/\|Z\| \in \mathbb{S}_H)}\frac{P(\|Z\| > tu)}{P(\|Z\| > u)} \\ &\to \frac{\sigma_S(A)}{\sigma_S(\mathbb{S}_H)}t^{-p} \quad (u \to \infty) \\ &= \Gamma(A)t^{-p},\end{aligned}$$

for any continuity set A of σ_S (equivalently, of Γ). Thus condition (ii) in Proposition 4.2.1 holds, which implies that Z is regularly varying with index p.

Next we assume that Z is regularly varying with index p, and show that (4.3.17) and (4.3.18) will hold. Using condition (ii) in Proposition 4.2.1, we have

$$\frac{P\left(\|Z\| > u, Z/\|Z\| \in A\right)}{P\left(\|Z\| > u, Z/\|Z\| \in A^\star\right)} = \frac{P(\|Z\| > u, Z/\|Z\| \in A)}{P\left(\|Z\| > u\right)} \frac{P\left(\|Z\| > u\right)}{P\left(\|Z\| > u, Z/\|Z\| \in A^\star\right)}$$
$$\to \frac{\Gamma(A)}{\Gamma(A^\star)} = \frac{\sigma_S(A)}{\sigma_S(A^\star)}, \quad (u \to \infty)$$

for all continuity sets A, $A^\star \in \mathcal{B}(\mathbb{S}_H)$ with $\sigma_S(A^\star) > 0$. Then, with the set A_m defined by (4.3.22),

$$\begin{aligned}\frac{P\left(\|\pi_m(Z)\| > tu\right)}{P\left(\|Z\| > u\right)} &= \frac{P(t^{-1}u^{-1}Z \in A_m)}{P(u^{-1}Z \in A_1)} \\ &= \frac{P(\|Z\| > u)}{P(u^{-1}Z \in A_1)} \frac{P(t^{-1}u^{-1}Z \in A_m)}{P(\|Z\| > tu)} \frac{P\left(\|Z\| > tu\right)}{P\left(\|Z\| > u\right)} \\ &\to \frac{\mu_Z(A_m)}{\mu_Z(A_1)} t^{-p} =: \frac{c_m}{c_1} t^{-p}, \quad (u \to \infty)\end{aligned}$$

where the above convergence follows from (4.2.9) provided we can show that $A_m,\ m \geq 1$ are continuity sets of the measure μ_Z. We do that next.

By the definition of A_m in (4.3.22) and since π_m is continuous and homogeneous, we have

$$\partial A_m = \{z \in H : \|\pi_m(z)\| = 1\} \quad \text{and} \quad \partial(rA_m) = r\partial A_m = \{z \in H : \|\pi_m(z)\| = r\}.$$

Furthermore, we have that $r_1 A_m \supset r_2 A_m$ for all $0 < r_1 < r_2$. This implies that $A_m = \cup_{r>1} \partial(rA_m)$, where the sets $\partial(rA_m)$ are all disjoint in r. By the homogeneity of μ_Z, however, (recall (4.2.10)) it follows that $\mu_Z(\partial(rA_m)) = r^{-p}\mu_Z(\partial A_m)$. In particular,

$$\mu_Z(A_m) \geq \sum_i \mu_Z(\partial(r_i A_m)) = \sum_i r_i^{-p} \mu_Z(\partial A_m),$$

for any sequence $r_i > 1$. If $\mu_Z(\partial A_m) > 0$, then by taking r_i's such that $\sum_i r_i^{-p} = \infty$, we obtain $\mu_Z(A_m) = \infty$, which is not possible since A_m is bounded away from zero. We have thus shown that $\mu_Z(\partial A_m) = 0$, i.e., A_m is a continuity set of μ_Z for all $m \geq 1$.

To complete the proof of (4.3.17), it remains is to show that $c_m = \mu_Z(A_m) \to 0$, as $m \to \infty$. Notice that $A_m \supset A_{m+1}$ and thus $\lim_{m\to\infty} \mu_Z(A_m) = \mu_Z(\cap_{m=1}^{\infty} A_m)$, since $\mu(A_1) < \infty$. It is easy to see that $\cap_{m=1}^{\infty} A_m = \emptyset$. Indeed, for each $z \in H$, we have $\|z\|^2 = \sum_{j=1}^{\infty} \langle z, e_j \rangle^2 < \infty$ and therefore

$$\|\pi_m(z)\|^2 = \sum_{j=m}^{\infty} \langle z, e_j \rangle^2 \to 0, \qquad \text{as } m \to \infty.$$

If $z \in \cap_{m\geq 1} A_m$, then $\|\pi_m(z)\| > 1$ for each $m \geq 1$, which is impossible.

Proof of Proposition 4.3.2

Since $\|Y\|_{\mathcal{S}} = \|X\|^2$ and $P(\|X\| > u) = u^{-\alpha} L(u)$, we conclude that

$$P(\|Y\|_{\mathcal{S}} > u) = u^{-\alpha/2} L(u^{1/2}).$$

Notice that $u \mapsto L(u^{1/2})$ is a slowly varying function. Thus, by Proposition 4.2.1 (iii), to establish the regular variation of Y it remains to show that there must exist a probability measure Γ_Y on $\mathbb{S}_{\mathcal{S}}$ such that

$$P\left(\|Y\|_{\mathcal{S}}^{-1} Y \in A | \, \|Y\|_{\mathcal{S}} > u\right) \to \Gamma_Y(A), \;\; u \to \infty,$$

for every Γ_Y-continuity set A. The operator Y takes values only in a small subset of $\mathbb{S}_{\mathcal{S}}$, namely in

$$\mathbb{S}_{\mathcal{S}}(1) = \{\Psi \in \mathbb{S}_{\mathcal{S}} : \Psi = x \otimes x \text{ for some } x \in \mathbb{S}_H\}. \tag{4.6.40}$$

The set $\mathbb{S}_{\mathcal{S}}(1)$ is closed in $\mathbb{S}_{\mathcal{S}}$ and its Borel subsets have the form $B \otimes B$, where B is a Borel subset of $\mathbb{S}_H$. We know that

$$\Gamma^{(u)}(B) := P\left(X/\|X\| \in B | \, \|X\| > u\right) \to \Gamma(B), \;\; u \to \infty,$$

for every Γ-continuity set $B \in \mathbb{S}_H$. Denote by ξ_u a random element of H taking values in $\mathbb{S}_H$ whose distribution is $\Gamma^{(u)}$. Then we have

$$\xi_u \xrightarrow{d} \xi, \quad u \to \infty, \tag{4.6.41}$$

where ξ has distribution Γ. Furthermore, denote by η_u a random element of $\mathcal{S}$ taking values in $\mathbb{S}_{\mathcal{S}}(1)$ whose distribution is

$$P(\eta_u \in A) = \frac{P\left(\|Y\|_{\mathcal{S}}^{-1} Y \in A, \|Y\|_{\mathcal{S}} > u\right)}{P(\|Y\|_{\mathcal{S}} > u)}, \quad A \in \mathbb{S}_{\mathcal{S}}(1).$$

We want to identify a random element η such that

$$\eta_u \xrightarrow{d} \eta, \quad u \to \infty, \tag{4.6.42}$$

whose distribution will be the desired measure Γ_Y.

We first verify that

$$\eta_u \stackrel{d}{=} \xi_{u^{1/2}} \otimes \xi_{u^{1/2}}. \tag{4.6.43}$$

Relation (4.6.43) is equivalent to

$$\frac{P\left(\|Y\|_{\mathcal{S}}^{-1} Y \in A, \|Y\|_{\mathcal{S}} > u\right)}{P(\|Y\|_{\mathcal{S}} > u)} = P\left(\xi_{u^{1/2}} \otimes \xi_{u^{1/2}} \in A\right), \quad \forall A \in \mathbb{S}_{\mathcal{S}}(1). \tag{4.6.44}$$

Set $A = B \otimes B$. Since $\|Y\|_{\mathcal{S}} = \|X\|^2$, the left–hand side of (4.6.44) is

$$\begin{aligned}
\frac{P\left(\|Y\|_{\mathcal{S}}^{-1} Y \in A, \|Y\|_{\mathcal{S}} > u\right)}{P(\|Y\|_{\mathcal{S}} > u)} &= \frac{P\left(\left(\|X\|^{-1} X\right) \otimes \left(\|X\|^{-1} X\right) \in B \otimes B, \|X\| > u^{1/2}\right)}{P(\|X\| > u^{1/2})} \\
&= \frac{P\left(\|X\|^{-1} X \in B, \|X\| > u^{1/2}\right)}{P(\|X\| > u^{1/2})} \\
&= \Gamma^{(u^{1/2})}(B),
\end{aligned}$$

while the right–hand side of (4.6.44) is

$$P\left(\xi_{u^{1/2}} \otimes \xi_{u^{1/2}} \in A\right) = P\left(\xi_{u^{1/2}} \in B, \xi_{u^{1/2}} \in B\right) = P\left(\xi_{u^{1/2}} \in B\right) = \Gamma^{(u^{1/2})}(B). \tag{4.6.45}$$

Therefore, (4.6.43) holds. It remains to show that

$$\eta_u \stackrel{d}{=} \xi_{u^{1/2}} \otimes \xi_{u^{1/2}} \stackrel{d}{\to} \xi \otimes \xi =: \eta, \quad u \to \infty.$$

The above relation holds because by (4.6.45) and (4.6.41),

$$P\left(\xi_{u^{1/2}} \otimes \xi_{u^{1/2}} \in A\right) = \Gamma^{(u^{1/2})}(B) \to \Gamma(B) = P\left(\xi \in B\right) = P\left(\eta \in A\right),$$

provided B is a continuity set of Γ. Using the relation $\|y \otimes z\|_{\mathcal{S}} = \|y\| \, \|z\|$, it is easy to check that $x_n \otimes x_n \to x \otimes x$ in $\mathcal{S}$ if and only if $x_n \to x$ in H. Hence, $\partial A = \partial B \otimes \partial B$, so the continuity sets of the distribution of η have the form $B \otimes B$ with $\Gamma(\partial B) = 0$.

Proof of Theorem 4.3.2

By Proposition 4.3.2, the operators $X_i \otimes X_i$ are iid regularly varying elements of $\mathcal{S}$, whose index of regular variation is $\alpha/2 \in (1, 2)$. In order to use Theorem 4.3.1, we first verify that $\mu_{X \otimes X}(A_m) > 0$, cf. Proposition 4.3.1. This is where Assumption 4.2.2 comes into play. An orthonormal basis of $L(X \otimes X)$ is $\{v_i \otimes v_j, i, j \geq 1\}$, where the v_j are the FPCs of X. Set

$$A_{n,m} = \left\{ \Psi \in \mathcal{S} : \left\| \sum_{i=n}^{\infty} \sum_{j=m}^{\infty} \langle \Psi, v_i \otimes v_j \rangle_{\mathcal{S}} \, v_i \otimes v_j \right\|_{\mathcal{S}} > 1 \right\}.$$

We must thus verify that $\mu_{X \otimes X}(A_{n,m}) > 0$. By (4.2.9),

$$\mu_{X \otimes X}(A_{n,m}) = \lim_{u \to \infty} \frac{P(X \otimes X \in u A_{n,m})}{P(\|X \otimes X\|_{\mathcal{S}} > u)}.$$

Clearly

$$P(\|X \otimes X\|_{\mathcal{S}} > u) = P(\|X\|^2 > u) = P\left(\sum_{j=1}^{\infty} \xi_j^2 > u \right),$$

which is the denominator of Q_{nm} in Assumption 4.2.2. Turning to the numerator, observe that $X \otimes X \in u A_{nm}$ iff

$$\left\| \sum_{i=n}^{\infty} \sum_{j=m}^{\infty} \langle X \otimes X, v_i \otimes v_j \rangle_{\mathcal{S}} \, v_i \otimes v_j \right\|_{\mathcal{S}} > u.$$

Direct verification, which uses the definition of the inner product in $\mathcal{S}$ and the orthonormality of the v_j, shows that $\langle X \otimes X, v_i \otimes v_j \rangle_{\mathcal{S}} = \xi_i \xi_j$. It follows that $X \otimes X \in uA_{nm}$ iff

$$\left\| \sum_{i=n}^{\infty} \sum_{j=m}^{\infty} \xi_i \xi_j v_i \otimes v_j \right\|_{\mathcal{S}}^2 > u^2.$$

Using the definition of the Hilbert–Schmidt norm and the orthogonality of the v_j again, we see that the above inequality is equivalent to $\sum_{i=n}^{\infty} \xi_i^2 \sum_{j=m}^{\infty} \xi_j^2 > u^2$, so $P(X \otimes X \in uA_{nm})$ is equal to the numerator of Q_{nm}.

It remains to show that the normalizing sequences can be chosen as specified in (4.3.24). It is easy to check that $k_N \to \infty$ and $\frac{k_N}{k_{N+1}} \to 1$. We will show that

$$N k_N^{-2} E \left(\|X\|^4 I_{\left\{ ||X||^2 \le k_N \right\}} \right) \to 1, \tag{4.6.46}$$

which in view of (4.3.19) would yield (4.3.23), where the spectral measure of the limit S is normalized so that $\lambda_p \sigma_S(\mathbb{S}_{\mathcal{S}}) = 1$ with λ_p in (4.3.20).

Observe that by the Tonelli-Fubini Theorem, we have

$$\begin{aligned} E\left[\|X\|^4 I_{\left\{ ||X||^2 \le k_N \right\}} \right] &= E\left[\int_0^{\infty} I_{\{x < \|X\|^4 \le k_N^2\}} dx \right] \\ &= \int_0^{k_N^2} \left[P(\|X\|^4 > x) - P(\|X\|^2 > k_N) \right] dx \\ &= \int_0^{k_N^2} x^{-\alpha/4} L(x^{1/4}) dx - k_N^2 k_N^{-\alpha/2} L(k_N^{1/2}), \end{aligned}$$

where we used the fact that $P(\|X\| > x) = x^{-\alpha} L(x)$. Now, by applying Karamata's theorem (Lemma 4.6.1 (iii)) to the integral in the last expression, we obtain

$$E\left[\|X\|^4 I_{\{||X||^2\le k_N\}}\right] \sim \frac{1}{(1-\alpha/4)} k_N^{2-\alpha/2} L(k_N^{1/2}) - k_N^{2-\alpha/2} L(k_N^{1/2})$$
$$= \left(\frac{4}{(4-\alpha)} - 1\right) k_N^{2-\alpha/2} L(k_N^{1/2})$$
$$= \frac{\alpha}{(4-\alpha)} k_N^{2-\alpha/2} L(k_N^{1/2}), \tag{4.6.47}$$

as $k_N \to \infty$, where $c_N \sim d_N$ means that $c_N/d_N \to 1$.

In view of (4.2.12) by taking $A = \{x : \|x\| > 1\}$, we obtain

$$NP(\|X\| > a_N) = N a_N^{-\alpha} L(a_N) \to 1, \tag{4.6.48}$$

since μ is normalized so that $\mu(A) = 1$. Thus, multiplying (4.6.47) by Nk_N^{-2} and recalling (4.3.24), we obtain

$$Nk_N^{-2} E\left(\|X\|^4 I_{\{||X||^2\le k_N\}}\right) \sim c_\alpha^\alpha k_N^{-\alpha/2} L(k_N^{1/2}) = a_N^{-\alpha} L(c_\alpha a_N),$$

where $c_\alpha = (\alpha/(4-\alpha))^{1/\alpha}$. Since L is a slowly varying function, we have $L(c_\alpha a_N) \sim L(a_N)$ as $a_N \to \infty$, and therefore by (4.6.48), we obtain (4.6.46). This completes the proof.

Proof of Theorem 4.3.3

Observe that by (4.3.14),

$$Nk_N^{-1}\left(\widehat{C} - C\right) = k_N^{-1}\left(\sum_{n=1}^{N} X_n \otimes X_n - \psi_N\right) + k_N^{-1} N E\left[(X\otimes X) I_{\{\|X\|^2 > k_N\}}\right], \tag{4.6.49}$$

with k_N and ψ_N as in Theorem 4.3.2. The first term converges to S, so we must verify the existence of the second term, show that it converges, and describe its limit. The issue is subtle because $k_N \to \infty$ implies that $k_N^{-1} N\left[(X\otimes X) I_{\{\|X\|^2\ge k_N\}}\right] \to 0$ with probability 1, yet the expected value does not tend to zero even in the case of scalar observations, see Theorem 2.2 of [61]. It is convenient to approach the problem in a slightly more general setting.

Suppose Y is a regularly varying element of a separable Hilbert space whose index of regular variation is $p,\ p \in (1,2)$. In our application, $Y = X \otimes X$, the Hilbert space is $\mathcal{S}$ and $p =$

$\alpha/2$. Denote by μ_Y the exponent measure of Y and by u_N a regularly varying sequence such that $NP(\|Y\| > u_N) \to 1$, so that

$$\mu_{N,Y}(A) := \frac{P(Y \in u_N A)}{P(\|Y\| > u_N)} \to \mu_Y(A), \tag{4.6.50}$$

with the usual restrictions on the set A, cf. Proposition 4.2.1. Set

$$Y_N = u_N^{-1} N Y I_{\{\|Y\| > u_N\}}$$

and observe that $E[Y_N]$ exists in the sense of Bochner. Indeed, by (4.2.8) and the Potter bounds (Lemma 4.6.1), we have

$$P(\|Y\| > u) = u^{-p} L(u) = o(u^{-p+\delta}), \qquad \text{as } u \to \infty,$$

for an arbitrarily small $\delta > 0$. Since $p \in (1, 2)$, by taking $p - \delta > 1$, we obtain $E[\|Y\|] = \int_0^\infty P(\|Y\| > y) dy < \infty$ and the expectation of Y and hence Y_N is well-defined.

Now set $M_N = E[Y_N]$. We want to identify $M \in H$ such that $\|M_N - M\| \to 0$. We will show that the above convergence holds with

$$M = \int_{\mathbb{B}^c} y \mu_Y(dy), \tag{4.6.51}$$

where $\mathbb{B} = \{y : \|y\| \leq 1\}$. Recall that Y is regularly varying and by (4.2.11) its exponent and angular measures are related as follows

$$\mu_Y(dy) = p r^{-p-1} dr \Gamma_Y(d\theta), \tag{4.6.52}$$

where $r := \|y\|$ and $\theta := y/\|y\|$ are polar coordinates in H. Thus, in polar coordinates, we obtain

$$\int_{\mathbb{B}^c} \|y\| \mu_Y(dy) = \int_1^\infty \int_{\mathbb{S}} r\|\theta\|\Gamma_Y(d\theta) p r^{-p-1} dr \tag{4.6.53}$$
$$= \left(p \int_1^\infty r^{-p} dr\right) \int_{\mathbb{S}} \|\theta\| \Gamma_Y(d\theta) \tag{4.6.54}$$
$$= \frac{p}{p-1}. \tag{4.6.55}$$

This shows that the Bochner integral in (4.6.51) is well defined and in fact equals

$$M = \frac{p}{p-1} \int_{\mathbb{S}} \theta \Gamma_Y(d\theta).$$

In view of Remark 4.3.2, by taking $Y = X \otimes X$ and $p = \alpha/2$, we then obtain

$$M = \frac{\alpha}{\alpha - 2} \int_{\mathbb{S}_H} (\theta \otimes \theta)\, \Gamma_X(d\theta),$$

which is the expression for the offset in (4.3.25).

Observe that by the definition (4.6.50) of $\mu_{N,Y}$, since $NP(\|Y\| > u_N) \to 1$, for any Bochner integrable mapping of the Hilbert space into itself, or to the real line,

$$NE[f(u_N^{-1}Y)] \sim \int f(y) \mu_{N,Y}(dy). \tag{4.6.56}$$

Therefore,

$$M_N = NE\left[u_N^{-1} Y I_{\mathbb{B}^c}(u_N^{-1}Y)\right] \sim \int_{\mathbb{B}^c} y \mu_{N,Y}(dy).$$

Observe that $\mu_{N,Y}(\mathbb{B}^c) = 1$, and by (4.6.52),

$$\mu_Y(\mathbb{B}^c) = \int_1^\infty \int_{\mathbb{S}} p r^{-p-1} dr \Gamma_Y(d\theta) = \sigma_Y(\mathbb{S}) = 1.$$

Thus $\mu_{N,Y}$ and μ_Y are probability measures on $\mathbb{B}^c$, and we want to show that

$$\int_{\mathbb{B}^c} y \mu_{N,Y}(dy) \to \int_{\mathbb{B}^c} y \mu_Y(dy).$$

Since $\mu_{N,Y}$ converges weakly to μ_Y, it suffices to verify that

$$\sup_{N\geq 1}\int_{\mathbb{B}^c} \|y\|^{1+\delta}\,\mu_{N,Y}(dy) < \infty, \tag{4.6.57}$$

for some $\delta > 0$ (this implies strong uniform integrability). Observe that by (4.6.56),

$$\begin{aligned}\int_{\mathbb{B}^c} \|y\|^{1+\delta}\,\mu_{N,Y}(dy) &= NE\left[\left\|u_N^{-1}Y\right\|^{1+\delta} I_{\mathbb{B}^c}(u_N^{-1}Y)\right]\\ &= Nu_N^{-1-\delta}E_N(\delta),\end{aligned} \tag{4.6.58}$$

where

$$E_N(\delta) = E\left[\|Y\|^{1+\delta} I_{\{||Y||>u_N\}}\right].$$

By the Tonelli–Fubini theorem, we have

$$\begin{aligned}E_N(\delta) &= E\left(\int_{u_N^{1+\delta}} I_{\{\|Y\|^{1+\delta}>x\}}dx\right) = \int_{u_N^{1+\delta}}^{\infty} P\left(\|Y\|^{1+\delta} > x\right) dx\\ &= \int_{u_N^{1+\delta}}^{\infty} x^{-p/(1+\delta)} L(x^{1/(1+\delta)})dx.\end{aligned}$$

Now, by picking $\delta > 0$ such that $\eta := p/(1+\delta) > 1$ and applying the Karamata Theorem (Lemma 4.6.1(iii)), for the right-hand side of (4.6.58), we obtain

$$\begin{aligned}Nu_N^{-1-\delta}E_N(\delta) &\sim Nu_N^{-1-\delta}\frac{1}{\eta-1}\left(u_N^{1/(1+\delta)}\right)^{1-p/(1+\delta)} L(u_N)\\ &\sim \frac{1}{\eta-1}Nu_N^{-p}L(u_N) = \frac{1}{\eta-1}NP(\|Y\| > u_N) \to \frac{1}{\eta-1},\end{aligned}$$

where the last convergence follows from the definition of the sequence u_N. This shows that the supremum in (4.6.57) is finite, which completes the proof.

4.6.2 Proofs of the results of Section 4.4

Proof of Theorem 4.4.1

The results of this section require Assumptions 4.4.1 and 4.4.2.

Before stating Theorem 4.4.1, we referred to Lemma 4.6.2 which ensures that the the series

$$T_{j,N} = \sum_{k\neq j}(\lambda_j - \lambda_k)^{-1} \langle Z_N, v_j \otimes v_k\rangle v_k;$$
$$T_j = \sum_{k\neq j}(\lambda_j - \lambda_k)^{-1} \langle Z, v_j \otimes v_k\rangle v_k.$$

converge a.s. in L^2. These series play a fundamental role in our arguments.

LEMMA 4.6.2. *Suppose* $\Psi \in \mathcal{S}$. *For* $1 \leq j \leq p$, *set*

$$g_j(\Psi) = \sum_{k\neq j}(\lambda_j - \lambda_k)^{-1} \langle \Psi, v_j \otimes v_k\rangle v_k.$$

Then, the series defining $g_j(\Psi)$ *converges in* L^2.

Proof. Since the v_k are orthonormal, it is enough to check that

$$\sum_{k\neq j}(\lambda_j - \lambda_k)^{-2} \langle \Psi, v_j \otimes v_k\rangle^2 < \infty.$$

Since the system $\{v_j \otimes v_k,\ j,k \geq 1\}$ forms an orthonormal basis in $\mathcal{S}$

$$\sum_{j,k\geq 1} \langle \Psi, v_j \otimes v_k\rangle^2 = \|\Psi\|_{\mathcal{S}}^2 < \infty.$$

Therefore,

$$\sum_{k\neq j}(\lambda_j - \lambda_k)^{-2} \langle \Psi, v_j \otimes v_k\rangle^2 \leq \alpha_j^{-2} \|\Psi\|_{\mathcal{S}}^2\,,$$

withe α_j defined in (4.5.32). □

We will use the following lemma, which is analogous to Lemma 1 in [15], whose fully analogous proof, based on algebraic manipulations, is omitted.

LEMMA 4.6.3. *For any* $j \geq 1$*,*

$$\langle \hat{v}_j - v_j, v_j \rangle = -\frac{1}{2} \| \hat{v}_j - v_j \|^2 .$$

For any $j, k \geq 1$ *such that* $j \neq k$ *and* $\hat{\lambda}_j \neq \lambda_k$*,*

$$\langle \hat{v}_j - v_j, v_k \rangle = r_N^{-1} (\hat{\lambda}_j - \lambda_k)^{-1} \langle Z_N, \hat{v}_j \otimes v_k \rangle .$$

By Assumption 4.4.1, $\|\widehat{C} - C\|_{\mathcal{S}} = O_P(r_N^{-1})$. Using the well–known inequalities

$$|\hat{\lambda}_j - \lambda_j| \leq \|\widehat{C} - C\|_{\mathcal{S}}, \quad \|\hat{v}_j - v_j\| \leq \frac{2\sqrt{2}}{\alpha_j} \|\widehat{C} - C\|_{\mathcal{S}},$$

(see e.g. Lemmas 2.2 and 2.3 in [6]), we obtain the following Lemma.

LEMMA 4.6.4. *For* $1 \leq j \leq p$*,*

$$\|\widehat{C} - C\|_{\mathcal{S}} = O_P(r_N^{-1}), \quad |\hat{\lambda}_j - \lambda_j| = O_P(r_N^{-1}), \quad \|\hat{v}_j - v_j\| = O_P(r_N^{-1}).$$

LEMMA 4.6.5. *For* $1 \leq j \leq p$*,*

$$\| r_N (\hat{v}_j - v_j) - T_{j,N} \| = O_P \left(r_N^{-1} \right) .$$

Proof. The same arguments apply to any fixed $j \in \{1, 2, \ldots, p\}$, so to reduce the number of indexes used, we present them for $j = 1$. Set

$$d_{N,k} = \langle r_N (\hat{v}_1 - v_1) - T_{1,N}, v_k \rangle ,$$

where

$$T_{1,N} = \sum_{\ell \geq 2} (\lambda_1 - \lambda_\ell)^{-1} \langle Z_N, v_1 \otimes v_\ell \rangle v_\ell.$$

By Parseval's identity,

$$\|r_N(\hat{v}_j - v_j) - T_{j,N}\|^2 = \sum_{k=1}^{\infty} d_{N,k}^2.$$

Focusing on the first term, $k = 1$, observe that

$$\langle T_{1,N}, v_1 \rangle = \sum_{\ell \geq 2} (\lambda_1 - \lambda_\ell)^{-1} \langle Z_N, v_1 \otimes v_\ell \rangle \langle v_\ell, v_\ell \rangle = 0$$

and, by Lemmas 4.6.3 and 4.6.4,

$$\langle r_N(\hat{v}_1 - v_1), v_1 \rangle = -\frac{r_N}{2} \|\hat{v}_1 - v_1\|^2 = O_P(r_N^{-1}).$$

We conclude that $d_{N,1}^2 O_P(r_N^{-2})$, and it remain to show that

$$\sum_{k=2}^{\infty} d_{N,k}^2 = O_P(r_N^{-2}). \tag{4.6.59}$$

In the remainder of the proof it is assumed that $k \geq 2$. Since

$$\langle T_{1,N}, v_k \rangle = (\lambda_1 - \lambda_k)^{-1} \langle Z_N, v_1 \otimes v_k \rangle,$$

by Lemma 4.6.3,

$$d_{N,k} = (\hat{\lambda}_1 - \lambda_k)^{-1} \langle Z_N, \hat{v}_1 \otimes v_k \rangle - (\lambda_1 - \lambda_k)^{-1} \langle Z_N, v_1 \otimes v_k \rangle.$$

Using a common denominator and rearranging the numerator, we obtain

$$d_{N,k} = \frac{\left\langle (\lambda_1 - \lambda_k) Z_N(\hat{v}_1 - v_1) + (\lambda_1 - \hat{\lambda}_1) Z_N(v_1)\ ,\ v_k \right\rangle}{(\hat{\lambda}_1 - \lambda_k)^2 (\lambda_1 - \lambda_k)^2}.$$

It is convenient to decompose the sum in (4.6.59) as

$$\sum_{k=2}^{\infty} d_{N,k}^2 = D_{N,1} + D_{N,2} + D_{N,3},$$

where

$$D_{N,1} = \sum_{k\geq 2} \frac{\langle Z_N(\hat{v}_1 - v_1), v_k\rangle^2}{(\hat{\lambda}_1 - \lambda_k)^2},$$
$$D_{N,2} = \sum_{k\geq 2} \frac{2(\lambda_1 - \hat{\lambda}_1)\langle Z_N(\hat{v}_1 - v_1), v_k\rangle \langle Z_N(v_1), v_k\rangle}{(\hat{\lambda}_1 - \lambda_k)^2(\lambda_1 - \lambda_k)},$$
$$D_{N,3} = \sum_{k\geq 2} \frac{(\lambda_1 - \hat{\lambda}_1)^2 \langle Z_N(v_1), v_k\rangle^2}{(\hat{\lambda}_1 - \lambda_k)^2(\lambda_1 - \lambda_k)^2}.$$

Since $\hat{\lambda}_1 - \lambda_k \geq \hat{\lambda}_1 - \lambda_2$, by Parseval's identity,

$$D_{N,1} \leq \frac{1}{(\hat{\lambda}_1 - \lambda_2)^2} \sum_{k\geq 2} \langle Z_N(\hat{v}_1 - v_1), v_k\rangle^2 \leq \frac{\|Z_N(\hat{v}_1 - v_1)\|^2}{(\hat{\lambda}_1 - \lambda_2)^2}.$$

By Lemma 4.6.4, the denominator converges in probability to $(\lambda_1 - \lambda_2)^2$, and the numerator is bounded above by $\|Z_N\|^2\|(\hat{v}_1 - v_1)\|^2 = O_P(r_N^{-2})$.

A similar argument shows that

$$|D_{N,2}| \leq \left| \frac{2(\lambda_1 - \hat{\lambda}_1)}{(\hat{\lambda}_1 - \lambda_2)^2(\lambda_1 - \lambda_2)} \right| |\langle Z_N(\hat{v}_1 - v_1), Z_N(v_1)\rangle| .$$

The denominator again converges to a positive constant. By the Cauchy–Schwarz inequality,

$$|\langle Z_N(\hat{v}_1 - v_1), Z_N(v_1)\rangle| \leq \|Z_N(\hat{v}_1 - v_1)\|\|Z_N(v_1)\| \leq \|Z_N\|^2\|\hat{v}_1 - v_1\|.$$

We see that $D_{N,2} = O_P(r_N^{-2})$.

The above method also shows that $D_{N,3} = O_P(r_N^{-2})$. □

PROOF OF THEOREM 4.4.1: To prove the first relation, we use the decomposition

$$r_N(\hat{v}_j - v_j) = T_{j,N} + (r_n(\hat{v}_j - v_j) - T_{j,N})\,.$$

By Lemma 4.6.5, it suffices to show that the $T_{j,n}$ converge jointly in distribution to the T_j. Consider the operator $\mathbf{g} : \mathcal{S} \to (L^2)^p$ defined by

$$\mathbf{g}(\Psi) = [g_1(\Psi), g_2(\Psi), \ldots, g_p(\Psi)]^\top,$$

with the functions g_j defined in Lemma 4.6.2. The proof of Lemma 4.6.2 shows that $\|g_j(\Psi)\| \le \alpha_j^{-1}\|\Psi\|_{\mathcal{S}}$, so each g_j is a continuous linear operator. Hence $\mathbf{g}$ is continuous, and so $\mathbf{g}(Z_N) \xrightarrow{d} \mathbf{g}(Z)$. Since, $g_j(Z_N) = T_{j,N}$ and $g_j(Z) = T_j$, the required convergence follows.

Now we turn to the convergence of the eigenvalues. We will derive an analogous decomposition,

$$r_N(\hat{\lambda}_j - \lambda_j) = \langle Z_N(v_j), v_j\rangle + \beta_N(j), \tag{4.6.60}$$

and show that for each $j = 1, 2, \ldots, p$, $\beta_N(j) = O_P(r_N^{-1})$. Since the projections

$$\mathcal{S} \ni \Psi \mapsto \langle \Psi(v_j), v_j\rangle = \langle \Psi, v_j \otimes v_j\rangle_{\mathcal{S}}$$

are continuous, the claim will follow.

Observe that

$$\begin{aligned}
(\hat{\lambda}_j - \lambda_j)v_j &= \hat{\lambda}_j v_j - \hat{\lambda}_j \hat{v}_j + \hat{\lambda}_j \hat{v}_j - \lambda_j v_j \\
&= \hat{\lambda}_j(v_j - \hat{v}_j) + \widehat{C}(\hat{v}_j) - C(v_j) \\
&= (\widehat{C} - C)(\hat{v}_j) + C(\hat{v}_j - v_j) - \hat{\lambda}_j(\hat{v}_j - v_j).
\end{aligned}$$

It follows that

$$r_N(\hat{\lambda}_j - \lambda_j)v_j = Z_N(\hat{v}_j) + r_N\left\{C(\hat{v}_j - v_j) - \hat{\lambda}_j(\hat{v}_j - v_j)\right\}.$$

We decompose the first term as $Z_N(\hat{v}_j) = Z_N(v_j) + Z_N(\hat{v}_j - v_j)$ and get (4.6.60) with

$$\begin{aligned}
\beta_N(j) &= \langle Z_N(\hat{v}_j - v_j), v_j\rangle + r_N\left\langle C(\hat{v}_j - v_j) - \hat{\lambda}_j(\hat{v}_j - v_j), v_j\right\rangle \\
&= r_N\left\langle \left[(\widehat{C} - C) + C - \hat{\lambda}_j\right](\hat{v}_j - v_j), v_j\right\rangle \\
&= r_N\left\langle \left[(\widehat{C} - C) + (C - \lambda_j) - (\hat{\lambda}_j - \lambda_j)\right](\hat{v}_j - v_j), v_j\right\rangle
\end{aligned}$$

By Lemma 4.6.4,

$$\left\langle (\widehat{C} - C)(\hat{v}_j - v_j), v_j\right\rangle = O_P(r_N^{-2})$$

and

$$\left\langle (\hat{\lambda}_j - \lambda_j)(\hat{v}_j - v_j), v_j\right\rangle = O_P(r_N^{-2}).$$

Since C is symmetric

$$\langle (C - \lambda_j)(\hat{v}_j - v_j), v_j\rangle = \langle \hat{v}_j - v_j, (C - \lambda_j)(v_j)\rangle = 0.$$

This shows that $\beta_N(j) = O_P(r_N^{-1})$, and completes the proof.

Proof of Theorem 4.4.2

We start with a simple lemma, custom formulated for our needs.

LEMMA 4.6.6. *Suppose $\{X_n\}$ and $\{Y_n\}$ are sequences of nonnegative random variables and $\{a_n\}$ is a convergent sequence of nonnegative numbers. Suppose $X_n \leq Y_n + a_n$. If the Y_n are uniformly integrable, then so are the X_n.*

Proof. We will establish a more general result under the assumption that $C := \sup_{n\in\mathbb{N}} a_n < \infty$. Recall that a sequence $\{X_n\}$ is uniformly integrable if and only if the following two conditions hold *(i)* We have $\sup_{n\in\mathbb{N}} E|X_n| < \infty$.

(ii) For all $\epsilon > 0$, there exists a $\delta > 0$, such that

$$\sup_{n\in\mathbb{N}} E\left(|X_n| 1_A\right) < \epsilon,$$

for all events such that $P(A) < \delta$ (see, e.g., Theorem 6.5.1 on page 184 in [78]).

Since $\{Y_n\}$ is uniformly integrable, we have $\sup_{n\in\mathbb{N}} E|Y_n| < \infty$ and Condition (i) above follows from the triangle inequality and the boundedness of the sequence $\{a_n\}$. To show that Condition (ii) holds, observe that by the triangle inequality

$$\sup_{n\in\mathbb{N}} E\left(|X_n| 1_A\right) \leq \sup_{n\in\mathbb{N}} E\left(|Y_n| 1_A\right) + CP(A). \tag{4.6.61}$$

Using the uniform integrability of $\{Y_n\}$, for every $\epsilon > 0$, one can find $\delta' > 0$ such that the first term in the right-hand side of (4.6.61) is less than $\epsilon/2$, provided $P(A) < \delta'$. By setting $\delta := \min\{\delta', \epsilon/(2C)\}$, we also ensure that the second term therein is less than $\epsilon/2$ for all $P(A) < \delta \leq \delta'$. This completes the proof of the uniform integrability of $\{X_n\}$. □

In the following, we assume that γ is a fixed number in $(0, \alpha/2)$. Theorem 6.1 of [79] implies that, in the notation of Theorem 4.3.1, cf. (4.3.16),

$$\lim_{N\to\infty} E\left\|b_N^{-1}\left(\sum_{i=1}^{N} Z_i - \gamma_N\right)\right\|^{\gamma} = E\left\|S\right\|^{\gamma}.$$

Applying the above result to (4.3.23), we obtain

$$\lim_{N\to\infty} E\left\|S_N\right\|^{\gamma} = E\left\|S\right\|^{\gamma}, \tag{4.6.62}$$

where

$$S_N = k_N^{-1}\left(\sum_{i=1}^{N} X_i \otimes X_i - \psi_N\right).$$

In the framework of Theorem 4.3.3, set

$$M = \int_{\mathbb{B}_{\mathcal{S}}^c} y \mu_{X \otimes X}(dy)$$

and

$$M_N = k_N^{-1} N E \left[(X \otimes X) I_{\{\|X\|^2 \geq k_N\}} \right],$$

so that (4.6.49) becomes

$$N k_N^{-1} \left(\widehat{C} - C \right) = S_N - M_N$$

with $S_N \stackrel{d}{\to} S$ and $\|M_N - M\|_{\mathcal{S}} \to 0$. We now explain why we can conclude that

$$E \|S_N - M_N\|_{\mathcal{S}}^{\gamma} \to E \|S - M\|_{\mathcal{S}}^{\gamma}. \tag{4.6.63}$$

Since $S_N - M_N \stackrel{d}{\to} S - M$ in $\mathcal{S}$, $\|S_N - M_N\|_{\mathcal{S}}^{\gamma} \stackrel{d}{\to} \|S - M\|_{\mathcal{S}}^{\gamma}$ in $\mathbb{R}$. Convergence (4.6.63) will follow if we can assert that the nonnegative random variables $\|S_N - M_N\|_{\mathcal{S}}^{\gamma}$ are uniformly integrable. Since $\|S_N\|_{\mathcal{S}}^{\gamma} \stackrel{d}{\to} \|S\|_{\mathcal{S}}^{\gamma}$ and (4.6.62) holds, Theorem 3.6 in [80] implies that the random variables $\|S_N\|_{\mathcal{S}}^{\gamma}$ are uniformly integrable. Relation (4.6.63) thus follows from the inequality

$$\|S_N - M_N\|_{\mathcal{S}}^{\gamma} \leq C_{\gamma} \left\{ \|S_N\|_{\mathcal{S}}^{\gamma} + \|M_N\|_{\mathcal{S}}^{\gamma} \right\}$$

and Lemma 4.6.6. Relation (4.6.63) implies the first relation in Theorem 4.4.2 with $L_{\gamma}(N) = L_0^{-2\gamma}(N)$.

Since $|\hat{\lambda}_j - \lambda_j| \leq \|\widehat{C} - C\|_{\mathcal{S}}$ (see e.g. Lemma 2.2 in [6]), the second relation follows from the first. Under Assumption 4.4.2, $\|\hat{v}_j - v_j\| \leq a_j \|\widehat{C} - C\|_{\mathcal{S}}$ (see e.g. Lemma 2.3 in [6] or Lemma 4.3 in [5]), so the third relation also follows from the first.

4.6.3 Proof of Theorem 4.5.1

Since $\|\Psi_{KL} - \Psi\|_{\mathcal{L}} \to 0$ by (4.5.28) and (4.5.29), it is enough to show that

$$\left\| \widehat{\Psi}_{KL} - \Psi_{KL} \right\|_{\mathcal{L}} \stackrel{a.s.}{\to} 0. \tag{4.6.64}$$

The operators Ψ_{KL} and $\widehat{\Psi}_{KL}$ have the following expansions:

$$\widehat{\Psi}_{KL}(x) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \frac{\hat{\sigma}_{\ell k}}{\hat{\lambda}_\ell} \langle \hat{v}_\ell, x \rangle \hat{u}_k, \quad \Psi_{KL}(x) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \frac{\sigma_{\ell k}}{\lambda_\ell} \langle v_\ell, x \rangle u_k.$$

Introduce the sample analogs of the subspaces $\mathcal{V}_L$ and $\mathcal{U}_K$,

$$\widehat{\mathcal{V}}_L = \text{span}\{\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_L\}, \quad \widehat{\mathcal{U}}_K = \text{span}\{\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_K\},$$

and consider the following projections:

$$\pi^L = \text{projection onto } \mathcal{V}_L, \quad \hat{\pi}^L = \text{projection onto } \widehat{\mathcal{V}}_L;$$

$$\pi^K = \text{projection onto } \mathcal{U}_K, \quad \hat{\pi}^K = \text{projection onto } \widehat{\mathcal{U}}_K.$$

Observe that

$$\widehat{\Psi}_{KL} = \hat{\pi}^K D_N \widehat{C}^{-1} \hat{\pi}^L, \; \Psi_{KL} = \pi^K D C^{-1} \pi^L,$$

where

$$D = E\left[X \otimes Y\right], \; D_N = \frac{1}{N} \sum_{i=1}^{N} X_i \otimes Y_i,$$

and

$$C = \sum_{j=1}^{\infty} \lambda_j v_j \otimes v_j, \; \widehat{C} = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{v}_j \otimes \hat{v}_j, \; C^{-1} = \sum_{j=1}^{\infty} \lambda_j^{-1} v_j \otimes v_j, \; \widehat{C}^{-1} = \sum_{j=1}^{\infty} \hat{\lambda}_j^{-1} \hat{v}_j \otimes \hat{v}_j.$$

Notice that for any $y = \pi^L(x)$ or $y = \hat{\pi}^L(x)$, $C^{-1}(y)$ and $\widehat{C}^{-1}(y)$ exist.

For $x \in L^2$, consider the decomposition

$$\begin{aligned}
\left(\widehat{\Psi}_{KL} - \Psi_{KL}\right)(x) =& \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \hat{\lambda}_j^{-1} \langle \hat{v}_j, x \rangle \hat{v}_j \right) - \pi^K D \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle v_j, x \rangle v_j \right) \\
=& \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \left(\hat{\lambda}_j^{-1} - \lambda_j^{-1} \right) \langle \hat{v}_j, x \rangle \hat{v}_j \right) \\
&+ \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle \hat{v}_j - v_j, x \rangle \hat{v}_j \right) \\
&+ \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle v_j, x \rangle (\hat{v}_j - v_j) \right) \\
&+ \left(\hat{\pi}^K D_N - \pi^K D \right) \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle v_j, x \rangle v_j \right) \\
=:& \, a_N(x) + b_N(x) + c_N(x) + d_N(x),
\end{aligned}$$

where

$$\begin{aligned}
a_N(x) &= \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \left(\hat{\lambda}_j^{-1} - \lambda_j^{-1} \right) \langle \hat{v}_j, x \rangle \hat{v}_j \right), \\
b_N(x) &= \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle \hat{v}_j - v_j, x \rangle \hat{v}_j \right), \\
c_N(x) &= \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle v_j, x \rangle (\hat{v}_j - v_j) \right), \\
d_N(x) &= \left(\hat{\pi}^K D_N - \pi^K D \right) \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle v_j, x \rangle v_j \right).
\end{aligned}$$

Relation (4.6.64) will follow from Lemmas 4.6.8, 4.6.9, 4.6.10 and 4.6.13. The first two of these lemmas use the following result.

LEMMA 4.6.7. *Under the assumptions of Theorem 4.5.1,*

$$\left\| \hat{\pi}^K D_N(\hat{v}_j) \right\| \leq \hat{\lambda}_j^{1/2} \left(\frac{1}{N} \sum_{i=1}^{N} \|Y_i\|^2 \right)^{1/2}.$$

Proof. For each integer ℓ, we have

$$
\begin{aligned}
|\left\langle \hat{\pi}^K D_N(\hat{v}_j), \hat{u}_\ell \right\rangle| =& \left| \left\langle \sum_{k=1}^{K} \frac{1}{N} \sum_{i=1}^{N} \langle X_i, \hat{v}_j \rangle \langle Y_i, \hat{u}_k \rangle \hat{u}_k, \hat{u}_\ell \right\rangle \right| \\
=& \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, \hat{v}_j \rangle \langle Y_i, \hat{u}_\ell \rangle \right| \\
\leq& \frac{1}{N} \left(\sum_{i=1}^{N} \langle X_i, \hat{v}_j \rangle^2 \right)^{1/2} \left(\sum_{i=1}^{N} \langle Y_i, \hat{u}_\ell \rangle^2 \right)^{1/2} \\
=& \left(\left\langle \widehat{C}(\hat{v}_j), \hat{v}_j \right\rangle \right)^{1/2} \left(\left\langle \widehat{C}_Y(\hat{u}_\ell), \hat{u}_\ell \right\rangle \right)^{1/2} \\
=& \hat{\lambda}_j^{1/2} \hat{\gamma}_\ell^{1/2}.
\end{aligned}
$$

Therefore,

$$
\left\| \hat{\pi}^K D_N(\hat{v}_j) \right\| = \sum_{\ell=1}^{\infty} \left\langle \hat{\pi}^K D_N(\hat{v}_j), \hat{u}_\ell \right\rangle^2 \leq \hat{\lambda}_j \sum_{\ell=1}^{\infty} \hat{\gamma}_\ell,
$$

and

$$
\sum_{\ell=1}^{\infty} \hat{\gamma}_\ell = \sum_{\ell=1}^{\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \langle Y_i, \hat{u}_\ell \rangle^2 \right) = \frac{1}{N} \sum_{i=1}^{N} \|Y_i\|^2 .
$$

Hence the claim holds. □

LEMMA 4.6.8. *Under the assumptions of Theorem 4.5.1,* $\|a_N\|_{\mathcal{L}} \overset{a.s.}{\to} 0$.

Proof. Observe that

$$
\begin{aligned}
\|a_N(x)\| =& \left\| \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \left(\hat{\lambda}_j^{-1} - \lambda_j^{-1} \right) \langle \hat{v}_j, x \rangle \hat{v}_j \right) \right\| \\
\leq& \sum_{j=1}^{L} \frac{|\hat{\lambda}_j - \lambda_j|}{\hat{\lambda}_j \lambda_j} | \langle \hat{v}_j, x \rangle | \left\| \hat{\pi}^K D_N(\hat{v}_j) \right\| .
\end{aligned}
$$

By Lemma 4.6.7, Lemma 2.2 of [6] and the Cauchy-Schwarz inequality, we obtain the bound

$$
\begin{aligned}
\|a_N(x)\| \leq& \sum_{j=1}^{L} \lambda_j^{-1} \hat{\lambda}_j^{-1/2} | \langle \hat{v}_j, x\rangle | \left(\frac{1}{N} \sum_{i=1}^{N} \|Y_i\|^2 \right)^{1/2} \left\| \widehat{C} - C \right\|_{\mathcal{L}} \\
\leq& \lambda_L^{-1} \hat{\lambda}_L^{-1/2} \|x\| L^{1/2} \left(\frac{1}{N} \sum_{i=1}^{N} \|Y_i\|^2 \right)^{1/2} \left\| \widehat{C} - C \right\|_{\mathcal{L}}.
\end{aligned}
$$

By Corollary 4.3.1, for $N > N_1$ (random),

$$
\hat{\lambda}_L \geq \lambda_L - \left\| \widehat{C} - C \right\|_{\mathcal{L}} \geq \lambda_L / 2.
$$

Then we have

$$
\|a_N\|_{\mathcal{L}} \leq \sqrt{2} \left(\frac{1}{N} \sum_{i=1}^{N} \|Y_i\|^2 \right)^{1/2} \lambda_L^{-3/2} L^{1/2} \left\| \widehat{C} - C \right\|_{\mathcal{L}}.
$$

Corollary 4.3.1 implies that, for any $\gamma \in (1, \alpha/2)$, $N^{1-1/\gamma} \|\widehat{C} - C\|_{\mathcal{S}} \stackrel{a.s.}{\to} 0$, and by the strong law of large numbers

$$
\frac{1}{N} \sum_{i=1}^{N} \|Y_i\|^2 \stackrel{a.s.}{\to} E \|Y\|^2 \leq 2 \left(\|\Psi\|_{\mathcal{S}}^2 E \|X\|^2 + E \|\varepsilon\|^2 \right) < \infty.
$$

The claim thus follows from condition (4.5.34). □

LEMMA 4.6.9. *Under the assumptions of Theorem 4.5.1,* $\|b_N\|_{\mathcal{L}} \stackrel{a.s.}{\to} 0$.

Proof. Lemma 4.6.7 implies that

$$
\begin{aligned}
\|b_N(x)\| =& \left\| \hat{\pi}^K D_N \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle \hat{v}_j - v_j, x \rangle \hat{v}_j \right) \right\| \\
\leq& \sum_{j=1}^{L} \lambda_j^{-1} | \langle \hat{v}_j - v_j, x \rangle | \left\| \hat{\pi}^K D_N(\hat{v}_j) \right\| \\
\leq& \sum_{j=1}^{L} \lambda_j^{-1} \hat{\lambda}_j^{1/2} \|x\| \|\hat{v}_j - v_j\| \left(\frac{1}{N} \sum_{i=1}^{N} \|Y_i\|^2 \right)^{1/2}.
\end{aligned}
$$

Lemma 2.3 of [6] yields the relation

$$\|\hat{v}_j - v_j\| \le 2\sqrt{2}\alpha_j^{-1} \left\|\widehat{C} - C\right\|_{\mathcal{L}},$$

with the α_i defined in (4.5.32). Hence,

$$\|b_N\|_{\mathcal{L}} \le 2\sqrt{2}\lambda_L^{-1}\hat{\lambda}_1^{1/2} \left(\sum_{j=1}^{L} \alpha_j^{-1}\right) \left(\frac{1}{N}\sum_{i=1}^{N} \|Y_i\|^2\right)^{1/2} \left\|\widehat{C} - C\right\|_{\mathcal{L}}.$$

Since, for $N > N_2$ (random),

$$\hat{\lambda}_1 \le \lambda_1 + \left\|\widehat{C} - C\right\|_{\mathcal{L}} \le \frac{3}{2}\lambda_1,$$

we have

$$\|b_N\|_{\mathcal{L}} \le 2\sqrt{3}\lambda_L^{-1}\lambda_1^{1/2} \left(\sum_{j=1}^{L} \alpha_j^{-1}\right) \left(\frac{1}{N}\sum_{i=1}^{N} \|Y_i\|^2\right)^{1/2} \left\|\widehat{C} - C\right\|_{\mathcal{L}}.$$

By Corollary 4.3.1 and the strong law of large numbers, the claim follows from (4.5.35). □

LEMMA 4.6.10. *Under the assumptions of Theorem 4.5.1,* $\|c_N\|_{\mathcal{L}} \overset{a.s.}{\to} 0$.

Proof. Observe that

$$\begin{aligned}
\|c_N(x)\| &= \left\|\hat{\pi}^K D_N \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle v_j, x\rangle (\hat{v}_j - v_j)\right)\right\| \\
&\le \left\|\hat{\pi}^K D_N\right\|_{\mathcal{L}} \sum_{j=1}^{L} \lambda_j^{-1} |\langle v_j, x\rangle| \, \|\hat{v}_j - v_j\| \\
&\le \left\|\hat{\pi}^K D_N\right\|_{\mathcal{L}} \left(\sum_{j=1}^{L} \lambda_j^{-1}\alpha_j^{-1} |\langle v_j, x\rangle|\right) \left\|\widehat{C} - C\right\|_{\mathcal{L}} \\
&\le \left\|\hat{\pi}^K D_N\right\|_{\mathcal{L}} \lambda_L^{-1} \|x\| \left(\sum_{j=1}^{L} \alpha_j^{-1}\right) \left\|\widehat{C} - C\right\|_{\mathcal{L}}.
\end{aligned}$$

Therefore,

$$\|c_N\|_{\mathcal{L}} \le \|D_N\|_{\mathcal{L}} \lambda_L^{-1} \left(\sum_{j=1}^{L} \alpha_j^{-1}\right) \left\|\widehat{C} - C\right\|_{\mathcal{L}}.$$

Since, by the law of large numbers, $\left\|\hat{\pi}^K D_N\right\|_{\mathcal{L}} \stackrel{a.s.}{\to} \|D\|_{\mathcal{L}}$, the claim follows from condition (4.5.35). □

To deal with the last term, we need additional lemmas.

LEMMA 4.6.11. *Under the assumptions of Theorem 4.5.1,* $N^{1-1/\gamma}\|D_N - D\|_{\mathcal{S}} \stackrel{a.s.}{\to} 0$.

Proof. The decomposition

$$\frac{1}{N}\sum_{i=1}^{N} X_i \otimes Y_i = \frac{1}{N}\sum_{i=1}^{N} X_i \otimes \Psi(X_i) + \frac{1}{N}\sum_{i=1}^{N} X_i \otimes \varepsilon_i$$

and the identities

$$X_i \otimes \Psi(X_i) = \Psi(X_i \otimes X_i), \;\; E[X \otimes \Psi(X)] = \Psi E[X \otimes X], \;\; E[X \otimes \varepsilon] = 0$$

imply that

$$\begin{aligned} \|D_N - D\|_{\mathcal{S}} &= \left\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes Y_i - E[X \otimes Y] \right\|_{\mathcal{S}} \\ &\le \|\Psi\|_{\mathcal{S}} \left\| \widehat{C} - C \right\|_{\mathcal{S}} + \left\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes \varepsilon_i \right\|_{\mathcal{S}}. \end{aligned}$$

For any $1 \le \gamma < 2$,

$$\left\| \frac{1}{N^{1/\gamma}}\sum_{i=1}^{N} X_i \otimes \varepsilon_i \right\|_{\mathcal{S}} \stackrel{a.s.}{\to} 0.$$

The above convergence follows from Theorem 4.1 of [74] which implies that in any separable Banach space of Rademacher type γ, $1 \le \gamma < 2$, $N^{-1/\gamma}\sum_{i=1}^{N} Y_i \stackrel{a.s.}{\to} 0$, provided the Y_i are iid with $E\|Y_i\|^\gamma < \infty$ and $EY_i = 0$. In our case, the Banach space is the Hilbert space $\mathcal{S}$ (a Hilbert space has Rademacher type γ for any $\gamma \le 2$, see e.g. Theorems 3.5.2 and 3.5.7 of [75]). Clearly, $E[X_i \otimes \varepsilon_i] = 0$ and $E\|X_i \otimes \varepsilon_i\|_{\mathcal{S}}^\gamma = E\|X_i\|^\gamma E\|\varepsilon_i\|^\gamma < \infty$. Another application of Corollary 4.3.1 completes the proof. □

LEMMA 4.6.12. *Under the assumptions of Theorem 4.5.1,* $\lambda_L^{-1} \left\| \hat{\pi}^K D_N - \pi^K D \right\|_{\mathcal{L}} \stackrel{a.s.}{\to} 0.$

Proof. By the triangle inequality,

$$\left\| \hat{\pi}^K D_N - \pi^K D \right\|_{\mathcal{L}} \leq \left\| \hat{\pi}^K D_N - \hat{\pi}^K D \right\|_{\mathcal{L}} + \left\| \hat{\pi}^K D - \pi^K D \right\|_{\mathcal{L}}.$$

For the first term, we have

$$\begin{aligned} \left\| \hat{\pi}^K D_N - \hat{\pi}^K D \right\|_{\mathcal{L}} &= \sup_{||x|| \leq 1} \left\| \sum_{k=1}^{K} \langle (D_N - D)(x), \hat{u}_k \rangle \hat{u}_k \right\| \\ &\leq \sup_{||x|| \leq 1} \left(\sum_{k=1}^{K} \left| \langle (D_N - D)(x), \hat{u}_k \rangle \right| \right) \\ &\leq K^{1/2} \left\| D_N - D \right\|_{\mathcal{L}}. \end{aligned}$$

Thus, $\lambda_L^{-1} \left\| \hat{\pi}^K D_N - \hat{\pi}^K D \right\|_{\mathcal{L}} \stackrel{a.s.}{\to} 0$ by Lemma 4.6.11 and condition (4.5.36).

Turning to the second term, observe first that

$$D(x) = E[\langle X, x \rangle Y] = \Psi(E[\langle X, x \rangle X]) = \Psi(C(x)).$$

Setting $y = \Psi(C(x))$ we thus have

$$\pi^K D(x) = \sum_{k=1}^{K} \langle y, u_k \rangle u_k, \quad \hat{\pi}^K D(x) = \sum_{k=1}^{K} \langle y, \hat{u}_k \rangle \hat{u}_k.$$

Consequently, $\hat{\pi}^K D(x) - \pi^K D(x) = D_1(x) + D_2(x)$, where

$$D_1(x) = \sum_{k=1}^{K} \langle y, u_k - \hat{u}_k \rangle u_k, \quad D_2(x) = \sum_{k=1}^{K} \langle y, \hat{u}_k \rangle (u_k - \hat{u}_k).$$

Next,

$$\|D_1(x)\| \leq \|y\| \left\{ \sum_{k=1}^{K} \|u_k - \hat{u}_k\|^2 \right\}^{1/2} \leq 2\sqrt{2}\|y\| \left\| \widehat{C}_Y - C_Y \right\|_{\mathcal{L}} \left\{ \sum_{k=1}^{K} \frac{1}{\beta_k^2} \right\}^{1/2}$$

and

$$\|D_2(x)\| \leq \sum_{k=1}^{K} |\langle y, \hat{u}_k \rangle| \|u_k - \hat{u}_k\| \leq 2\sqrt{2}\|y\| \left\| \widehat{C}_Y - C_Y \right\|_{\mathcal{L}} \sum_{k=1}^{K} \frac{1}{\beta_k}.$$

We see that condition (4.5.37) implies that $\lambda_L^{-1} \left\| \hat{\pi}^K D - \hat{\pi}^K D \right\|_{\mathcal{L}} \overset{a.s.}{\to} 0$. □

LEMMA 4.6.13. *Under the assumptions of Theorem 4.5.1,* $\|d_N\|_{\mathcal{L}} \overset{a.s.}{\to} 0$.

Proof. Observe that

$$\begin{aligned}
\|d_N(x)\|^2 &= \left\| \left(\hat{\pi}^K D_N - \pi^K D\right) \left(\sum_{j=1}^{L} \lambda_j^{-1} \langle v_j, x \rangle v_j \right) \right\|^2 \\
&\leq \left\| \hat{\pi}^K D_N - \pi^K D \right\|_{\mathcal{L}}^2 \left(\sum_{j=1}^{L} \lambda_j^{-2} \langle v_j, x \rangle^2 \right) \\
&\leq \left\| \hat{\pi}^K D_N - \pi^K D \right\|_{\mathcal{L}}^2 \lambda_L^{-2} \left(\sum_{j=1}^{L} \langle v_j, x \rangle^2 \right) \\
&\leq \left\| \hat{\pi}^K D_N - \pi^K D \right\|_{\mathcal{L}}^2 \lambda_L^{-2} \|x\|^2 .
\end{aligned}$$

Consequently, $\|d_N\|_{\mathcal{L}} \leq \left\| \hat{\pi}^K D_N - \pi^K D \right\|_{\mathcal{L}} \lambda_L^{-1}$, so the claim follows from Lemma 4.6.12 and condition (4.5.34). □

Bibliography

[1] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2005.

[2] T. Hsing and R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, 2015.

[3] J. Cederbaum, M. Pouplier, P. Hoole, and S. Greven. Functional linear mixed models for irregularly or sparsely sampled data. *Statistical Modelling*, 16(1):67–88, 2016.

[4] R. J. Hyndman and M.S. Ullah. Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956, 2007.

[5] D. Bosq. *Linear Processes in Function Spaces*. Springer, 2000.

[6] L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, 2012.

[7] J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for principal component analysis of a vector random function. *Journal of Multivariate Analysis*, 12:136–154, 1982.

[8] S. Hörmann and P. Kokoszka. Weakly dependent functional data. *The Annals of Statistics*, 38:1845–1884, 2010.

[9] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987.

[10] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, 1997.

[11] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2006.

[12] L. de Haan and A. Ferreira. *Extreme Value Theory: an Introduction*. Springer, 2006.

[13] R.A. Fisher and L.H.C. Tippett. *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, volume 24. Mathematical Proceedings of the Cambridge Philosophical Society, 1928.

[14] A. Mas. Weak convergence for the covariance operators of a hilbertian linear process. *Stochastic processes and their applications*, 99(1):117–135, 2002.

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

[16] J. Kiefer. K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests. *The Annals of Mathematical Statistics*, pages 420–447, 1959.

[17] M.R. Leadbetter. On extreme values in stationary sequences. *Probability theory and related fields*, 28(4):289–303, 1974.

[18] H. Rootzén. The rate of convergence of extremes of stationary normal sequences. *Advances in Applied Probability*, 15:54–80, 1983.

[19] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):121–140, 2006.

[20] T. Meiguet. *Heavy Tailed Functional Time Series*. Universite catholigue de Louvain, 2010.

[21] T. Meinguet and J. Segers. Regularly varying time series in Banach spaces. Technical report, Universite catholigue de Louvain, 2010. arXiv:1001.3262.

[22] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, 2006.

[23] I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society (B)*, 71:927–946, 2009.

[24] R. L. Smith. Extreme value analysis of environmental time series: An application to trend detection in ground–level ozone. *Statistical Science*, 4:367–377, 1989.

[25] G. Dierckx and J. L. Teugels. Change point analysis of extreme values. *Environmetrics*, 21:661–686, 2010.

[26] W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847, 1987.

[27] B. Efron. Regression percentiles using asymmetric squared loss. *Statistica Sinica*, 1:93–125, 1991.

[28] G. De Rossi and A. Harwey. Quantiles, expectiles and splines. *Journal of Econometrics*, 152:179–185, 2009.

[29] S. K. Schnabel and P. H. C. Eilers. Optimal expectile smoothing. *Computational Statistics & Data Analysis*, 53:4168–4177, 2009.

[30] P. H. C. Eilers and B. D. Marx. Flexible smoothing with B–splines and penalties. *Statistical Science*, 11:89–121, 1996.

[31] M. Guo and W. K. Härdle. Simultaneous confidence bands for expectile functions. *AStA Advances in Statistical Analysis*, 96:517–541, 2012.

[32] F. Sobotka, G. Kauermann, L. Schulze-Waltrup, and T. Kneib. On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, 23:135–148, 2013.

[33] M. Guo, L. Zhou, J. Z. Huang, and W. K. Härdle. Functional data analysis of generalized regression quantiles. *Statistics and Computing*, 25:189–202, 2015.

[34] J. Taylor. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6:231–252, 2008.

[35] F. Sobotka, S. Schnabel, L. Schulz-Waltrup, P. Eilers, T. Kneib, and G. Kauermann. *R package: expectreg*, 2014. Version: 0.39, published: March 05, 2014.

[36] S. K. Schnabel. *Expectile smoothing: new perspectives on asymmetric least squares. An application to life expectancy*. PhD thesis, Utrecht University, 2011.

[37] S. K. Schnabel and P. H. C. Eilers. Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, 97:77–87, 2013.

[38] B. E. Brodsky and B. S. Darkhovsky. *Nonparametric Methods in Change–Point Problems*. Kluwer, 1993.

[39] M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Wiley, 1997.

[40] J. Chen and A. K. Gupta. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Birkhäuser, 2011.

[41] J. P. Kossin, T. L. Olander, and K. R. Knapp. Trend analysis with a new global record of tropical cyclone intensity. *Journal of Climate*, 26:9960–9976, 2013.

[42] R. Fraiman, A. Justel, R. Liu, and P. Llop. Detecting trends in time series of functional data: A study of antarctic climate change. *The Canadian Journal of Statistics*, 42:597–609, 2014.

[43] O. Gromenko and P. Kokoszka. Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination. *Computational Statistics and Data Analysis*, 59:82–94, 2013.

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

[45] S. Hörmann, L. Kidziński, and M. Hallin. Dynamic functional principal components. *Journal of the Royal Statistical Society (B)*, 77:319–348, 2015.

[46] L. Torgovitsky. *Hilbert space valued signal plus noise models: Analysis of structural breaks under high dimensionality and temporal dependence*. PhD thesis, Universität zu Köln, 2016.

[47] UNISYS. Data in atlantic and west pacific, 2015. Unisys Weather Information Systems, http://weather.unisys.com/hurricane/index.php, Accessed: February 20, 2015.

[48] I. Berkes, L. Horváth, and G. Rice. Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications*, 123:385–403, 2013.

[49] P. Kokoszka and M. Reimherr. *Introduction to Functional Data Analysis*. CRC Press, 2017.

[50] T. Hsing, J. Hüsler, and R-D. Reiss. The extremes of a triangular array of normal random variables. *The Annals of Applied Probability*, 6:671–686, 1996.

[51] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37:4046–4087, 2009.

[52] X. Zhang. White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics*, 194:76–95, 2016.

[53] W. B. Wu. Strong invariance principles for dependent random variables. *The Annals of Probability*, 35:2294–2320, 2007.

[54] X. Shao and W. B. Wu. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35:1773–1801, 2007.

[55] M. R. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer, 1983.

[56] N. N. Vakhaniia, V. I. Tarieladze, and S. A. Chobanian. *Probability Distributions on Banach Spaces*. Springer, 1987.

[57] Ph. Barbe and W. P. McCormick. Second-order expansion for the maximum of some stationary Gaussian sequences. *Stochastic Processes and their Applications*, 110:315–342, 2004.

[58] R. A. Davis and S. I. Resnick. Tail estimates motivated by extreme value theory. *The Annals of Statistics*, 12:1467–1487, 1984.

[59] P. Hall and M. Hosseini-Nasab. On properties of functional principal components. *Journal of the Royal Statistical Society (B)*, 68:109–126, 2006.

[60] R. A. Davis and S. I. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, 13(1):179–195, 1985.

[61] R. A. Davis and S. I. Resnick. Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics*, 14(2):533–558, 1986.

[62] C. Klüppelberg and T. Mikosch. Some limit theory for the self-normalised periodogram of stable processes. *Scandinavian Journal of Statistics*, 21:485–491, 1994.

[63] T. Mikosch, T. Gadrich, C. Klüppelrberg, and R. Adler. Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics*, 23:305–326, 1995.

[64] P. Kokoszka and M. Taqqu. Parameter estimation for infinite variance fractional ARIMA. *The Annals of Statistics*, 24:1880–1913, 1996.

[65] P. L. Anderson and M. M. Meerschaert. Periodic moving averages of random variables with regularly varying tails. *Annals of Statistics*, 24:771–785, 1997.

[66] D. O. Lucca and E. Moench. The pre-FOMC announcement drift. *The Journal of Finance*, 70:329–371, 2015.

[67] L. Horváth, P. Kokoszka, and G. Rice. Testing stationarity of functional time series. *Journal of Econometrics*, 179:66–82, 2014.

[68] S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, 2006.

[69] M. M. Meerschaert and H-P. Scheffler. *Limit distributions for sums of independent random vectors*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 2001. Heavy tails in theory and practice.

[70] M. M. Meerschaert. *Multivariate Domains of Attraction and Regular Variation*. PhD thesis, University of Michigan, Ann Arbor, 1984.

[71] J. Kuelbs and V. Mandrekar. Domains of attraction of stable measures on a Hilbert space. *Studia Mathematica*, 50:149–162, 1974.

[72] E. L. Rvačeva. On domains of attraction of multi-dimensional distributions. In *Select. Transl. Math. Statist. and Probability, Vol. 2*, pages 183–205. American Mathematical Society, 1962.

[73] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, 1994.

[74] A. de Acosta. Inequalities for B–valued random vectors with applications to the strong law of large numbers. *The Annals of Probability*, 9:157–161, 1981.

[75] W. Linde. *Probability in Banach Spaces - Stable and Infinitely Divisible Distributions*. Wiley, 1986.

[76] F. Yao, H-G. Müller, and J-L. Wang. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33:2873–2903, 2005b.

[77] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

[78] S. I. Resnick. *A Probability Path*. Birkhäuser, 1999.

[79] A. de Acosta and E. Giné. Convergence of moments and related functionals in the general central limit theorem in Banach spaces. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 48:213–231, 1979.

[80] P. Billingsley. *Convergence of Probability Measures; Second Edition*. Wiley, New York, 1999.

[81] E. J. Gumbel. *Statistics of Extremes*. Courier Corporation, 2012.

Appendix A

Elaboration on condition (3.2.10)

The normalizing constants $a_{j,N}$ defined in (3.2.9) satisfy condition (3.2.10) for normal, exponential, and any gamma distribution, and for any $\kappa > 0$. In the following we drop the subscript j. Condition (3.2.10) is thus equivalent to

$$N^{1-\kappa} F'(b_N) \to 0, \quad N \to \infty, \tag{A.0.1}$$

cf. [12], p. 18. Recall that $N = 1/\left(1 - F(b_N)\right)$.

For the normal $N(\mu, \sigma^2)$ distribution, the b_N satisfy

$$\int_{b_N}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{N}.$$

Example 1.1.7 in [12] implies that

$$b_N = \sigma \tilde{b}_N + \mu \sim \sigma \left(2 \log N\right)^{1/2},$$

where $\tilde{b}_N \sim \left(2 \log N\right)^{1/2}$ is the normalizing constants for the standard normal distribution. Observe that for any $\kappa > 0$, as $N \to \infty$,

$$N^{1-\kappa} F'(b_N) = N^{1-\kappa} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(b_N-\mu)^2/2\sigma^2} \sim \frac{1}{\sqrt{2\pi\sigma^2}} N^{-\kappa} \to 0.$$

Now consider the gamma distribution with the density function

$$F'(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \ \alpha > 0, \beta > 0.$$

For the reduced gamma distribution $\Gamma(\alpha, 1)$, the normalizing constants $\tilde{b}_N$ satisfy

$$\int_{\tilde{b}_N}^{\infty} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx = \frac{1}{N}.$$

By the calculation on p. 145 of [81], $\tilde{b}_N \sim \log\left(N/\Gamma(\alpha)\right)$. For the $\Gamma(\alpha, \beta)$ distribution, the b_N satisfy

$$\int_{b_N}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{1}{N}.$$

Therefore,

$$b_N = \frac{1}{\beta} \tilde{b}_N \sim \frac{1}{\beta} \log\left(N/\Gamma(\alpha)\right).$$

Observe that for any $\kappa > 0$, as $N \to \infty$,

$$N^{1-\kappa} F'(b_N) = N^{1-\kappa} \frac{\beta^\alpha}{\Gamma(\alpha)} b_N^{\alpha-1} e^{-\beta b_N} \sim N^{-\kappa} \beta \left(\log\left(\frac{N}{\Gamma(\alpha)} \right) \right)^{\alpha-1} \to 0.$$