Inference on linear processes in Hilbert and Banach spaces. Statistical analysis of high-dimensional data

A DISSERTATION PRESENTED BY

JAVIER ÁLVAREZ LIÉBANA

AND SUPERVISED BY

M. DOLORES RUIZ-MEDINA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

PHILOSOPHIÆ DOCTOR (PH. D.)

IN THE DOCTORAL PROGRAMME OF APPLIED AND MATHEMATICAL STATISTICS

UNIVERSIDAD DE GRANADA

DEPARTMENT OF STATISTICS AND O.R.. UNIVERSITY OF GRANADA. GRANADA (SPAIN)

15TH JUNE 2018

Editor: Universidad de Granada. Tesis Doctorales
Autor: Javier Álvarez Liébana
ISBN: 978-84-9163-952-7
URI: http://hdl.handle.net/10481/52557

This dissertation has been supervised by my advisor M. Dolores Ruiz Medina, full professor at University of Granada.

TRIBUNAL THESIS

Jose Miguel Ángulo Ibáñez (President). University of Granada.
Ana María Aguilera del Pino (Registrar). University of Granada.
Wenceslao González Manteiga (external referee). University of Santiago de Compostela.
María Dolores Ugarte Martínez (external referee). Public University of Navarre.
Florence Merlevède (international referee). University of Paris-Est Marne-la-Vallée (France).

El doctorando Javier Álvarez Liébana, y la directora de la tesis M. Dolores Ruiz Medina, garantizamos, al firmar la presente tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de la directora de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

The doctoral candidate Javier Álvarez Liébana, and the thesis advisor M. Dolores Ruiz Medina, guarantee, by signing this dissertation, that the work has been done by the doctoral candidate under the direction and supervision of the thesis advisor and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Directora de tesis / Thesis advisor

Doctorando / Doctoral candidate

Lugar y fecha / Place and date: Granada, 15th June 2018

Mathematics knows no races or geographic boundaries; for mathematics, the cultural world is one country

David Hilbert (23rd January 1862 – 14th February 1943)

Contents

Acknowledgments

I would strongly like to thank my advisor, the full professor M. Dolores Ruiz Medina, for her tireless and invaluable effort. This dissertation has been possible thanks to her support and experience. I look forward to continuing to work together since it has been a great pleasure during this period. This dissertation has been supported in part by projects MTM2012-32674 and MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

I should also like to thank the full professor Denis Bosq, for welcoming me during my two research stays at the University Paris VI (Paris, France), and to recognize, for their responsiveness filling the reports, the external referees Sophie Dabo–Niang (Lille University), Jeff Goldsmith (Columbia University) and Ingrid Van Keilegom (KU Leuven University). It has been a delight to belong to the Department of Statistics and O.R., at the University of Granada.

Thanks to Sandra, family and friends. I am very grateful to all those colleagues who have contributed to this research experience.

Los nadies: los hijos de nadie, los dueños de nada. Los nadies: los ningunos, los ninguneados. Que no son, aunque sean. Que no hablan idiomas, sino dialectos. Que no son seres humanos, sino recursos humanos. Los nadies, que cuestan menos que la bala que los mata.

E. Galeano (3rd September 1940 - 13th April 2015)

Agradecimientos

Decía Gabriel Celaya que "la poesía es un arma cargada de futuro", como lo es la ciencia de un país, tan maltratada en el nuestro, un patrimonio que debemos proteger ya que, aunque no nos prometa la felicidad como dijo Emilé Zola, es la única que nos "ha prometido la verdad". Esta tesis va dedicada, en primer, lugar a todos ellos, a los/as nadies. A cada investigador/a predoctoral que hace de la ciencia un lugar mejor. A todos los maestros, a todas las maestras que alumbran una luz en la mente de un niño. A Andrés Cassinello, mi profesor de matemáticas en un pequeño instituto de Carabanchel: porque fueron, somos.

Quiero agradecer enormemente al Departamento de Estadística e Investigación Operativa de la Universidad de Granada por el trato durante estos casi 4 años. Gracias a mis compañeros/as Xu Meng, Esquivel, David, José Luis, Antonio, Irene, Doris y Ramón. Por su amistad durante nuestro doctorado, gracias a la ya doctora Beatriz Cobo, por los cafés tomados y las confidencias compartidas. Dicen que uno es lo que sus experiencias le han dado. Y yo no sería lo que soy sin mi paso por Bologna ni por la Universidad Complutense. A Rocío, Aída, Ari, Laurita, Marta. A Estefanía, a Gabriel Valverde, a Lucía, a Diana y a Paloma. Gracias a ti, Laura, por lo vivido, por empezar juntos la carrera y por seguir ahí.

La ciencia es siempre un trabajo en el que debemos apoyarnos en "hombros de gigantes" para poder vislumbrar lo que hay al otro lado. Pero no siempre esos gigantes son viejas glorias o afamados investigadores. La mayoría de las veces son personas con las que te vas cruzando de congreso en congreso y vas forjando un trato de cariño y, en algunos casos, una amistad. Esta tesis también va dedicada a vosotros/as, a los/as que hacéis de cada evento una experiencia. Gracias a toda la gente con la que me he cruzado, y en especial a los participantes del SYSORM y la SEIO de Asturias. Muchísimas gracias por los momentos compartidos a Beatriz Bueno y a Yolanda. Gracias a Paula por esa manera de ser y a ti Jenifer, mi gallega favorita. Eternamente agradecido a Sonia por su ayuda y su sonrisa.

Esta tesis no solo le debe mucho a mis orígenes (a Sara, a Mauricio, a Andrea, a Ana, a Rebeca), sino también a quienes han compartido amistad conmigo en Granada. Gracias a Ana, a Lucía, a Berta, a Arturo y a Marina. Gracias a mis amigas de teatro en Fuenlabrada, las que tanta comedia y drama me han aportado. Gracias a Mónica. A veces los recién llegados son los que más te desgarran el corazón cuando les ves marchar. Gracias Mar, por mantener viva la revolución. Gracias a mis "conos", por todo lo reído, por todo lo llorado, por hacer que mi estancia en París fuera la mejor posible. Esta tesis es vuestra (incluso tuya, Silvia): Alba Martínez, Miguel López-Unzu, Miguel Montiel, Alba Fernández, Elena, Álvaro, Oriol, Carlos, Juanan, Andrea, Ismael y Pere. Gràcies Núria, per estar sense demanar.

Gracias infinitas (no numerables) a Lola por estos 4 años: hubiera sido incapaz de publicar un solo artículo sin tu trabajo, sin tu ayuda, sin tu esfuerzo día y noche para que esta tesis pudiera salir adelante. Gracias por las horas empleadas, por la paciencia, por tu rigor y por la confianza. Eres una persona y una investigadora admirable. Ojalá algún día pueda llegar a donde tú lo has hecho.

Gracias a mi familia por darme todo, por criarme y por estar siempre ahí, de forma incondicional. Esta tesis va dedicada a mis abuelos, a los que aún tengo (Pepa y Paco) y a los que ya se fueron (Tina y Goyo). Estoy donde estoy porque me disteis todo y siempre me habéis apoyado en cada paso, incluso el de estudiar Matemáticas siendo los dos de letras puras. Por ello, porque lo que tengo es gracias a vosotros, gracias a mis padres Paloma y Goyo. Porque la mitad de este triunfo os corresponde. Gracias a ti, María, por ser mi hermana y mi compañera en tantos momentos de la vida, por sacarnos una sonrisa. Gracias por seguir el camino marcado en las Matemáticas y no tengo duda de que serás la mejor profesora, porque siempre has podido y siempre podrás.

Por aguantarme, por sostenerme, por tu cariño, por tus abrazos, por ser mi compañera de vida tantos años, por los viajes, por las risas, por los lloros, por quererme, por nuestra historia. Gracias Sandra.

List of figures

List of tables

RESUMEN

Esta tesis proporciona nuevos resultados en el contexto de la estimación y predicción funcional, a partir de modelos autorregresivos Hilbertianos, o bien, con valores en espacios de Banach separables. El objetivo fundamental es proporcionar herramientas adecuadas para modelizar relaciones lineales entre variables aleatorias funcionales, que dependen de un índice temporal. Se ha adoptado un enfoque paramétrico, en la estimación funcional, basado en proyectar sobre bases ortonormales adecuadas. Los resultados derivados, sobre propiedades asintóticas de los estimadores considerados, se aplican al contexto de la regresión lineal funcional, con errores correlados en el tiempo, y con valores funcionales en espacios de Hilbert separables. En particular, se considera un análisis funcional de la varianza para dichos modelos. Adicionalmente, se introduce un enfoque Bayesiano en la derivación de la aproximación considerada, componente a componente, para el operador de autocorrelación, bajo condiciones menos restrictivas. El enfoque no paramétrico se contempla en la clasificación de datos funcionales con soporte espacial. Las contribuciones de esta tesis se pueden resumir, fundamentalmente, en los siguientes puntos:

- La derivación de nuevos resultados sobre consistencia débil y fuerte de estimadores de proyección del operador de autocorrelación, en modelos autorregresivos Hilbertianos de orden 1 (modelos $\mathrm{ARH}(1)$), respecto a diferentes normas, tales como la norma definida sobre el espacio de operadores lineales y acotados, la norma en el espacio de operadores de Hilbert–Schmidt y la norma para operadores traza. Bajo el mismo escenario, se obtiene la consistencia del correspondiente predictor funcional plug–in. Se considera, en esta derivación, el caso de autovectores conocidos y desconocidos. Como caso especial, se aborda el problema de predicción funcional del proceso conocido como Ornstein–Uhlenbeck, con valores en espacios de Hilbert y Banach separables. Este aspecto motiva el siguiente bloque de contribuciones.

- La extensión de los resultados derivados previamente en el contexto $\mathrm{ARH}(1)$ al contexto $\mathrm{ARB}(1)$,

siendo B un espacio de Banach abstracto y separable. Esta extensión también proporciona una metodología más flexible en el contexto de los procesos autorregresivos funcionales, dado que, hasta el momento, los espacios considerados por excelencia, en este ámbito, han sido los espacios de funciones continuas sobre un intervalo acotado, dotados con la norma del supremo, y el espacio de funciones continuas a la derecha, con límite por la izquierda, dotado con la geometría de Skorokhod. La metodología desarrollada se basa en la construcción del Lema 2.1 en Kuelbs [1970], donde se establece que, para cualquier espacio de Banach separable, se puede definir un espacio de Hilbert con topología más débil, bajo condiciones apropiadas. En este contexto, se genera una nueva sucesión de espacios de Hilbert y Banach encajados de forma continua, que permite extender los resultados existentes, sobre consistencia, a un contexto más general.

- La introducción de un enfoque Bayesiano en la estimación componente a componente de los autovalores del operador de autocorrelación, estableciendo la eficiencia asintótica y la equivalencia entre el estimador clásico y Bayesiano. Asimismo, se establece la equivalencia asintótica de los predictores asociados.

- La aplicación de los resultados derivados al contexto de modelos FANOVA, con término de error ARH(1), es también considerada. En particular, se introducen nuevos modelos de operadores de covarianza matricial, cuyas entradas funcionales, fuera de la diagonal, poseen un espectro puntual no separable.

- Se consideran, en todos los casos, amplios estudios de simulación, con el objeto de comparar con otros enfoques las propiedades asintóticas de los estimadores analizados, así como derivar numéricamente nuevas razones de convergencia en relación con la eficiencia asintótica y la consistencia.

- Se ilustra, en particular, la implementación práctica para el análisis de datos de elevada dimensión, de las metodologías de estimación y predicción funcional adoptadas, en todos los ámbitos estudiados, mediante aplicaciones en términos de datos reales.

SUMMARY

This PhD thesis focuses on statistical estimation and prediction from temporal correlated functional data. We adopt the functional time series framework, considering, in particular, autoregressive processes in Hilbert and Banach spaces (ARH(1) and ARB(1) processes). Our primary objective is the statistical estimation of the conditional mean, from temporal correlated data, considering linear models in a parametric framework. That is the case, for example, of the estimation of the functional response in linear regression, with functional regressors and correlated errors, lying in Hilbert or Banach spaces. Some extensions to the Bayesian framework are derived as well. Nonparametric classification is also considered, in the special case of spatially supported uncorrelated functional data. Specifically, the main contributions of this PhD thesis can be summarized as follows:

- The derivation of new weak– and strong– consistency results, for componentwise estimators of the autocorrelation operator of an ARH(1) process, in the norms of bounded linear, Hilbert–Schmidt and trace operators. Under the same setting of conditions, consistency of the corresponding plug–in predictors is derived as well. The cases of known and unknown eigenvectors are studied. Some particular examples are also analysed, such as the Ornstein–Uhlenbeck process in Hilbert and Banach spaces, as motivation of the subject summarized in the next paragraph.

- The extension of the results previously derived on functional prediction, based on AR$\mathcal{C}(1)$ and AR$\mathcal{D}(1)$ processes, with respective values in the space of continuous functions and in the Skorokhod space, to the case of an abstract separable Banach space. Specifically, sufficient conditions are obtained for the strong–consistency of the componentwise estimator of the autocorrelation operator, and the associated plug–in predictor. The methodological approach proposed, in the derivation of these results, is based on the construction appearing in [Kuelbs, 1970, Lemma 2.1], and the definition of continuous embeddings between suitable Banach and Hilbert spaces.

- The introduction of the Bayesian statistical perspective, in the componentwise estimation of the autocorrelation operator of an ARH(1) process, with the consideration of the corresponding ARH(1) plug–in predictor, under weaker setting of conditions than before for its asymptotic efficiency. The asymptotic equivalence of both, the classical and Bayesian estimators and plug–in predictors, is studied as well.

- The FANOVA analysis of functional fixed effect models in Hilbert spaces, under correlated errors, having values in a separable Hilbert space. In this context, non–separable point spectrum matrix covariance operator models are analysed.

- A wide range of simulation studies have been undertaken, for comparative purposes, in relation to the existing functional prediction methodologies in the ARH(1), ARB(1) and nonparametric frameworks.

- Some real–data applications are considered to illustrate the implementation of the proposed functional estimation and prediction methodologies in practice.

Mathematics is the art of giving the same name to different things

Henri Poincaré (29th April 1854 - 17th July 1912)

1
INTRODUCTION

This introduction provides the reader with a comprehensive review about the state of the art of functional data analysis, focusing on the functional time series context. Aimed at clarifying, the main elements concerning the notation adopted in the following sections are previously presented.

NOTATION

FDA Functional Data Analysis.

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ Sets of natural, integers, rational, real and complex numbers, respectively.

$A \times B$ Cartesian product of sets A and B.

$(\Omega, \mathcal{A}, \mathcal{P})$ Probability space, being Ω a non–empty set, $\mathcal{A}$ a σ–algebra and $\mathcal{P}$ a probability measure.

$i.i.d.r.v.$ Independent and identically distributed random variables.

$\mathcal{B}(a, b)$ A Beta distribution with shape parameters a and b.

$\mathrm{E}\{X\}, \mathrm{Var}\{X\}$ Expectation and variance, respectively, of a random variable X.

$\mathrm{E}\{X|Y\}$ Conditional expectation of the random variable X, depending on the random variable Y.

$C_X, C_{X,Y}$ Covariance and cross–covariance operators of the random variables X and Y.

$X \perp Y$ X and Y are weakly orthogonal between them.

$\boldsymbol{Id}_n$ Identity matrix (real–valued matrix or matrix of identity operators) of dimension $n \times n$.

$\mathcal{N}(\mu, \sigma^2)$ Real–valued normal distribution with expectation μ and variance σ^2.

$\mathcal{N}(\mu, C)$ Normal distribution with expectation μ and covariance operator C, valued in a function space.

$\ln(x)$ Natural (naperian) logarithm.

$\lceil x \rceil$ Integer part of x.

$\{x_i,\ i = 1, \ldots, n\}$ Sequence indexed from $i = 1$ to $i = n$.

$\boldsymbol{A} = \{a_{i,j}\}_{i=1,\ldots,n}^{j=1,\ldots,m}$ Matrix which entries are given by $a_{i,j}$, for each $i = 1, \ldots, n,\ j = 1, \ldots, m$.

$\mathbf{1}$. Indicator function: $\mathbf{1}_A(x) = 1$ if $x \in A$; $\mathbf{1}_A(x) = 0$ if x does not belong to A.

$X \equiv Y$ X is equivalent to Y.

$X \simeq Y$ X is approximated by Y.

■ End of proof.

$u_n \sim o(v_n)$ Little–o notation: $\lim_{n\to\infty} \frac{u_n}{v_n} = 0$.

$u_n \sim \mathcal{O}(v_n)$ Big–$\mathcal{O}$ notation: for n_0 large enough, there exists a finite constant M such that $u_n \leq M v_n$, for each $n \geq n_0$.

$\longrightarrow^p$ Converge in probability.

$\longrightarrow^{a.s.}$ Almost sure convergence.

$\mathrm{span}(x_i,\ i \in I)$ Linear space generated by $\{x_i,\ i \in I\}$.

$\delta_{k,p}$ Delta function such that $\delta_{k,p} = 0$ when $k \neq p$, and $\delta_{k,p} = 1$ when $k = p$.

$\mathrm{Tr}(\boldsymbol{A})$ Trace of matrix $\boldsymbol{A}$.

$\mathrm{sgn}(x)$ Sign of the value x, such that $\mathrm{sgn}(x) = \mathbf{1}_{x\geq 0} - \mathbf{1}_{x<0}$.

$\mathcal{C}((a,b))$ Space of continuous functions, whose support is defined on the real interval (a,b).

$(H, \langle\cdot,\cdot\rangle_H)$ Hilbert space H with its associated inner product $\langle\cdot,\cdot\rangle_H$.

$(B, \|\cdot\|_B)$ Banach space B with its associated norm $\|\cdot\|_B$.

B^* Topological dual of a Banach space B.

ℓ^* Adjoint of the operator ℓ.

$\mathcal{L}(B, B')$ Space of continuous and bounded linear operators from B to B', associated with the usual uniform norm.

$\left(\mathcal{L}(B), \|\cdot\|_{\mathcal{L}(B)}\right)$ Space of continuous and bounded linear operators from the Banach space B to itself, associated with the usual uniform norm $\|\cdot\|_{\mathcal{L}(B)}$.

$\left(\mathcal{S}(H), \|\cdot\|_{\mathcal{S}(H)}\right)$ Space of Hilbert–Schmidt operators over H, associated with the Hilbert–Schmidt norm $\|\cdot\|_{\mathcal{S}(H)}$.

$\left(\mathcal{N}(H), \|\cdot\|_{\mathcal{N}(H)}\right)$ Space of nuclear operators over H, associated with the nuclear norm $\|\cdot\|_{\mathcal{N}(H)}$.

$f \otimes g$ Tensorial or Kronecker product of Hilbert–valued functions f and g, given by $(f \otimes g)(x) = \langle f, x \rangle_H g(x)$ for each $x \in H$, and belonging to the space $\mathcal{L}(H)$.

INTRODUCTION

The classical theory of statistical inference from time series models (see, e.g., Brillinger [1981]; Brockwell and Davis [1987]; Hamilton [1994]; Rao et al. [2012]) has emerged as a powerful tool for the analysis of data correlated in time, motivated by a wide range of applications, in diverse fields such as electricity consumption (see Abdel-Aal and Al-Garni [1997]), environmental data (see Mills [2013]) and term structure forecasting (see Diebold and Li [2006]). The aim of the current dissertation has been to establish alternative frameworks on the statistical inference on infinite–dimensional time series, whose underlying phenomena are continuous in nature. Specifically, we provide new theoretical results and alternative numerical implementations on the estimation and prediction of time series in Hilbert and Banach spaces.

Similar techniques to those used in multivariate data analysis (MVA) framework (see Anderson [2003]; Izenman [2008], among others) have been adapted to the functional data analysis, where measures are being gathered with an increasing frequency. As exposed in the monograph by Ramsay and Silverman [2005], these recent technological developments have led to the formulation of alternative methodologies for dealing with high–dimensional data problems (see, e.g., Bouveyron [2004]; Bühlmann and de Geer [2011]). Data may even be continuously collected so that high–dimensional data become functional data (term coined by Ramsay [1982]), constituted by infinite–dimensional objects, such as curves or surfaces, including a richer source of information, but also facing more complex challenges. This one of the main motivations, in the extensive literature developed in the last few decades on functional data analysis (FDA). Hence, FDA covers a wide range of problems, such as temporal gene expression data (see Leng and Müller [2006]), spectrometric absorbance curves (see Ferraty and Vieu [2006]) or climatological data (see Besse et al. [2000]). This is the case, for instance, of the analysis of air pollutants, which are measured every hour but the pattern of daily concentration curves provides key information (see, e.g., Febrero-Bande et al. [2008]; Ignaccolo et al. [2014]; Slini et al. [2006]; Stadlober et al. [2008]). Since a finite discretization will be required, from a computational perspective, one might be tempted to handle it under a MVA or high–dimensional methodologies, but ill–posed problems arise. This fact has raised the formulation of a formally mathematical framework. We refer to comprehensive monographs and surveys by Bosq [2000]; Cuevas [2014]; Ferraty and Vieu [2006]; Goia and Vieu [2016]; Horváth and Kokoszka [2012]; Hsing and Eubank [2015], where some of the FDA foundations have been introduced, deriving fundamental results for the analysis of uncorrelated and correlated functional data, such as functional limit theorems (see Ledoux and Talagrand [2011]), semi–metrics well–adapted for functional data (see Berrendero et al. [2018b]), outliers detection (see Kuhnt and Rehage [2016]), functional classification (see Álvarez-Liébana and Ruiz-Medina [2015]; Baíllo et al. [2011]; James and Hastie [2001]) and functional time series (see, e.g., Bosq [2000]; Ferraty et al. [2002]). The current dissertation is mainly aimed at providing alternative estimation methodologies, and asymptotic results, related to the latter context, as well as extensions to more general frameworks.

Even though it is not the scope of this dissertation, it deserves to be mentioned the fact that several methodologies can be found for the analysis of functional data, based on Functional Principal Components Analysis (so–called FPCA; see, e.g., Boente and Fraiman [2000]; Febrero-Bande et al. [2017]). In particular, this technique has been applied to the estimation of the functional slope $\beta(\cdot)$ of a functional linear regression model, under uncorrelated or independent observation errors. Hall and Hosseini-Nasab [2006] derived the conditional mean–square error of the estimator of β. Rates of convergence were provided in Hall and Horowitz [2007], and the asymptotic behaviour of the so–called regularised estimator, avoiding the disruptions caused by small eigenvalues, was analysed (see also Pezzulli and Silverman [1993]). We refer to Escabias et al. [2014], where the FPCA–based estimation of a functional generalized logit model with multi–category response variables was addressed (see Albaqshi [2017]). As an alternative to the infinite-dimensional decomposition in terms of the FPC scores, the Functional Partial Least Squares Regression (so–called FPLSR) methodology, early proposed by Preda and Saporta [2005], is aimed at finding the more relevant set of uncorrelated random variables on the forecasting of the real–valued response (Delaigle and Hall [2012]). See also the proposals in Gabrys et al. [2010]; Gabrys and Kokoszka [2007] for testing the i.i.d. assumption of the error terms.

The main objective of this dissertation has been the statistical analysis of temporal correlated functional data. Thereby, functional prediction and estimation in Hilbert and Banach spaces, from a parametric linear time series framework is, indeed, the central topic in this monograph. Before discussing functional time series approaches based on moment estimation of the dependence structure (commonly involving mixing conditions; see Bosq [2000]), we refer to some general approaches beyond the state equation based modelling. A first attempt can be found in Hörmann and Kokoszka [2010], where the notion of m–dependent processes is introduced. In this framework, the concept of $\mathcal{L}^p$–m–approximability plays a crucial role (see also Diaconis and Freeman [1999]; Wu and Shao [2004], where functional stochastic processes were reinterpreted as a collection of iterated random functions). Some applications of m–dependence to non–linear sequences were raised by Berkes et al. [2011]; Hörmann [2008]. We also refer to the results outlined in [Horváth and Kokoszka, 2012, Section 16] concerning change–point detection tests for the functional mean of $\mathcal{L}^p$–m–approximable processes, extending the results by Berkes et al. [2009]. Aimed at extending the mixing case mostly adopted (but unrealistic in some situations), the dependence structure of a sequence of random variables can also be determined by its ergodicity properties (functional ergodicity was addressed by Laïb [2005]; Laïb and Louani [2011], among others). Although, here, in Appendix A5, more general situations are analysed, beyond the usual ergodicity assumptions, data sampled from stationary and ergodic stochastic processes also contitutes a central topic in FDA. In this context, a nonparametric kernel–based regression was suggested by Laïb and Louani [2010] and its asymptotic properties were therein studied. Chaouch et al. [2017] proposed a nonparametric kernel–based estimator for the conditional mode, providing uniformly consistent rates of convergence. In the introduced context of mode regression, the asymptotic normality of the conditional mode estimator was proved in Ezzahrioui and Ould-Saïd [2008, 2010], in both i.i.d. and mixing scenarios. Motivated by data coming from, for instance, sampling surveys, in which data are rarely fully observed, Ling et al. [2017] formulated a well–adapted conditional mode estimator, under stationary ergodic and *missing at random* (MAR) responses (see also Ferraty et al. [2013b]).

Aditionally to conditional mode estimation, several authors have contributed to conditional density, hazard function or conditional distribution estimation. Conditional density estimation for a scalar response variable, depending on a functional regressor, was addressed in Ferraty et al. [2010] (see also Mörters and Peres [2010]). With the same aim, Kara et al. [2017a] formulated a k–Nearest–Neighbours (kNN) methodology for i.i.d. pairs $\{(X_i, Y_i)\,,\ i = 1, \ldots, n\}$, where $\{X_i,\ i = 1, \ldots, n\}$ are valued in a semi–metric space and $\{Y_i,\ i = 1, \ldots, n\} \subset \mathbb{R}$, on the estimation of three conditional models: regression (i.e., $\mathrm{E}\{Y|X = x\}$), conditional density (i.e., $F^x(\cdot) = \mathcal{P}(Y \leq \cdot|X = x)$) and conditional distribution function (i.e., $f^x(\cdot) = (F^x)'(\cdot)$). The three conditional models above referred are also tackled in Kara et al. [2017b], providing uniform asymptotic results on the choice of the bandwidth involved. Median notions and depth–based measures have also been studied (see, e.g., Cuesta-Albertos and Nieto-Reyes [2008]; Cuevas and Fraiman [2009]; López-Pintado and Romo [2006, 2009]).

According to the space where response and covariates take their values, functional linear regression models (FLRM) can be grouped (see Horváth and Kokoszka [2012]) into three categories: scalar response FLRM, functional response FLRM and fully FLRM. Cardot et al. [1999] firstly addressed the former scenario, applying a FPCA methodology in order to achieve a strongly–consistent componentwise estimator of the functional slope (that approach was later modified by Cardot and Sarda [2011]). Regularization penalized least–squares approaches, based on a roughness penalty, have also been formulated to solve the dimensionality problem. In Li and Hsing [2007], a regression operator, in terms of a Fourier basis expansion, is considered, such that an estimator was built by solving a penalized minimization problem. A spline basis expansion, instead of Fourier bases, was regarded in Crambes et al. [2009] (see also Cardot et al. [2007]). Reiss and Ogden [2007] combined both perspectives, providing a componentwise framework such that FPC scores of regressors are obtained by using a roughness penalty. Berrendero et al. [2018a] recently proposed a variable selection procedure, suggesting that the entire paths of functional regressors can be replaced by a set of their more relevant instants. From a semi–parametric perspective, Goia and Vieu [2015] recently proposed a two–terms semi–parametric approach, formulating the so–called Partitioned Functional Single Index Model (see the surveys by Goia and Vieu [2014, 2016] and the references therein). This methodology revolved around considering an additive decomposition of the regression operator in terms of its restrictions into a finite set of sub–intervals. These restricted operators depend on unknown real link functions and unknown directions, which are estimated by a nonparametric kernel–based iterative procedure. This non–linear link function setting is motivated by the categorization of classification problems (for instance, in a binary code as done in Müller and Stadtmüller [2005]). Link functions can be assumed to be known (see, e.g., Wang et al. [2010]) or unknown (see, e.g., Chen et al. [2011]; Wang et al. [2016]).

In a nonparametric context, Ferraty and Vieu [2006, 2011] consider, for instance, a functional formulation of the Nadaraya–Watson kernel estimator. See also Biau et al. [2010]; Burba et al. [2009], where kNN nonparametric regression estimators were established (a similar strategy can be found in Laloë [2008], where its weak–consistency was studied). The asymptotic normality of the mentioned Nadaraya–Watson kernel estimator was derived by Masry [2005]. In the Bayesian framework, Lian et al. [2016] provided a theoretical benchmark for the analysis of the asymptotic properties of a Bayesian scalar response in FLRM, illustrating how the choice of the prior distributions can imply the minimax rate in prediction risk (see also the Bayesian approach developed in Shang [2013] for estimating the bandwidth in nonparametric kernel–based functional regression, when nonparametric local–weighting–based methodology is adopted). Grollemund et al. [2018] has recently derived a novel Bayesian methodology for the estimation of the support of the functional explanatory variable. The case of functional response in FLRM has been mainly addressed by Chiou et al. [2004, 2003]. Their results were motivated by testing how different dietary doses levels (scalar regressors) affects to the egg–laying curves in flies; i.e., how the number of eggs laid every day are affected by the dietary conditions. The same experiment was studied by Müller and Stadtmüller [2005], where the above referred semi–parametric generalized functional linear model, based on a link function, was applied to classify the period of life of those flies (see also Li et al. [2010]) (1 is assigned if flies live, at least, 44 days, and 0 otherwise).

Lastly, the problem of fully FLRM has been deeply studied. This framework was firstly analysed by Ramsay and Dalzell [1991], providing a penalized least–squares method to estimate, in a strongly–consistent way, the kernel of the regression operator. Cuevas et al. [2002] considered Hilbert–valued explanatory and response variables, assuming a fixed and triangular design. The referred insights established in Ferraty and Vieu [2006], in the context of nonparametric functional prediction, were extended by Ferraty et al. [2011] to the fully FLRM, when the functional response and regressors are valued, respectively, in a Hilbert and a semi–metric space. The asymptotic normality of this functional kernel–based predictor was later studied in Ferraty et al. [2012]. Beyond the flexibility obtained from nonparametric proposals, as commented before, their main drawback lies in the dimensionality problem, and on the selection problem, regarding the choice of the kernel function and the bandwidth parameter. A fractal approach was proposed in Ferraty et al. [2002] for solving the so–called *curse of dimensionality*, by imposing a concentration assumption about the distribution of stochastic processes.

Functional testing procedures have also played a mayor role. Motivated by the analysis of the perturbations displayed in the magnetic field in polar regions, an approach to test the lack of effect of the regression operator of a fully FLRM was proposed in Kokoszka et al. [2008]; Maslova et al. [2010]. We also refer to

Horváth and Reeder [2012], where the problem of detecting a change in the regression operator, throughout a time period, was reached. Collazos et al. [2016] formulated a proposal for simultaneously testing which set of regressors or covariates is more relevant (see also Bunea et al. [2006]), commonly applied in the weather forecasting. The extension of the well–known ANOVA techniques to the FDA framework has engaged the effort of many authors, since their crucial role on the medical imaging, among other areas. Fan [1996] firstly formulated significance tests applied to the framework of functional data, based on the sparsity achieved in the context of wavelet–based signal decomposition. Shen and Faraway [2004] formulated an F test for functional response FLRM, aimed at testing different functional linear models. A FANOVA model, for the particular case when error term has independent and homoscedastic components, was proposed in Zoglat [2008]. This proposal was extended by Ruiz-Medina [2016] to the case of multivariate Hilbert–valued fixed effects model, in which the functional components of the Gaussian error term are assumed to be correlated. See also the non–separable approach in Álvarez-Liébana and Ruiz-Medina [2017], displayed in Appendix A4, and applied to the analysis of functional magnetic resonance images (Aston and Kirch [2011a,b]) as discussed below.

Functional time series can be seen as a particular case of the fully FLRM, where functional regressors are given by their own past values. As well known, functional time series framework could also be viewed as an alternative *semi–parametric* approach (parametric in time and without parametric assumptions in the function space where the process takes its values). The existing methodologies may be divided into time–domain and frequency–domain approaches. In the frequency–domain context, spectral analysis plays a major role. For instance, discrete Fourier analysis was used in Panaretos and Tavakoli [2013b] for the purely infinite–dimensional inference of the entire second–order dependence structure of a stationary weakly dependent functional time series, by estimating the spectral density operator. This estimation was achieved by smoothing the periodogram operator, providing central limit theorems for the functional mean and the long–run covariance operator. Shishebor et al. [2011] studied the spectral asymptotic properties and distribution of the periodogram operator of periodically correlated processes. Soltani et al. [2010] proved that the periodogram, for weakly and strongly second–order periodically correlated processes, is unbiased, in an asymptotically way (see also Panaretos and Tavakoli [2013a]; Soltani and Shishebor [2007]).

As commented before, the current dissertation focuses on the time–domain approaches. Specifically, for the forecasting of a continuous–time zero–mean stochastic process $\xi = \{\xi_t,\ t \in \mathbb{R}^+\}$, we will apply the methodology proposed in Bosq [1991]. As suggested, giving a real interval $[0, \delta]$, the stochastic process ξ is turned into a set of zero–mean functional random variables $X = \{X_n(t) := \xi_{n\delta+t},\ 0 \leq t \leq \delta,\ n \in \mathbb{Z}\}$, with support on the interval $[0, \delta]$. The first attempt on the forecasting of functional time series was achieved by Bosq [1991], where $X = \{X_n,\ n \in \mathbb{Z}\}$ is assumed to be valued in a separable Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$. Bosq [1991] firstly established the state equation of a zero–mean Hilbertian autoregressive linear process of order one (so–called ARH(1) process), as follows:

$$X_n(t) = \rho\left(X_{n-1}\right)(t) + \varepsilon_n(t), \quad X_n,\ \varepsilon_n \in H, \quad \rho:\ H \longrightarrow H, \quad t \in [0, \delta],\ n \in \mathbb{Z}, \tag{1.1}$$

arising a natural extension of the multivariate time series framework, going beyond the finite–dimensional structure of the state space. The autocorrelation operator ρ appeared in (1.1) was imposed to belong to the space of bounded linear operators on H (i.e., $\rho \in \mathcal{L}(H)$), being $\varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}$ a H–valued strong white noise with $\sigma_\varepsilon^2 = \mathrm{E}\left\{\|\varepsilon_n\|_H^2\right\} < \infty$ (and uncorrelated with the random initial condition). The ARH(1) model displayed in (1.1) has been applied during the last decades to a wide range of real–data problems, such as forecasting of electricity consumption (see Cavallini et al. [1994]), analysis of electrocardiograms (see Bernard [1997]), forecasting of sulfur dioxide levels (see de Castro et al. [2005]) and credit card transactions (see Horváth et al. [2010]), among other areas of application. Approaches in functional time series have been derived mainly adopting a parametric, semi–parametric or nonparametric framework.

In the parametric linear context, the ARH(1) estimation methodology established in Bosq [1991] was focused on extending the well–known moment–based estimation approach, by projecting the whole trajectories into a finite subspace spanned by a complete orthonormal eigenvectors system $\{\phi_j,\ j \geq 1\}$ of the associated autocovariance operator C. As stated in the real–valued framework, the stationarity of the ARH(1) process above defined has had to be established (see [Bosq, 2000, Lemma 3.1], where sufficient conditions were provided). From the infinite–dimensional nature of the underlying trajectories, the main difficult to be solved in this particular benchmark relies on the fact that C cannot be inverted in the whole domain, and therefore, empirical truncated estimators of the autocorrelation operator ρ must be formulated (see the proposal for the choice of the truncation parameter in [Hörmann and Kidziński, 2015, Section 2.5]).

Once stated the referred stationary zero–mean ARH(1) linear model, from the asymptotic results derived in Bosq [1999a,b]; Merlevède [1996a]; Merlevède et al. [1997], among others, and from the Close Graph Theorem and the compactness of C, Mas [1999] paid attention to the adjoint of ρ instead of itself, proving the asymptotic normality of a componentwise estimator, under the strictly positiveness of C. In keeping with the framework firstly formulated in Bosq [1991], a comprehensive asymptotic estimation theory was insightfully developed in Bosq [2000], under the major assumption that ρ belongs to the class of Hilbert–Schmidt operators. Specifically, the following strongly–consistent estimator was therein provided, when the eigenvectors of C are unknown, and therefore, empirical eigenvectors $\{\phi_{n,j},\ j \geq 1\}$ and eigenvalues $\{C_{n,j},\ j \geq 1\}$ must be computed:

$$\widehat{\rho}_{k_n}(\cdot) = \sum_{j=1}^{k_n} \frac{1}{C_{n,j}} \langle \cdot, \phi_{n,j} \rangle_H \widetilde{\Pi}^{k_n} D_n(\phi_{n,j}),$$

where D_n denotes the empirical estimator of the cross–covariance operator, being $\widetilde{\Pi}^{k_n}$ the orthogonal projector into the finite–dimensional spanned by $\{\phi_{n,j},\ 1 \leq j \leq k_n\}$. During the last two decades, several proposals to improve the referred parametric componentwise framework can be found, achieving, for instance, the asymptotic efficiency of the estimator of ρ established in Bosq [2000] (see the regularization procedure proposed in Guillas [2001]), the derivation of the asymptotic properties of the covariance operators (see the large and moderate deviations established in Mas and Menneteau [2003a] for the empirical mean and empirical covariance operators, in the case of Gaussian ARH(1) processes), the consistency of the estimator in Mas [1999] (see the developments provided by Mas [2004]) or its weak–convergence (see sufficient conditions established in Mas [2007]). In the current literature, consistency results have been mostly derived in terms of the norm of bounded linear operators, but convergence in the Hilbert–Schmidt and trace operator norms had been barely tackled, up to our knowledge. With this purpose, in the particular scenario when ρ admits a diagonal spectral decomposition in terms of the eigenvectors of C (easily reached when wavelet–based and shrinkage procedures are applied to the covariance operators), sufficient conditions for the convergence in the mean–square sense of the formulatd estimator to ρ, in the Hilbert–Schmidt norm, were theoretically established in Álvarez-Liébana et al. [2017] (see Appendix A3). In the same context of consistency in the norm of Hilbert–Schmidt operators, asymptotic efficient results, in the referred norm, have been achieved in Ruiz-Medina and Álvarez-Liébana [2017] (see Appendix A5), beyond the Hilbert–Schmidt assumption over the autocorrelation operator, commonly adopted in the above mentioned works.

One of the first works which tackled alternative bases to the ones involved in the spectral decomposition of the covariance operator was the one carried out by Bensmain and Mourid [2001], where the Grenander's theory of the so–called sieves was adapted to the ARH(1) framework. Specifically, a Fourier basis expansion was proposed, in a manner that the estimator of ρ was reached by forecasting its associated coefficients, restricted to a finite–dimensional subset of the parametric space (i.e., restricted to a sieve), where ρ takes values. Regularization procedures based on smoothing functions, B–spline and wavelet bases can also be implemented (see, e.g., Antoniadis and Sapatinas [2003]; Laukaitis and Rackauskas [2002]; Mas [2000]). The work by Kargin and Onatski [2008] was a path–breaking reference on the forecasting of an ARH(1) process, since their reduced–rank–searching–based proposal paid attention to the spectral decomposition of ρ, finding the directions more relevant to forecasting, instead of focusing on the covariance operators themselves (see the comparative study addressed in Didericksen and Kokoszka [2012], between approaches in Bosq [2000] and Kargin and Onatski [2008]). Weighted extensions of the FPLSR and FPCA methodologies have also been formulated in Hyndman and Shang [2009]; Hyndman and Ullah [2007].

As an extension of the asymptotic efficiency results derived in Bosq and Ruiz-Medina [2014] for l_2–valued Poisson processes, the mentioned framework established in Ruiz-Medina and Álvarez-Liébana [2017] has also been intended to develop an alternative Beta–prior–based Bayesian componentwise estimation framework. Besides the referred Bayesian proposals in the more general context of FLRM, new results on the Bayesian networks were formulated in Shojaie and Michailidis [2009], aimed at testing changes in the gene expression levels, such that the associated adjacency matrix, which represents the relations among genes and proteins, was efficiently estimated. Those results can be adapted for the estimation of the eigenvalues of the autocorrelation operator of an ARH(1) process, which can be reinterpreted as the influence matrix of a graph, closely linked to the adjacency matrix (see also the Markov–Chains–based methodology proposed in Petris [2013] for the Bayesian estimation of functional time series). Fahrmeir and Kneib [2011] addressed the Bayesian–based componentwise analysis of high–dimensional longitudinal and spatial data. Blanke and Bosq [2015] recently formulated a Bayesian methodology for the estimation and prediction of an Ornstein–Uhlenbeck process (see also Álvarez-Liébana et al. [2016] and Appendix A2, where a strongly–consistent estimator of an Ornstein–Uhlenbeck was established). In the particular case of monotonic functional time series, a Bayesian approach, focused on the forecasting of Italian natural gas market, was recently proposed in Canale and Ruggiero [2016], based on population genetics strategies (we also refer to Kowal [2017]).

Besides the results formulated for testing the stationarity, as well as the nullity of the autocorrelation operator and its change point detection (see, e.g., Horváth et al. [2010, 2014]; Kokoszka et al. [2008]; Zhang et al. [2011]), the ARH(1) parametric framework was gradually extended to alternative dependence structures, such as ARH(p) processes with p greater than one (see Guillas [2000]). See also the procedure for testing the lag order p proposed in Kokoszka and Reimherr [2013b] and the automatic selection approach in [Aue et al., 2015, Section 3]. Aimed at incorporing richer sources of information, Guillas [2000] firstly formulated the additive inclusion of exogenous variables (ARHX processes) in the dependence structure, which was subsequently addressed in Damon and Guillas [2002, 2005]. See also ARHD processes stated in Marion and Pumo [2004]; Mas and Pumo [2007], by including the first derivatives as the exogenous variables. Exogenous information can also be incorporated in a non–additive way, as proposed in Cugliari [2013]; Guillas [2002], where ρ was defined in terms of the realization of a sequence of random variables. The randomness feature of ρ was addressed in Mourid [2004] by conditioning the referred operator to each element of the sample space. Note that the autocorrelation operator can also be given as a function of an unknown real–valued parameter (see, e.g., Kara-Terki and Mourid [2016]). The interest of this last framework relies on, for instance, the forecasting of interest rates modelling by the Vasicek's model. This model, also known as Ornstein–Uhlenbeck process, can be characterized as a stationary ARH(1) process (as well as a stationary Banach–valued autoregressive ARB(1) process), driven by an autocorrelation operator depending on its scale parameter. This characterization, jointly with its consistent prediction, is the common thread among the already referred paper by Álvarez-Liébana et al. [2016]. Hilbertian moving average processes (MAH processes), ARMAH processes and Hilbertian general linear processes (LPH processes) have also been analysed in recent works (see, for instance, Bosq [2007]; Bosq and Blanke [2007]; Chen et al. [2016]; Dedecker and Merlevède [2003]; Turbillon et al. [2008]). Soltani and Hashemi [2011] addressed the estimation of an ARH(1) process, in which the autocorrelation operator is periodically correlated (PCARH(1) processes; see also Kidziński et al. [2016]), such that $\rho_n = \rho_{n+T}$ (i.e., periodically correlated with period T). Hörmann et al. [2016] focused on testing the presence of a periodic component in functional time series in both scenarios, when a periodic signal valued in a function space is disrupted by a functional white noise, and in the case of weakly dependent contaminated processes.

It deserves mentioning the another path–breaking work by Ruiz-Medina [2011], where a new branch in the field of functional time series was exhibited, on the inference of spatial Hilbert–valued autoregressive processes of order one (SARH(1) processes). That work provided the sufficient conditions required in the stationarity of a SARH(1) process, which methodology revolved around the so–called Markov property of the three points for a spatial stochastic process. This novel approach was applied in Ruiz-Medina [2012] to the context of spatial functional prediction of ocean surface temperature, providing the moment–based estimators of the functional parameters involved (see also the spatio–temporal analysis of epidemiological data addressed by Ruiz-Medina et al. [2014]). Some insights about spatial functional data can also be found in Delicado et al. [2010]; Giraldo et al. [2010]; Nerini et al. [2010], among others.

Some authors have also been focused on extending the univariate non–linear models to the FDA framework (see the non–linear representations for functional data provided by Chen and Müller [2012], as well as the detection break points methodology proposed in Davis et al. [2008]). Note that the referred nonparametric proposals are usually addressed involving non–linear regression or autocorrelation operators. The same issue should be noted for $\mathcal{L}^p$–m–approximable stochastic processes, which are not constraint to linear models, and therefore, they can be adapted for the analysis of non–linear functional time series. For instance, Kokoszka and Reimherr [2013a] established the asymptotic normality of the empirical autocovariance operator, in the context of non–linear and $\mathcal{L}^p$–m–approximable processes. Hörmann et al. [2013] extended, to the functional context, the well–known ARCH model, which state equation is non–linear respect to the error terms. Bootstrap methods well–adapted for non–linear functional time series have been recently developed by Shang [2018], on the estimation of the long–run covariance under nonparametric kernel–based techniques. A special semi–functional framework was introduced in Wang [2008], where a real–valued non–linear ARIMA(p,d,q) model was modified, including functional MA coefficients.

Most of the works on the inference of functional time series are formulated in Hilbert spaces. This choice is not arbitrary since a Hilbert space naturally provides a generalization of classical Euclidean spaces, with an inner–product–based structure. However, Hilbert–valued framework is constraint, e.g., Hilbert spaces like L^2 spaces are not suitable for measuring local regularity or singularity of functions. We now refer to this more flexible Banach space framework. The main challenge facing when Banach–valued time series are analysed lies in the fact that the notion of orthogonality disappears. The first attempt was carried out by Pumo [1992, 1998], as an initial extension of the ARH(1) framework. The referred works consider, as a Banach space $(B, \|\cdot\|_B)$, the space of continuous functions on $[0,1]$ (so–called $\mathcal{C} = \mathcal{C}([0,1])$), with the supremum norm. Particularly, they formulated an AR$\mathcal{C}(1)$ process, which can be continuously extended to the Hilbert space $L^2([0,1])$, enabling to apply the asymptotic componentwise Hilbertian results already established. Whereas the natural extension of AR$\mathcal{C}(1)$ to AR$\mathcal{C}$(p) processes (with p greater than one) was raised by Mourid [1993, 1996] and later analysed in Benyelles and Mourid [2001], strongly–consistent results on the estimation and prediction of AR$\mathcal{C}(1)$ processes were proved in Bosq [2000], under strongly–mixing conditions. The referred Banach space was also addressed by Haan and Lin [2001] in the study of extreme value distributions in infinite–dimensional spaces. A broad literature concerning the main properties of Banach–valued autoregressive processes have also been developed (see, e.g., the book by Mas and Pumo [2010]). Specifically, the functional second–order moments of a Banach–valued autoregressive process were empirically estimated by Bosq [2002]. In contrast with the orthogonal innovations stated in Bosq [2002], Dehling and Sharipov [2005] studied an extension of that theoretically results to the case of weakly dependent innovations. A new estimator of the autocorrelation operator of a Banach–valued autoregressive process was established in Rachedi and Mourid [2003] and its strong–consistency deeply studied. Since the Reproducing Kernel Hilbert Spaces (RKHS) theory will play a key role in Appendix A8, the work by Mokhtari and Mourid [2003] deserves to be mentioned, achieving the forecasting of $\mathcal{C}$–valued stochastic processes in the RKHS framework via Parzen's optimal approach (see Parzen [1961]). Bueno-Larraz and Klepsch [2018] have recently addressed the dimension reduction problem on the forecasting of an AR$\mathcal{C}$ process by adapting the variable selection proposal established in Berrendero et al. [2018a], based on the RKHS theory.

Paying attention to meteorological data, usually driven by jump diffusion processes (see, e.g., Grivas and Chaloulakou [2006]; Ignaccolo et al. [2014]; Stadlober et al. [2008]), which underlying curves are rather irregular than functions in $\mathcal{C}([0,1])$, Hajj [2011, 2013] firstly formulated the state equation of

$\mathcal{D}$–valued autoregressive processes of order one (so–called AR$\mathcal{D}(1)$ processes), being $\mathcal{D} = \mathcal{D}([0,1])$ the Skorokhod space of right–continuous functions on $[0,1]$ (see properly definitions in Skorokhod [1956]), having limit to the left at each $t \in [0,1]$. Previously, Davis and Mikoschb [2008] investigated the extreme value behaviour of the space–time process with values in the Skorokhod space $\mathcal{D}\left([0,1]^d\right)$, equipped with the J_1–topology. Specifically, Hajj [2013] formulated a consistent estimator, in the $\mathcal{D}$–norm, of the autocorrelation operator of an AR$\mathcal{D}(1)$ process, and established the state equation of a $\mathcal{D}$–valued moving average process of order one. In the referred work, the estimators were established in terms of the detection of the jumps involved, and the estimation of their amplitudes. With this purpose, Blanke and Bosq [2014] provided exponential bounds for those discontinuities, which were later used in Blanke and Bosq [2016]. In the referred work, the setback of $\mathcal{D}([0,1])$–valued ARMA(1,1) processes were examined, considering different scenarios: fixed instants with a given but unknown probability of jumps (deterministic case), random instants with ordered intensities (random case), and random instants with non ordered intensities (fully random case). Motivated by providing an extension of the referred $\mathcal{D}$–valued framework, Ruiz-Medina and Álvarez-Liébana [2018b] (see Appendix A8) goes beyond the regularity assumptions commonly arising in the $\mathcal{C}$– and $\mathcal{D}$–valued contexts, allowing to predict ARB(1) processes in general Banach spaces (only the separability property is assumed). A strongly–consistent estimator of an ARB(1) process, in the norm of $\mathcal{L}(B)$, and of its associated plug–in predictor, in the norm of B, are theoretically derived, respectively. This work can also be understood as an extension of the work by Labbas and Mourid [2002].

As commented, despite the good properties derived in parametric frameworks, the fitting to a particular model in some statistical problems could be unclear. This fact motivated the existence of semi–parametric and nonparametric methodologies, since those approaches provide us a greater flexibility with values of the regressors and responses in more general spaces, like semi–metric spaces (see the recent survey by Goia and Vieu [2016] and the overview about nonparametric and semi–metric regression theory by Härdle et al. [2004]). However, a dimensionality problem must be solved, as discussed above and pointed out in the surveys by Geenens [2011]; Vieu [2018]. For instance, a functional extension of the Nadaraya–Watson kernel–based predictor was performed in Besse et al. [2000]; Poggi [1994], among others. The use of semi–metrics for measuring the closeness between realizations of functional random elements can overcome the referred *curse of dimensionality* (see Álvarez-Liébana and Ruiz-Medina [2015]; Ferraty and Vieu [2006]).

Beyond the mostly adopted Nadaraya–Watson functional predictor, Antoniadis et al. [2006] proposed a two–steps prediction approach (later improved by Antoniadis et al. [2012]; Cugliari [2011]), based on a wavelet basis decomposition, which scaling coefficients are estimated by means of a nonparametric kernel–based predictor (see also Aneiros-Pérez et al. [2011]).

Aimed at finding a balance between the prediction accuracy and the flexibility of the models, some enhancements can be found in the current literature. Ruiz-Medina and Álvarez-Liébana [2018a] (see Appendix A6) recently derived the conditions required for establishing a strongly–consistent parametric componentwise predictor of an ARH(1) process, in the Hilbert–Schmidt and trace operator norms. In the later case, since only diagonal spectral eigenelements are involved when the trace norm is computed, a crucial dimension reduction is raised. The same aim was behind the new diagonal strongly–consistent componentwise estimator formulated in the referred paper, in terms of the singular value decomposition of the autocorrelation operator, depending on right and left eigenvector systems and its associated singular values. A semi–parametric literature has recently provided, motivated by capturing both the flexibility naturally provided by nonparametric frameworks and the advantages on the performance of parametric models. In this context, Aneiros-Pérez and Vieu [2006] proposed a semi–functional partial linear model, in which a real–valued response variable is inferred from an additional functional random variable and a linear combination of a set of real–valued explanatory variables. The forecasting of the response variable was therein addressed by a Nadaraya–Watson kernel–based predictor of the errors derived from the linear part, after computing a least–squares estimator of the vector parameters acting on the real–valued explanatory variables. The referred semi–functional methodology was adapted by Aneiros-Pérez and Vieu [2008] to the particular functional linear regression context of functional time series. As referred before, see also the Single Functional Index and sparse nonparametric functional regression models introduced in Ferraty et al. [2003] and Aneiros and Vieu [2017], respectively, involving nonparametric time series modelling in the random part.

As already mentioned, the main aim of the current dissertation has been to formulate new asymptotic results on the inference for temporal stochastic processes with values in infinite–dimensional spaces. Specifically, this thesis has contributed with new tools in the field of functional time series in Hilbert and Banach spaces, paying attention to the consistent estimation and prediction of ARH(1) and ARB(1) processes. Namely, the current dissertation provides novel techniques to solve the dimension reduction problem arises in the functional time series context, providing sufficient conditions on the functional estimators and predictors to ensure good asymptotic properties, as an alternative to the existing references in the semi–parametric and nonparametric frameworks. The achievements herein discussed can be grouped into four main branches. On the one hand, Appendices A2–A3 and A5–A7 establish diverse methodologies for the estimation and prediction of stationary ARH(1) processes, such as the convergence in the mean–square sense and the asymptotic efficiency in the norm of Hilbert–Schmidt operators, as well as the strong consistency in the three operator norms, under different sets of conditions. The formulation of a strongly–consistent estimator of an ARB(1) process, valued in an abstract separable Banach space, has also been carried out in Appendix A8. On the other hand, we have also derived flexible theoretical and numerical frameworks, on the supervised classification of random functions (see Appendix A1) and on testing the functional significance in FANOVA models (see Appendix A4).

The current dissertation is divided into six main chapters, broadly providing the objectives (see Chapter 2) and the methodology (see Chapter 3) followed in each one of the articles included in Appendices A1–A8, as well as the results (see Chapter 4) and conclusions (see Chapter 5) achieved. In accordance with the regulation in force, conclusions can also be found in Spanish language (see Chapter 6). Open research lines have been briefly discussed in Chapter 7. An exhaustive bibliography is listed at the end of this document.

I do not see that the sex of the candidate is an argument against her admission as a teaching assistant. After all, we are a university and not a bathing establishment

Amalie E. Noether (23rd March 1882 – 14th April 1935)

2 OBJECTIVES

We summarize here the main objectives addressed in the research articles and developments which constituted the current dissertation. Results and conclusions reached in Appendices A1–A8 below will be discussed in Chapter 4 and Chapters 5 –6, respectively.

- **Appendix A1.** Ferraty and Vieu [2006] proposed a supervised nonparametric curves classification technique. The application of these techniques in the classification of high–dimensional functional data requires the development of suitable numerical integration methods for n–dimensional supported random functions. This task has constituted the main objective of Appendix A1 (see Álvarez-Liébana and Ruiz-Medina [2015]). In particular, the implementation of semi–metrics to measure the closeness between n–dimensional elements is required, since a kernel–based approach, for the estimation of the posterior probabilities of the membership classes, is adopted. The purpose of this appendix has been to illustrate the performance of the functional classification methodology proposed, by the simulation study undertaken, and the real–data application addressed, in the field of railway engineering, in which the goal is to classify rail roughness surfaces.

- **Appendix A2.** The work in Álvarez-Liébana et al. [2016], reflected in Appendix A2 below, has as primary aim the theoretically formulation of a consistent predictor of the Ornstein–Uhlenbeck process (so–called O.U. process; see Uhlenbeck and Ornstein [1930]; Wang and Uhlenbeck [1945]). Specifically, the O.U. process $\{\xi_t,\ t \in \mathbb{R}\}$ satisfies the following stochastic equation in (2.1), driven by standard bilateral Wiener process $W = \{W_t,\ t \in \mathbb{R}\}$ (see also Doob [1942]):

$$d\xi_t = \theta\left(\mu - \xi_t\right) dt + \sigma dW_t, \quad \theta,\ \sigma > 0, \quad t \in \mathbb{R}. \tag{2.1}$$

  Note that the O.U. process is the only non–trivial stochastic process which is stationary, Gaussian and Markovian. We suggest characterizing the process reflected in (2.1) as ARH(1) and AR$\mathcal{C}$(1) processes. Our proposal revolves around establishing a consistent estimator of the scale parameter θ, characterizing the autocorrelation operator. Thereby, the final goal has been to enlighten the theoretical results derived, in terms of a simulation study, illustrating, in particular, the normality and asymptotic efficiency of the MLE of θ (see Kleptsyna and Breton [2002]).

- **Appendix A3.** Once formulated in Bosq [1991] the state equation of a zero–mean ARH(1) linear model, the derivation of sufficient conditions, for the weak– and strong– consistency of the estimators of the functional parameters, has constituted a central topic in the literature, such as in the monograph by Bosq [2000]. From these results, functional prediction in a consistent way was performed. The weakest operator norm is mostly considered. That is, the supremum norm in the space of bounded linear operators. Appendix A3 has as primary goal the derivation of the convergence in the mean–square sense, in the Hilbert–Schmidt norm, to the theoretical operator ρ, of its componentwise estimator (denoted as $\widehat{\rho}_{k_n}$ in Appendix A3; see also Álvarez-Liébana et al. [2017]). Thereby, for a properly truncation parameter k_n, the following results have been pursued (see Propositions A3.3.1–A3.3.2 and Remark A3.3.3 in Appendix A3 below):

$$\lim_{n\to\infty} \mathrm{E}\left\{\|\rho-\widehat{\rho}_{k_n}\|^2_{\mathcal{S}(H)}\right\}=0, \quad \lim_{n\to\infty} \mathrm{E}\left\{\|(\rho-\widehat{\rho}_{k_n})(X_n)\|_H\right\}=0.$$

The referred diagonal design is intended to deal with the *curse of dimensionality*, a central issue around which functional time series revolves. Beyond the classical FPCA and FPLSR techniques, one of the most recurrent strategies in the parametric context has been the use of wavelet bases, which are well suited for providing almost diagonal spectral decompositions of the covariance operators. Relying on these sparsity and whitening properties of unconditional bases (see Nason [2008]; Wand and Ormerod [2011]), Appendix A3 is devoted to provide a significant dimension reduction.

- **Appendix A4.** The work published in Álvarez-Liébana and Ruiz-Medina [2017] (see Appendix A4), is mainly aimed at establishing an explicit class of matrix covariance operators with non–separable point spectra, characterizing its functional entries, to model the temporal correlation of the error term, in the FANOVA model introduced in Ruiz-Medina [2016]. Specifically, in that work, separable point–spectral covariance structures were adopted for the correlated functional error components, extending the work by Zoglat [2008] (where independent and homoscedastic error components were assumed). Particularly, we propose a multivariate Hilbert–valued fixed effect model with ARH(1) errors (see equations (A4.2)–(A4.3) in Appendix A4 below):

$$\begin{aligned}
\boldsymbol{Y}(\cdot) &= \boldsymbol{X}\boldsymbol{\beta}(\cdot)+\boldsymbol{\varepsilon}(\cdot), \quad \boldsymbol{X}\in\mathbb{R}^{n\times p},\ \boldsymbol{\beta}(\cdot)\in H^p,\ \boldsymbol{Y}(\cdot)=[Y_1(\cdot),\dots,Y_n(\cdot)]^T\in H^n,\\
\boldsymbol{\varepsilon}(\cdot) &= [\varepsilon_1(\cdot),\dots,\varepsilon_n(\cdot)]^T\in H^n, \quad \varepsilon_m(\cdot)=\rho(\varepsilon_{m-1})(\cdot)+\nu_m(\cdot), \quad m\in\mathbb{Z},
\end{aligned}$$

Developments in Appendix A4 are also encouraged to attach a second objective, namely, to solve a second gap: the numerically implementation of the functional statistical test proposed in Ruiz-Medina [2016], formulated in terms of a statistics characterized by an infinite–dimensional non–computable chi–squared distribution, under the null hyphotesis. With this objective, the random–direction based testing framework considered in Cuesta-Albertos et al. [2007] is adopted, in a multivariate framework. As a final purpose, the flexibility of our approach is tested in a real–data example on fMRI analysis.

- **Appendix A5.** In the parametric ARH(1) estimation context, the Hilbert–Schmidt condition over the autocorrelation has commonly been imposed (see, e.g., Bosq [2000]; Guillas [2001]; Mas [1999]). The main purpose of Appendix A5, whose results are published in Ruiz-Medina and Álvarez-Liébana [2017], has consisted in providing an alternative ARH(1) estimation framework. Specifically, a more flexible class of standard Gaussian ARH(1) processes has been studied, characterized by a symmetric and compact autocorrelation operator ρ that might not satisfy the Hilbert–Schmidt condition, as usual, whose eigenvalues have an accumulation point at one. A second goal draws forth the results displayed in Appendix A5. We have also aimed the introduction of a Bayesian framework in the componentwise estimation of the autocorrelation operator of an ARH(1) process. The asymptotic efficiency, and the asymptotic equivalence of both, classical and Bayesian componentwise estimators, is proved as well.

- **Appendix A6.** Up to our knowledge, there exists no proposals for the strongly–consistent estimation in the Hilbert–Schmidt and trace operator norms. The main objective of the article (under minor revision) by Ruiz-Medina and Álvarez-Liébana [2018a] (see also Appendix A6) has been to demonstrate that the componentwise estimator of ρ established in Bosq [2000] also satisfies the strong consistency in the mentioned alternative operator norms, when ρ is a Hilbert–Schmidt or trace operator, respectively. The aforementioned dimension reduction challenge, addressed when the trace operator norm is adopted (see Chapter 1), is also the guiding thread in Appendix A6.4, mainly aimed at providing a new diagonal componentwise strongly–consistent estimator of ρ, which is assumed to be compact but not necessarily symmetric. Specifically, it has been supposed that ρ can be defined in terms of its singular value decomposition, associated with right and left eigenvectors systems $\{\psi_j,\ j \geq 1\}$

and $\left\{\widetilde{\psi}_j,\ j \geq 1\right\}$, respectively, as follows, adopting the notation displayed in Appendix A6:

$$\rho(x) = \sum_{j=1}^{\infty} \rho_j \langle x, \psi_j \rangle_H \widetilde{\psi}_j, \quad \rho(\psi_j) = \rho_j \widetilde{\psi}_j, \quad \rho_j \in \mathbb{C}. \tag{2.2}$$

The diagonal design reached in (2.2) constitutes an alternative to the classical parametric dimension reduction techniques for functional linear processes.

- **Appendix A7.** The survey presented in Álvarez-Liébana [2017] (submitted; see also Appendix A7) pretends to provide the beginners with a complete overview on the inference of functional time series valued in Hilbert spaces. Thereby, a comparative study between the most remarkable methodologies has been performed. From the aspects widely discussed in the referred simulation study, the reader can acquire a very comprehensive picture about how functional time series can be implemented under different settings.

- **Appendix A8.** The main goal of the proposal formulated in Ruiz-Medina and Álvarez-Liébana [2018b] (under already submitted minor revision; see also Appendix A8), has consisted of the introduction of a suitable theoretical framework for autoregressive functional estimation and prediction, in abstract Banach spaces. Nuclear spaces arise as an important special case, where the scale of fractional Besov and, in particular, Sobolev spaces, can be considered, for the description of the (regular/singular) local behaviour of the functional data analysed. In particular, the results derived hold beyond the usual regularity assumptions, characterizing the functions lying in the spaces $\mathcal{C}([0,1])$ and Skorokhod spaces $\mathcal{D}([0,1])$ (see Bosq [2000]; Hajj [2011]). Our methodology can also be interpreted as an extension of the approach formulated in Labbas and Mourid [2002], where strongly–consistent estimation results were achieved in the norm of $\mathcal{L}(\widetilde{H})$, being $\widetilde{H}$ a continuous extension of B built by the Kuelb's Lemma (see [Kuelbs, 1970, Lemma 2.1]). This extension has played a crucial role in our approach. Specifically, as discussed in Chapter 3, the methodology applied is based on the following continuous inclusions:

$$\mathcal{H}(X) \hookrightarrow \widetilde{H}^* \hookrightarrow B^* \hookrightarrow H \hookrightarrow B \hookrightarrow \widetilde{H} \hookrightarrow [\mathcal{H}(X)]^*,$$

where, according to the notation in Ruiz-Medina and Álvarez-Liébana [2018b], $\hookrightarrow$ denotes a continuous embedding, B^* and $\widetilde{H}^*$ are the topological dual spaces of B and $\widetilde{H}$, respectively, $\mathcal{H}(X)$ is the Reproducing Kernel Hilbert Space associated with the autocovariance operator of the extended ARB(1) process, and $B^* \hookrightarrow H \hookrightarrow B$ constitutes a Rigged–Hilbert–Space structure.

Algebra is nothing more than geometry, in words; geometry is nothing more than algebra, in pictures

Sophie Germain (1st April 1776 - 27th June 1831)

3 METHODOLOGY

We herein describe broadly the methodology followed in Appendices A1–A8, aimed at reaching the objectives previously exposed in Chapter 2, providing the results and conclusions argued in Chapter 4 and Chapters 5–6 below, respectively.

The methodology adopted in Appendices A1–A8 is now described, following the notation already introduced in Chapter 1.

- **Appendix A1.** Inspired by the work of Ferraty and Vieu [2006] (see the R software freely available at `https://www.math.univ-toulouse.fr/~ferraty/SOFTWARES/NPFDA/index.html`) and the supervised nonparametric kernel–based curve classification methodology therein proposed, the membership probabilities in our classification approach are assigned as follows (see equations (A1.3)–(A1.5) in Appendix A1):

$$y(\chi) = \arg\max_{g \in \overline{G}} p_g(\chi), \quad p_g(\chi) = \mathcal{P}(Y = g | \boldsymbol{\chi} = \chi) = \mathrm{E}\{\mathbf{1}_{Y=g} | \boldsymbol{\chi} = \chi\}.$$

$$\widehat{p}_g(\chi) \equiv \widehat{p}_{g,h}(\chi) = \frac{\sum_{i=1}^{m} K\left(\frac{d(\chi, \chi_i)}{h(\chi)}\right) \mathbf{1}_{y_i = g}}{\sum_{i=1}^{m} K\left(\frac{d(\chi, \chi_i)}{h(\chi)}\right)} = \frac{\sum_{\{i:\, y_i = g\}} K\left(\frac{d(\chi, \chi_i)}{h(\chi)}\right)}{\sum_{i=1}^{m} K\left(\frac{d(\chi, \chi_i)}{h(\chi)}\right)},$$

from a sample of functional random variables $\{\chi_i,\ i = 1, \ldots, m\}$ and their class memberships $\{y_i,\ i = 1, \ldots, m\} \subset \overline{G} = \{1, \ldots, g\}$, being $K(\cdot)$ a kernel function. The more relevant choice, jointly with the bandwidth $h(\chi)$, depending on χ, has been to decide which distances $d(\cdot, \cdot)$ must be implemented. In the case of curves classification, semi–metrics based on Functional Principal Component Analysis (FPCA) and Functional Partial Least Squares Regression (FPLSR), as well as a semi–metric based on derivatives, have usually been considered, and herein implemented, in terms of a suitable numerical approximation (see also details in Febrero-Bande et al. [2017]).

  FPCA and FPLSR semi–metrics have been extended to the context of random functions with n–dimensional support (such as surfaces). For this purpose, since FPCA and FPLSR semi–metrics are approximated in terms of a numerical integral, an extension of the so–called Smolyak (univariate) quadrature rule of order k, to the integration of n–dimensional random functions, has been implemented (see equation (A1.9) and Definition A1.3.2 in Appendix A1.3.2 below):

$$I^n(f) \simeq \bigotimes_{j=1}^{n} U_{l_j}^{(j)}(f) = Q_k^n, \quad l_j \leq k, \quad f \in \mathcal{C}^r\left(\prod_{j=1}^{n} I_j\right), \quad I \subset \mathbb{R},$$

where $\left\{U_{l_j}^{(j)},\ j=1,\dots,n\right\}$ denotes a sequence of univariate k_{l_j}–point quadrature rules with $k_{l_j} = 2^{l_j-1}-1$, for each $j = 1,\dots,n$. These univariate quadrature rules, at dimension j, are given by

$$I\left(f_j\right) \simeq U_{l_j}^{(j)}\left(f_j\right) := \sum_{h=1}^{k_{l_j}} w_h^{(j)} f_j\left(x_h^{(j)}\right), \quad f_j \in \mathcal{C}^r\left(I_j\right), \tag{3.1}$$

verifying $I_p = U_{l_j}^{(j)}(p)$, with p being a polynomial of degree at most k_{l_j}. In equation (3.1), the sets $\left\{w_h^{(j)},\ h=1,\dots,k_{l_j}\right\}$ and $\left\{x_h^{(j)},\ h=1,\dots,k_{l_j}\right\}$ denote, respectively, the weights and the nodes provided by the univariate rule $U_{l_j}^{(j)}$, at dimension j, for each $j = 1,\dots,n$. Particularly, we will focus on the Trapezoidal and the Clenshaw–Curtis univariate quadrature rules (see Gerstner [2007]). Thereby, we are able to implement semi–metrics for measuring the closeness between objects with n–dimensional support.

- **Appendix A2.** The methodology herein adopted, on the estimation and prediction of an O.U. process, has been formulated in the framework of ARH(1) and AR$\mathcal{C}$(1) processes (see, e.g., Bosq [2000]). Specifically, applying the method of separation of variables, the O.U. process can be rewritten as the following zero–mean process, with $\sigma = 1$ (see Álvarez-Liébana et al. [2016] and equations (A2.2)–(A2.3) in Appendix A2.1):

$$\begin{aligned} X_n(t) &= \xi_{nh+t} = \int_{-\infty}^{nh+t} e^{-\theta(nh+t-s)} dW_s = \rho_\theta\left(X_{n-1}\right)(t) + \varepsilon_n(t), \\ \rho_\theta(x)(t) &= e^{-\theta t} x(h), \quad \varepsilon_n(t) = \int_{nh}^{nh+t} e^{-\theta(nh+t-s)} dW_s, \quad \theta > 0, \quad 0 \leq t \leq h, \quad n \in \mathbb{Z}, \end{aligned}$$

whose probability density function satisfies the well–known Fokker–Planck scalar equation (see Kadanoff [2000]). Since our autocorrelation operator depends on the unknown scale parameter θ, we have studied the maximum likelihood strongly–consistent estimator of θ proposed in [Kleptsyna and Breton, 2002, Propositions 2.2–2.3] and [Kutoyants, 2004, p. 63 and p. 117] (see equation (A2.7) in Appendix A2.2):

$$\widehat{\theta}_T = \frac{1 + \frac{\xi_0^2}{T} - \frac{\xi_T^2}{T}}{\frac{2}{T}\displaystyle\int_0^T \xi_t^2 dt}, \quad T > 0.$$

On the one hand, in the context of ARH(1) processes, a real separable Hilbert space is stated as

$$H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right), \quad \langle f, g\rangle_H = \sqrt{\int_0^h f(t)\overline{g(t)}dt + f(h)\overline{g(h)}}, \tag{3.2}$$

being λ and $\delta(h)$ the Lesbesgue and Dirac measures (at point h), respectively, while $\beta_{[0,h]}$ represents the Borel $\sigma-$algebra generated by the subintervals included in $[0,h]$. Note that the Hilbert space in (3.2) determines a set of equivalence classes such that $f \sim_{\lambda+\delta_{(h)}} g$ as long as

$$\left(\lambda + \delta_{(h)}\right)\left(\{t: \ f(t) \neq g(t)\}\right) = 0.$$

On the other hand,

$$B = \mathcal{C}\left([0,h]\right), \quad \|f\|_B = \sup_{0\leq x\leq h} |f(x)|,$$

will be identified as the Banach space.

The strategy adopted in the simulation study is based on the so–called Euler–Maruyama's approach (see, among others, Kloeden and Platen [1992]), discretizing the stochastic linear differential equation as follows:

$$\widehat{\xi}_{i+1} = \widehat{\xi}_i - \theta\widehat{\xi}_i + \Delta W_i, \widehat{\xi}_0 = 0, \quad i = 0, 1, \ldots, p, \quad \Delta W_i \sim \sqrt{\Delta t}\mathcal{N}\left(0,1\right),$$

being $\Delta t = 0.02$ the discretization step and $0 = t_0 < t_1 < \ldots < t_p = T$ the discretized interval, in which $\widehat{\xi}$ is valued.

- **Appendix A3.** Aimed at providing a rate of convergence in the mean–square sense, in the Hilbert–Schmidt norm, for the estimator of the autocorrelation operator of an ARH(1) process (as well as the mean absolute convergence for the associated plug–in predictor), we propose a diagonal componentwise estimation methodology. Particularly, the following ARH(1) model is assumed (see details Álvarez-Liébana et al. [2017] in Appendix A3)

$$X_n(t) = \rho\left(X_{n-1}\right)(t) + \varepsilon_n(t), \quad \|\rho\|_{\mathcal{L}(H)} < 1, \quad \sigma_\varepsilon^2 = \mathrm{E}\left\{\|\varepsilon_n\|_H^2\right\} < \infty, \quad n \in \mathbb{Z},$$

being $\|\cdot\|_{\mathcal{L}(H)}$ the norm in the space of bounded linear operators on H. As usual, the autocovariance operator is defined as a self–adjoint, trace and positive operator (see **Assumption A1** in Appendix A3.2). Our framework is based on the assumption that the autocorrelation operator ρ belongs to the Hilbert–Schmidt and symmetric operator class (i.e., it can be diagonally decomposed), which eigenvectors $\{\phi_j,\ j \geq 1\}$ are the same as those of the autocovariance operator:

Assumption A2. (See Appendix A3.2) The autocorrelation operator ρ can be decomposed as follows:

$$\rho = \sum_{j=1}^{\infty} \rho_j \phi_j \otimes \phi_j, \quad \sum_{j=1}^{\infty} \rho_j^2 < \infty, \quad \|\rho\|_{\mathcal{L}(H)} = \sup_{j\geq 1} |\rho_j| < 1, \quad C_{\varepsilon} = \sum_{j=1}^{\infty} C_j \left(1 - \rho_j^2\right) \phi_j \otimes \phi_j, \tag{3.3}$$

being C_{ε} the autocovariance operator of the error term.

Under the assumption that $\{\phi_j,\ j \geq 1\}$ are known, our methodology revolves around estimating each one of the eigenvalues of ρ, adopting a truncated version of the spectral decomposition in (3.3). These eigenvalues can be understood as the autocorrelation parameters of projected stationary AR(1) processes $X_{n,j} = \langle X_n, \phi_j \rangle_H$, for each $j \geq 1$ and $n \in \mathbb{Z}$. Hence, the infinite–dimensional ARH(1) estimation problem of ρ has been dropped to the moment–based parameter estimation of k_n real–valued AR(1) processes, as follows (see Álvarez-Liébana et al. [2017] and equations (A3.15)–(A3.16) in Appendix A3.3):

$$\widehat{\rho}_{k_n} = \sum_{j=1}^{\infty} \widehat{\rho}_{n,j} \phi_j \otimes \phi_j, \quad \widehat{\rho}_{n,j} = \frac{n}{n-1} \frac{\displaystyle\sum_{i=0}^{n-2} \langle X_i, \phi_j \rangle_H \langle X_{i+1}, \phi_j \rangle_H}{\displaystyle\sum_{i=0}^{n-1} \langle X_i, \phi_j \rangle_H^2},$$

for a suitable truncation parameter k_n, ensuring the desirable asymptotic properties of our estimator. Thus, **Assumptions A1–A2** imposed in Appendix A3.2, jointly with an optimal k_n, and a condition over the second–order moments of the projections into $\{\phi_j,\ j \geq 1\}$ (see **Assumptions A3–A4** in Appendix A3.3), are imposed. Concerning the methodology carried out in the comparative study (see Appendix A3.5.2), we have compared the accuracy of the referred approach with those given in Bosq [2000]; Guillas [2001], testing sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and different discretization steps and truncation rules. For smaller sample sizes, the approaches in Antoniadis and Sapatinas [2003]; Besse et al. [2000] are also tested.

- **Appendix A4.** We propose a methodology aimed at extending the Hilbertian fixed effect model formulated in Ruiz-Medina [2016], allowing a non–separable point spectrum of the cross–covariance operator of the multivariate functional error term, which, in this case, it displays a temporal autoregressive dynamics of order one. In **Assumptions A0–A1** imposed in Álvarez-Liébana and Ruiz-Medina [2017] (also displayed in Appendix A4.2 below), we consider the orthogonality condition $\mathrm{E}\left\{\langle \varepsilon_i, \phi_k\rangle_H \langle \varepsilon_j, \phi_p\rangle_H\right\} = \delta_{k,p}$, for each $i, j, k, p \geq 1$, jointly with a semiorthogonal non–square design matrix (i.e., $\mathbf{X}^T\mathbf{X} = \mathbf{Id}_p$, with $\mathbf{X} \in \mathbb{R}^{n\times p}$). Under the referred conditions, we have computed a generalized componentwise least–squares estimator of $\boldsymbol{\beta}(\cdot)$, in the norm of the Reproducing Kernel Hilbert Space (RKHS). Aimed at compensating the divergence of the inverse of the eigenvalues of the covariance operators, a linear transformation $\mathbf{W}$ has been applied, providing the almost surely finiteness of both, the explained and the residual variability (see equations (A4.13)–(A4.15) in Appendix A4.2):

$$\mathbf{WY} = \mathbf{WX}\beta + \mathbf{W}\varepsilon, \quad \mathbf{W} = \begin{pmatrix} \sum_{k=1}^{\infty} w_{k11}\phi_k \otimes \phi_k & \dots & \sum_{k=1}^{\infty} w_{k1n}\phi_k \otimes \phi_k \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{\infty} w_{kn1}\phi_k \otimes \phi_k & \dots & \sum_{k=1}^{\infty} w_{knn}\phi_k \otimes \phi_k \end{pmatrix}.$$

Concerning testing the significance of functional multivariate fixed effect parameters $\boldsymbol{\beta}(\cdot) = [\beta_1(\cdot), \dots, \beta_p(\cdot)]^T \in H^p$, we cannot directly apply the functional significance test proposed in Ruiz-Medina [2016]. As noted in the referred paper, since the distribution of the statistics, under the null hypothesis, has been generated as a quadratic form, involving infinite–dimensional multivariate Gaussian measures, and it can be explicitly defined only through its characteristic functional, its explicit definition is not possible. The methodology adopted has therefore consisted in randomly selecting a direction $\boldsymbol{h} = (h, \ldots, h)^T_{p\times 1}$ for projecting the functional fixed effect parameters. We then come back to the formulation of a real–valued multivariate fixed effect test of significance (see Cuesta-Albertos et al. [2007] for the univariate case and Appendix A4.3 below):

$$H_0^{\boldsymbol{h}} : \langle \beta_1 (\cdot), h(\cdot) \rangle_H = \ldots = \langle \beta_p (\cdot), h(\cdot) \rangle_H, \tag{3.4}$$

If $H_0 : \beta_1 (\cdot) = \ldots = \beta_p (\cdot)$ fails, then, for μ–almost every function $\boldsymbol{h} \in H^p$, $H_0^{\boldsymbol{h}}$ in (3.4) also fails. Thus, a statistical test at level α to test $H_0^{\boldsymbol{h}}$ is equivalent to a statistical test at the same level α to test H_0.

With the purpose of illustrating and motivating the flexibility of our proposal, the fMRI response to external stimuli (brain is scanned at 16 depth levels, constituted each of them by a grid of 64×64 tridimensional pixels) is analysed, and the significance of functional fixed effect parameters has been tested, adapting the software implemented by Worsley et al. [2002].

- **Appendix A5.** The methodology established in Appendix A5 lies in the more flexible setting of conditions imposed, when Hilbert–Schmidt assumption over the autocorrelation operator might not be verified, providing an alternative ARH(1) framework. With this aim, we will attempt to compensate the decay rate of the eigenvalues $\{\rho_j, \ j \geq 1\}$ of the autocorrelation operator ρ by the regularity of the autocovariance operator of the innovation process (see **Assumption A1** in Appendix A5.2), whereas the autocovariance operator C is symmetric, strictly positive and belonging to the trace class (see **Assumption A2** in Appendix A5.2). Both C and ρ can be diagonally decomposed, in a weak–sense, in terms of an orthonormal system of eigenvectors $\{\phi_j, \ j \geq 1\}$, being $\{C_j, \ j \geq 1\}$ the eigenvalues of C and $\{\rho_j, \ j \geq 1\}$ the eigenvalues of ρ. Under **Assumptions A1–A2** imposed in Ruiz-Medina and Álvarez-Liébana [2017a] and Appendix A5.2, our estimation problem can be formulated in a componentwise way, being $\left\{\sigma_j^2 = \left(1 - \rho_j^2\right) C_j, \ j \geq 1\right\}$ the eigenvalues of the covariance operator of the innovation process. These eigenvalues must compensate the slow decay rate of eigenvalues $\{\rho_j, \ j \geq 1\}$ of ρ (see **Assumption A2B** below, already appeared in Ruiz-Medina et al. [2016]).

Assumption A2B. The following asymptotic behaviour is held $\frac{\sigma_j^2}{C_j} = \mathcal{O}(j^{-1-\gamma})$, with $\gamma > 0$ and $\frac{\sigma_j^2}{C_j} \leq 1$, for each $j \geq 1$, leading to $\lim_{j\to\infty} \rho_j^2 = 1$.

We also assume that the cross–covariance operator is diagonally decomposed in terms of $\{\phi_j,\ j \geq 1\}$ (see **Assumption A4** in Appendix A5.2). Under **Assumptions A1, A2, A2B** and **A4** regarded in Ruiz-Medina and Álvarez-Liébana [2017a], the asymptotic efficiency of the componentwise estimator of ρ

$$\widehat{\rho}_n = \sum_{j=1}^{\infty} \widehat{\rho}_{n,j}\phi_j \otimes \phi_j, \quad \widehat{\rho}_{n,j} = \frac{\sum_{i=0}^{i-1} \langle X_i, \phi_j\rangle_H \langle X_{i+1}, \phi_j\rangle_H}{\sum_{i=0}^{i-1} \langle X_i, \phi_j\rangle_H^2} = \frac{\sum_{i=0}^{i-1} X_{i,j}X_{i+1,j}}{\sum_{i=0}^{i-1} X_{i,j}^2}$$

is achieved. Our methodology is also intended to establish an alternative to the above classical componentwise estimator of ρ, based on the formulation of a Beta–prior–based Bayesian diagonal componentwise estimator $\widetilde{\rho}_n = \sum_{j=1}^{\infty} \widetilde{\rho}_{n,j}\phi_j \otimes \phi_j$, according to the notation in Ruiz-Medina and Álvarez-Liébana [2017a]. The equivalence and asymptotic efficiency in the Hilbert–Schmidt norm of both, Bayesian and classical estimators, is derived as well. Particularly, for every $j \geq 1$, the random components of the generalized maximum likelihood estimator are given by

$$\begin{aligned} \widetilde{\rho}_{n,j} &= \frac{\left[\sum_{i=1}^{n} X_{i-1,j}X_{i,j} + X_{i-1,j}^2\right]}{2\sum_{i=1}^{n} X_{i-1,j}^2} \\ &\pm \frac{\sqrt{\left[\sum_{i=1}^{n} X_{i-1,j}X_{i,j} - X_{i-1,j}^2\right]^2 - 4\sigma_j^2\left[\sum_{i=1}^{n} X_{i-1,j}^2\right][2-(a_j+b_j)]}}{2\sum_{i=1}^{n} X_{i-1,j}^2}, \end{aligned} \tag{3.5}$$

assuming the following Beta prior distribution

$$\rho_j \sim \mathcal{B}\left(a_j, b_j\right),$$

with shape parameters (a_j, b_j) satisfying $a_j + b_j \geq 2$, to ensure the asymptotic efficiency, and then, $\mathrm{E}\left\{\rho_j\right\} = \frac{a_j}{a_j+b_j}$. Note that the Bayesian methodology proposed is based on the assumption that the projections $\{\theta_j,\ j \geq 1\}$, into $\{\phi_j,\ j \geq 1\}$, of the unknown functional parameter θ satisfy $\theta_j \perp \{X_{n,k},\ n \geq 1,\ k \neq j\}$. Under such an assumption, the asymptotic equivalence of both, frequentist and Bayesian proposals, also holds in our infinite-dimensional framework (see Theorems A5.4.1–A5.4.2 in Appendix A5.4).

- **Appendix A6.** The methodology adopted in the first part of Appendix A6 is devoted to demonstrate that essentially, under the same conditions assumed in [Bosq, 2000, Chapter 8], for the case of unknown eigenvectors, the componentwise estimator of ρ converges to the theoretical functional value almost surely, not only in the norm of bounded linear operators but also in the norm of Hilbert–Schmidt and trace operators. This estimator was explicitly formulated in Ruiz-Medina and Álvarez-Liébana [2018a] as follows:

$$\widetilde{\rho}_{k_n}(x) = \left(\widetilde{\Pi}^{k_n} D_n C_n^{-1} \widetilde{\Pi}^{k_n}\right)(x) = \left(\sum_{j=1}^{k_n} \frac{1}{C_{n,j}} \langle x, \phi_{n,j}\rangle_{\widetilde{H}} \widetilde{\Pi}^{k_n} D_n(\phi_{n,j})\right), \quad x \in H, \tag{3.6}$$

where $\widetilde{\Pi}^{k_n}$ is the orthogonal projector into the empirical eigenvectors $\{\phi_{n,j},\ j \geq 1\}$ of C, associated with eigenvalues $\{C_{n,j},\ j \geq 1\}$ of the empirical autocovariance operator C_n, with n denoting the functional sample size. Note that D_n denotes the empirical cross–covariance operator (see Appendix A6). According to the framework therein proposed, we have assumed the almost surely boundedness of the initial condition (see **Assumption A1** in Appendix A6.2). For n large enough, the strictly positiveness of the empirical estimator of the autocovariance operator also holds (see **Assumption A2** in Appendix A6.2). Under **Assumptions A1–A2** imposed in the referred appendix, estimator in (3.6) is proved to be strongly–consistent in the trace norm $\|\cdot\|_1$ and in the Hilbert–Schmidt norm $\|\cdot\|_{\mathcal{S}(H)}$, when ρ is a trace or Hilbert–Schmidt operator, respectively (see Theorem A6.3.1 and Remark A6.3.1). In these results, a suitable value of the truncation parameter k_n has been adopted (see equation (A6.2) and Lemma A6.3.1 in Appendix A6.3 below):

$$k_n \Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right), \quad \Lambda_{k_n} = \sup_{1 \leq j \leq k_n} (C_j - C_{j+1})^{-1},$$

being $\{C_j,\ j \geq 1\}$ the set of eigenvalues of C. The trace norm $\|\cdot\|_1$ of an operator $\mathcal{K}$ is herein defined as

$$\|\mathcal{K}\|_1 = \sum_{j=1}^{\infty} \left\langle \sqrt{\mathcal{K}^* \mathcal{K}}(v_j), v_j \right\rangle_H,$$

being $\{v_j,\ j \geq 1\}$ an orthonormal basis of H.

When ρ is compact but not Hilbert–Schmidt, not trace, neither symmetric operator, a new diagonal componentwise estimator is also formulated (see Ruiz-Medina and Álvarez-Liébana [2018a] and equation (A6.5) in Appendix A6.4), based on the right and left eigenvectors of ρ involved in its singular value decomposition, denoted as $\{\psi_j, j \geq 1\}$ and $\left\{\widetilde{\psi}_j, j \geq 1\right\}$, respectively:

$$\rho(\cdot) = \sum_{j=1}^{\infty} \rho_j \langle \cdot, \psi_j \rangle_H \widetilde{\psi}_j, \quad \rho(\psi_j) = \rho_j \widetilde{\psi}_j, \quad \rho_j \in \mathbb{C}, \tag{3.7}$$

with $\{\rho_j,\ j \geq 1\}$ being the associated singular values. The spectral decomposition in (3.7) is performed under the compactness of ρ. From the trace property of the cross–covariance operator D (see **Assumption A3** in Appendix A6.4), the singular value decompositions of D and its empirical estimator D_n can also be considered, for n sufficiently large:

$$D(\cdot) = \sum_{j=1}^{\infty} d_j \langle \cdot, \varphi_j \rangle_H \widetilde{\varphi}_j, \quad D_n(\cdot) = \sum_{j=1}^{\infty} d_{n,j} \langle \cdot, \varphi_{n,j} \rangle_H \widetilde{\varphi}_{n,j},$$

for the singular values $\{d_j,\ j \geq 1\}$ and $\{d_{n,j},\ j \geq 1\}$, associated with the right and left theoretical eigenvectors $\{\varphi_j,\ j \geq 1\}$ and $\{\widetilde{\varphi}_j,\ j \geq 1\}$, respectively, and their empirical counterparts

$\{\varphi_{n,j},\ j \geq 1\}$ and $\{\widetilde{\varphi}_{n,j},\ j \geq 1\}$. From [Bosq, 2000, Lemma 4.2, p. 103], for n large enough,

$$D_n C_n^{-1}(\cdot) = \sum_{j=1}^{n} \widehat{\rho}_{n,j} \langle \cdot, \psi_{n,j} \rangle_H \widetilde{\psi}_{n,j} \tag{3.8}$$

where $\{\psi_{n,j}, j \geq 1\}$ and $\left\{\widetilde{\psi}_{n,j}, j \geq 1\right\}$ denote the right and left empirical eigenvectors of ρ, and $\{\widehat{\rho}_{n,j}, j \geq 1\}$ its associated singular values. As usual, the estimator of ρ, denoted as $\widehat{\rho}_{k_n}$, is defined as a suitable k_n–dependent truncated version of (3.8). The strong consistency of $\widehat{\rho}_{k_n}$ is derived under **Assumptions A1–A4** and the following conditions on the separation of the modulus of the eigenvalues of ρ :

$$k_n \Lambda_{k_n}^{\rho} = o\left(\frac{1}{\|D_n C_n^{-1} - DC^{-1}\|_{\mathcal{L}(H)}}\right), \quad \Lambda_{k_n}^{\rho} = \sup_{1 \leq j \leq k_n} \left\{\left(|\rho_j|^2 - |\rho_{j+1}|^2\right)^{-1}\right\}.$$

- **Appendix A7.** As usual in surveys, a complete review has been developed in Appendix A7 (see Álvarez-Liébana [2017]) throughout the main references existing in the literature about the parametric componentwise estimation and prediction of ARH processes (see Appendix A7.2), in terms of the eigenvectors of the autocovariance operator, as well as in terms of alternative bases (see Appendix A7.4). Based on the works by Damon and Guillas [2002, 2005]; Guillas [2002]; Marion and Pumo [2004], among others, extensions of the classical ARH(1) model are covered in Appendix A7.3. Beyond the stiffness of parametric approaches, alternative nonparametric and semi–parametric methodologies have been also sketched.

  In addition to the detailed overview and comparative study implemented, Appendix A7.7 is intended to provide the conditions required on the formulation of a diagonal componentwise strongly–consistent estimator of the autocorrelation operator of an ARH(1) process, achieving an important dimension reduction. The methodology adopted is based on the strictly positiveness of the autocovariance operator and the symmetry and Hilbert–Schmidt condition over the autocorrelation, which admits a diagonal spectral decomposition in terms of the eigenvectors of the autocovariance operator (see **Assumptions A1–A2** in Appendix A7.7 below). The following assumption concerning its empirical eigenvalues $\{C_{n,j},\ j = 1, \ldots, k_n\}$ is imposed (being k_n a suitable truncation parameter), jointly with the almost surely boundedness of the initial condition (see **Assumption A3** in Appendix A7.7).

Assumption A4. $C_{n,k_n} > 0\ a.s$, where $k_n < n$ is a truncation parameter with $\lim_{n\to\infty} k_n = \infty$.

The componentwise estimator proposed, under **Assumptions A1–A4**, is given as follows (see Álvarez-Liébana [2017]):

$$\widetilde{\rho}_{k_n} = \sum_{j=1}^{k_n} \widetilde{\rho}_{n,j}\phi_{n,j} \otimes \phi_{n,j}, \quad \widetilde{\rho}_{n,j} = \frac{D_{n,j}}{C_{n,j}} = \frac{n}{n-1}\frac{\sum_{i=0}^{n-2} \widetilde{X}_{i,n,j}\widetilde{X}_{i+1,n,j}}{\sum_{i=0}^{n-1} \widetilde{X}_{i,n,j}^2},\ j \geq 1,\ n \geq 2,$$

being $\widetilde{X}_{i+1,n,j} = \langle X_i, \phi_{n,j}\rangle_H$, for each $i = 0, \ldots, n-1$, $n \geq 2$ and $j \geq 1$. Its strong consistency has been provided in Proposition A7.7.1 below (see also [Bosq, 2000, Chapter 8.3], in the non-diagonal point spectral scenario).

- **Appendix A8.** Kuelb's Lemma (see [Kuelbs, 1970, Lemma 2.1]) formulates a solution to overcome the difficulties arising in normed spaces, by the lack of the orthogonality condition associated with the existing inner–product in Hilbert spaces. Our approach exploits the solution given in the referred lemma, considering a separable Hilbert space $\widetilde{H}$, where the Banach space studied is continuously embedded. Thus, the extended version of our ARB(1) process X is defined as follows (see Appendices A8.2–A8.3):

$$X_n \underset{\widetilde{H}}{=} \sum_{j=1}^{\infty} \langle X_n, v_j\rangle_{\widetilde{H}} v_j\ a.s., \quad \langle x, y\rangle_{\widetilde{H}} = \sum_{n=1}^{\infty} t_n F_n(x) F_n(y), \quad x, y \in \widetilde{H},$$

for any orthonormal basis $\{v_j,\ j \geq 1\}$ of $\widetilde{H}$, being $\{x_n,\ n \in \mathbb{N}\} \subset B$ a dense sequence, under the construction of a sequence $\{F_n,\ n \in \mathbb{N}\}$, belonging to the dual space B^*, such that $F_n(x_n) = \|x_n\|_B$ and $\|x\|_B = \sup_{n\geq 1} |F_n(x)|$. Here, $\{t_n,\ n \in \mathbb{N}\}$ is an absolute summable sequence of positive numbers, with sum equal to 1.

The elements appearing in the above definition of $\widetilde{H}$ are considered in the construction of a Rigged–Hilbert–Space structure (B^*, H, B) (also known as a Gelfand triple). Specifically, H is defined as a closed Hilbert subspace of B, constituted by the elements of B satisfying

$$\left\{ x \in B : \quad \sum_{n=1}^{\infty} [F_n(x)]^2 < \infty \right\}.$$

Thus, the inner product in H is given by

$$\langle x, y \rangle_H = \sum_{n=1}^{\infty} F_n(x) F_n(y), \quad x, y \in H.$$

Therefore, $B^* \hookrightarrow H \hookrightarrow B$ and the following continuous embeddings are established (see Ruiz-Medina and Álvarez-Liébana [2018b] and Lemma A8.3.1 in Appendix A8.3 below):

$$\mathcal{H}(X) \hookrightarrow \widetilde{H}^* \hookrightarrow B^* \hookrightarrow H \hookrightarrow B \hookrightarrow \widetilde{H} \hookrightarrow [\mathcal{H}(X)]^*, \tag{3.9}$$

where $\mathcal{H}(X)$ denotes the RKHS generated by the autocovariance operator C of the extended ARB(1) process. Specifically,

$$\begin{aligned} \widetilde{H}^* &= \left\{ x \in B; \sum_{n=1}^{\infty} \frac{1}{t_n} \left\{F_n(x)\right\}^2 < \infty \right\}, \quad \langle f, g \rangle_{\widetilde{H}^*} = \sum_{n=1}^{\infty} \frac{1}{t_n} F_n(f) F_n(g), \\ \mathcal{H}(X) &= \left\{ x \in \widetilde{H};\ \left\langle C^{-1}(x), x \right\rangle_{\widetilde{H}} < \infty \right\}, \quad [\mathcal{H}(X)]^* = \left\{ x \in \widetilde{H};\ \langle C(x), x \rangle_{\widetilde{H}} < \infty \right\}, \end{aligned}$$

where, as denoted in Ruiz-Medina and Álvarez-Liébana [2018b], $[\mathcal{H}(X)]^*$ constitutes the dual space of the RKHS $\mathcal{H}(X)$ and, as before, B denotes a separable Banach space and B^* its dual.

Additionally, to the above Hilbert–based construction, enveloping B and B^*, similar assumptions to those ones required in the Hilbert space context are assumed for the extended ARB(1) process, allowing the derivation of new parallel results in the B–norm, like, for example, the strong consistency of the empirical eigenvectors in the Banach norm (see Ruiz-Medina and Álvarez-Liébana [2018b] , and Lemmas A8.2.1–A8.3.8 and Remarks A8.3.4–A8.3.6 in Appendix A8.3). As final result, it is obtained the strong consistency, in the space of bounded linear operators on B, of the following component-wise estimator of ρ :

$$\widetilde{\rho}_{k_n}(x) = \left(\widetilde{\Pi}^{k_n} D_n C_n^{-1} \widetilde{\Pi}^{k_n}\right)(x) = \left(\sum_{j=1}^{k_n} \frac{1}{C_{n,j}} \langle x, \phi_{n,j} \rangle_{\widetilde{H}} \widetilde{\Pi}^{k_n} D_n(\phi_{n,j})\right), \quad x \in B, \quad (3.10)$$

where, as before, $\widetilde{\Pi}^{k_n}$ denotes the orthogonal projector into the empirical eigenvectors $\{\phi_{n,j},\ j \geq 1\}$ of C, associated with eigenvalues $\{C_{n,j},\ j \geq 1\}$, and D_n is the empirical cross–covariance operator.

Concerning the simulation study carried out, and based on the wavelet–based characterization of Besov spaces, we adopt the following functional spaces (see Supplementary Material provided in Ruiz-Medina and Álvarez-Liébana [2018b] and Appendix A8.8)

$$B = B^0_{\infty,\infty}([0,1]), \quad B^* = B^0_{1,1}([0,1])$$

in the scale of Besov spaces

$$\left\{\left(B^s_{p,q}, \|\cdot\|_{p,q,s}\right), \quad 1 \leq p, q \leq \infty,\ s \in \mathbb{R}\right\},$$

as well as the fractional Sobolev space $\widetilde{H} = H_2^{-\beta}([0,1])$.

I have had my results for a long time: but I do not yet know how I am to arrive at them

J. C. F. Gauss (30th April 1777 – 23rd February 1855)

4
RESULTS

THE MAIN RESULTS ACHIEVED THROUGHOUT THE CURRENT DISSERTATION WILL BE SUMMARIZED AND DISCUSSED IN THIS CHAPTER, DETAILING BOTH THEORETICAL AND NUMERICAL RESULTS ADDRESSED IN APPENDICES A1–A8. GENERAL CONCLUSIONS AND CONSEQUENCES OF THOSE RESULTS ARE DISCUSSED IN CHAPTER 5 (SEE CHAPTER 6 IN SPANISH LANGUAGE). CURRENT RESEARCH LINES CAN BE FOUND IN CHAPTER 7.

- **Appendix A1.** We have implemented an extension of the Smolyak quadrature rule to random functions with n–dimensional support, providing the nodes and weights required for its numerical integration. This numerical result is applied to the classification of uncorrelated spectrometric curves, such that a similar miss–classification average rate (MCAR) is displayed in comparison with Ferraty and Vieu [2006]. In the simulation study undertaken, in the context of supervised classification of random surfaces, two families, with very close mean functions, have been generated. Specifically, a MCAR of 0.3 is gained, when the FPCA semi–metric is performed in terms of the Clenshaw–Curtis quadrature rule, while a MCAR of 0.12 is noticed for the Trapezoidal rule. In addition, FPLSR semi–metric clearly outperforms these results. In the field of railway engineering, we classify 12 classes of deterministic irregularities (disrupted by a zero–mean Gaussian device error) and 4 classes of purely random Gaussian railway roughness. In the former scenario, a MCAR of 0.065 is reached for the Trapezoidal rule. In the second case, a MCAR of 0.35 is gotten for weakly–dependent spatial correlation models, and 0.48 for strongly–dependent models.

- **Appendix A2.** The main theoretical result provided in Álvarez-Liébana et al. [2017] (see also Appendix A2) has been the estimation and prediction of an O.U. process. From the referred insights and some auxiliary theoretical results established in Lemmas A2.2.1–A2.2.3 and Theorems A2.2.1–A2.2.2 in Appendix A2.5 below, the following asymptotic upper bounds are settled (see Proposition A2.2.1, Remark A2.2.3 and Corollaries A2.2.1–A2.2.2):

$$\begin{aligned}
\left\| \rho_\theta - \rho_{\widehat{\theta}_n} \right\|_{\mathcal{L}(H)} &\leq_{a.s.} \left| \theta - \widehat{\theta}_n \right| h \sqrt{\frac{h}{3} + 1}, \quad \mathrm{E}\left\{ \left\| \rho_\theta - \rho_{\widehat{\theta}_n} \right\|_{\mathcal{L}(H)}^2 \right\} \leq G\left(\theta, \widehat{\theta}_n, n\right), \\
\left\| \rho_\theta - \rho_{\widehat{\theta}_n} \right\|_{\mathcal{L}(B)} &\leq_{a.s.} h \left| \theta - \widehat{\theta}_n \right|, \quad \mathrm{E}\left\{ \left\| \rho_\theta - \rho_{\widehat{\theta}_n} \right\|_{\mathcal{L}(B)}^2 \right\} \leq G\left(\theta, \widehat{\theta}_n, n\right),
\end{aligned}$$

where $G\left(\theta, \widehat{\theta}_n, n\right) = \mathcal{O}\left(\frac{2\theta}{n}\right)$, as n goes to infinity. Hence, from the results in Kleptsyna and Breton [2002] in relation to θ, a strongly–consistent estimator of the θ–dependent autocorrelation operator is formulated, in the norm of $\mathcal{L}(H)$ and $\mathcal{L}(B)$, respectively (see Proposition A2.2.1 and Lemma A2.2.4). The convergence in probability of the associated plug–in predictor is directly reached, in the norm of H and B, respectively (see Corollaries A2.2.1–A2.2.2). From the simulation study, the empirical errors for the MLE of θ lie within the band $\pm 3\sqrt{\frac{2\theta}{n}}$, at least, 99.33% of the simulations, at each one of the scenarios generated. The consistency of the plug–in predictor is also enlightened.

- **Appendix A3.** Results tackled in Álvarez-Liébana et al. [2017] and Appendix A3 are focused on the formulation of a decay rate for the mean–square convergence, in the Hilbert–Schmidt norm, under **Assumptions A1–A4** imposed in Appendices A3.2–A3.3 below. On the derivation of the L^2–convergence of the estimator of ρ (denoted as $\widehat{\rho}_{k_n}$ in Álvarez-Liébana et al. [2017]), as well as the L^1–convergence of its associated plug–in predictor (see properly definitions in equations (A3.15)–(A3.16)), the following theoretical insights have been proved (see Propositions A3.3.1–A3.3.2):

$$\mathrm{E}\left\{\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{S}(H)}\right\} \leq g(n) = \mathcal{O}\left(\frac{1}{C_{k_n}^2 n}\right), \quad \mathrm{E}\left\{\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H\right\} \leq \sqrt{g(n)},$$

being k_n a suitable truncation parameter and $\{C_j,\ j \geq 1\}$ the eigenvalues of the autocovariance operator. A simulation study has reflected the performance of the above estimator and plug–in predictor. Particularly, for truncation rules $k_n = \lceil n^{1/\alpha} \rceil$, with $\alpha = 5$ and $\alpha = 6$ and sample sizes from $n = 15000$ to $n = 395000$, we obtain empirical functional mean–square errors of order 10^{-4} and absolute prediction errors of order 10^{-3}. Curves $n^{-3/4}$ and $n^{-1/3}$ are numerically fitted as the decay rates, respectively. In the comparative study implemented, when theoretical eigenvectors are known, prediction errors of order 10^{-3} are observed for our parametric approach and those ones in Bosq [2000]; Guillas [2001], but a better performance of our methodology is observed when small sample sizes are tested. In the case when the eigenvectors of C are, as usual, unknown, our approach outperforms, with errors of order 10^{-2}, those ones in Bosq [2000]; Guillas [2001]. On the other hand, for smaller sample sizes, wavelet–based predictor in Antoniadis and Sapatinas [2003] and non-parametric and penalized predictors in Besse et al. [2000] are also compared. Slightly improvements are observed for the kernel–based predictor, while the proposal by Antoniadis and Sapatinas [2003] only outperforms our predictor when a very small number of parameters must be estimated; i.e., for small truncation values of the parameter k_n.

- **Appendix A4.** Two major theoretical results achieved in Appendix A4 deserve to be mentioned. The first one has consisted in providing explicit matrix covariance operators for a multivariate H–valued fixed effects model, under a non–separable cross–covariance point spectrum design, with fast decay velocity. Specifically, a generalized least–squares estimator of the fixed effects parameters is formulated (see properly definitions in equations (A4.2)–(A4.9) in Appendix A4 below):

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\sum_{k=1}^{\infty} \widehat{\beta}_{k1}\phi_k, \ldots, \sum_{k=1}^{\infty} \widehat{\beta}_{kp}\phi_k\right)^T, \quad \left(\widehat{\beta_{k1}}, \ldots, \widehat{\beta_{kp}}\right)^T = \left(\mathbf{X}^T \boldsymbol{\Lambda}_k^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Lambda}_k^{-1}\mathbf{Y}_k, \\
\boldsymbol{Y}_k &= \left(\langle Y_1, \phi_k\rangle_H, \ldots, \langle Y_n, \phi_k\rangle_H\right)^T, \quad \boldsymbol{\Lambda}_k = \{\Lambda_{kij}\}_{i=1,\ldots,n}^{j=1,\ldots,n}, \\
\Lambda_{kii} &= \lambda_k\left(R_0\right), \quad \Lambda_{kij} = \lambda_k\left(R_1\right), \; j = i+1, \quad \Lambda_{kij} = \lambda_k\left(R_1^*\right), \; i = j+1,
\end{aligned}
$$

with R_0 and R_1 now denoting the autocovariance and cross–covariance operators of the ARH(1) process, respectively, according to the notation used in Appendix A4 (see also Álvarez-Liébana and Ruiz-Medina [2017]).

Here, $\{\phi_j,\ j \geq 1\}$ are the eigenvectors of the autocovariance operator R_0 of the error term. Particularly, in rectangular domains, the empirical functional mean–squares error of the functional estimator of fixed effect parameter vector displays a performance of order 10^{-3}, while an accuracy of 10^{-2} can be observed for the estimated functional response. In the case of circular domains, a better accuracy is attached. Values of F statistics greater than 1 can be observed at each rectangular scenario, without substantially differences between them. In contrast, F statistics reaches values of order 10^7 when circular sector domains are analysed. On the other hand, since functional significance tests in Ruiz-Medina [2016] depend on a non–computable infinite–dimensional probability distribution of the statistics, under the null hypothesis, we have implemented a multivariate extension of the random directions functional test approach in Cuesta-Albertos and Febrero-Bande [2010]; Cuesta-Albertos et al. [2007]. In rectangular domains, the null hypothesis fails for at least 99.75% of the simulations, at level $\alpha = 0.05$, from the observations generated under the models considered, for the different scenarios analysed. Similar rejection values are observed from the functional observations generated, in the case of the circular sectors. For illustration purposes, Appendix A4.5 provides a real–data application, in relation to the fMRI analysis. According to Worsley et al. [2002], brain is scanned at tridimensional pixels of dimensions $3.75 \times 3.75 \times 7\, mm$. Specifically, concerning the significance of the two–dimensional functional effects, p–values are, at most, of order 10^{-4}, in the most of marginal distributions selected (i.e., in the most of projected random directions).

- **Appendix A5.** The classical moment–based componentwise estimator $\widehat{\rho}_n$ and the Beta–prior–based Bayesian estimator $\widetilde{\rho}_n^-$ of the autocorrelation operator of an ARH(1) process (according to the notation displayed in equations (A5.15) and (A5.20)–(A5.22), respectively) are both proved to be equivalent, in the sense of the asymptotic efficiency displayed. Thus, in Ruiz-Medina and Álvarez-Liébana [2017a], under **Assumptions A1, A2, A2B** and **A4**, the following results are derived (see Theorems A5.4.1–A5.4.2 and Remarks A5.2.2–A5.2.3 in Appendix A5 below):

$$\lim_{n\to\infty} n\mathrm{E}\left\{\|\widetilde{\rho}_n^- - \rho\|^2_{\mathcal{S}(H)}\right\} = \lim_{n\to\infty} n\mathrm{E}\left\{\|\widehat{\rho}_n - \rho\|^2_{\mathcal{S}(H)}\right\} = \sum_{j=1}^{\infty}\frac{\sigma_j^2}{C_j} < \infty,$$

where $\rho_j \sim \mathcal{B}\left(a_j, b_j\right)$, denoting, as before, $\mathcal{B}\left(a_j, b_j\right)$ the beta distribution with parameters a_j and b_j, with $a_j + b_j \geq 2$, for each $j \geq 1$. Similar results have been demonstrated for the associated plug–in predictors (see Theorems A5.4.1–A5.4.2). In the simulation study undertaken, different decay rates of the eigenvalues of the autocovariance operator are analysed. In particular, three scenarios are generated. For tested sample sizes in the range $[250, 2000]$, the magnitudes of empirical functional mean–squares prediction errors are of order 10^{-3} for Examples 1–2 (considering a truncation order $k_n = 5$; see Appendices A5.5.1–A5.5.2), while errors of order 10^{-4} are displayed in Example 3 for $k_n = \lceil n^{1/4.1} \rceil$ (see Appendix A5.5.3).

- **Appendix A6.** We have derived alternative asymptotic properties of the componentwise estimator of ρ formulated in Bosq [2000]. Under **Assumptions A1–A2** imposed in Appendix A6, and $k_n\Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right)$ (see Ruiz-Medina and Álvarez-Liébana [2018a] and Theorem A6.3.1 in Appendix A6), the strong consistency in the norms of Hilbert–Schmidt and trace operators has been established, when ρ belongs to the class of Hilbert–Schmidt operators and to the class of nuclear operators, respectively. Alternatively, when ρ is compact and not necessarily symmetric, a new diagonal componentwise estimator is formulated (denoted in Appendix A6.4 as $\widehat{\rho}_{k_n}$). Under **Assumptions A1–A4**, we have (see Ruiz-Medina and Álvarez-Liébana [2018a], and Remark A6.4.1 and Theorem A6.4.1 in Appendix A6.4)

$$\|\widehat{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} \longrightarrow_{a.s.} 0, \quad n \to \infty, \qquad \text{as long as} \quad k_n\Lambda^{\rho}_{k_n} = o\left(\frac{1}{\|D_nC_n^{-1} - DC^{-1}\|_{\mathcal{L}(H)}}\right),$$

where $\Lambda_{k_n}^{\rho} = \sup_{1\leq j\leq k_n} \left\{ \left(|\rho_j|^2 - |\rho_{j+1}|^2 \right)^{-1} \right\}$, being $\{\rho_j,\ j \geq 1\}$ the singular values of ρ.

- **Appendix A7.** Appendix A7 reviews the main contributions in the ARH(1) framework, as well as provides a comparative study. Besides the wide review throughout the existing parametric, semi–parametric and nonparametric methodologies, in Appendix A7.7, the main asymptotics of a diagonal componentwise estimator of the autocorrelation operator, in the lines reflected in the monograph by [Bosq, 2000, Chapter 8], is analysed, for the case of unknown eigenvectors (see Álvarez-Liébana [2017] and Proposition A7.7.1 in Appendix A7.7). A simulation study is undertaken as well.

- **Appendix A8.** Under **Assumptions A1–A5** in Ruiz-Medina and Álvarez-Liébana [2018b], and from Lemmas A8.2.1–A8.3.8 and Remarks A8.3.4–A8.3.6 in Appendix A8, large deviations inequalities and then, the strong consistency of the componentwise estimator displayed in (3.10), are derived, in the norm of bounded linear operators on B, under suitable conditions. The strong consistency of the corresponding plug–in predictor follows. Specifically, the following main results are obtained, assuming that k_n is such that $\sqrt{k_n}\Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right)$, as $n \to \infty$ (see Theorem A8.5.1 and equation (A8.38) in Appendix A8):

$$\mathcal{P}\left(\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(B)} \geq \eta \right) \leq \mathcal{K} \exp\left(-\frac{n\eta^2}{Q_n} \right), \quad \eta > 0, \quad Q_n = \mathcal{O}\left(\left\{ C_{k_n}^{-1} k_n \sum_{j=1}^{k_n} a_j \right\}^2 \right),$$

In addition, under an extra condition over the truncation parameter, $\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(B)} \to_{a.s} 0$, as $n \to \infty$ (see Ruiz-Medina and Álvarez-Liébana [2018b] and the referred theorem in Appendix A8). The approach presented is illustrated in terms of the scale of Besov spaces of fractional order. Continuous embeddings theorems between Besov spaces are then applied. Large–sample behaviour of the ARB(1) plug–in predictor, with truncation rule $k_n = \lceil \ln(n) \rceil$, and sample sizes from 2500 to 165000, has been analysed, such that the asymptotic efficiency is raised and $n^{-1/4}$ has been numerically fitted as the rate of convergence. Aimed at displaying how the asymptotic behaviour is similar when discretely observed processes are regarded, a set of discretization steps, decreasing to zero, are implemented. The same accuracy as before has been appreciated, such that the percentage of empirical errors greater than the upper bound theoretically derived decreases to zero.

If your experiment needs a statistician, you need a better experiment

E. Rutherford (30th August 1871 - 19th October 1937)

5 CONCLUSIONS

Herein we discuss the more important conclusions derived from the theoretical and numerical results presented in Appendices A1–A8 below, and already summarized in Chapter 4. These conclusions will allow us to propose different open research lines in Chapter 7 which could be addressed in the future. Conclusions in Spanish language can be found in Chapter 6.

- **Appendix A1.** As a conclusion, we have explicitly detailed, and numerically implemented, a new proposal for the kernel–based classification of random functions whose support is n–dimensional. With this purpose, we have implemented a numerical integration method for n–dimensional supported functions, particularly applied to the computation of FPCA and FPLSR semi–metrics for random surfaces. In the light of the findings, we can conclude that the choice of the univariate quadrature rule is not as trivial as it might seem at first sight, since the accuracy between semi–metrics differs most notably when the Clenshaw–Curtis rule is tested. In fact, a better performance is gained when Trapezoidal rule is implemented. This fact may come from the definition of the Clenshaw–Curtis quadrature rule, in which expansions depending on trigonometric functions play a key role. Since FPLSR semi–metric depends not just on the explanatory variables, but also on the response variable, this gap can be easily appreciated for the FPCA semi–metric.

- **Appendix A2.** Achievements in Appendix A2 allow us to conclude that an O.U. process can be characterized as a stationary ARH(1) and ARB(1) process, with an autocorrelation operator depending on a fixed but unknown scale parameter θ. Thereby, consistent functional predictors in Hilbert and Banach spaces of the O.U. process can then be formulated, which, in particular, is of interest in the finance context. From the simulation study, the asymptotic efficiency of the MLE of θ can be numerically concluded. Particularly, the $\sqrt{n}$–strong consistency of the estimator of the autocorrelation operator is provided in both Hilbert and Banach spaces.

- **Appendix A3.** The major contribution in Appendix A3, firstly derived up to our knowledge, has been to establish the set of conditions required on the derivation of the convergence in the mean–square sense of a diagonal componentwise estimator of the autocorrelation operator of an ARH(1) process, in the norm of Hilbert–Schmidt operators, with the determination of a minimum rate of convergence. Note that the set of assumptions has been mainly focused on adopting a diagonal spectral decomposition for ρ, in terms of the eigenvectors of the autocovariance operator C, which can be found working under shrinkage scenarios inducing sparsity provided by wavelet bases. The Gaussian case addressed in Álvarez-Liébana et al. [2017] (see also Appendix A3.4 below) satisfies the assumptions imposed, in the derivation of the theoretical results. The large–sample properties obtained in Appendix A3 also describe the n–asymptotic behaviour of the point spectral tail of the autocorrelation operator. This information is loss, when other weaker norms are adopted.

- **Appendix A4.** As conclusions, we obtain that the FANOVA analysis performed is affected by the geometry of the domain, that defines the support of the functional values of the observed data, in the multivariate functional fixed effects context analysed. Regarding significance hypothesis testing, a multivariate version of Theorem 4.1 in Cuesta-Albertos et al. [2007] has been applied to implement a statistical significance test in practice. Neuroimaging analysis seems to be a suitable field of application of the theoretical results derived, as tested in the real–data example analysed.

- **Appendix A5.** A more flexible framework is considered than in Bosq [2000], regarding the point spectral asymptotic of the autocorrelation operator of an ARH(1) process. Thus, a wider class of autocorrelation operators is estimated in an asymptotic efficient way, from a classical and Bayesian perspectives. The regularity of the autocovariance operator of the innovation process plays a key role in the presented approach. The asymptotic equivalence of both, classical and Bayesian approximations, is proved, as expected. As conclusion, the regularity of the autocovariance operator of the innovation process allows the consideration of a more flexible class (probably, more singular) of autocorrelation operators, in this standard ARH(1) framework. Despite the fact that n–asymptotic efficiency is obtained, for a given percentage of the explained functional variance, a larger truncation order should be considered when slower decay rates of the eigenvalues of C are used. These differences are more evident when small sizes are given.

- **Appendix A6.** We have demostrated how the componentwise estimator of the autocorrelation operator of an ARH(1) process formulated in Bosq [2000] is strongly–consistent also in the trace and Hilbert–Schmidt norms, provided that ρ belongs to these operator classes. In the context of alternative techniques to solve the so–called *curse of dimensionality*, a novel strongly–consistent estimation technique has been proposed, assuming that ρ is compact but not necessarily Hilbert–Schmidt or symmetric, based on the singular value decomposition of ρ. Thus, an important dimension reduction is reached. The set of required conditions, concerning the truncation parameter and its relationship with the decay rate and separation of the eigenvalues of ρ, has been as well established.

- **Appendix A7.** The survey in Appendix A7 has been focused on providing the reader with a comprehensive overview about the crucial aspects concerning the estimation and prediction of Hilbertian time series. Our diagonal approach therein formulated outperforms those ones included in Bosq [2000]; Guillas [2000] when the truncation rule proposed in Bosq [2000] is used. As noticed, Guillas [2001] ends up being the best performance when the truncation rule therein proposed is fixed. In addition, even when small sample sizes are compared, a better accuracy can be appreciated, when our prediction approach is applied to pseudo–diagonal point spectral autocorrelation scenarios. For small sample sizes, only our approach and those ones formulated in Besse et al. [2000] seem to reach the strong consistency when $k_n = \lceil \ln(n) \rceil$. The penalized predictor in Besse et al. [2000] has been shown to be the most accurate, being less influenced by the regularity conditions. Despite these numerical results, we would stress, as a drawback, that methodologies in Antoniadis and Sapatinas [2003]; Besse et al. [2000] require large computational times.

- **Appendix A8.** A general abstract separable Banach context is studied in Appendix A8, beyond the space of continuous functions on an interval, with the supremum norm, and the space of right continuous functions, with limit at the left, with the Skorokhod topology. In particular, an extension of the results derived in [Bosq, 2000, Chapter 8] and Labbas and Mourid [2002] is obtained. Note that continuity (or right–continuity) is an usual minimal regularity assumption satisfied by the functions in those spaces. It is well–known that the Banach context is traditionally intended, in linear functional time series framework, to find a finer scale of norms for measuring local regularity. In our case, the opposite motivation is also exploited, since, in some practical problems (see, for example, meteorological data problems addressed by Febrero-Bande et al. [2008]; Ignaccolo et al. [2014]; Slini et al. [2006]; Stadlober et al. [2008], among others), the local singularity displayed by functional data should be measured in an accurately way. Thus, our more flexible framework leads to the strongly–consistent estimation of ARB(1) processes, whose functional values could neither be continuous nor differentiable. That is the case, for example, of the solution to integro–differential or pseudo–differential equations of fractional order. In particular, the smoothing kernel norms appearing, for example, in Besov or Sobolev spaces of negative order, allow the consideration of a wider class of autocovariance operators beyond the usual trace condition with respect to the L^2–norm. The interest of our approach, in the statistical analysis of functional time series, with values in nuclear spaces, is illustrated, as commented before, in the simulation study undertaken (see Appendix A8.6 and the Supplementary Material provided in Appendix A8.8). In particular, the scale of fractional Besov spaces is considered, and wavelet bases are selected for projection.

A scientist in his laboratory is not a mere technician: he is also a child confronting natural phenomena that impress him as though they were fairy tales

Maria Sklodowska (7th November 1867 - 4th July 1934)

6
CONCLUSIONES

Este capítulo estará dedicado a la discusión de las conclusiones derivadas de los resultados presentados en los trabajos aquí incluidos (ver Apéndices A1–A8). Dichas conclusiones nos permitirán plantear posibles líneas futuras de investigación (ver Capítulo 7), las cuales quedan fuera del alcance de la presente tesis.

- **Apéndice A1.** En el Apéndice A1 se ha detallado de forma explícita una propuesta para la clasificación no paramétrica (de tipo núcleo) de funciones aleatorias cuyo soporte es n–dimensional. Con este propósito, hemos implementado un método de integración numérica para funciones con soporte n–dimensional, particularizado al cálculo numérico de las semi–métricas FPCA y FPLSR para superficies aleatorias. En base a los resultados obtenidos se concluye la vital importancia que tiene la elección de la regla de cuadratura, decisión que no es tan trivial como pudiera antojarse a primera vista. Esta diferencia en cuanto a la calidad de la clasificación se hace más evidente cuando se aplica la regla de Clenshaw–Curtis, lo cual parece lógico ya que dicha regla de cuadratura viene definida por desarrollos en serie, en términos de funciones trigonométricas. Dado que la semi–métrica FPLSR no solo depende de las variables explicativas sino también de la variable respuesta, esta discrepancia en la proporción de mal clasificados puede apreciarse más fácilmente para la semi–métrica FPCA.

- **Apéndice A2.** Los resultados derivados en el Apéndice A2 nos permiten concluir que un proceso Ornstein–Uhlenbeck puede caracterizarse como un proceso estacionario ARH(1) y ARB(1), de forma que su operador de autocorrelación dependa de un parámetro de escala θ, fijo pero desconocido. En particular, se han formulado predictores de carácter funcional del proceso en espacios de Hilbert y Banach. Como se detalla en secciones anteriores, estos resultados pueden ser de gran interés en el contexto financiero. A partir del estudio de simulación realizado se ha obtenido de forma númerica la eficiencia asintótica del estimador de máxima verosimilitud del parámetro anteriormente denotado como θ. En particular, hemos probado la consistencia fuerte, con ratio de convergencia $\sqrt{n}$, del estimador del operador de autocorrelación involucrado, tanto en espacios de Hilbert como de Banach.

- **Apéndice A3.** La mayor contribución del Apéndice A3 ha sido la de establecer el conjunto de condiciones necesarias en la derivación de la convergencia en media cuadrática de un estimador diagonal, definido componente a componente, del operador de autocorrelación de un proceso estacionario ARH(1), en la norma de los operadores de Hilbert–Schmidt, obteniendo un ratio mínimo de convergencia. Cabe notar que el conjunto de hipótesis se ha planteado con el fin de obtener un operador de autocorrelación que admita una descomposición espectral diagonal, en términos de los autovectores del operador de autocovarianza C. Este diseño diagonal puede ser fácilmente alcanzado en escenarios donde existan representaciones espectrales diagonales (o pseudodiagonales) inducidas por descomposiciones en función de bases de tipo wavelet. Por otro lado, se demuestra como el escenario Gaussiano abordado en Álvarez-Liébana et al. [2017] (ver también Apéndice A3.4) satisface las hipótesis planteadas en la derivación de los resultados teóricos aquí expuestos, derivando así el comportamiento n–asintótico de las colas del espectro puntual del operador de autocorrelación, resultados que no se suelen inferir cuando se adoptan normas más débiles.

- **Apéndice A4.** Los resultados aquí desarrollados nos permiten deducir la importancia que tiene la geometría de los dominios cuando llevamos a cabo un análisis funcional de la varianza, ya que dicha geometría determina el soporte de los valores funcionales observados, en el modelo multivariante de efectos fijos funcionales analizado. Con respecto a los test de hipótesis para significación de los parámetros funcionales de efectos fijos, se ha formulado una versión multivariante de los resultados incluidos en el Teorema 4.1 establecido en Cuesta-Albertos et al. [2007]. La ilustración de dichos resultados se ha realizado mediante una aplicación, con datos reales, para el análisis estadístico de resonancias magnéticas de tipo fMRI.

- **Apéndice A5.** En dicho apéndice hemos considerado un conjunto de condiciones que nos proporciona un escenario más flexible que el propuesto en Bosq [2000], planteando condiciones asintóticas alternativas respecto al espectro puntual del operador de autocorrelación de un proceso ARH(1). Así, hemos conseguido estimar una clase más amplia de operadores de autocorrelación, de forma asintóticamente eficiente y desde una perspectiva tanto clásica (frequentista) como Bayesiana, jugando un papel fundamental la regularidad impuesta sobre el operador de autocovarianza de las innovaciones. Como en el caso real–valuado, hemos constatado la equivalencia asintótica de ambas aproximaciones. Podemos concluir, por tanto, que la regularidad del operador de autocovarianza de las innovaciones nos ha permitido trabajar con una clase más flexible de operadores de autocorrelación, que la comúnmente adoptada en procesos ARH(1) estándar. Los resultados del estudio de simulación nos indican que debe considerarse un orden de truncamiento mayor cuando se utilizan caídas de los autovalores de C más lentas, para un porcentaje dado de varianza funcional explicada. Estas diferencias se hacen más evidentes cuando implementamos tamaños muestrales pequeños.

- **Apéndice A6.** Hemos probado cómo el estimador componente a componente del operador de autocorrelación de un proceso ARH(1) formulado en Bosq [2000] es también fuertemente consistente en las normas traza y de Hilbert–Schmidt, siempre que ρ pertenezca a dichas clases de operadores. En particular, se ha establecido la consistencia fuerte en las tres normas de operadores. Por otro lado, en el contexto de técnicas alternativas para resolver el problema de dimensionalidad inherente, se ha propuesto una estimación diagonal fuertemente consistente del operador de autocorrelación ρ basada en la descomposición de valores singulares del mismo, asumiendo que dicho operador es compacto pero no necesariamente Hilbert–Schmidt ni simétrico, consiguiendo reducir la dimensión considerablemente. De la misma forma, se han derivado condiciones suficientes sobre el parámetro de truncamiento, de acuerdo a la caída del módulo de los valores singulares de ρ, y a la separación de los mismos, para garantizar la consistencia fuerte de dicho estimador.

- **Apéndice A7.** La revisión bibliográfica realizada en el Apéndice A7 se ha centrado en ofrecer, a los que se inician en el campo de las series temporales funcionales, una perspectiva general de los aspectos cruciales en cuanto a la estimación y predicción de series temporales Hilbertianas. El estimador diagonal propuesto obtiene mejores resultados predictivos que los formulados en Bosq [2000]; Guillas [2000] cuando se utiliza la regla de truncamiento propuesta en Bosq [2000] y se aplica a escenarios diagonales o pseudodiagonales, mientras que si usamos la regla de truncamiento establecida en Guillas [2001], el predictor allí formulado termina siendo el mejor en términos predictivos. Para tamaños muestrales pequeños, solo nuestra propuesta y aquellos formulados en Besse et al. [2000] parecen ser fuertemente consistentes cuando $k_n = \lceil \ln(n) \rceil$. De hecho, el predictor mediante penalizaciones de Besse et al. [2000] se ha demostrado numéricamente como el más preciso, siendo menos influenciado por las condiciones de regularidad. Remarcar que, como desventaja, las metodologías derivadas en Antoniadis and Sapatinas [2003]; Besse et al. [2000] requieren mucho más tiempo de computación.

- **Apéndice A8.** En el Apéndice A8 se ha adoptado el contexto de espacios de Banach abstractos separables, más allá de los espacios comúnmente estudiados, que contienen a las funciones continuas sobre un intervalo, dotados con la norma del supremo, y a las funciones continuas por la derecha con límite por la izquierda, con la topología de Skorokhod. En particular, nuestra metodología puede entenderse como una extensión de los resultados ya derivados en [Bosq, 2000, Capítulo 8] y Labbas and Mourid [2002]. Como es bien sabido, el análisis estadístico de datos funcionales con valores en espacios de Banach ha venido tradicionalmente motivado, dentro del contexto de series temporales lineales funcionales, por la búsqueda de una escala de normas más fina que la topología usual L^2 en el contexto de espacios de Hilbert separables, para dotarnos de herramientas que nos permitan medir la regularidad local. En nuestro caso particular, nuestra motivación ha sido justo la contraria: la de conseguir capturar y analizar de forma adecuada la singularidad local de datos funcionales, la cual es fundamental en problemas tales como los abordados en el campo de la meteorología (ver, por ejemplo, Febrero-Bande et al. [2008]; Ignaccolo et al. [2014]; Slini et al. [2006]; Stadlober et al. [2008]). Este marco teórico flexible nos ha permitido derivar un estimador fuertemente consistente del operador de autocorrelación de un proceso ARB(1), cuyas trayectorias funcionales no tengan porque ser continuas, ni siquiera diferenciables, satisfaciendo, por ejemplo, una ecuación integro–diferencial o pseudo–diferencial de orden fraccionario. Destacar como el suavizamiento inducido por las normas involucradas, por ejemplo, los espacios de Besov y Sobolev de orden negativo, nos permite considerar una gama más amplia de operadores de autocovarianza, al margen de los clásicos operadores asumidos como traza en la norma L^2. Como ya se ha mencionado en secciones anteriores, el interés de nuestro enfoque se ha ilustrado numéricamente (ver el estudio de simulación llevado a cabo en el Apéndice A8.6 y el Material Suplementario aportado en el Apéndice A8.8) mediante el análisis estadístico de series temporales funcionales, con valores en espacios nucleares. En particular, se ha considerado la escala continua de espacios de Besov de orden fraccionario, cuya caracterización ha sido llevada a cabo en función de bases wavelet.

A recent survey has demonstrated that one in seven billion human beings is you

L. Piedrahita (19th February 1977 -)

7

OPEN RESEARCH LINES

We briefly discussed the major current research lines in which we are working and that could be raised in the future, in keeping with the results and conclusions herein reached.

- **Appendix A1.** Inspired by the good accuracy achieved by semi–metrics based on derivatives in curves classification (see Ferraty and Vieu [2006] and the numerical results in Appendix A1.4), and beyond FPCA and FPLSR semi–metrics, these derivatives–based semi–metrics may be implemented in the context of random surfaces. For this purpose, non–uniform rational B–splines (NURBS) could be adopted (see, e.g., Schneider [2014]; Schoenberg [2012]). The extension of the results obtained to the case of correlated random surfaces constitutes the subject of the subsequent investigation.

- **Appendix A2.** The same insights established in the current dissertation could be reformulated in the context of ARH(p) processes (see [Bosq, 2000, Chapter 5] and Damon and Guillas [2002, 2005]), moving–average and linear processes, valued in the referred function spaces. It would also be worth the implementation of a real–data application, in the context of the Vasicek's model. Lastly, the more complex framework of truncated O.U. process, where time is restricted to the positive real line, is being developed.

- **Appendix A3.** The main extension that might be considered consists in establishing the same asymptotic results, concerning the convergence in the mean–square sense, in the Hilbert–Schmidt norm, when the eigenvectors of the autocovariance operator do not diagonalize (respectively, pseudo–diagonalize) the autocorrelation operator. A more extensive comparative study could be implemented, including parametric and nonparametric techniques.

- **Appendix A4.** The extension of the results derived to a more general modelling context, regarding the functional errors, displaying correlations in time or/and space, could be addressed. For example, the approach by Hörmann and Kokoszka [2010], on weakly dependent processes, could be adopted here, for modelling temporal correlated errors. The extension of the studied multivariate Hilbert–valued fixed effect model to the mixed effect model context constitutes another future research line for investigation.

- **Appendix A5.** The extension of the results derived, beyond the restriction on the existence of a common eigenvectors system, diagonalizing the autocovariance and autocorrelation operators, should also be addressed. The case of alternative prior distributions could also be analysed.

- **Appendices A6–A7.** The main open research line may be to develop a comprehensive review, where the more flexible theoretical framework provided, in the linear parametric functional framework, could outperform the asymptotic properties of nonparametric and semi–parametric estimators, with an important dimension reduction of the problem. A survey about some of the most interesting real–data applications in functional time series could be addressed.

- **Appendix A8.** One of the first subjects to address in the near future, in this research line, will be the analysis of functional (or high–dimensional) real–data applications, where local singular behaviours are observed, and must be measured in a properly way, for functional prediction of the magnitude of interest. That is the case, for example, of functional data related to circadian rhythms and sleep quality, as well as physical activity tracking (see, e.g., Gruen et al. [2017]; Lee et al. [2017]; Sathyanarayana et al. [2016]). Currently, we are working on the estimation of ARBX(1) processes; i.e., ARB(1) processes with exogenous variables. This framework is motivated by the forecasting of pollutants particles, such as PM_{10} pollutants, displaying an erratic local behaviour in time (see the cokriging approach in Bohorquez et al. [2017]), which are heavily dependent on meteorological variables (see, e.g., Poggi and Portier [2011]). Since PM_{10} are inhalable atmospheric pollution particles, its forecasting has became crucial aimed at adopting efficient public transport policies. The ARBX(1) estimation is being carried out by adopting the referred Markovian matrix representation proposed in [Bosq, 2000, Chapter 5, p. 128] to transform ARH(p) processes into ARH(1) models valued in H^p, which can be extended to the framework of nuclear Banach spaces.

A1
FUNCTIONAL STATISTICAL CLASSIFICATION OF NON-LINEAR DYNAMICS AND RANDOM SURFACES ROUGHNESS IN CONTROL SYSTEMS

ÁLVAREZ-LIÉBANA, J.; RUIZ-MEDINA, M. D.: Functional statistical classification of non-linear dynamics and random surfaces roughness in control systems. Int. J. Math. Models Methods Appl. Sciences 9 (2015), pp. 1–20. ISSN: 1998-0140

ABSTRACT

This paper addresses, in a nonparametric functional statistical framework, the problem of classification of non–linear features of curve and surface data in control systems. Specifically, on the one hand, in the detection of non–linear dynamic features, wavelength absorbance curve data are analysed for different meat pieces to discriminate between two categories of meat in quality control in food industry. On the other hand, in the nonparametric functional classification of deterministic and random surface roughness and irregularities, in the field of railway engineering, train deterministic and random vibrations are analysed to discriminate between different non–linear features characterizing roughness and irregularities of railway.

A1.1 INTRODUCTION

Non–linear dynamics and features in the data can be captured and suitable analysed within the functional statistical framework. Temporal and spatial functional statistics are relatively recent branches of Statistics, where non-parametric statistical techniques are now been developing to approximate the non-linear functional form of the probability distribution underlying to a sequence of random curves, surfaces, etc. In this context, new criteria for curve classification are proposed (see Ferraty and Vieu [2006]; Ramsay and Silverman [2005], among others). These procedures for random curve classification are designed in the absence or in the presence of interactions between different individuals, as well as between different times (see Aach and Church [2001]; Ferraty and Vieu [2006]; Hall et al. [2001]; James and Hastie [2001]; Liu and Müller [2003]; Müller and Stadtmüller [2005]). In James and Hastie [2001], a variant of linear discriminant analysis, in terms of the curve projections assuming a Gaussian distribution with common covariance matrix for all classes, is considered in the setting of filtering methods. Specifically, minimization of the distance to the group mean is the criterion adopted in this functional classification methodology. In Hall et al. [2001] a likelihood-based approach based on quadratic discriminant analysis is presented. They propose a fully nonparametric density estimation, and, in practice, multivariate Gaussian densities are considered. Dealing with non–linear discriminant algorithms, the learning optimal kernel for Kernel Fisher Discriminant Analysis (KFDA) is proposed in Ge and Fan [2013] to be able to optimize a combination of weight coefficients and kernels. In a generalized linear model framework, the model-based functional classification procedures proposed in Hidalgo and Ruiz-Medina [2012]; Leng and Müller [2006] are implemented. Specifically, for dimension reduction, Functional Principal Component Analysis (FPCA), and local wavelet-vaguelette decomposition are considered. K-nearest neighbor method is applied to Fourier coefficients in Biau et al. [2003]. Wavelet bases are selected for projection in Berlinet et al. [2008]. In James and Sugar [2003] spline bases are considered in a random effect model context, combining the best properties of filtering and regularization methods. These methods are effective when the observations are sparse, irregularly spaced or occur at different time points for each subject (see also Abraham et al. [2003], where B-splines bases are previously chosen for projection in the application of k-means-based classification procedure). In Ghosh and Kaabouch [2014], a support vector machine is used to scene classification, in order to construct an effective clustering procedure for real time applications, in particular, for image sequence classification depending on several factors.

Functional nonparametric statistical classification procedures, based on kernels, are extensively developed in the context of statistical learning methods (see, for example, Scholkopf and Smola [2002]). In this framework, the unknown function is estimated, considering its optimal approximation in a functional class given by a Reproducing Kernel Hilbert Space (RKHS), under some prescribed criterion. Chaos game representation and multifractal analysis can also be considered in the classification of functional protein sequences displaying singular features (see, for instance, Yang et al. [2009a,b]).

This paper deals with the functional statistical nonparametric classification of non-linear random functions with n–dimensional support (e.g., curves, surfaces, etc). They are assumed to be uncorrelated random functions. As motivation for illustration of the proposed functional nonparametric statistical methodology, we address two problems in the applied areas of food industry and railway engineering. Specifically, fat content is first analysed for classification of meat pieces, from the observation of spectrometric curve data corresponding to the absorbance measured at 100 wavelengths. On the other hand, in the random surface discrimination context, the statistical analysis of train deterministic and random vibrations is achieved from the nonparametric functional statistical classification of rail roughness and irregularities. The results obtained, after the implementation of the proposed classification methodology are showed in Appendices A1.4 and A1.6, respectively. In such an implementation, an extended version of the classification algorithm formulated in Ferraty and Vieu [2006] is derived. Namely, numerical integration is performed by applying the Smolyak quadrature rule, after interpolation over a finer n–dimensional grid the values observed at a coarser grid, which constitutes our actual functional dataset. Different semi-metrics can then be applied, mainly based on FPCA and Functional Partial Least Squares Regression (FPLSR), which is an extension of Partial Least Squares technique (see, e.g., Oladunni [2013]). In addition, the kernel estimation of the posterior probability of belonging to each one of the categories defining the response provides us a rule for classification of the observed n–dimensional supported functional data in a nonparametric statistical context.

The resulting classification procedure for non-linear random functions with n-dimensional support, in the context of nonparametric functional statistics, allows discrimination in a more flexible framework. In particular, this paper provides an extension to the two–dimensional case of the one–dimensional models proposed in Fryba [1999]; Mohammadzadeh et al. [2013]; Youcef et al. [2013], among others, for the analysis of imperfections of railway track. These irregularities are the second source of bridge vibrations and the first one of train vibrations, and can be classified into non-random and random irregularities (as the roughness of the rails). The dynamics of these railway tracks under moving trains must be taken into account in order to construct and design the railway bridges and beams, as well as to locate and construct the railway stations and surrounding buildings. The effects of rail roughness and rail irregularities on the dynamic behaviour of bridge and vehicles are considered in Mohammadzadeh et al. [2013]; Youcef et al. [2013]. In this paper, the non-random imperfections are represented in terms of a two-dimensional function perturbed by Gaussian white noise, reflecting the measurement device error, while the random ones will be defined in terms of zero-mean Gaussian random surfaces, displaying different non-linear spatial patterns according to their spatial correlation structure.

The outline of the paper is as follows. Appendix A1.2 presents some preliminaries definitions and elements involved in the functional statistical nonparametric classification algorithm studied in Ferraty and Vieu [2006]. Appendix A1.3 establishes the main steps of the proposed classification algorithm for n-dimensional supported non-linear random functions, and in particular, for random curves and surfaces. The application of this algorithm to spectrometric curve data for meat piece classification according to fat content is illustrated in Appendix A1.4. Appendix A1.5 provides a training simulation study to discriminate between different trend surfaces in Gaussian random surface classification. A simulation study is undertaken in Appendix A1.6 for illustration of the proposed functional classification methodology for perturbed deterministic and random irregularities in the surface of railway track. Conclusions are drawn in Appendix A1.7.

A1.2 PRELIMINARIES ABOUT FUNCTIONAL NON-PARAMETRIC CLASSIFICATION

Let us first introduce the preliminary elements and definitions, as well as the required notation for the description of the curve statistical functional classification algorithm proposed by Ferraty and Vieu [2006] in a nonparametric framework.

Assume that $T = (t_{min}, t_{max})$ is an interval in $\mathbb{R}$. We shall use the notation:

- $\boldsymbol{\chi} = \{\boldsymbol{\chi}(t),\ t \in T\}$ for representing a functional random variable (f.r.v.); that is, a random variable $\boldsymbol{\chi}$ that takes values in an infinite–dimensional space.
- χ functional data (f.d.) denotes an observation of $\boldsymbol{\chi}$.
- We shall denote a funcional dataset (f.dat.) $\{\chi_i,\ i = 1, \ldots, n\}$ as the observation of n–sample f.r.v

$$\{\boldsymbol{\chi}_i,\ i = 1, \ldots, n\} \sim \boldsymbol{\chi}.$$

Different families of semi-metrics mainly based on FPCA (see Jackson [2004], among others), FPLSR, and derivatives are commonly used to measure distances between curves. In the context of infinite–dimensional spaces, they are usually computed by numerical integration, considering, in our case, n–dimensional integration based on suitable quadrature rules.

A1.2.1 Functional Principal Component Analysis (FPCA)

This technique is based on projection into the eigenvector system of the covariance operator, obtaining a series expansion of the f.r.v. defining our data set, in terms of uncorrelated r.v., with scale parameters given by the square root of the associated eigenvalues. It is well-known that PCA (with euclidean metric) is formulated as follows:

$$\begin{aligned} z_i &= \frac{\langle \boldsymbol{v}_i, \boldsymbol{x} \rangle}{\|\boldsymbol{v}_i\|} = \frac{1}{\|\boldsymbol{v}_i\|} \sum_{j=1}^{p} v_{i,j} x_j = \frac{1}{\|\boldsymbol{v}_i\|} \boldsymbol{v}_i^T \boldsymbol{x}, \quad z_i \in \mathbb{R}, \quad \boldsymbol{v}_i, \boldsymbol{x} \in \mathbb{R}^p, \quad i = 1, \ldots, p, \\ \boldsymbol{x} &= \sum_{j=1}^{p} e_j x_j \equiv \sum_{j=1}^{p} v_j z_j, \quad \boldsymbol{x}, \boldsymbol{e} \in \mathbb{R}^p \end{aligned}$$

being $\boldsymbol{e} = (e_1, \dots, e_p)$ and $\boldsymbol{v}_i = (v_{i,1}, \dots, v_{i,p})$ orthonormal bases in $\mathbb{R}^p$, where, for $i = 1, \dots, p$,

$$\mathrm{E}\left\{z_i^2\right\} = \lambda_i, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p.$$

In the infinite–dimensional case, we consider the spaces L^p with respect to a measure μ, introduced in terms of the semi-norm $\|\cdot\|_p$, given by

$$\|f\|_p := \left(\int |f(x)|^p \mu(dx)\right)^{\frac{1}{p}}.$$

In particular, we concentrate in the case of $p = 2$, where we have a Hilbert space structure. Recall the fundamental definitions associated with this case.

Definition A1.2.1 *Let A be a linear operator. A function $f \neq 0$ is an eigenfunction of A if and only if $A(f) = \lambda f$.*

Definition A1.2.2 *Let $(H, \langle\cdot,\cdot\rangle_H)$ be a real valued pre-Hilbert space with the inner product*

$$\langle f, g\rangle_H = \int f(x)g(x)w(x)dx, \quad \forall f, g \in H,$$

where w is a weight function. Two functions f, g are then orthogonal if and only if

$$\langle f, g\rangle_H = 0.$$

The resulting series expansions in PCA (on left) and FPCA (on right) are given as follows, when $\{\boldsymbol{v}_j,\ j \geq 1\}$ are normalized:

$$\begin{aligned} z_j &= \langle \boldsymbol{v}_j, \boldsymbol{x}\rangle_H, \qquad z_j = \int \boldsymbol{\chi}(x)v_j(x), \\ \boldsymbol{x} &= \sum_{j=1}^{p} \boldsymbol{v}_j z_j, \qquad \boldsymbol{\chi}(x) = \sum_{j=1}^{\infty} v_j(x) z_j. \end{aligned}$$

Thus, for the infinite-dimensional case we have

$$\boldsymbol{\chi}(z) = \sum_{j=1}^{\infty}\left(\int \boldsymbol{\chi}(x)v_j(x)dx\right)v_j(z),$$

and its truncated version can be written as

$$\widehat{\boldsymbol{\chi}}^{(q)}(z) = \sum_{j=1}^{q}\left(\int \boldsymbol{\chi}(x)\, v_j(x)\, dx\right)v_j(z). \tag{A1.1}$$

From (A1.1), the following semi–norm can be defined:

$$d_q^{FPCA}\left(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2\right) = \sqrt{\sum_{j=1}^{q}\left(\int \left[\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\right](x)\, v_j(x)\, dx\right)^2}$$

From a practical point of view, the above integrals are approximated by a quadrature rule. Specifically, for the observed discretized curves, namely, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the following numerical approximation is computed:

$$d_q^{FPCA}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) = \sqrt{\sum_{j=1}^{q}\left(\sum_{i=1}^{I} w_i \left[\boldsymbol{x}_1 - \boldsymbol{x}_2\right](t_i)\, v_{ji}\right)^2} \tag{A1.2}$$

where $\{t_i,\ i = 1, \ldots, I\}$ are the nodes, $1 \leq q \leq n$ the number of components chosen and

$$\boldsymbol{\Sigma}_{\boldsymbol{\chi}}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\chi}_i(s)\, \boldsymbol{\chi}_i(t)$$

the empirical version of the covariance kernel (i.e., its empirical matrix approximation), being $\{\boldsymbol{v}_j = (v_{j1}, \ldots, v_{jI}),\ j = 1, \ldots, q\}$ the empirical eigenvectors of

$$\mathbf{W}^{1/2}\boldsymbol{\Sigma}\mathbf{W}^{1/2}, \quad \boldsymbol{W} = diag\left(w_1, \ldots, w_I\right)$$

being a diagonal matrix with non–null entries given by the quadrature weights provided by a quadrature rule.

A1.2.2 Functional Partial Least Squares Regression (FPLSR)

The Multivariate Partial Least Squares Regression (MPLSR) is an extension of PLSR motivated by dealing with multivariate response or when the number of predictors is very large in comparison with the number of observations.

We can apply MPLSR with only one scalar response but it would be inadequate with regard to the complexity of functional data. Hence, we are going to construct a multivariate response binary matrix where each column j represents if the i–th observation belongs to class j. Such as FPCA technique, we can extend MPLSR to FPLSR in functional framework, providing us g components depending on a number of factors q, which plays similar role to the number of dimensions retained in FPCA. The main difference between FPCA and FPLSR comes from the fact that the FPCA explains only the predictors, whereas the FPLSR approach computes a simultaneous decomposition of the set of predictors and responses, being able to explain both predictors and responses. Thus, we get a similar FPCA formula:

$$d_q^{FPLSR}\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) = \sqrt{\sum_{j=1}^{g}\left(\sum_{i=1}^{I} w_i \left[\boldsymbol{x}_1 - \boldsymbol{x}_2\right](t_i)\, v_{ji}^q\right)^2}$$

where $\boldsymbol{v}_1^q, \ldots, \boldsymbol{v}_g^q$ are performed by FPLSR.

A1.2.3 Semi–metrics based on derivatives

Lastly, we introduce the semi–metric based on derivatives. That is, the L^2 distance between the derivatives of different orders of two given curves is established as a measure of closeness in the following way:

$$d_q^{deriv}(\chi_1, \chi_2) = \sqrt{\int \left(\chi_1^{(q)} - \chi_2^{(q)}\right)^2 (x)\, dx}$$

where $\chi^{(q)}$ is the q–th derivative of χ and $d_0^{deriv} = d_{L^2}$.

To avoid stability problems with derivatives, a B–spline basis approximation is usually considered (see, e.g., de Boor [1978]; Schumaker [1981]). Using the discretized curve $\boldsymbol{x}_i = (\chi_i(t_1), \ldots, \chi_i(t_I))$, we obtain the following approximation:

$$\widehat{\chi}_i(\cdot) = \sum_{b=1}^{B} \widehat{\beta}_{ib} B_b(\cdot) \quad \widehat{\chi}_i^{(q)}(\cdot) = \sum_{b=1}^{B} \widehat{\beta}_{ib} B_b^{(q)}(\cdot)$$

where $\{B_1, \ldots, B_B\}$ is a B–spline basis. Thus, for numerical approximation of

$$d_q^{deriv}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sqrt{\int \left(\widehat{\chi}_1^{(q)}(x) - \widehat{\chi}_2^{(q)}(x)\right)^2 dx},$$

a quadrature rule is considered. Note that B–spline basis allow to work even with unbalanced data sets.

A1.2.4 Numerical integration: quadrature rules

To define all of these semi–metrics in a functional space, numerical integration in terms of a quadrature rules is required. Let see a brief about them.

There is a large variety of one–dimensional numerical integration procedures, as the trapezoidal rule (see Gerstner [2007]), the Clenshaw–Curtis rule (see, e.g, Kaarnioja [2013]; Novak and Ritter [1998]) and rules introduced in Burkardt [2011]. We could also use stochastic simulation applying methods such as Monte Carlo (MC) and Quasi-Monte Carlo methods (QMC) (see, for example, Gerstner and Griebel [1998]). We will restrict our attention to numerical integration, since a set of weights is needed.

According to Gerstner [2007]; Kaarnioja [2013], in the following, we consider functions $f(x)$ from a regular function class:

$$\mathcal{C}^r(\Omega) := \left\{ f : \Omega \subset \mathbb{R}^n \to \mathbb{R}, \left\| \frac{\partial^s f}{\partial x^s} \right\|_\infty < \infty, \quad s \leq r \right\}.$$

As we will see, the goal is to approximate the integral $\displaystyle\int_\Omega f(x)dx$ in a subset $\Omega \subset \mathbb{R}^n$, by a sequence of n_l–point quadrature, with $n_l = 2^{l-1} + 1$.

A1.2.5 Functional nonparametric supervised classification of random curves

As described in Ferraty and Vieu [2006], we now observe a f.r.v $\boldsymbol{\chi}$ and a categorical response $\boldsymbol{y}$ that represents the class membership of each element. The main aim is to be able to predict the class membership of a new f.d., by means of a nonparametric rule.

Denoting by (E, d) a semi–metric space and $\overline{G} = \{1, \dots, G\}$ a set of integers, we consider

$$(\boldsymbol{\chi}_i, \boldsymbol{y}_i) \sim \{\boldsymbol{\chi}, \boldsymbol{y},\ i = 1, \dots, n\},$$

to be a sample of n independent pairs in $E \times \overline{G}$. Thus, (χ_i, y_i) denotes an observation of $(\boldsymbol{\chi}_i, \boldsymbol{y}_i)_{i=1,\dots,n}$, and $(\boldsymbol{x}_i, y_i)$, with $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,I})$ being the discretization of (χ_i, y_i).

Applying the Bayes rule, our goal is estimate $p_g(\chi) = P(Y = g | \boldsymbol{\chi} = \chi) = \mathrm{E}\{\mathbf{1}_{Y=g} | \boldsymbol{\chi} = \chi\}$ $(g \in \overline{G})$, doing the assignment:

$$\widehat{y}(\chi) = arg \max_{g \in \overline{G}} \widehat{\boldsymbol{p}}_g(\chi) \tag{A1.3}$$

where $\widehat{\boldsymbol{p}}_{\boldsymbol{g}}(\chi) = (\widehat{p}_1(\chi), \dots, \widehat{p}_G(\chi))$ are the estimate posterior probabilities and $\mathbf{1}_{Y=g}$ is the indicator function.

Let K be a kernel function and $\Lambda : \mathbb{R}^p \to \mathbb{R}$ a function (an operator in the infinite–dimensional case) which we want to estimate. We define the kernel smoother as:

$$K_h(\chi, \boldsymbol{\chi}_i) := K\left(\frac{d(\chi, \boldsymbol{\chi}_i)}{h(\chi)}\right),$$

where K is a positive kernel function that decreasing with the distance between $\boldsymbol{\chi}_i$ and χ, $h(\chi)$ is a positive bandwidth, depending on χ. Therefore, we can use the truncated kernel regression estimator of Λ proposed in Nadaraya [1964]; Watson [1964], in an infinite–dimensional setting, as follows:

$$\widehat{\Lambda}(\chi) := \frac{\sum_{i=1}^{n} K_h(\chi, \boldsymbol{\chi}_i) \Lambda(\boldsymbol{\chi}_i)}{\sum_{i=1}^{n} K_h(\chi, \boldsymbol{\chi}_i)}, \tag{A1.4}$$

where $\Lambda(\boldsymbol{\chi}_i) = \mathrm{E}\{\mathbf{1}_{Y_i=g} | \chi_i = \boldsymbol{\chi}_i\} = \mathbf{1}_{yi=g} = p_g(\boldsymbol{\chi}_i) = 1$. Thus, according to (A1.4):

$$\widehat{p}_{g,h}(\chi) = \frac{\sum_{i=1}^{n} K\left(\frac{d(\chi, \boldsymbol{\chi}_i)}{h(\chi)}\right) \mathbf{1}_{yi=g}}{\sum_{i=1}^{n} K\left(\frac{d(\chi, \boldsymbol{\chi}_i)}{h(\chi)}\right)} = \sum_{\{i: y_i = g\}} w_{i,h}(\chi) \tag{A1.5}$$

with $w_{i,h}(\chi) = \dfrac{K\left(\dfrac{d(\chi, \boldsymbol{\chi}_i)}{h(\chi)}\right)}{\sum_{i=1}^{n} K\left(\dfrac{d(\chi, \boldsymbol{\chi}_i)}{h(\chi)}\right)}$.

If we choose a kernel that $K(x) = 0$ if $|x| < 1$ results:

$$\widehat{p}_{g,h}(\chi) = \sum_{i \in \mathcal{J}} w_{i,h}(\chi)$$

where $\mathcal{J} = \{i : y_i = g\} \cap \{i : d(\chi, \boldsymbol{\chi}_i) < h\}$.

A1.2.6 Bandwidth selection

Finally, we have to choose h with the goal of minimizing a loss function that depends on $\widehat{p}_{g,h}(\chi_i, y_i)$'s and y_i's:

$$h_{Loss} = arg \inf_{h} Loss(h) \tag{A1.6}$$

With this aim, we will replace the choice of h among an infinite set $\mathcal{H}$ with an integer parameter k among a finite subset $\mathcal{K}$, by the consideration of k-Nearest Neighborhood (kNN) discretized version of (A1.6):

$$\widehat{p}_{g,k}(\boldsymbol{x}) = \frac{\sum_{i \in \mathcal{J}} K\left(\dfrac{d(\boldsymbol{x}, \boldsymbol{x}_i)}{h_k(\boldsymbol{x})}\right)}{\sum_{i=1}^{n} K\left(\dfrac{d(\boldsymbol{x}, \boldsymbol{x}_i)}{h_k(\boldsymbol{x})}\right)}$$

where h_k is such that $\{i : d(\boldsymbol{x}, \boldsymbol{x}_i) < h_k\} = k$. Thus, we have to find $k_{Loss} = arg \min_{k \in \mathcal{K}} Loss(k)$. From now on, we consider $\widehat{p}_{g,k}$ the estimator of $\widehat{p}_g$.

If we use the cross–validation procedure proposed in Ferraty and Vieu [2006] and choose as loss function

$$Loss(k) = LCV(k, i_0) = \sum_{g=1}^{G} \left(\mathbf{1}_{y_{i_0}=g} - p_{g,k}^{(-i_0)}(\boldsymbol{x}_{i_0})\right)^2, \tag{A1.7}$$

where

$$p_{g,k}^{(-i_0)}(\boldsymbol{x}_{i_0}) = \frac{\sum_{i \in \mathcal{J},\, i \neq i_0} K\left(\dfrac{d(\boldsymbol{x}_{i_0}, \boldsymbol{x}_i)}{h_k(\boldsymbol{x}_{i_0})}\right)}{\sum_{i=1,\, i \neq i_0}^{n} K\left(\dfrac{d(\boldsymbol{x}_{i_0}, \boldsymbol{x}_i)}{h_k(\boldsymbol{x}_{i_0})}\right)},$$

and $\boldsymbol{x}_{i_0}$ is the nearest neighbour of $\boldsymbol{x}$, so we denote $i_0 = arg \min_{i=1,\ldots,n} d(\boldsymbol{x}, \boldsymbol{x}_i)$. Hence, the local choice is the following:

$$\begin{aligned} k_{LCV}\left(\boldsymbol{x}_{i_0}\right) &= arg\min_k LCV\left(k, i_0\right) \\ k_{LCV}\left(\boldsymbol{x}_{i_0}\right) &\longrightarrow h_k = h_{LCV}\left(\boldsymbol{x}_{i_0}\right) \\ Miss.\,Rate &= \frac{\sum_{i=1}^{n} \mathbf{1}_{y_i \neq y_i^{LCV}}}{n} \end{aligned}$$

A1.3 NONPARAMETRIC CLASSIFICATION OF UNCORRELATED SURFACES

Let us consider

$$\boldsymbol{\psi} = \left\{\boldsymbol{\psi}\left(x_1, \ldots, x_n\right), \quad \left(x_1, \ldots, x_n\right) \in \mathbb{R}^n\right\}$$

a random n–dimensional supported f.r.v. The observed realization ψ of $\boldsymbol{\psi}$ is referred a n–dimensional f.d. In the particular case of $n = 2$, that is, of $\mathbb{R}^2$, a regular grid is chosen with nodes having coordinates $\left(\left(x_1, y_1\right), \ldots, \left(x_N, y_M\right)\right)$. Hence, in the following, we refer to an $M \times N$ rectangular regular grid.

A1.3.1 Reformulation of semi–metrics

The corresponding reformulation of semi-metric based on FPCA is straightforward. In particular, when $n = 2$, we have

$$d_q^{FPCA}\left(\psi_1, \psi_2\right) = \sqrt{\sum_{j=1}^{q}\left(\sum_{i=1}^{I} w_i x_i^* v_{ji}\right)^2},$$

where $1 \leq q \leq n$ the number of components chosen, $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}\left(s, t\right) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\chi}_i\left(s\right)\boldsymbol{\chi}_i\left(t\right)$ is the empirical version of the covariance kernel, $\left\{\boldsymbol{v}_j, \ j = 1, \ldots, q\right\}$, are the orthonormal eigenvectors (corresponding to the components chosen) of empirical covariance matrix

$$\mathbf{W}^{1/2}\boldsymbol{\Sigma}\mathbf{W}^{1/2}_{I\times I}, \quad \boldsymbol{W} = diag\left(w_1, \ldots, w_I\right)$$

whose diagonal entries are two–dimensional quadrature weights, and

$$\left(x_1^*, \ldots, x_I^*\right) = \left(\left(\psi_1 - \psi_2\right)\left(x_1, y_1\right), \ldots, \left(\psi_1 - \psi_2\right)\left(x_N, y_M\right)\right), \tag{A1.8}$$

remains being a real vector, with $I = N \times M$. As previously, $\left(x_i, y_j, \ i = 1, \ldots, N, \ j = 1, \ldots, M\right) \in D \subset \mathbb{R}^2$ represents the set of nodes of a regular rectangular grid, with associated discretized functional value of the observed f.d. given by $\left(\psi\left(x_1, y_1\right), \ldots, \psi\left(x_N, y_M\right)\right)$, which can also be treated as a real vector associated with the discrete observation of $\boldsymbol{\psi}$.

Reformulation of FPLSR in the two–dimensional case can be derived in a similar way. Thus,

$$d_q^{FPLSR}(\psi_1, \psi_2) = \sqrt{\sum_{j=1}^{g} \left(\sum_{i=1}^{I} w_i x_i^* v_{ji}^q \right)^2}$$

where $\{x_i^*,\, i = 1, \ldots, I\}$, are given as in equation (A1.8), and $\left(\boldsymbol{v}_1^q, \ldots, \boldsymbol{v}_g^q\right)$ are performed by FPLSR.

Although it is out of our scope, semi-metrics based on derivatives can be also reformulated by considering the corresponding L^2 norm of the corresponding partial derivatives. In particular, for $n = 2$, non-uniform rational B–spline (NURBS) can be used (see, e.g., Schneider [2014]; Schoenberg [2012]).

A1.3.2 Smolyak quadrature

We will describe the n–dimensional version of the Smolyak quadrature rule to obtain a set of weights, defining, in particular, the metric $\boldsymbol{W}^{1/2}\Sigma\boldsymbol{W}^{1/2}$, in the numerical approximation of the integral by a weighted sum of values of the integrand at certain nodes (see, e.g.,Gerstner and Griebel [1998]; Kaarnioja [2013]).

The main goal is to approximate

$$I_W^n f := \int_{\prod_{i=1}^{n} I_i} f(x_1, \ldots, x_n) \prod_{i=1}^{n} W_i(x_i)\, dx_i$$

by a n–sequence of k_{l_j}–point quadratures, with $k_{l_j} = 2^{l_j - 1} + 1$ and $j \in \{1, \ldots, n\}$:

$$U_{l_j} := \sum_{i=1}^{k_{l_j}} w_i f(x_i) = \sum_{i=1}^{2^{l_j-1}+1} w_i f(x_i)$$

with $l_j \geq 1$. Smolyak rule combines, by means of tensor products, univariate quadratures rules $\left\{U_{l_j},\, j = 1, \ldots, n\right\}$, respectively associated with each dimension j, for $j = 1, \ldots, n$ (e.g., Trapezoidal rule, Clenshaw–Curtis's rule, Gauss–Legendre's rule, Gauss-Patterson's rule, etc).

Definition A1.3.1 *Let $S : \mathcal{C}(\Omega) \to \mathbb{R}$ and $T : \mathcal{C}(\Xi) \to \mathbb{R}$ be operators that admit a representation of the form:*

$$\begin{aligned} Sf(\boldsymbol{x}) &= \sum_{i=1}^{m} a_i f(x_i) \\ Tg(\boldsymbol{y}) &= \sum_{j=1}^{n} b_j g(y_j), \end{aligned}$$

with positive weights, $\boldsymbol{x} = (x_1, \ldots, x_m)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$. The tensor product of S and T is the linear operator $S \otimes T : \mathcal{C}(\Omega \times \Xi) \to \mathbb{R}$ defined by:

$$Sf \otimes Tg(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j f(x_i) g(y_j).$$

Let $\left\{U_{l_j}^{(j)},\ j = 1, \ldots, n\right\}$ be a sequence of univariate quadrature rules, where j represents the dimension in which we are integrating and $k_{l_j} = 2^{l_j - 1} + 1$ the number of evaluation points. This univariate rules are chosen in such a way such that $I_{W_j}^1 p = U_{l_j}^{(j)}$, where p is a polynomial of degree at most k_{l_j}.

We denote as $\left\{w_i^{(j)},\ i = 1, \ldots, k_{l_j}\right\}$ and $\left\{x_i^{(j)},\ i = 1, \ldots, k_{l_j}\right\}$ the weights and the nodes, respectively, of the univariate rule $U_{l_j}^{(j)}$, for $j = 1, \ldots, n$. Thus, the original problem can be approximated in tensor product form:

$$I_W^n f \approx \bigotimes_{j=1}^{n} U_{l_j}^{(j)} f = Q_k^n \tag{A1.9}$$

with $\left\{U_{l_j}^{(j)}\right\} = k_{l_j} = 2^{l_j - 1} + 1$ and $\boldsymbol{l} = (l_1, \ldots, l_n)$, with $l_j \leq k\ \forall j \in \{1, \ldots, n\}$.

In fact, Smolyak quadrature rule proposed in Gerstner [2007]; Kaarnioja [2013] uses difference operators instead of directly applying the tensor product.

Definition A1.3.2 *Let $\left\{U_i^{(j)},\ i = 1, \ldots, \infty\right\}$ be a sequence of univariate rules in I_j. We define the difference operators in I_j as:*

$$\Delta_0^{(j)} = 0,\ \Delta_1^{(j)} = U_1^{(j)}, \quad \Delta_{i+1}^{(j)} = U_{i+1}^{(j)} - U_i^{(j)}.$$

Thus, Smolyak quadrature rule of order k in the n–dimensional rectangle $I_1 \times \ldots \times I_n$ (for simplicity we assume $I^n = I \times \ldots \times I$) can be defined as the operator:

$$Q_k^n = \sum_{\|\boldsymbol{\alpha}\|_1 \leq k} \bigotimes_{j=1}^{n} \Delta_{\alpha_j}^{(j)} \tag{A1.10}$$

where $\boldsymbol{\alpha} \in \mathbb{N}^n$ and $\alpha_j > 0$ (which implies that $k \geq n$). Remark that in the case of $n = 1,\ Q_k^1 = U_k^{(1)}$.

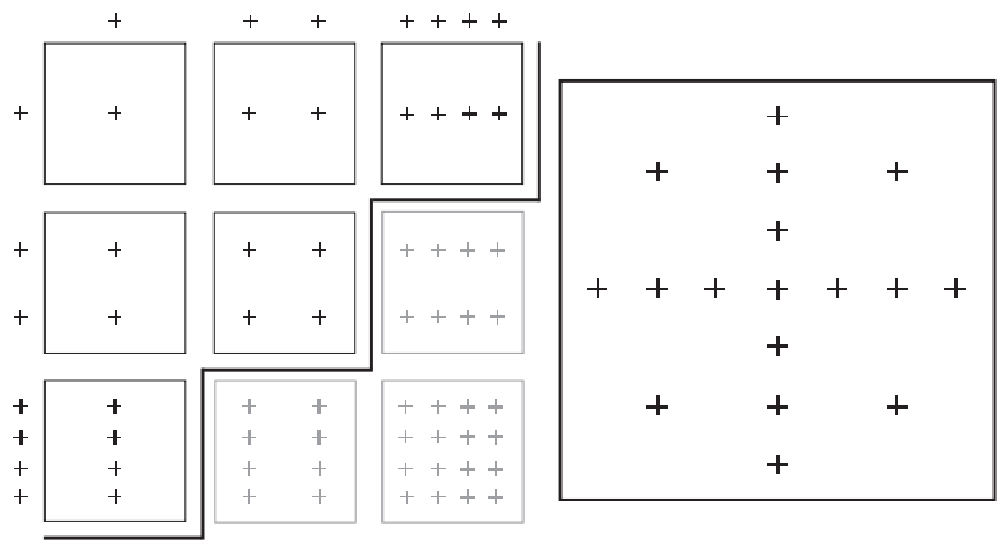


Figure A1.3.1: Product grids $X_{i_1}^{(1)} \times X_{i_2}^{(1)}$ such that $\|(i_1, i_2)\|_\infty \leq 3$ (on left) and the Q_4^2 grid (on right).

Since there are many terms that are removed, we shall also present a combination method of Smolyak rules (see, e.g.,Wasilkowski and Wozniakowski [1995]):

$$Q_k^n = \sum_{\substack{m \leq \|\boldsymbol{\alpha}\|_1 \leq k \\ \boldsymbol{\alpha} \in \mathbb{N}^n,\, \boldsymbol{\alpha} \geq \mathbf{1}}} (-1)^{k-\|\boldsymbol{\alpha}\|_1} \binom{n-1}{k-\|\boldsymbol{\alpha}\|_1} \bigotimes_{j=1}^{n} U_{\alpha_j}^{(j)}$$

with $m = max\,\{n, k-n+1\}$.

Rewritting (A1.10) and using (A1.9), we obtain:

$$Q_k^n = \sum_{l=m}^{k} \sum_{\substack{\|\boldsymbol{\alpha}\|_1 = l \\ \boldsymbol{\alpha} \in \mathbb{N}^n,\, \boldsymbol{\alpha} \geq \mathbf{1}}} \sum_{j_1=1}^{k_{\alpha_1}} \ldots \sum_{j_n=1}^{k_{\alpha_n}} c(k,n,l) w_{j,\boldsymbol{\alpha}} f\left(\boldsymbol{x}_{j,\boldsymbol{\alpha}}\right) \tag{A1.11}$$

where $c(k,n,l) = (-1)^{k-l}\binom{n-1}{k-l}$, $w_{j,\boldsymbol{\alpha}} = w_{j_1}^{(\alpha_1)} \ldots w_{j_n}^{(\alpha_n)}$ and $\boldsymbol{x}_{j,\boldsymbol{\alpha}} = \left(x_{j_1}^{(\alpha_1)} \ldots x_{j_n}^{(\alpha_n)}\right)$.

A1.3.2.1 Numerical implementation

The main steps and auxiliary functions in the implementation of the Smolyak's quadrature are the following :

- **Step 1** Define the function that provides us univariate nodes and weights (univariate quadrature rules at each dimension).

- **Step 2** Generate all multi-indices satisfying restrictions established in the algorithm proposed in Gerstner [2007]. For instance, if $n = 3$ and $k = 5$, $\boldsymbol{\alpha}$ could be $(1,1,1)$, $(1,1,2)$, $(1,2,1)$, $(2,1,1)$, $(1,1,3)$, $(1,3,1)$, $(3,1,1)$, $(1,2,2)$, $(2,1,2)$ and $(2,2,1)$.

- **Step 3** Determine, for any vector sequence $\left(\boldsymbol{v}^{(i)}\right)_{i=1}^{l}$, with $\boldsymbol{v}^{(i)} \in \mathbb{R}^{n_i}$, $i = 1, \ldots, l$, its vector combination. Thus, we define inductively $c_{\boldsymbol{v},l} = \text{combvec}\left(\left(\boldsymbol{v}^{(i)}\right)_{i=1}^{l}\right)$ as follows:

$$c_{\boldsymbol{v},l} = \begin{pmatrix} c_{\boldsymbol{v},l-1} \ldots c_{\boldsymbol{v},l-1} \ldots c_{\boldsymbol{v},l-1} \ldots c_{\boldsymbol{v},l-1} \\ v_1^{(l)} \underset{(n_l-1)}{\ldots} v_1^{(l)} \ldots v_{n_l}^{(l)} \underset{(n_l-1)}{\ldots} v_{n_l}^{(l)} \end{pmatrix}$$

with $c_{\boldsymbol{v},1} = \boldsymbol{v}_{(1)} = (v_1^{(1)} \ldots v_{n_1}^{(1)})$.

In addition, we have implemented two more functions. A function that groups weights associated at the same node, and auxiliary function that deletes the nodes with total weight equal to zero. Smolyak's nodes are different from the nodes where we have our observations, so we previously interpolate our f.dat. considering locally polynomials or k–Nearest Neighbourhood Smoother. The assignment of weights is done in two ways: To each interpolated node, we assign the weight corresponding to the Nearest Neighbour Smolyak's node; or, alternatively, we assign the weight defined by the average of the weights associated with the $k_{Smolyak}$–Nearest Neighbourhood Smolyak's nodes.

A1.4 FUNCTIONAL CLASSIFICATION RESULTS OF CURVES

methodology, as well as of the one formulated in Ferraty and Vieu [2006] is now compared in terms of their implementation from a spectrometric curve dataset available at `http://lib.stat.cmu.edu/datasets/tecator`. This dataset is related to quality control in food industry. It corresponds to a sample of finely chopped meat. For each unit i, among 215 pieces, we observe one spectrometric curve which corresponds to the absorbance measured at 100 wavelengths. Moreover, we have measured its fat content $\{y_i,\ i = 1, \ldots, 215\}$, obtained by an analytical chemical processing.

In the implementation of the classification procedure for validation purposes, our f.dat. sample has been randomly split into two sub–samples respectively corresponding to the training f.dat. sample, which constitutes a 70% of the total dataset, and a f.dat. validation sample or test sample, which in our case constitutes a 30% of the total sample.

Figure A1.4.1 shows spectrometric f.dat. The magnitude plotted is absorbance versus wavelength for different pieces, where 100 channel spectrum of absorbances are showed. Hence, each data appears as a discretized curve in 100 points, and interpolation is performed to get the corresponding values in a finer partition of the set containing the 100 points within the same wavelength interval $850 - 1050$ (see Figure A1.4.3). Two categories or groups are distinguished in advance: fat content under 20 $(y_i = 1)$ and over 20 $(y_i = 2)$ (see Figure A1.4.2).

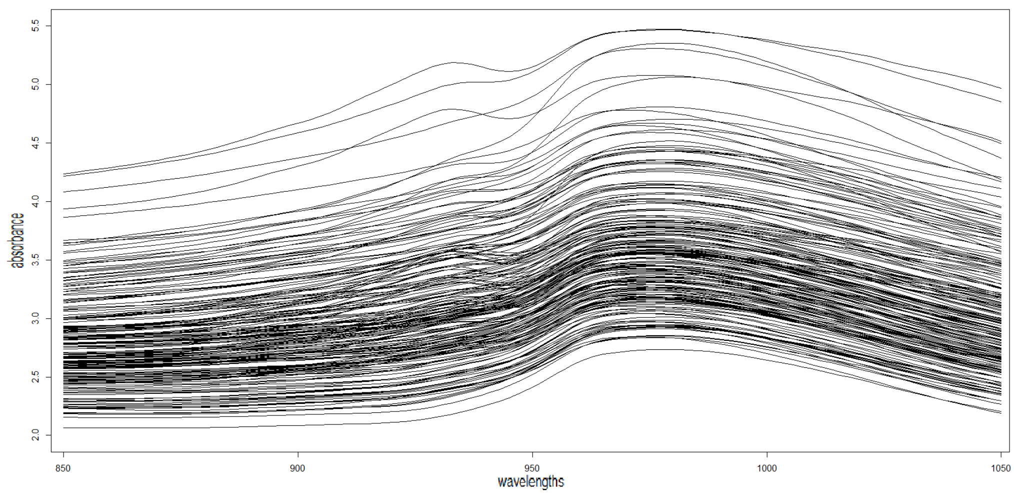


Figure A1.4.1: Discretized spectometric curves.

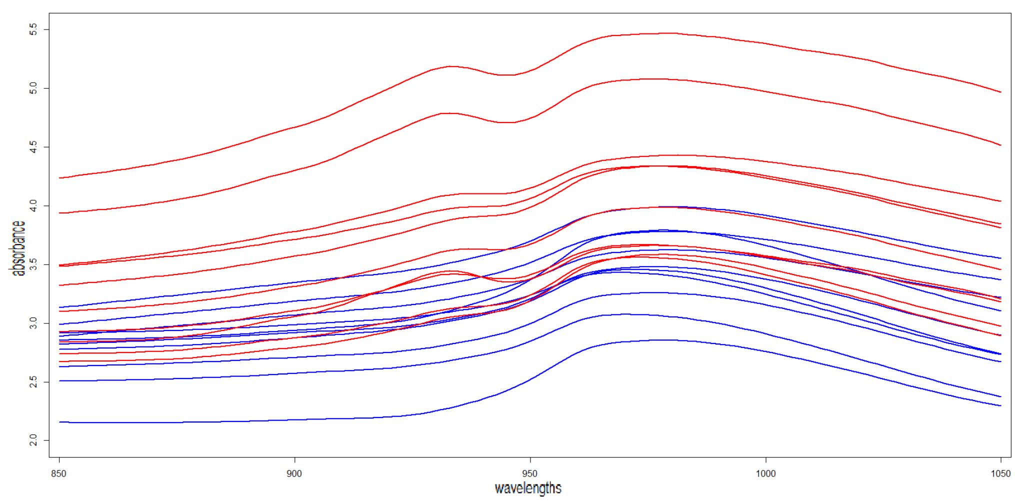


Figure A1.4.2: Discretized curves splitted by groups: the blue ones belong to class 1 (low fat content) and the red ones belong to class 2 (higher fat content).

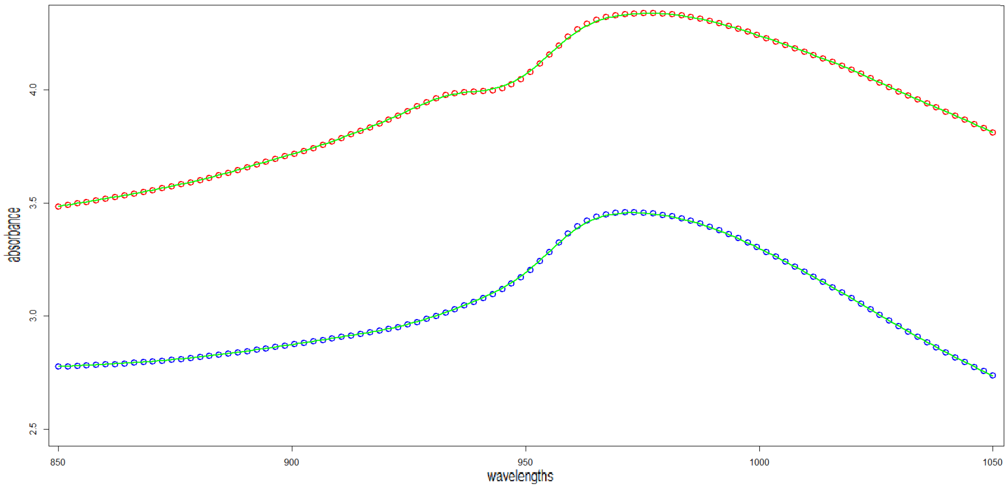


Figure A1.4.3: Accuracy of interpolation of a curve at each category, with step called $step_{mesh}$.

Figure A1.4.4 shows the results obtained using FPCA semi–metric, when different kernels (quadratic, indicator and triangle) and inputs (components, factors or orders) are considered, using the methodology given in Ferraty and Vieu [2006], by means of 50 simulations.

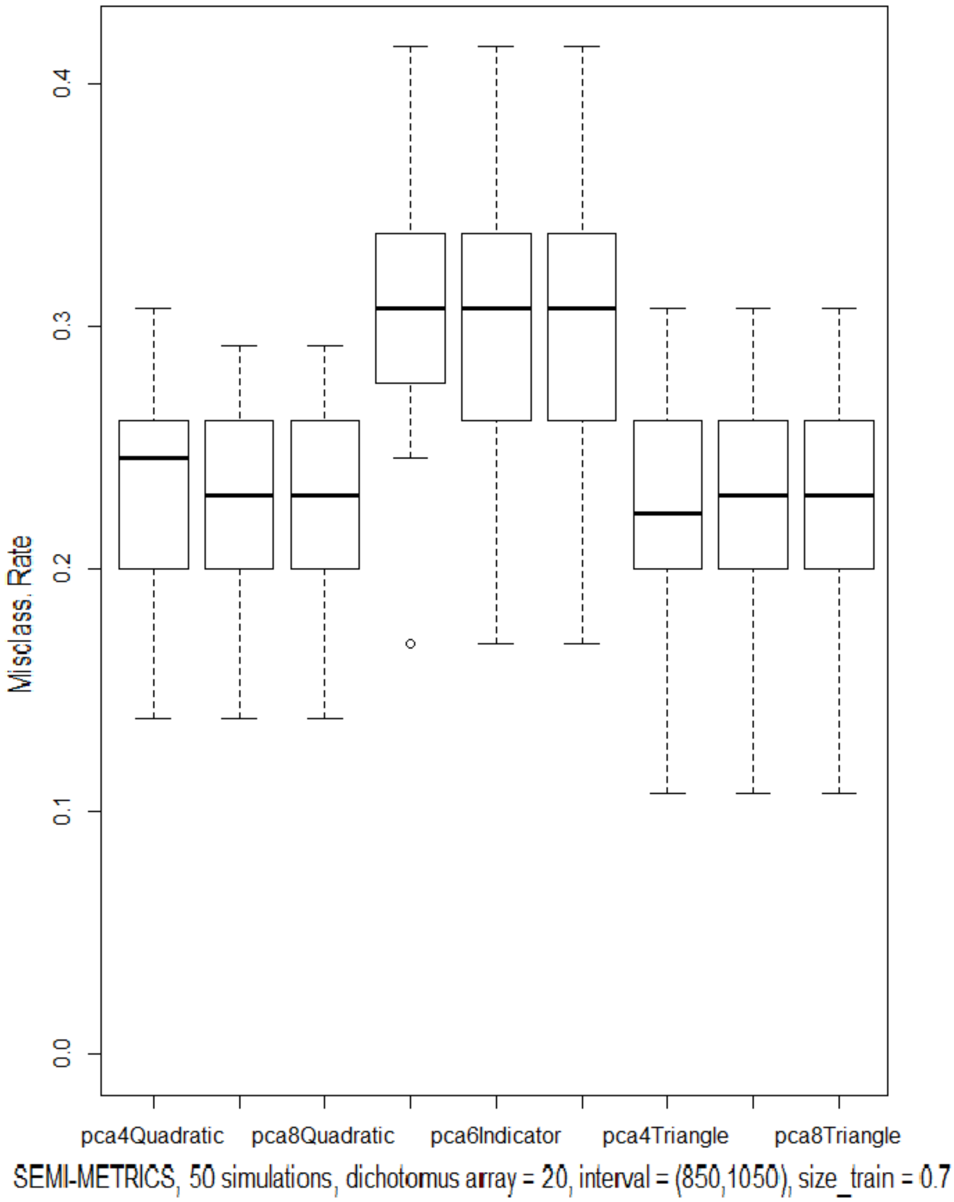


Figure A1.4.4: Misclassification rate of functional classification using the method proposed in Ferraty and Vieu [2006], with FPCA semi–metric.

Figure A1.4.5 shows the results obtained using FPLSR semi–metric, when different kernels and inputs are considered, using the methodology given in Ferraty and Vieu [2006].

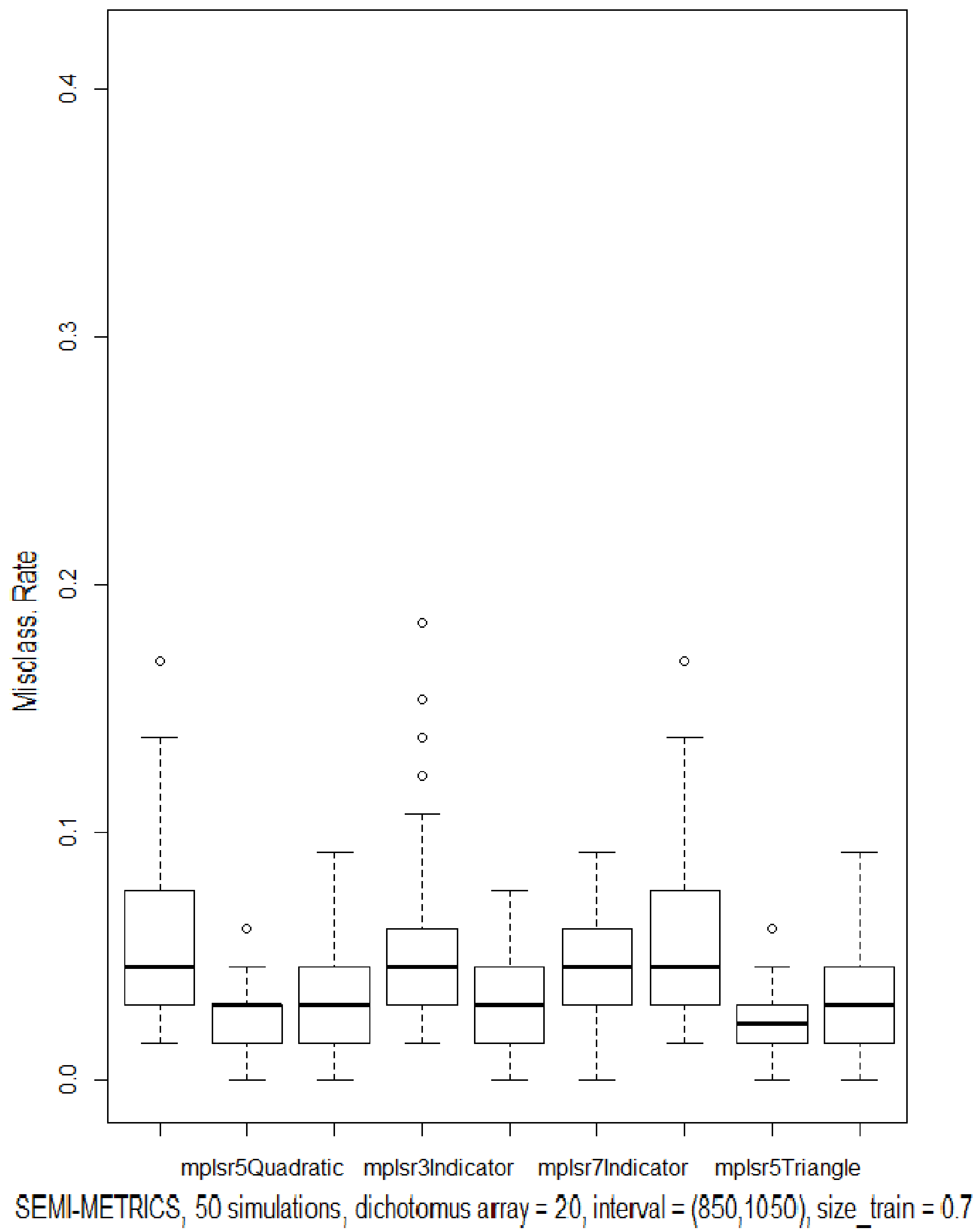


Figure A1.4.5: Misclassification rate of functional classification using the method proposed in Ferraty and Vieu [2006], with FPLSR semi–metric.

Figure A1.4.6 shows the results obtained using a semi–metric based on derivatives, when different kernels and inputs are considered.

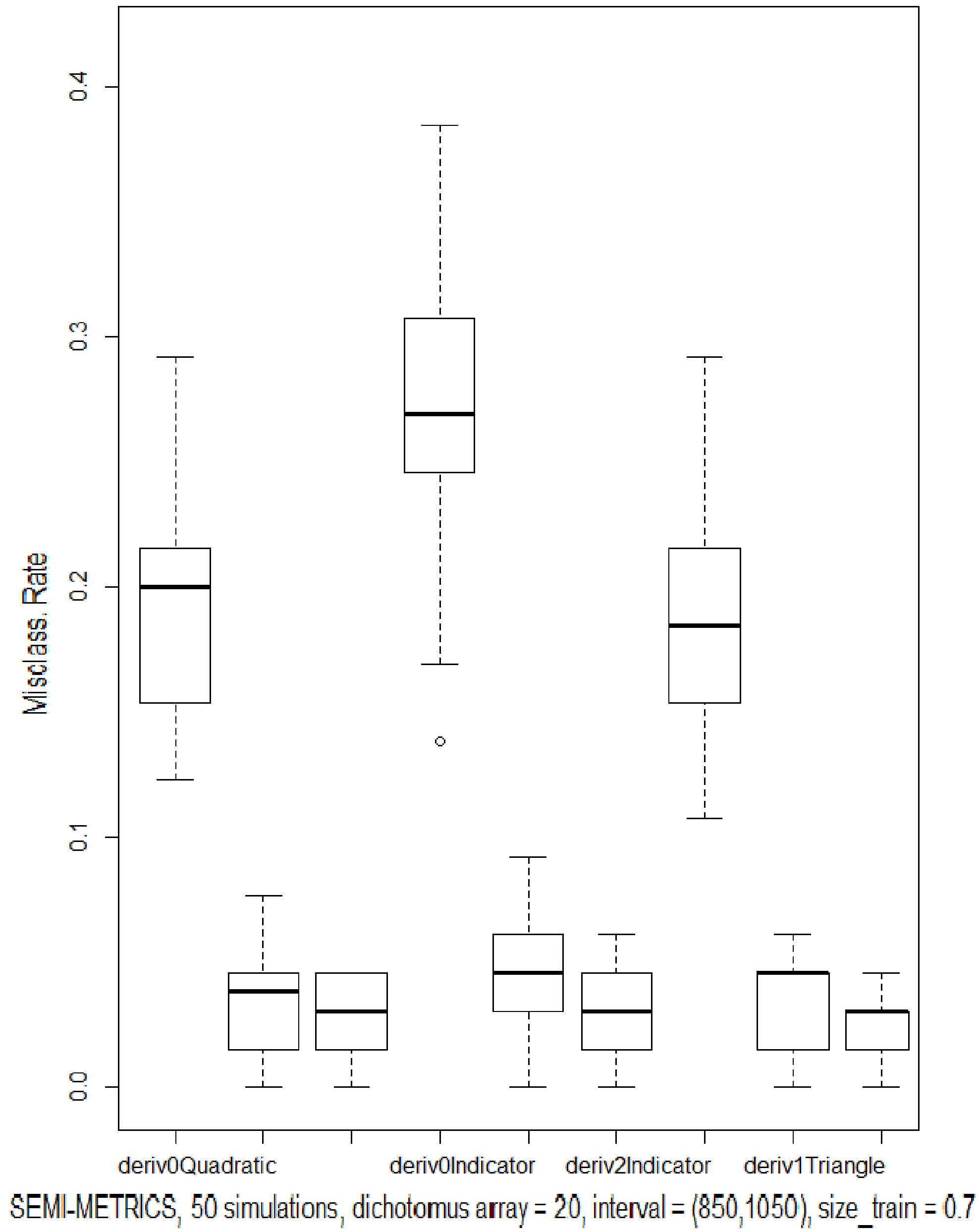


Figure A1.4.6: Misclassification rate of functional classification using the method proposed in Ferraty and Vieu [2006], with semi–metric based on derivatives.

At each one of these box–plots, we reflect results obtained with implementation of a quadratic kernel in the first three ones, the next three ones reflect results with indicator kernel, and the three last ones show the results with triangle kernel. Alternatively, Figures A1.4.7–A1.4.9 display the results using our methodology in terms of the Smolyak's quadrature rule considering three neighbours, implementing the Trapezoidal rule with $k = 5$ and using discretization step equal to 0.25.

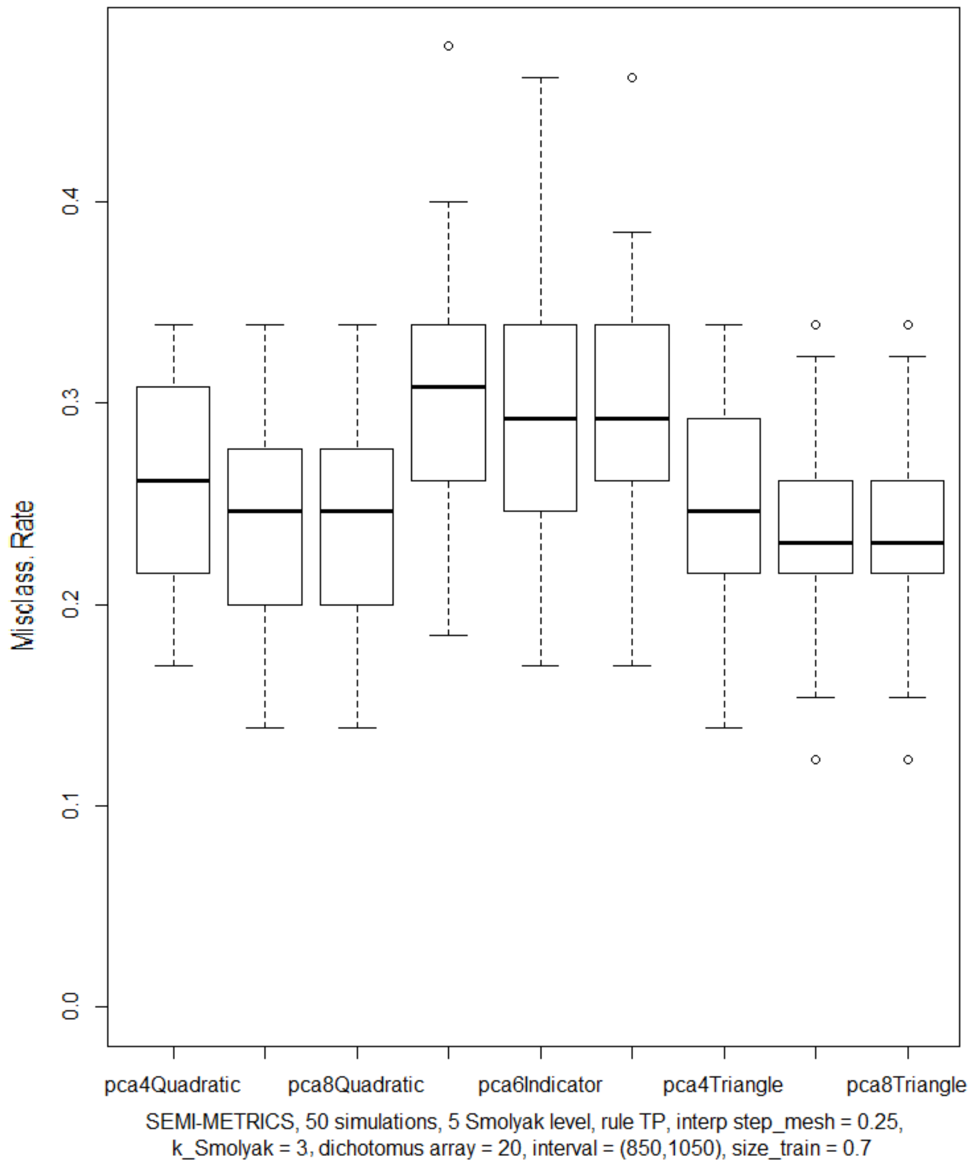


Figure A1.4.7: Results obtained with our implementation using the Trapezoidal rule (at level 5), with discretization step equal to 0.25 and 3 neighbours, with FPCA.

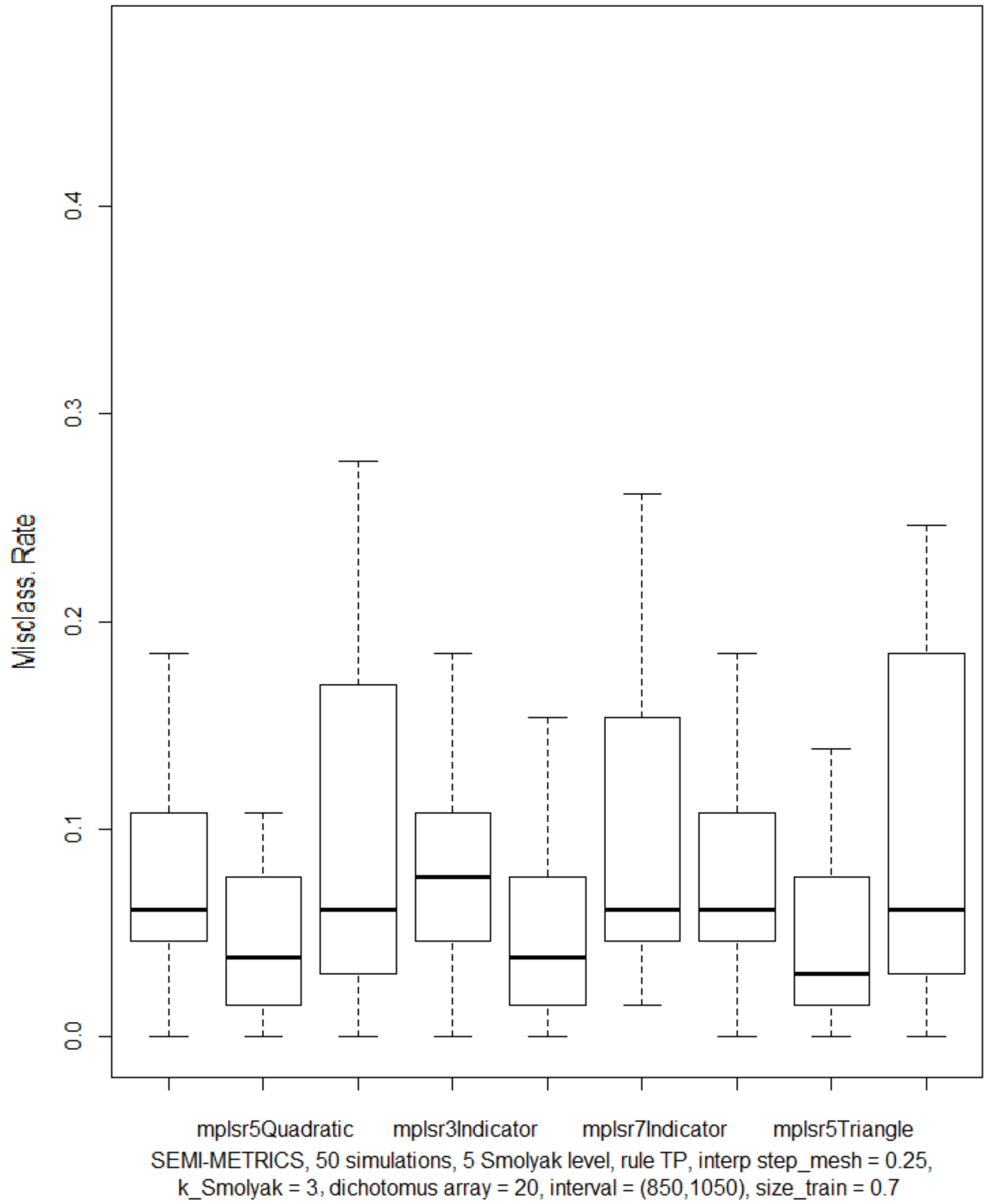


Figure A1.4.8: Results obtained with our implementation using the Trapezoidal rule (level 5), with discretization step equal to 0.25 and 3 neighbours, with FPLSR.

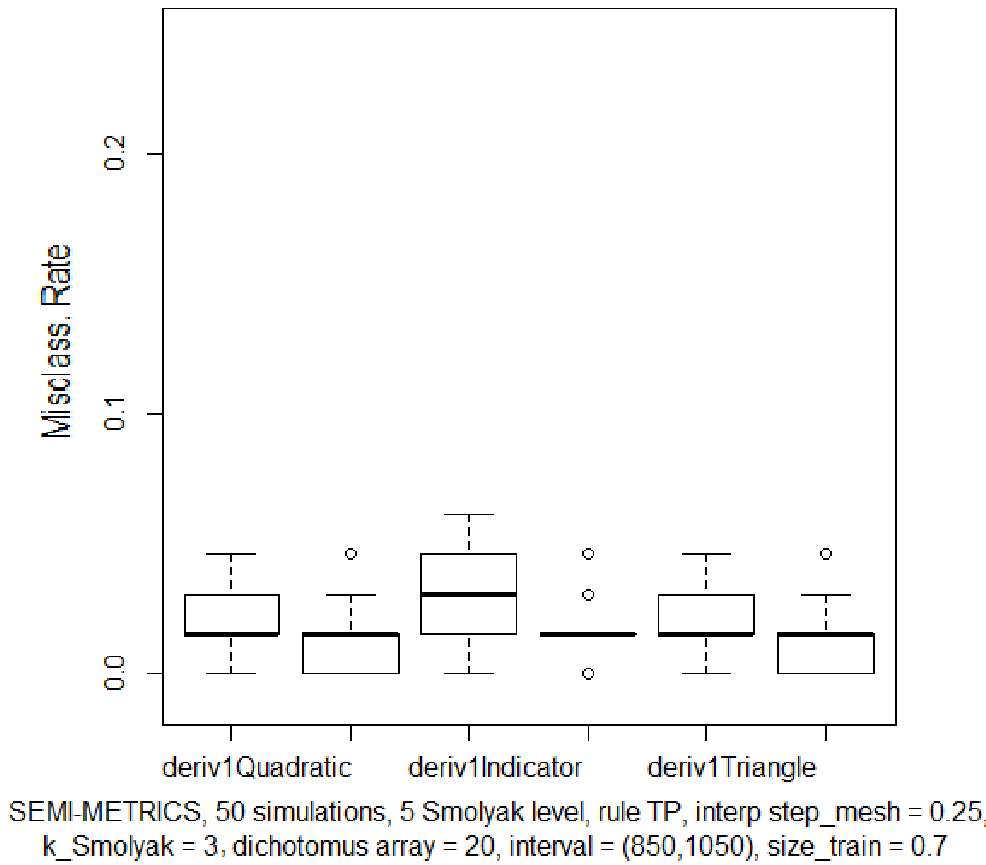


Figure A1.4.9: Results obtained with our implementation using the Trapezoidal rule (level 5), with discretization step equal to 0.25 and 3 neighbours, with semi–metric based on derivatives.

Figures A1.4.10–A1.4.12 show different implementations of our methodology with different inputs such as the Clenshaw–Curtis's quadrature rule or doing directly the assignment of Smolyak's weights. A similar performance is obtained in comparison with the previous results displayed. One can observe that our methodology is more flexible than the one presented in Ferraty and Vieu [2006]. However, our methodology is also affected by the interpolation error, and the error associated with the rule considered for the assigning of weights. This fact can also be observed in Figures A1.4.13–A1.4.14, where we have used a greater interpolation step. Note that a slight improvement in the accuracy can be appreciated. Summarizing, we have to look for a compromise between precision in the numerical approximation of the integral, increasing the number of points in the sample by interpolation, and the associated interpolation and weight allocation errors.

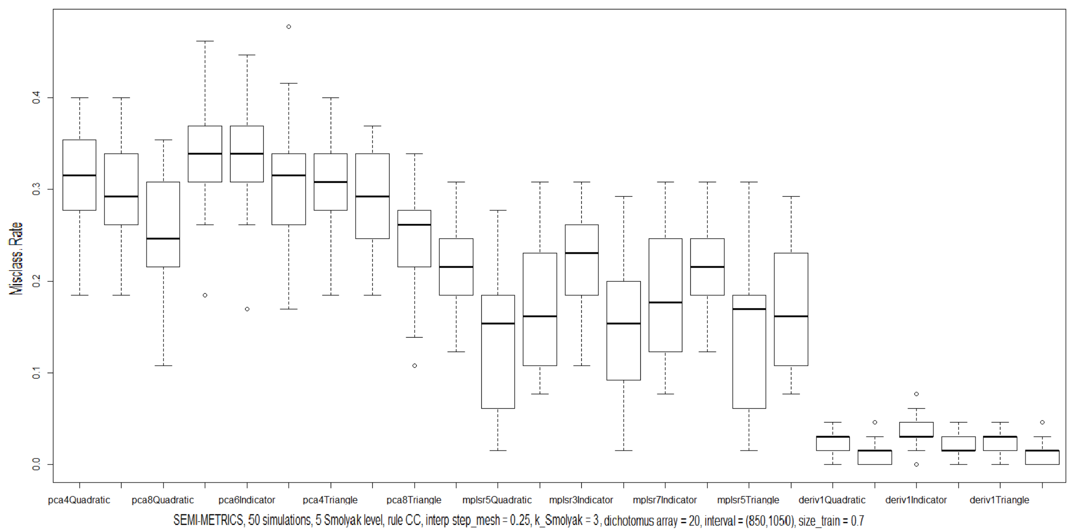


Figure A1.4.10: Results obtained with our implementation using the Clenshaw–Curtis's rule (at level 5), with discretization step equal to 0.25 and 3 neighbours.

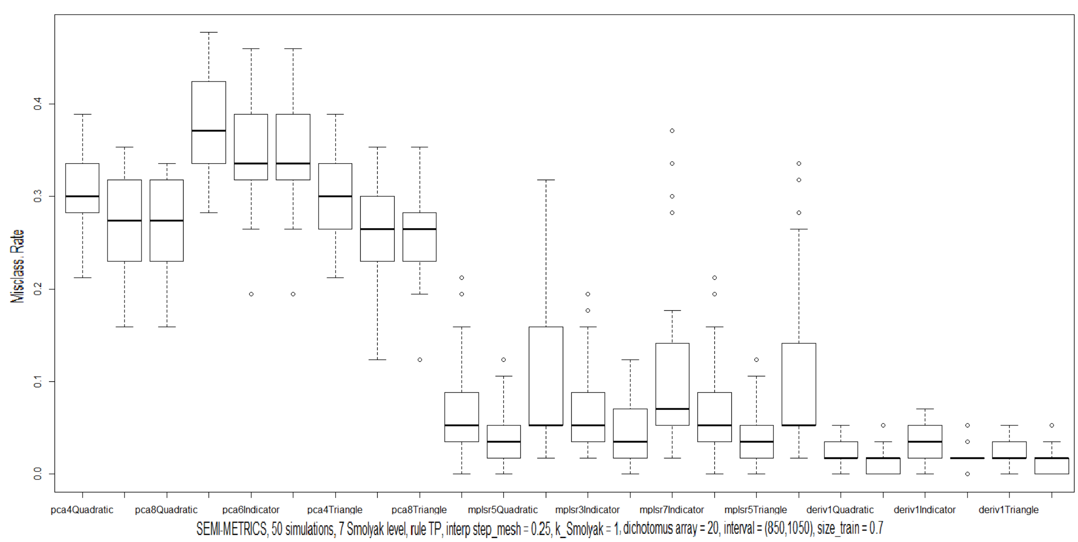


Figure A1.4.11: Results obtained with our implementation using the Trapezoidal rule (at level 7), with discretization step equal to 0.25.

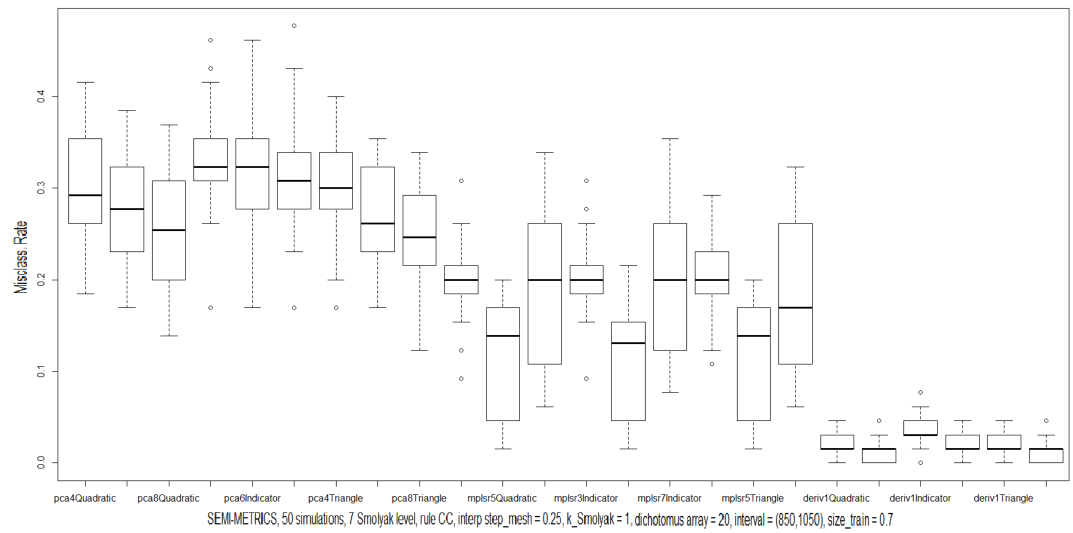


Figure A1.4.12: Results obtained with our implementation using the Clenshaw–Curtis's rule (at level 7), with discretization step equal to 0.25.

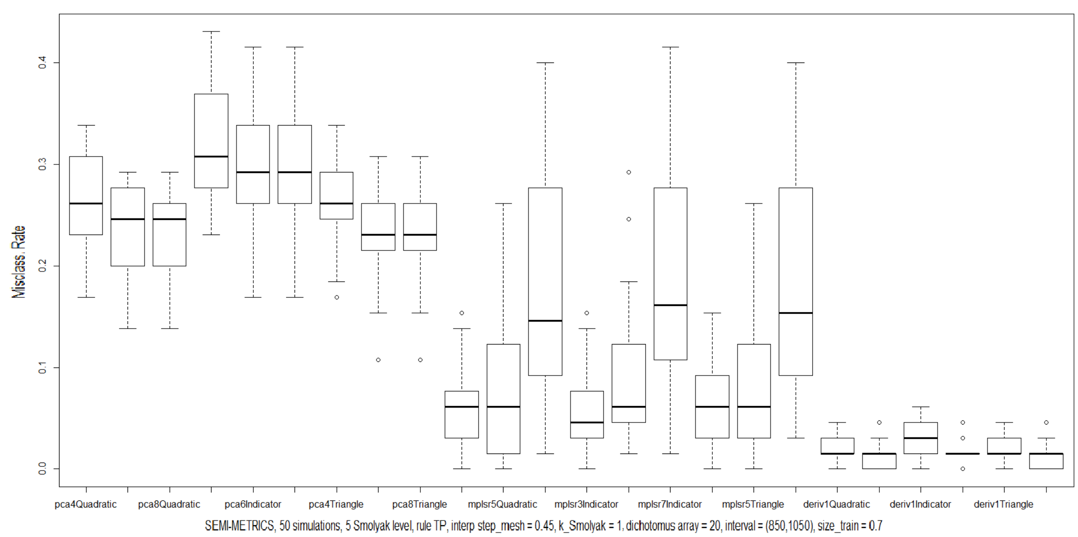


Figure A1.4.13: Results obtained with our implementation using the Trapezoidal rule (at level 5), with discretization step equal to 0.45.

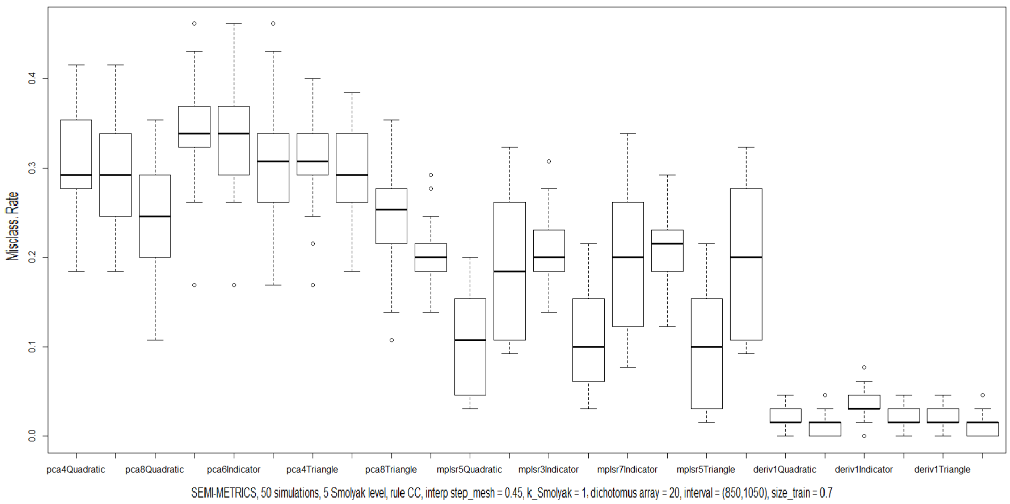


Figure A1.4.14: Results obtained with our implementation using the Clenshaw–Curtis's rule (at level 5), with discretization step equal to 0.45.

A1.5 NUMERICAL EXAMPLE FOR FUNCTIONAL CLASSIFICATION OF TREND IN RANDOM GAUSSIAN SURFACES

A sample of 200 Gaussian random surfaces is generated, over a regular grid within the square $[1, 5] \times [1, 5]$, with the same integral covariance operator defined by the isotropic Gaussian kernel in two dimensions. These Gaussian surfaces have two different (which lead to the definition of our two groups), but very close, functional means (see Figure A1.5.1). Our problem consists in discriminating between different trends defining the mean value of Gaussian surfaces. This numerical example is considered previously to our main simulation study developed in the next section where, among other subjects, we solve the problem of discrimination between different spatial correlation functions characterizing the infinite–dimensional distribution of zero–mean Gaussian random surfaces.

Let us then consider the following two groups of Gaussian random surfaces:

$$\begin{aligned} \boldsymbol{\chi}^1 &\sim \mathcal{N}\left(\boldsymbol{\mu}_1 = \{h_i,\, i = 1, \ldots, k\}, \boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}^T\right) \\ \boldsymbol{\chi}^2 &\sim \mathcal{N}\left(\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \boldsymbol{v}, \boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}^T\right) \end{aligned} \tag{A1.12}$$

where

$$h_i = \frac{i}{N \times M} + 2, \quad i = 1, \ldots, k, \quad k = N \times M,$$

with $\boldsymbol{v} = (0.5, 0.5, \ldots, 0.5) \in \mathbb{R}^{N \times M}$, and $C_{ij} = e^{-\|(x_i, y_i) - (x_j, y_j)\|_2^2}$, where $\boldsymbol{\chi}^j$ denotes a random surface of type $j = 1, 2$, with values defined over a regular grid given by $((x_1, y_1), \ldots, (x_N, y_M))$. We have imposed a minimum number of surfaces belonging to each class.

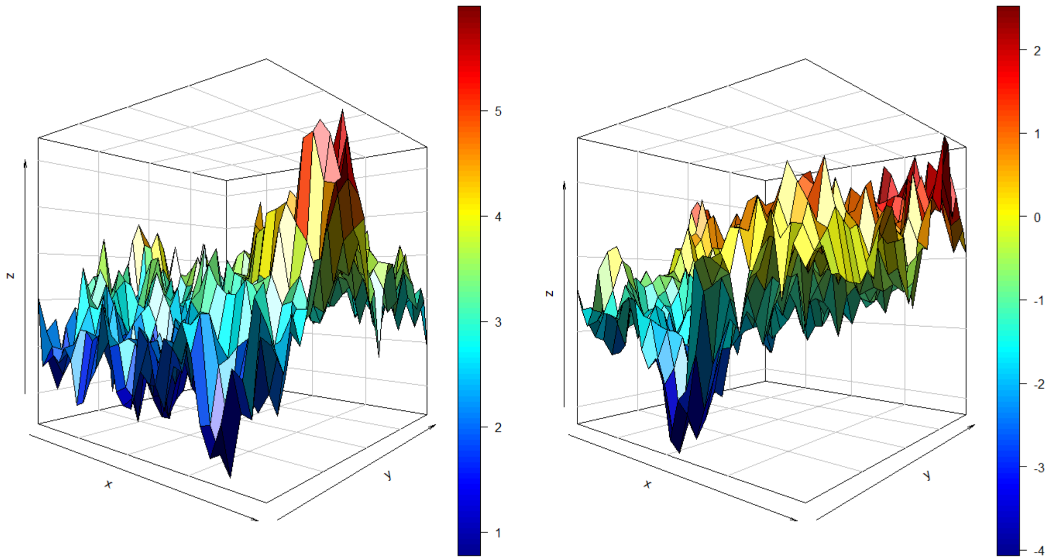


Figure A1.5.1: Surfaces plotted: on left surface belongs to category 1, on right surface belongs to category 2.

Note that now we have not to interpolate since we can generate surfaces as finely as wanted. A minimum number of surfaces belonging to each class is fixed to ensure the representativeness of the groups. As commented, in the previous implementation of our methodology in terms of curves, we restrict our attention to the FPCA and FPLSR semi-metrics. Figures A1.5.2–A1.5.3 then display the derived classification results, reflecting a good performance of our methodology for discrimination between different trends of Gaussian surfaces, keeping in mind that the two categories distinguished are very close.

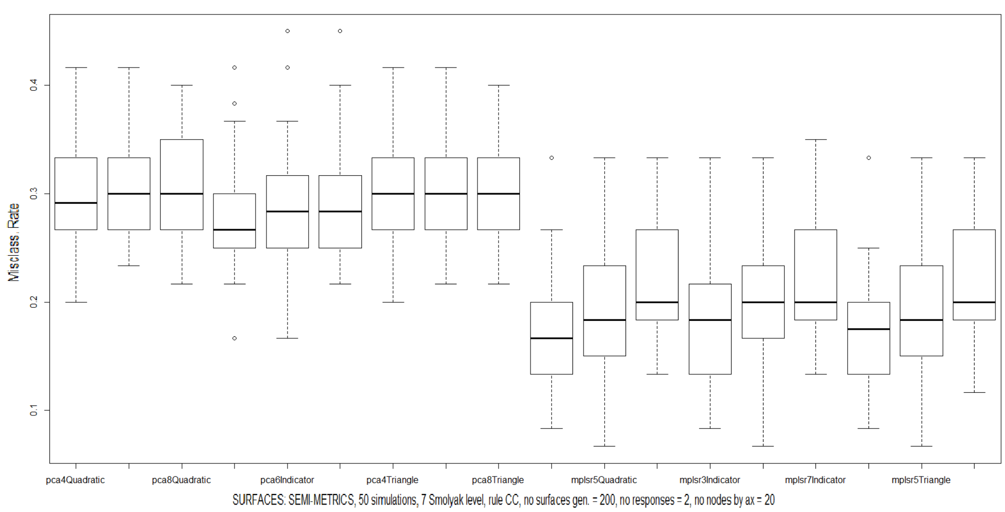


Figure A1.5.2: Results obtained with our implementation for surfaces using the Clenshaw–Curtis's rule (at level 7) on a 20×20 spatial regular grid.

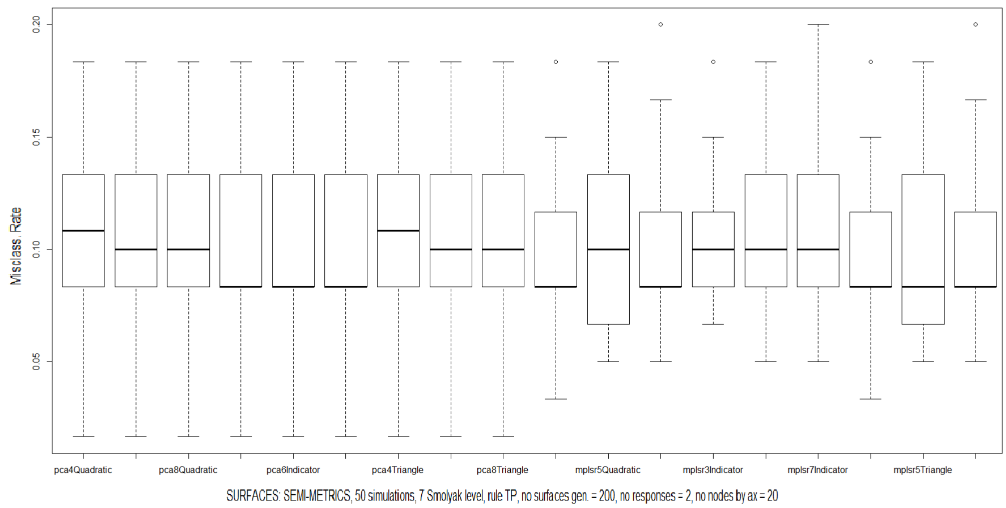


Figure A1.5.3: Results obtained with our implementation for surfaces using the Trapezoidal rule (at level 7) on a 20×20 spatial regular grid.

Consider now two groups respectively based on a linear and a non-linear, cosine type, trend surfaces:

$$\boldsymbol{\mu}_2 = \cos\left(\boldsymbol{\mu}_1 \frac{\pi}{2}\right),$$

where $\boldsymbol{\mu}_1$ is given as before. For a lower resolution level, namely for a 12×12 regular grid, FPLSR clearly outperforms FPCA (see Figure A1.5.4). Thus, FPLS is more suitable for well-differentiated groups when numerical integration must be performed from a low quality discrete version of our surface dataset.

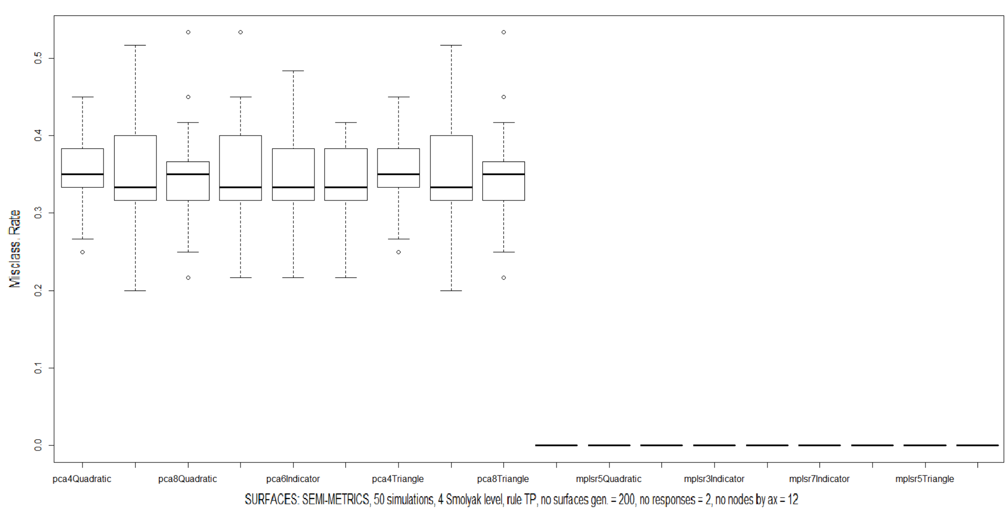


Figure A1.5.4: Results obtained with our implementation for surfaces using Trapezoidal rule (at level 4) on a 12×12 spatial regular grid.

A1.6 FUNCTIONAL CLASSIFICATION RESULTS OF RANDOM AND NON–RANDOM SURFACE IRREGULARITIES OF RAILWAY TRACK

The problem of deterministic and random vibration classification from the observation of surface irregularities of railway track will be addressed in this section, which constitutes a key problem in the field of railway engineering. As commented in Mohammadzadeh et al. [2013], it is very important to modeling these irregularities since the created loads resulting from them cause fatigue in the vehicles and rail beams. According to Youcef et al. [2013], the rail irregularities are the second leading cause of bridge vibrations, and the first one of train vibrations.

Two types of rail irregularities are studied in Mohammadzadeh et al. [2013]; Youcef et al. [2013]: random and non–random irregularities. Random irregularities include the roughness of the rails. Here, these irregularities are represented in terms of zero-mean Gaussian surfaces with different spatial functional correlations. Deterministic irregularities are usually represented in terms of a irregularity function of the railway $r(x)$ (see, for example, Fryba [1999]).

A1.6.1 Non-random surfaces irregularities

As proposed in Mohammadzadeh et al. [2013]; Youcef et al. [2013] approach, an one–dimensional railway track is firstly considered. We can see in the example shown in FigureA1.6.1 that a simple beam of span length $L = 50\,m$ is analysed. We denote as B the distance from the origin to the first irregularity, and A the constant rail length between two imperfections (see Figure A1.6.1).

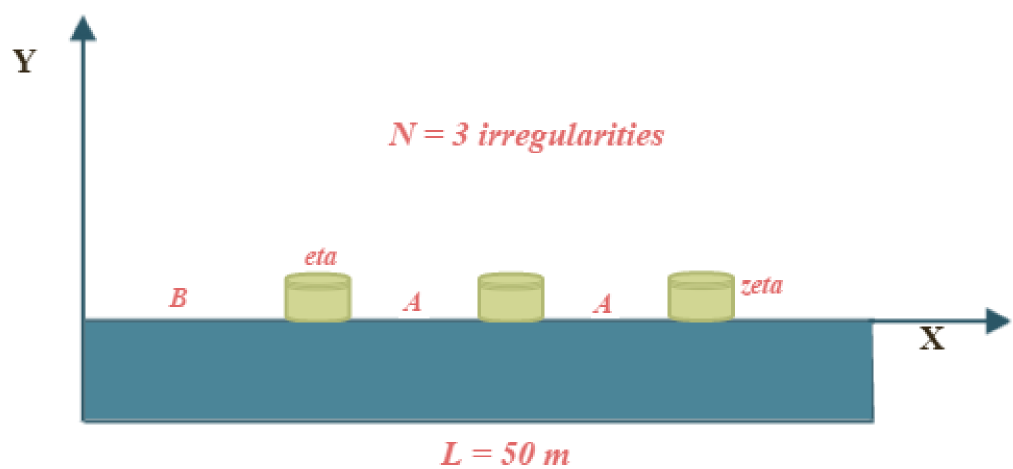


Figure A1.6.1: Illustrative and very simple example of the approach, with $L = 50\ m$ and $N = 3$.

Setting the number of these irregularities in the railway track of length L, denoted as N, and considering the depth and the length of the imperfections (ζ and η respectively, as shown in Figure A1.6.1), we can establish the following formula:

$$N = \frac{L - B}{A + \eta} \tag{A1.13}$$

Let us now consider three different models of imperfections: $N_1 = 3$ and $B_1 = 4.5$; $N_2 = 4$ and $B_2 = 2.5$ and $N_3 = 5$ and $B_3 = 1$. We divide each one of them into two models using different values of A, and using two different values of ζ, the final set of models is given in Table A1.6.1:

Table A1.6.1: Final models.

Models	N	B (m)	A (m)	ζ (m)
Model 1	3	4.5	3.5	0.007
Model 2	3	4.5	5.2	0.007
Model 3	3	4.5	3.5	0.015
Model 4	3	4.5	5.2	0.015
Model 5	4	2.5	3.5	0.007
Model 6	4	2.5	5.2	0.007
Model 7	4	2.5	3.5	0.015
Model 8	4	2.5	5.2	0.015
Model 9	5	1	3.5	0.007
Model 10	5	1	5.2	0.007
Model 11	5	1	3.5	0.015
Model 12	5	1	5.2	0.015

Choosing any of them, and using formula (A1.13), we can get the corresponding set of values of η (see Table A1.6.2):

Table A1.6.2: Set of values of η.

Models	η (m)
Model 1	11.667
Model 2	9.967
Model 3	11.667
Model 4	9.967
Model 5	8.375
Model 6	6.675
Model 7	8.375
Model 8	6.675
Model 9	6.300
Model 10	4.600
Model 11	6.300
Model 12	4.600

As proposed in Fryba [1999], for each one of these models, denoted as $\{M_i,\ i = 1, \ldots, 12\}$, the non–random irregularities can be mathematically defined by the following function:

$$r\left(x\right) = \begin{cases} \frac{\varsigma}{2}\left(1 - cos\left(\frac{2\pi x}{\eta}\right)\right) & \text{if } C \leq x \leq C + \eta \\ 0 & \text{elsewhere} \end{cases} \tag{A1.14}$$

where $C = B + k\left(A + \eta\right),\ k = 0, 1, \ldots, N.$

As we want to deal with surfaces, in this paper we shall extend this approach to the two-dimensional framework. Such as the rail width is quite smaller than L, we use a anisotropic model where the imperfections are deployed through the x-axis (see Figures A1.6.2).

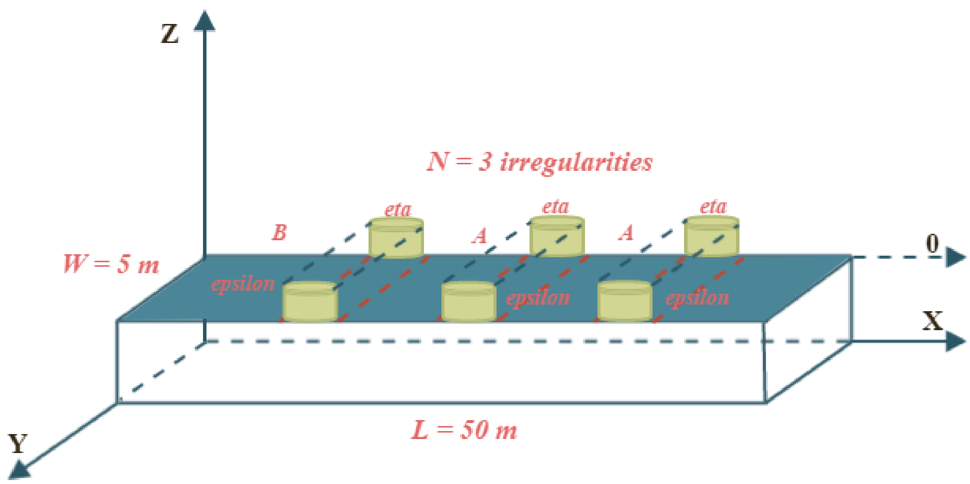


Figure A1.6.2: Illustrative and very simple example of the two-dimensional approach, with $L = 50\ m$, $W = 2.5\ m$ and $N = 3$.

Extending equation (A1.14) and setting $W = 2.5$, we have

$$r\left(x,y\right) = \begin{cases} \frac{\varsigma}{2}\left(1 - cos\left(\frac{2\pi x}{\eta}\right)\right) & \text{if } C \leq x \leq C + \eta \\ 0 & \text{elsewhere} \end{cases}$$

where $C = B + k\left(A + \eta\right),\ k = 0, 1, \ldots, N$ and $y \in \left[0, W\right]$.

As well as in previous sections we have been working with a square regular grid, where the length coincides with the width, a rectangular grid, with $L = 50\ m$ and $W = 2.5\ m$ is used now. In Appendix A1.3.2, for simplicity we have assumed $I^n = I \times \cdots \times I$, but this implementation used in the previous section is not valid here, and we have to recalculate all the steps of the proposed numerical integration algorithm for functional classification of noisy Gaussian surfaces. We have then obtained from formula (A1.11):

$$Q_k^n = \sum_{l=m}^{k} \sum_{\substack{\|\boldsymbol{\alpha}\|_1 = l \\ \boldsymbol{\alpha}\in\mathbb{N}^n \\ \boldsymbol{\alpha}\geq\mathbf{1}}} \sum_{j_1=1}^{k_{\alpha_1}} \cdots \sum_{j_n=1}^{k_{\alpha_n}} c(k,n,l) w_{j,\boldsymbol{\alpha}} f\left(\boldsymbol{x}_{j,\boldsymbol{\alpha}}\right) \tag{A1.15}$$

where $c(k,n,l) = (-1)^{k-l}\binom{n-1}{k-l}$, $w_{j,\boldsymbol{\alpha}} = w_{j_1}^{(\alpha_1)} \ldots w_{j_n}^{(\alpha_n)}$ and $\boldsymbol{x}_{j,\boldsymbol{\alpha}} = \left(x_{j_1}^{(\alpha_1)} \ldots x_{j_n}^{(\alpha_n)}\right)$.

Note that, in the previous section, $x_{j_i}^{(\alpha_i)} \in I$, for all $i = 1, \ldots, n$. However, we now compute $x_{j_i}^{(\alpha_i)}$ such as $x_{j_i}^{(\alpha_i)} \in I_i\ \forall i = 1, \ldots, n$. Rewritting (A1.15), we obtain:

$$Q_k^{n,\boldsymbol{L}} = \sum_{l=m}^{k} \sum_{\substack{\|\boldsymbol{\alpha}\|_1 = l \\ \boldsymbol{\alpha}\in\mathbb{N}^n \\ \boldsymbol{\alpha}\geq\mathbf{1}}} \sum_{j_1=1}^{k_{\alpha_1}} \cdots \sum_{j_n=1}^{k_{\alpha_n}} c(k,n,l) w_{j,\boldsymbol{\alpha}} f\left(\boldsymbol{x}_{j,\boldsymbol{\alpha}}\right)$$

where $c(k,n,l) = (-1)^{k-l}\binom{n-1}{k-l}$, $w_{j,\boldsymbol{\alpha}} = w_{j_1}^{(\alpha_1)} \dots w_{j_n}^{(\alpha_n)}$ are the weights of the $U_{l_j}^{(j)}$ univariate quadrature in I_j, $\boldsymbol{x}_{j,\boldsymbol{\alpha}} = \left(x_{j_1}^{(\alpha_1)} \dots x_{j_n}^{(\alpha_n)}\right) \in I_1 \times \cdots \times I_n$ and $\boldsymbol{L}$ is an interval matrix where $L_i j = a_{ij},\ i = 1,\dots,n,\ j = 1,2$, with $I_i = (a_{i1}, a_{i2})$.

Figures A1.6.3–A1.6.5 provide a zoom of the generated irregularity models for the two-dimensional deterministic case.

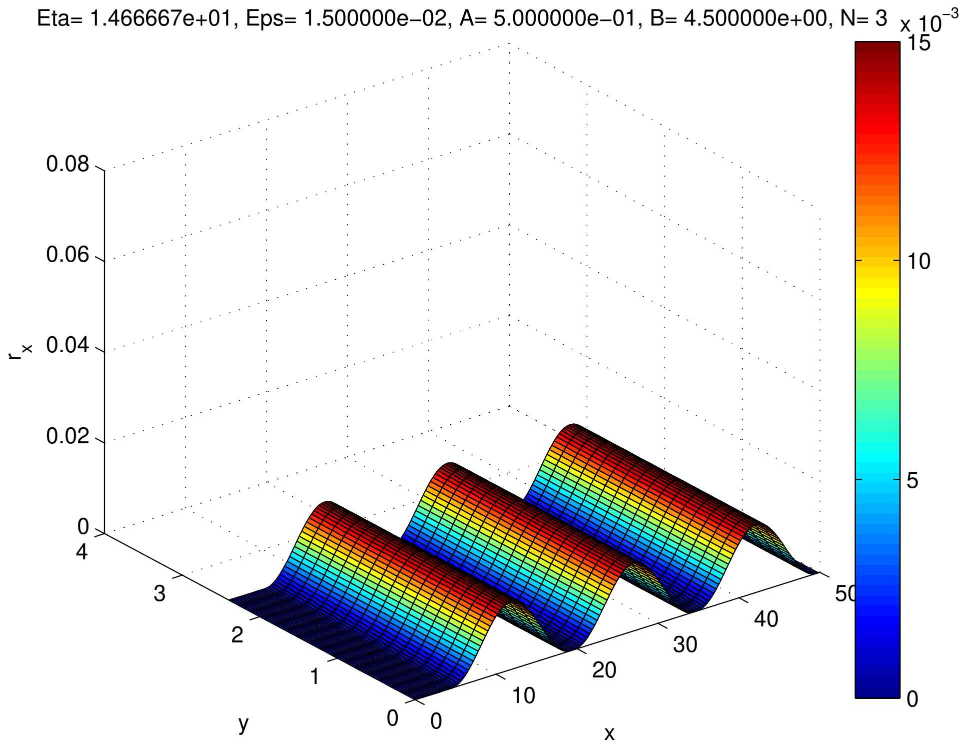


Figure A1.6.3: Irregularity belongs to model M_3.

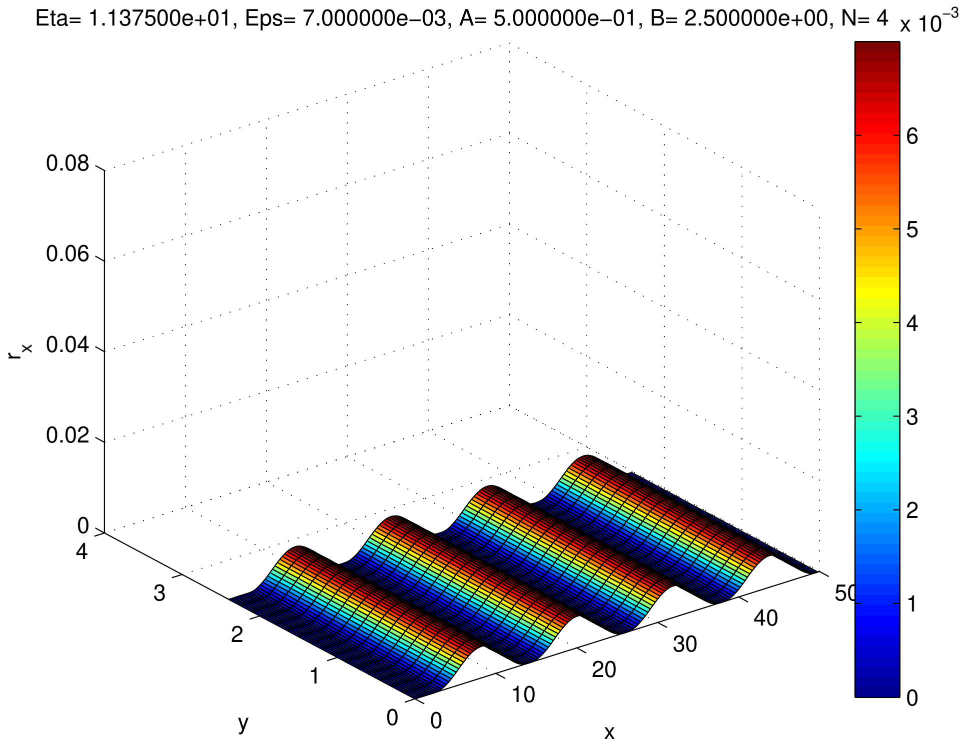


Figure A1.6.4: Irregularity belongs to model M_5.

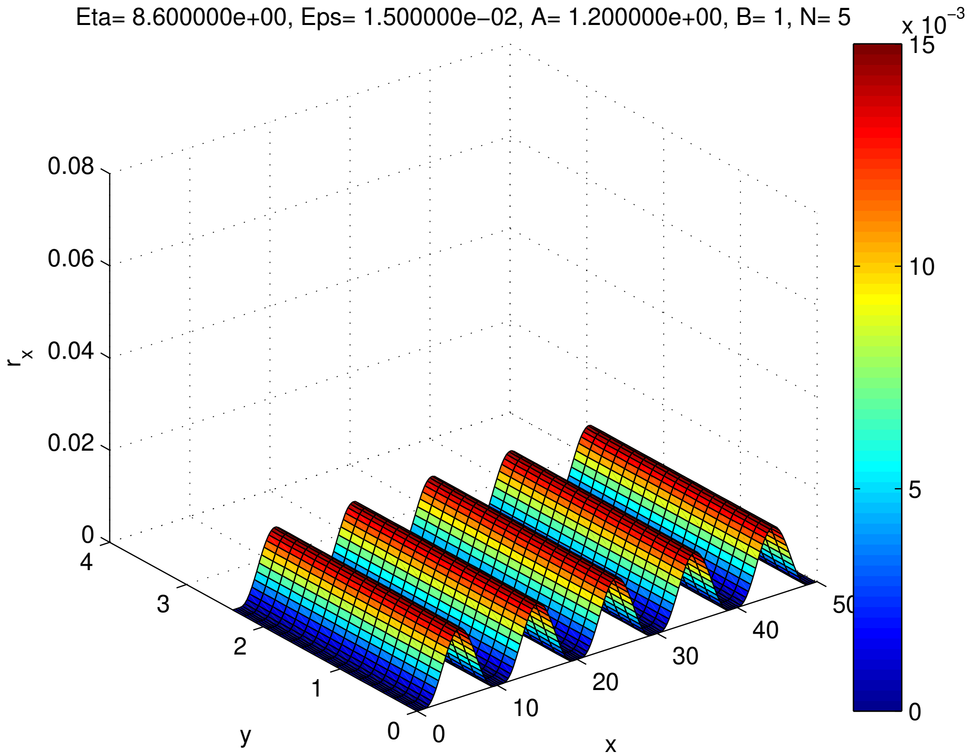


Figure A1.6.5: Irregularity belongs to model M_{12}.

It is assumed that our observed irregularities are measured by a device that introduces and additive zero-mean Gaussian noise. That is, they are perturbed by such a noise as follows:

$$S_i\left(x,y\right) = r\left(x,y\right) + \varepsilon\left(x,y\right) \tag{A1.16}$$

for $i = 1,\ldots,12$, models considered, and for $\varepsilon\left(x,y\right) \sim N\left(\boldsymbol{\mu} = 0, \boldsymbol{\Sigma} = \sigma_i^2 \boldsymbol{Id}\right)$ being a Gaussian white noise with standard deviation $\sigma_i = \frac{\eta_i}{2}$, $i = 1,\ldots,12$. Figures A1.6.6–A1.6.8 show again a zoom of the perturbed Gaussian surfaces.

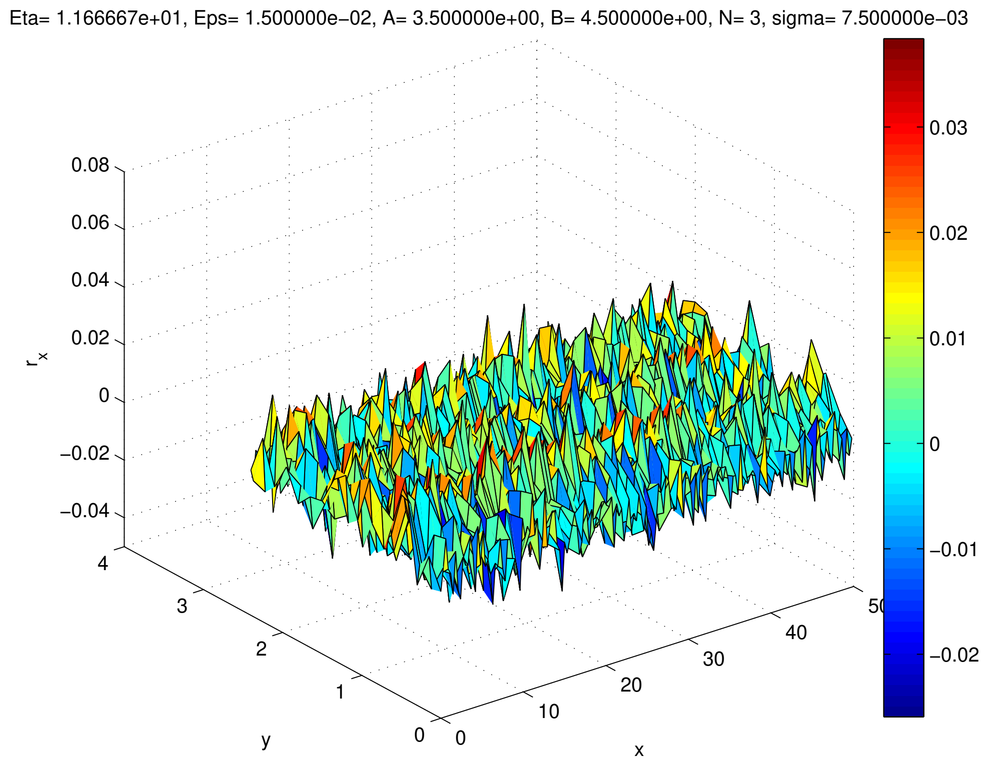


Figure A1.6.6: Irregularity perturbed belongs to model M_3.

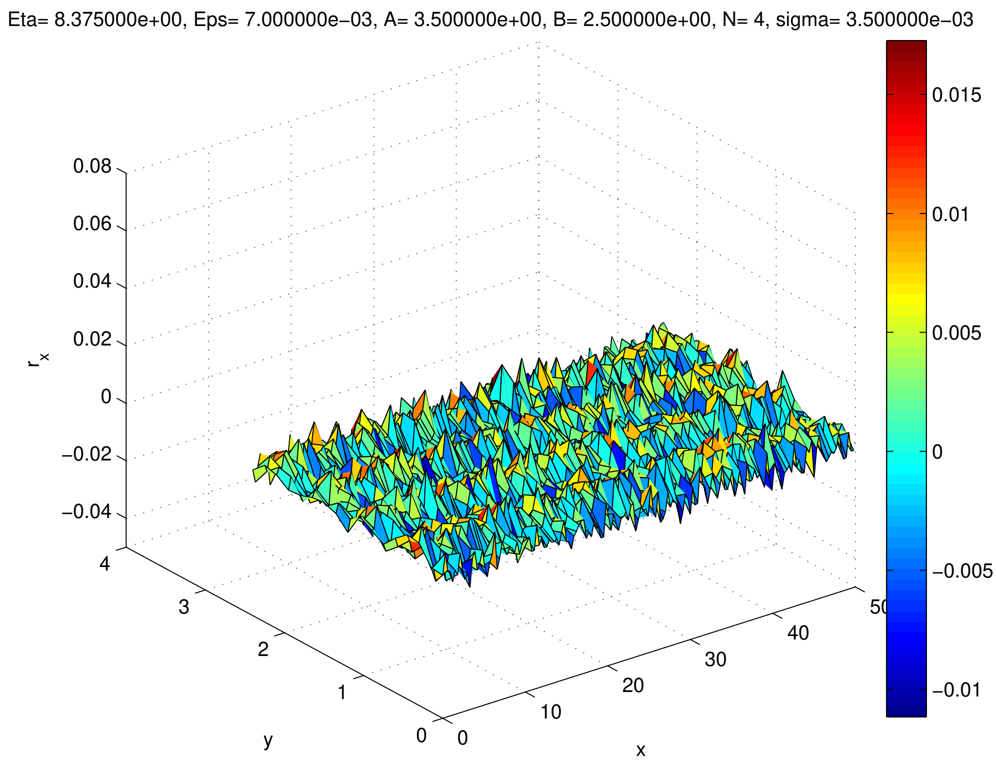


Figure A1.6.7: Irregularity perturbed belongs to model M_5.

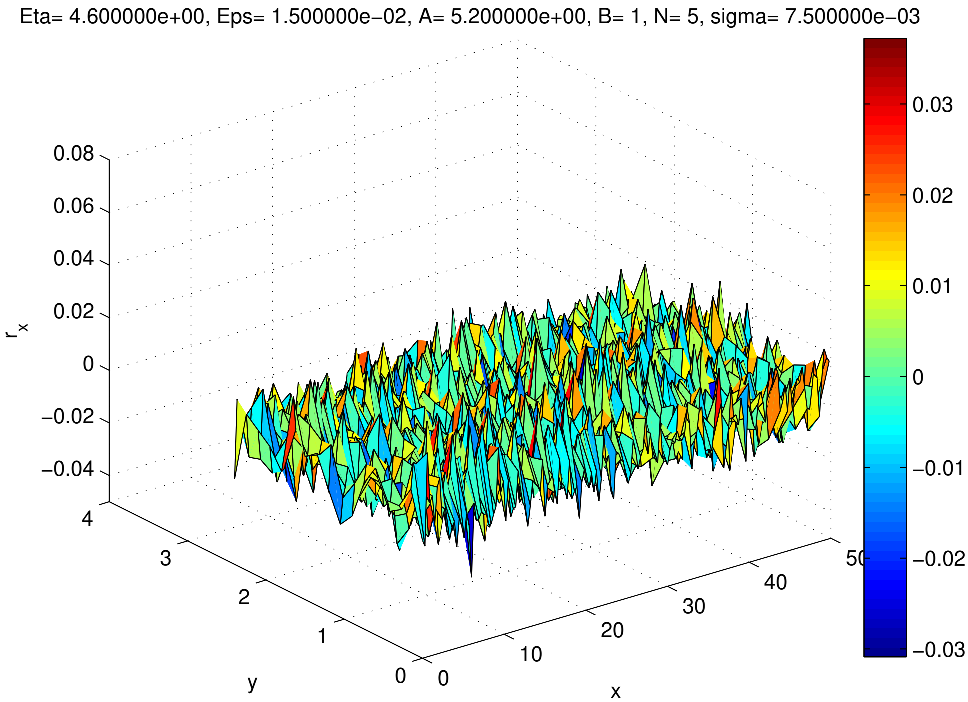


Figure A1.6.8: Irregularity perturbed belongs to model M_{12}.

We consider a regular grid corresponding to discretization steps 1.3 in length, and 0.3 in width. A minimum number of surfaces belonging to each group has been set and 50 simulations have been running. Applying the same methodology as the one used in Appendix A1.5 with a sample of 500 surfaces, we obtain the results shown in Figure A1.6.9.

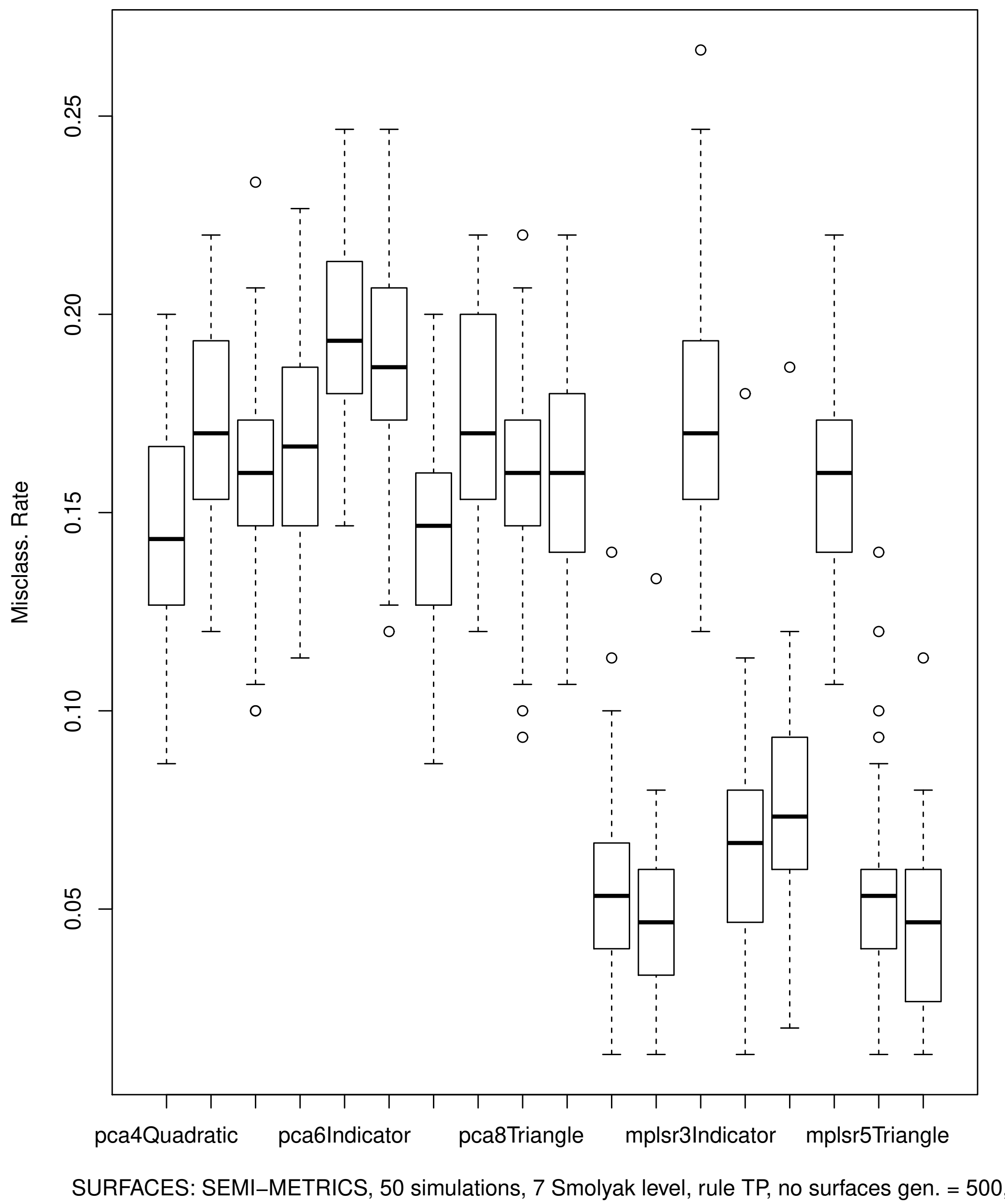


Figure A1.6.9: Results obtained with our implementation for non–random irregularities using the Trapezoidal rule (at level 7).

Remark that the accuracy depends on the magnitude of σ_i, $i = 1, \ldots, 12$, and the length of the gap between the irregularities (A). One can observe that to a greater σ_i, $i = 1, \ldots, 12$, corresponds a better performance. For the same reason, we get a better accuracy using a greater value of A (see Figure A1.6.10).

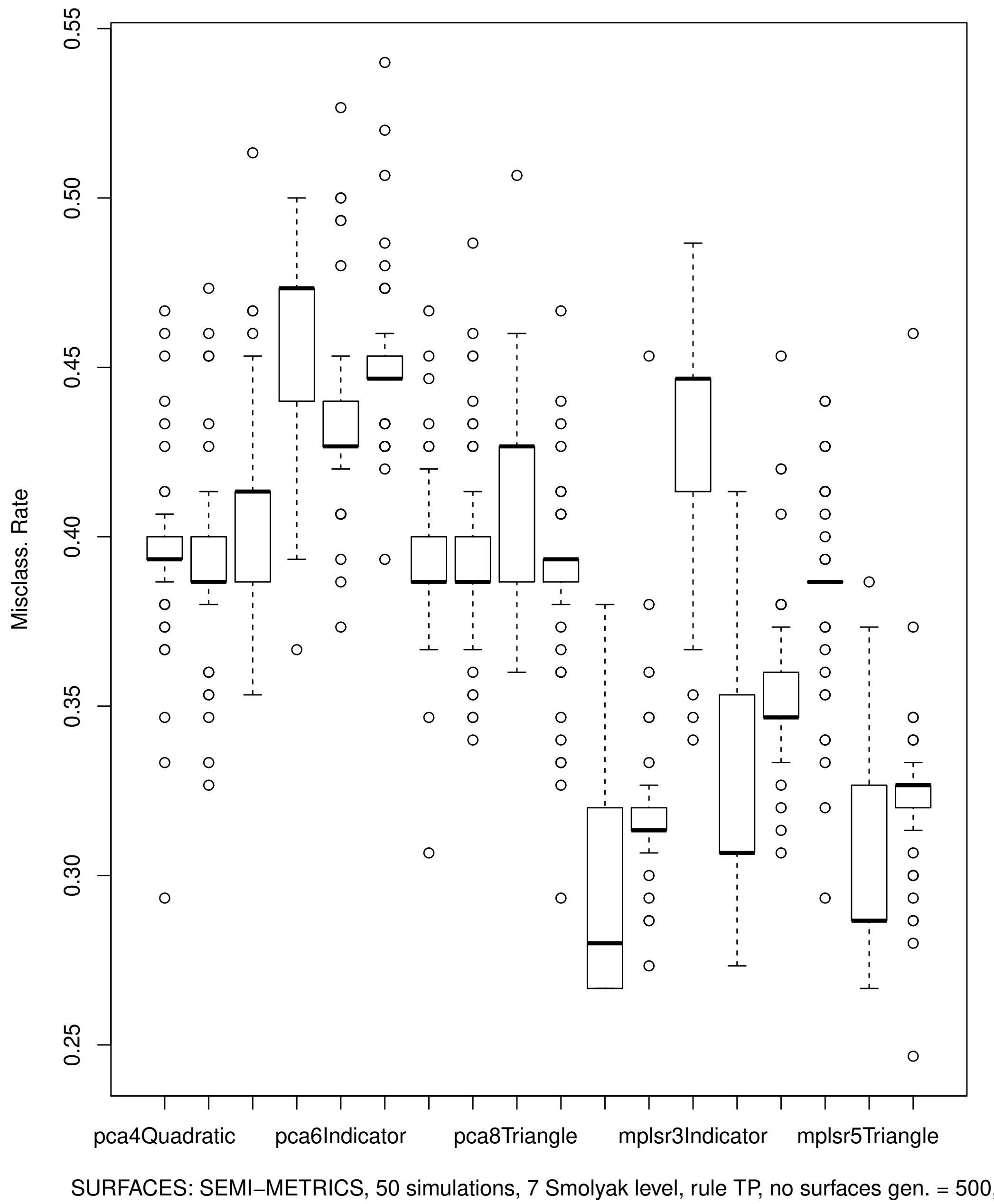


Figure A1.6.10: Results obtained with our implementation for non-random irregularities using the Trapezoidal rule (at level 7) and $\boldsymbol{A}_{new} = (1.5, 2.2)$ instead of previous $\boldsymbol{A} = (3.5, 5.2)$.

A1.6.2 Random surfaces irregularities

As commented before, rail imperfections can be divided into deterministic and random imperfections. Different factors may be the cause of these random irregularities, as imperfections in material or in rail joints, errors during design, among others.

We are going to focus on the little roughness of the rails, which is included in random imperfections, by means of Gaussian surfaces. Since we will consider little roughness, distributions with null mean will be considered, taking into account that the origin of ordinate axis is represented by the rail. Generating a sample of 200 Gaussian surfaces, we will distinguish the following four categories of roughness (see Figures A1.6.11–A1.6.14):

$$\boldsymbol{\chi}^h \sim \mathcal{N}\left(\boldsymbol{\mu}_h = 0, \boldsymbol{\Sigma} = \boldsymbol{C}_h \boldsymbol{C}_h^T\right) \tag{A1.17}$$

where $h = 1, 2, 3, 4$ identifies our categories, and

$$C_{hij} = \frac{k_h}{LW} e^{\frac{-\left\| \left(\frac{x_i}{L}, \frac{y_i}{W}\right) - \left(\frac{x_j}{L}, \frac{y_j}{W}\right) \right\|_2}{k_h}} \quad (h = 1, 3)$$

represents the correlation structure model for each group $h = 1, 3$, within the family of Ornstein–Uhlenbeck covariance kernels, and

$$C_{hij} = \frac{k_h}{LW} e^{\frac{-\left\| \left(\frac{x_i}{L}, \frac{y_i}{W}\right) - \left(\frac{x_j}{L}, \frac{y_j}{W}\right) \right\|_2^2}{k_h}} \quad (h = 2, 4)$$

within the family of spatial correlations functions given by the non–linear isotropic Gaussian kernel, using a vector of scales $\boldsymbol{k}_h = (0.04, 0.04, 0.06, 0.06)$. Both correlation models correspond to weak dependence in space (see Figures A1.6.11–A1.6.14). As previously, a minimum number of surfaces belonging to each class has been fixed, and the two-dimensional rectangle $[0, L] \times [0, W]$ has been considered, with discretization step size 1.3 in length, and discretization step size 0.3 in width.

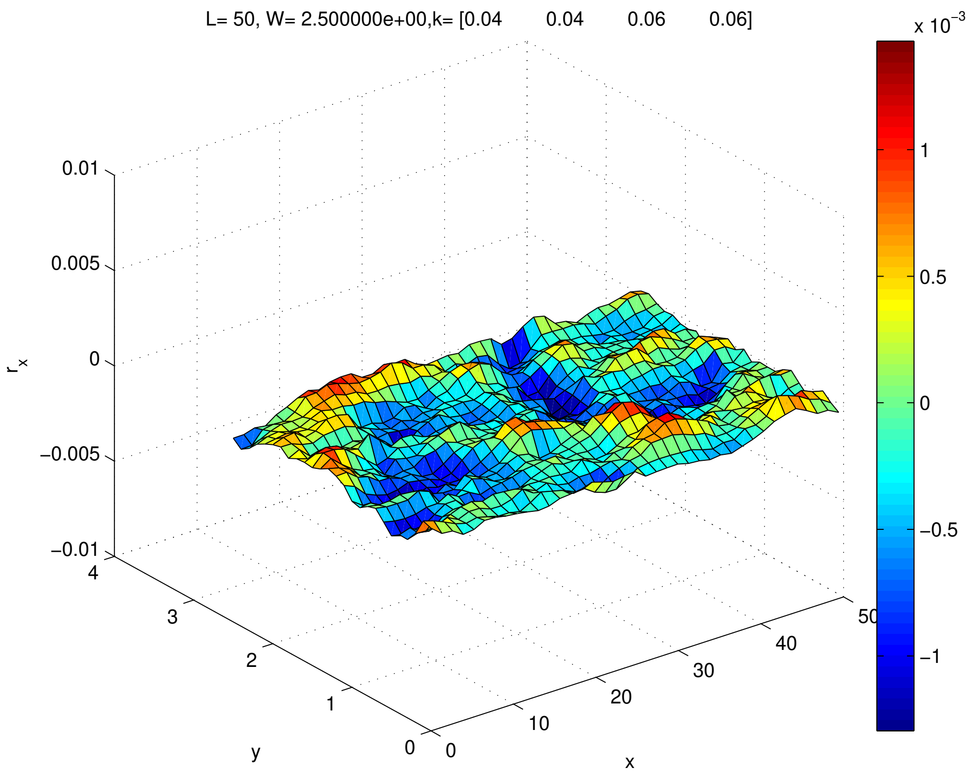


Figure A1.6.11: Random surface belongs to category 1, using the isotropic Ornstein–Uhlenbeck covariance kernel and $k_h = 0.04$ (weak correlated model).

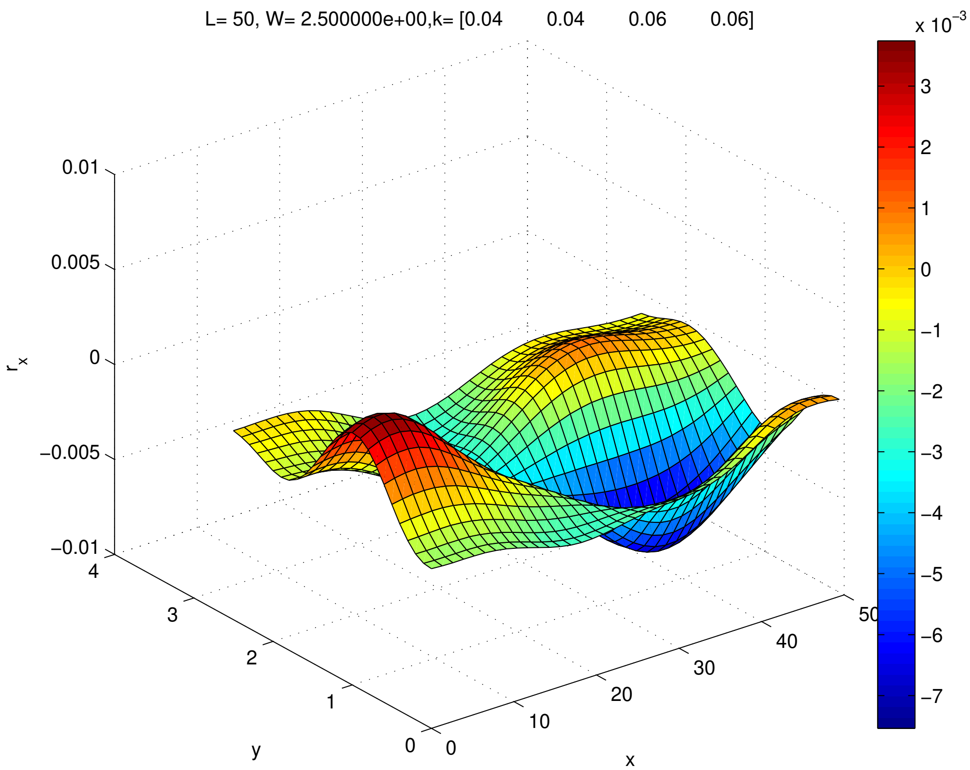


Figure A1.6.12: Random surface belongs to category 2, using the isotropic Gaussian covariance kernel and $k_h = 0.04$ (weak correlated model).

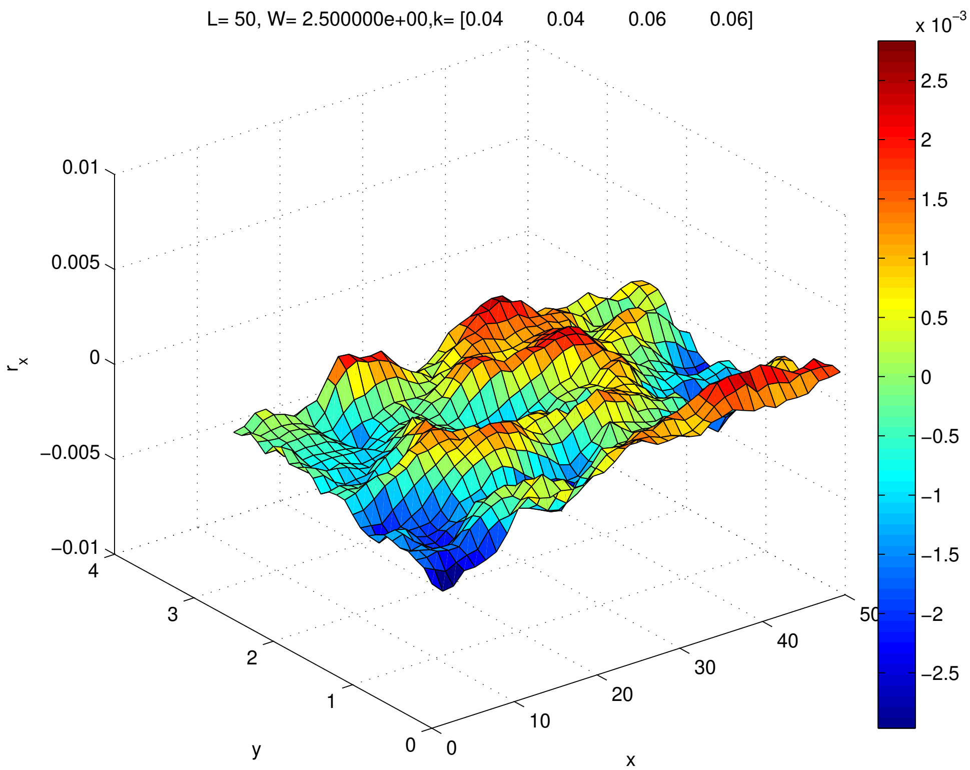


Figure A1.6.13: Random surface belongs to category 3, using the isotropic Ornstein–Uhlenbeck covariance kernel and $k_h = 0.06$ (strong correlated model).

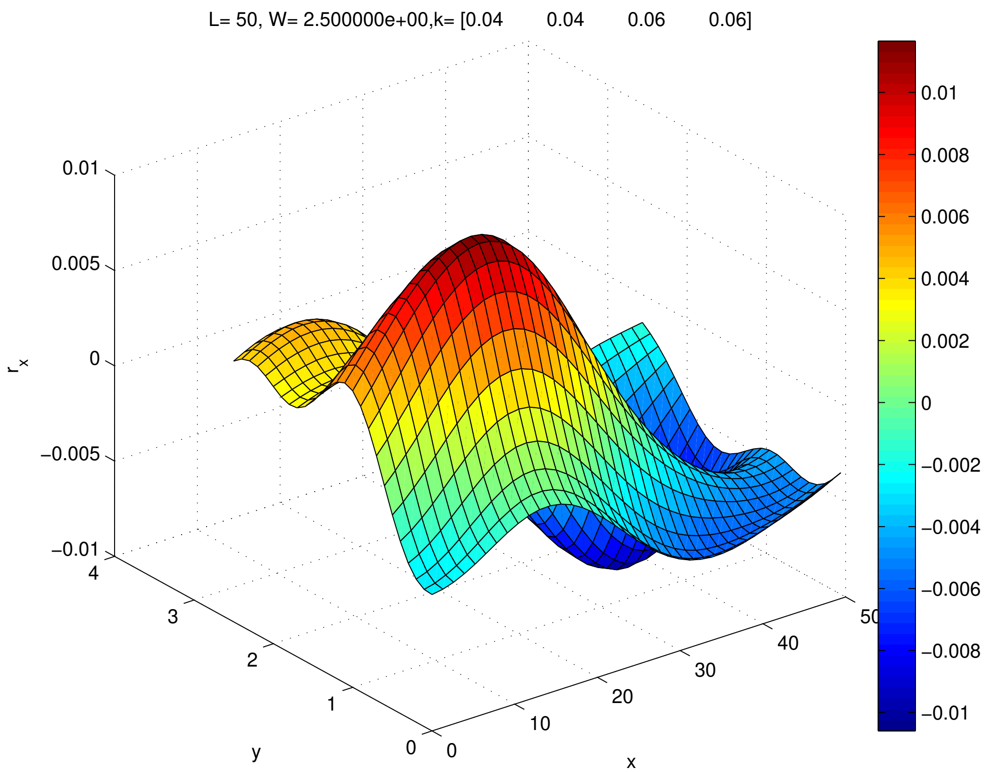


Figure A1.6.14: Random surface belongs to category 4, using the isotropic Gaussian covariance kernel and $k_h = 0.06$ (strong correlated model).

One can observe from the random surfaces displayed in Figures A1.6.11–A1.6.14, that the random surfaces with covariance matrix given by the isotropic Gaussian kernel display a smoother local behaviour than the ones with Ornstein–Uhlenbeck correlation kernel. Note that the first ones display stronger spatial correlations (see equation (A1.17)). Parameter k_h within each spatial functional correlation family represents the spatial dependence range (scale parameter) of each random surface class. Applying our methodology as in Appendix A1.6.1 with a sample of 200 surfaces, the following results are obtained (see Figure A1.6.15):

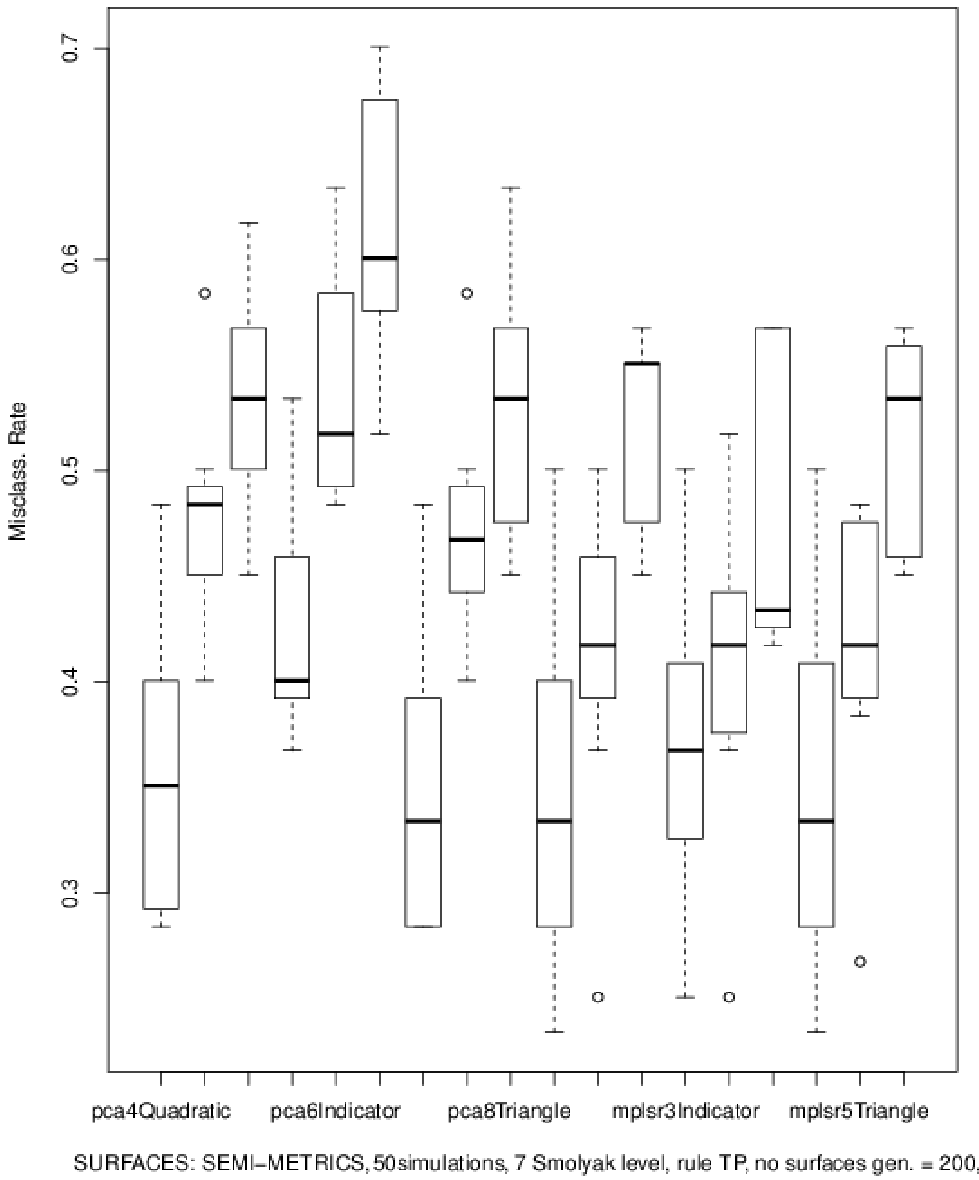


Figure A1.6.15: Results obtained with our implementation for random irregularities using Trapezoidal rule (at level 7), using a weak correlated model.

Weak correlated surfaces, e.g., $k_h = (0.04, 0.04, 0.5, 0.5)$, are displayed in Figures A1.6.16–A1.6.17, while Figures A1.6.18–A1.6.19 show strong-correlated surfaces.

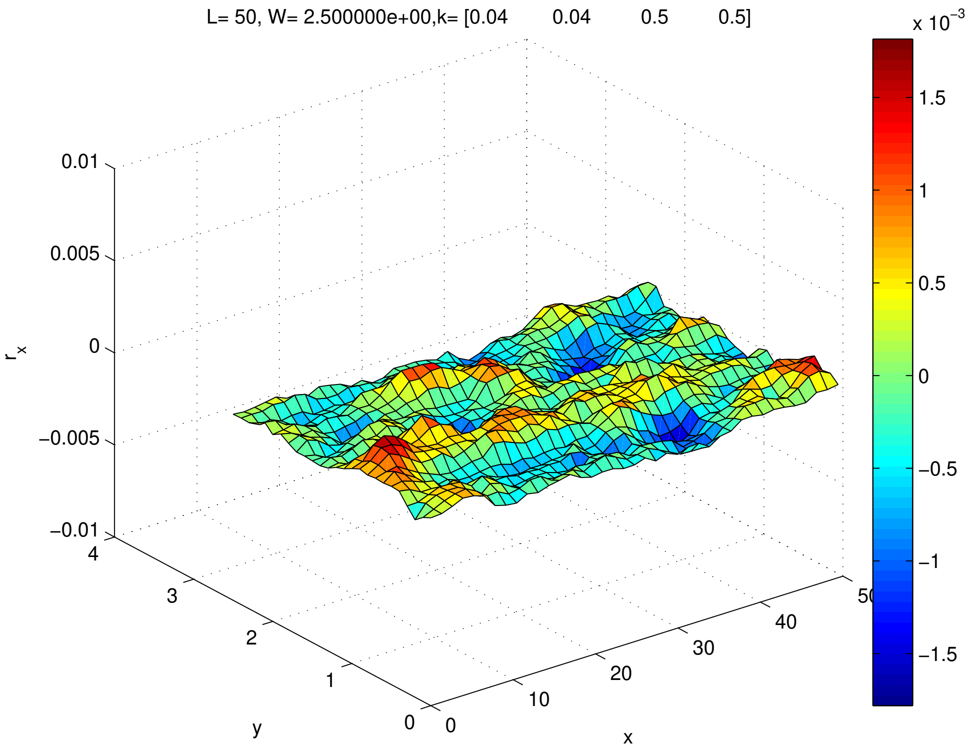


Figure A1.6.16: Random surface belongs to category 1, using the isotropic Ornstein–Uhlenbeck covariance kernel and $k_h = 0.04$ (weak spatial correlated surfaces).

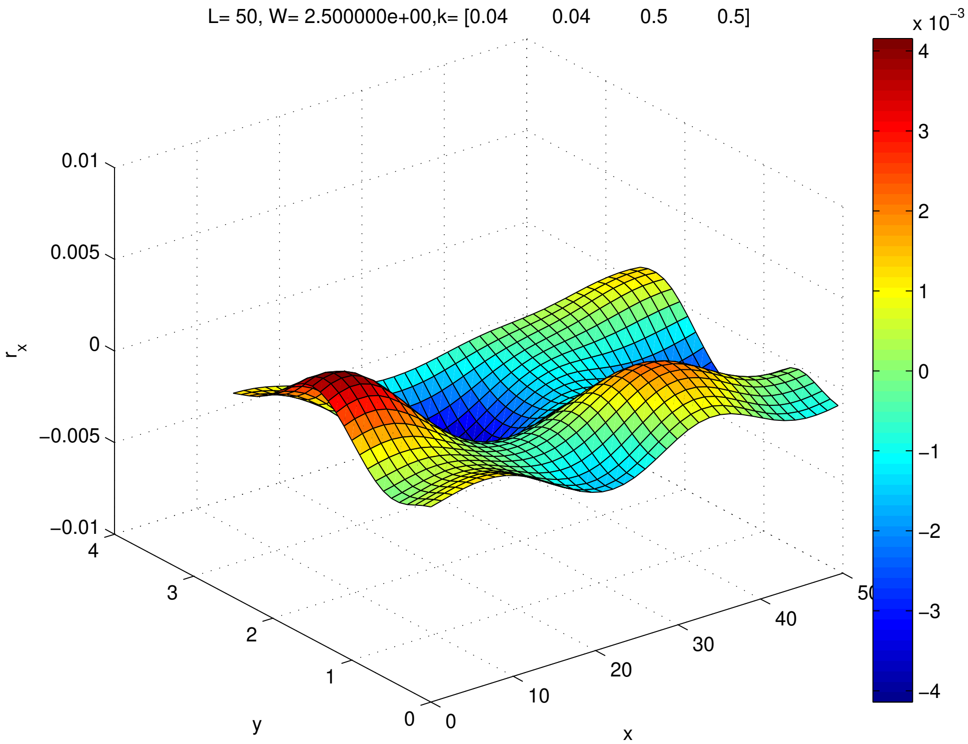


Figure A1.6.17: Random surface belongs to category 2, using the isotropic Gaussian covariance kernel and $k_h = 0.04$ (weak spatial correlated surfaces).

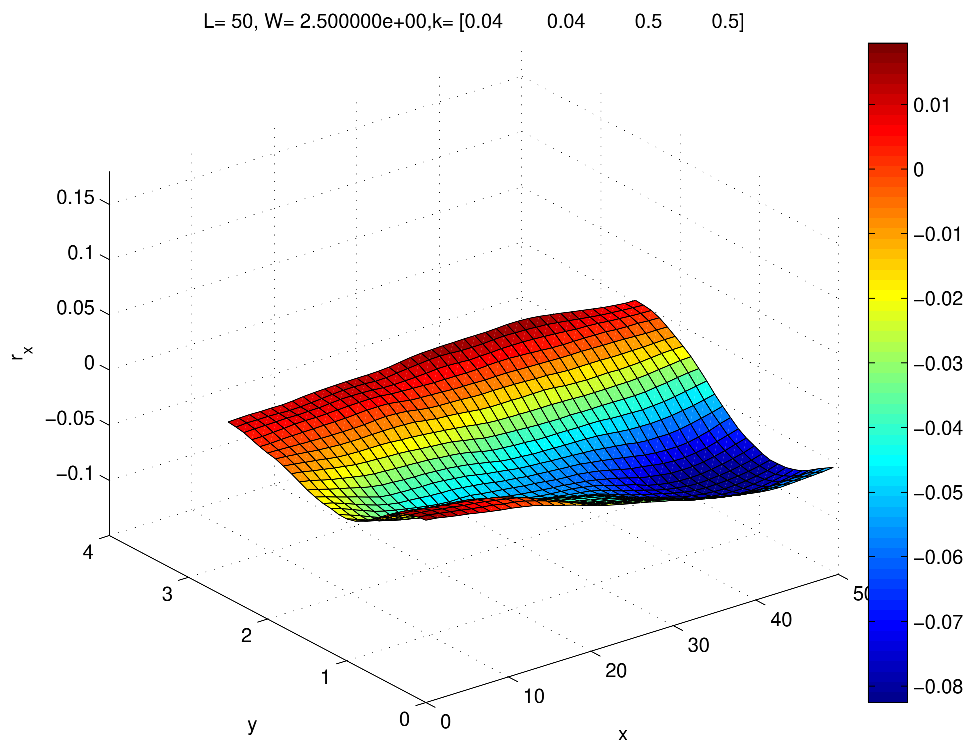


Figure A1.6.18: Random surface belongs to category 3, using the isotropic Ornstein–Uhlenbeck covariance kernel and $k_h = 0.5$ (strong spatial correlated surfaces).

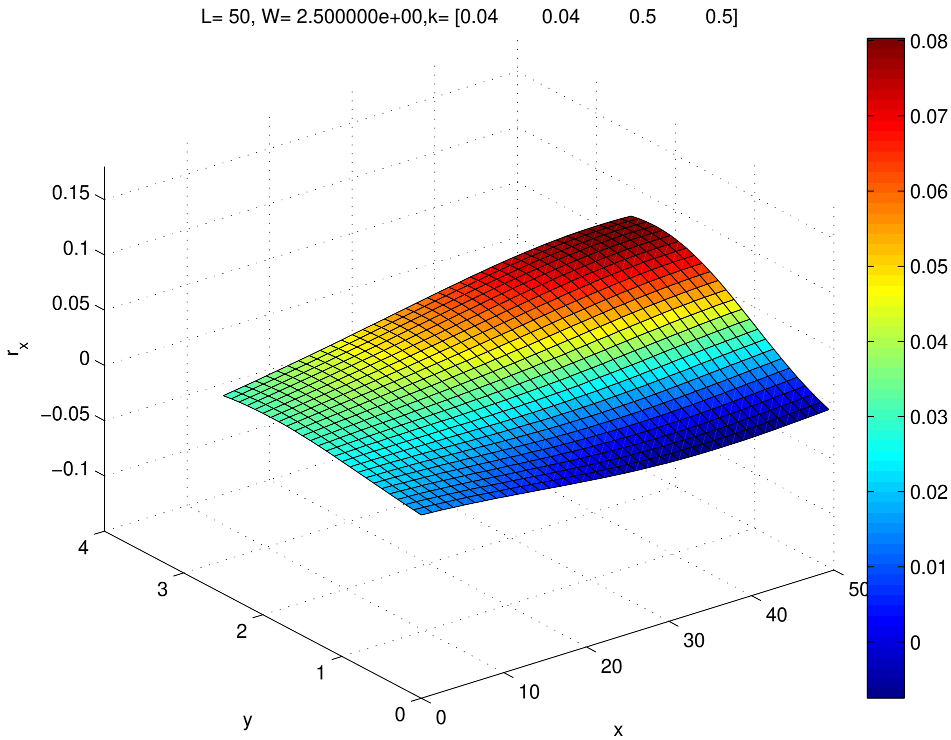


Figure A1.6.19: Random surface belongs to category 4, using the isotropic Gaussian covariance kernel and $k_h = 0.5$ (strong spatial correlated surfaces).

The classification results are displayed in Figure A1.6.20, from the implementation of our proposed functional statistical methodology to discriminate between strong and weak correlated Gaussian surfaces.

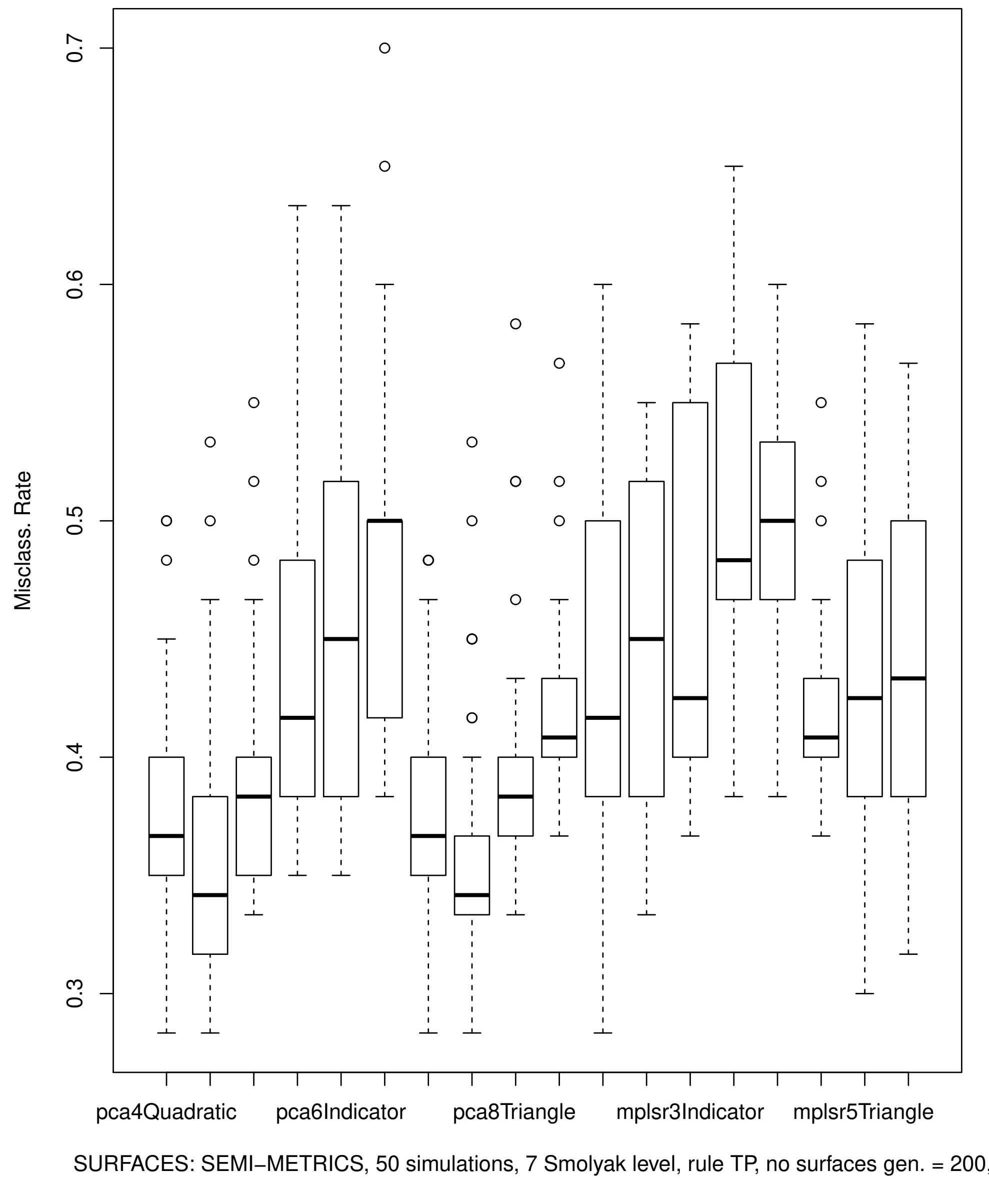


Figure A1.6.20: Results obtained with our implementation for random irregularities using the Trapezoidal rule (at level 7), using a weak spatial correlation model.

Finally, to discriminate between strong spatial correlated surfaces (smoother surfaces), the following values of parameter k_h are considered $k_h = (0.2, 0.2, 0.6, 0.6)$ (see Figures A1.6.21–A1.6.24). The coresponding classification results are showed in Figure A1.6.25.

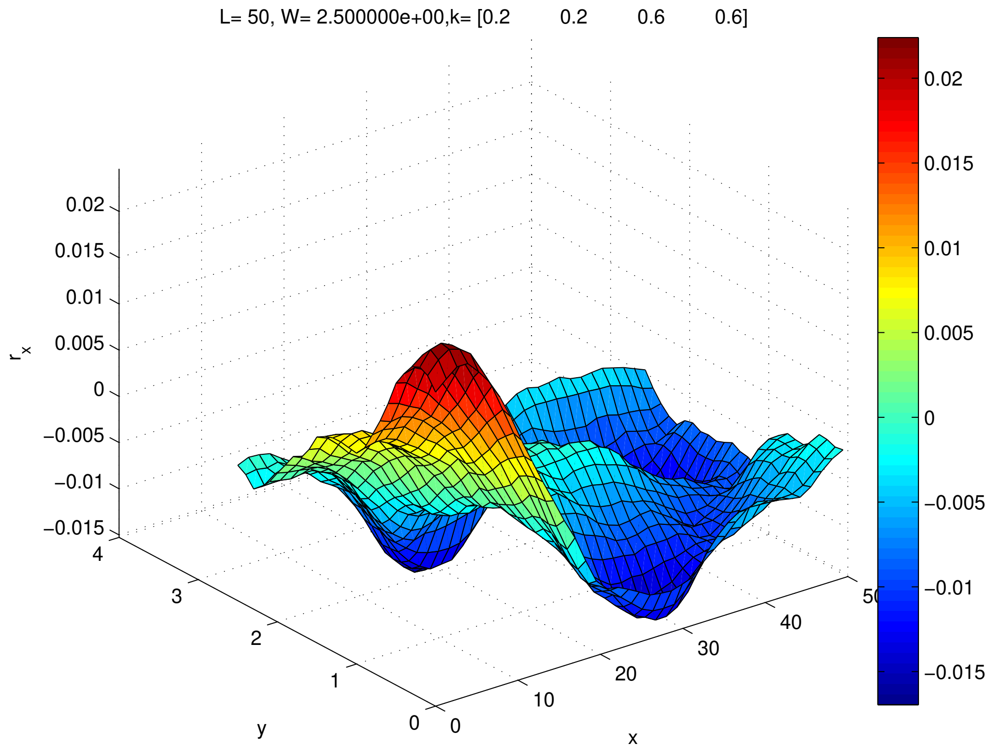


Figure A1.6.21: Random surface belongs to category 1, using the isotropic Ornstein–Uhlenbeck covariance kernel and $k_h = 0.2$ (strong spatial correlated random surface).

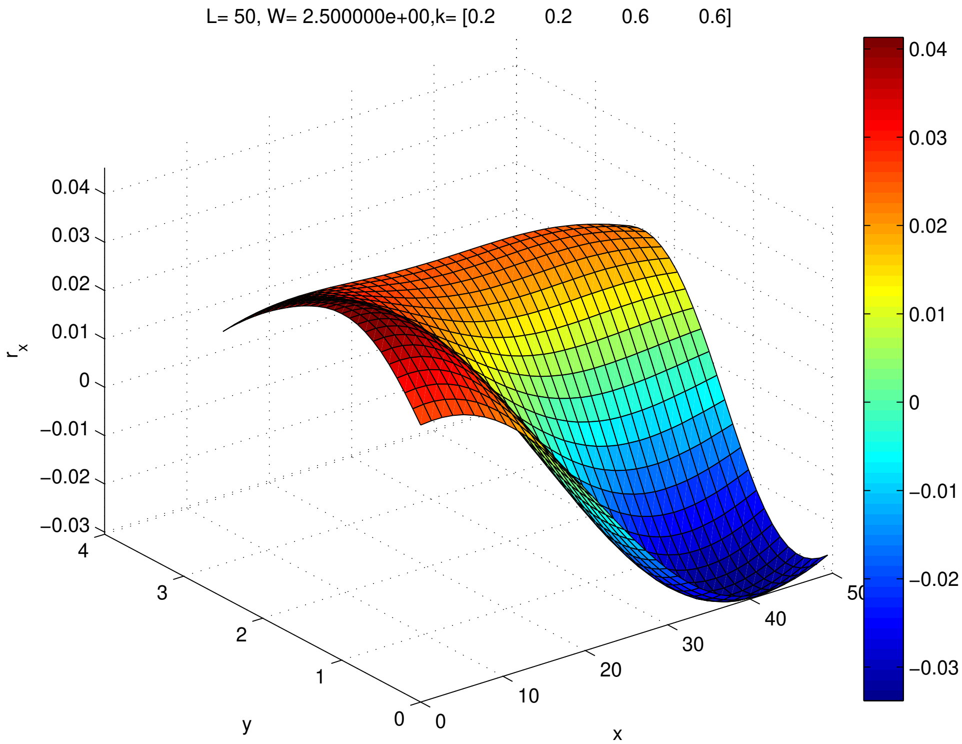


Figure A1.6.22: Random surface belongs to category 2, using the isotropic Gaussian covariance kernel and $k_h = 0.2$ (strong spatial correlated random surface).

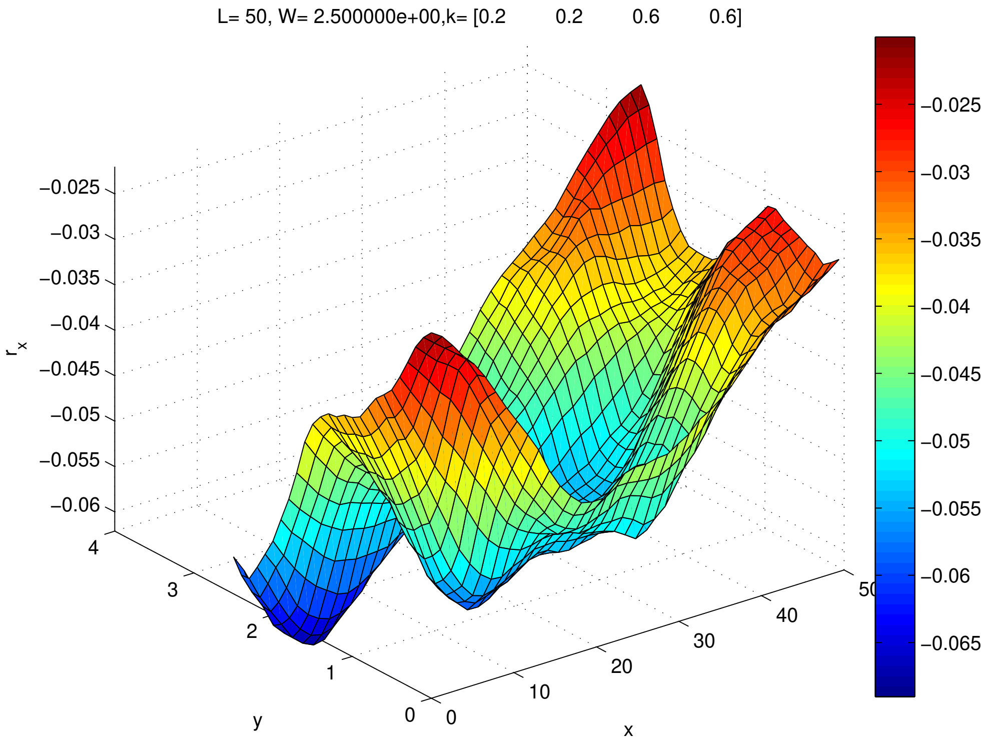


Figure A1.6.23: Random surface belongs to category 3, using the isotropic Ornstein–Uhlenbeck covariance kernel and $k_h = 0.6$ (strong spatial correlated random surface).

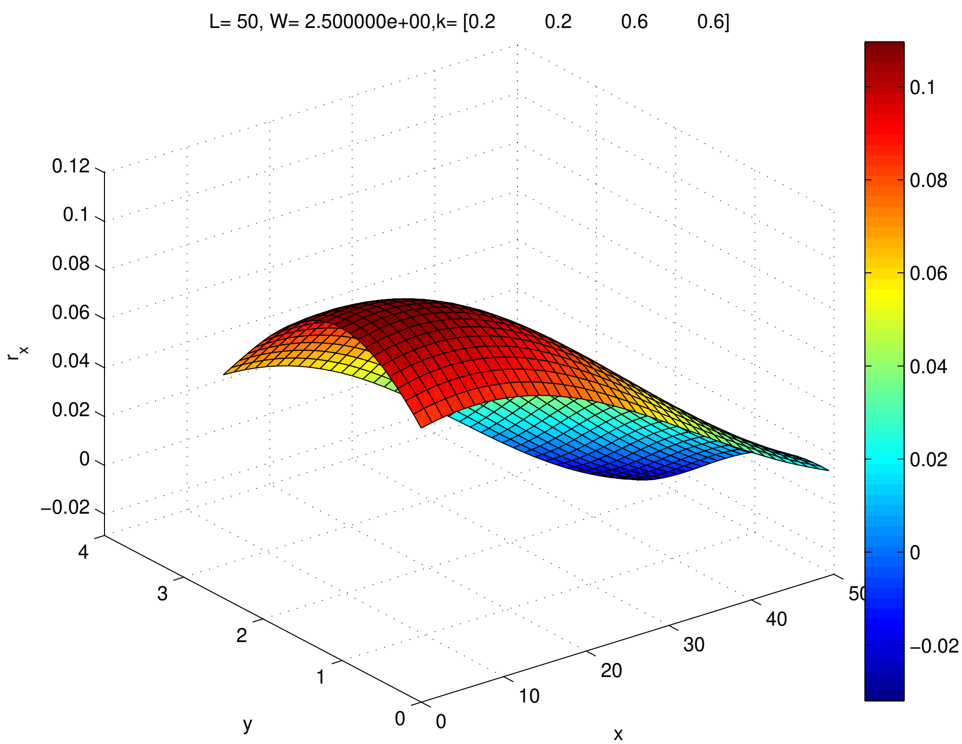


Figure A1.6.24: Random surface belongs to category 4, using the isotropic Gaussian covariance kernel and $k_h = 0.6$ (strong spatial correlated random surface).

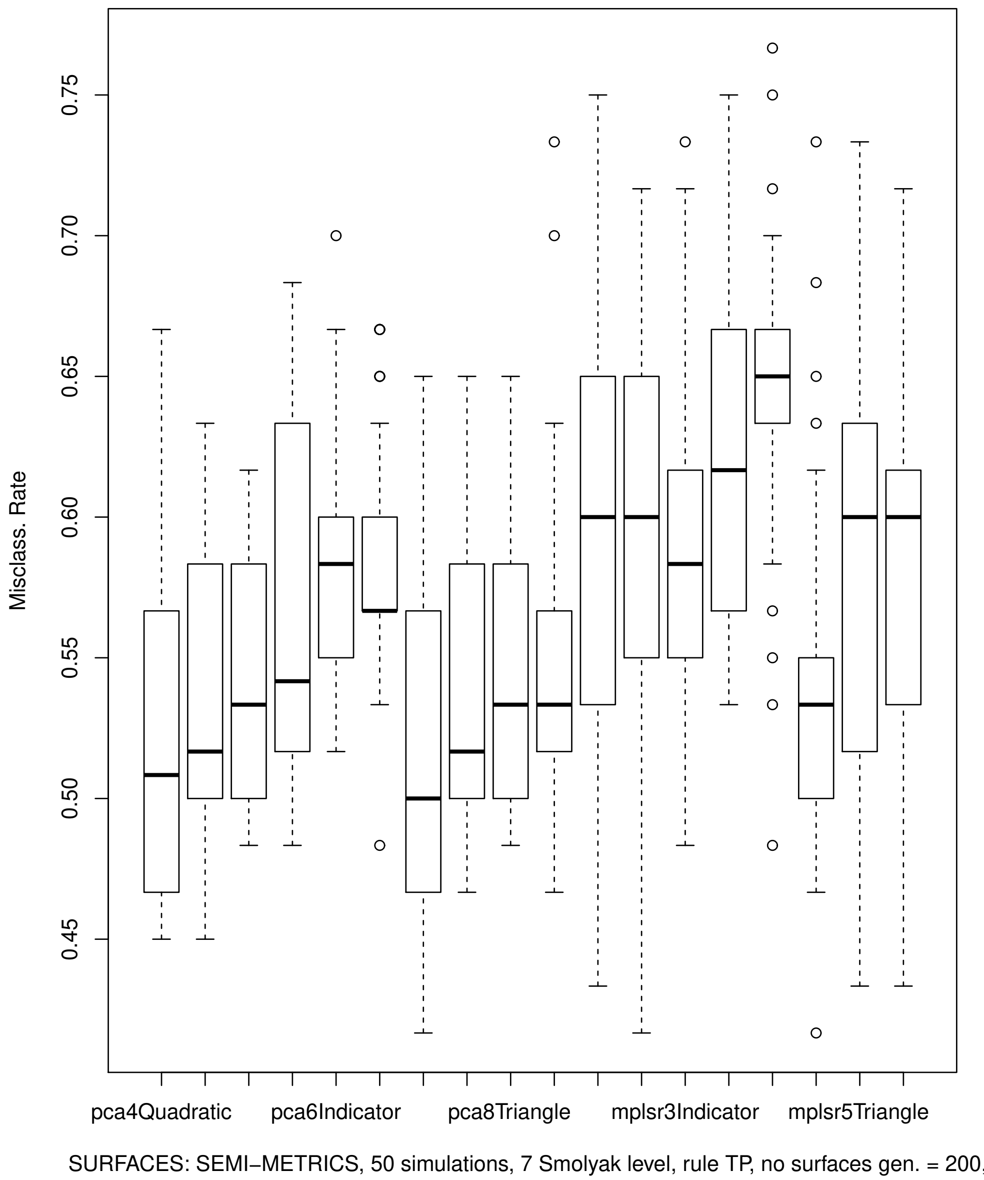


Figure A1.6.25: Results obtained with our implementation for random irregularities using the Trapezoidal rule (at level 7), from strong spatial correlated random surfaces.

We can appreciate a better performance of the functional classification methodology proposed when the non-linear Gaussian random surfaces analysed display weak correlation, inducing a higher degree of local singularity which facilitates the detection of such a more pronounced roughness in the railway track.

A1.7 CONCLUSIONS

In all our implementations, different kernels have been considered, as quadratic, indicator and triangle kernels; and different inputs have been used. We improved the accuracy when we increase the number of evaluations in the Smolyak's quadrature rule in both the Trapezoidal and Clenshaw–Curtis's rule.

In Appendices A1.4–A1.5, we obtain a better performance using the Trapezoidal rule. This is explained by the fact that the Clenshaw–Curtis's quadrature rule is a truncated expanding in the series of trigonometric functions; thus, it looks natural that we obtain less accuracy. This became increasingly evident using FPCA semi–metric (see also Appendix A1.6). With these results, we notice that the choice of univariate quadrature rule is not as trivial as it might seem at first sight. Such as the FPCA semi–metric only depends on the data, its accuracy is more affected by the choice of nodes. Meanwhile, the MPLSR semi–metric also depends on responses that are not affected by the quadrature rule. For that reason, FPLSR semi–metric provides us a better performance than FPCA case. Note also that that the semi-metric based on derivatives is the more accuracy (see Appendix A1.4).

One can observe that with greater interpolation step (see Appendix A1.4) or a finer grid (see Appendices A1.5–A1.6), a slight improvement is obtained due to associated interpolation error and weight allocation error.

In Appendix A1.5, note also that the two categories distinguished in surfaces classification are very close. Also, for well–differentiated categories or groups, FPLSR outperforms FPCA when low–quality data are available or when numerical integration rules are applied at low resolution levels.

The noisy non–random irregularities studied in Appendix A1.6 provide us 12 categories to discriminate, corresponding to 12 irregularity models $\{M_i,\ i = 1, \ldots, 12\}$. Despite having a large number of categories and the closeness between perturbed surfaces, Figure A1.6.9 show us a good performance of our algorithm, such as we obtain a relatively low missclassification rate. In the light of the results shown in Figure A1.6.10, it was concluded that the smaller the distance A between irregularities, the greater the missclassification rate is, since they are more difficult to distinguish to each other.

The surface classification problem addressed in Appendix A1.6 leads us to the following general conclusion: the best performance of our proposed functional classification methodology is obtained when deterministic surfaces perturbed by additive Gaussian white noise are considered (non–linear model with random perturbation). While, in the Gaussian random surface case considered, a better performance is achieved when weak spatial correlated surfaces must be discriminated against strong spatial correlated surfaces. On the other hand, the worst performance is observed when we have to discriminate between smoother random surfaces, corresponding to strong spatial correlated zero–mean Gaussian surfaces.

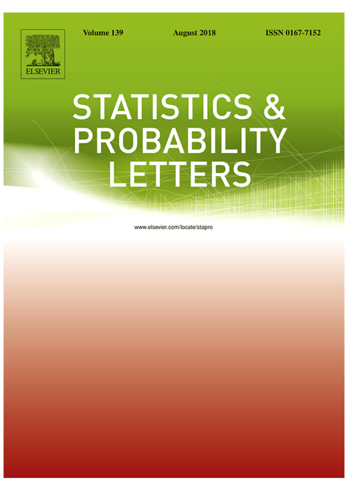
Summarizing, the real–data based and the numerical results showed allow us to confirm that our proposed implementation for general n–dimensional supported non–linear random and deterministic functions can offer an extended version of the previous implementation by Ferraty and Vieu [2006], in a more flexible way, allowing classification of n–dimensional supported deterministic and random surfaces in particular.

A2

CONSISTENCY OF THE PLUG–IN FUNCTIONAL PREDICTOR OF THE ORNSTEIN–UHLENBECK PROCESS IN HILBERT AND BANACH SPACES

ÁLVAREZ-LIÉBANA, J.; BOSQ, D.; RUIZ-MEDINA, M. D.: Consistency of the plug–in functional predictor of the Ornstein–Uhlenbeck process in Hilbert and Banach spaces. Statist. Probab. Letters 117 (2016), pp. 12–22.

DOI: `doi.org/10.1016/j.spl.2016.04.023`



Year	Categ.	Cites	Impact Factor (5 years)	Quartil
2016	Statist. & Probab.	3769	0.636	Q4

ABSTRACT

New results on functional prediction of the Ornstein–Uhlenbeck process, in autoregressive Hilbert-valued and Banach–valued frameworks, are derived. Specifically, consistency of the maximum likelihood estimator of the autocorrelation operator, and of its associated plug–in predictors, is obtained in both frameworks.

A2.1 INTRODUCTION

This paper derives new results in the context of linear processes in function spaces. An extensive literature has been developed in this context in the last few decades (see, for example, Bosq [2000]; Ferraty and Vieu [2006]; Ramsay and Silverman [2005]; among others). In particular, the problem of functional prediction of linear processes in Hilbert and Banach spaces has been widely addressed. We refer to the reader to the papers by Bensmain and Mourid [2001], Bosq [1996, 2002, 2004, 2007], Guillas [2000, 2001], Mas [2002, 2004, 2007], Mas and Menneteau [2003a]; Menneteau [2005], Labbas and Mourid [2002]; Mokhtari and Mourid [2003]; Mourid [2002, 2004] Rachedi [2004, 2005]; Rachedi and Mourid [2003], Dedecker and Merlevède [2003]; Dehling and Sharipov [2005]; Glendinning and Fleet [2007]; Kargin and Onatski [2008]; Ruiz-Medina [2012], Marion and Pumo [2004]; Pumo [1998] and Turbillon et al. [2008, 2007]; and the references therein. In the above–mentioned papers, different projection methodologies have been adopted in the derivation of the main asymptotic properties of the formulated functional parameter estimators and predictors. Particularly, Bosq [2000]; Bosq and Blanke [2007] apply Functional Principal Component Analysis (FPCA); Antoniadis et al. [2006]; Antoniadis and Sapatinas [2003]; Laukaitis and Vasilecas [2009] propose wavelet–bases–based estimation methods. Applications of these functional estimation results can be found in the papers by Antoniadis and Sapatinas [2003]; Damon and Guillas [2002]; Hörmann and Kokoszka [2011]; Laukaitis [2008]; Ruiz-Medina and Salmerón [2009]; among others.

We here pay attention to the problem of functional prediction of the Ornstein–Uhlenbeck (O.U.) process (see, for example, Uhlenbeck and Ornstein [1930]; Wang and Uhlenbeck [1945], for its introduction and properties). See also Doob [1942] for the classical definition of O.U. process from the Langevin (linear) stochastic differential equation. We can find in Kutoyants [2004]; Liptser and Shiraev [2001] an explicit expression of the maximum likelihood estimator (MLE) of the scale parameter θ, characterizing its covariance function. Its strong consistency is proved, for instance, in Kleptsyna and Breton [2002]. We formulate here the O.U. process as an autoregressive Hilbertian process of order one (so–called ARH(1) process), and as an autoregressive Banach–valued process of order one (so–called ARB(1) process). Consistency of the MLE of θ is applied to prove the consistency of the corresponding MLE of the autocorrelation operator of the O.U. process. We adopt the methodology applied in Bosq [1991], since our interest relies on forecasting the values of the O.U. process over an entire time interval. Specifically, considering the O.U. process $\{\xi_t,\ t \in \mathbb{R}\}$ on the basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$, we can define

$$X_n(t) = \xi_{nh+t}, \quad 0 \leq t \leq h, \quad n \in \mathbb{Z}, \tag{A2.1}$$

satisfying

$$X_n(t) = \xi_{nh+t} = \int_{-\infty}^{nh+t} e^{-\theta(nh+t-s)} dW_s = \rho_\theta (X_{n-1})(t) + \varepsilon_n(t), \quad n \in \mathbb{Z}, \tag{A2.2}$$

with

$$\begin{aligned} \rho_\theta(x)(t) &= e^{-\theta t} x(h), \quad \rho_\theta(X_{n-1})(t) = e^{-\theta t}\int_{-\infty}^{nh} e^{-\theta(nh-s)} dW_s, \\ \varepsilon_n(t) &= \int_{nh}^{nh+t} e^{-\theta(nh+t-s)} dW_s, \end{aligned} \tag{A2.3}$$

for $0 \leq t \leq h$, where $W = \{W_t,\ t \in \mathbb{R}\}$ is a standard bilateral Wiener process (see Supplementary Material A2.5). Thus, $X = \{X_n,\ n \in \mathbb{Z}\}$ satisfies the ARH(1) equation (A2.2) (see also equation (A2.4) below for its general definition). The real separable Hilbert space H is given by $H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$, where $\beta_{[0,h]}$ is the Borel σ-algebra generated by the subintervals in $[0,h]$, λ is the Lebesgue measure and $\delta_{(h)}(s) = \delta(s-h)$ is the Dirac measure at point h. The associated norm

$$\|f\|_H = \sqrt{\int_0^h (f(t))^2 dt + (f(h))^2}, \quad f \in H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right),$$

establishes the equivalent classes of functions given by the relationship $f \sim_{\lambda+\delta_{(h)}} g$ if and only if

$$\left(\lambda + \delta_{(h)}\right)(\{t : f(t) \neq g(t)\}) = 0,$$

with

$$\left(\lambda + \delta_{(h)}\right)(\{t : f(t) \neq g(t)\}) = 0 \Leftrightarrow \lambda(\{t : f(t) \neq g(t)\}) = 0 \text{ and } f(h) = g(h),$$

where, as before, $\delta_{(h)}$ is the Dirac measure at point h. We will prove, in Lemma A2.2.1 below, that $X = \{X_n,\ n \in \mathbb{Z}\}$, constructed in (A2.1) from the O.U. process, satisfying equations (A2.2)–(A2.3), is the unique stationary solution to equation (A2.2), in the space $H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$, admitting a MAH$(\infty)$ representation. Similarly, in Lemma A2.2.4 below, we will prove that $X = \{X_n,\ n \in \mathbb{Z}\}$, constructed in (A2.1) from the O.U. process, satisfying equations (A2.2)–(A2.3), is the unique stationary solution to equation (A2.2), admitting a MAB(∞) representation, in the space $B = \mathcal{C}([0,h])$, the real separable Banach space of continuous functions, whose support is the interval $[0,h]$, with the supremum norm.

The main results of this paper provide the almost surely convergence to ρ_θ of its MLE $\rho_{\widehat{\theta}}$, in the norm of $\mathcal{L}(H)$, the space of bounded linear operators in the Hilbert space H (respectively, in the norm of $\mathcal{L}(B)$, the space of bounded linear operators in the Banach space B). The convergence in probability of the associated plug–in ARH(1) and ARB(1) predictors (i.e., the convergence in probability of $\rho_{\widehat{\theta}}(X_{n-1})$ to $\rho_{\theta}(X_{n-1})$ in H and B, respectively) is proved as well.

The outline of this paper is as follows. In Appendix A2.2, the main results of this paper are obtained. Specifically, Appendix A2.2.1 provides the definition of an O.U. process as an ARH(1) process. Strong consistency in $\mathcal{L}(H)$ of the estimator of the autocorrelation operator is derived in Appendix A2.2.2. Consistency in H of the associated plug–in ARH(1) predictor is then established in Appendix A2.2.3. The corresponding results in Banach spaces are given in Appendix A2.2.4. For illustration purposes, a simulation study is undertaken in Appendix A2.3. Final comments can be found in Appendix A2.4. The basic preliminary elements, applied in the proof of the main results of this paper, and the proof of Lemma A2.2.1, can be found in the Supplementary Material A2.5.

A2.2 PREDICTION OF O.U. PROCESSES IN HILBERT AND BANACH SPACES

In this section, we consider H to be a real separable Hilbert space. Recall that a zero–mean ARH(1) process $X = \{X_n,\ n \in \mathbb{Z}\}$, on the basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$, satisfies (see Bosq [2000])

$$X_n(t) = \rho\left(X_{n-1}\right)(t) + \varepsilon_n(t), \quad n \in \mathbb{Z}, \quad \rho \in \mathcal{L}(H), \tag{A2.4}$$

where ρ denotes the autocorrelation operator of process X. Here, $\varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}$ is assumed to be a strong–white noise; i.e., ε is a Hilbert–valued zero-mean stationary process, with independent and identically distributed components in time, with $\sigma^2 = \mathrm{E}\left\{\|\varepsilon_n\|_H^2\right\} < \infty$, for all $n \in \mathbb{Z}$.

A2.2.1 O.U. processes as ARH(1) processes

As commented in Appendix A2.1, equations (A2.1)–(A2.3) provide the definition of an O.U. process as an ARH(1) process, with $H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$. The norm in the space H of $\rho_\theta(x)$, with ρ_θ introduced in (A2.3) and $x \in H$, is given by

$$\|\rho_\theta(x)\|_H^2 = \int_0^h \left(\rho_\theta\left(x\right)(t)\right)^2 d\left(\lambda + \delta_{(h)}\right)(t) = \int_0^h \left(\rho_\theta\left(x\right)(t)\right)^2 dt + \left(\rho_\theta\left(x\right)(h)\right)^2,$$

for each $h > 0$. The following lemma provides, for each $k \geq 1$, the exact value of the norm of ρ_θ^k, in the space of bounded linear operators on H. As a direct consequence, the existence of an integer k_0 such that $\|\rho_\theta^k\|_{\mathcal{L}(H)} < 1$, for $k \geq k_0$, is also derived for $\theta > 0$.

Lemma A2.2.1 *Let us consider $\theta > 0$ and $X = \{X_n,\ n \in \mathbb{Z}\}$ satisfying equations (A2.1)–(A2.3). For each $k \geq 1$, the uniform norm of ρ_θ^k is given by*

$$\|\rho_\theta^k\|_{\mathcal{L}(H)} = \sqrt{e^{-2\theta(k-1)h}\left(\frac{1 + e^{-2\theta h}\left(2\theta - 1\right)}{2\theta}\right)} = e^{-\theta(k-1)h}\|\rho_\theta\|_{\mathcal{L}(H)}. \tag{A2.5}$$

Furthermore, for $k \geq k_0 = \left[\frac{1}{\theta} + 1\right]^+$,
$$\|\rho_\theta^k\|_{\mathcal{L}(H)} < 1, \tag{A2.6}$$
where $[t]^+$ *denotes the closest upper integer of* t, *for every* $t \in \mathbb{R}_+$.

The proof of this lemma can be found in the Supplementary Material A2.5 provided.

Remark A2.2.1 *From equation (A2.6), applying [Bosq, 2000, Theorem 3.1], Lemma A2.2.1 implies that* X *constructed in (A2.1) from an O.U. process, defines the unique stationary solution to equation (A2.2) in the space* $H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$, *admitting the MAH(∞) representation*

$$X_n = \sum_{k=0}^{+\infty} \rho_\theta^k (\varepsilon_{n-k}), \quad n \in \mathbb{Z}, \quad \rho_\theta \in \mathcal{L}(H).$$

Remark A2.2.2 *Note that, for all* $x \in H$, *and* $k \geq 2$, $\|\rho_\theta^k\|_{\mathcal{L}(H)} \leq \|\rho_\theta\|_{\mathcal{L}(H)}^k$.

A2.2.2 Functional parameter estimation and consistency

We now prove the strong consistency of the estimator $\rho_{\widehat{\theta}_n}$ of operator ρ_θ in $\mathcal{L}(H)$, with, as before, $H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$, and $\widehat{\theta}_n$ denoting the MLE of θ, based on the observation of an O.U. process on the interval $[0,T]$, with $T = nh$. Note that, from equation (A2.3), for all $x \in H$, and for a given sample size n,
$$\rho_{\widehat{\theta}_n}(x) = e^{-\widehat{\theta}_n t} x(h),$$
where the MLE of θ is given, for $T = nh$, by

$$\widehat{\theta}_T = \frac{1 + \frac{\xi_0^2}{T} - \frac{\xi_T^2}{T}}{\frac{2}{T}\int_0^T \xi_t^2 dt}, \quad T > 0, \tag{A2.7}$$

with $\{\xi_t,\ t \in [0,T]\}$ being the observed values of the O.U. process over the interval $[0,T]$. Thus, $\rho_{\widehat{\theta}_n}$ is introduced in an abstract way, since it can only be explicitly computed, for each particular function $x \in H$ considered. However, the norm $\|\rho_\theta - \rho_{\widehat{\theta}_n}\|_{\mathcal{L}(H)}$ is explicitly computed in equation (A2.8) below.

The following results will be applied in the proof of Proposition A2.2.1.

Lemma A2.2.2 *If* $t \in [0, +\infty)$, *it holds that*

$$|e^{-ut} - e^{-vt}| \leq |u - v| t, \quad u, v \geq 0.$$

The proof of this lemma is given in the Supplementary Material A2.5.

Theorem A2.2.1 *(See also [Kleptsyna and Breton, 2002, Proposition 2.2] and [Kutoyants, 2004, p. 63 and p. 117]). The MLE of θ defined in equation (A2.7) is strongly consistent; i.e.,*

$$\widehat{\theta}_T \longrightarrow \theta \quad a.s., \quad T \to \infty.$$

The proof follows from the Ibragimov–Khasminskii's Theorem.

Proposition A2.2.1 *Let H be the space $L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$. Then, the estimator $\rho_{\widehat{\theta}_n}$ of operator ρ_θ, based on the MLE $\widehat{\theta}_n$ of θ, is strongly consistent in the norm of $\mathcal{L}(H)$; i.e.,*

$$\|\rho_\theta - \rho_{\widehat{\theta}_n}\|_{\mathcal{L}(H)} \longrightarrow 0 \quad a.s., \quad n \to \infty.$$

Proof. The following straightforward almost surely identities are obtained:

$$\begin{aligned}
\|\rho_\theta - \rho_{\widehat{\theta}_n}\|_{\mathcal{L}(H)} &= \sup_{x \in H} \left\{ \frac{\|\left(\rho_\theta - \rho_{\widehat{\theta}_n}\right)(x)\|_H}{\|x\|_H} \right\} \\
&= \sup_{x \in H} \left\{ \sqrt{\frac{\int_0^h \left(\left(\rho_\theta - \rho_{\widehat{\theta}_n}\right)(x)(t)\right)^2 d\left(\lambda + \delta_{(h)}\right)(t)}{\int_0^h (x(t))^2 d\left(\lambda + \delta_{(h)}\right)(t)}} \right\} \\
&= \sup_{x \in H} \left\{ \sqrt{(x(h))^2 \frac{\int_0^h \left(e^{-\theta t} - e^{-\widehat{\theta}_n t}\right)^2 dt + \left(e^{-\theta h} - e^{-\widehat{\theta}_n h}\right)^2}{\int_0^h (x(t))^2 dt + (x(h))^2}} \right\} \\
&= \sqrt{\int_0^h \left(e^{-\theta t} - e^{-\widehat{\theta}_n t}\right)^2 dt + \left(e^{-\theta h} - e^{-\widehat{\theta}_n h}\right)^2}, \qquad \text{(A2.8)}
\end{aligned}$$

where the last identity is obtained in a similar way to equation (A2.5) in Lemma A2.2.1 (see Supplementary Material A2.5).

From Lemma A2.2.2 and equation (A2.8), for n sufficiently large, we have

$$\begin{aligned}
\|\rho_\theta - \rho_{\widehat{\theta}_n}\|_{\mathcal{L}(H)} &\leq \sqrt{\int_0^h t^2|\theta - \widehat{\theta}_n|^2 dt + h^2|\theta - \widehat{\theta}_n|^2} = |\theta - \widehat{\theta}_n|\sqrt{\int_0^h t^2 dt + h^2} \\
&= |\theta - \widehat{\theta}_n| h\sqrt{\frac{h}{3} + 1} \quad a.s. \qquad \text{(A2.9)}
\end{aligned}$$

The strong–consistency of $\rho_{\widehat{\theta}_n}$ in $\mathcal{L}(H)$ directly follows from Theorem A2.2.1 and equation (A2.9). ■

Remark A2.2.3 From [Kleptsyna and Breton, 2002, Proposition 2.3] (see also Theorem A2.2.2 below), the MLE $\widehat{\theta}_T$ of θ satisfies

$$\mathrm{E}\left\{\left(\theta-\widehat{\theta}_T\right)^2\right\}=\mathcal{O}\left(\frac{2\theta}{T}\right),\quad T\to\infty. \tag{A2.10}$$

In addition, from equation (A2.9), considering $T=nh$, $h>0$,

$$\mathrm{E}\left\{\|\rho_\theta-\rho_{\widehat{\theta}_n}\|^2_{\mathcal{L}(H)}\right\}\leq \mathrm{E}\left\{|\theta-\widehat{\theta}_n|^2\right\}h^2\left(\frac{h}{3}+1\right). \tag{A2.11}$$

Equations (A2.10)–(A2.11) lead to

$$\mathrm{E}\left\{\|\rho_\theta-\rho_{\widehat{\theta}_n}\|^2_{\mathcal{L}(H)}\right\}\leq G(\theta,\widehat{\theta}_n,h),$$

with

$$G(\theta,\widehat{\theta}_n,h)=\mathcal{O}\left(\frac{2\theta}{n}\right),\quad n\to\infty.$$

Therefore, the functional parameter estimator $\rho_{\widehat{\theta}_n}$ is $\sqrt{n}$–consistent.

A2.2.3 Consistency of the plug–in ARH(1) predictor

Let us consider the plug–in ARH(1) predictor $\widehat{X}_n$, constructed from the MLE $\rho_{\widehat{\theta}_n}$ of ρ_θ in Proposition A2.2.1, given by

$$\widehat{X}_n(t)=\rho_{\widehat{\theta}_n}(X_{n-1})(t)=e^{-\widehat{\theta}_n t}X_{n-1}(h),\quad 0\leq t\leq h,\quad n\in\mathbb{Z}. \tag{A2.12}$$

Corollary A2.2.1 below provides the consistency of $\widehat{X}_n$, given in equation (A2.12), from Proposition A2.2.1 by applying the following lemma and theorem.

Lemma A2.2.3 *Let $\{Z_n,\ n\in\mathbb{Z}\}$ be a sequence of random variables such that*

$$Z_n\sim\mathcal{N}\left(0,\frac{1}{2\theta}\right),\quad \theta>0,$$

and let $\{Y_n,\ n\in\mathbb{Z}\}$ be another sequence of random variables such that

$$\sqrt{\ln(n)}Y_n\longrightarrow^p 0,\quad n\to\infty.$$

Then,

$$Y_n |Z_n| \longrightarrow^p 0, \quad n \to \infty,$$

where, as usual, $\longrightarrow^p$ *indicates convergence in probability.*

The proof of this lemma can be found in the Supplementary Material A2.5.

Theorem A2.2.2 *Let* $\widehat{\theta}_T$ *be the MLE of* θ *defined in equation (A2.7), with* $\theta > 0$. *Hence,*

$$\mathrm{E}\left\{ \left(\theta - \widehat{\theta}_T\right)^2 \right\} = \mathcal{O}\left(\frac{2\theta}{T}\right), \quad T \to \infty. \tag{A2.13}$$

In particular,

$$\lim_{T \to \infty} \mathrm{E}\left\{ \left(\theta - \widehat{\theta}_T\right)^2 \right\} = 0.$$

The proof of this result is given in [Kleptsyna and Breton, 2002, Proposition 2.3].

Corollary A2.2.1 *Let* $H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$ *be the Hilbert space introduced above. Then, the plug–in ARH(1) predictor (A2.12) of an O.U. process is consistent in* H*; i.e.,*

$$\left\| \left(\rho_\theta - \rho_{\widehat{\theta}_n}\right)(X_{n-1}) \right\|_H \longrightarrow^p 0.$$

Proof. By definition,

$$\left\| \left(\rho_\theta - \rho_{\widehat{\theta}_n}\right)(X_{n-1}) \right\|_H = |X_{n-1}(h)| \sqrt{\int_0^h \left(e^{-\theta t} - e^{-\widehat{\theta}_n t}\right)^2 dt + \left(e^{-\theta h} - e^{-\widehat{\theta}_n h}\right)^2}. \tag{A2.14}$$

From equations (A2.8)–(A2.9) and (A2.14), we then obtain, for n sufficiently large,

$$\left\| \left(\rho_\theta - \rho_{\widehat{\theta}_n}\right)(X_{n-1}) \right\|_H \leq |X_{n-1}(h)| \left|\theta - \widehat{\theta}_n\right| h \sqrt{\frac{h}{3} + 1} \quad a.s. \tag{A2.15}$$

Let us set

$$\{Y_n,\, n \in \mathbb{Z}\} = \left\{ |\theta - \widehat{\theta}_n| h \sqrt{\frac{h}{3} + 1},\, n \in \mathbb{Z} \right\}, \quad \{Z_n,\, n \in \mathbb{Z}\} = \{X_{n-1}(h),\, n \in \mathbb{Z}\},$$

with $Z_n \sim \mathcal{N}\left(0, \frac{1}{2\theta}\right)$, for every $n \in \mathbb{Z}$. From Theorem A2.2.1,

$$Y_n \longrightarrow 0 \quad a.s., \quad n \to \infty.$$

Hence, to apply Lemma A2.2.3, we need to prove that

$$\sqrt{\ln(n)}Y_n \longrightarrow^p 0, \quad n \to \infty.$$

From the Chebyshev's inequality and Theorem A2.2.2, we get, for all $\varepsilon > 0$,

$$\lim_{n\to 0} \mathcal{P}\left(|\theta - \widehat{\theta}_n|\sqrt{\ln(n)}h\sqrt{\frac{h}{3}+1} \geq \varepsilon\right) \leq \frac{h^2\left(\frac{h}{3}+1\right)\ln(n)\,\mathrm{E}\left\{\left|\theta-\widehat{\theta}_n\right|^2\right\}}{\varepsilon^2} = 0.$$

Therefore, from Lemma A2.2.3, we obtain the convergence in probability of $\left\|\left(\rho_\theta - \rho_{\widehat{\theta}_n}\right)(X_{n-1})\right\|_H$ to zero.

■

A2.2.4 Prediction of O.U. processes in $B = \mathcal{C}([0,h])$

As before, let B be now the Banach space of continuous functions, whose support is the interval $[0,h]$, with the supremum norm, denoted as $\mathcal{C}([0,h])$. The following lemma states that $\|\rho_\theta^k\|_{\mathcal{L}(B)} \leq 1$, for $\theta > 0$, and for every $k \geq 1$, with $\mathcal{L}(B)$ being the space of bounded linear operators on the Banach space $B = \mathcal{C}([0,h])$, and ρ_θ being introduced in equation (A2.3). Consequently, from [Bosq, 2000, Theorem 6.1], $X = \{X_n,\ n \in \mathbb{Z}\}$, constructed in (A2.1) from the O.U. process, defines the unique stationary solution to equation (A2.2), in the Banach space $B = \mathcal{C}([0,h])$, admitting a MAB(∞) representation.

Lemma A2.2.4 *Let ρ_θ introduced in (A2.3), defined on $B = \mathcal{C}([0,h])$. Then, for $k \geq 1$, $\|\rho_\theta^k\|_{\mathcal{L}(B)} \leq 1$, with $\theta > 0$.*

Proof.
From

$$\rho_\theta^k(x)(t) = e^{-\theta t}e^{-\theta(k-1)h}x(h),$$

for each $k \geq 1$ and $\theta > 0$, we have

$$\begin{aligned}
\|\rho_\theta^k\|_{\mathcal{L}(B)} &= \sup_{x\in B}\left\{\frac{\|\rho_\theta^k(x)\|_B}{\|x\|_B}\right\} = \sup_{x\in B}\left\{\frac{\sup_{0\leq t\leq h}\left\{\left|e^{-\theta t}e^{-\theta(k-1)h}x(h)\right|\right\}}{\sup_{0\leq t\leq h}|x(t)|}\right\} \\
&= \sup_{x\in B}\left\{\frac{|x(h)|\,e^{-\theta(k-1)h}\sup_{0\leq t\leq h}e^{-\theta t}}{\sup_{0\leq t\leq h}|x(t)|}\right\} \leq \sup_{x\in B}\left\{\frac{|x(h)|\sup_{0\leq t\leq h}e^{-\theta t}}{|x(h)|}\right\} \\
&= \sup_{0\leq t\leq h} e^{-\theta t} = 1.
\end{aligned} \tag{A2.16}$$

■

We now check the strong consistency of the MLE $\rho_{\widehat{\theta}_n}$ of ρ_θ in $\mathcal{L}(B)$. From equation (A2.16),

$$\|\rho_\theta - \rho_{\widehat{\theta}_n}\|_{\mathcal{L}(B)} \leq \sup_{0\leq t\leq h} \left\{ \left| e^{-\theta t} - e^{-\widehat{\theta}_n t} \right| \right\} \quad a.s.$$

From Lemma A2.2.2, for n sufficiently large, we then have

$$\|\rho_\theta - \rho_{\widehat{\theta}_n}\|_{\mathcal{L}(B)} \leq h \left| \theta - \widehat{\theta}_n \right| \quad a.s. \tag{A2.17}$$

Theorem A2.2.1 then leads to the desired result on strong consistency of the estimator $\rho_{\widehat{\theta}_n}$ of ρ_θ in $\mathcal{L}(B)$. Furthermore, from Theorem A2.2.2 , in a similar way to Remark A2.2.3, the $\sqrt{n}$–consistency of $\rho_{\widehat{\theta}_n}$ in $\mathcal{L}\left(B\right)$ also follows from equations (A2.13) and (A2.17).

Similarly to Corollary A2.2.1, in the following result, the consistency, in the Banach space $B = C([0,h])$, of the plug–in predictor (A2.12) is obtained.

Corollary A2.2.2 *The ARB(1) plug–in predictor (A2.12) of a zero–mean O.U. process is consistent in* $B = C([0,h])$*; i.e., as* $n \to \infty$,

$$\left\| \left(\rho_\theta - \rho_{\widehat{\theta}_n}\right) (X_{n-1}) \right\|_B \longrightarrow^p 0.$$

Proof. From Lemma A2.2.2, for n sufficiently large, and for each $h>0$,

$$\| \left(\rho_\theta - \rho_{\widehat{\theta}_n}\right) (X_{n-1}) \|_B = \sup_{0\leq t\leq h} \left\{ \left| e^{-\theta t} - e^{-\widehat{\theta}_n t} \right| |X_{n-1}(h)| \right\} \leq h|\theta - \widehat{\theta}_n||X_{n-1}(h)| \quad a.s. \tag{A2.18}$$

As derived in the proof of Corollary A2.2.1, from Theorem A2.2.2, the random sequence $\{Y_n,\, n\in\mathbb{Z}\} = \left\{ h|\theta - \widehat{\theta}_n|,\, n\in\mathbb{Z} \right\}$ is such that

$$\sqrt{\ln(n)}Y_n \leq \sqrt{\frac{h}{3}+1}\sqrt{\ln(n)}Y_n \longrightarrow^p 0, \quad n\to\infty.$$

Moreover, $\{Z_n,\, n\in\mathbb{Z}\} = \{X_{n-1}(h),\, n\in\mathbb{Z}\}$ is such that $Z_n \sim \mathcal{N}\left(0, \frac{1}{2\theta}\right)$. Lemma A2.2.3 then leads, as $n\to\infty$, to the desired convergence result from equation (A2.18):

$$\| \left(\rho_\theta - \rho_{\widehat{\theta}_n}\right) (X_{n-1}) \|_B \leq Y_n|Z_n| \longrightarrow^p 0.$$

A2.3 SIMULATIONS

In this section, a simulation study is undertaken to illustrate the asymptotic results presented in this paper about the MLE $\widehat{\theta}_n$ of θ, and the consistency of the ML functional parameter estimators of the autocorrelation operator, and the associated plug–in predictors, in the ARH(1) and ARB(1) frameworks.

A2.3.1 Estimation of the scale parameter θ

On the simulation of the sample–paths of an O.U. process, an extension of the Euler's method, the so–called Euler–Murayama's method (see Kloeden and Platen [1992]) is applied, from the Langevin stochastic differential equation satisfied by the O.U. process $\{\xi_t,\ t \in [0,T]\}$

$$d\xi_t = -\theta\xi_t + dW_t, \quad \theta > 0, \quad t \in [0,T], \quad \xi_0 = x_0. \tag{A2.19}$$

Thus, let $0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of the real interval $[0,T]$. Then, (A2.19) can be discretized as

$$\widehat{\xi}_{i+1} = \widehat{\xi}_i - \theta\widehat{\xi}_i + \Delta W_i, \quad \widehat{\xi}_0 = \xi_0 = 0, \tag{A2.20}$$

where $\{\Delta W_i,\ i = 0,\dots,n-1\}$ are i.i.d. Wiener increments; i.e.,

$$\Delta W_i \sim \mathcal{N}(0,\Delta t) = \sqrt{\Delta t}\mathcal{N}(0,1), \quad i = 0,\dots,n-1.$$

In the following, we take $\Delta t = 0.02$ as discretization step size, considering $N = 1000$ simulations of the O.U. process. In particular, Figure A2.3.1 shows some realizations of the discrete version of the solution to (A2.19) generated from (A2.20).

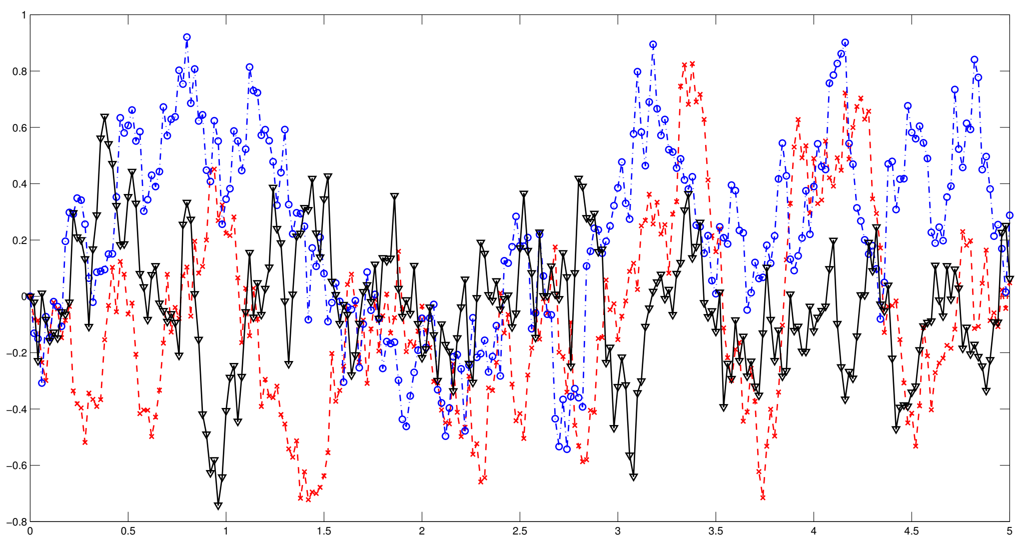


Figure A2.3.1: Sample paths of an O.U. process $\{\xi_t,\ 0 \le t \le T\}$ generated with $T = 5$, $\Delta t = 0.02$, $\theta = 5$ and $\widehat{\xi}_0 = 0$.

Let us first illustrate the asymptotic normal distribution of $\widehat{\theta}_T$; i.e., for T sufficiently large, we can consider $\widehat{\theta}_T \sim \mathcal{N}\left(\theta, \frac{2\theta}{T}\right)$ (see Theorem A2.5.1 in the Supplementary Material A2.5). From equation (A2.7), we take

$$\widehat{\theta}_T = \frac{-\displaystyle\int_0^T \xi_t d\xi_t}{\displaystyle\int_0^T \xi_t^2 dt},$$

(see also Supplementary material A2.5), to compute the following approximation of the MLE $\widehat{\theta}_T$ of θ, for each one of the $N = 1000$ simulations performed, and for each one of the six values of parameter θ considered:

$$\widehat{\theta}_T \simeq \frac{-\displaystyle\sum_{i=0}^{n-1} \widehat{\xi}_{t_i,s}(\theta)\left(\widehat{\xi}_{t_{i+1},s}(\theta) - \widehat{\xi}_{t_i,s}(\theta)\right)}{\displaystyle\sum_{i=0}^{n-1} \widehat{\xi}^2_{t_i,s}(\theta)\Delta t}, \quad t_0 = 0,\ t_n = T,\ \Delta t = 0.02,\ \ s = 1, \dots, N, \quad \text{(A2.21)}$$

where $\widehat{\xi}_{t_i,s}(\theta)$ represents the s–th discrete generation of the O.U. process, evaluated at time t_i, with covariance scale parameter θ, for

$$\theta = [0.1, 0.4, 0.7, 1, 2, 5]\,.$$

Table A2.3.1 displays the empirical probabilities of the error $\widehat{\theta}_T - \theta$ to be within the band $\pm 3\sqrt{\frac{2\theta}{T}}$, from $N = 1000$ discrete simulations of the O.U. process, considering different sample sizes $\{T_l = 12000 + 1000(l-1),\ l = 1, \dots, 7\}$. Figure A2.3.2 displays the cases $\theta = 0.1$ (at the top) and $\theta = 5$ (at the bottom). It can be observed that, for each one of the sample sizes considered, $\{T_l = 12000 + 1000(l-1),\ l = 1,$ approximately a 99% of the realizations of $\widehat{\theta}_T - \theta$ lie within the band $\pm 3\sqrt{\frac{2\theta}{T}}$, which supports the asymptotic Gaussian distribution.

Table A2.3.1: Empirical probabilities of the error of the MLE of θ to lie within the band $\pm 3\sigma = \pm 3\sqrt{\frac{2\theta}{T}}$, for different sample sizes T, and values of parameter θ.

	Parameter θ					
T	0.1	0.4	0.7	1	2	5
12000	0.998	1	0.998	0.998	1	0.998
13000	0.997	0.998	0.998	1	0.995	1
14000	0.998	0.997	1	0.997	1	0.998
15000	0.998	0.997	0.998	0.998	1	0.998
16000	0.997	0.995	0.997	0.998	1	1
17000	0.993	0.998	1	0.997	0.995	1
18000	0.997	0.997	0.995	1	1	0.998

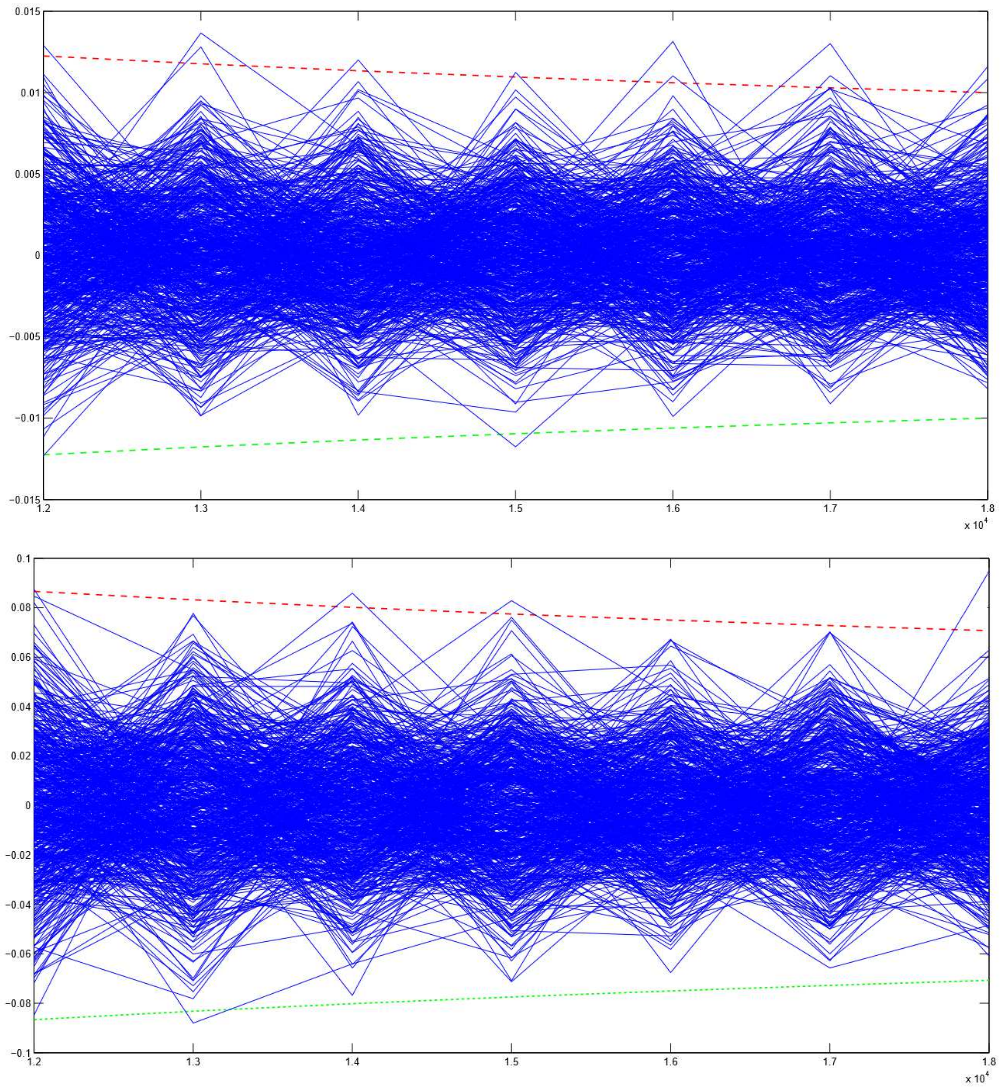


Figure A2.3.2: The values of $\widehat{\theta}_T - \theta$, based on $N = 1000$ simulations of the O.U. process over the interval $[0, T]$, for $\{T_l = 12000 + (l-1)1000,\ l = 1, \dots, 7\}$, are represented against the confidence bands given by $+3\sigma = 3\sqrt{\frac{2\theta}{T}}$ (upper red dotted line) and $-3\sigma = -3\sqrt{\frac{2\theta}{T}}$ (lower green dotted line), for values $\theta = 0.1$ (at the top) and $\theta = 5$ (at the bottom).

Regarding asymptotic efficiency stated in Theorem A2.2.2, from $N = 1000$ simulations of the O.U. process over the interval $[0, T]$, for $\{T_l = 50 + 250(l-1),\ l = 1, \dots, 25\}$, the corresponding empirical mean square errors

$$\mathrm{EMSE}(N, T, \theta) = \frac{1}{N} \sum_{s=1}^{N} \left(\theta - \widehat{\theta}_T(\omega_s) \right)^2, \quad N = 1000, \quad \theta = [0.1, 0.4, 0.7, 1],$$

are displayed in Figure A2.3.3. Here, $\widehat{\theta}_T(\omega_s)$, with $\omega_s \in \Omega$, $s = 1, \dots, N$, represent the respective approximated values (A2.21) of the MLE of θ, computed from $\xi_{t_i,s}$, $s=1,\ldots,N$, $t_i\in[0,T]$, $i=1,\ldots,n$. It can be observed, from the results displayed in Figure A2.3.3, that Theorem A2.2.2 holds for T sufficiently large.

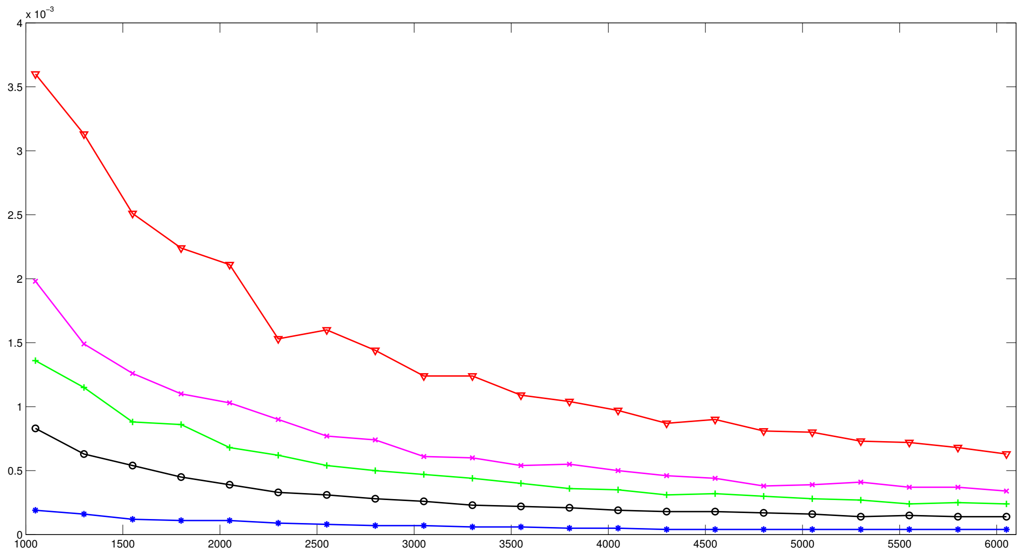


Figure A2.3.3: $\mathrm{EMSE}(N,T,\theta)$ based on $N=1000$ generations of O.U. process, for different sample sizes and values $\theta=0.1$ (blue star line), $\theta=0.4$ (black circles line), $\theta=0.7$ (green plus line), $\theta=1$ (magenta cross line) and $\theta=2$ (red triangle line).

A2.3.2 Consistency of $\rho_{\widehat{\theta}_T}=\rho_{\widehat{\theta}_n}$ in $\mathcal{L}(H)$ and $\mathcal{L}(B)$

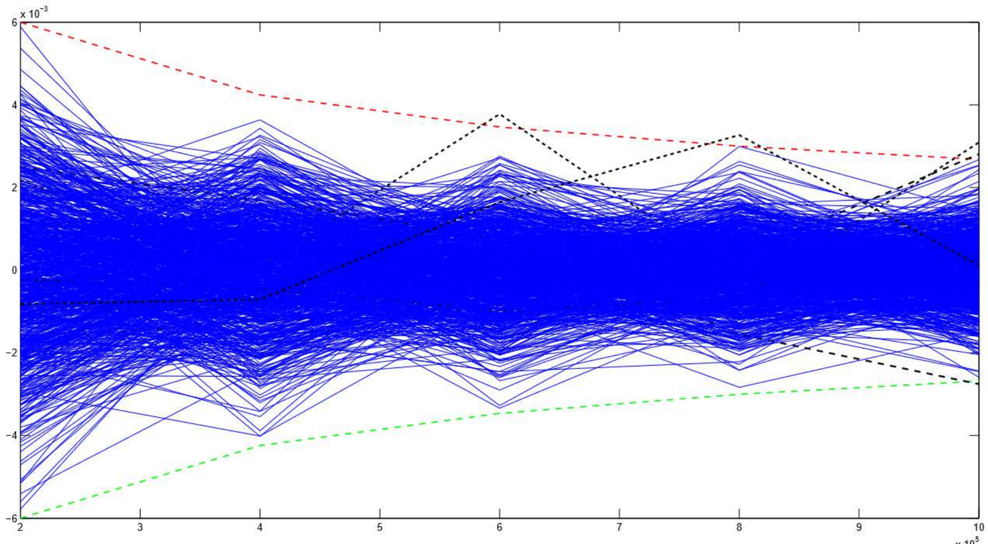
The strong–consistency of $\rho_{\widehat{\theta}_n}$ in $\mathcal{L}(H)$ is derived in Proposition A2.2.1 from the following almost surely upper bound

$$\|\rho_\theta-\rho_{\widehat{\theta}_n}\|^2_{\mathcal{L}(H)}\leq|\theta-\widehat{\theta}_n|h\sqrt{\frac{h}{3}+1}\quad a.s. \tag{A2.22}$$

Here, from $N=1000$ simulations of the O.U. process on the interval $[0,T]$, with sample sizes $T=nh=n=\{200000+(l-1)200000,\ l=1,\ldots,5\}$, the corresponding values of $\widehat{\theta}_T-\theta=\widehat{\theta}_n-\theta$ are computed, considering the cases $\theta=[0.4,0.7,1]$. Table A2.3.2 shows the empirical probability of $\widehat{\theta}_T-\theta$ to lie within the band $\pm3\sqrt{\frac{2\theta}{T}}$, for each one of sample sizes and cases $\theta=[0.4,0.7,1]$ regarded. It can be observed that for the sample sizes studied, in the case of $\theta=1$, the empirical probabilities are equal to one. Thus, the almost surely convergence to zero of the upper bound (A2.22) holds, with an approximated convergence rate of $\sqrt{T}=\sqrt{n}$. Note that, for the other two cases, $\theta=0.4$ and $\theta=0.7$, the empirical probabilities are also very close to one (see also Table A2.3.1 for smaller sample sizes, where we can also observe the empirical probabilities very close to one for the same band). In particular, Figure A2.3.4 displays the cases $\theta=0.4$ (at the top) and $\theta=1$ (at the bottom).

Table A2.3.2: Empirical probability of $\widehat{\theta}_T - \theta$ to be within the band $\pm 3\sigma = \pm 3\sqrt{\frac{2\theta}{T}}$, from $N = 1000$ simulations of an O.U. process over the interval $[0, T]$, with $\{T_l = 200000 + (l-1)200000,\ l = 1, \ldots, 5\}$, considering the cases $\theta = [0.4, 0.7, 1]$.

	Parameter θ		
T	0.4	0.7	1
200000	1	1	1
400000	1	1	1
600000	0.999	1	1
800000	0.999	0.999	1
1000000	0.998	1	1



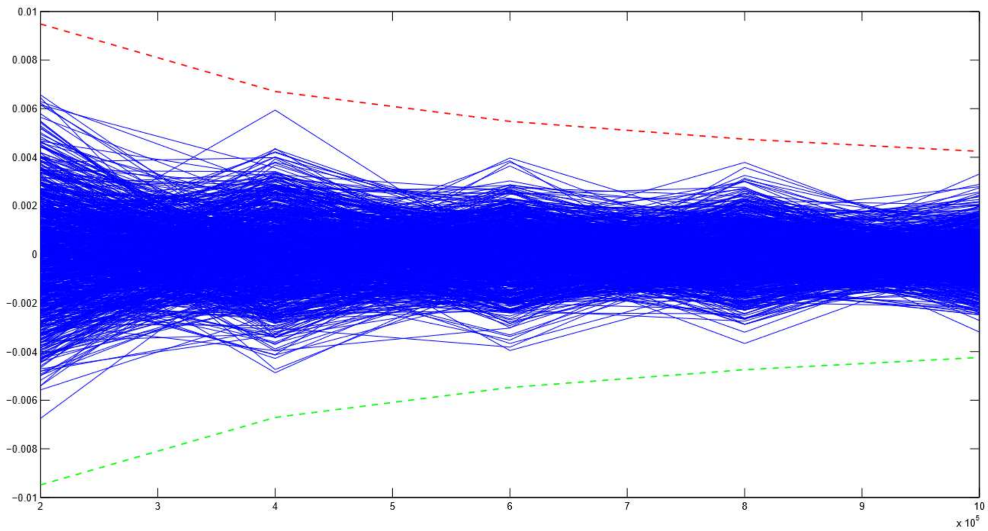


Figure A2.3.4: The values of $\widehat{\theta}_T - \theta$ are represented, corresponding to $N = 1000$ simulations of an O.U. process over the interval $[0, T]$, with $\{T_l = 200000 + (l-1)200000,\ l = 1, \ldots, 5\}$, considering the cases $\theta = 0.4$ (at the top), and $\theta = 1$ (at the bottom). The upper red dotted line is $+3\sqrt{\frac{2\theta}{T}}$ and the lower green dotted line is $-3\sqrt{\frac{2\theta}{T}}$.

It can be observed from Table A2.3.2 that a better performance is obtained for the largest values of θ, which corresponds to the weakest dependent case. Furthermore, from the upper bound in (A2.17), the strong consistency of $\rho_{\widehat{\theta}_n}$ in $\mathcal{L}(B)$, with, as before, $B = \mathcal{C}([0, h])$, is also illustrated from the results displayed in Table A2.3.2 and Figure A2.3.4.

A2.3.3 Consistency of the ARH(1) and ARB(1) plug–in predictors for the O.U. process

Let us now consider the derived upper bounds in (A2.15) and (A2.18) in Corollaries A2.2.1–A2.2.2, for the ARH(1) and ARB(1) predictors, respectively. From the generation of $N = 1000$ discrete realizations of an O.U. process over the interval $[0, T]$, for $\{T_l = 200000 + (l-1)200000,\ l = 1, \ldots, 5\}$, the upper bounds (A2.15) and (A2.18) are evaluated, for the cases $\theta = [0.4, 0.7, 1]$. The following empirical probabilities for $\epsilon = 0.008$, are reflected in Table A2.3.3

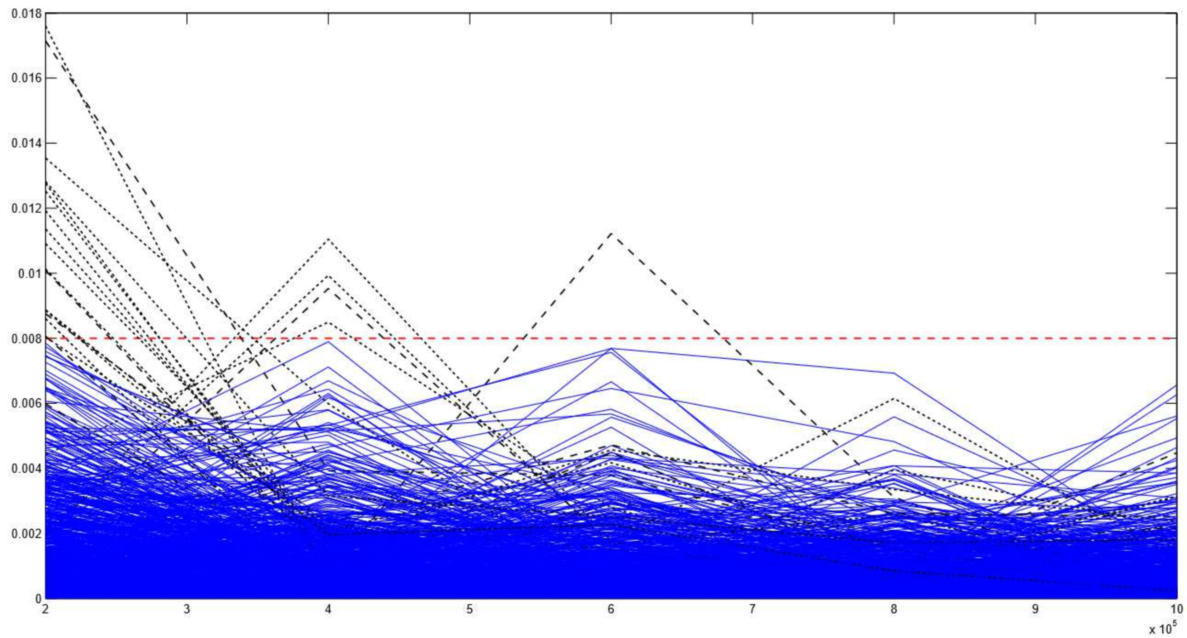
$$\widehat{\mathcal{P}}^H(N, T, \theta) = 1 - \widehat{\mathcal{P}}\left(|X_{n-1}(h)|\,|\theta - \widehat{\theta}_n| h \sqrt{\frac{h}{3} + 1} > \epsilon \right), \tag{A2.23}$$

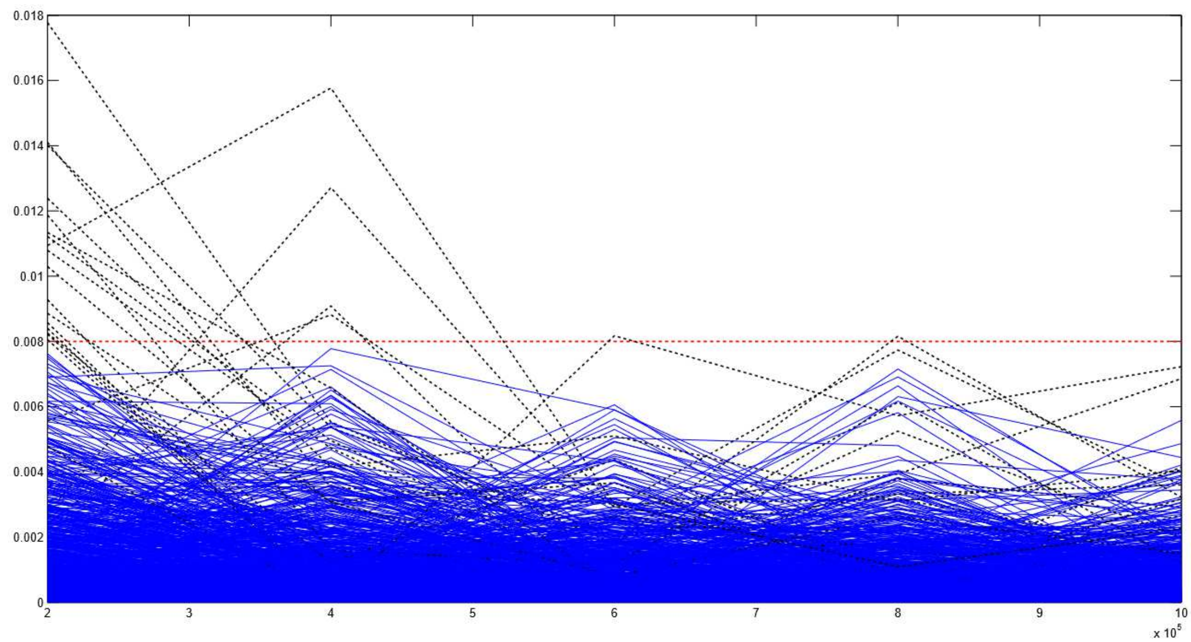
$$\widehat{\mathcal{P}}^B(N, T, \theta) = 1 - \widehat{\mathcal{P}}\left(|X_{n-1}(h)|\,|\theta - \widehat{\theta}_n| h > \epsilon \right), \tag{A2.24}$$

with $N = 1000$, $\{T_l = 200000 + (l-1)200000,\ l = 1, \ldots, 5\}$ and $\theta = [0.4, 0.7, 1]$, for the Hilbert–valued and Banach–valued (see (A2.15) and (A2.18)) frameworks (see also Figure A2.3.5). It can be observed that the empirical probabilities are equal to one in both frameworks for the largest sample sizes, in any of the cases considered.

Table A2.3.3: Empirical probabilities (A2.23)–(A2.24), based on $N = 1000$ simulations of the O.U. process over the interval $[0, T]$, for $\{T_l = 200000 + (l-1)200000,\ l = 1, \dots, 5\}$, considering the cases $\theta = [0.4, 0.7, 1]$, and $\epsilon = 0.008$.

	Parameter θ					
	Hilbert-valued case			Banach-valued case		
T	0.4	0.7	1	0.4	0.7	1
200000	0.980	0.980	0.980	0.987	0.991	0.987
400000	0.995	0.995	0.995	0.997	0.998	0.9977
600000	0.999	0.998	0.999	0.999	0.999	1
800000	1	0.999	0.999	1	1	1
1000000	1	1	1	1	1	1


0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
2
3
4
5
6
7
8
9
10
x 10^5


0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
2
3
4
5
6
7
8
9
10
x 10^5

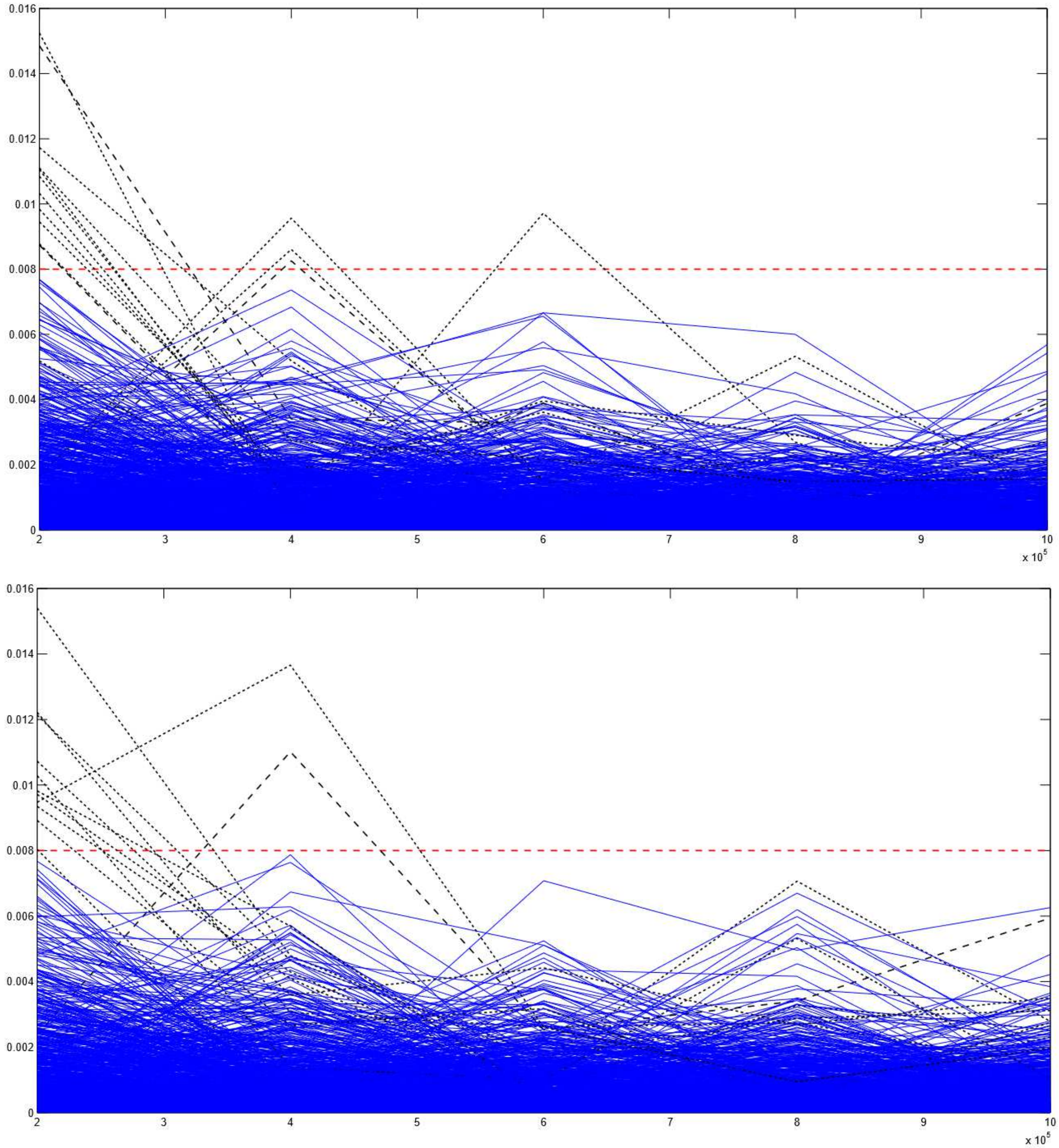


Figure A2.3.5: The values of $|X_{n-1}(h)||\theta-\widehat{\theta}_n|h\sqrt{\frac{h}{3}+1}$ (first and second figure) and $|X_{n-1}(h)||\theta-\widehat{\theta}_n|h$ (third and fourth figure) are represented, based on $N=1000$ generations of O.U. process over the interval $[0,T]$, for $\{T_l = 200000+(l-1)200000,\ l=1,\dots,5\}$, against $\epsilon = 0.008$ (red dotted line), considering $\theta = 0.4$ (first and third figure) and $\theta = 1$ (second and fourth figure).

The strong–consistency of the MLE of θ and of the autocorrelation operator of the O.U. process, in Banach and Hilbert spaces, has been first illustrated. The almost surely rate of convergence to zero is shown as well. The numerical results on the consistency of the associated ARH(1) and ARB(1) plug–in predictors then follow, from the computation of the corresponding empirical probabilities for the derived upper bounds. Note that the numerical results displayed in Appendix A2.3 are obtained under generation of sample sizes ranging from 12000 up to a million of time instants, considering 1000 repetitions for each one of such sample sizes. In all these simulations performed, the discretization step size considered has been $\Delta t = 0.02$.

A2.4 FINAL COMMENTS

The problem of functional prediction of the O.U. process could be of interest in several applied fields. For example, in finance, in the context of the Vasicek's model (see Vasicek [1977]) the results derived allow to predict the curve representing the interest rate over a temporal interval, in a consistent way. Note that, in this context, the MLE computed for parameter θ provides a consistent approximation of the speed reversion, which definitely determines the proposed functional predictor of the interest rate.

Summarizing, this paper addresses the problem of functional prediction of the O.U. process from ARH(1) and ARB(1) perspectives. Specifically, considering the O.U. process as an ARH(1) and an ARB(1) process, new results on strong consistency (almost surely convergence to the true parameter value), in the spaces $\mathcal{L}(H)$ and $\mathcal{L}(B)$ of the MLE of its autocorrelation operator are derived. Consistency results (convergence in probability to the true value) of the associated plug–in predictors are obtained as well. The numerical results shown, in addition, the normality and the asymptotic efficiency of the MLE of the scale parameter θ of the covariance function of the O.U. process.

A2.5 SUPPLEMENTARY MATERIAL

The definition and properties of an O.U. process are given here, as well as the proof of Lemma A2.2.1.

A2.5.1 Ornstein–Uhlenbeck process

Let $\xi(\omega) = \{\xi_t(\omega), \ t \in \mathbb{R}\}, \omega \in \Omega$, be a real–valued sample–path continuous stochastic process defined on the basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$, with index set the real line $\mathbb{R}$. As demonstrated in Doob [1942], process ξ is an O.U. process if it provides the Gaussian solution to the following stochastic linear Langevin differential equation:

$$d\xi_t = \theta(\mu - \xi_t)\, dt + \sigma dW_t, \quad \theta, \ \sigma > 0, \quad t \in \mathbb{R}, \tag{A2.25}$$

where $W = \{W_t, \ t \in \mathbb{R}\}$ is a standard bilateral Wiener process; i.e.,

$$W_t = W_t^{(1)} \mathbf{1}_{\mathbb{R}^+}(t) + W_{-t}^{(2)} \mathbf{1}_{\mathbb{R}^-}(t),$$

with $W_t^{(1)}$ and $W_{-t}^{(2)}$ being independent standard Wiener processes, and $\mathbf{1}_{\mathbb{R}^+}$ and $\mathbf{1}_{\mathbb{R}^-}$ respectively denoting the indicator functions over the positive and negative real line. Applying, in equation (A2.25), the method of separation of variables, considering $f(\xi_t, t) = \xi_t e^{\theta t}$, we obtain

$$\xi_t \;=\; \mu + \int_{-\infty}^{t} \sigma e^{-\theta(t-s)} dW_s, \quad \theta, \quad \sigma > 0, \quad t \in \mathbb{R}, \tag{A2.26}$$

where the integral is understood in the Itô sense (see Ash and Gardner [1975]; Sobczyk [1991] for more details). Particularizing to $\xi = \{\xi_t,\ t \in \mathbb{R}^+\}$, the O.U. process is transformed into

$$\xi_t = \xi_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \int_0^t \sigma e^{-\theta(t-s)} dW_s, \quad \theta, \sigma > 0, \quad t \in \mathbb{R}^+. \tag{A2.27}$$

It is well–known that the solution $\xi = \{\xi_t,\ t \in \mathbb{R}\}$ to the stochastic differential equation

$$d\xi_t = \mu(\xi_t, t)\, dt + \sqrt{D(\xi_t, t)} dW_t, \quad t \in \mathbb{R},$$

has marginal probability density function $f(x,t)$, satisfying the following Fokker–Planck's scalar equation (see, for example, Kadanoff [2000]):

$$\frac{\partial}{\partial t} f(x,t) = \frac{-\partial}{\partial x}\left[\mu(x,t) f(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[D(x,t) f(x,t)\right], \quad t \in \mathbb{R}.$$

In the case of O.U. process, the stationary solution ($\frac{\partial}{\partial t} f(x,t) = 0$), under $f(x, x_0) = \delta(x - x_0)$, adopts the form

$$f(x,t) = \sqrt{\frac{\theta}{\pi\sigma^2}} e^{\frac{-\theta(x-\mu)^2}{\sigma^2}}, \quad \theta, \sigma > 0, \quad t \in \mathbb{R},$$

which corresponds to the probability density function of a Gaussian distribution with mean μ and variance $\frac{\sigma^2}{2\theta}$, i.e., which corresponds to the probability density function of a random variable X such that

$$X \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2\theta}\right).$$

From (A2.26), the mean and covariance functions of O.U. process (see, for instance, Doob [1942]; Uhlenbeck and Ornstein [1930]) can be computed as follows:

$$\begin{aligned}
\mu_\xi(t) &= \mathrm{E}\{\xi_t\} = \mu + \sigma \mathrm{E}\left\{\int_{-\infty}^{t} e^{-\theta(t-s)} dW_s\right\} = \mu, \quad t \in \mathbb{R}, \\
C_\xi(t,s) &= \mathrm{Cov}(\xi_s, \xi_t) = \mathrm{E}\{(\xi_s - \mu)(\xi_t - \mu)\} = \sigma^2 e^{-\theta(t+s)} \mathrm{E}\left\{\int_{-\infty}^{t} e^{\theta u} dW_u \int_{-\infty}^{s} e^{\theta v} dW_v\right\} \\
&= \sigma^2 e^{-\theta(t+s)} \int_{-\infty}^{\infty} e^{2\theta u} \mathbf{1}_{[-\infty,t]}(u)\, \mathbf{1}_{[-\infty,s]}(u)\, du = \sigma^2 e^{-\theta(t+s)} \int_{-\infty}^{\min\{s,t\}} e^{2\theta u} du \\
&= \frac{\sigma^2}{2\theta} e^{-\theta(t+s)} e^{2\theta \min\{s,t\}} = \frac{\sigma^2}{2\theta} e^{-\theta|t-s|}, \quad t, s \in \mathbb{R},
\end{aligned} \tag{A2.28}$$

where $\mathrm{Cov}\,(X,Y)$ denotes the covariance between random variables X and Y. Additionally, from (A2.27), we obtain the following identities:

$$\begin{aligned} \mathrm{E}\left\{\xi_t\right\} &= \mu e^{-\theta t}+\mu\left(1-e^{-\theta t}\right)=\mu, \quad \mathrm{E}\left\{\xi_t|\xi_0=c\right\}=\mu+e^{-\theta t}\left(c-\mu\right), \quad t\in\mathbb{R}^+, \\ \mathrm{Cov}\left(\xi_s,\xi_t|\xi_0=c\right) &= \frac{\sigma^2}{2\theta}e^{-\theta|t-s|}+\left(c^2-2c\mu+\mu^2\right)e^{-\theta(s+t)}, \quad t,s\in\mathbb{R}^+, \end{aligned}$$

where c is a constant. In the subsequent development, we will consider $\mu=0$ and $\sigma=1$.

A2.5.2 Maximum likelihood estimation of the covariance scale parameter θ

The MLE of θ in (A2.28) is given by (see Graczyk and Jakubowski [2006]; [Kutoyants, 2004, p. 63]; [Liptser and Shiraev, 2001, p. 265])

$$\widehat{\theta}_T=\frac{-\displaystyle\int_0^T\xi_t d\xi_t}{\displaystyle\int_0^T\xi_t^2dt}=\frac{\theta\displaystyle\int_0^T\xi_t^2dt-\int_0^T\xi_t dW_t}{\displaystyle\int_0^T\xi_t^2dt}=\theta-\frac{\displaystyle\int_0^T\xi_t dW_t}{\displaystyle\int_0^T\xi_t^2dt},\quad \theta,T>0. \tag{A2.29}$$

Thus, equation (A2.29) becomes

$$\widehat{\theta}_T = \frac{1+\frac{\xi_0^2}{T}-\frac{\xi_T^2}{T}}{\frac{2}{T}\displaystyle\int_0^T\xi_t^2dt},\quad T>0. \tag{A2.30}$$

We will assume that T is large enough such that $\widehat{\theta}_T>0$ almost surely. It is well–known that the MLE $\widehat{\theta}_T$ of θ is strongly consistent (see details in [Kleptsyna and Breton, 2002, Proposition 2.2]; [Kutoyants, 2004, p. 63 and p. 117]).

Theorem A2.5.1 *The following limit in distribution sense holds for the MLE $\widehat{\theta}_T$ of θ, given in equation (A2.30):*

$$\lim_{T\to\infty}\sqrt{T}\left(\widehat{\theta}_T-\theta\right)=\lim_{T\to\infty}\frac{-\sqrt{T}\displaystyle\int_0^T\xi_t dW_t}{\displaystyle\int_0^T\xi_t^2dt}=Z,\quad \textit{with}\quad Z\sim\mathcal{N}\left(0,2\theta\right).$$

Results in [Jiang, 2012, Theorem 1.1 and Corollary 1.1] lead to the following almost surely identities (see also [Bosq, 2000, Theorem 2.10];[Ledoux and Talagrand, 2011, pp. 196–203], in relation to the law of the iterated logarithm)

$$
\begin{aligned}
&\limsup_{T\to+\infty} \frac{\widehat{\theta}_T - \theta}{\sqrt{\frac{4\theta}{T}\ln\left(\ln\left(T\right)\right)}} = 1 \quad a.s.,\\
&-\liminf_{T\to+\infty} \frac{\widehat{\theta}_T - \theta}{\sqrt{\frac{4\theta}{T}\ln\left(\ln\left(T\right)\right)}} = 1 \quad a.s.,\\
&|\theta - \widehat{\theta}_T| = \mathcal{O}\left(\sqrt{\frac{4\theta\ln\left(\ln\left(T\right)\right)}{T}}\right) \quad a.s.
\end{aligned}
$$

A2.5.3 Preliminary inequalities and results

In this section we recall some inequalities and well–known convergence results on random variables, as well as basic deterministic inequalities, that have been applied in the derivation of the main results displayed above.

Lemma A2.5.1 *Let X be a zero–mean normal distributed random variable, i.e., $X \sim \mathcal{N}\left(0,\sigma^2\right)$, with $\sigma > 0$. Then,*

$$
\mathcal{P}\left(|X| \geq x\right) \leq e^{-\frac{x^2}{2\sigma^2}}, \quad x \geq 0.
$$

Proof. Let X' be such that $X' \sim \mathcal{N}\left(0,1\right)$. Then,

$$
\mathcal{P}\left(|X'| \geq x\right) = 2F_{X'}\left(-x\right) = \sqrt{\frac{2}{\pi}}\int_x^{\infty} e^{-\frac{t^2}{2}}dt, \ \forall x \geq 0. \tag{A2.31}
$$

Let us set

$$
\begin{aligned}
g\left(x\right) &= e^{-\frac{x^2}{2}} - \sqrt{\frac{2}{\pi}}\int_x^{\infty} e^{-\frac{t^2}{2}}dt, \quad g\left(0\right) = 0, \quad \lim_{x\to\infty} g\left(x\right) = 0,\\
g'\left(x\right) &= -xe^{-\frac{x^2}{2}} + \sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2}} = e^{-\frac{x^2}{2}}\left(\sqrt{\frac{2}{\pi}} - x\right).
\end{aligned} \tag{A2.32}
$$

Function g is monotone increasing over $\left(0, \sqrt{\frac{2}{\pi}}\right)$, and g is monotone decreasing over $\left(\sqrt{\frac{2}{\pi}}, \infty\right)$.

From equations (A2.31)–(A2.32),

$$
\mathcal{P}\left(|X'| \geq x\right) \leq e^{-\frac{x^2}{2}}, \quad x \geq 0.
$$

Now, consider $X' = \frac{X}{\sigma}$, with $X \sim \mathcal{N}(0, \sigma^2)$, then,

$$\mathcal{P}(|X| \geq x) \leq e^{-\frac{x^2}{2\sigma^2}}, \quad x \geq 0.$$

■

A2.5.3.1 Proof of Lemma 1

Proof.

Let us first consider the case $k = 1$, from

$$\begin{aligned} \rho_\theta(x)(t) &= e^{-\theta t} x(h), \quad \rho_\theta(X_{n-1})(t) = e^{-\theta t} \int_{-\infty}^{nh} e^{-\theta(nh-s)} dW_s, \\ \varepsilon_n(t) &= \int_{nh}^{nh+t} e^{-\theta(nh+t-s)} dW_s, \end{aligned}$$

and

$$\|\rho_\theta(x)\|_H^2 = \int_0^h (\rho_\theta(x)(t))^2 \, d\left(\lambda + \delta_{(h)}\right)(t) = \int_0^h (\rho_\theta(x)(t))^2 \, dt + (\rho_\theta(x)(h))^2,$$

we have

$$\|\rho_\theta\|_{\mathcal{L}(H)} = \sup_{x \in H} \left\{ \frac{\|\rho_\theta(x)\|_H}{\|x\|_H} \right\} = \sup_{x \in H} \left\{ \sqrt{\frac{\left(\int_0^h e^{-2\theta t} dt + e^{-2\theta h}\right)(x(h))^2}{\int_0^h (x(t))^2 \, dt + (x(h))^2}} \right\}. \quad \text{(A2.33)}$$

Furthermore,

$$\|\rho_\theta\|_{\mathcal{L}(H)} = \sup_{x \in H} \left\{ \sqrt{\frac{\left(\int_0^h e^{-2\theta t} dt + e^{-2\theta h}\right)(x(h))^2}{\int_0^h (x(t))^2 \, dt + (x(h))^2}} \right\} \leq \sqrt{\int_0^h e^{-2\theta t} dt + e^{-2\theta h}}. \quad \text{(A2.34)}$$

Additionally, the function $x_0 : [0, h] \longrightarrow \mathbb{R}$, given by

$$x_0(t) = \chi_{\mathcal{M}}(t), \quad h \in \mathcal{M} \subset [0, h], \quad \int_{\mathcal{M}} dt = 0, \quad \text{(A2.35)}$$

with $\mathbf{1}_{\mathcal{M}}$, denoting the indicator function of set $\mathcal{M}$, belongs to $H = L^2\left([0,h], \beta_{[0,h]}, \lambda + \delta_{(h)}\right)$, since

$$x_0^2(h) = 1, \quad \int_0^h x_0^2(t)dt = 0 \quad \|x_0\|_H^2 = \int_0^h x_0^2(s)ds + x_0^2(h) = 1.$$

Thus, by definition of $\|\rho_\theta\|_{\mathcal{L}(H)}$,

$$\frac{\|\rho_\theta(x_0)\|_H}{\|x_0\|_H} = \sqrt{\int_0^h e^{-2\theta t}dt + e^{-2\theta h}} \leq \|\rho_\theta\|_{\mathcal{L}(H)} \tag{A2.36}$$

Equations (A2.33)–(A2.36) lead to

$$\|\rho_\theta\|_{\mathcal{L}(H)} = \sqrt{\int_0^h e^{-2\theta t}dt + e^{-2\theta h}} = \sqrt{\frac{1 + e^{-2\theta h}(2\theta - 1)}{2\theta}}. \tag{A2.37}$$

We are now going to compute $\|\rho_\theta^k\|_{\mathcal{L}(H)}$, for $k \geq 2$. Since, for all $x \in H$,

$$\rho_\theta^k(x)(t) = e^{-\theta t}e^{-\theta(k-1)h}x(h),$$

we obtain

$$\|\rho_\theta^k\|_{\mathcal{L}(H)} = \sup_{x\in H}\left\{\sqrt{\frac{\left[e^{-2\theta(k-1)h}\displaystyle\int_0^h e^{-2\theta t}dt + e^{-2\theta kh}\right](x(h))^2}{\displaystyle\int_0^h (x(t))^2\,dt + (x(h))^2}}\right\}.$$

Considering function x_0 defined in equation (A2.35), applying similar arguments to those given in the computation of $\|\rho_\theta\|_{\mathcal{L}(H)}$, we have

$$\|\rho_\theta^k\|_{\mathcal{L}(H)} = \sqrt{e^{-2\theta(k-1)h}\frac{1 + e^{-2\theta h}(2\theta - 1)}{2\theta}} = e^{-\theta(k-1)h}\|\rho_\theta\|_{\mathcal{L}(H)}.$$

Now, from equation (A2.37),

$$\|\rho_\theta\|_{\mathcal{L}(H)} < 1 \Longleftrightarrow 1 - e^{-2\theta h} < 2\theta\left(1 - e^{-2\theta h}\right) \Longleftrightarrow \theta > \frac{1}{2}.$$

Furthermore, for $\theta \in (0, 1/2]$,

$$\|\rho_\theta\|_{\mathcal{L}(H)} = \sqrt{\alpha(\theta)} < \sqrt{1+h},$$

since $\sqrt{\alpha(\theta)}$ is a monotonically decreasing function on $(0, 1/2]$, with $\alpha(\theta) = 1$ if $\theta = \frac{1}{2}$ and

$\alpha(\theta) \to 1+h$, when $\theta \to 0$. Hence, if $\theta(k-1) \geq 1$,

$$\|\rho_\theta^k\|_{\mathcal{L}(H)} = e^{-\theta(k-1)h}\sqrt{\alpha(\theta)} \leq e^{-h}\sqrt{\alpha(\theta)} < \frac{\sqrt{1+h}}{e^h} < 1, \quad h > 0,$$

which implies that $\|\rho_\theta^{k_0}\|_{\mathcal{L}(H)} < 1$, when $k_0 \geq \frac{1}{\theta} + 1$.

■

A2.5.3.2 Proof of Lemma 2

Proof. Let us first assume that $x \geq y > 0$.

From the Mean Value Theorem applied over e^z, there exists $0 < \alpha < 1$ such that

$$\frac{e^{z+h} - e^z}{h} = e^{z+\alpha h}.$$

Taking $z = -xt$ and $z + h = -yt$, we get the following inequalities:

$$|e^{-xt} - e^{-yt}| = |x-y|te^{-xt+\alpha(x-y)t} = |x-y|te^{xt(\alpha-1)}e^{-y\alpha t} \leq |x-y|te^{-y\alpha t} \leq |x-y|t.$$

Similar inequalities are obtained for the case $y \geq x > 0$, by applying the Mean Value Theorem over the interval $[x,\ y]$, instead of $[y,\ x]$.

■

A2.5.3.3 Proof of Lemma 3

Proof. Considering the indicator function $\mathbf{1}_\cdot$, it holds

$$Y_n|Z_n| = Y_n|Z_n|\mathbf{1}_{\{|Z_n|<a_n\}} + Y_n|Z_n|\mathbf{1}_{\{|Z_n|\geq a_n\}} \leq Y_n a_n + Y_n|Z_n|\mathbf{1}_{\{|Z_n|\geq a_n\}}, \tag{A2.38}$$

where $\{a_n,\ n\in\mathbb{Z}\}$ is a sequence of positive numbers such that the event $\left\{Y_n|Z_n|\mathbf{1}_{\{|Z_n|\geq a_n\}},\ n\in\mathbb{Z}\right\}$ is equivalent to $\{|Z_n| \geq a_n,\ n\in\mathbb{Z}\}$. From (A2.38) and Lemma A2.5.1, if we take $a_n > \frac{\varepsilon}{2}$, for all $n \in \mathbb{Z}$, we get, for each $\varepsilon > 0$,

$$\mathcal{P}(Y_n|Z_n| \geq \varepsilon) \leq \mathcal{P}\left(Y_n a_n \geq \frac{\varepsilon}{2}\right) + \mathcal{P}(|Z_n| \geq a_n) \leq \mathcal{P}\left(Y_n a_n \geq \frac{\varepsilon}{2}\right) + e^{-\theta a_n^2}. \tag{A2.39}$$

For $a_n = c\sqrt{\ln(n)} > \frac{\varepsilon}{2}$, with $\frac{1}{\sqrt{\theta}} < c < +\infty$,

$$\sum_{n\in\mathbb{Z}} \mathcal{P}(|Z_n| \geq a_n) \leq \sum_{n\in\mathbb{Z}} e^{-\theta a_n^2} = \sum_{n\in\mathbb{Z}} \frac{1}{n^{\theta c^2}} < +\infty,$$

which implies that

$$\lim_{n\to\infty} \mathcal{P}(|Z_n| \geq a_n) = 0$$

in equation (A2.39).

On the other hand, since $\sqrt{\ln(n)}Y_n \longrightarrow^p 0$, for every $\varepsilon > 0$,

$$0 = \lim_{n\to\infty} \mathcal{P}\left(\sqrt{\ln(n)}Y_n \geq \frac{\varepsilon}{2}\right) = \lim_{n\to\infty} \mathcal{P}\left(Y_n \frac{a_n}{c} \geq \frac{\varepsilon}{2}\right).$$

Thus, $Y_n|Z_n| \longrightarrow^p 0$.

■

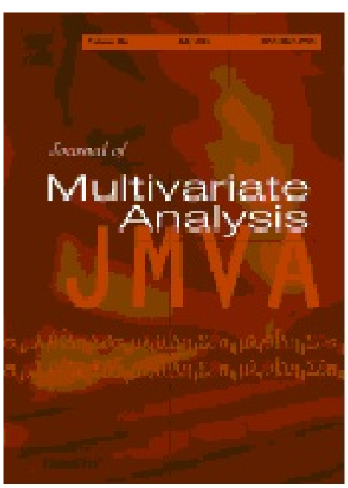
Acknowledgments

This work has been supported in part by projects MTM2012-32674 and MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

A3
ASYMPTOTIC PROPERTIES OF A COMPONENTWISE ARH(1) PLUG–IN PREDICTOR

ÁLVAREZ-LIÉBANA, J.; BOSQ, D.; RUIZ-MEDINA, M. D.: Asymptotic properties of a component–wise ARH(1) plug–in predictor. J. Multivariate Anal. 155 (2017), pp. 12–34.

DOI: doi.org/10.1016/j.jmva.2016.11.009



Year	Categ.	Cites	Impact Factor (5 years)	Quartil
2016	Statist. & Probab.	3938	1.229	Q3

ABSTRACT

This paper presents new results on the prediction of linear processes in function spaces. The autoregressive Hilbertian process framework of order one (ARH(1) framework) is adopted. A componentwise estimator of the autocorrelation operator is derived from the moment–based estimation of its diagonal coefficients with respect to the orthogonal eigenvectors of the autocovariance operator, which are assumed to be known. Mean–square convergence to the theoretical autocorrelation operator is proved in the space of Hilbert–Schmidt operators. Consistency then follows in that space. Mean absolute convergence, in the underlying Hilbert space, of the ARH(1) plug–in predictor to the conditional expectation is obtained as well. A simulation study is undertaken to illustrate the large–sample behaviour of the formulated componentwise estimator and predictor. Additionally, alternative componentwise (with known and unknown eigenvectors), regularized, wavelet–based penalized, and nonparametric kernel estimators of the autocorrelation operator are compared with the one presented here, in terms of prediction.

A3.1 INTRODUCTION

In the last few decades, an extensive literature on statistical inference from functional random variables has emerged. This work was motivated in part by the statistical analysis of high–dimensional data, as well as data of a continuous (infinite-dimensional) nature; see, e.g., Bosq [2000, 2007]; Dedecker and Merlevède [2003]; Ferraty and Vieu [2006]; Merlevède [1996b, 1997]; Ramsay and Silverman [2005]; Ruiz-Medina [2012]. New developments in functional data analysis are described, e.g., in Bongiorno et al. [2014]; Cuevas [2014]; Horváth and Kokoszka [2012]; Hsing and Eubank [2015], and in a recent Special Issue of this journal Goia and Vieu [2016].

These references include a nice summary on the statistics theory for functional data, contemplating covariance operator theory and eigenfunction expansion, perturbation theory, smoothing and regularization, probability measures on a Hilbert spaces, functional principal component analysis, functional counterparts of the multivariate canonical correlation analysis, the two sample problem and the change point problem, functional linear models, functional test for independence, functional time series theory, spatially distributed curves, software packages and numerical implementation of the statistical procedures discussed, among other topics.

The special case of functional regression models, in which the predictor is a random function and the response is scalar, has been particularly well studied. Various specifications of the functional regression parameter arise in fields such as biology, climatology, chemometrics, and economics. To avoid the computational (high–dimensional) limitations of the nonparametric approach, several parametric and semi–parametric methods have been proposed; see, e.g., Ferraty et al. [2012] and the references therein. In Ferraty et al. [2012], a combination of a spline approximation and the one–dimensional Nadaraya–Watson approach was proposed to avoid high dimensionality issues. Generalizations to the case of more regressors (all functional, or both functional and real) were also addressed in the nonparametric, semi–parametric, and parametric frameworks; for an overview, see Aneiros-Pérez and Vieu [2006]; Febrero-Bande and González-Manteiga [2013]; Ferraty and Vieu [2009].

In the nonparametric regression framework, the case where the covariates and the response are functional was considered by Ferraty et al. [2012], where a functional version of the Nadaraya–Watson estimator was proposed for the estimation of the regression operator and shown to be point–wise asymptotically normal. Resampling techniques were used to overcome the difficulties arising in the estimation of the asymptotic bias and variance. Semi–functional partial linear regression, introduced in Aneiros-Pérez and Vieu [2008], allows the prediction of a real-valued random variable from a set of real–valued explanatory variables, and a time–dependent functional explanatory variable. Motivated by genetic and environmental applications, a semi–parametric maximum likelihood method for the estimation of odds ratio association parameters was developed by Chen et al. [2012] in a high–dimensional data context.

In the autoregressive Hilbertian time series framework, several estimation and prediction procedures have been proposed and studied. Mas [1999] established, under suitable conditions, the asymptotic normal distribution of the formulated estimator of the autocorrelation operator, based on projection into the theoretical eigenvectors. In Bosq [2000]; Bosq and Blanke [2007], the problem of prediction of linear processes in function spaces was addressed. In particular, sufficient conditions for the consistency of the empirical autocovariance and cross–covariance operators were obtained. The asymptotic normal distribution of the empirical autocovariance operator was also derived. Moreover, the asymptotic properties of the empirical eigenvalues and eigenvectors were analysed.

Guillas [2001] established the efficiency of a componentwise estimator of the autocorrelation operator, based on projection into the empirical eigenvector system of the autocovariance operator. Consistency, in the space of bounded linear operators, of the formulated estimator of the autocorrelation operator, and of its associated ARH(1) plug–in predictor was later proved by Mas [2004]. He derived sufficient conditions for the weak convergence of the ARH(1) plug–in predictor to a Hilbert–valued Gaussian random variable (see Mas [2007]). Simultaneously, Mas and Menneteau [2003a] obtained high deflection results or large and moderate deviations for infinite–dimensional autoregressive processes. Furthermore, the law of the iterated logarithm for the covariance operator estimator was formulated by Menneteau [2005].

The main properties of the class of autoregressive Hilbertian processes with random coefficients were investigated by Mourid [2004]. Kargin and Onatski [2008] gave interesting extensions of the autoregressive Hilbertian framework, based on the spectral decomposition of the autocorrelation operator, and not of the autocovariance operator. The first generalization on autoregressive processes of order greater than one was proposed by Mourid [1993], in order to improve prediction. ARHX(1) models; i.e., autoregressive Hilbertian processes with exogenous variables were studied by Damon and Guillas [2002, 2005]. In Guillas [2000, 2001] a doubly stochastic formulation of the autoregressive Hilbertian process was investigated. The ARHD model was introduced by Marion and Pumo [2004], taking into account the regularity of trajectories through the derivatives. The conditional autoregressive Hilbertian process (CARH process) was considered by Cugliari [2011], developing parallel projection estimation methods to predict such processes. In the Banach–valued context, we refer to the papers by Bensmain and Mourid [2001]; Dehling and Sharipov [2005]; Pumo [1992, 1998], among others.

In this paper, we assume that the autocorrelation operator belongs to the Hilbert–Schmidt class, and admits a diagonal spectral decomposition in terms of the orthogonal eigenvector system of the autocovariance operator. Such is the case, e.g., of an autocorrelation operator defined as a continuous function of the autocovariance operator. A componentwise estimator of the autocorrelation operator is then constructed in terms of the eigenvectors of the autocovariance operator, which are assumed to be known. This occurs when the random initial condition is defined as the solution, in the mean–square sense, of a stochastic differential equation driven by white noise. Beyond this case, the sparse representation and whitening properties of wavelet bases can be exploited to obtain a diagonal representation of the autocovariance and cross–covariance operators, in terms of a common and known wavelet basis. Unconditional bases, like wavelet bases, also allow the diagonal spectral series representation of the distributional kernels of Calderón-Zygmund operators.

Under the assumptions stated in Appendices A3.2–A3.4, we establish the convergence in the $\mathcal{L}^2$-sense of a componentwise estimator of the autocorrelation operator in the space of Hilbert–Schmidt operators $\mathcal{S}(H)$, i.e., $\mathcal{L}^2_{\mathcal{S}(H)}(\Omega, \mathcal{A}, \mathcal{P})$, is derived. Consistency then follows in $\mathcal{S}(H)$. Under the same conditions, consistency in H of the associated ARH(1) plug–in predictor is obtained, from its convergence in the $\mathcal{L}^1$-sense in the Hilbert space H, i.e., in the space $\mathcal{L}^1_H(\Omega, \mathcal{A}, \mathcal{P})$. The Gaussian framework is analysed in Appendix A3.4 and illustrated in Appendix A3.5, where examples show the behaviour of the proposed componentwise autocorrelation operator estimator, and associated predictor, for large sample sizes. We also present there a comparative study with alternative ARH(1) prediction techniques, including componentwise parameter estimation of the autocorrelation operator, from known and unknown eigenvectors, as well as kernel (nonparametric) functional estimation, and penalized, spline and wavelet, estimation. Final comments on the application of the proposed approach from real data are provided in Appendix A3.6.

A3.2 PRELIMINARIES

This section contains the preliminary definitions and lemmas that will be used to derive the main results of this paper. In the following, H denotes a real separable Hilbert space. Recall that, from Bosq [2000], a zero–mean ARH(1) process $X = \{X_n,\ n \in \mathbb{Z}\}$ satisfies, for all $n \in \mathbb{Z}$, the equation

$$X_n = \rho(X_{n-1}) + \varepsilon_n, \tag{A3.1}$$

where ρ denotes the autocorrelation operator of the process X, which belongs to the space $\mathcal{L}(H)$ of bounded linear operators, such that $\|\rho^k\|_{\mathcal{L}(H)} < 1$, for all integers $k \geq k_0$ beyond a certain $k_0 \geq 1$, with $\|\cdot\|_{\mathcal{L}(H)}$ denoting the norm in the space $\mathcal{L}(H)$. The Hilbert–valued innovation process $\varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}$ is assumed to be a strong–white noise which is uncorrelated with the random initial condition. That is, ε is a Hilbert–valued zero–mean stationary process, with independent and identically distributed components in time, with $\sigma^2_\varepsilon = \mathrm{E}\{\|\varepsilon_n\|^2_H\} < \infty$, for all $n \in \mathbb{Z}$. We restrict our attention here to the case where ρ is such that

$$\|\rho\|_{\mathcal{L}(H)} < 1.$$

The following assumptions are made.

Assumption A1. The autocovariance operator

$$C = \mathrm{E}\{X_n \otimes X_n\} = \mathrm{E}\{X_0 \otimes X_0\}, \quad n \in \mathbb{Z},$$

is a positive, self–adjoint and trace operator. As a result, it admits the following diagonal spectral representation

$$C=\sum_{j=1}^{\infty}C_j\phi_j\otimes\phi_j,$$

in terms of an orthonormal system $\{\phi_j,\ j\geq 1\}$ of eigenvectors which are known. Here,

$$C_1\geq C_2\geq\cdots\geq C_j\geq\cdots>0$$

denote the real positive eigenvalues of C arranged in decreasing order of magnitude and

$$\sum_{j=1}^{\infty}C_j<\infty.$$

Assumption A2. The autocorrelation operator ρ is a self–adjoint and Hilbert–Schmidt operator, admitting the diagonal spectral decomposition

$$\rho=\sum_{j=1}^{\infty}\rho_j\phi_j\otimes\phi_j,\qquad \sum_{j=1}^{\infty}\rho_j^2<\infty,$$

where $\{\rho_j,\ j\geq 1\}$ is the system of eigenvalues of the autocorrelation operator ρ, with respect to the orthonormal system of eigenvectors $\{\phi_j,\ j\geq 1\}$ of the autocovariance operator C.

Note that, under **Assumption A2**,

$$\|\rho\|_{\mathcal{L}(H)}=\sup_{j\geq 1}|\rho_j|<1.$$

Remark A3.2.1 ***Assumption A2*** *holds, in particular, when operator* ρ *is defined as a continuous function of operator C (see [Dautray and Lions, 1990, pp. 119–140] and Remark A3.2.4).*

In the following, for any $n\in\mathbb{Z}$, let

$$D=\mathrm{E}\{X_n\otimes X_{n+1}\}=\mathrm{E}\{X_0\otimes X_1\}$$

be the cross–covariance operator of the ARH(1) process X.

Remark A3.2.2 *Under* ***Assumptions A1–A2****, it follows from equation (A3.1) that*

$$C_{\varepsilon}=C_{\rho}C\rho=\sum_{j=1}^{\infty}C_j\left(1-\rho_j^2\right)\phi_j\otimes\phi_j=\sum_{j=1}^{\infty}\sigma_j^2\phi_j\otimes\phi_j.$$

By projecting equation (A3.1) into the orthonormal system $\{\phi_j,\ j \geq 1\}$, we also have, for each $j \geq 1$ and all $n \in \mathbb{Z}$, the AR(1) equation

$$X_{n,j} = \rho_j X_{n-1,j} + \varepsilon_{n,j}, \quad n \in \mathbb{Z}, \tag{A3.2}$$

where $X_{n,j} = \langle X_n, \phi_j \rangle_H$ and $\varepsilon_{n,j} = \langle \varepsilon_n, \phi_j \rangle_H$, for all $n \in \mathbb{Z}$. From equation (A3.2), we have, for each $j \geq 1$ and all $n \in \mathbb{Z}$,

$$\begin{aligned} \rho_j &= \rho(\phi_j)(\phi_j) = \left\langle \phi_j, DC^{-1}(\phi_j) \right\rangle_H = \langle D(\phi_j), \phi_j \rangle_H \left\langle C^{-1}(\phi_j), \phi_j \right\rangle_H \\ &= \frac{\mathrm{E}\left\{X_{n,j} X_{n-1,j}\right\}}{\mathrm{E}\left\{X_{n-1,j}^2\right\}} = \frac{D_j}{C_j}, \quad n \in \mathbb{Z}, \end{aligned} \tag{A3.3}$$

where

$$D_j = \langle D(\phi_j), \phi_j \rangle_H = \mathrm{E}\left\{X_{n,j} X_{n-1,j}\right\}, \quad C_j^{-1} = [\mathrm{E}\left\{X_{n-1,j}^2\right\}]^{-1}, \quad X_{n,j} = \langle X_n, \phi_j \rangle_H,$$

given that, for all $j \geq 1$,

$$D = \sum_{j=1}^{\infty} D_j \phi_j \otimes \phi_j, \quad D_j = \rho_j C_j, \quad j \geq 1. \tag{A3.4}$$

Let us now consider the Banach space $L_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P})$ of the equivalence classes of $\mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P})$, the space of zero–mean second–order Hilbert–valued random variables ($\mathcal{H}$–valued random variables) with finite seminorm given by

$$\|Z\|_{\mathcal{L}_{\mathcal{H}}^2(\Omega,\mathcal{A},\mathcal{P})} = \sqrt{\mathrm{E}\left\{\|Z\|_{\mathcal{H}}^2\right\}}, \quad \forall Z \in \mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P}).$$

That is, for $Z, Y \in \mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P})$, Z and Y belong to the same equivalence class if and only if

$$\mathrm{E}\left\{\|Z - Y\|_{\mathcal{H}}\right\} = 0.$$

The convergence in the seminorm of $\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})$ will be considered in Proposition A3.2.1, where $\mathcal{H} = \mathcal{S}(H)$ denotes the Hilbert space of Hilbert–Schmidt operators on a Hilbert space H.

For each $n \in \mathbb{Z}$, let us consider the following biorthogonal representation of the functional value X_n of the ARH(1) process $X = \{X_n,\ n \in \mathbb{Z}\}$, and of the functional value ε_n of its innovation process:

$$X_n = \sum_{j=1}^{\infty} \sqrt{C_j} \frac{\langle X_n, \phi_j \rangle_H}{\sqrt{C_j}} \phi_j = \sum_{j=1}^{\infty} \sqrt{C_j} \eta_j(n) \phi_j, \tag{A3.5}$$

$$\varepsilon_n = \sum_{j=1}^{\infty} \sigma_j \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} \phi_j = \sum_{j=1}^{\infty} \sigma_j \widetilde{\eta}_j(n) \phi_j, \tag{A3.6}$$

where

$$\eta_j(n) = \frac{\langle X_n, \phi_j \rangle_H}{\sqrt{C_j}} = \frac{X_{n,j}}{\sqrt{C_j}}, \quad \widetilde{\eta}_j(n) = \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} = \frac{\varepsilon_{n,j}}{\sigma_j}, \quad n \in \mathbb{Z},\ j \geq 1.$$

Here, under **Assumptions A1–A2**, for $C_\varepsilon = \mathrm{E}\left\{\varepsilon_n \otimes \varepsilon_n\right\} = \mathrm{E}\left\{\varepsilon_0 \otimes \varepsilon_0\right\},\ n \in \mathbb{Z}$,

$$C_\varepsilon\left(\phi_j\right) = \sigma_j^2 \phi_j, \quad j \geq 1,$$

where, as before, $\{\phi_j,\ j \geq 1\}$ denotes the system of eigenvectors of the autocovariance operator C, and

$$\sum_{j=1}^{\infty} \sigma_j^2 = \sigma_\varepsilon^2 = \mathrm{E}\left\{\|\varepsilon_n\|_H^2\right\},$$

for all $n \in \mathbb{Z}$.

The following lemma provides the convergence, in the seminorm of $\mathcal{L}_H^2(\Omega, \mathcal{A}, \mathcal{P})$, of the series expansions (A3.5)–(A3.6).

Lemma A3.2.1 *Let $X = \{X_n,\ n \in \mathbb{Z}\}$ be a zero–mean ARH(1) process. Under* **Assumptions A1–A2**, *for any $n \in \mathbb{Z}$, the following limit holds*

$$\lim_{M \to \infty} \mathrm{E}\left\{\left\|X_n - \widehat{X}_{n,M}\right\|_H^2\right\} = 0,$$

where $\widehat{X}_{n,M} = \sum_{j=1}^{M} \sqrt{C_j}\eta_j(n)\phi_j$. Furthermore,

$$\lim_{M \to \infty} \left\|\mathrm{E}\left\{\left(X_n - \widehat{X}_{n,M}\right) \otimes \left(X_n - \widehat{X}_{n,M}\right)\right\}\right\|_{\mathcal{S}(H)}^2 = 0.$$

Similar assertions hold for the biorthogonal series representation

$$\varepsilon_n = \sum_{j=1}^{\infty} \sigma_j \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} \phi_j = \sum_{j=1}^{\infty} \sigma_j \widetilde{\eta}_j(n) \phi_j.$$

Proof.

Under **Assumption A1**, from the trace property of C, the sequence

$$\left\{\widehat{X}_{n,M} = \sum_{j=1}^{M} \sqrt{C_j}\eta_j(n)\phi_j,\ M \geq 1\right\}$$

satisfies, for M sufficiently large, and $L > 0$, arbitrary,

$$\begin{aligned}
\|\widehat{X}_{n,M+L} - \widehat{X}_{n,M}\|^2_{\mathcal{L}^2_H(\Omega,\mathcal{A},P)} &= \mathrm{E}\left\{\|\widehat{X}_{n,M+L} - \widehat{X}_{n,M}\|^2_H\right\} \\
&= \sum_{j=M+1}^{M+L}\sum_{k=M+1}^{M+L}\sqrt{C_j}\sqrt{C_k}\mathrm{E}\left\{\eta_j(n)\eta_k(n)\right\}\langle\phi_j,\phi_k\rangle_H \\
&= \sum_{j=M+1}^{M+L} C_j \to 0, \quad \text{when } M\to\infty,
\end{aligned} \tag{A3.7}$$

since, under **Assumption A1**, $\sum_{j=1}^{\infty} C_j < \infty$. Hence, $\left\{\sum_{j=1}^{M} C_j,\ M\geq 1\right\}$ is a Cauchy sequence. Thus,

$$\lim_{M\to\infty}\sum_{j=M+1}^{M+L} C_j = 0,$$

for $L > 0$ arbitrary. From equation (A3.7),

$$\left\{\widehat{X}_{n,M} = \sum_{j=1}^{M}\sqrt{C_j}\eta_j(n)\phi_j,\ M\geq 1\right\}$$

is also a Cauchy sequence in $\mathcal{L}^2_H(\Omega,\mathcal{A},P)$. Thus, the sequence $\left\{\widehat{X}_{n,M},\ M\geq 1\right\}$ has finite limit in $\mathcal{L}^2_H(\Omega,\mathcal{A},\mathcal{P})$, for all $n\in\mathbb{Z}$.

Furthermore,

$$\begin{aligned}
\lim_{M\to\infty}\mathrm{E}\left\{\left\|X_n - \widehat{X}_{n,M}\right\|^2_H\right\} &= \mathrm{E}\left\{\|X_n\|^2_H\right\} + \lim_{M\to\infty}\sum_{j=1}^{M}\sum_{h=1}^{M}\sqrt{C_j}\sqrt{C_h}\mathrm{E}\left\{\eta_j(n)\eta_h(n)\right\}\langle\phi_j,\phi_h\rangle_H \\
&- 2\lim_{M\to\infty}\sum_{j=1}^{M}\sqrt{C_j}\mathrm{E}\left\{\langle X_n,\eta_j(n)\phi_j\rangle_H\right\} = \sigma^2_X \\
&- \lim_{M\to\infty}\sum_{j=1}^{M} C_j = 0.
\end{aligned} \tag{A3.8}$$

In the derivation of the identities in (A3.7)–(A3.8), we have applied that, for every $j,\ h\geq 1$,

$$
\begin{aligned}
C\left(\phi_j\right) &= C_j\phi_j, \qquad \sigma_X^2 = \mathrm{E}\left\{\|X_n\|_H^2\right\} = \sum_{j=1}^{\infty} C_j < +\infty, \quad \langle \phi_j, \phi_h\rangle_H = \delta_{j,h}, \\
\mathrm{E}\left\{\eta_j(n)\eta_h(n)\right\} &= \delta_{j,h}, \quad \mathrm{E}\left\{\langle X_n, \eta_j(n)\phi_j\rangle_H\right\} = \sqrt{C_j}.
\end{aligned}
\tag{A3.9}
$$

Moreover, from identities in (A3.9),

$$
\begin{aligned}
& \left\|\mathrm{E}\left\{\left(X_n - \lim_{M\to\infty}\widehat{X}_{n,M}\right)\otimes\left(X_n - \lim_{M\to\infty}\widehat{X}_{n,M}\right)\right\}\right\|_{\mathcal{S}(H)}^2 \\
= & \left\|\mathrm{E}\left\{X_n\otimes X_n\right\} + \lim_{M\to\infty}\sum_{j=1}^{M}\sum_{h=1}^{M}\sqrt{C_j}\sqrt{C_h}\phi_j\otimes\phi_h\mathrm{E}\left\{\eta_j(n)\eta_h(n)\right\}\right. \\
& \left. -2\lim_{M\to\infty}\sum_{j=1}^{M}\mathrm{E}\left\{X_n\otimes\sqrt{C_j}\eta_j(n)\phi_j\right\}\right\|_{\mathcal{S}(H)}^2 \\
= & \left\|\mathrm{E}\left\{X_n\otimes X_n\right\} + \lim_{M\to\infty}\left[\sum_{j=1}^{M}C_j\phi_j\otimes\phi_j - 2\sum_{j=1}^{M}C_j\phi_j\otimes\phi_j\right]\right\|_{\mathcal{S}(H)}^2 \\
= & \left\|\mathrm{E}\left\{X_n\otimes X_n\right\} - \lim_{M\to\infty}\sum_{j=1}^{M}C_j\phi_j\otimes\phi_j\right\|_{\mathcal{S}(H)}^2 = 0.
\end{aligned}
\tag{A3.10}
$$

In a similar way, we can derive the convergence to ε_n, in $\mathcal{L}_H^2(\Omega, \mathcal{A}, \mathcal{P})$, of the series $\sum_{j=1}^{\infty}\sigma_j\widetilde{\eta}_j(n)\phi_j$, for every $n\in\mathbb{Z}$, since ε is assumed to be strong–white noise, and hence, its covariance operator C_ε is in the trace class. We can also obtain an analogous to equation (A3.10).

■

In equations (A3.5)–(A3.6), for every $n\in\mathbb{Z}$,

$$
\begin{aligned}
&\mathrm{E}\left\{\eta_j(n)\right\} = 0, \quad \mathrm{E}\left\{\eta_j(n)\eta_h(n)\right\} = \delta_{j,h}, \quad j,h\geq 1, \quad n\in\mathbb{Z}, \\
&\mathrm{E}\left\{\widetilde{\eta}_j(n)\right\} = 0, \quad \mathrm{E}\left\{\widetilde{\eta}_j(n)\widetilde{\eta}_h(n)\right\} = \delta_{j,h}, \quad j,h\geq 1, \quad n\in\mathbb{Z}.
\end{aligned}
\tag{A3.11}
$$

Note that, from **Assumption A2** for each $j\geq 1$, $\{X_{n,j},\ n\in\mathbb{Z}\}$ in equation (A3.2) defines a stationary and invertible AR(1) process. In addition, from equations (A3.5) and (A3.9), for every $n\in\mathbb{Z}$, and $j, p \geq 1$,

$$
\begin{aligned}
X_n &= \sum_{j=1}^{\infty} X_{n,j}\phi_j, \\
\mathrm{E}\left\{X_{n,j}X_{n,p}\right\} &= \sum_{k=0}^{\infty}\sum_{h=0}^{\infty}\rho_j^k\rho_p^h \mathrm{E}\left\{\varepsilon_{n-k,j}\varepsilon_{n-h,p}\right\} = \delta_{j,p}\sum_{k=0}^{\infty}\rho_j^{2k}\sigma_j^2 = \delta_{j,p}\frac{\sigma_j^2}{1-\rho_j^2}, \\
\mathrm{E}\left\{\|X_n\|_H^2\right\} &= \sum_{j=1}^{\infty}\mathrm{E}\left\{X_{n,j}^2\right\} = \sum_{j=1}^{\infty}\langle C(\phi_j), \phi_j\rangle_H = \sum_{j=1}^{\infty} C_j = \sigma_X^2 < \infty,
\end{aligned}
\tag{A3.12}
$$

which implies that

$$
C_j = \frac{\sigma_j^2}{1-\rho_j^2}, \quad j \geq 1.
$$

In particular, we obtain, for each $j \geq 1$, and for every $n \in \mathbb{Z}$,

$$
\begin{aligned}
\mathrm{E}\left\{\eta_j(n)\eta_j(n+1)\right\} &= \mathrm{E}\left\{\frac{X_{n,j}}{\sqrt{C_j}}\frac{X_{n+1,j}}{\sqrt{C_j}}\right\} = \frac{\mathrm{E}\left\{X_{n,j}X_{n+1,j}\right\}}{C_j} \\
&= \frac{\sum_{k=0}^{\infty}\sum_{h=0}^{\infty}\rho_j^{k+h}\mathrm{E}\left\{\varepsilon_{n-k,j}\varepsilon_{n+1-h,j}\right\}}{C_j} \\
&= \frac{\sum_{k=0}^{\infty}\rho_j^{2k+1}\sigma_j^2}{C_j} = \frac{\sigma_j^2}{C_j}\frac{\rho_j}{1-\rho_j^2} = \rho_j.
\end{aligned}
\tag{A3.13}
$$

Remark A3.2.3 From equation (A3.2) and Lemma A3.2.1, keeping in mind that

$$
C_j = \frac{\sigma_j^2}{1-\rho_j^2}, \quad j \geq 1,
$$

the following invertible and stationary AR(1) process can be defined:

$$
\eta_j(n) = \rho_j\eta_j(n-1) + \sqrt{1-\rho_j^2}\,\widetilde{\eta}_j(n), \quad 0 < \rho_j^2 \leq \rho_j < 1, \tag{A3.14}
$$

where, for each $j \geq 1$, $\{\eta_j(n),\ n \in \mathbb{Z}\}$ and $\{\widetilde{\eta}_j(n),\ n \in \mathbb{Z}\}$ are respectively introduced in equations (A3.5)-(A3.6). In the following, for each $j \geq 1$, we assume that

$$
\mathrm{E}\left\{(\widetilde{\eta}_j(n))^4\right\} < \infty, \quad n \in \mathbb{Z},
$$

to ensure ergodicity for all second–order moments, in the mean–square sense; see, e.g., [Hamilton, 1994, pp. 192–193].

Furthermore,

$$\begin{aligned} D = \mathrm{E}\{X_n \otimes X_{n+1}\} &= \sum_{j=1}^{\infty}\sum_{p=1}^{\infty} \mathrm{E}\left\{\langle X_n, \phi_j\rangle_H \langle X_{n+1}, \phi_p\rangle_H\right\} \phi_j \otimes \phi_p \\ &= \sum_{j=1}^{\infty}\sum_{p=1}^{\infty} \sqrt{C_j}\sqrt{C_p}\frac{\mathrm{E}\left\{\langle X_n, \phi_j\rangle_H \langle X_{n+1}, \phi_p\rangle_H\right\}}{\sqrt{C_j}\sqrt{C_p}} \phi_j \otimes \phi_p \\ &= \sum_{j=1}^{\infty}\sum_{p=1}^{\infty} \sqrt{C_j}\sqrt{C_p}\mathrm{E}\{\eta_j(n)\eta_p(n+1)\} \phi_j \otimes \phi_p. \end{aligned}$$

Remark A3.2.4 *In particular, **Assumption A2** holds if the following orthogonality condition is satisfied, for all $n \in \mathbb{Z}$ and $j, p \geq 1$,*

$$\mathrm{E}\{\eta_j(n)\eta_p(n+1)\} = \delta_{j,p},$$

where $\delta_{j,p}$ denotes the Kronecker Delta function. In practice, unconditional bases, e.g., wavelet bases, lead to a sparse representation for functional data; see, e.g., Nason [2008]; Ogden [1997]; Vidakovic [1998] for statistically-oriented treatments. Wavelet bases are also designed for sparse representation of kernels defining integral operators, in L^2 spaces with respect to a suitable measure (see Mallat [2009]). The Discrete Wavelet Transform (DWT) approximately decorrelates or whitens *data (see Vidakovic [1998]). In particular, operators C and D could admit an almost diagonal representation with respect to the self-tensorial tensorial product of a suitable wavelet basis.*

A3.3 Estimation and prediction results.

A componentwise estimator of the autocorrelation operator and of the associated ARH(1) plug–in predictor are formulated in this section. Their convergence to the corresponding theoretical functional values are derived in the spaces $\mathcal{L}^2_{\mathcal{S}(H)}(\Omega, \mathcal{A}, \mathcal{P})$ and $\mathcal{L}_H(\Omega, \mathcal{A}, \mathcal{P})$, respectively. Their consistency in the spaces $\mathcal{S}(H)$ and H then follows.

From equation (A3.3), for each $j \geq 1$, and for a given sample size n, one can consider the usual respective moment–based estimators $\widehat{D}_{n,j}$ and $\widehat{C}_{n,j}$ of D_j and C_j, in the AR(1) framework, given by

$$\widehat{D}_{n,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,j}X_{i+1,j}, \quad \widehat{C}_{n,j} = \frac{1}{n}\sum_{i=0}^{n-1} X_{i,j}^2.$$

The following truncated componentwise estimator of ρ is then formulated:

$$\widehat{\rho}_{k_n} = \sum_{j=1}^{k_n} \widehat{\rho}_{n,j} \phi_j \otimes \phi_j, \tag{A3.15}$$

where, for each $j \geq 1$,

$$\widehat{\rho}_{n,j} = \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}} = \frac{\frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\frac{1}{n}\sum_{i=0}^{n-1} X_{i,j}^2} = \frac{n}{n-1} \frac{\sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\sum_{i=0}^{n-1} X_{i,j}^2}. \tag{A3.16}$$

Here, the truncation parameter indicates that we have considered the first k_n eigenvectors associated with the first k_n eigenvalues, arranged in decreasing order of their modulus magnitude. Furthermore, k_n is such that

$$\lim_{n\to\infty} k_n = \infty, \qquad \frac{k_n}{n} < 1, \qquad n \geq 2. \tag{A3.17}$$

The following additional condition will be assumed on k_n for the derivation of the subsequent results:

Assumption A3. The truncation parameter k_n in (A3.15) is such that

$$\lim_{n\to\infty} C_{k_n}\sqrt{n} = \infty.$$

Remark A3.3.1 ***Assumption A3*** *has also been considered in [Bosq, 2000, p. 217], to ensure weak consistency of the proposed estimator of ρ, as well as, in [Mas, 1999, Proposition 4], in the derivation of asymptotic normality.*

From Remark A3.2.3, for each $j \geq 1$, $\eta_j = \{\eta_j(n),\ n \in \mathbb{Z}\}$ in equation (A3.14) defines a stationary and invertible AR(1) process, ergodic in the mean–square sense; see, e.g., Bartlett [1946]. Therefore, in view of equations (A3.11) and (A3.13), for each $j \geq 1$, there exist two positive constants $K_{j,1}$ and $K_{j,2}$ such that the following identities hold:

$$\lim_{n\to\infty} \frac{\mathrm{E}\left\{\left[1 - \frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2(i)\right]^2\right\}}{\frac{1}{n}} = K_{j,1}, \tag{A3.18}$$

$$\lim_{n\to\infty} \frac{\mathrm{E}\left\{\left[\rho_j - \frac{1}{n-1}\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)\right]^2\right\}}{\frac{1}{n}} = K_{j,2}. \tag{A3.19}$$

Equations (A3.18)-(A3.19) imply, for n sufficiently large,

$$\operatorname{Var}\left\{\frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2(i)\right\} \leq \frac{\widetilde{K}_{j,1}}{n}, \tag{A3.20}$$

$$\operatorname{Var}\left\{\frac{1}{n-1}\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)\right\} \leq \frac{\widetilde{K}_{j,2}}{n}, \tag{A3.21}$$

for certain positive constants $\widetilde{K}_{j,1}$ and $\widetilde{K}_{j,2}$, for each $j \geq 1$. Equivalently, for n sufficiently large,

$$\mathrm{E}\left\{\left(1-\frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2(i)\right)^2\right\} \leq \frac{\widetilde{K}_{j,1}}{n}, \tag{A3.22}$$

$$\mathrm{E}\left\{\left(\rho_j-\frac{1}{n-1}\sum_{i=0}^{n-1}\eta_j(i)\eta_j(i+1)\right)^2\right\} \leq \frac{\widetilde{K}_{j,2}}{n}, \tag{A3.23}$$

The following assumption is now considered.

Assumption A4. We assume that

$$S = \sup_{j\geq 1}\left(\widetilde{K}_{j,1}+\widetilde{K}_{j,2}\right) < \infty.$$

Remark A3.3.2 *From equation (A3.16), applying the Cauchy–Schwarz's inequality, we obtain, for each* $j \geq 1$,

$$\begin{aligned}
|\widehat{\rho}_{n,j}| &= \frac{n}{n-1}\left|\frac{\sum_{i=0}^{n-2}X_{i,j}X_{i+1,j}}{\sum_{i=0}^{n-1}X_{i,j}^2}\right| \leq \frac{n}{n-1}\frac{\sqrt{\sum_{i=0}^{n-2}X_{i,j}^2\sum_{i=0}^{n-2}X_{i+1,j}^2}}{\sum_{i=0}^{n-1}X_{i,j}^2} \\
&\leq \frac{n}{n-1}\sqrt{\frac{\sum_{i=0}^{n-2}X_{i+1,j}^2}{\sum_{i=0}^{n-1}X_{i,j}^2}} \leq \frac{n}{n-1}\,a.s.
\end{aligned} \tag{A3.24}$$

A3.3.1 Convergence in $\mathcal{L}^2_{\mathcal{S}(H)}(\Omega, \mathcal{A}, \mathcal{P})$

Next, the convergence of $\widehat{\rho}_{k_n}$ to ρ, in the space $\mathcal{L}^2_{\mathcal{S}(H)}(\Omega, \mathcal{A}, \mathcal{P})$, is derived under the setting of conditions formulated in the previous sections.

Proposition A3.3.1 *Let $X = \{X_n,\ n \in \mathbb{Z}\}$ be a zero–mean standard ARH(1) process. Under* ***Assumptions A1–A4****, the following limit holds:*

$$\lim_{n\to\infty} \|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} = 0. \tag{A3.25}$$

Specifically,

$$\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} \leq g(n), \quad \textit{with} \quad g(n) = \mathcal{O}\left(\frac{1}{C^2_{k_n} n}\right), \quad n \to \infty. \tag{A3.26}$$

Remark A3.3.3 *[Bosq, 2000, Corollary 4.3] can be applied to obtain weak convergence results, in terms of weak expectation, using the empirical eigenvectors . See definition of weak expectation at the beginning of [Bosq, 2000, Section 1.3, p. 27]).*

Proof. For each $j \geq 1$, the following almost surely inequality is satisfied:

$$\begin{aligned} |\rho_j - \widehat{\rho}_{n,j}| &= \left|\frac{D_j}{C_j} - \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}}\right| = \left|\frac{D_j\widehat{C}_{n,j} - \widehat{D}_{n,j}C_j}{C_j\widehat{C}_{n,j}}\right| \\ &= \left|\frac{D_j\widehat{C}_{n,j} - \widehat{D}_{n,j}C_j + \widehat{C}_{n,j}\widehat{D}_{n,j} - \widehat{C}_{n,j}\widehat{D}_{n,j}}{C_j\widehat{C}_{n,j}}\right| \\ &= \left|\frac{D_j - \widehat{D}_{n,j}}{C_j} + \frac{\widehat{C}_{n,j} - C_j}{C_j}\frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}}\right| \leq \frac{1}{C_j}\left(|\widehat{\rho}_{n,j}|\left|C_j - \widehat{C}_{n,j}\right| + \left|D_j - \widehat{D}_{n,j}\right|\right). \end{aligned}$$

Thus, under **Assumptions A1–A2**, from equation (A3.24), for each $j \geq 1$,

$$\begin{aligned} (\rho_j - \widehat{\rho}_{n,j})^2 &\leq \frac{1}{C_j^2}\left(|\widehat{\rho}_{n,j}|\left|C_j - \widehat{C}_{n,j}\right| + \left|D_j - \widehat{D}_{n,j}\right|\right)^2 \\ &\leq \frac{2}{C_j^2}\left((\widehat{\rho}_{n,j})^2\left(C_j - \widehat{C}_{n,j}\right)^2 + \left(D_j - \widehat{D}_{n,j}\right)^2\right) \\ &\leq \frac{2}{C_j^2}\left(\left(\frac{n}{n-1}\right)^2\left(C_j - \widehat{C}_{n,j}\right)^2 + \left(D_j - \widehat{D}_{n,j}\right)^2\right) \ a.s., \end{aligned}$$

which implies, for each $j \geq 1$,

$$\mathrm{E}\left\{(\rho_j - \widehat{\rho}_{n,j})^2\right\} \leq \frac{2}{C_j^2}\left(\left(\frac{n}{n-1}\right)^2 \mathrm{E}\left\{\left(C_j - \widehat{C}_{n,j}\right)^2\right\} + \mathrm{E}\left\{\left(D_j - \widehat{D}_{n,j}\right)^2\right\}\right). \quad \text{(A3.27)}$$

Under **Assumption A2**, from equations (A3.15) and (A3.27),

$$\begin{aligned}
\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} &= \mathrm{E}\left\{\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{S}(H)}\right\} = \sum_{j=1}^{k_n} \mathrm{E}\left\{(\rho_j - \widehat{\rho}_{n,j})^2\right\} + \sum_{j=k_n+1}^{\infty} \mathrm{E}\left\{\rho_j^2\right\} \\
&\leq \sum_{j=1}^{k_n} \frac{2}{C_j^2}\left(\left(\frac{n}{n-1}\right)^2 \mathrm{E}\left\{\left(C_j - \widehat{C}_{n,j}\right)^2\right\}\right. \\
&\quad \left. + \mathrm{E}\left\{\left(D_j - \widehat{D}_{n,j}\right)^2\right\}\right) + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\
&\leq \frac{2}{C_{k_n}^2} \sum_{j=1}^{k_n} \left(\frac{n}{n-1}\right)^2 \left(\mathrm{E}\left\{\left(C_j - \widehat{C}_{n,j}\right)^2\right\}\right. \\
&\quad \left. + \mathrm{E}\left\{\left(D_j - \widehat{D}_{n,j}\right)^2\right\}\right) + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\
&\leq \frac{2\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2} \sum_{j=1}^{k_n} \left(\mathrm{E}\left\{\left(C_j - \widehat{C}_{n,j}\right)^2\right\} + \mathrm{E}\left\{\left(D_j - \widehat{D}_{n,j}\right)^2\right\}\right) \\
&+ \sum_{j=k_n+1}^{\infty} \rho_j^2. \qquad \text{(A3.28)}
\end{aligned}$$

Furthermore, from (A3.5) and (A3.16), for each $j \geq 1$,

$$\widehat{C}_{n,j} = \frac{1}{n}\sum_{i=0}^{n-1} X_{i,j}^2 = \frac{1}{n}\sum_{i=0}^{n-1} C_j \eta_j^2(i), \quad \text{(A3.29)}$$

$$\widehat{D}_{n,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,j}X_{i+1,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} C_j \eta_j(i)\eta_j(i+1), \quad \text{(A3.30)}$$

where, considering equation (A3.4),

$$D_j = \mathrm{E}\left\{X_{n,j}X_{n+1,j}\right\} = C_j \mathrm{E}\left\{\eta_j(n)\eta_j(n+1)\right\} = C_j\rho_j, \quad \text{(A3.31)}$$

for each $j \geq 1$. Equations (A3.28)–(A3.31) then lead to

$$\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} \leq \frac{2\left(\frac{n}{n-1}\right)^2}{C^2_{k_n}} \sum_{j=1}^{k_n} C_j^2 \left(\mathrm{E}\left\{ \left(1 - \frac{1}{n}\sum_{i=0}^{n-1} \eta_j^2(i)\right)^2 \right\} \right.$$
$$\left. + \mathrm{E}\left\{ \left(\rho_j - \frac{1}{n-1}\sum_{i=0}^{n-2} \eta_j(i+1)\eta_j(i)\right)^2 \right\} \right)$$
$$+ \sum_{j=k_n+1}^{\infty} \rho_j^2.$$

For each $j \geq 1$, and for n sufficiently large, considering equations (A3.22)–(A3.23), under **Assumption A4**,

$$\begin{aligned} \mathrm{E}\left\{ \|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{S}(H)} \right\} &\leq \frac{2\left(\frac{n}{n-1}\right)^2}{C^2_{k_n}} \sum_{j=1}^{k_n} C_j^2 \left(\frac{\widetilde{K}_{j,1} + \widetilde{K}_{j,2}}{n} \right) + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\ &\leq \frac{2S\left(\frac{n}{n-1}\right)^2}{C^2_{k_n} n} \sum_{j=1}^{k_n} C_j^2 + \sum_{j=k_n+1}^{\infty} \rho_j^2. \end{aligned} \tag{A3.32}$$

From the trace property of operator C,

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} C_j^2 = \sum_{j=1}^{\infty} C_j^2 < \infty, \tag{A3.33}$$

and from the Hilbert–Schmidt property of ρ,

$$\lim_{n\to\infty} \sum_{j=k_n+1}^{\infty} \rho_j^2 = 0. \tag{A3.34}$$

Thus, in view of equations (A3.32)–(A3.34),

$$\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} = \mathrm{E}\left\{ \|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{S}(H)} \right\} \leq g(n) = \mathcal{O}\left(\frac{1}{C^2_{k_n} n} \right), \; n \to \infty, \tag{A3.35}$$

where

$$g(n) = \frac{2S\left(\frac{n}{n-1}\right)^2}{C^2_{k_n} n} \sum_{j=1}^{k_n} C_j^2 + \sum_{j=k_n+1}^{\infty} \rho_j^2. \tag{A3.36}$$

Under **Assumption A3**, equations (A3.35)–(A3.36) imply

$$\lim_{n\to\infty} \|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} = 0,$$

as we wanted to prove.

■

Note that consistency of $\widehat{\rho}_{k_n}$ in the space $\mathcal{S}(H)$ directly follows from equation (A3.25) in Proposition A3.3.1.

Corollary A3.3.1 *Let $X = \{X_n,\ n \in \mathbb{Z}\}$ be a zero–mean standard ARH(1) process. Under* ***Assumptions A1–A4****, as long as $n \to \infty$,*

$$\|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{S}(H)} \to^p 0,$$

where, as usual, $\to^p$ denotes the convergence in probability.

A3.3.2 Consistency of the ARH(1) plug–in predictor.

Let us consider $\mathcal{L}(H)$ the space of bounded linear operators on H, with the norm

$$\|\mathcal{A}\|_{\mathcal{L}(H)} = \sup_{x\in H} \frac{\|\mathcal{A}(x)\|_H}{\|x\|_H},$$

for every $\mathcal{A} \in \mathcal{L}(H)$. In particular, for each $x \in H$,

$$\|\mathcal{A}(x)\|_H \leq \|\mathcal{A}\|_{\mathcal{L}(H)} \|x\|_H. \tag{A3.37}$$

In the following, we denote by

$$\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1}) \tag{A3.38}$$

as usual, the ARH(1) plug–in predictor of X_n, as an estimator of the conditional expectation $\mathrm{E}\{X_n | X_{n-1}\} = \rho(X_{n-1})$. The following proposition provides the consistency of $\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1})$ in H.

Proposition A3.3.2 *Let $X = \{X_n,\ n \in \mathbb{Z}\}$ be a zero–mean standard ARH(1) process. Under* ***Assumptions A1–A4****,*

$$\lim_{n\to\infty} \mathrm{E}\{\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H\} = 0.$$

Specifically,

$$\mathrm{E}\{\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H\} \leq h(n), \quad h(n) = \mathcal{O}\left(\frac{1}{C_{k_n}\sqrt{n}}\right), \quad n \to \infty.$$

In particular,

$$\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H \to^p 0,$$

where, as usual, $\to^p$ denotes the convergence in probability.

Proof.

From (A3.37) and Proposition A3.3.1, for n sufficiently large, the following almost surely inequality holds:

$$\left\|\rho(X_{n-1}) - \widehat{X}_n\right\|_H \leq \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}(H)} \|X_{n-1}\|_H,$$

where, as given in equation (A3.38), $\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1})$. Thus,

$$\mathrm{E}\left\{\left\|\rho(X_{n-1}) - \widehat{X}_n\right\|_H\right\} \leq \mathrm{E}\left\{\|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}(H)} \|X_{n-1}\|_H\right\}. \tag{A3.39}$$

From the Cauchy-Schwarz's inequality, keeping in mind that, for a Hilbert–Schmidt operator $\mathcal{K}$, it always holds that $\|\mathcal{K}\|_{\mathcal{L}(H)} \leq \|\mathcal{K}\|_{\mathcal{S}(H)}$, we have from equation (A3.39),

$$\begin{aligned}
\mathrm{E}\left\{\left\|X_n - \widehat{X}_n\right\|_H\right\} &\leq \sqrt{\mathrm{E}\left\{\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}(H)}\right\}}\sqrt{\mathrm{E}\left\{\|X_{n-1}\|^2_H\right\}} \\
&\leq \sqrt{\mathrm{E}\left\{\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{S}(H)}\right\}}\sqrt{\mathrm{E}\left\{\|X_{n-1}\|^2_H\right\}} \\
&= \sqrt{\mathrm{E}\left\{\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{S}(H)}\right\}}\sigma_X,
\end{aligned} \tag{A3.40}$$

where, as before, $\sigma_X^2 = \mathrm{E}\left\{\|X_{n-1}\|^2_H\right\} = \sum_{j=1}^{\infty} C_j < \infty, \quad n \in \mathbb{Z}$ (see equation (A3.9)).

Since from Proposition A3.3.1 (see equation (A3.26)),

$$\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} \leq g(n), \quad \text{with} \quad g(n) = \mathcal{O}\left(\frac{1}{C_{k_n}^2 n}\right), \quad n \to \infty,$$

from equation (A3.40), we obtain,

$$\mathrm{E}\left\{\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H\right\} \leq h(n),$$

where $h(n) = \sigma_X \sqrt{g(n)}$, with $g(n)$ being given in (A3.36). In particular, under **Assumption A3**,

$$\lim_{n\to\infty} \mathrm{E}\left\{\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H\right\} = 0,$$

which implies that

$$\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H = \left\|\rho(X_{n-1}) - \widehat{X}_n\right\|_H \to^p 0, \quad n \to \infty.$$

A3.4 The Gaussian case.

In this section, we prove that, in the Gaussian ARH(1) context, **Assumptions A1–A2** and **A4** also hold. From equation (A3.11), for $n \geq 1$,

$$\mathrm{E}\left\{\frac{\sum_{i=0}^{n-1}\eta_j^2(i)}{n}\right\} = 1.$$

Furthermore, for each $j \geq 1$ and $n \geq 2$, the $n \times 1$ random vector $\boldsymbol{\eta}_j^T = (\eta_j(0), \ldots, \eta_j(n-1))$ follows a Multivariate Normal distribution with null mean vector, and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_j & 0 & \ldots & \ldots & 0 \\ \rho_j & 1 & \rho_j & 0 & \ldots & 0 \\ 0 & \rho_j & 1 & \rho_j & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \rho_j & 1 \end{pmatrix}_{n\times n}. \tag{A3.41}$$

It is well–known (see, for example, Gurland [1956]) that the variance of a quadratic form defined from a multivariate Gaussian vector $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, and a symmetric matrix $\mathbf{Q}$ is given by:

$$\mathrm{Var}\left\{\mathbf{y}^T\boldsymbol{Q}\mathbf{y}\right\} = 2\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}\boldsymbol{\Lambda}\right) + 4\boldsymbol{\mu}^T\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}\boldsymbol{\mu}. \tag{A3.42}$$

For each $j \geq 1$, applying equation (A3.42), with $\mathbf{y} = \boldsymbol{\eta}_j$, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}$ in (A3.41), and $\boldsymbol{Q} = \boldsymbol{Id}_n$, the $n \times n$ identity matrix, keeping in mind $\mathrm{E}\left\{\eta_j(i)\eta_j(i+1)\right\} = \rho_j$, for every $i \in \mathbb{Z}$,

$$\mathrm{Var}\left\{\boldsymbol{\eta}_j^T\boldsymbol{Id}_n\boldsymbol{\eta}_j\right\} = \mathrm{Var}\left\{\sum_{i=0}^{n-1}\eta_j^2(i)\right\} = 2\mathrm{Tr}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}\right) = 2\left(n + 2(n-1)\rho_j^2\right). \tag{A3.43}$$

Furthermore, from equation (A3.43), for each $j \geq 1$,

$$\mathrm{Var}\left\{\frac{\sum_{i=0}^{n-1}\eta_j^2(i)}{n}\right\} = \frac{2}{n^2}\left(n + 2(n-1)\rho_j^2\right) = \frac{2}{n} + 4\left(\frac{1}{n} - \frac{1}{n^2}\right)\rho_j^2. \tag{A3.44}$$

We then obtain, from equation (A3.44),

$$\begin{aligned}\lim_{n\to\infty} \operatorname{Var}\left\{\frac{\sum_{i=0}^{n-1}\eta_j^2(i)}{n}\right\} &= \lim_{n\to\infty} \mathrm{E}\left\{\left(1-\frac{\sum_{i=0}^{n-1}\eta_j^2(i)}{n}\right)^2\right\} \\ &= \lim_{n\to\infty}\frac{2}{n}+4\left(\frac{1}{n}-\frac{1}{n^2}\right)\rho_j^2 = 0. \end{aligned} \tag{A3.45}$$

Equation (A3.45) leads to

$$\lim_{n\to\infty}\frac{\operatorname{Var}\left\{\frac{\sum_{i=0}^{n-1}\eta_j^2(i)}{n}\right\}}{\frac{1}{n}} = 2+4\rho_j^2.$$

Hence, for each $j \geq 1$, $K_{j,1}$ in equation (A3.18) is given by

$$K_{j,1} = 2+4\rho_j^2,$$

and, from equation (A3.44),

$$\operatorname{Var}\left\{\frac{\sum_{i=0}^{n-1}\eta_j^2(i)}{n}\right\} \leq 2+4\left(\frac{1}{n}-\frac{1}{n^2}\right)\rho_j^2 \leq 2+4\rho_j^2 \leq 6.$$

Thus, for every $j \geq 1$, $\widetilde{K}_{j,1}$ in equation (A3.20) satisfies

$$\widetilde{K}_{j,1} \leq 6.$$

Remark A3.4.1 *Note that, from Lemma A3.2.1, for each* $j \geq 1$ *and* $i \in \mathbb{Z}$*,*

$$\mathrm{E}\left\{\widetilde{\eta}_j^4(i)\right\} = 3.$$

Thus, the assumption considered in Remark A3.2.3 holds, and for each $j \geq 1$*, the AR(1) process* $\eta_j = \{\eta_j(n),\ n \in \mathbb{Z}\}$ *is ergodic for all second-order moments, in the mean-square sense; see [Hamilton, 1994, pp. 192–193].*

For $n \geq 2$, and for each $j \geq 1$, we are now going to compute $K_{j,2}$ in (A3.19). The $(n-1) \times 1$ random vectors

$$\boldsymbol{\eta}_j^{\star} = (\eta_j(0), \ldots, \eta_j(n-2))^T, \quad \boldsymbol{\eta}_j^{\star\star} = (\eta_j(1), \ldots, \eta_j(n-1))^T$$

are multivariate Normal distributed, with null mean vector, and covariance matrix

$$\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 1 & \rho_j & 0 & \ldots & \ldots & 0 \\ \rho_j & 1 & \rho_j & 0 & \ldots & 0 \\ 0 & \rho_j & 1 & \rho_j & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \rho_j & 1 \end{pmatrix}_{(n-1)\times(n-1)} . \tag{A3.46}$$

From equation (A3.13), for each $j \geq 1$,

$$\mathrm{E}\left\{\sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1)\right\} = \sum_{i=0}^{n-2} \rho_j = (n-1)\rho_j = \mathrm{Tr}\left(\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star}[\boldsymbol{\eta}_j^{\star\star}]^T\right\}\right), \tag{A3.47}$$

where

$$\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star}[\boldsymbol{\eta}_j^{\star\star}]^T\right\} = \mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\} = \rho_j \boldsymbol{Id}_{n-1}, \tag{A3.48}$$

with, as before, $\boldsymbol{Id}_{n-1}$ denoting the $(n-1) \times (n-1)$ identity matrix.

However, the variance of

$$\sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1)$$

depends greatly on the distribution of $\boldsymbol{\eta}_j^{\star}$ and $\boldsymbol{\eta}_j^{\star\star}$. In the Gaussian case, keeping in mind that

$$\boldsymbol{\eta}_j^{\star} = (\eta_j(0), \ldots, \eta_j(n-2))^T, \quad \boldsymbol{\eta}_j^{\star\star} = (\eta_j(1), \ldots, \eta_j(n-1))^T$$

are zero–mean multivariate Normal distributed vectors with covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ given in (A3.46), and having cross–covariance matrix in (A3.48), we can compute the variance of $\sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1)$, from (A3.47)–(A3.48), as follows. First,

$$\begin{aligned} \mathrm{Var}\left\{[\boldsymbol{\eta}_j^{\star}]^T \boldsymbol{Id}_{n-1}\boldsymbol{\eta}_j^{\star\star}\right\} &= \mathrm{E}\left\{[\boldsymbol{\eta}_j^{\star}]^T \boldsymbol{Id}_{n-1}\boldsymbol{\eta}_j^{\star\star}[\boldsymbol{\eta}_j^{\star}]^T \boldsymbol{Id}_{n-1}\boldsymbol{\eta}_j^{\star\star}\right\} \\ &- \left(\mathrm{E}\left\{[\boldsymbol{\eta}_j^{\star}]^T \boldsymbol{Id}_{n-1}\boldsymbol{\eta}_j^{\star\star}\right\}\right)^2 . \end{aligned}$$

This can be rewritten as

$$\sum_{i=0}^{n-2}\sum_{p=0}^{n-2} \mathrm{E}\left\{\eta_j(i)\eta_j(i+1)\eta_j(p)\eta_j(p+1)\right\} - \left(\mathrm{E}\left\{[\boldsymbol{\eta}_j^{\star}]^T \boldsymbol{Id}_{n-1}\boldsymbol{\eta}_j^{\star\star}\right\}\right)^2,$$

which is equal to

$$\begin{aligned}
\sum_{i=0}^{n-2} \mathrm{E}\left\{\eta_j(i)\eta_j(i+1)\right\} \sum_{p=0}^{n-2} \mathrm{E}\left\{\eta_j(p)\eta_j(p+1)\right\} \quad &+ \quad \sum_{i=0}^{n-2}\sum_{p=0}^{n-2} \mathrm{E}\left\{\eta_j(i)\eta_j(p)\right\} \mathrm{E}\left\{\eta_j(i+1)\eta_j(p+1)\right\} \\
&+ \quad \sum_{i=0}^{n-2}\sum_{p=0}^{n-2} \mathrm{E}\left\{\eta_j(i)\eta_j(p+1)\right\} \mathrm{E}\left\{\eta_j(i+1)\eta_j(p)\right\} \\
&- \quad \left(\mathrm{E}\left\{[\boldsymbol{\eta}_j^{\star}]^T \boldsymbol{Id}_{n-1} \boldsymbol{\eta}_j^{\star\star}\right\}\right)^2.
\end{aligned}$$

This then reduces to

$$\begin{aligned}
\left[\mathrm{Tr}\left(\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\right)\right]^2 \quad &+ \quad \mathrm{Tr}\left(\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}\right) \\
&+ \quad \mathrm{Tr}\left(\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\left[\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\right]^T\right) - \left[\mathrm{Tr}\left(\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\right)\right]^2,
\end{aligned} \tag{A3.49}$$

which is the same as

$$\begin{aligned}
\mathrm{Tr}\left(\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}\right) \quad &+ \quad \mathrm{Tr}\left(\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\left[\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\right]^T\right) \\
&= \quad (n-1) + 2(n-2)\rho_j^2 + (n-1)\rho_j^2,
\end{aligned}$$

where, from (A3.48),

$$\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\left[\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star} \otimes \boldsymbol{\eta}_j^{\star\star}\right\}\right]^T = \begin{pmatrix} \rho_j^2 & 0 & \ldots & \ldots & 0 \\ 0 & \rho_j^2 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ldots & \ddots & \ddots & \rho_j^2 \end{pmatrix} = \rho_j^2 \boldsymbol{Id}_{n-1}.$$

From (A3.49),

$$\mathrm{Var}\left\{\frac{\sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1)}{n-1}\right\} = \frac{(n-1) + 2(n-2)\rho_j^2 + (n-1)\rho_j^2}{(n-1)^2}. \tag{A3.50}$$

Therefore, for each $j \geq 1$,

$$\lim_{n\to\infty} n\mathrm{Var}\left\{\frac{\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)}{n-1}\right\} = 1 + 3\rho_j^2.$$

Thus, for each $j \geq 1$, $K_{j,2}$ in (A3.19) is given by $K_{j,2} = 1 + 3\rho_j^2$. From equation (A3.50),

$$\mathrm{Var}\left\{\frac{\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)}{n-1}\right\} \leq 1 + 3\rho_j^2 \leq 4.$$

Hence, for every $j \geq 1$, $\widetilde{K}_{j,2}$ in equation (A3.21) satisfies

$$\widetilde{K}_{j,2} \leq 4.$$

Therefore, the constant S in **Assumption A4** is such that $S \leq 6 + 4 = 10$.

A3.5 Simulation study

A simulation study is undertaken to illustrate the behaviour of the formulated componentwise estimator of the autocorrelation operator, and of its associated ARH(1) plug–in predictor for large sample sizes. The results are reported in Appendix A3.5.1. In Appendix A3.5.2, a comparative study is developed, from the implementation of the ARH(1) plug–in prediction techniques proposed in Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001]. In the subsequent sections, we restrict our attention to the Gaussian case

A3.5.1 Behaviour of $\widehat{\rho}$ and $\widehat{X}_n$ for large sample sizes

Let $(-\Delta)_{(a,b)}$ be the Dirichlet negative Laplacian operator on (a, b) given by

$$\begin{aligned}
(-\Delta)_{(a,b)}(f)(x) &= -\frac{d^2}{dx^2}f(x), \quad x \in (a,b) \subset \mathbb{R}, \\
f(a) &= f(b) = 0.
\end{aligned}$$

The eigenvectors $\{\phi_j,\ j \geq 1\}$ and eigenvalues $\left\{\lambda_j\left((-\Delta)_{(a,b)}\right),\ j \geq 1\right\}$ of $(-\Delta)_{(a,b)}$ satisfy, for each

$j \geq 1$ and for each $x \in (a, b)$,

$$(-\Delta)_{(a,b)}\phi_j(x) = \lambda_j\left((-\Delta)_{(a,b)}\right)\phi_j(x), \quad \phi_j(a) = \phi_j(b) = 0. \tag{A3.51}$$

For each $j \geq 1$ and $x \in [a, b]$, the solution to equation (A3.51) is given by (see [Grebenkov and Nguyen, 2013, p. 6]):

$$\phi_j(x) = \sqrt{\frac{2}{b-a}}\sin\left(\frac{\pi j x}{b-a}\right), \quad \forall x \in [a, b], \quad \lambda_j\left((-\Delta)_{(a,b)}\right) = \frac{\pi^2 j^2}{(b-a)^2}. \tag{A3.52}$$

We consider here the operator C defined as

$$C = \left((-\Delta)_{(a,b)}\right)^{-2(1-\gamma_1)}, \quad \gamma_1 \in (0, 1/2).$$

From [Dautray and Lions, 1990, pp. 119–140], the eigenvectors of C coincide with the eigenvectors of $(-\Delta)_{(a,b)}$, and its eigenvalues $\{C_j,\ j \geq 1\}$ are given by:

$$C_j = \left[\lambda_j\left((-\Delta)_{(a,b)}\right)\right]^{-2(1-\gamma_1)} = \left[\frac{\pi^2 j^2}{(b-a)^2}\right]^{-2(1-\gamma_1)}. \tag{A3.53}$$

Additionally, considering

$$\rho = \left[\frac{(-\Delta)_{(a,b)}}{\lambda_1\left((-\Delta)_{(a,b)}\right) - \epsilon}\right]^{-(1-\gamma_2)}, \quad \gamma_2 \in (0, 1/2),$$

for certain positive constant $\epsilon < \lambda_1\left((-\Delta)_{(a,b)}\right)$ close to zero, ρ is a positive self–adjoint Hilbert–Schmidt operator, whose eigenvectors coincide with the eigenvectors of $(-\Delta)_{(a,b)}$, and whose eigenvalues $\{\rho_j,\ j \geq 1\}$ are such that $\rho_j < 1$, for every $j \geq 1$, and

$$\rho_j^2 = \left[\frac{\lambda_j\left((-\Delta)_{(a,b)}\right)}{\lambda_1\left((-\Delta)_{(a,b)}\right) - \epsilon}\right]^{-2(1-\gamma_2)}, \quad \rho_j^2 \in (0, 1), \quad \gamma_2 \in (0, 1/2), \tag{A3.54}$$

where, as before, $\left\{\lambda_j\left((-\Delta)_{(a,b)}\right),\ j \geq 1\right\}$ are given in equation (A3.52).

From (A3.12), the eigenvalues $\left\{\sigma_j^2,\ j \geq 1\right\}$ of C_ε are then defined, for each $j \geq 1$, as

$$\sigma_j^2 = C_j\left(1 - \rho_j^2\right) = \left[\lambda_j\left((-\Delta)_{(a,b)}\right)\right]^{-2(1-\gamma_1)} - \frac{\left[\lambda_j\left((-\Delta)_{(a,b)}\right)\right]^{-2(2-\gamma_1-\gamma_2)}}{\left[\lambda_1\left((-\Delta)_{(a,b)}\right) - \epsilon\right]^{-2(1-\gamma_2)}}.$$

Note that C_ε is in the trace class, since the trace property of C, and the fact that $\rho_j^2 < 1$, for every $j \geq 1$, implies

$$\sum_{j=1}^{\infty}\sigma_j^2 = \sum_{j=1}^{\infty} C_j\left(1 - \rho_j^2\right) < \sum_{j=1}^{\infty} C_j < \infty.$$

For this particular example of operator C, we have considered truncation parameter k_n of the form

$$k_n = n^{1/\alpha}, \tag{A3.55}$$

for a suitable $\alpha > 0$, which, in particular, allows verification of (A3.17). From equation (A3.53), one has, for $\gamma_1 \in (0, 1/2)$,

$$\sqrt{n} C_{k_n} = \sqrt{n} \left[\lambda_{k_n} \left(-\Delta_{(a,b)}\right)\right]^{-2(1-\delta_1)} = \sqrt{n} \left(\frac{\pi k_n}{b-a}\right)^{-4(1-\delta_1)}, \quad \delta_1 > 1.$$

From equation (A3.55), **Assumption A3** is then satisfied if

$$1/2 - \frac{4(1-\gamma_1)}{\alpha} > 0, \quad \text{i.e., if} \quad \alpha > 8(1-\gamma_1) > 4. \tag{A3.56}$$

since $\gamma_1 \in (0, 1/2)$. Fix $\gamma_1 = 0.4$ and $\gamma_2 = 9/20$. Then, from equation (A3.56), $\alpha > 48/10$. In particular, the values $\alpha_1 = 5$ and $\alpha_2 = 6$ have been tested, in Table A3.5.1 below, for $H = L^2((a,b))$, and $(a,b) = (0,4)$, where $L^2((a,b))$ denotes the space of square integrable functions on (a,b).

The computed empirical truncated functional mean square error $\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$ of the estimator $\widehat{\rho}_{k_n}$ of ρ, for a sample size n, is given by:

$$\mathrm{EMSE}_{\widehat{\rho}_{k_n}} = \frac{1}{N} \sum_{w=1}^{N} \sum_{j=1}^{k_n} \left(\rho_j - \widehat{\rho}_{n,j}^{w}\right)^2, \tag{A3.57}$$

$$\widehat{\rho}_{n,j}^{w} = \frac{\widehat{D}_{n,j}^{w}}{\widehat{C}_{n,j}^{w}} = \frac{\frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j}^{w} X_{i+1,j}^{w}}{\frac{1}{n} \sum_{i=0}^{n-1} \left(X_{i,j}^{w}\right)^2}, \tag{A3.58}$$

where N denotes the number of simulations, and for each $j = 1, \dots, k_n$, $\widehat{\rho}_{n,j}^{w}$ represents the estimator of ρ_j, based on the w–th generation of the values $X_{0,j}^{w}, \dots, X_{n-1,j}^{w}$, with $X_{i,j}^{w} = \langle X_i^w, \phi_j \rangle_H$, for $w = 1, \dots, 700$, and $i = 0, \dots, n-1$.

For the plug–in predictor $\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1})$, we compute the empirical version $\mathrm{UB(EMAE)}_{\widehat{X}_n^{k_n}}$ of the derived upper bound (A3.40), which, for each $n \in \mathbb{Z}$, is given by

$$\mathrm{UB(EMAE)}_{\widehat{X}_n^{k_n}} = \sqrt{\frac{1}{N} \sum_{w=1}^{N} \sum_{j=1}^{k_n} \left(\rho_j - \widehat{\rho}_{n,j}^{w}\right)^2 \mathrm{E}\left\{\widehat{\left\|X_{n-1}^{w}\right\|_H^2}\right\}}. \tag{A3.59}$$

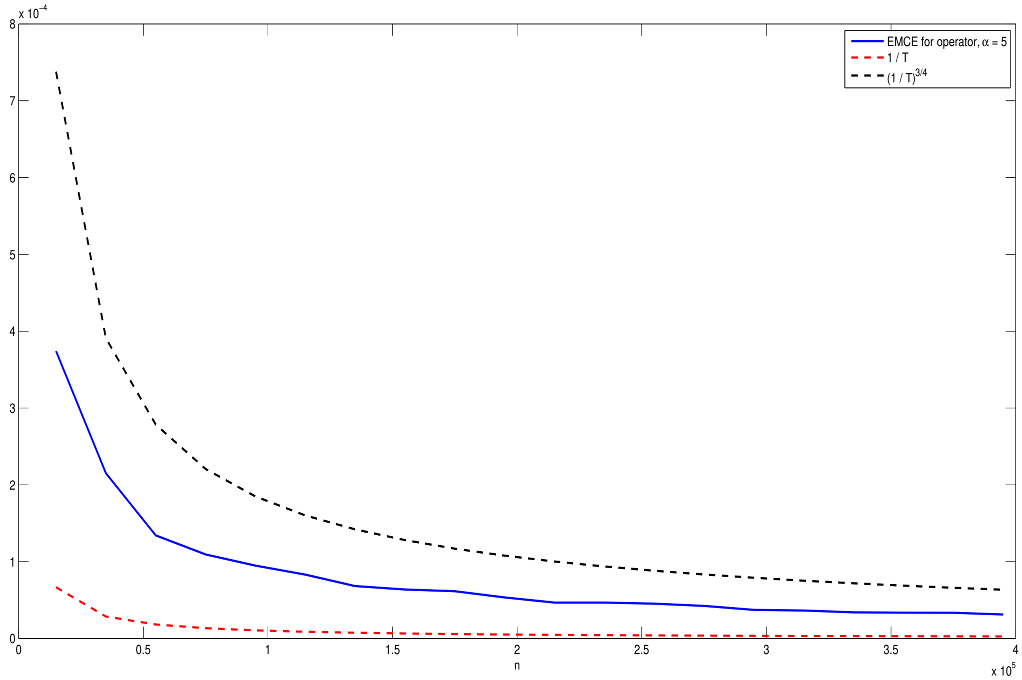
From $N = 700$ realizations, for each one of the elements of the sequence of sample sizes

$$\{n_t,\, t = 1, \dots, 20\} = \{15000 + 20000(t-1),\, t = 1, \dots, 20\},$$

the $\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$ and $\mathrm{UB}(\mathrm{EMAE})_{\widehat{X}_n^{k_n}}$ values, for $\alpha = 5$ and $\alpha = 6$, are displayed in Table A3.5.1, where the abbreviated notations $\mathrm{MSE}_{\widehat{\rho}_{k_{n,1}}}$, for $\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$, and $\mathrm{UB}_{\widehat{X}_n{}^{k_{n,1}}}$, for $\mathrm{UB}(\mathrm{EMAE})_{\widehat{X}_n^{k_n}}$, are used (see also Figures A3.5.1–A3.5.2).

Table A3.5.1: $\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$ (here, $\mathrm{MSE}_{\widehat{\rho}_{k_{n,i}}}$), and $\mathrm{UB}(\mathrm{EMAE})_{\widehat{X}_n^{k_n}}$ (here, $\mathrm{UB}_{\widehat{X}_n{}^{k_{n,i}}}$) values, in (A3.57)–(A3.59), based on $N = 700$ simulations, for $\gamma_1 = 0.4$ and $\gamma_2 = 9/20$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and the corresponding $k_{n,1}$ and $k_{n,2}$ values, for $\alpha_1 = 5$ and $\alpha_2 = 6$.

n	$k_{n,1}$	$\mathrm{MSE}_{\widehat{\rho}_{k_{n,1}}}$	$\mathrm{UB}_{\widehat{X}_n{}^{k_{n,1}}}$	$k_{n,2}$	$\mathrm{MSE}_{\widehat{\rho}_{k_{n,2}}}$	$\mathrm{UB}_{\widehat{X}_n{}^{k_{n,2}}}$
$n_1 = 15000$	6	$3.74\,(10)^{-4}$	$2.87\,(10)^{-2}$	4	$2.45\,(10)^{-4}$	$2.25\,(10)^{-2}$
$n_2 = 35000$	8	$2.15\,(10)^{-4}$	$2.21\,(10)^{-2}$	5	$1.35\,(10)^{-4}$	$1.71\,(10)^{-2}$
$n_3 = 55000$	8	$1.34\,(10)^{-4}$	$1.75\,(10)^{-2}$	6	$1.03\,(10)^{-4}$	$1.51\,(10)^{-2}$
$n_4 = 75000$	9	$1.09\,(10)^{-4}$	$1.57\,(10)^{-2}$	6	$7.55\,(10)^{-5}$	$1.29\,(10)^{-2}$
$n_5 = 95000$	9	$9.48\,(10)^{-5}$	$1.47\,(10)^{-2}$	6	$5.86\,(10)^{-5}$	$1.14\,(10)^{-2}$
$n_6 = 115000$	10	$8.31\,(10)^{-5}$	$1.39\,(10)^{-2}$	6	$5.16\,(10)^{-5}$	$1.07\,(10)^{-2}$
$n_7 = 135000$	10	$6.81\,(10)^{-5}$	$1.25\,(10)^{-2}$	7	$4.86\,(10)^{-5}$	$1.04\,(10)^{-2}$
$n_8 = 155000$	10	$6.37\,(10)^{-5}$	$1.21\,(10)^{-2}$	7	$3.88\,(10)^{-5}$	$9.66\,(10)^{-3}$
$n_9 = 175000$	11	$6.14\,(10)^{-5}$	$1.19\,(10)^{-2}$	7	$3.87\,(10)^{-5}$	$9.65\,(10)^{-3}$
$n_{10} = 195000$	11	$5.34\,(10)^{-5}$	$1.11\,(10)^{-2}$	7	$3.42\,(10)^{-5}$	$8.79\,(10)^{-3}$
$n_{11} = 215000$	11	$4.67\,(10)^{-5}$	$1.03\,(10)^{-2}$	7	$3.40\,(10)^{-5}$	$8.74\,(10)^{-3}$
$n_{12} = 235000$	11	$4.66\,(10)^{-5}$	$1.03\,(10)^{-2}$	7	$2.92\,(10)^{-5}$	$8.12\,(10)^{-3}$
$n_{13} = 255000$	12	$4.53\,(10)^{-5}$	$1.02\,(10)^{-2}$	7	$2.77\,(10)^{-5}$	$7.95\,(10)^{-3}$
$n_{14} = 275000$	12	$4.24\,(10)^{-5}$	$9.95\,(10)^{-3}$	8	$2.77\,(10)^{-5}$	$7.94\,(10)^{-3}$
$n_{15} = 295000$	12	$3.72\,(10)^{-5}$	$9.32\,(10)^{-3}$	8	$2.67\,(10)^{-5}$	$7.76\,(10)^{-3}$
$n_{16} = 315000$	12	$3.62\,(10)^{-5}$	$9.21\,(10)^{-3}$	8	$2.55\,(10)^{-5}$	$7.64\,(10)^{-3}$
$n_{17} = 335000$	12	$3.39\,(10)^{-5}$	$8.91\,(10)^{-3}$	8	$2.28\,(10)^{-5}$	$7.04\,(10)^{-3}$
$n_{18} = 355000$	12	$3.34\,(10)^{-5}$	$8.86\,(10)^{-3}$	8	$2.20\,(10)^{-5}$	$7.04\,(10)^{-3}$
$n_{19} = 375000$	13	$3.34\,(10)^{-5}$	$8.86\,(10)^{-3}$	8	$2.04\,(10)^{-5}$	$6.84\,(10)^{-3}$
$n_{20} = 395000$	13	$3.12\,(10)^{-5}$	$8.56\,(10)^{-3}$	8	$1.92\,(10)^{-5}$	$6.65\,(10)^{-3}$



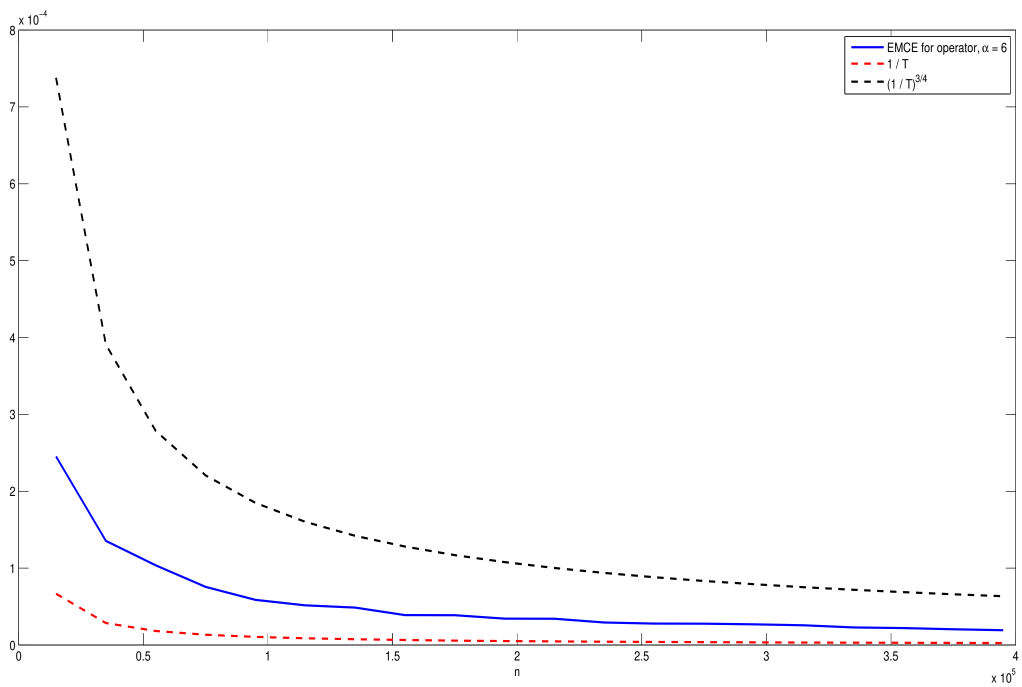
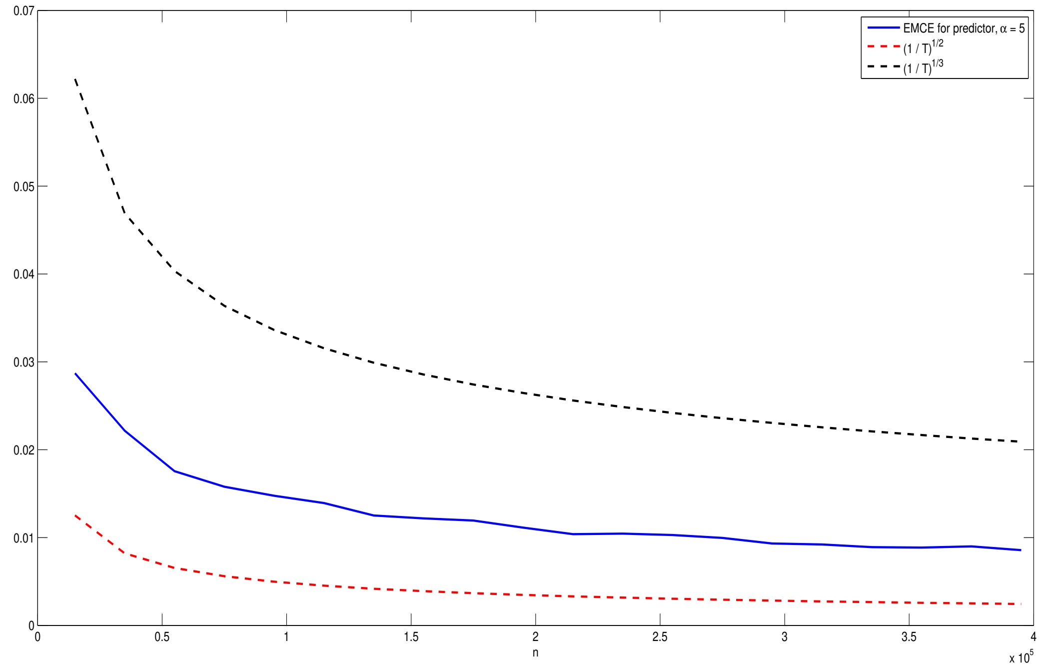


Figure A3.5.1: $\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$ values (blue line), in (A3.57)–(A3.58), based on $N = 700$ simulations, for $\gamma_1 = 0.4$ and $\gamma_2 = 9/20$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and the corresponding $k_{n,1}$ and $k_{n,2}$ values, for $\alpha_1 = 5$ (left-hand side) and $\alpha_2 = 6$ (right-hand side), against curves $(1/n_t)^{3/4}$ (black dot line) and $1/n_t$ (red dot line).



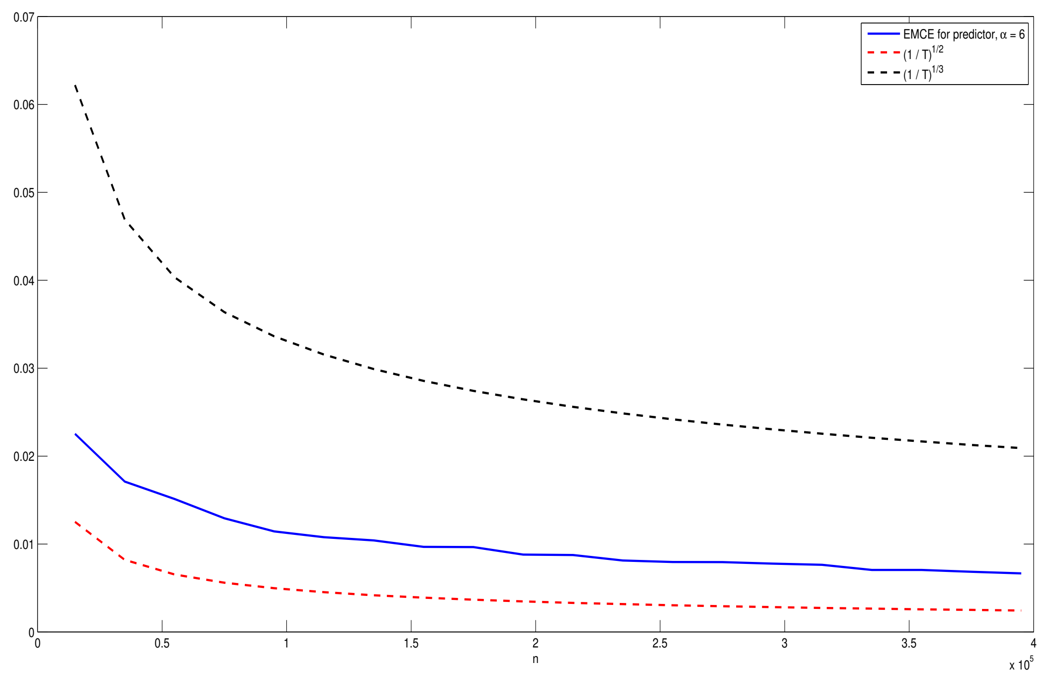


Figure A3.5.2: $\mathrm{UB}(\mathrm{EMAE})_{\widehat{X}_n^{k_n}}$ values (blue line), in (A3.59), based on $N = 700$ simulations, for $\gamma_1 = 0.4$ and $\gamma_2 = 9/20$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and the corresponding $k_{n,1}$ and $k_{n,2}$ values, for $\alpha_1 = 5$ (left-hand side) and $\alpha_2 = 6$ (right-hand side), against curves $(1/n_t)^{1/2}$ (red dot line) and $(1/n_t)^{1/3}$ (black dot line).

In this paper, a one–parameter model of k_n is selected depending on parameter α. In [Guillas, 2001, Example 2], in the same spirit, for an equivalent spectral class of operators C, a three–parameter model is established for k_n to ensure convergence in quadratic mean in the space $\mathcal{L}(H)$ of the componentwise estimator of ρ constructed from the known eigenvectors of C. The numerical results displayed in Table A3.5.1 and Figures A3.5.1–A3.5.2 illustrate the fact that the proposed componentwise estimator $\widehat{\rho}_{k_n}$ presents a speed of convergence to ρ, in quadratic mean in $S(H)$, faster than $n^{-1/3}$, which corresponds to the optimal case for the componentwise estimator of ρ proposed in Guillas [2001], in the case of known eigenvectors of C; see, in particular, [Guillas, 2001, Theorem 1, Remark 2 and Example 2]. For larger values of the parameters γ_1 than 2.4, and α than 6, a faster velocity of convergence of $\widehat{\rho}_{k_n}$ to ρ, in quadratic mean in the space $S(H)$, will be obtained. However, larger sample sizes are required for larger values of α, in order to estimate a given number of coefficients of ρ. A more detailed discussion about comparison of the rates of convergence of the ARH(1) plug–in predictors proposed in Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001] can be found in the next section.

A3.5.2 A comparative study

In this section, the performance of our approach is compared with those ones given in Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001], including the case of unknown eigenvectors of C. In the last case, our approach and the approaches presented in Bosq [2000]; Guillas [2001] are implemented in terms of the empirical eigenvectors.

A3.5.2.1 Theoretical–eigenvector–based componentwise estimators

Let us first compare the performance of our ARH(1) plug–in predictor, defined in (A3.38), and the ones formulated in Bosq [2000]; Guillas [2001], in terms of the theoretical eigenvectors $\{\phi_j,\ j \geq 1\}$ of C. Note that, in this first part of our comparative study, we consider the previous generated Gaussian ARH(1) process, with autocovariance and autocorrelation operators defined from equations (A3.53) and (A3.54), for different rates of convergence to zero of parameters C_j and $\rho_j^2, j \geq 1$, with both sequences being summable sequences. Since we restrict our attention to the Gaussian case, conditions A_1, B_1 and C_1, formulated in [Bosq, 2000, pp. 211–212] are satisfied by the generated ARH(1) process. Similarly, Conditions H_1–H_3 in [Guillas, 2001, p. 283] are satisfied as well.

In [Bosq, 2000, Section 8.2] the following estimator of ρ is proposed

$$\widehat{\rho}_n(x) = \left(\Pi^{k_n} D_n \widehat{C}_n^{-1} \Pi^{k_n}\right)(x) = \sum_{l=1}^{k_n} \widehat{\rho}_{n,l}(x)\phi_l, \quad x \in H, \tag{A3.60}$$

$$\widehat{\rho}_{n,l}(x) = \frac{1}{n-1} \sum_{i=0}^{n-2} \sum_{j=1}^{k_n} \frac{1}{\widehat{C}_{n,j}} \langle \phi_j, x \rangle_H X_{i,j} X_{i+1,l}, \tag{A3.61}$$

in the finite dimensional subspace

$$H_{k_n} = \text{span}\,(\phi_1, \ldots, \phi_{k_n})$$

of H, where Π^{k_n} is the orthogonal projector over H_{k_n}, and, as before, $X_{i,j} = \langle X_i, \phi_j \rangle_H$, for $j \geq 1$.

A modified estimator of ρ is studied in [Guillas, 2001, Section 2], given by

$$\widehat{\rho}_{n,a}(x) = \left(\Pi^{k_n} D_n \widehat{C}_{n,a}^{-1} \Pi^{k_n}\right)(x) = \sum_{l=1}^{k_n} \widehat{\rho}_{n,a,l}(x)\phi_l, \quad x \in H, \tag{A3.62}$$

$$\widehat{\rho}_{n,a,l}(x) = \frac{1}{n-1}\sum_{i=1}^{n-1}\sum_{j=1}^{k_n} \frac{1}{\max\left(\widehat{C}_{n,j},\, a_n\right)} \langle \phi_j, x\rangle_H X_{i,j} X_{i+1,l}, \tag{A3.63}$$

where

$$\widehat{C}_{n,a}^{-1}(x) = \sum_{j=1}^{k_n} \frac{1}{\max\left(\widehat{C}_{n,j},\, a_n\right)} \langle \phi_j, x\rangle_H \phi_j \; a.s.$$

Here, $\{a_n,\ n \in \mathbb{N}\}$ is such that (see [Guillas, 2001, Theorem 1])

$$\alpha \frac{C_{k_n}^{\gamma}}{n^{\varepsilon}} \leq a_n \leq \beta \lambda_{k_n}, \quad \alpha > 0, \quad 0 < \beta < 1, \quad \varepsilon < 1/2, \quad \gamma \geq 1.$$

Tables A3.5.2–A3.5.3 display the truncated, for two different k_n rules, empirical values of $\mathrm{E}\left\{\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}(X_{n-1})\|_H\right\}$, based on $N = 700$ generations of each one of the functional samples considered with sizes $n_t = 15000 + 20000(t-1)$, $t = 1, \dots, 20$, when

$$C_j = b_C j^{-\delta_1}, \quad b_C > 0, \quad \rho_j^2 = b_\rho j^{-\delta_2}, \quad b_\rho > 0.$$

Specifically, $\widehat{\rho}_{k_n}$ is computed from equations (A3.15)–(A3.16) (see third column), $\widehat{\rho}_{k_n} = \widehat{\rho}_n$, with $\widehat{\rho}_n$ being given in equations (A3.60)–(A3.61) (see fourth column), and $\widehat{\rho}_{k_n} = \widehat{\rho}_{n,a}$, with $\widehat{\rho}_{n,a}$ being defined in (A3.62)–(A3.63) (see fifth column).

In Table A3.5.2, $\delta_1 = 2.4$ $\delta_2 = 1.1$, and $k_n = \lceil n^{1/\alpha} \rceil$, for $\alpha = 6$, according to our **Assumption A3**, which is also considered in [Bosq, 2000, p. 217] to ensure weak consistency of the proposed estimator of ρ. In Table A3.5.3, the same empirical values are displayed for $\delta_1 = \frac{61}{60}$, $\delta_2 = 1.1$, and k_n is selected according to [Guillas, 2001, Example 2]. Thus, in Table A3.5.3,

$$k_n = \lceil n^{\frac{1-2\epsilon}{\delta_1(4+2\gamma)}} \rceil, \quad \gamma \geq 1,\ \epsilon < 1/2. \tag{A3.64}$$

In particular we have chosen $\gamma = 2$, and $\epsilon = 0.04\delta_1$. Note that, from [Guillas, 2001, Theorem 1 and Remark 1], for the choice made of k_n in Table A3.5.3, convergence to ρ, in quadratic mean in the space $\mathcal{L}(H)$, holds for $\widehat{\rho}_{n,a}$ given in (A3.62)–(A3.63).

Table A3.5.2: Truncated empirical values of $\mathrm{E}\|\rho\,(X_{n-1}) - \widehat{\rho}_{k_n}(X_{n-1})\|_H$, for $\widehat{\rho}_{k_n}$ given in equations (A3.15)-(A3.16) (third column), in equations (A3.60)–(A3.61) (fourth column), and in equations (A3.62)–(A3.63) (fifth column), based on $N = 700$ simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \ldots, 20\}$ and the corresponding $k_n = \lceil n^{1/\alpha} \rceil$ values, for $\alpha = 6$.

n	k_n	Our Approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	4	$2.25\,(10)^{-2}$	$2.57\,(10)^{-2}$	$2.36\,(10)^{-2}$
$n_2 = 35000$	5	$1.71\,(10)^{-2}$	$1.72\,(10)^{-2}$	$1.84\,(10)^{-2}$
$n_3 = 55000$	6	$1.51\,(10)^{-2}$	$1.65\,(10)^{-2}$	$1.53\,(10)^{-2}$
$n_4 = 75000$	6	$1.29\,(10)^{-2}$	$1.46\,(10)^{-2}$	$1.37\,(10)^{-2}$
$n_5 = 95000$	6	$1.14\,(10)^{-2}$	$1.20\,(10)^{-2}$	$1.16\,(10)^{-2}$
$n_6 = 115000$	6	$1.07\,(10)^{-2}$	$1.10\,(10)^{-2}$	$1.11\,(10)^{-2}$
$n_7 = 135000$	7	$1.04\,(10)^{-2}$	$1.06\,(10)^{-2}$	$1.07\,(10)^{-2}$
$n_8 = 155000$	7	$9.66\,(10)^{-3}$	$9.91\,(10)^{-3}$	$1.01\,(10)^{-2}$
$n_9 = 175000$	7	$9.65\,(10)^{-3}$	$9.79\,(10)^{-3}$	$9.68\,(10)^{-3}$
$n_{10} = 195000$	7	$8.79\,(10)^{-3}$	$9.12\,(10)^{-3}$	$8.93\,(10)^{-3}$
$n_{11} = 215000$	7	$8.74\,(10)^{-3}$	$8.79\,(10)^{-3}$	$8.83\,(10)^{-3}$
$n_{12} = 235000$	7	$8.12\,(10)^{-3}$	$8.69\,(10)^{-3}$	$8.75\,(10)^{-3}$
$n_{13} = 255000$	7	$7.95\,(10)^{-3}$	$8.53\,(10)^{-3}$	$8.73\,(10)^{-3}$
$n_{14} = 275000$	8	$7.94\,(10)^{-3}$	$8.52\,(10)^{-3}$	$8.58\,(10)^{-3}$
$n_{15} = 295000$	8	$7.76\,(10)^{-3}$	$8.49\,(10)^{-3}$	$8.36\,(10)^{-3}$
$n_{16} = 315000$	8	$7.64\,(10)^{-3}$	$7.88\,(10)^{-3}$	$8.13\,(10)^{-3}$
$n_{17} = 335000$	8	$7.04\,(10)^{-3}$	$7.24\,(10)^{-3}$	$7.59\,(10)^{-3}$
$n_{18} = 355000$	8	$7.04\,(10)^{-3}$	$7.23\,(10)^{-3}$	$6.92\,(10)^{-3}$
$n_{19} = 375000$	8	$6.84\,(10)^{-3}$	$6.89\,(10)^{-3}$	$6.90\,(10)^{-3}$
$n_{20} = 395000$	8	$6.65\,(10)^{-3}$	$6.67\,(10)^{-3}$	$6.85\,(10)^{-3}$

Table A3.5.3: Truncated empirical values of $\mathrm{E}\|\rho(X_{n-1}) - \widehat{\rho}_{k_n}(X_{n-1})\|_H$, for $\widehat{\rho}_{k_n}$ given in equations (A3.15)–(A3.16) (third column), in equations (A3.60)–(A3.61) (fourth column), and in equations (A3.62)–(A3.63) (fifth column), based on $N = 700$ simulations, for $\delta_1 = \frac{61}{60}$ and $\delta_2 = 1.1$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and the corresponding k_n given in (A3.64).

n	k_n	Our Approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	2	$9.91\,(10)^{-3}$	$1.39\,(10)^{-2}$	$1.26\,(10)^{-2}$
$n_2 = 35000$	3	$8.78\,(10)^{-3}$	$1.34\,(10)^{-2}$	$1.24\,(10)^{-2}$
$n_3 = 55000$	3	$7.89\,(10)^{-3}$	$1.15\,(10)^{-2}$	$1.14\,(10)^{-2}$
$n_4 = 75000$	3	$6.49\,(10)^{-3}$	$1.01\,(10)^{-2}$	$8.58\,(10)^{-3}$
$n_5 = 95000$	3	$6.36\,(10)^{-3}$	$9.09\,(10)^{-3}$	$8.29\,(10)^{-3}$
$n_6 = 115000$	3	$6.14\,(10)^{-3}$	$7.65\,(10)^{-3}$	$7.26\,(10)^{-3}$
$n_7 = 135000$	3	$5.91\,(10)^{-3}$	$7.03\,(10)^{-3}$	$6.69\,(10)^{-3}$
$n_8 = 155000$	3	$5.73\,(10)^{-3}$	$6.77\,(10)^{-3}$	$6.54\,(10)^{-3}$
$n_9 = 175000$	3	$5.44\,(10)^{-3}$	$6.74\,(10)^{-3}$	$6.16\,(10)^{-3}$
$n_{10} = 195000$	3	$5.10\,(10)^{-3}$	$6.69\,(10)^{-3}$	$5.97\,(10)^{-3}$
$n_{11} = 215000$	4	$5.01\,(10)^{-3}$	$6.48\,(10)^{-3}$	$5.94\,(10)^{-3}$
$n_{12} = 235000$	4	$4.85\,(10)^{-3}$	$6.45\,(10)^{-3}$	$5.83\,(10)^{-3}$
$n_{13} = 255000$	4	$4.17\,(10)^{-3}$	$6.17\,(10)^{-3}$	$5.68\,(10)^{-3}$
$n_{14} = 275000$	4	$4.64\,(10)^{-3}$	$5.99\,(10)^{-3}$	$5.60\,(10)^{-3}$
$n_{15} = 295000$	4	$4.55\,(10)^{-3}$	$5.94\,(10)^{-3}$	$5.58\,(10)^{-3}$
$n_{16} = 315000$	4	$4.48\,(10)^{-3}$	$5.69\,(10)^{-3}$	$5.50\,(10)^{-3}$
$n_{17} = 335000$	4	$4.38\,(10)^{-3}$	$5.58\,(10)^{-3}$	$5.44\,(10)^{-3}$
$n_{18} = 355000$	4	$4.16\,(10)^{-3}$	$5.45\,(10)^{-3}$	$5.42\,(10)^{-3}$
$n_{19} = 375000$	4	$3.91\,(10)^{-3}$	$5.34\,(10)^{-3}$	$5.32\,(10)^{-3}$
$n_{20} = 395000$	4	$3.86\,(10)^{-3}$	$5.29\,(10)^{-3}$	$5.26\,(10)^{-3}$

One can observe in Table A3.5.2 a similar performance of the three methods compared with the truncation order kn satisfying **Assumption A3**, with slightly worse results being obtained from the estimator defined in (A3.62)–(A3.63), specially, for the sample size $n_8 = 155000$. Furthermore, in Table A3.5.3, a better performance of our approach is observed for the smallest sample sizes (from $n_1 = 15000$ until $n_4 = 75000$). For the remaining largest sample sizes, only slight differences are observed, with, again, a better performance of our approach, very close to the other two approaches presented in Bosq [2000]; Guillas [2001].

A3.5.2.2 Empirical–eigenvector–based componentwise estimators

In this section, we address the case where $\{\phi_j,\ j \geq 1\}$ are unknown, as is often the case in practice. Specifically, for a given sample size n, let $\{\phi_{n,j},\ j \geq 1\}$ be the empirical counterpart of the theoretical eigenvectors $\{\phi_j,\ j \geq 1\}$, satisfying, for every $j \geq 1$,

$$C_n(\phi_{n,j}) = \frac{1}{n}\sum_{i=0}^{n-1}\langle X_i, \phi_{n,j}\rangle_H X_i = C_{n,j}\phi_{n,j},$$

where $\{C_{n,j},\ j \geq 1\}$ denotes the system of eigenvalues associated with the system of empirical eigenvectors $\{\phi_{n,j},\ j \geq 1\}$. We then consider the following estimators for comparison purposes

$$\widetilde{\rho}_{n,j} = \frac{\frac{1}{n-1}\sum_{i=0}^{n-2}\widetilde{X}_{i,j}\widetilde{X}_{i+1,j}}{\frac{1}{n}\sum_{i=0}^{n-1}\left(\widetilde{X}_{i,j}\right)^2}, \quad \widetilde{\rho}_{k_n} = \sum_{j=1}^{k_n}\widetilde{\rho}_{n,j}\phi_{n,j}\otimes\phi_{n,j}, \tag{A3.65}$$

$$\widetilde{\rho}_n(x) = \left(\widetilde{\Pi}^{k_n}D_nC_n^{-1}\widetilde{\Pi}^{k_n}\right)(x) = \sum_{l=1}^{k_n}\widetilde{\rho}_{n,l}(x)\phi_{n,l}, \quad x \in H,$$

$$\widetilde{\rho}_{n,l}(x) = \frac{1}{n-1}\sum_{i=0}^{n-2}\sum_{j=1}^{k_n}\frac{1}{C_{n,j}}\langle\phi_{n,j}, x\rangle_H\widetilde{X}_{i,j}\widetilde{X}_{i+1,l}, \tag{A3.66}$$

$$\widetilde{\rho}_{n,a}(x) = \left(\widetilde{\Pi}^{k_n}D_nC_{n,a}^{-1}\widetilde{\Pi}^{k_n}\right)(x) = \sum_{l=1}^{k_n}\widetilde{\rho}_{n,a,l}(x)\phi_{n,l}, \quad x \in H,$$

$$\widetilde{\rho}_{n,a,l}(x) = \frac{1}{n-1}\sum_{i=0}^{n-2}\sum_{j=1}^{k_n}\frac{1}{\max(C_{n,j}, a_n)}\langle\phi_{n,j}, x\rangle_H\widetilde{X}_{i,j}\widetilde{X}_{i+1,l}, \tag{A3.67}$$

where, for $i \in \mathbb{Z}$, and $j \geq 1$, $\widetilde{X}_{i,j} = \langle X_i, \phi_{n,j}\rangle_H$, $\widetilde{\Pi}^{k_n}$ denotes the orthogonal projector into the space

$$\widetilde{H}_{k_n} = \operatorname{span}(\phi_{n,1}, \ldots, \phi_{n,k_n}).$$

The Gaussian ARH(1) process is generated under **Assumptions A1–A2**, as well as C_1' in [Bosq, 2000, p. 218]. Note that conditions A_1 and B_1' in Bosq [2000] already hold. Moreover, as given in [Bosq, 2000, Theorem 8.8 and Example 8.6], for

$$C_j = b_C j^{-\delta_1}, \quad b_C > 0, \quad \delta_1 > 0,$$

with, in particular, $\delta_1 = 2.4$, and for

$$\rho_j = b_\rho j^{-\delta_2}, \quad b_\rho > 0,$$

with $\delta_2 = 1.1$,, the estimator $\widetilde{\rho}_n$ converges almost surely to ρ under the condition

$$\frac{nC_{kn}^2}{\ln(n)\left(\sum_{j=1}^{k_n} b_j\right)^2} \longrightarrow \infty,$$

where

$$b_1 = 2\sqrt{2}\,(C_1 - C_2)^{-1}, \quad b_j = 2\sqrt{2}\max\left\{(C_{j-1} - C_j)^{-1}, (C_j - C_{j+1})^{-1}\right\}, \; j \geq 2.$$

In Table A3.5.4, $k_n = \lceil \ln(n) \rceil$ has been tested; see [Bosq, 2000, Example 8.6].

Table A3.5.4: Truncated empirical values of $\mathrm{E}\{\|\rho(X_{n-1}) - \widetilde{\rho}_{k_n}(X_{n-1})\|_H\}$, for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_{k_n}$ given in equation (A3.65) (third column), $\widetilde{\rho}_{k_n} = \widetilde{\rho}_n$ defined in equation (A3.66) (fourth column) and $\widetilde{\rho}_{k_n} = \widetilde{\rho}_{n,a}$ defined in equation (A3.67) (fifth column), based on $N = 700$ simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and $k_n = \lceil \ln(n) \rceil$.

n	k_n	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	9	$8.42\,(10)^{-2}$	1.061	1.035
$n_2 = 35000$	10	$5.51\,(10)^{-2}$	1.019	1.005
$n_3 = 55000$	10	$4.75\,(10)^{-2}$	1.017	0.999
$n_4 = 75000$	11	$4.43\,(10)^{-2}$	1.015	0.995
$n_5 = 95000$	11	$3.68\,(10)^{-2}$	1.013	0.988
$n_6 = 115000$	11	$3.51\,(10)^{-2}$	1.011	0.963
$n_7 = 135000$	11	$3.23\,(10)^{-2}$	1.008	0.925
$n_8 = 155000$	11	$2.95\,(10)^{-2}$	1.007	0.912
$n_9 = 175000$	12	$2.94\,(10)^{-2}$	1.006	0.911
$n_{10} = 195000$	12	$2.80\,(10)^{-2}$	0.995	0.891
$n_{11} = 215000$	12	$2.71\,(10)^{-2}$	0.902	0.862
$n_{12} = 235000$	12	$2.59\,(10)^{-2}$	0.890	0.820
$n_{13} = 255000$	12	$2.58\,(10)^{-2}$	0.878	0.800
$n_{14} = 275000$	12	$2.35\,(10)^{-2}$	0.872	0.783
$n_{15} = 295000$	12	$2.28\,(10)^{-2}$	0.860	0.778
$n_{16} = 315000$	12	$2.27\,(10)^{-2}$	0.842	0.747
$n_{17} = 335000$	12	$2.16\,(10)^{-2}$	0.822	0.714
$n_{18} = 355000$	12	$2.14\,(10)^{-2}$	0.800	0.707
$n_{19} = 375000$	12	$2.09\,(10)^{-2}$	0.778	0.687
$n_{20} = 395000$	12	$2.06\,(10)^{-2}$	0.769	0.662

A better performance of our estimator (A3.65) in comparison with estimator (A3.66), formulated in Bosq [2000], and estimator (A3.67), formulated in [Guillas, 2001, Example 4 and Remark 4], is observed in Table A3.5.4. Note that, in particular, in [Guillas, 2001, Example 4 and Remark 4], smaller values of k_n than $\ln(n)$ are required for a given sample size n, to ensure convergence in quadratic mean, and, in particular, weak–consistency. However, considering a smaller discretization step size $\Delta t = 0.015$ than in Table A3.5.4, where $\Delta t = 0.08$, and for $k_n = \lceil n^{1/6} \rceil$, (i.e., $\alpha = 6$), we obtain in Table A3.5.5, for the same parameter values $\delta_1 = 2.4$ and $\delta_2 = 1.1$, better results than in Table A3.5.4, since a smaller number of coefficients of ρ (parameters) to be estimated is considered in Table A3.5.5, from a richer sample information (coming from the smaller discretization step size considered). One can also observe in Table A3.5.5 a similar performance of the three approaches studied. In Table A3.5.6, the value $k_n = \lceil e' n^{1/(8\delta_1+2)} \rceil$, with $e' = \frac{17}{10}$ proposed in [Guillas, 2001, Example 4 and Remark 4] is considered to compute the truncated empirical values of $\mathrm{E}\left\{ \|\rho(X_{n-1}) - \widetilde{\rho}_{k_n}(X_{n-1})\|_H \right\}$, for $\widetilde{\rho}_{k_n}$ defined in equation (A3.65) (third column), for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_n$ given in equation (A3.66) (fourth column), and for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_{n,a}$ in equation (A3.67) (fifth column). A similar performance of the three approaches is observed, with the exception of $n_{20} = 395000$, where the approach presented in Guillas [2001] displays a slightly better performance

Table A3.5.5: Truncated empirical values of $\mathrm{E}\left\{ \|\rho(X_{n-1}) - \widetilde{\rho}_{k_n}(X_{n-1})\|_H \right\}$, for $\widetilde{\rho}_{k_n}$ defined in equation (A3.65) (third column), for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_n$ given in equation (A3.66) (fourth column), and for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_{n,a}$ in equation (A3.67) (fifth column), based on $N = 200$ (due to high-dimensionality) simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and $k_n = \lceil n^{1/6} \rceil$.

n	k_n	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	4	$9.88\,(10)^{-2}$	$9.25\,(10)^{-2}$	0.106
$n_2 = 35000$	5	$9.52\,(10)^{-2}$	$9.07\,(10)^{-2}$	$9.86\,(10)^{-2}$
$n_3 = 55000$	6	$9.12\,(10)^{-2}$	$8.92\,(10)^{-2}$	$9.39\,(10)^{-2}$
$n_4 = 75000$	6	$8.48\,(10)^{-2}$	$8.64\,(10)^{-2}$	$8.98\,(10)^{-2}$
$n_5 = 95000$	6	$7.61\,(10)^{-2}$	$8.30\,(10)^{-2}$	$8.46\,(10)^{-2}$
$n_6 = 115000$	6	$7.05\,(10)^{-2}$	$7.96\,(10)^{-2}$	$8.04\,(10)^{-2}$
$n_7 = 135000$	7	$6.99\,(10)^{-2}$	$7.84\,(10)^{-2}$	$7.82\,(10)^{-2}$
$n_8 = 155000$	7	$6.70\,(10)^{-2}$	$7.45\,(10)^{-2}$	$7.40\,(10)^{-2}$
$n_9 = 175000$	7	$6.49\,(10)^{-2}$	$7.03\,(10)^{-2}$	$7.07\,(10)^{-2}$
$n_{10} = 195000$	7	$5.88\,(10)^{-2}$	$6.74\,(10)^{-2}$	$6.80\,(10)^{-2}$
$n_{11} = 215000$	7	$5.63\,(10)^{-2}$	$6.46\,(10)^{-2}$	$6.57\,(10)^{-2}$
$n_{12} = 235000$	7	$5.30\,(10)^{-2}$	$6.28\,(10)^{-2}$	$6.37\,(10)^{-2}$
$n_{13} = 255000$	7	$5.05\,(10)^{-2}$	$6.19\,(10)^{-2}$	$6.24\,(10)^{-2}$
$n_{14} = 275000$	8	$4.88\,(10)^{-2}$	$5.99\,(10)^{-2}$	$6.15\,(10)^{-2}$
$n_{15} = 295000$	8	$4.58\,(10)^{-2}$	$5.74\,(10)^{-2}$	$6.04\,(10)^{-2}$
$n_{16} = 315000$	8	$4.24\,(10)^{-2}$	$5.52\,(10)^{-2}$	$5.93\,(10)^{-2}$
$n_{17} = 335000$	8	$3.86\,(10)^{-2}$	$5.24\,(10)^{-2}$	$5.70\,(10)^{-2}$
$n_{18} = 355000$	8	$3.70\,(10)^{-2}$	$5.02\,(10)^{-2}$	$5.53\,(10)^{-2}$
$n_{19} = 375000$	8	$3.55\,(10)^{-2}$	$4.88\,(10)^{-2}$	$5.36\,(10)^{-2}$
$n_{20} = 395000$	8	$3.46\,(10)^{-2}$	$4.70\,(10)^{-2}$	$5.23\,(10)^{-2}$

Table A3.5.6: Truncated empirical values of $\mathrm{E}\left\{\|\rho\left(X_{n-1}\right)-\widetilde{\rho}_{k_n}\left(X_{n-1}\right)\|_H\right\}$, for $\widetilde{\rho}_{k_n}$ defined in equation (A3.65) (third column), for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_n$ given in equation (A3.66) (fourth column), and for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_{n,a}$ in equation (A3.67) (fifth column), based on $N = 200$ (due to high-dimensionality) simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and $k_n = \lceil e' n^{1/(8\delta_1+2)} \rceil,\ e' = \frac{17}{10}$.

n	k_n	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	2	$6.78\,(10)^{-2}$	$8.77\,(10)^{-2}$	$6.64\,(10)^{-2}$
$n_2 = 35000$	2	$6.72\,(10)^{-2}$	$8.61\,(10)^{-2}$	$6.30\,(10)^{-2}$
$n_3 = 55000$	2	$6.46\,(10)^{-2}$	$8.48\,(10)^{-2}$	$6.17\,(10)^{-2}$
$n_4 = 75000$	2	$6.24\,(10)^{-2}$	$8.20\,(10)^{-2}$	$5.76\,(10)^{-2}$
$n_5 = 95000$	2	$5.42\,(10)^{-2}$	$7.84\,(10)^{-2}$	$5.03\,(10)^{-2}$
$n_6 = 115000$	2	$4.84\,(10)^{-2}$	$7.34\,(10)^{-2}$	$4.56\,(10)^{-2}$
$n_7 = 135000$	2	$4.27\,(10)^{-2}$	$6.95\,(10)^{-2}$	$3.94\,(10)^{-2}$
$n_8 = 155000$	2	$3.64\,(10)^{-2}$	$6.60\,(10)^{-2}$	$3.65\,(10)^{-2}$
$n_9 = 175000$	3	$3.51\,(10)^{-2}$	$6.52\,(10)^{-2}$	$3.42\,(10)^{-2}$
$n_{10} = 195000$	3	$3.38\,(10)^{-2}$	$6.16\,(10)^{-2}$	$3.24\,(10)^{-2}$
$n_{11} = 215000$	3	$3.16\,(10)^{-2}$	$5.78\,(10)^{-2}$	$2.85\,(10)^{-2}$
$n_{12} = 235000$	3	$2.98\,(10)^{-2}$	$5.53\,(10)^{-2}$	$2.60\,(10)^{-2}$
$n_{13} = 255000$	3	$2.83\,(10)^{-2}$	$5.15\,(10)^{-2}$	$2.34\,(10)^{-2}$
$n_{14} = 275000$	3	$2.50\,(10)^{-2}$	$4.85\,(10)^{-2}$	$2.05\,(10)^{-2}$
$n_{15} = 295000$	3	$2.23\,(10)^{-2}$	$4.46\,(10)^{-2}$	$1.83\,(10)^{-2}$
$n_{16} = 315000$	3	$2.15\,(10)^{-2}$	$4.30\,(10)^{-2}$	$1.58\,(10)^{-2}$
$n_{17} = 335000$	3	$2.06\,(10)^{-2}$	$4.14\,(10)^{-2}$	$1.40\,(10)^{-2}$
$n_{18} = 355000$	3	$1.98\,(10)^{-2}$	$3.95\,(10)^{-2}$	$1.24\,(10)^{-2}$
$n_{19} = 375000$	3	$1.89\,(10)^{-2}$	$3.77\,(10)^{-2}$	$1.05\,(10)^{-2}$
$n_{20} = 395000$	3	$1.82\,(10)^{-2}$	$3.70\,(10)^{-2}$	$9.93\,(10)^{-3}$

A3.5.2.3 Kernel–based nonparametric and penalized estimation

In practice, curves are observed in discrete times, and should be approximated by smooth functions. In Besse et al. [2000], the following optimization problem is considered:

$$\widehat{X}_i = argmin \left\| L\widehat{X}_i \right\|_{L^2}^2,\ \widehat{X}_i(t_j) = X_i(t_j), \quad j = 1, \dots, p,\ i = 0, \dots, n-1, \tag{A3.68}$$

where L is a linear differential operator of order d. Our interpolation is computed by Matlab *smoothingspline* method. Non-linear kernel regression is then considered, in terms of the smoothed functional data, solution to (A3.68), as follows:

$$\widehat{X}_n^{h_n} = \widehat{\rho}_{h_n}(X_{n-1}), \quad \widehat{\rho}_{h_n}(x) = \frac{\sum_{i=0}^{n-2} \widehat{X}_{i+1} K\left(\frac{\left\|\widehat{X}_i - x\right\|_{L^2}^2}{h_n}\right)}{\sum_{i=0}^{n-2} K\left(\frac{\left\|\widehat{X}_i - x\right\|_{L^2}^2}{h_n}\right)},$$

where K is the usual Gaussian kernel, and

$$\left\|\widehat{X}_i - x\right\|_{L^2}^2 = \int (\widehat{X}_i(t) - x(t))^2 dt, \quad i = 0, \ldots, n-2.$$

Alternatively, in Besse et al. [2000], prediction, in the context of functional autoregressive processes (FAR(1) processes), under the linear assumption on ρ, which is considered to be a compact operator, with $\|\rho\| < 1$, is also studied, from smooth data $\widehat{X}_1, \ldots, \widehat{X}_n$, solving the optimization problem

$$\min_{\widehat{X}_i \in H_q} \frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{1}{p} \sum_{j=1}^{p} \left(X_i(t_j) - \widehat{X}_i^{q,l}(t_j) \right)^2 + l \left\| D^2 \widehat{X}_i^{q,l} \right\|_{L^2}^2 \right), \tag{A3.69}$$

where l is the smoothing parameter, H_q is the q–dimensional functional subspace spanned by the leading eigenvectors of the autocovariance operator C associated with its largest eigenvalues. Thus, smoothness and rank constraint are considered in the computation of the solution to the optimization problem (A3.69). Such a solution is obtained by means of functional PCA.

The following regularized empirical estimators of C and D are then considered, with inversion of C in the subspace H_q:

$$\widehat{C}_{q,l} = \frac{1}{n} \sum_{i=0}^{n-1} \widehat{X}_i \otimes \widehat{X}_i, \quad \widehat{D}_{q,l} = \frac{1}{n-1} \sum_{i=0}^{n-2} \widehat{X}_i \otimes \widehat{X}_{i+1}.$$

Thus, the regularized estimator of ρ is given by

$$\widehat{\rho}_{q,l} = \widehat{D}_{q,l} \widehat{C}_{q,l}^{-1},$$

and the predictor

$$\widehat{X}_n^{q,l} = \widehat{\rho}_{q,l} X_{n-1}.$$

Due to computational cost limitations, in Table A3.5.7, the following statistics are evaluated to compare the performance of the two above-referred prediction methodologies:

$$EMAE_{\widehat{X}_n}^{h_n} = \frac{1}{p} \sum_{j=1}^{p} \left(X_n(t_j) - \widehat{X}_n^{h_n}(t_j) \right)^2, \tag{A3.70}$$

$$EMAE_{\widehat{X}_n}^{q,l} = \frac{1}{p}\sum_{j=1}^{p}\left(X_n(t_j) - \widehat{X}_n^{q,l}(t_j)\right)^2. \quad \text{(A3.71)}$$

Table A3.5.7: $EMAE_{\widehat{X}_n}^{h_{n,i}}$, $i = 1, 2$, and $EMAE_{\widehat{X}_n}^{q,l}$ values (see (A3.70) and (A3.71), respectively), with $q = 7$, based on $N = 200$ simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering now the sample sizes $\{n_t = 750 + 500(t-1),\ t = 1, \ldots, 13\}$ $h_{n,1} = 0.1$ and $h_{n,2} = 0.3$.

n	$EMAE_{\widehat{X}_n}^{h_{n,1}}$	$EMAE_{\widehat{X}_n}^{h_{n,2}}$	$EMAE_{\widehat{X}_n}^{q,l}$
$n_1 = 750$	$8.57\,(10)^{-2}$	$8.85\,(10)^{-2}$	$8.99\,(10)^{-2}$
$n_2 = 1250$	$7.67\,(10)^{-2}$	$8.43\,(10)^{-2}$	$8.69\,(10)^{-2}$
$n_3 = 1750$	$7.15\,(10)^{-2}$	$7.12\,(10)^{-2}$	$8.05\,(10)^{-2}$
$n_4 = 2250$	$7.09\,(10)^{-2}$	$6.87\,(10)^{-2}$	$7.59\,(10)^{-2}$
$n_5 = 2750$	$6.87\,(10)^{-2}$	$6.67\,(10)^{-2}$	$7.31\,(10)^{-2}$
$n_6 = 3250$	$6.52\,(10)^{-2}$	$5.92\,(10)^{-2}$	$7.28\,(10)^{-2}$
$n_7 = 3750$	$6.20\,(10)^{-2}$	$5.56\,(10)^{-2}$	$7.13\,(10)^{-2}$
$n_8 = 4250$	$6.06\,(10)^{-2}$	$5.32\,(10)^{-2}$	$7.06\,(10)^{-2}$
$n_9 = 4750$	$5.67\,(10)^{-2}$	$5.25\,(10)^{-2}$	$6.47\,(10)^{-2}$
$n_{10} = 5250$	$5.24\,(10)^{-2}$	$5.12\,(10)^{-2}$	$6.08\,(10)^{-2}$
$n_{11} = 5750$	$5.01\,(10)^{-2}$	$4.82\,(10)^{-2}$	$5.75\,(10)^{-2}$
$n_{12} = 6250$	$4.90\,(10)^{-2}$	$4.49\,(10)^{-2}$	$5.33\,(10)^{-2}$
$n_{13} = 6750$	$4.87\,(10)^{-2}$	$3.87\,(10)^{-2}$	$4.97\,(10)^{-2}$

It can be observed a similar performance of the kernel–based and penalized FAR(1) predictors, from smooth functional data, which is also comparable, considering one realization, to the performance obtained in Table A3.5.6, from the empirical eigenvectors.

A3.5.2.4 Wavelet–based prediction for ARH(1) processes

The approach presented in Antoniadis and Sapatinas [2003] is now studied. Specifically, wavelet-based regularization is applied to obtain smooth estimates of the sample paths. The projection onto the space V_J, generated by translations of the scaling function $\phi_{Jk},\ k = 0, \ldots, 2^J - 1$, at level J, associated with a multiresolution analysis of H, is first considered. For a given primary resolution level j_0, with $j_0 < J$, the following wavelet decomposition at $J - j_0$ resolution levels can be computed for any projected curve $\Phi_{V_J} X_i$, in the space V_J, for $i = 0, \ldots, n-1$:

$$\Phi_{V_J} X_i = \sum_{k=0}^{2^{j_0}-1} c_{j_0 k}^{i} \phi_{j_0 k} + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j - 1} d_{jk}^{i} \psi_{jk},$$
$$c_{j_0 k}^{i} = \langle \Phi_{V_J} X_i, \phi_{j_0 k}\rangle_H,\ d_{jk}^{i} = \langle \Phi_{V_J} X_i, \psi_{jk}\rangle_H.$$

For $i = 0, \ldots, n-1$, the following variational problem is solved to obtain the smooth estimate of the curve X_i :

$$\inf_{f^i \in H} \left\{ \left\| \Phi_{V_J} X_i - f^i \right\|_{L^2}^2 + \lambda \left\| \Phi_{V_{j_0}^{\perp}} f \right\|^2 ; \ f \in H \right\}, \tag{A3.72}$$

where $\Phi_{V_{j_0}^{\perp}}$ denotes the orthogonal projection operator of H onto the orhogonal complement of V_{j_0}, and for $i = 0, 1 \ldots n-1$,

$$f^i = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k}^i \phi_{j_0 k} + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{jk}^i \psi_{jk}.$$

Using the equivalent sequence of norms of fractional Sobolev spaces of order s with $s > 1/2$, on a suitable interval (in our case, $s = \delta_1$), the minimization of (A3.72) is equivalent to the optimization problem, for $i = 0, \ldots, n-1$,

$$\sum_{k=0}^{2^{j_0}-1} (\alpha_{j_0 k}^i - c_{j_0 k}^i)^2 + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (d_{jk}^i - \beta_{jk}^i)^2 + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \lambda 2^{js} [\beta_{jk}^i]^2. \tag{A3.73}$$

The solution to (A3.73) is given by, for $i = 0, \ldots, n-1$,

$$\begin{aligned}
\widehat{\alpha_{j_0 k}^i} &= c_{j_0 k}^i, \quad k = 0, 1, \ldots, 2^{j_0} - 1, \\
\widehat{\beta_{j_0 k}^i} &= \frac{d_{jk}^i}{(1 + \lambda 2^{2sj})}, \quad j = j_0, \ldots, J-1, \ k = 0, 1, \ldots, 2^j - 1, \\
\widehat{\beta_{j_0 k}^i} &= 0, \quad j \geq J, \ k = 0, 1, \ldots, 2^j - 1.
\end{aligned}$$

In particular, in the subsequent computations, we have considered the following value of the smoothing parameter λ (see Angelini et al. [2003]):

$$\widehat{\lambda}^M = \frac{\left(\sum_{j=1}^{M} \sigma_j^2 \right) \left(\sum_{j=1}^{M} C_j \right)}{n}.$$

The following smoothed data are then computed

$$\widetilde{X}_{i, \widehat{\lambda}^M} = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha_{j_0 k}^i} \phi_{j_0 k} + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \widehat{\beta_{j_0 k}^i} \psi_{jk},$$

removing the trend

$$\widetilde{a}_{n, \widehat{\lambda}^M} = \frac{1}{n} \sum_{i=0}^{n-1} \widetilde{X}_{i, \widehat{\lambda}^M}$$

to obtain

$$\widetilde{Y}_{i,\widehat{\lambda}^M} = \widetilde{X}_{i,\widehat{\lambda}^M} - \widetilde{a}_{n,\widehat{\lambda}^M}, \quad i = 0, \ldots, n-1,$$

for the computation of

$$\begin{aligned}
\widetilde{\rho}_{n,\widehat{\lambda}^M}(x) &= \left(\widetilde{\Pi}^{k_n}_{\widehat{\lambda}^M} \widetilde{D}_{n,\widehat{\lambda}^M} \widetilde{C}^{-1}_{n,\widehat{\lambda}^M} \widetilde{\Pi}^{k_n}_{\widehat{\lambda}^M}\right)(x) = \sum_{l=1}^{k_n} \widetilde{\rho}_{n,\widehat{\lambda}^M,l}(x) \widetilde{\phi}^M_l, \; x \in H, \\
\widetilde{\rho}_{n,\widehat{\lambda}^M,l}(x) &= \sum_{j=1}^{k_n} \frac{1}{n-1} \sum_{i=0}^{n-2} \frac{1}{\widetilde{C}_{n,\widehat{\lambda}^M,j}} \langle \widetilde{\phi}^M_j, x \rangle_H \widetilde{Y}_{i,\widehat{\lambda}^M,j} \widetilde{Y}_{i+1,\widehat{\lambda}^M,l},
\end{aligned}$$

for $x \in H$ and

$$\widetilde{C}_{n,\widehat{\lambda}^M} = \frac{1}{n} \sum_{i=0}^{n-1} \widetilde{Y}_{i,\widehat{\lambda}^M} \otimes \widetilde{Y}_{i,\widehat{\lambda}^M},$$

where

$$\widetilde{Y}_{i,\widehat{\lambda}^M,j} = \left\langle \widetilde{Y}_{i,\widehat{\lambda}^M}, \widehat{\phi}_{j,\widehat{\lambda}^M} \right\rangle,$$

and

$$\widetilde{C}_{n,\widehat{\lambda}^M,j} = \left\langle \widetilde{C}_{n,\widehat{\lambda}^M} \widehat{\phi}_{j,\widehat{\lambda}^M} \right\rangle,$$

for every $j \geq 1$. Table A3.5.8 displays the empirical truncated approximation of the expectation $\mathrm{E}\left\{\|\widetilde{\rho}_{k_n}(X_{n-1}) - \rho(X_{n-1})\|_H\right\}$ and $\mathrm{E}\left\{\|\widetilde{\rho}_{n,\widehat{\lambda}^M}(X_{n-1}) - \rho(X_{n-1})\|_H\right\}$, respectively obtained applying our approach, and the approach in Antoniadis and Sapatinas [2003], in the estimation of the autocorrelation operator ρ. Here, we have tested $k_{n_i} = \lceil n^{1/\alpha_i} \rceil$, $i = 1, 2$, with $\alpha_1 = 6$, according to **Assumption A3**, and $\alpha_2 > 4\delta_1$, according to

$$H_4 : \; nC^4_{k_n} \to \infty$$

in [Antoniadis and Sapatinas, 2003, p. 149]. In particular, we have considered $\delta_1 = 2.4$, and $\alpha_2 = 10$. From the results displayed in Table A3.5.8, one can observe a similar performance for the two truncation rules implemented, and approaches compared, for the small sample sizes tested. A similar accuracy is also displayed by the approaches presented in Besse et al. [2000], for such small sample sizes (see Table A3.5.7).

Table A3.5.8: Truncated empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1}) - \widetilde{\rho}_{k_n}(X_{n-1})\|_H\right\}$, with $\widetilde{\rho}_{k_n}$ defined in equation (A3.65), and of $\mathrm{E}\left\{\|\widetilde{\rho}_{n,\widehat{\lambda}^M}(X_{n-1}) - \rho(X_{n-1})\|_H\right\}$,, based on $N = 200$ simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $\{n_t = 750 + 500(t-1),\ t = 1, \dots, 13\}$, using $\widehat{\lambda}_M,\ M = 50$, and the corresponding $k_{n,i} = \lceil n^{1/\alpha_i} \rceil$, for $\alpha_1 = 6$ and $\alpha_2 = 10$. Here, O.A. means *Our Approach*.

n	$k_{n,1}$	O.A.	Antoniadis and Sapatinas [2003]	$k_{n,2}$	O.A.	Antoniadis and Sapatinas [2003]
750	3	0.070	0.091	1	0.064	0.059
1250	3	0.055	0.087	2	0.051	0.043
1750	3	0.047	0.080	2	0.045	0.039
2250	3	0.041	0.079	2	0.041	0.038
2750	3	0.037	0.073	2	0.036	0.035
3250	3	0.034	0.072	2	0.033	0.031
3750	3	0.033	0.068	2	0.033	0.029
4250	4	0.033	0.067	2	0.031	0.029
4750	4	0.032	0.066	2	0.031	0.026
5250	4	0.031	0.064	2	0.028	0.023
5750	4	0.030	0.060	2	0.020	0.019
6250	4	0.028	0.058	2	0.017	0.015
6750	4	0.028	0.056	2	0.019	0.014

A3.6 Final comments

As noted before, in this paper, the eigenvectors of C are considered to be known in the derivation of the results on consistency. This assumption is satisfied, e.g., when the random initial condition is given as the solution, in the mean-square sense, of a stochastic differential equation driven by white noise (e.g., the Wiener measure), since the eigenvectors of the differential operator involved in that equation coincide with the eigenvectors of the autocovariance operator of the ARH(1) process. In the case where the eigenvectors of the autocovariance operator are unknown, the numerical results displayed in Tables A3.5.4–A3.5.6 illustrate the fact that our approach displays, in terms of the empirical eigenvectors, very similar prediction results to those obtained with the implementation of the componentwise estimators proposed in Bosq [2000]; Guillas [2001], with a better performance of our approach in the more unfavorable case, corresponding to a large discretization step size, and truncation order (see Table A3.5.4 computed for $k_n = \lceil \ln(n) \rceil$).

Regarding **Assumption A2**, Remark A3.2.1 provides an example where this assumption is satisfied. However, our approach can still be applied in a wider range of situations. Wavelet bases are well suited for sparse representation of functions; recent work has considered combining them with sparsity-inducing penalties, both for semiparametric regression (see, e.g., Wand and Ormerod [2011]), and for regression with functional or kernel predictors (see Wand and Ormerod [2011]; Zhao et al. [2015, 2012], among others). The latter papers focused on ℓ_1 penalization, also known as the lasso (see Tibshirani [1996]), in the wavelet domain. Alternatives to the lasso include the SCAD penalty by Fan and Li [2001], and the adaptive lasso by Zou [2006]. The ℓ_1 penalty in the elastic net criterion has the effect of shrinking small coefficients to zero. This can be interpreted as imposing a prior that favors a sparse estimate. The above mentioned smoothing techniques, based on wavelets, can be applied to obtain a smooth sparse approximation $\widehat{X}_1, \dots \widehat{X}_n$ of the functional data $X_1, \dots, X_n$, whose empirical auto-covariance operator

$$\widehat{C}_n = \frac{1}{n}\sum_{i=0}^{n-1} \widehat{X}_i \otimes \widehat{X}_i$$

and cross-covariance operator

$$\widehat{D}_n = \frac{1}{n-1}\sum_{i=0}^{n-2} \widehat{X}_i \otimes \widehat{X}_{i+1}$$

admits a diagonal representation in terms of wavelets.

In the literature, shrinkage approaches for estimating a high–dimensional covariance matrix are employed to circumvent the limitations of the sample covariance matrix. In particular, a new family of nonparametric Stein–type shrinkage covariance estimators is proposed in Touloumis [2015] (see also references therein), whose members are written as a convex linear combination of the sample covariance matrix and of a predefined invertible diagonal target matrix. These results can be applied to our framework, considering the shrinkage estimators of the autocovariance and cross-covariance operators, with respect to a common suitable wavelet basis, which can lead to an empirical diagonal representation of both operators.

In the Supplementary Material provided (see Appendix A3.7), a numerical example is provided to illustrate the performance of our approach, in the case of a pseudo–diagonal autocorrelation operator.

A3.7 Supplementary Material: Non-diagonal autocorrelation operator

This Section provides as a numerical example where the methodology proposed in such paper still works beyond the considered **Assumption A2**. In particular, this section illustrates the performance of the proposed estimation methodology, when **Assumption A2** is not satisfied, but ρ is close to be diagonal in some sense. The numerical results obtained are compared to those ones derived from the computation of the ARH(1) predictors, based on the componentwise estimators proposed in Bosq [2000]; Guillas [2001] where this diagonal assumption is not required. The Gaussian ARH(1) process generated has autocorrelation operator ρ with coefficients $\rho_{j,h}$ with respect to the basis $\{\phi_j \otimes \phi_h,\ j,h \geq 1\}$, given by

$$\rho_{j,j}^2 \quad = \quad \left(\frac{\lambda_j\left((-\Delta)_{(a,b)}\right)}{\lambda_1\left((-\Delta)_{(a,b)}\right) - \epsilon}\right)^{-\delta_2}, \tag{A3.74}$$

in the diagonal, and outside of the diagonal

$$\rho_{j,j+a}^2 \quad = \quad \frac{0.01}{5a^2},\ a = 1,\dots,5, \quad \rho_{j+a,j}^2 = \frac{0.02}{5a^2},\ a = 1,\dots,5, \tag{A3.75}$$

where $\rho_{j,j+a}^2 = \rho_{j+a,j}^2 = 0$ when $a \geq 6$. The coefficients of the autocovariance operator C_ε of the innovation process ε, with respect to the mentioned basis $\{\phi_j \otimes \phi_h,\ j, h \geq 1\}$, are given by

$$\sigma_{j,j}^2 = C_j \left(1 - \rho_{j,j}^2\right)$$

in the diagonal, and outside of the diagonal by

$$\sigma_{j,j+a}^2 \quad = \quad \frac{0.015}{5a^2},\ a = 1, 2, 3, 4, 5, \quad \sigma_{j+a,j}^2 = \frac{0.01}{5a^2},\ a = 1, 2, 3, 4, 5, \qquad \text{(A3.76)}$$

where $\sigma_{j,j+a}^2 = \sigma_{j+a,j}^2 = 0$ when $a \geq 6$. Table A3.7.1 below displays the empirical truncated values of $\mathrm{E}\left\{\left\|\rho(X_{n-1}) - \widehat{\rho}_{k_n}^{ND}(X_{n-1})\right\|_H\right\}$ based on $N = 200$ simulations of each one of the 20 functional samples considered, with sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \ldots, 20\}$, for the corresponding k_n values obtained, in this case, by the rule $k_n = \lceil n^{1/\alpha} \rceil$, with $\alpha = 6$. We have considered parameter $\delta_1 = 2.4$ in the definition of the eigenvalues of C; but, in this case, as noted before, operators ρ and C_ε are non-diagonal (see equations A3.75–A3.76). The estimators of ρ and the associated plug–in predictors are computed, for the three approaches compared, under the assumption that the eigenvectors of C are known.

As expected, in Table A3.7.1, an outperformance of the approaches in Bosq [2000]; Guillas [2001] is observed in comparison with our methodology. However, for large sample sizes, the ARH(1) prediction methodology proposed here still can be applied with an order of magnitude of 10^{-2} for the empirical errors associated with $\widehat{\rho}_{k_n}$ given in equation A3.65. Thus, in the pseudodiagonal autocorrelation operator case, in some sense, our approach could still be considered. As referred in our paper, an example is given in the case where the autocovariance and autocorrelation operators admit a sparse representation in terms of a suitable orthonormal wavelet basis (see, for instance, Angelini et al. [2003]; Antoniadis and Sapatinas [2003]).

Table A3.7.1: Truncated empirical values of $\mathrm{E}\left\{\left\|\rho(X_{n-1}) - \widehat{\rho}_{k_n}^{ND}(X_{n-1})\right\|_H\right\}$, for $\widehat{\rho}_{k_n}^{ND}$ given in equations (A3.15)–(A3.16) (third column), in equations (A3.60)–(A3.61) (fourth column), and in equations (A3.62)–(A3.63) (fifth column), from the non–diagonal data generated by equations (A3.74)–(A3.76), based on $N = 200$ (due to high–dimensionality) simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $\{n_t = 15000 + 20000(t-1),\ t = 1, \dots, 20\}$ and the corresponding $k_n = \lceil n^{1/\alpha} \rceil,\ \alpha = 6$ values. The eigenvectors $\{\phi_j,\ j \geq 1\}$ are assumed to be known.

n	k_n	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	4	0.581	$8.94\,(10)^{-2}$	0.1055
$n_2 = 35000$	5	0.560	$7.05\,(10)^{-2}$	$9.49\,(10)^{-2}$
$n_3 = 55000$	6	0.548	$6.67\,(10)^{-2}$	$9.14\,(10)^{-2}$
$n_4 = 75000$	6	0.532	$6.24\,(10)^{-2}$	$8.85\,(10)^{-2}$
$n_5 = 95000$	6	0.512	$5.89\,(10)^{-2}$	$8.47\,(10)^{-2}$
$n_6 = 115000$	6	0.498	$5.62\,(10)^{-2}$	$8.04\,(10)^{-2}$
$n_7 = 135000$	7	0.495	$5.57\,(10)^{-2}$	$7.66\,(10)^{-2}$
$n_8 = 155000$	7	0.481	$5.28\,(10)^{-2}$	$7.24\,(10)^{-2}$
$n_9 = 175000$	7	0.474	$5.01\,(10)^{-2}$	$6.78\,(10)^{-2}$
$n_{10} = 195000$	7	0.461	$4.90\,(10)^{-2}$	$6.30\,(10)^{-2}$
$n_{11} = 215000$	7	0.442	$4.69\,(10)^{-2}$	$6.07\,(10)^{-2}$
$n_{12} = 235000$	7	0.425	$4.45\,(10)^{-2}$	$5.82\,(10)^{-2}$
$n_{13} = 255000$	7	0.411	$4.25\,(10)^{-2}$	$5.54\,(10)^{-2}$
$n_{14} = 275000$	8	0.408	$4.14\,(10)^{-2}$	$5.16\,(10)^{-2}$
$n_{15} = 295000$	8	0.381	$4.09\,(10)^{-2}$	$4.81\,(10)^{-2}$
$n_{16} = 315000$	8	0.360	$3.85\,(10)^{-2}$	$4.53\,(10)^{-2}$
$n_{17} = 335000$	8	0.349	$3.56\,(10)^{-2}$	$4.29\,(10)^{-2}$
$n_{18} = 355000$	8	0.330	$3.29\,(10)^{-2}$	$3.98\,(10)^{-2}$
$n_{19} = 375000$	8	0.320	$2.90\,(10)^{-2}$	$3.75\,(10)^{-2}$
$n_{20} = 395000$	8	0.318	$2.62\,(10)^{-2}$	$3.44\,(10)^{-2}$

Acknowledgments

This work has been supported in part by project MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

A4

THE EFFECT OF THE SPATIAL DOMAIN IN FANOVA MODELS WITH ARH(1) ERROR TERM

ÁLVAREZ–LIÉBANA, J.; RUIZ–MEDINA, M. D.: The effect of the spatial domain in FANOVA models with ARH(1) error term. Stat. Interface 10 (2017), pp. 607–628.

DOI: doi.org/10.4310/SII.2017.v10.n4.a7

Year	Categ.	Cites	Impact Factor (5 years)	Quartil
2016	Math. & Comput. Biology	421	0.808	Q4
2016	Math. Interdiscip. Appl.	421	0.808	Q4

ABSTRACT

Functional Analysis of Variance (FANOVA) from Hilbert–valued correlated data with spatial rectangular or circular supports is analysed, when Dirichlet conditions are assumed on the boundary. Specifically, a Hilbert–valued fixed effect model, with error term defined from an autoregressive Hilbertian process of order one (ARH(1) process) is considered, extending the formulation given in Ruiz-Medina [2016]. A new statistical test is also derived to contrast the significance of the functional fixed effect parameters. The Dirichlet conditions established at the boundary affect the dependence range of the correlated error term. While the rate of convergence to zero of the eigenvalues of the covariance kernels, characterizing the Gaussian functional error components, directly affects the stability of the generalized least–squares parameter estimation problem. A simulation study and a real–data application, related to fMRI analysis, are undertaken to illustrate the performance of the parameter estimator and statistical test derived.

A4.1 INTRODUCTION

In the last few decades, functional data analysis techniques have grown significantly given the new technologies available, in particular, in the field of medicine (see, for instance, Sorensen et al. [2013]). High–dimensional data, which are functional in nature, are generated, for example, from measurements in time, over spatial grids or images with many pixels (e.g., data on electrical activity of the heart, data on electrical activity along the scalp, data reconstructed from medical imaging, expression profiles in genetics and genomics, monitoring of continuity activity through accelerometers, etc). Effective experimental design and modern functional statistics have led to recent advances in medical imaging, improving, in particular, the study of human brain function (see, for example, Delzell et al. [2012]). Magnetic Resonance Imaging (MRI) data have been analysed with different aims. For example, we refer to the studies related with cortical thickness (see Lerch and Evans [2005]), where magnetic resonance imaging data are analysed to detect the spatial locations of the surface of the brain, where the cortical thickness is correlated with an independent variable, such as age or gender (see also Shaw et al. [2006]). Cortical thickness is usually previously smoothed along the surface of the brain (see Chung et al. [2005]). Thus, it can be considered as a functional random variable with spatial circular support. In general, the following linear model is considered, for cortical thickness $Y_i(s)$ on subjects $i = 1, \ldots, n,$:

$$Y_i(\mathrm{s}) = \boldsymbol{x}_i \boldsymbol{\beta}(\mathrm{s}) + Z_i(\mathrm{s})\sigma_i(\mathrm{s}), \quad \mathrm{s} \in S, \tag{A4.1}$$

where $\boldsymbol{x}_i$ is a vector of known p regressors, and for each $\mathrm{s} \in S$, with S denoting the surface of the brain, parameter $\boldsymbol{\beta}(\mathrm{s})$ is an unknown p–vector of regression coefficients. The errors $\{Z_1, \ldots, Z_n\}$ are independent zero-mean Gaussian random fields. In Taylor and Worsley [2007], this model is also considered to detect how the regressors are related to the data at spatial location s, by testing contrasts in $\boldsymbol{\beta}(\mathrm{s})$, for $\mathrm{s} \in S$. The approach presented in this paper allows the formulation of model (A4.1) in a functional (Hilbert–valued) framework, incorporating possible correlations between subjects, due to genetic characteristics, breed, geographic location, etc.

The statistical analysis of functional magnetic resonance image (fMRI) data has also generated an important activity in research about brain activity, where the functional statistical approach implemented in this paper could lead to important spatio–temporal analysis improvements. It is well–known that fMRI techniques have been developed to address the unobserved effect of scanner noise in studies of the auditory cortex. A penalized likelihood approach to magnetic resonance image reconstruction is presented in Bulaevskaya and Oehlert [2007]. A new approach which incorporates the spatial information from neighbouring voxels, as well as temporal correlation within each voxel, which makes use of regional kriging is derived in Christensen and Yetkin [2005]. Conditional autoregressive and Markov random field modelling involves some restrictions in the characterization of spatially contiguous effect regions, and, in general, in the representation of the spatial dependence between spatially connected voxels (see, for example, Banerjee et al. [2004]; Besag [1986]). Multiscale adaptive regression models assume spatial independence to construct a weighted likelihood parameter estimate. At each scale, the weights determine the amount of information that observations in a neighborhood voxel provides on the parameter vector to be estimated at a given voxel, under the assumption of independence between the conditional distributions of the responses at the neighborhood voxels, for each scale. The weights are sequentially computed through different scales, for adaptively update of the parameter estimates and test statistics (see, for example, Li et al. [2011]).

In Zhu et al. [2012], a multivariate varying coefficient model is considered for neuroimaging data, under a mixed effect approach, to reflect dependence within–curve and between–curve, in the case where coefficients are one–parameter functions, although extension to higher dimension is straightforward. The approach presented in this paper adopts a functional framework to analyse multivariate varying coefficient models in higher dimensions (two–dimensional design points), under the framework of multivariate fixed effect models in Hilbert spaces. Namely, the response is a multivariate functional random variable reflecting dependence within-surface (between voxels), and between-surface (between different times), with Hilbert–valued multivariate Gaussian distribution. Hence, the varying coefficients are estimated from the application of an extended version of generalized least–squares estimation methodology, in the multivariate Hilbert–valued context (see Ruiz-Medina [2016]), while, in Zhu et al. [2012], local linear regression is applied to estimate the coefficient functions. The dependence structure of the functional response is estimated here from the moment–based parameter estimation of the ARH(1) error term (see Bosq [2000]). In Zhu et al. [2012], local linear regression technique is employed to estimate the random effects, reflecting dependence structure in the varying coefficient mixed effect model. An extended formulation of the varying coefficient model considered in Zhu et al. [2012] is given in Zhu et al. [2014], combining a univariate measurement mixed effect model, a jumping surface model, and a functional component analysis model. In the approach presented in this paper, we have combined a nonparametric surface model with a multivariate functional principal component approach in the ARH(1) framework. Thus, a continuous spatial variation of the fMRI response is assumed, incorporating temporal and spatial correlations (across voxels), with an important dimension reduction in the estimation of the varying coefficient functions.

The above–referred advances in medicine are supported by the extensive literature on linear models in function spaces developed in parallel in the last few decades. We particularly refer to the functional linear regression context (see, for example, Cai and Hall [2006]; Cardot et al. [2003]; Cardot and Sarda [2011]; Chiou et al. [2004]; Crambes et al. [2009]; Cuevas et al. [2002]; Ferraty et al. [2013a]; Kokoszka et al. [2008], among others). See also Bosq [2000, 2007]; Ruiz-Medina [2011, 2012], in the functional time series context, and Ferraty and Vieu [2006, 2011] in the functional nonparametric regression framework.

Functional Analysis of Variance (FANOVA) techniques for high–dimensional data with a functional background have played a crucial role, within the functional linear model literature as well. Related work has been steadily growing (see, for example, Angelini et al. [2003]; Dette and Derbort [2001]; Gu [2002]; Huang [1998]; Kaufman and Sain [2010]; Kaziska [2011]; Lin [2000]; Ramsay and Silverman [2005]; Spitzner et al. [2003]; Stone et al. [1997]; Wahba et al. [1995]). The paper Ruiz-Medina [2016] extends the results in Zoglat [2008] from the $L^2([0,1])$–valued context to the separable Hilbert–valued space framework, and from the case of independent homocedastic error components to the correlated heteroscedastic case. In the context of hypothesis testing from functional data, tests of significance based on wavelet thresholding are formulated in Fan [1996], exploiting the sparsity of the signal representation in the wavelet domain, for dimension reduction. A maximum likelihood ratio based test is suggested for functional variance components in mixed–effect FANOVA models in Guo [2002]. From classical ANOVA tests, an asymptotic approach is derived in Cuevas et al. [2004], for studying the equality of the functional means from k independent samples of functional data. The testing problem for mixed–effect functional analysis of variance models is addressed in Abramovich and Angelini [2006]; Abramovich et al. [2004], developing asymptotically optimal (minimax) testing procedures for the significance of functional global trend, and the functional fixed effects. The wavelet transform of the data is again used in the implementation of this approach (see also Antoniadis and Sapatinas [2007]). Recently, in the context of functional data defined by curves, considering the L^2–norm, an up–to–date overview of hypothesis testing methods for functional data analysis is provided in Zhang [2013], including functional ANOVA, functional linear models with functional responses, heteroscedastic ANOVA for functional data, and hypothesis tests for the equality of covariance functions, among other related topics.

In this paper, the model formulated in Ruiz-Medina [2016] is extended to the case where the error term is an ARH(1) process. Furthermore, an alternative test to contrast the significance of the functional fixed effect parameters is formulated, based on a sharp form of the Cramér–Wold's Theorem derived in Cuesta-Albertos et al. [2007], for Gaussian measures on a separable Hilbert space. The simulation study undertaken illustrates the effect of the boundary conditions and the geometry of the domain on the spatial dependence range of the functional vector error term. Specifically, in that simulations, we consider the case where the Gaussian error components satisfy a stochastic partial differential equation, given in terms of a fractional power of the Dirichlet negative Laplacian operator. The autocovariance and cross–covariance operators of the functional error components are then defined in terms of the eigenvectors of the Dirichlet negative Laplacian operator. The eigenvectors of the Dirichlet negative Laplacian operator vanish continuously at the boundary, in the case of the regular domains studied (the rectangle, disk and circular sector), with decay velocity determined by the boundary conditions and the geometry of the domain. Thus, the boundary conditions and the geometry of the domain directly affect the dependence range of the error components, determined by the rate of convergence to zero of the Dirichlet negative Laplacian eigenvectors at the boundary. The influence of the truncation order is studied as well, since the rate of convergence to zero of the eigenvalues of the spatial covariance kernels, that define the matrix covariance operator of the error term, could affect the stability of the generalized least–squares estimation problem addressed here. Furthermore, in the fMRI data problem considered, the presented functional fixed effect model, with ARH(1) error term, is fitted. In that case, the temporal dependence range of the error term is controlled by the ARH(1) dynamics, while the spatial dependence range is controlled by the boundary conditions. Thus, the performance of the functional least–squares estimator and the functional significance test introduced in this paper is illustrated in both cases, the simulation study and the real–data example considered. A comparative study with the classical approach presented in Worsley et al. [2002] is also achieved for the fMRI data set analysed (freely available at `http://www.math.mcgill.ca/keith/fmristat/`).

The outline of this paper is as follows. The functional fixed effect model with ARH(1) error term is formulated in Appendix A4.2. The main results obtained on generalized least–squares estimation of the Hilbert–valued vector of fixed effect parameters, and the functional analysis of variance are also collected in this appendix. Linear hypothesis testing is derived in Appendix A4.3. The results obtained from the simulation study undertaken are displayed in Appendix A4.4. Functional statistical analysis of fMRI data is given in Appendix A4.5. Conclusions and open research lines are provided in Appendix A4.6. Finally, the Supplementary Material in Appendix A4.7 introduces the required preliminary elements on eigenvectors and eigenvalues of the Dirichlet negative Laplacian operator on the rectangle, disk and circular sector.

A4.2 Multivariate Hilbert–valued fixed effect model with ARH(1) error term

This section provides the extended formulation of the multivariate Hilbert–valued fixed effect model studied in Ruiz-Medina [2016], to the case where the correlated functional components of the error term satisfy an ARH(1) state equation. In that formulation, compactly supported non–separable autocovariance and cross–covariance kernels are considered for the functional error components, extending the separable case studied in Ruiz-Medina [2016].

Denote by H a real separable Hilbert space with the inner product $\langle\cdot,\cdot\rangle_H$, and the associated norm $\|\cdot\|_H$. Let us first introduce the multivariate Hilbert–valued fixed effect model with ARH(1) error term

$$\mathbf{Y}(\cdot) = \mathbf{X}\boldsymbol{\beta}(\cdot) + \boldsymbol{\varepsilon}(\cdot), \tag{A4.2}$$

where $\mathbf{X}$ is a real-valued $n \times p$ matrix, the fixed effect design matrix,

$$\boldsymbol{\beta}(\cdot) = [\beta_1(\cdot), \dots, \beta_p(\cdot)]^T \in H^p$$

represents the vector of fixed effect parameters,

$$\mathbf{Y}(\cdot) = [Y_1(\cdot), \dots, Y_n(\cdot)]^T$$

is the H^n-valued Gaussian response, with $\mathrm{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$. The H^n-valued error term

$$\boldsymbol{\varepsilon}(\cdot) = [\varepsilon_1(\cdot), \dots, \varepsilon_n(\cdot)]^T$$

is assumed to be an ARH(1) process on the basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$; i.e., a stationary in time Hilbert–valued Gaussian process satisfying (see Bosq [2000])

$$\varepsilon_m(\cdot) = \rho(\varepsilon_{m-1})(\cdot) + \nu_m(\cdot), \quad m \in \mathbb{Z}, \tag{A4.3}$$

where $\mathrm{E}\{\varepsilon_m\} = 0$, for each $m \in \mathbb{Z}$, and ρ denotes the autocorrelation operator of the error process ε, which belongs to the space of bounded linear operators on H. Here, $\nu = \{\nu_m,\ m \in \mathbb{Z}\}$ is assumed to be a Gaussian strong white noise; i.e., ν is a Hilbert–valued zero–mean stationary process, with independent and identically distributed components in time, and with $\sigma^2 = \mathrm{E}\left\{\|\nu_m\|_H^2\right\} < \infty$, for all $m \in \mathbb{Z}$. Thus, in (A4.2), the components of the vector error term $[\varepsilon_1(\cdot), \ldots, \varepsilon_n(\cdot)]^T$ corresponding to observations at times $t_1, \ldots, t_n$, obey the functional state equation (A4.3), under suitable conditions on the point spectrum of the autocorrelation operator ρ. Hence, the non–null functional entries of the matrix covariance operator $\mathbf{R}_{\boldsymbol{\varepsilon\varepsilon}}$ of

$$\boldsymbol{\varepsilon}(\cdot) = [\varepsilon_1(\cdot), \ldots, \varepsilon_n(\cdot)]^T$$

are then constituted by the elements located at the three main diagonals. Specifically,

$$\mathrm{E}\left\{\varepsilon_i \otimes \varepsilon_j\right\} = R_1, \quad \text{if } j - i = 1, \quad \mathrm{E}\left\{\varepsilon_i \otimes \varepsilon_j\right\} = R_1^*, \quad \text{if } i - j = 1,$$

and

$$\mathrm{E}\left\{\varepsilon_i \otimes \varepsilon_i\right\} = R_0, \quad \text{if } i = j,$$

where R_1 and R_1^* denote, respectively, the cross–covariance operator and its adjoint for the ARH(1) process $\varepsilon = \{\varepsilon_i,\ i \in \mathbb{Z}\}$, and R_0 represents its autocovariance operator. Note that, in this appendix, it is assumed that ρ is sufficiently regular. In particular, ρ is such that $\|\rho^2\|_{\mathcal{L}(H)} \simeq 0$.

Equivalently, the matrix covariance operator $\mathbf{R}_{\boldsymbol{\varepsilon\varepsilon}}$ is given by

$$\begin{aligned}
\mathbf{R}_{\boldsymbol{\varepsilon\varepsilon}} &= \mathrm{E}\left\{[\varepsilon_1(\cdot), \ldots, \varepsilon_n(\cdot)]^T [\varepsilon_1(\cdot), \ldots, \varepsilon_n(\cdot)]\right\} \\
&= \begin{pmatrix} \mathrm{E}\left\{\varepsilon_1 \otimes \varepsilon_1\right\} & \ldots & \mathrm{E}\left\{\varepsilon_1 \otimes \varepsilon_n\right\} \\ \vdots & \ddots & \vdots \\ \mathrm{E}\left\{\varepsilon_n \otimes \varepsilon_1\right\} & \ldots & \mathrm{E}\left\{\varepsilon_n \otimes \varepsilon_n\right\} \end{pmatrix} \\
&\simeq \begin{pmatrix} R_0 & R_1 & 0_H & 0_H & \ldots & 0_H & 0_H & 0_H \\ R_1^* & R_0 & R_1 & 0_H & \ldots & 0_H & 0_H & 0_H \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0_H & 0_H & 0_H & 0_H & \ldots & R_1^* & R_0 & R_1 \\ 0_H & 0_H & 0_H & 0_H & \ldots & 0_H & R_1^* & R_0 \end{pmatrix},
\end{aligned}$$

where 0_H denotes the approximation by zero in the corresponding operator norm, given the conditions imposed on ρ.

In the space $\mathcal{H} = H^n$, we consider the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle_{H^n} = \sum_{i=1}^{n} \langle f_i, g_i \rangle_H, \qquad \mathbf{f}, \mathbf{g} \in H^n.$$

It is well–known that the autocovariance operator R_0 of an ARH(1) process is in the trace class (see [Bosq, 2000, pp. 27–36]). Therefore, it admits a diagonal spectral decomposition

$$R_0 = \sum_{k=1}^{\infty} \lambda_k \phi_k \otimes \phi_k,$$

in terms of a complete orthogonal eigenvector system $\{\phi_k,\ k \geq 1\}$, defining in H a resolution of the identity $\sum_{k=1}^{\infty} \phi_k \otimes \phi_k$. Here, for each $k \geq 1$, $\lambda_k = \lambda_k(R_0)$ is the k–th eigenvalue of R_0, with $R_0(\phi_k) = \lambda_k(R_0)\phi_k$. The following series expansion then holds, in the mean–square sense:

$$\varepsilon_i = \sum_{k=1}^{\infty} \langle \varepsilon_i, \phi_k \rangle_H \phi_k = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\eta_k(i)\phi_k, \quad i = 1, \dots, n,$$

where $\eta_k(i) = \frac{1}{\sqrt{\lambda_k}} \langle \varepsilon_i, \phi_k \rangle_H$, for $k \geq 1$ and $i \in \mathbb{N}$.

The following assumption is made:

Assumption A0. The standard Gaussian random variable sequences $\{\eta_k(i),\ k \geq 1,\ i \in \mathbb{N}\}$, with, for each $k \geq 1$,

$$\sqrt{\lambda_k}\eta_k(i) = \langle \varepsilon_i, \phi_k \rangle_H,$$

for every $i \in \mathbb{N}$, satisfy the following orthogonality condition, for every $i, j \in \mathbb{N}$,

$$\mathrm{E}\{\eta_k(i)\eta_p(j)\} = \delta_{k,p}, \quad k, p \in \mathbb{N},$$

where δ denotes the Kronecker delta function, and

$$R_1 = \sum_{k=1}^{\infty} \lambda_k(R_1)\phi_k \otimes \phi_k, \quad R_1^* = \sum_{k=1}^{\infty} \lambda_k(R_1^*)\phi_k \otimes \phi_k.$$

Under **Assumption A0**, the computation of the generalized least–squares estimator of $[\beta_1(\cdot), \dots, \beta_p(\cdot)]^T$ is achieved by projection into the orthogonal basis of eigenvectors $\{\phi_k,\ k \geq 1\}$ of the autocovariance operator R_0 of the ARH(1) process $\varepsilon = \{\varepsilon_i,\ i \in \mathbb{Z}\}$. Denote by $\mathbf{\Phi}^*$ the projection operator into the eigenvector system $\{\phi_k,\ k \geq 1\}$, acting on a vector function $\mathbf{f} \in \mathcal{H} = H^n$ as follows:

$$\begin{aligned} \mathbf{\Phi}^*(\mathbf{f}) &= \{\mathbf{\Phi}_k^*(\mathbf{f}),\ k \geq 1\} = \left\{(\langle f_1, \phi_k \rangle, \dots, \langle f_n, \phi_k \rangle)^T,\ k \geq 1\right\} \\ &= \left\{(f_{k1}, \dots, f_{kn})^T,\ k \geq 1\right\} = \left\{\mathbf{f}_k^T,\ k \geq 1\right\}, \end{aligned} \tag{A4.4}$$

where $\mathbf{\Phi\Phi}^* = \mathbf{Id}_{\mathcal{H}=H^n}$, with

$$\mathbf{\Phi}\left(\{\mathbf{f}_k^T,\ k \geq 1\}\right) = \left(\sum_{k=1}^{\infty} f_{k1}\phi_k, \dots, \sum_{k=1}^{\infty} f_{kn}\phi_k\right)^T.$$

For $\mathbf{A} = \{A_{i,j}\}_{i=1,\dots,n}^{j=1,\dots,n}$ be a matrix operator such that, for each $i, j = 1, \dots, n$, its functional entries are given by

$$A_{i,j} = \sum_{k=1}^{\infty} \gamma_{kij} \phi_k \otimes \phi_k$$

with $\sum_{k=1}^{\infty} \gamma_{kij}^2 < \infty$. The following identities are straightforward:

$$\mathbf{\Phi}^* \mathbf{A} \mathbf{\Phi} = \{\mathbf{\Gamma}_k, \ k \geq 1\}, \quad \mathbf{\Phi}\left(\{\mathbf{\Gamma}_k, \ k \geq 1\}\right) \mathbf{\Phi}^* = \mathbf{A}, \tag{A4.5}$$

where, for each $k \geq 1$, the entries of $\mathbf{\Gamma}_k$ are $\Gamma_{kij} = \gamma_{kij}$, for $i, j = 1, \ldots, n$.

Applying (A4.4)–(A4.5), we directly obtain

$$\begin{aligned}
\mathbf{\Phi}^* \mathbf{R}_{\varepsilon\varepsilon} \mathbf{\Phi} &= \{\mathbf{\Lambda}_k, \ k \geq 1\}, \quad \mathbf{\Phi}^* \mathbf{R}_{\varepsilon\varepsilon}^{-1} \mathbf{\Phi} = \left\{\mathbf{\Lambda}_k^{-1}, \ k \geq 1\right\}, \\
\mathbf{R}_{\varepsilon\varepsilon}^{-1}(\mathbf{f}, \mathbf{g}) &= \mathbf{\Phi}^* \mathbf{R}_{\varepsilon\varepsilon}^{-1} \mathbf{\Phi}\left(\mathbf{\Phi}^* \mathbf{f}, \mathbf{\Phi}^* \mathbf{g}\right) = \langle \mathbf{f}, \mathbf{g} \rangle_{\mathbf{R}_{\varepsilon\varepsilon}^{-1}} = \sum_{k=1}^{\infty} \mathbf{f}_k^T \mathbf{\Lambda}_k^{-1} \mathbf{g}_k, \quad \mathbf{f}, \ \mathbf{g} \in \mathbf{R}_{\varepsilon\varepsilon}^{1/2}\left(H^n\right), \\
\|\mathbf{f}\|_{\mathbf{R}_{\varepsilon\varepsilon}^{-1}}^2 &= \sum_{k=1}^{\infty} \mathbf{f}_k^T \mathbf{\Lambda}_k^{-1} \mathbf{f}_k, \quad \mathbf{f} \in \mathbf{R}_{\varepsilon\varepsilon}^{1/2}\left(H^n\right),
\end{aligned} \tag{A4.6}$$

where, for each $k \geq 1$, $\mathbf{\Lambda}_k = \mathbf{\Phi}_k^* \mathbf{R}_{\varepsilon\varepsilon} \mathbf{\Phi}_k$ is given by

$$\mathbf{\Lambda}_k = \begin{bmatrix} \lambda_k(R_0) & \lambda_k(R_1) & 0 & \ldots & 0 & 0 \\ \lambda_k(R_1^*) & \lambda_k(R_0) & \lambda_k(R_1) & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_k(R_1^*) & \lambda_k(R_0) \end{bmatrix}, \tag{A4.7}$$

with $\mathbf{\Lambda}_k^{-1}$ denoting its inverse matrix.

Remark A4.2.1 *In Appendix A4.4, we restrict our attention to the functional error model studied in Ruiz-Medina [2016], considering the Hilbert–valued stochastic partial differential equation system framework. In that framework, matrices* $\{\mathbf{\Lambda}_k, \ k \geq 1\}$ *, are known, since they are defined from the eigenvalues of the differential operators involved in the equation system. Particularly, in that section, for each* $k \geq 1$*, matrix* $\mathbf{\Lambda}_k$ *is considered to have entries* Λ_{kij} *given by*

$$\begin{aligned}
\Lambda_{kij} &= \exp\left(-\frac{|i-j|}{\lambda_{ki} + \lambda_{kj}}\right), \quad \text{if } i \neq j, \\
\Lambda_{kii} &= \lambda_{ki} = \lambda_k\left([f_i(-\mathbf{\Delta}_{D_l})]^2\right) = \lambda_k\left((-\mathbf{\Delta}_{D_1})^{-2(d-\gamma_i)}\right) = \left[\lambda_k\left((-\mathbf{\Delta}_{D_1})\right)\right]^{-2(d-\gamma_i)},
\end{aligned} \tag{A4.8}$$

with

$$\gamma_i \in (0, d/2), \quad i = 1, \ldots, n,$$

and $(-\boldsymbol{\Delta}_{D_l})$ *representing the Dirichlet negative Laplacian operator on domain* D_l*, for* $l = 1$ *(the rectangle),* $l = 2$ *(the disk) and* $l = 3$ *(the circular sector). However, in practice, as shown in Appendix A4.5 in the analysis of fMRI data, matrices* $\{\boldsymbol{\Lambda}_k,\ k \geq 1\}$ *, are not known, and should be estimated from the data. Indeed, in that real–data example, we approximate the entries of* $\{\boldsymbol{\Lambda}_k,\ k \geq 1\}$ *, from the coefficients (eigenvalues and singular values), that define the diagonal spectral expansion of the empirical autocovariance* $\widehat{R_0}$ *and cross covariance* $\widehat{R_1}$ *operators, given by (see Bosq [2000])*

$$\widehat{R_0} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \otimes \varepsilon_i, \quad \widehat{R_1} = \frac{1}{n-1}\sum_{i=1}^{n-1} \varepsilon_i \otimes \varepsilon_{i+1}, \quad \widehat{R_1^*} = \frac{1}{n-1}\sum_{i=2}^{n} \varepsilon_i \otimes \varepsilon_{i-1}. \tag{A4.9}$$

We also consider here the following semi–orthogonal condition for the non-square design matrix $\mathbf{X}$:

Assumption A1. The fixed effect design matrix $\mathbf{X}$ is a semi–orthogonal non–square matrix. That is,

$$\mathbf{X}^T\mathbf{X} = \mathbf{Id}_p, \quad \mathbf{Id}_p \in \mathbb{R}^{p \times p}.$$

Remark A4.2.2 ***Assumption A1*** *implies (see Ruiz-Medina [2016])*

$$\sum_{k=1}^{\infty} \mathrm{Tr}\left(\mathbf{X}^T\boldsymbol{\Lambda}_k^{-1}\mathbf{X}\right)^{-1} < \infty.$$

The generalized least–squares estimation of $[\beta_1(\cdot), \ldots, \beta_p(\cdot)]^T$ is achieved by minimizing the loss quadratic function in the norm of the Reproducing Kernel Hilbert Space (RKHS norm). Note that, for an $\mathcal{H}$–valued zero–mean Gaussian random variable with autocovariance operator R_Z, the RKHS of Z is defined by

$$\mathcal{H}(Z) = R_Z^{1/2}(\mathcal{H})$$

(see, for example, Prato and Zabczyk [2002]).

From equation (A4.6) we get

$$\mathrm{E}\left\{\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\boldsymbol{R}_{\varepsilon\varepsilon}^{-1}}\right\} = \boldsymbol{R}_{\varepsilon\varepsilon}^{-1}(\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon}) = \sum_{k=1}^{\infty}\mathrm{E}\left\{\|\boldsymbol{\varepsilon}_k(\boldsymbol{\beta}_k)\|^2_{\boldsymbol{\Lambda}_k^{-1}}\right\} \simeq \sum_{k=1}^{\infty}\mathrm{E}\left\{\|\boldsymbol{\varepsilon}_k(\boldsymbol{\beta}_k)\|^2_{\widehat{\boldsymbol{\Lambda}}_k^{-1}}\right\}, \tag{A4.10}$$

where, in the last identity, for each $k \geq 1$, matrix $\widehat{\boldsymbol{\Lambda}}_k$ represents the empirical counterpart of $\boldsymbol{\Lambda}_k$, constructed from the eigenelements of $\widehat{R_0}$, $\widehat{R_1}$ and $\widehat{R_1}^*$, considered when R_0 and R_1 are unknown. Here,

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}, \quad \boldsymbol{\varepsilon}_k(\boldsymbol{\beta}_k) = \boldsymbol{\Phi}_k^*(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad k \geq 1.$$

The minimum of equation (A4.10) is attached if, for each $k \geq 1$, the expectation

$$\mathrm{E}\left\{\|\boldsymbol{\varepsilon}_k\left(\boldsymbol{\beta}_k\right)\|^2_{\boldsymbol{\Lambda}_k^{-1}}\right\}$$

is minimized, with, as before, $\boldsymbol{\Lambda}_k^{-1}$ defining the inverse of matrix $\boldsymbol{\Lambda}_k$ given in (A4.7) (and approximated by $\widehat{\boldsymbol{\Lambda}}_k$, when R_0 and R_1 are unknown). That is,

$$\widehat{\boldsymbol{\beta}}_k \;=\; \left(\widehat{\beta_{k1}},\ldots,\widehat{\beta_{kp}}\right)^T = \left(\mathbf{X}^T\boldsymbol{\Lambda}_k^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Lambda}_k^{-1}\mathbf{Y}_k,$$

and given by

$$\left(\widetilde{\beta_{k1}},\ldots,\widetilde{\beta_{kp}}\right)^T = \left(\mathbf{X}^T\widehat{\boldsymbol{\Lambda}}_k^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\widehat{\boldsymbol{\Lambda}}_k^{-1}\mathbf{Y}_k, \tag{A4.11}$$

in the case where R_0 and R_1 are unknown. Here, $\mathbf{Y}_k = \boldsymbol{\Phi}_k^*\left(\mathbf{Y}\right)$ is the vector of projections into ϕ_k of the components of $\mathbf{Y}$, for each $k \geq 1$.

In the remaining of this section, we restrict our attention to the case where R_0 and R_1 are known. In that case,

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\Phi}\left(\left\{\widehat{\boldsymbol{\beta}}_k,\; k \geq 1\right\}\right) = \left(\sum_{k=1}^{\infty}\widehat{\beta_{k1}}\phi_k,\ldots,\sum_{k=1}^{\infty}\widehat{\beta_{kp}}\phi_k\right)^T.$$

The estimated response is then given by $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$. Under **Assumption A1**,

$$\mathrm{E}\left\{\sum_{k=1}^{\infty}\sum_{i=1}^{p}\widehat{\beta}_{ki}^2\right\} \;=\; \sum_{k=1}^{\infty}\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\Lambda}_k^{-1}\mathbf{X})^{-1} + \|\boldsymbol{\beta}\|^2_{H^p} < \infty, \tag{A4.12}$$

i.e., $\widehat{\boldsymbol{\beta}} \in H^p$ almost surely (see Ruiz-Medina [2016] for more details).

Remark A4.2.3 In the case where R_0 and R_1 are unknown, under the conditions assumed in [Bosq, 2000, Corollary 4.2, pp. 101–102], strong consistency of the empirical autocovariance operator $\widehat{R_0}$ holds. Moreover, under the conditions assumed in [Bosq, 2000, Theorem 4.8, pp. 116–117], the empirical cross–covariance operator $\widehat{R_0}$ is strongly–consistent. Therefore, the plug–in functional parameter estimator (A4.11) satisfies (A4.12), for n sufficiently large.

The Functional Analysis of Variance of model in (A4.2)–(A4.3) can be achieved as described in Ruiz-Medina [2016]. Specifically, a linear transformation of the functional data should be considered, for the almost surely finiteness of the functional components of variance, in the following way:

$$\mathbf{WY} = \mathbf{WX}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\varepsilon}, \tag{A4.13}$$

where $\mathbf{W}$ is such that

$$\mathbf{W} = \begin{pmatrix} \sum_{k=1}^{\infty} w_{k11}\phi_k \otimes \phi_k & \dots & \sum_{k=1}^{\infty} w_{k1n}\phi_k \otimes \phi_k \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{\infty} w_{kn1}\phi_k \otimes \phi_k & \dots & \sum_{k=1}^{\infty} w_{knn}\phi_k \otimes \phi_k \end{pmatrix},$$

and satisfies

$$\sum_{k=1}^{\infty} \mathrm{Tr}\left(\mathbf{\Lambda}_k^{-1}\mathbf{W}_k\right) < \infty. \tag{A4.14}$$

Here, for each $k \geq 1$, $\mathbf{\Lambda}_k$ is defined in (A4.7). The functional components of variance associated with the transformed model (A4.13) are then given by

$$\begin{aligned}
\widetilde{SST} &= \langle \mathbf{WY}, \mathbf{WY} \rangle_{R_{\varepsilon\varepsilon}^{-1}} = \sum_{k=1}^{\infty} \mathbf{Y}_k^T \mathbf{W}_k^T \mathbf{\Lambda}_k^{-1} \mathbf{W}_k \mathbf{Y}_k, \\
\widetilde{SSE} &= \langle \mathbf{W}\left(\mathbf{Y} - \widehat{\mathbf{Y}}\right), \mathbf{W}\left(\mathbf{Y} - \widehat{\mathbf{Y}}\right) \rangle_{R_{\varepsilon\varepsilon}^{-1}} = \sum_{k=1}^{\infty} \left(\mathbf{M}_k \mathbf{W}_k \mathbf{Y}_k\right)^T \mathbf{\Lambda}_k^{-1} \mathbf{M}_k \mathbf{W}_k \mathbf{Y}_k, \\
\widetilde{SSR} &= \widetilde{SST} - \widetilde{SSE}.
\end{aligned}$$

where $\mathbf{M}_k = \mathbf{Id}_{n\times n} - \mathbf{X}\left(\mathbf{X}^T \mathbf{\Lambda}_k^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{\Lambda}_k^{-1}$, for each $k \geq 1$.

The statistics

$$F = \frac{\widetilde{SSR}}{\widetilde{SSE}}, \tag{A4.15}$$

provides information on the relative magnitude between the empirical variability explained by the functional transformed model and the residual variability (see Appendix A4.4).

A4.3 Significance test from the Cramér–Wold's Theorem

In Ruiz-Medina [2016], a linear functional statistical test is formulated, with explicit definition of the probability distribution of the derived functional statistics under the null hypothesis:

$$H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{C},$$

against

$$H_1 : \mathbf{K}\boldsymbol{\beta} \neq \mathbf{C},$$

where $\mathbf{C} \in H^m$ and

$$\mathbf{K} : H^p \longrightarrow H^m$$

is a matrix operator such that its functional entries $\mathbf{K} = \{K_{ij}\}_{i=1,\ldots,m}^{j=1,\ldots,p}$, are given, for each $f, g \in H$, by

$$K_{ij}(f)(g) = \sum_{k=1}^{\infty} \lambda_k(K_{ij}) \langle \phi_k, g \rangle_H \langle \phi_k, f \rangle_H.$$

In particular,

$$\{(\mathbf{\Phi}_k^* \mathbf{K} \mathbf{\Phi}_k),\ k \geq 1\} = \{\mathbf{K}_k,\ k \geq 1\}$$

with

$$\mathbf{K}_k = \begin{pmatrix} \lambda_k(K_{11}) & \ldots & \lambda_k(K_{1p}) \\ \vdots & \ddots & \vdots \\ \lambda_k(K_{m1}) & \ldots & \lambda_k(K_{mp}) \end{pmatrix} \in \mathbb{R}^{m \times p}.$$

At level α, there exists a test ψ given by:

$$\psi = \begin{cases} 1 & \text{if} \quad S_{H_0}(\mathbf{Y}) > C(H_0, \alpha), \\ 0 & \text{otherwise,} \end{cases}$$

where

$$S_{H_0}(\mathbf{Y}) = \left\langle \mathbf{K}\widehat{\boldsymbol{\beta}} - \mathbf{C}, \mathbf{K}\widehat{\boldsymbol{\beta}} - \mathbf{C} \right\rangle_{\mathcal{H}=H^n}.$$

The constant $C(H_0, \alpha)$ is such that

$$\mathcal{P}\{S_{H_0}(\mathbf{Y}) > C(H_0, \alpha), \quad \mathbf{K}\boldsymbol{\beta} = \mathbf{C}\} = 1 - \mathcal{P}\{S_{H_0}(\mathbf{Y}) \leq C(H_0, \alpha),\ \mathbf{K}\boldsymbol{\beta} = \mathbf{C}\} = 1 - \mathbf{F}_\alpha = \alpha,$$

where the probability distribution $\mathbf{F}$ on $\mathcal{H} = H^n$ has characteristic functional given in [Ruiz-Medina, 2016, Proposition 4, Eq. (66)].

Alternatively, as an application of [Cuesta-Albertos et al., 2007, Theorem 4.1], a multivariate version of the significance test formulated in Cuesta-Albertos and Febrero-Bande [2010] is considered here, for the fixed effect parameters (see, in particular, [Cuesta-Albertos and Febrero-Bande, 2010, Theorem 2.1]. Specifically, we consider

$$H_0^{\mathbf{h}}:\ \mathbf{K}\boldsymbol{\beta}(\mathbf{h}) = \mathbf{C}, \tag{A4.16}$$

for $\mathbf{h} = (h, \ldots, h)_{p \times 1}^T$ defining a random vector in H^p, with h generated from a zero–mean Gaussian measure μ in H, with trace covariance operator R_μ (see, for example, Prato and Zabczyk [2002]). Here,

$$\boldsymbol{\beta}(\mathbf{h}) = \left(\langle \beta_1, h \rangle_H, \ldots, \langle \beta_p, h \rangle_H\right)_{p \times 1}^T,$$

$\mathbf{K}$ is given by

$$\mathbf{K} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{pmatrix} \in \mathbb{R}^{(p-1)\times p}, \tag{A4.17}$$

and $\mathbf{C}$ is a null $(p-1) \times 1$ functional vector; i.e.,

$$\mathbf{C} = (0, 0, \dots, 0)^T \in \mathbb{R}^{(p-1)\times 1}. \tag{A4.18}$$

From equations (A4.17)–(A4.18), for any $(p \times 1)$–dimensional functional random vector $\mathbf{h} = (h, \dots, h)^T_{p\times 1}$ generated from a Gaussian measure μ on H, $H_0^{\mathbf{h}}$ can then be equivalently expressed as

$$H_0^{\mathbf{h}} : \langle \beta_1, h \rangle_H = \langle \beta_2, h \rangle_H = \dots = \langle \beta_p, h \rangle_H . \tag{A4.19}$$

The test statistic to contrast (A4.19) is defined as

$$T_{\boldsymbol{h}} = \left(\mathbf{K}\widehat{\boldsymbol{\beta}}(\mathbf{h}) - \mathbf{C}\right)^T \left(\mathbf{K}\mathbf{Q}_{\mathbf{h}}\mathbf{K}^T\right)^{-1} \left(\mathbf{K}\widehat{\boldsymbol{\beta}}(\mathbf{h}) - \mathbf{C}\right), \tag{A4.20}$$

where $\mathbf{K}$ and $\mathbf{C}$ are respectively given in equations (A4.17)–(A4.18), and

$$\mathbf{Q}_{\mathbf{h}} = (\mathbf{X}^T \boldsymbol{\Lambda}_{\mathbf{h}} \mathbf{X})^{-1}, \quad \widehat{\boldsymbol{\beta}}(\mathbf{h}) = \left(\mathbf{X}^T \boldsymbol{\Lambda}_{\mathbf{h}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Lambda}_{\mathbf{h}}^{-1} \mathbf{Y}(\mathbf{h}), \tag{A4.21}$$

with

$$\mathbf{Y}(\mathbf{h}) = \left(\langle Y_1, h \rangle_H, \dots, \langle Y_n, h \rangle_H\right).$$

Here, $\boldsymbol{\Lambda}_{\mathbf{h}}$ is a $(n \times n)$–dimensional matrix with entries $\{\Lambda_{\mathbf{h}}(i,j)\}_{i=1,\dots,n}^{j=1,\dots,n}$, given by

$$\Lambda_{\mathbf{h}}(i,j) = \sum_{k=1}^{\infty} \left[\langle h, \phi_k \rangle_H\right]^2 \lambda_k(R_{ij}), \quad i,j = 1, \dots, n,$$

where, as before, $\lambda_k(R_{ij})$ denotes the k–th coefficient in the diagonal expansion of the covariance operator R_{ij} with respect to the basis $\{\phi_k \otimes \phi_k,\ k \geq 1\}$; i.e., in the diagonal expansion

$$R_{ij} = \sum_{k=1}^{\infty} \lambda_k(R_{ij}) \phi_k \otimes \phi_k, \quad i,j = 1, \dots, n.$$

Note that in the ARH(1) error term case described in Appendix A4.2, from equation (A4.7),

$$\lambda_k(R_{ij}) = 0, \quad \text{for } |i-j| > 1,\ k \geq 1.$$

Assuming that the autocovariance and cross–covariance operator of the ARH(1) error terms are known, under the null hypothesis $H_0^{\mathbf{h}}$, the conditional distribution of T_h in (A4.20), given $Y = h$, is a chi–square distribution with $p-1$ degrees of freedom. Here, Y is a zero-mean H–valued random variable with Gaussian probability measure μ on H, having trace covariance operator R_μ. Note that the last assertion directly follows from the fact that, in equation (A4.21), the conditional distribution of $\widehat{\boldsymbol{\beta}}(\mathbf{h})$ given $Y=h$ is

$$\widehat{\boldsymbol{\beta}}(\mathbf{h}) \sim \mathcal{N}(\boldsymbol{\beta}(\mathbf{h}), \mathbf{Q}_{\mathbf{h}}),$$

with $\mathbf{Q}_{\mathbf{h}}$ being introduced in equation (A4.21); i.e., the conditional distribution of $\widehat{\boldsymbol{\beta}}(\mathbf{h})$, given $Y=h$, is a multivariate Gaussian distribution with mean vector $\boldsymbol{\beta}(\mathbf{h})$ and covariance matrix $\mathbf{Q}_{\mathbf{h}}$.

From [Cuesta-Albertos et al., 2007, Theorem 4.1] and [Cuesta-Albertos and Febrero-Bande, 2010, Theorem 2.1], if

$$H_0: \beta_1(\cdot) = \beta_2(\cdot) = \cdots = \beta_p(\cdot)$$

fails, then, for μ-almost every function $h \in H$, $H_0^{\mathbf{h}}$ in (A4.16), or equivalently in (A4.19), also fails. Thus, a statistical test at level α to test $H_0^{\mathbf{h}}$ is a statistical test at the same level α to test H_0.

A4.4 Simulation study

In this section, we consider the real separable Hilbert space

$$H = L_0^2(D_l) = \overline{\mathcal{C}_0^\infty(D_l)}^{L^2(\mathbb{R}^2)},$$

the closure, in the norm of the square integrable functions in $\mathbb{R}^2$, of the space of infinitely differentiable functions with compact support contained in D_l, for each $l=1,2,3$. We restrict our attention to the family of error covariance operators given in (A4.8). Thus, for each $i,j=1,\dots,n$,

$$R_{\varepsilon_i\varepsilon_j} = R_{ij} = \mathrm{E}\{\varepsilon_i \otimes \varepsilon_j\} = \sum_{k=1}^{\infty}\left(\delta_{i,j}^{*}\exp\left(-\frac{|i-j|}{\lambda_{ki}+\lambda_{kj}}\right) + \delta_{i,j}\sqrt{\lambda_{ki}\lambda_{kj}}\right)\phi_k\otimes\phi_k, \tag{A4.22}$$

where $\delta_{i,j}^{*} = 1-\delta_{i,j}$, and $\delta_{i,j}$ is the Kronecker delta function. As before, for each $i,j=1,\dots,n$ and $k \geq 1$,

$$\lambda_{ki} = \lambda_k(R_{ii}), \quad \lambda_k(R_{ij}) = \exp\left(-\frac{|i-j|}{\lambda_{ki}+\lambda_{kj}}\right).$$

Note that the above error covariance operator models correspond to define, for $i=1,\dots,n$, the functional Gaussian error component ε_i as the solution, in the mean–square sense, of the stochastic partial differential equation

$$(-\boldsymbol{\Delta}_{D_l})^{(d-\gamma_i)}\varepsilon_i = \xi_i, \quad \gamma_i \in (0, d/2),$$

with ξ_i being spatial Gaussian white noise on $L^2(D_l)$, for $l=1,2,3$.

To approximate

$$\mathrm{FMSE}_{\boldsymbol{\beta}} = \mathrm{E}\left\{\|\boldsymbol{\beta}(\cdot) - \widehat{\boldsymbol{\beta}}(\cdot)\|_{H^p}^2\right\},$$

ν samples are generated for the computation of

$$\mathrm{EFMSE}_{\boldsymbol{\beta}} = \sum_{v=1}^{\nu} \frac{\sum_{s=1}^{p} \|\boldsymbol{\beta}_s^v(\cdot) - \widehat{\boldsymbol{\beta}}_s^v(\cdot)\|_H^2}{\nu}, \tag{A4.23}$$

the empirical functional mean–square error $\mathrm{EFMSE}_{\boldsymbol{\beta}}$ associated with the functional estimates

$$\left\{\widehat{\boldsymbol{\beta}}_s^v(\cdot) = \left(\widehat{\beta}_s^v(x_1, y_1), \dots, \widehat{\beta}_s^v(x_L, y_L)\right),\ s = 1, \dots, p,\ v = 1, \dots, \nu\right\}$$

of $\boldsymbol{\beta}$, where L is the number of nodes considered in the regular grid constructed over the domains $\{D_l,\ l = 1, 2, 3\}$.

Also, we will compute the following statistics:

$$L_{\boldsymbol{\beta}}^{\infty}(\cdot) = \sum_{v=1}^{\nu} \frac{\left(\|\boldsymbol{\varepsilon}_{\boldsymbol{\beta},v}^2(x_1, y_1)\|_{\infty}, \dots, \|\boldsymbol{\varepsilon}_{\boldsymbol{\beta},v}^2(x_L, y_L)\|_{\infty}\right)}{\nu},$$

where

$$\boldsymbol{\varepsilon}_{\boldsymbol{\beta},v}^2(x_j, y_j) = \left(\varepsilon_{\beta_1^v}^2(x_j, y_j), \dots, \varepsilon_{\beta_p^v}^2(x_j, y_j)\right), \quad j = 1, \dots, L,$$

and

$$\varepsilon_{\boldsymbol{\beta}_s^v}(x_j, y_j) = \beta_s^v(x_j, y_j) - \widehat{\beta}_s^v(x_j, y_j), \quad s = 1, \dots, p, \quad j = 1, \dots, L, \quad v = 1, \dots, \nu,$$

with $\|\cdot\|_{\infty}$ denoting the L^{∞}–norm.

Let

$$\{\mathbf{Y}_i^v(\cdot) = (Y_i^v(x_1, y_1), \dots, Y_i^v(x_L, y_L)),\ i = 1, \dots, n,\ v = 1, \dots, \nu\}$$

be the generated functional samples. The empirical approximation of

$$\mathrm{FMSE}_{\mathbf{Y}} = \mathrm{E}\left\{\|\mathbf{Y}(\cdot) - \widehat{\mathbf{Y}}(\cdot)\|_{H^n}^2\right\},$$

with $\mathrm{FMSE}_{\mathbf{Y}}$ being the FMSE of $\mathbf{Y}$, can be computed as follows:

$$\mathrm{EFMSE}_{\mathbf{Y}} = \sum_{v=1}^{\nu} \frac{\sum_{i=1}^{n} \|\mathbf{Y}_i^v(\cdot) - \widehat{\mathbf{Y}}_i^v(\cdot)\|_H^2}{\nu}. \tag{A4.24}$$

Also, we will consider the statistics

$$L_{\mathbf{Y}}^{\infty}(\cdot) = \sum_{v=1}^{\nu} \frac{\left(\|\boldsymbol{\varepsilon}_{\mathbf{Y},v}^2(x_1, y_1)\|_{\infty}, \dots, \|\boldsymbol{\varepsilon}_{\mathbf{Y},v}^2(x_L, y_L)\|_{\infty}\right)}{\nu},$$

where

$$\boldsymbol{\varepsilon}^2_{\mathbf{Y},v}(x_j,y_j) = \left(\varepsilon^2_{\mathbf{Y}_1^v}(x_j,y_j),\dots,\varepsilon^2_{\mathbf{Y}_n^v}(x_j,y_j)\right), \quad \boldsymbol{\varepsilon}_{\mathbf{Y}_i^v}(x_j,y_j) = \mathbf{Y}_i^v(x_j,y_j) - \widehat{\mathbf{Y}}_i^v(x_j,y_j),$$

for $i=1,\dots,n$, $j=1,\dots,L$, and $v=1,\dots,\nu$.

In the following numerical examples, the functional analysis of variance is implemented from a transformed functional data model, considering the matrix operator $\mathbf{W}$ such that, for each $k \geq 1$, $\mathbf{\Phi}_k^*\mathbf{W} = \mathbf{W}_k$ compensates the divergence of the eigenvalues of $\mathbf{\Lambda}_k^{-1}$. Thus, condition (A4.14) is satisfied. Hence, for all $k \geq 1$, $\mathbf{W}_k$ can be defined as

$$\mathbf{W}_k = \mathbf{\Psi}_k \mathbf{\Omega}(\mathbf{W}_k)\mathbf{\Psi}_k^T, \tag{A4.25}$$

where $\mathbf{\Omega}(\mathbf{W}_k) = diag(\omega_{k11},\dots,\omega_{knn})$ denoting a diagonal matrix, which elements are defined by

$$w_{kii} = \omega_i(\mathbf{\Lambda}_k) + \frac{1}{a_k},$$

under

$$\sum_{k=1}^{\infty} \frac{1}{a_k} < \infty.$$

We have chosen $a_k = k^2$. Here, for each $k \geq 1$, $\mathbf{\Psi}_k$ denotes the projection operator into the system $\{\psi_{lk}, l=1,\dots,n\}$ of eigenvectors of matrix $\mathbf{\Lambda}_k$, and $\{\omega_i(\mathbf{\Lambda}_k),\ i=1,\dots,n\}$ are the associated eigenvalues (see Ruiz-Medina [2016]).

In practice, the infinite series defining the generalized least–squares estimator, and the functional components of variance is truncated at TR. Specifically, in the rectangle, we work with a two–dimensional truncation parameter $TR = TR_1 \times TR_2$, and, for circular domains, we fix a one–dimensional parameter (the order k of Bessel functions), thus, $TR_1 = 1$, and move the second truncation parameter associated with the radius R (see Appendices A4.7.2–A4.7.3). We then have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &\simeq \mathbf{\Phi}\left(\left\{\widehat{\boldsymbol{\beta}_k},\ k=1,\dots,TR\right\}\right), && \text{(A4.26)}\\
\widetilde{SSE} &\simeq \sum_{k=1}^{TR} (\mathbf{M}_k\mathbf{W}_k\mathbf{Y}_k)^T \mathbf{\Lambda}_k^{-1}\mathbf{M}_k\mathbf{W}_k\mathbf{Y}_k, && \text{(A4.27)}\\
\widetilde{SST} &\simeq \sum_{k=1}^{TR} \mathbf{Y}_k^T\mathbf{W}_k^T\mathbf{\Lambda}_k^{-1}\mathbf{W}_k\mathbf{Y}_k, && \text{(A4.28)}\\
\widetilde{SSR} &= \widetilde{SST} - \widetilde{SSE}, && \text{(A4.29)}\\
\mathbf{\Lambda}_k &= \mathbf{\Psi}_k\mathbf{\Omega}(\mathbf{\Lambda}_k)\mathbf{\Psi}_k^T, \quad k=1,\dots,TR, && \text{(A4.30)}\\
\mathbf{W}_k &= \mathbf{\Psi}_k\mathbf{\Omega}(\mathbf{W}_k)\mathbf{\Psi}_k^T, \quad k=1,\dots,TR. && \text{(A4.31)}
\end{aligned}$$

From the transformed model (A4.13), the finite–dimensional approximations (A4.27)–(A4.31) of $\widetilde{SSE}$,

$\widetilde{SST}$, and $\widetilde{SSR}$, respectively, are computed to obtain the values of the statistics (A4.15), reflecting the relative magnitude between the empirical functional variability explained by the model and the residual variability.

In the computation of the test statistics $T_{\boldsymbol{h}}$, a truncation order is also considered in the calculation of the elements defining matrix $\boldsymbol{\Lambda}_{\boldsymbol{h}}$.

In all the subsequent sections, the truncation order TR has been selected according to the following criteria:

(i) The percentage of explained functional variance. In all the subsequent numerical examples, the TR values considered always ensure a percentage of explained functional variance larger or equal than 95%.

(ii) The rate of convergence to zero of the eigenvalues of the covariance operators, defining the functional entries of the matrix covariance operator of the H^n–valued error term. Specifically, in the simulation study undertaken, according to the asymptotic order (rate of convergence to zero) of such eigenvalues, we have selected the optimal TR to remove divergence of the spectra of the corresponding inverse covariance operators.

(iii) The functional form of the eigenvectors, depending on the geometry of the domain and the Dirichlet conditions on the boundary. Small truncation orders or values of TR are considered, when fast decay velocity to zero is displayed at the boundary, by the common eigenvectors of the autocovariance operators of the error components, since, in that case, the error dependence range is shorter.

Summarizing, lower truncation orders are required when a fast decay velocity to zero is displayed by the covariance kernel eigenvalues, since a sufficient percentage of explained variability is achieved with a few terms. Note that larger truncation orders can lead to a ill–posed nature of the functional parameter estimation problem, and associated response plug-in prediction. In the subsequent sections, applying criteria (i)–(iii), a smaller number of terms is required in circular domains than in rectangular domains.

A4.4.1 Rectangular domain

The H^n–valued zero–mean Gaussian error term is generated from the matrix covariance operator $\boldsymbol{R}_{\boldsymbol{\varepsilon\varepsilon}}$, whose functional entries $\left\{\boldsymbol{R}_{\varepsilon_i\varepsilon_j}\right\}_{i=1,\dots,n}^{j=1,\dots,n}$, are defined in equation (A4.22), with for $i = 1,\dots,n$, $\lambda_{ki} = \lambda_k(R_{ii})$ being given in equations (A4.8) and (A4.36). Specifically, $\{\phi_k,\ k \geq 1\}$ are the eigenvectors of the Dirichlet negative Laplacian operator on the rectangle, associated with the eigenvalues of such an operator (see equation (A4.36) in the Supplementary Material in Appendix A4.7), arranged in decreasing order of their modulus magnitude.

Let us now define the scenarios studied for the rectangular domain

$$D_1 = \prod_{i=1}^{2}\left[a_i, b_i\right],$$

where $\nu = 20$ functional samples of size $n = 200$ have been considered, for a given semi–orthogonal design matrix

$$\mathbf{X} \in \mathbb{R}^{n \times p}, \quad \mathbf{X}^T\mathbf{X} = \mathbf{Id}_p.$$

These scenarios are determined from the possible values of the vector variable (P_i, u, C_i), where P_i refers to the number of components of $\boldsymbol{\beta}$, specifically, for $i = 1, p = 4$ components, and for $i = 2, p = 9$ components. Here, u takes the values a, b, c, d respectively corresponding to the truncation orders $TR = 16$ $(u = a)$, $TR = 36$ $(u = b)$, $TR = 64$ $(u = c)$ and $TR = 144$ $(u = d)$. In addition, $\{C_i,\ i = 1, 2\}$ indicate the shape of $\boldsymbol{\beta}$. Specifically, we have considered

- $\beta_s\,(x, y) = \sin\left(\frac{\pi s x_{b_1}}{l_1}\right)\sin\left(\frac{\pi s y_{b_2}}{l_2}\right)$ (C1)
- $\beta_s\,(x, y) = \cos\left(\frac{x_{b_1}+x_{a_1}}{l_1}\right)\cos\left(\frac{y_{b_2}+y_{a_2}}{l_2}\right)$ (C2),

where

$$x_{b_1} = \frac{\pi}{2}\,(2s+1)\,(b_1 - x)\,,\ x_{a_1} = (x - a_1)\,, \quad y_{b_2} = \frac{\pi}{2}\,(2s+1)\,(b_2 - y)\,, \quad y_{a_2} = (y - a_2)$$

and $s = 1, \dots, p$.

A summary of the generated and analysed scenarios are displayed in Table A4.4.1 below.

Table A4.4.1: Scenarios for rectangular domain.

Cases	$a_1 = a_2$	$b_1 = b_2$	$h_x = h_y$	p	TR
$(\mathrm{P}_1,\mathrm{a},\mathrm{C}_1)$	-2	3	0.05	4	4×4
$(\mathrm{P}_1,\mathrm{b},\mathrm{C}_2)$	-2	3	0.05	4	6×6
$(\mathrm{P}_1,\mathrm{c},\mathrm{C}_2)$	-2	3	0.05	4	8×8
$(\mathrm{P}_1,\mathrm{d},\mathrm{C}_1)$	-2	3	0.05	4	12×12
$(\mathrm{P}_2,\mathrm{a},\mathrm{C}_2)$	-2	3	0.05	9	4×4
$(\mathrm{P}_2,\mathrm{b},\mathrm{C}_1)$	-2	3	0.05	9	6×6
$(\mathrm{P}_2,\mathrm{c},\mathrm{C}_1)$	-2	3	0.05	9	8×8
$(\mathrm{P}_2,\mathrm{d},\mathrm{C}_2)$	-2	3	0.05	9	12×12

In Table A4.4.1, h_x and h_y refer to the discretization step size at each dimension. In the cases $(\mathrm{P}_1,\mathrm{a},\mathrm{C}_1)$ and $(\mathrm{P}_2,\mathrm{a},\mathrm{C}_2)$, a generation of a functional value (surface) of the response is respectively represented in Figures A4.4.1–A4.4.2.

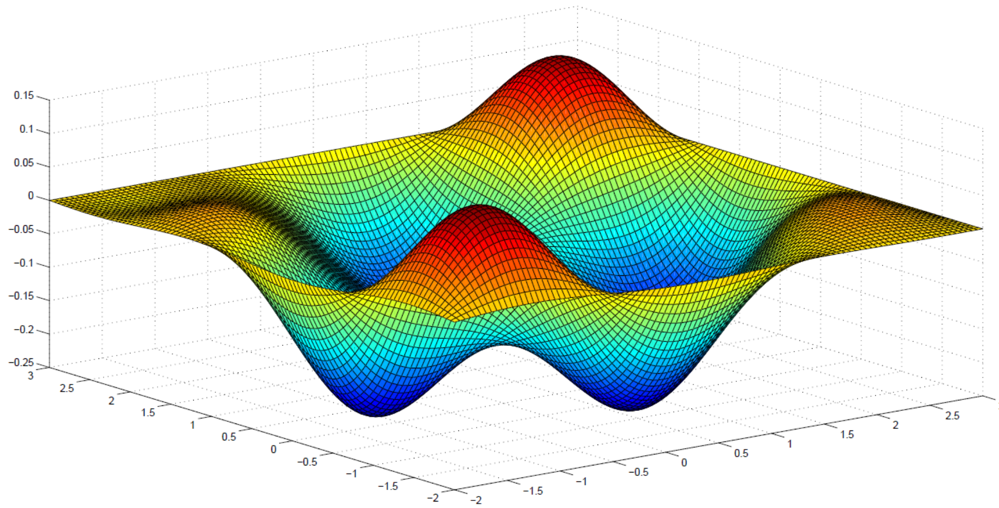


Figure A4.4.1: Case (P_1,a,C_1). Simulated response with $p = 4$, $TR = 16$ and $\boldsymbol{\beta}$ of type C_1.

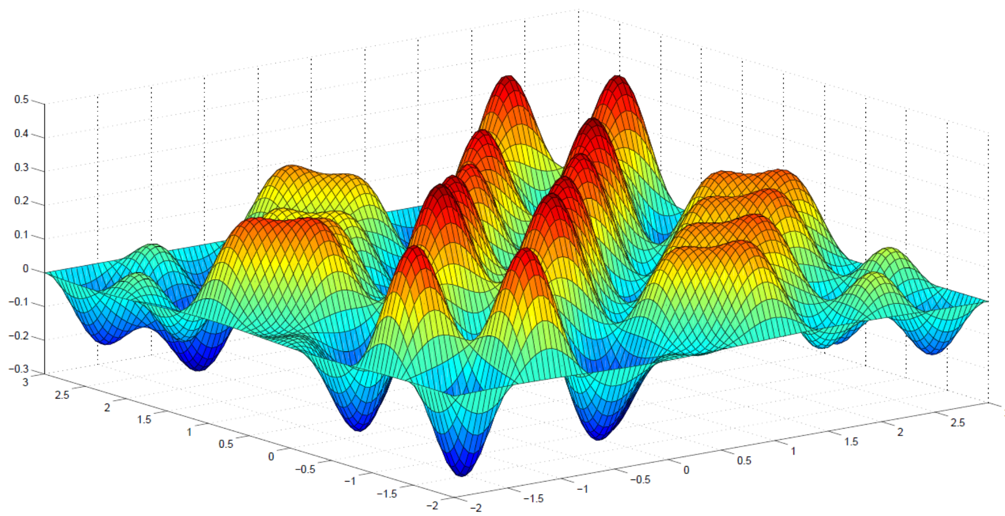


Figure A4.4.2: Case (P_2,a,C_2). Simulated response with $p = 9$, $TR = 16$ and $\boldsymbol{\beta}$ of type C_2.

Figures A4.4.3–A4.4.4 below show the respective functional estimates $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ of the responses displayed in Figures A4.4.1–A4.4.2 above.

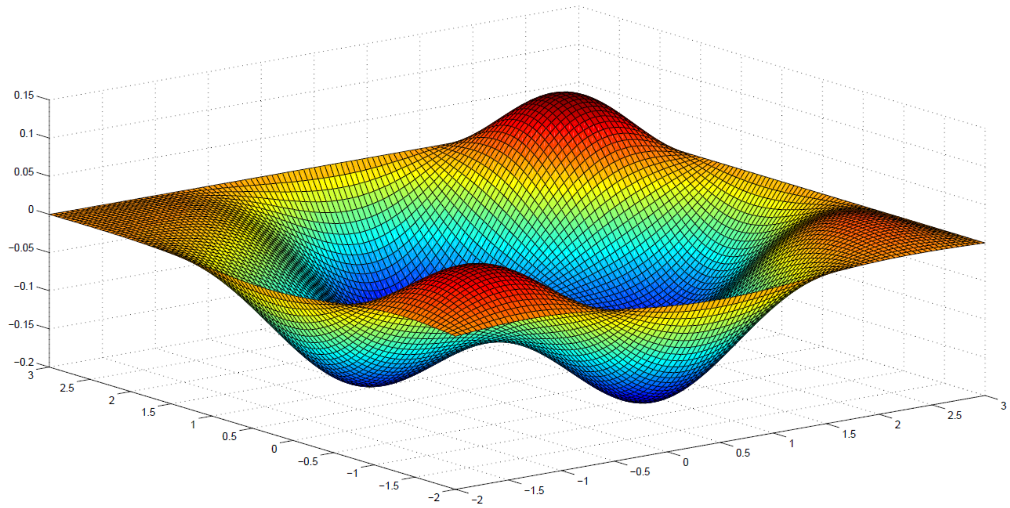


Figure A4.4.3: Case (P$_1$,a,C$_1$). Estimated response with $p = 4$, $TR = 16$ and $\boldsymbol{\beta}$ of type C$_1$.

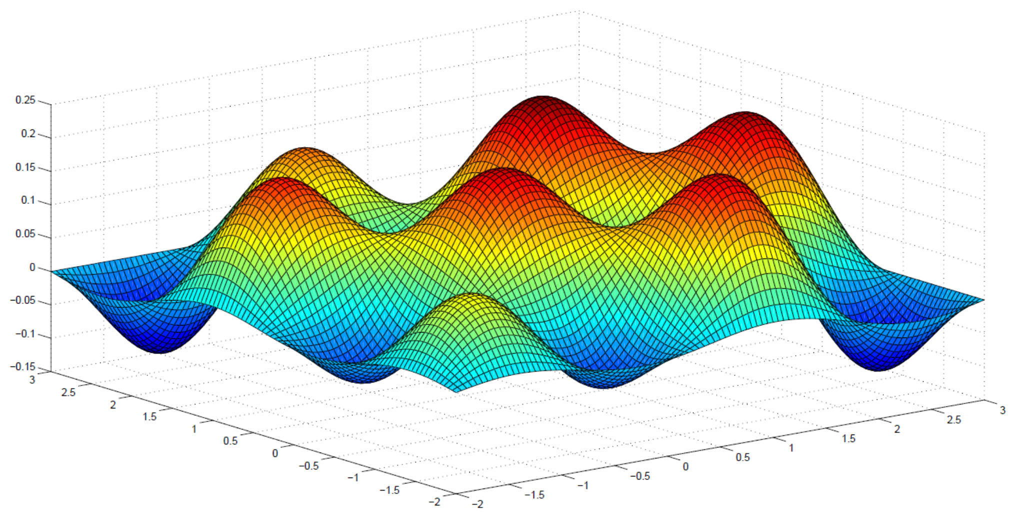


Figure A4.4.4: Case (P$_2$,a,C$_2$). Estimated response with $p = 9$, $TR = 16$ and $\boldsymbol{\beta}$ of type C$_2$.

The statistics (A4.23)–(A4.24) are evaluated in all the cases displayed in Table A4.4.1 (see Tables A4.4.2–A4.4.3 for the statistics $L^{\infty}_{\boldsymbol{\beta}}$ and $L^{\infty}_{\mathbf{Y}}$, respectively).

Table A4.4.2: $EFMSE_{\boldsymbol{\beta}}$ for rectangular domain.

$EFMSE_{\boldsymbol{\beta}}$			
(P_1,a,C_1)	(P_1,b,C_2)	(P_1,c,C_2)	(P_1,d,C_2)
$1.070\,(10)^{-3}$	$1.060\,(10)^{-3}$	$1.040\,(10)^{-3}$	$1.040\,(10)^{-3}$
(P_2,a,C_2)	(P_2,b,C_1)	(P_2,c,C_1)	(P_2,d,C_2)
$9.400\,(10)^{-4}$	$9.300\,(10)^{-4}$	$9.300\,(10)^{-4}$	$9.100\,(10)^{-4}$

Table A4.4.3: $EFMSE_{\mathbf{Y}}$ for rectangular domain.

$EFMSE_{\mathbf{Y}}$							
(P_1,a,C_1)	(P_1,b,C_2)	(P_1,c,C_2)	(P_1,d,C_2)	(P_2,a,C_2)	(P_2,b,C_1)	(P_2,c,C_1)	(P_2,d,C_2)
0.014	0.013	0.010	0.009	0.011	0.011	0.009	0.007

As expected, the results displayed in Table A4.4.2, corresponding to the empirical functional mean quadratic errors associated with the estimation of $\boldsymbol{\beta}$, are less than the ones obtained in Table A4.4.3 for the response, with order of magnitude 10^{-3} in all the scenarios generated. In Table A4.4.3, we can appreciate a better performance of the generalized least–squares estimator for the higher truncation orders. However, we have to note that, even for the smallest truncation order considered; i.e., for $TR = 4 \times 4 = 16$, a good performance is observed with associated empirical functional mean quadratic errors having order of magnitude 10^{-2} in all the cases displayed in Table A4.4.1 (see the above truncation order criteria (i)–(iii)). It can also be observed that the number of components of parameter $\boldsymbol{\beta}$, and their functional shapes do not affect the accuracy of the least–squares generalized estimations of the functional values of the response. It can also be observed that the number of components of parameter $\boldsymbol{\beta}$, and their functional shapes do not affect the accuracy of the least–squares generalized estimations of the functional values of the response.

The statistics (A4.15) is now computed, as an empirical approximation of the relative magnitude between the explained functional variability and the residual variability, after fitting the transformed Hilbert–valued fixed effect model (A4.13). The results obtained are given in Table A4.4.4. It can be observed that, in all the cases studied, the explained functional variability exceeds the residual functional variability. The truncation order, the number of components of $\boldsymbol{\beta}$, and the functional shape of such components do not substantially affect the goodness of fit of the transformed Hilbert–valued fixed effect model in (A4.13).

Table A4.4.4: F statistics (A4.15) for rectangular domain.

Cases	(P_1,a,C_1)	(P_1,b,C_2)	(P_1,c,C_2)	(P_1,d,C_1)	(P_2,a,C_2)	(P_2,b,C_1)	(P_2,c,C_1)	(P_2,d,C_2)
F	1.926	1.717	1.673	1.626	1.898	1.845	1.761	1.606

Let us now compute the statistics T_h in (A4.20) to contrast the significance of parameter vector $\boldsymbol{\beta}$ in Case C_1, when $p = 4$. To apply [Cuesta-Albertos et al., 2007, Theorem 4.1] and [Cuesta-Albertos and

Febrero-Bande, 2010, Theorem 2.1], we have generated eight realizations of a Gaussian random function $\boldsymbol{h}$, from the trajectories of the Gaussian random field ξ, solution, in the mean–square sense, of the following boundary value problem:

$$\begin{aligned} (-\Delta)\xi(\mathbf{x}) &= \varsigma(\mathbf{x}), \quad \mathbf{x} = (x_1, x_2) \in [-2,3] \times [-2,3], \\ \xi(-2, x_2) &= \xi(3, x_2) = \xi(x_1, -2) = \xi(x_1, 3) = 0, \quad x_1, x_2 \in [-2,3] \times [-2,3], \end{aligned} \tag{A4.32}$$

where ς denotes a zero–mean Gaussian white noise on $L^2([-2,3] \times [-2,3])$; i.e., a zero–mean generalized Gaussian process satisfying

$$\int_{[-2,3]\times[-2,3]} f(\mathbf{x}) \mathrm{E}\left\{\varsigma(\mathbf{y})\varsigma(\mathbf{x})\right\} d\mathbf{x} = f(\mathbf{y}), \quad \mathbf{y} \in [-2,3] \times [-2,3], \quad \forall f \in L^2([-2,3] \times [-2,3]).$$

Table A4.4.5 below reflects the percentage of successes, for $\alpha = 0.05$, and the averaged p–values over the 150 samples of the response generated with parameter $\boldsymbol{\beta}$ of C_1 type having $p = 4$ components, and with size $n = 150$, for $TR = 4 \times 4$.

Table A4.4.5: *Rectangle*. Percentage of successes for $\alpha = 0.05$, at the left–hand side, and averaged p–values at the right–hand side, for each one of the eight realizations considered of the Gaussian function $h \in L^2([-2,3] \times [-2,3])$.

D	$\%$ Success	p
1	100%	0
2	100%	0
3	99.75%	$1.998(10)^{-8}$
4	100%	0
5	99.8%	$7.541(10)^{-7}$
6	100%	0
7	100%	0
8	100%	$6.441(10)^{-10}$

A high percentage of successes and very small p–values are observed in Table A4.4.5; i.e., a good performance of the test statistics is observed.

A4.4.2 Disk domain

In the disk domain

$$D_2 = \left\{\mathbf{x} \in \mathbb{R}^2 : \, 0 < \|\mathbf{x}\| < R\right\},$$

the zero–mean Gaussian H^n–valued error term is generated from the matrix covariance operator $\boldsymbol{R}_{\varepsilon\varepsilon}$, whose functional entries are defined in equation (A4.22), considering the eigenvectors $\{\phi_k,\ k \geq 1\}$ of the Dirichlet negative Laplacian operator on the disk (see equation (A4.37) in the Supplementary Material in Appendix A4.7), arranged in decreasing order of the modulus magnitude of their associated eigenvalues. Specifically, for $i = 1, \ldots, n$, $\lambda_{ki} = \lambda_k(R_{ii})$ in (A4.22) is defined in equations (A4.8) and (A4.37). Again, $\nu = 20$ functional samples of size $n = 200$ of the response have been generated. The cases studied are summarized in terms of the vector (P_i, u, C_j), $i = 1, 2$, $j = 1, 2, 3$, with variable $u = a, b, c, d, e, f$. Namely, it is considered $u = a$ for $TR = 3$, $u = b$ for $TR = 5$, $u = c$ for $TR = 7$, $u = d$ for $TR = 15$, $u = e$ for $TR = 31$, and $u = f$ for $TR = 79$. Furthermore, P_i indicates the number of components of $\boldsymbol{\beta}$, with $p = 4$ for $i = 1$, and $p = 9$ for $i = 2$. Finally, the values of C_j, $j = 1, 2, 3$, refer to the shape of the components of $\boldsymbol{\beta}$, defined from their projections, in terms of the following equations:

$$
\begin{aligned}
\beta_{ks} &= \frac{(-1)^s}{k^{3.5}} e^{\left(\frac{k}{TR}\right)^{7.5+2s-1}} P(s,k)^{2.5+2s-1} \\
&+ e^{\left(\frac{k}{TR}\right)^{6.5+2s-1}} P(s,k)^{3.5+2s-1}, \quad k = 1, \ldots, TR, \quad s = 1, \ldots, p \quad \text{(C1)} \\
\beta_{ks} &= \frac{1}{R} e^{\frac{s+\frac{k}{R}}{n}} + k\cos\left((-1)^k 2\pi \frac{R}{k}\right), \quad k = 1, \ldots, TR, \quad s = 1, \ldots, p \quad \text{(C2)} \\
\beta_{ks} &= \frac{1}{k^{2.5+2s-1}} P(s,k)^{1.5+2s-1}, \quad k = 1, \ldots, TR, \quad s = 1, \ldots, p \quad \text{(C3)} \\
P(s,k) &= 1 + \left(\frac{k}{TR}\right)^2 + \left(\frac{TR-k+1}{TR}\right)^4, \quad k = 1, \ldots, TR, \quad s = 1, \ldots, p.
\end{aligned}
$$

Table A4.4.6 reflects a summary with all the cases analysed.

Table A4.4.6: Scenarios for disk domain.

Cases	R	h_R	h_ϕ	TR	p
(P_1,a,C_3)	12	$\frac{R}{145}$	$\frac{2\pi}{135}$	3	4
(P_1,b,C_2)	18	$\frac{R}{145}$	$\frac{2\pi}{135}$	5	4
(P_1,c,C_1)	25	$\frac{R}{145}$	$\frac{2\pi}{135}$	7	4
(P_1,d,C_1)	50	$\frac{R}{145}$	$\frac{2\pi}{135}$	15	4
(P_1,e,C_2)	100	$\frac{R}{145}$	$\frac{2\pi}{135}$	31	4
(P_1,f,C_3)	250	$\frac{R}{145}$	$\frac{2\pi}{135}$	79	4
(P_2,a,C_1)	12	$\frac{R}{145}$	$\frac{2\pi}{135}$	3	9
(P_2,b,C_2)	18	$\frac{R}{145}$	$\frac{2\pi}{135}$	5	9
(P_2,c,C_3)	25	$\frac{R}{145}$	$\frac{2\pi}{135}$	7	9
(P_2,d,C_3)	50	$\frac{R}{145}$	$\frac{2\pi}{135}$	15	9
(P_2,e,C_2)	100	$\frac{R}{145}$	$\frac{2\pi}{135}$	31	9
(P_2,f,C_1)	250	$\frac{R}{145}$	$\frac{2\pi}{135}$	79	9

Figures A4.4.5–A4.4.6 respectively reflect the generation of a functional value of the response in the cases (P_1,c,C_1) and (P_1,f,C_3).

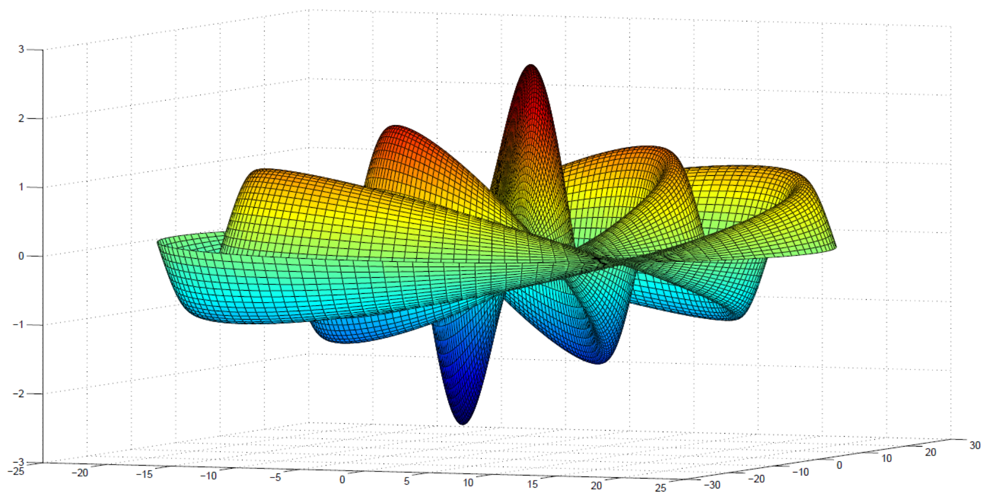


Figure A4.4.5: Case (P_1,c,C_1). Simulated response with $p = 4$, $R = 25$ and $\boldsymbol{\beta}$ of type C_1.

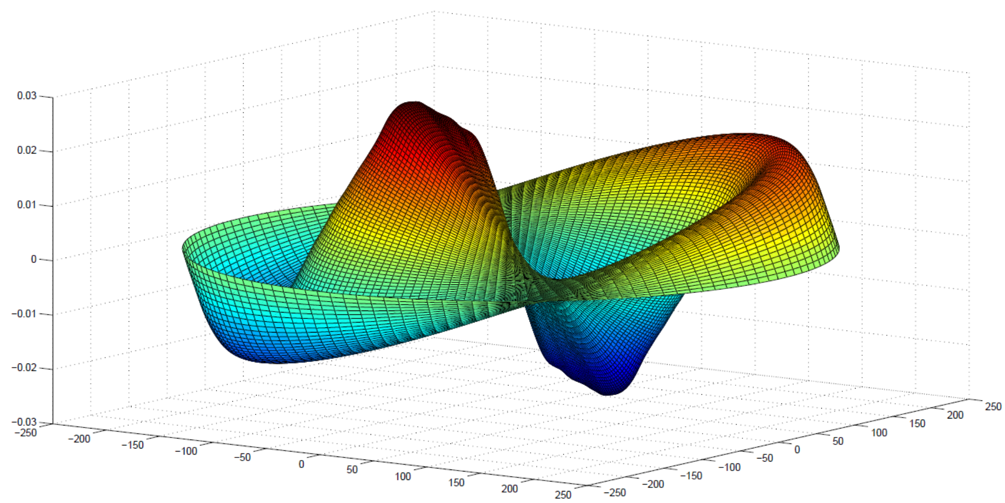


Figure A4.4.6: Case (P_1,f,C_3). Simulated response with $p = 4$, $R = 250$ and $\boldsymbol{\beta}$ of type C_3.

The respective generalized least–squares functional estimates are displayed in Figures A4.4.7–A4.4.8.

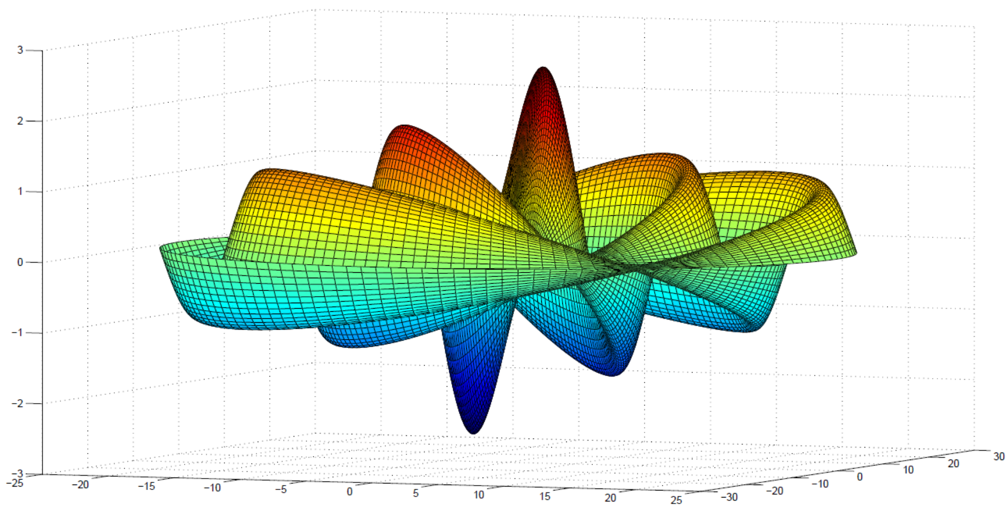


Figure A4.4.7: Case (P_1,c,C_1). Estimated response with $p = 4$, $R = 25$ and $\boldsymbol{\beta}$ of type C_1.

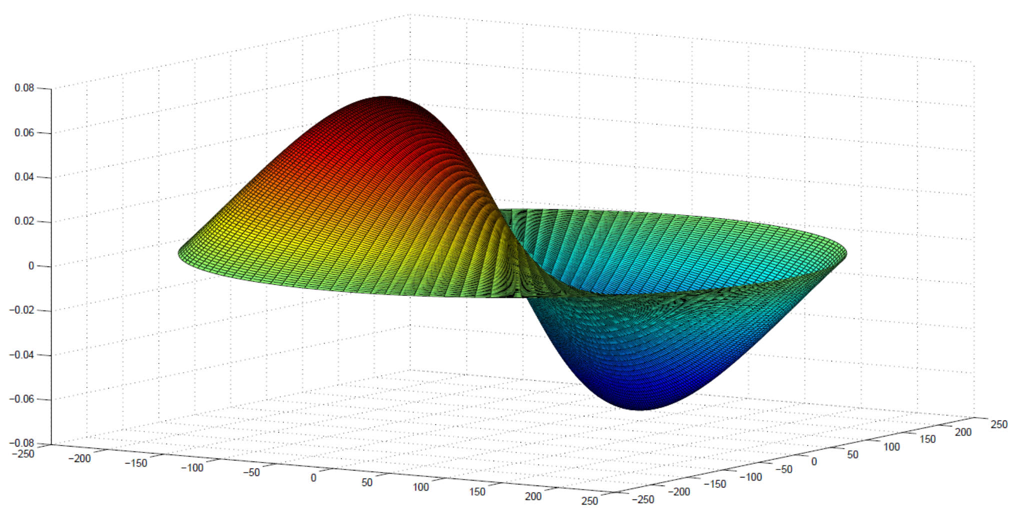


Figure A4.4.8: Case (P_1,f,C_3). Estimated response with $p = 4$, $R = 250$ and $\boldsymbol{\beta}$ of type C_3.

The empirical functional mean quadratic errors (see equations (A4.23)–(A4.24)) are displayed in Table A4.4.7, for the estimation of the functional parameter vector $\boldsymbol{\beta}$, and in Table A4.4.8 for the estimation of the response $\mathbf{Y}$. It can be observed, as in the rectangular domain, that the order of magnitude of the empirical functional mean quadratic errors associated with the estimation of $\boldsymbol{\beta}$ is of order 10^{-3}, and for the estimation of the response is 10^{-2}. However, the number of terms considered is less than in the case of the rectangle; i.e., a finite dimensional space with lower dimension than in the rectangle is required, according to criterion (iii) reflected in Appendix A4.4. It can also be appreciated that the number of components of $\boldsymbol{\beta}$ does not substantially affect the accuracy of the estimates.

Table A4.4.7: $EFMSE_{\boldsymbol{\beta}}$ for disk domain.

$EFMSE_{\boldsymbol{\beta}}$

(P_1,a,C_3)	(P_1,b,C_2)	(P_1,c,C_1)
$7.500\,(10)^{-4}$	$7.500\,(10)^{-4}$	$7.400\,(10)^{-4}$
(P_1,d,C_1)	(P_1,e,C_2)	(P_1,f,C_3)
$7.500\,(10)^{-4}$	$7.600\,(10)^{-4}$	$7.500\,(10)^{-4}$
(P_2,a,C_1)	(P_2,b,C_2)	(P_2,c,C_3)
$7.000\,(10)^{-4}$	$7.100\,(10)^{-4}$	$7.100\,(10)^{-4}$
(P_2,d,C_3)	(P_2,e,C_2)	(P_2,f,C_1)
$7.900\,(10)^{-4}$	$8.000\,(10)^{-4}$	$8.000\,(10)^{-4}$

Table A4.4.8: $EFMSE_{\mathbf{Y}}$ for disk domain.

$EFMSE_{\mathbf{Y}}$

(P_1,a,C_3)	(P_1,b,C_2)	(P_1,c,C_1)	(P_1,d,C_1)	(P_1,e,C_2)	(P_1,f,C_3)
0.048	0.048	0.048	0.048	0.048	0.048
(P_2,a,C_1)	(P_2,b,C_2)	(P_2,c,C_3)	(P_2,d,C_3)	(P_2,e,C_2)	(P_2,f,C_1)
0.050	0.050	0.050	0.049	0.050	0.050

The statistics (A4.15) is now computed (see Table A4.4.9), as an empirical approximation of the relative magnitude between the explained functional variability and the residual variability, after fitting the transformed Hilbert-valued fixed effect model (A4.13). It can be noticed that the values of $\frac{\widetilde{SSR}}{\widetilde{SST}}$ are very close to one in all the scenarios analysed. This fact induces large values of (A4.15) (see Table A4.4.9), since

$$F = \frac{\widetilde{SSR}}{\widetilde{SSE}} = \frac{\widetilde{SSR}/\widetilde{SST}}{1 - \widetilde{SSR}/\widetilde{SST}}.$$

It can be observed, one time more, from criterion (iii), reflected in Appendix A4.4, that the boundary conditions and the geometry of the domain allows in this case a more substantial dimension reduction than in the rectangular domain case, since with lower truncation orders a better model fitting is obtained.

Table A4.4.9: F statistics (A4.15) over the disk domain.

Cases	(P_1,a,C_3)	(P_1,b,C_2)	(P_1,c,C_1)
F	$1.100(10^2)$	$4.100(10^3)$	$1.200(10^5)$
Cases	(P_1,d,C_1)	(P_1,e,C_2)	(P_1,f,C_3)
F	$3.900(10^6)$	$6.300(10^6)$	$4.200(10^6)$
Cases	(P_2,a,C_1)	(P_2,b,C_2)	(P_2,c,C_3)
F	$2.200(10^3)$	$8.200(10^3)$	$7.600(10^7)$
Cases	(P_2,d,C_3)	(P_2,e,C_2)	(P_2,f,C_1)
F	$2.500(10^7)$	$1.400(10^7)$	$8.500(10^7)$

The statistics $T_{\boldsymbol{h}}$ in (A4.20) is computed to contrast the significance of the parameter vector $\boldsymbol{\beta}$ in case C_1, with $p = 4$ components. Again, eight realizations of Gaussian random functions $\boldsymbol{h}$ are considered, generated from a Gaussian random field ξ, solution, in the mean–square sense, of the following boundary value problem on the disk:

$$\begin{aligned}(-\Delta)\xi(\mathbf{x}) &= \varsigma(\mathbf{x}), \quad \mathbf{x} = (x_1, x_2) \in D_{25} = \{\mathbf{x} \in \mathbb{R}^2;\, 0 < \|\mathbf{x}\| < 25\},\\ \xi(\theta, 25) &= 0, \quad \forall\theta \in [0, 2\pi]\end{aligned}$$

where ς denotes a zero–mean Gaussian white noise on $L^2(D_{25})$; i.e., a zero–mean generalized Gaussian process satisfying

$$\int_{[0,2\pi]\times[0,25]} f(\varphi, v)\mathrm{E}\left\{\varsigma(\theta, r)\varsigma(\varphi, v)\right\} d\varphi dv = f(\theta, r), \quad (\theta, r) \in [0, 2\pi] \times [0, 25], \quad f \in L^2(D_{25}).$$

Table A4.4.10 reflects the percentage of successes, for $\alpha = 0.05$, and the averaged p–values over the 150 samples, generated with size $n = 150$, of the functional response having parameter vector $\boldsymbol{\beta}$ of type C_1 with $p = 4$ components, for $TR = 7$.

Table A4.4.10: *Disk*. Percentage of successes for $\alpha = 0.05$, at the left–hand side, and averaged p–values at the right–hand side, for each one of the eight realizations of the Gaussian function $h \in L^2(D_{25})$.

D	% Success	p
1	99.95%	$1.672(10)^{-8}$
2	99.5%	$9.746(10)^{-7}$
3	100%	0
4	99.9%	$8.546(10)^{-8}$
5	97.45%	$7.400(10)^{-7}$
6	100%	0
7	100%	$8.775(10)^{-9}$
8	100%	0

Table A4.4.10 again illustrates a good performance of the statistics $T_{\boldsymbol{h}}$ in (A4.20). Indeed, we can appreciate a high percentage of successes, and very small p–values, very close to zero, that support the significance of the functional parameter vector, considered in the generation of the data set analysed.

A4.4.3 Circular sector domain

In the circular sector

$$D_3 = \{(r\cos(\varphi), \, r\sin(\varphi)) : \, 0 < \|r\| < R, \, 0 < \varphi < \pi\theta\}$$

of radius R and angle $\pi\theta$, the zero-mean Gaussian vector error term is generated from the matrix covariance operator $\boldsymbol{R}_{\varepsilon\varepsilon}$, whose functional entries are defined in equation (A4.22). The eigenvectors $\{\phi_k, \; k \geq 1\}$ of the Dirichlet negative Laplacian operator on the circular sector are considered (see equation (A4.39) in the Supplementary Material in Appendix A4.7), arranged in decreasing order of the modulus magnitude of their associated eigenvalues. Specifically, here, $\boldsymbol{R}_{\varepsilon\varepsilon}$ is defined in equation (A4.22), with for $i = 1, \dots, n$, $\lambda_{ki} = \lambda_k(R_{ii})$ being given in equations (A4.8) and (A4.39).

As in the above examples, $\nu = 20$ functional samples of size $n = 200$ are generated. The cases studied are also summarized in terms of the values of the vector (P_i, u, C_j), $i = 1, 2$, $u = a, b, c, d, e, f$, and $j = 1, 2, 3$, with the values of u having the same meaning as in the disk domain. Again, values of P_i provide the number p of components of $\boldsymbol{\beta}$; i.e., $p = 4$ if $i = 1$, and $p = 9$ if $i = 2$. The values C_1, C_2 and C_3 respectively correspond to the following functions defining the components of $\boldsymbol{\beta}$, whose projections are given by:

$$\begin{aligned}
\beta_{sk} &= 1 + (k-1)s, \quad k = 1, \dots, TR, \quad s = 1, \dots, p \quad (\text{C1}) \\
\beta_{sk} &= \frac{1}{R} e^{\frac{s + \frac{k}{R}}{n}} + k\cos\left((-1)^k \, 2\pi \frac{R}{k}\right), \quad k = 1, \dots, TR, \quad s = 1, \dots, p \quad (\text{C2}) \\
\beta_{sk} &= \cos\left(\pi \frac{TR - k}{k}\right) \cos\left(\pi \frac{p - s}{s}\right), \quad k = 1, \dots, TR, \quad s = 1, \dots, p \quad (\text{C3}).
\end{aligned}$$

A summary of the cases analysed is given in Table A4.4.11.

Table A4.4.11: Scenarios for circular sector domain.

Cases	R	h_R	h_ϕ	TR	θ	p
(P_1,a,C_3)	12	$\frac{R}{145}$	$\frac{2\pi}{115}$	3	$\frac{2}{3}$	4
(P_1,b,C_2)	18	$\frac{R}{145}$	$\frac{2\pi}{115}$	5	$\frac{2}{3}$	4
(P_1,c,C_1)	25	$\frac{R}{145}$	$\frac{2\pi}{115}$	7	$\frac{2}{3}$	4
(P_1,d,C_1)	50	$\frac{R}{145}$	$\frac{2\pi}{115}$	15	$\frac{2}{3}$	4
(P_1,e,C_2)	100	$\frac{R}{145}$	$\frac{2\pi}{115}$	31	$\frac{2}{3}$	4
(P_1,f,C_3)	250	$\frac{R}{145}$	$\frac{2\pi}{115}$	79	$\frac{2}{3}$	4
(P_2,a,C_1)	12	$\frac{R}{145}$	$\frac{2\pi}{115}$	3	$\frac{2}{3}$	9
(P_2,b,C_2)	18	$\frac{R}{145}$	$\frac{2\pi}{115}$	5	$\frac{2}{3}$	9
(P_2,c,C_3)	25	$\frac{R}{145}$	$\frac{2\pi}{115}$	7	$\frac{2}{3}$	9
(P_2,d,C_3)	50	$\frac{R}{145}$	$\frac{2\pi}{115}$	15	$\frac{2}{3}$	9
(P_2,e,C_2)	100	$\frac{R}{145}$	$\frac{2\pi}{115}$	31	$\frac{2}{3}$	9
(P_2,f,C_1)	250	$\frac{R}{145}$	$\frac{2\pi}{115}$	79	$\frac{2}{3}$	9

Figures A4.4.9–A4.4.10 display the generation of a functional value of the response in the cases (P_2,e,C_2) and (P_1,f,C_3), respectively.

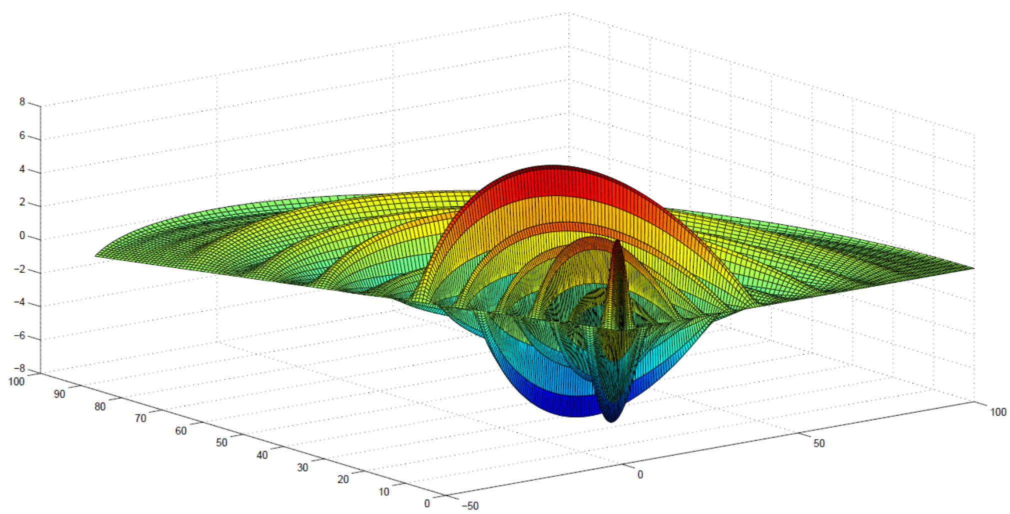


Figure A4.4.9: Case (P_2,e,C_2). Simulated response with $p = 9$, $R = 100$ and β of type C_2.

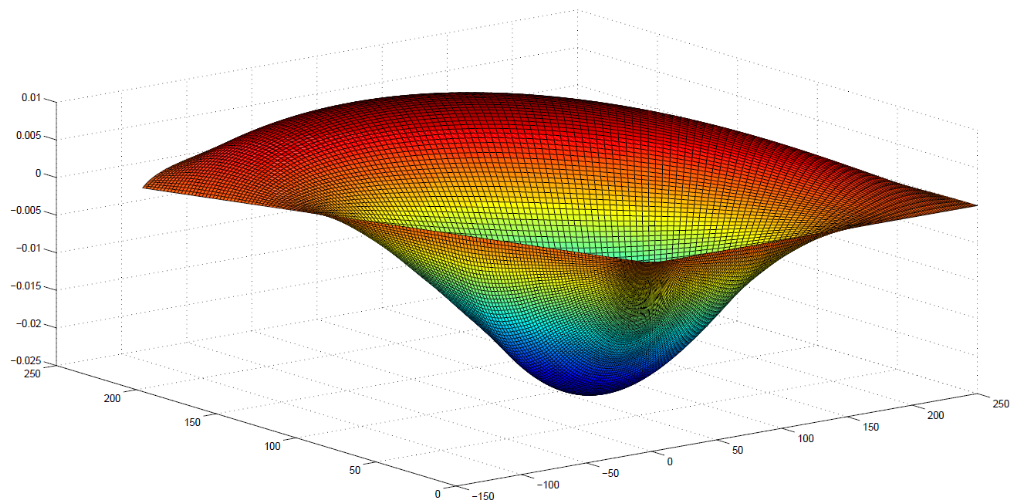


Figure A4.4.10: Case $(\mathrm{P}_1,\mathrm{f},\mathrm{C}_3)$. Simulated response with $p = 4$, $R = 250$ and $\boldsymbol{\beta}$ of type C_3.

The functional estimates obtained from the finite–dimensional approximation of the generalized least–squares estimator of $\boldsymbol{\beta}$ are now given in Figures A4.4.11–A4.4.12, for the cases $(\mathrm{P}_2,\mathrm{e},\mathrm{C}_2)$ and $(\mathrm{P}_1,\mathrm{f},\mathrm{C}_3)$, respectively.

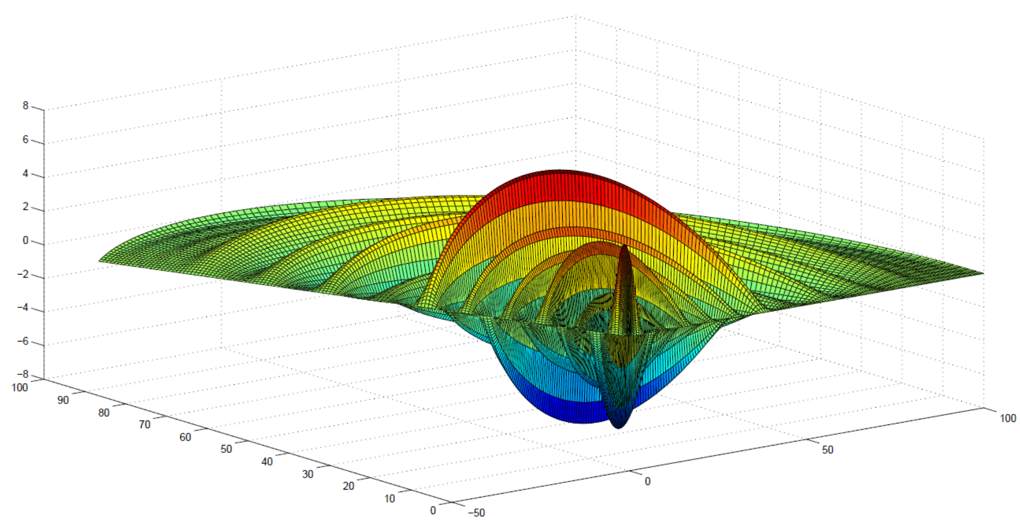


Figure A4.4.11: Case $(\mathrm{P}_2,\mathrm{e},\mathrm{C}_2)$. Estimated response with $p = 9$, $R = 100$ and $\boldsymbol{\beta}$ of type C_2.

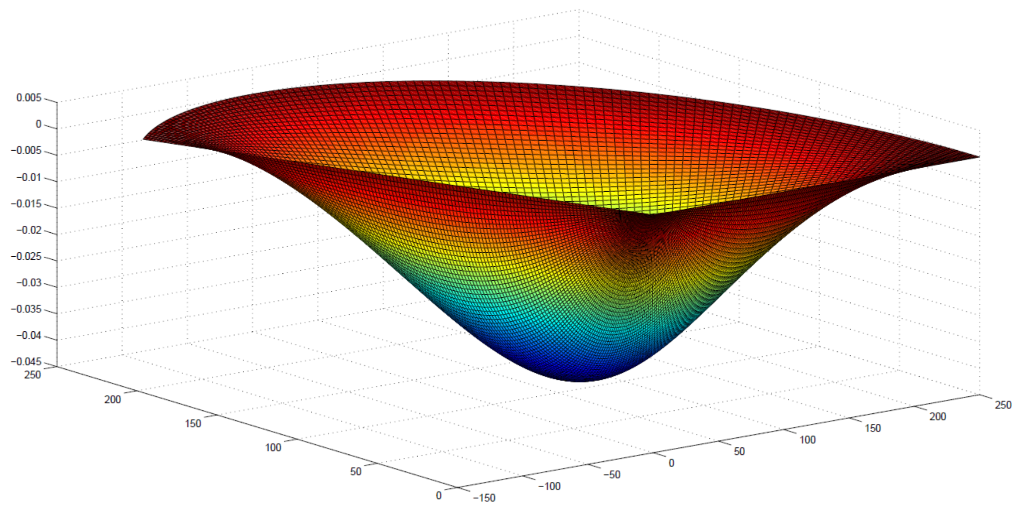


Figure A4.4.12: Case (P_1,f,C_3). Estimated response with $p = 4$, $R = 250$ and $\boldsymbol{\beta}$ of type C_3.

As in the previous sections, the empirical functional mean quadratic errors, associated with the estimation of $\boldsymbol{\beta}$ and $\mathbf{Y}$, are computed from equations (A4.23)–(A4.24). They are shown in Table A4.4.12, for $\boldsymbol{\beta}$, and in Table A4.4.13, for $\mathbf{Y}$.

These empirical functional mean quadratic errors are very stable through the different cases considered, and their order of magnitude is again 10^{-3} for the parameter $\boldsymbol{\beta}$, and 10^{-2} for the response. Here, the results displayed also correspond to the projection into lower finite–dimensional spaces than in the case of the rectangle, according to the functional form of the eigenvectors (see truncation order criterion (iii) in Appendix A4.4).

Table A4.4.12: $EFMSE_{\boldsymbol{\beta}}$ for the circular sector.

$EFMSE_{\boldsymbol{\beta}}$

(P_1,a,C_3)	(P_1,b,C_2)	(P_1,c,C_1)
$1.200\,(10)^{-4}$	$1.100\,(10)^{-4}$	$1.200\,(10)^{-4}$
(P_1,d,C_1)	(P_1,e,C_2)	(P_1,f,C_3)
$1.200\,(10)^{-4}$	$1.200\,(10)^{-4}$	$1.100\,(10)^{-4}$
(P_2,a,C_1)	(P_2,b,C_2)	(P_2,c,C_3)
$1.900\,(10)^{-4}$	$2.000\,(10)^{-4}$	$2.000\,(10)^{-4}$
(P_2,d,C_3)	(P_2,e,C_2)	(P_2,f,C_1)
$1.900\,(10)^{-4}$	$1.900\,(10)^{-4}$	$2.000\,(10)^{-4}$

Table A4.4.13: $EFMSE_{\mathbf{Y}}$ for the circular sector.

$EFMSE_{\mathbf{Y}}$

(P_1,a,C_3)	(P_1,b,C_2)	(P_1,c,C_1)
$8.770\,(10)^{-3}$	$8.810\,(10)^{-3}$	$8.820\,(10)^{-3}$
(P_1,d,C_1)	(P_1,e,C_2)	(P_1,f,C_3)
$8.820\,(10)^{-3}$	$8.820\,(10)^{-3}$	$8.810\,(10)^{-3}$
(P_2,a,C_1)	(P_2,b,C_2)	(P_2,c,C_3)
$9.630\,(10)^{-3}$	$9.670\,(10)^{-3}$	$9.670\,(10)^{-3}$
(P_2,d,C_3)	(P_2,e,C_2)	(P_2,f,C_1)
$9.670\,(10)^{-3}$	$9.680\,(10)^{-3}$	$9.660\,(10)^{-3}$

Statistics (A4.15) is now computed. Its values are displayed in Table A4.4.14. Again, as in the disk, the proportion of explained functional variability is very close to one leading to large values of statistics (A4.15), as it can be observed in Table A4.4.14 for all the cases analysed.

Table A4.4.14: F statistics (A4.15) for the circular sector.

Cases	(P_1,a,C_3)	(P_1,b,C_2)	(P_1,c,C_1)	(P_1,d,C_1)	(P_1,e,C_2)	(P_1,f,C_3)
F	$9.2(10^2)$	$3.1(10^3)$	$4.2(10^6)$	$4.8(10^8)$	$5.8(10^6)$)	$7.3(10^8)$
Cases	(P_2,a,C_1)	(P_2,b,C_2)	(P_2,c,C_3)	(P_2,d,C_3)	(P_2,e,C_2)	(P_2,f,C_1)
F	$1.8(10^3)$	$4.1(10^3)$	$2.6(10^7)$	$3.1(10^9)$	$6.8(10^6)$	$1.8(10^9)$

The statistics $T_{\boldsymbol{h}}$ in (A4.20) is computed to contrast the significance of the parameter vector $\boldsymbol{\beta}$ in case C_1 with $p = 4$ functional components. Eight realizations of a Gaussian random function $\boldsymbol{h}$ are considered from a Gaussian random field ξ, solution, in the mean-square sense, of the following boundary value problem on the circular sector

$$\begin{aligned} (-\Delta)\xi(\mathbf{x}) &= \varsigma(\mathbf{x}), \quad \mathbf{x} = (r\cos(\varphi),\, r\sin(\varphi)),\ 0 < \|r\| < R,\ 0 < \varphi < \pi\theta, \\ \xi(\varphi, 25) &= 0, \quad \varphi \in [0, \pi\theta], \end{aligned}$$

where $\theta = 2/3$, ς denotes a zero–mean Gaussian white noise on the circular sector such that

$$\int_{[0,\pi\theta]\times[0,25]} f(\varphi, v)\mathrm{E}\left\{\varsigma(\gamma, r)\varsigma(\varphi, v)\right\} d\varphi dv = f(\gamma, r), \quad (\gamma, r) \in [0, \pi\theta] \times [0, 25], \quad f \in L^2(CS),$$

with $L^2(CS)$ denoting the space of square–integrable functions on the circular sector. Table A4.4.15 reflects the percentage of successes, for $\alpha = 0.05$, and the averaged p–values over the 150 samples, with size $n = 150$, of the response, having C_1–type functional parameter vector $\boldsymbol{\beta}$ with $p = 4$ components, considering $TR = 7$.

Table A4.4.15: *Circular Sector.* Percentage of successes for $\alpha = 0.05$, at the left–hand side, and averaged p–values at the right–hand side, for each one of the eight realizations of the Gaussian function $h \in L^2(CS)$.

D	% Success	p
1	97.5%	$6.504(10)^{-6}$
2	100%	0
3	100%	$3.600(10)^{-8}$
4	100%	0
5	98%	$2.006(10)^{-6}$
6	99.5%	$9.807(10)^{-8}$
7	100%	0
8	99.5%	$4.111(10)^{-7}$

Table A4.4.15 again confirms the good performance of the test statistics $T_{\mathbf{h}}$, showing a high percentage of successes, and very small magnitudes for the averaged p–value (almost zero values), according to the significance of the parameter vector $\boldsymbol{\beta}$ considered in the generation of the analysed functional data set.

A4.5 Functional statistical analysis of FMRI data

In this section, we compare the results obtained from the application of the MatLab function *fmrilm.m* (see Liao et al. [2012] and Worsley et al. [2002]) from *fmristat.m* function set (available at `http://www.math.mcgill.ca/keith/fmristat`), with those ones provided by the implementation of our proposed functional statistical methodology, based on the Hilbert-valued fixed effect models with ARH(1) error term above introduced. The fMRI data set analysed is also freely available in AFNI format at `http://www.math.mcgill.ca/keith/fmristat/`. (AFNI Matlab toolbox can be applied to read such a data set). In the next section, structural information about such fMRI data is provided (see *BrikInfo.m* Matlab function).

The first step in the statistical analysis of fMRI data is to modeling the data response to an external stimulus. Specifically, at each voxel, denote by $x(t)$ the (noise-free) fMRI response at time t, and by $s(t)$ the external stimulus at that time. It is well–known that the corresponding fMRI response is not instantaneous, suffering a blurring and a delay of the peak response by about $6s$ (see, for example, Liao et al. [2012]). This fact is usually modelled by assuming that the fMRI response depends on the external stimulus by convolution with a hemodynamic response function $h(t)$ (which is usually assumed to be independent of the voxel), as follows:

$$x(t) = \int_0^{\infty} h(u)s(t-u)du. \tag{A4.33}$$

Several models have been proposed in the literature for the hemodynamic response function (hrf). For example, the gamma function (see Lange and Zeger [1997]), or the difference of two gamma functions, to model the slight intensity dip after the response has fallen back to zero (see Friston et al. [1998]).

The effects $(x_{i,1}, \ldots, x_{i,p})$ of p different types of stimuli on data, in scan i, is combined in terms of an additive model with different multiplicative coefficients $(\beta_1, \ldots, \beta_p)$ that vary from voxel to voxel. The combined fMRI response is then modeled as the linear model (see Friston et al. [1995])

$$x_{i,1}\beta_1(v) + \cdots + x_{i,p}\beta_p(v),$$

for each voxel v.

An important drift over time can be observed in fMRI time series data in some voxels. Such a drift is usually linear, or a more general slow variation function. In the first case, i.e., for a linear function

$$x_{i,k+1}\beta_{k+1}(v) + \cdots + x_{i,m}(v)\beta_m(v),$$

when the drift is not removed, it can be confounded with the fMRI response. Otherwise, it can be added to the estimate of the random noise ε, which, in the simplest case is assumed to be an AR(1) process at each voxel. In that case, the linear model fitted to the observed fMRI data is usually given by

$$Y_i(v) = x_{i,1}\beta_1(v) + \cdots + x_{i,p}\beta_p(v) + x_{i,k+1}\beta_{k+1}(v) + \cdots + x_{i,m}\beta_m(v) + \varepsilon_i(v), \quad i = 1, \ldots, n, \tag{A4.34}$$

for each one of the voxels v, in the real–valued approach presented in Worsley et al. [2002]. In (A4.34),

$$\varepsilon_i(v) = \rho(v)\varepsilon_{i-1}(v) + \xi_i(v), \quad |\rho(v)| < 1,$$

where $\{\xi_i(v),\ i = 1, \ldots, n\}$ are n random components of Gaussian white noise in time, for each voxel v. This temporal correlation structure for the noise has sense, under the assumption that the scans are equally spaced in time, and that the error from the previous scan is combined with fresh noise to produce the error for the current scan. In the presented Hilbert–valued approach, a similar reasoning can be applied to arrive to the fixed effect model with ARH(1) error term, introduced in Appendix A4.2. This model allows the representation of fMRI data in a functional spatially continuous form. Specifically, for the scan i, a continuous spatial variation is assumed underlying to the values of the noise across the voxels, reflected in the functional value of the ARH(1) process, representing the error term. In the same way, the H–valued components of the parameter vector $\boldsymbol{\beta}(\cdot)$ provide a continuous model to represent spatial variation over the voxels of the multiplicative coefficients $\beta_1(\cdot), \ldots, \beta_p(\cdot)$, independently of time. Since the fMRI response is subsampled at the n scan acquisition times $t_1, \ldots, t_n$, the fixed effect design matrix $\boldsymbol{X}$, constituted by the values of the fMRI response (A4.33) at such times, under the p different types of stimuli considered, has dimension $n \times p$. Note that in (A4.33) x is assumed to be independent of the voxel, according to the definition of the hrf.

A4.5.1 Description of the data set and the fixed effect design matrix

Brain scan measurements are represented on a set of $64 \times 64 \times 16$ voxels. Each one of such voxels represents a cube of $3.75 \times 3.75 \times 7\,mm$. At each one of the 16 depth levels or slices $\{S_i,\ i = 1, \ldots, 16\}$, the brain is scanned in 68 frames, $\{Fr_h,\ h = 1, \ldots, 68\}$. Equivalently, for $i = 1, \ldots, 16$, on the slice S_i, a 64×64 rectangular grid is considered, where measurements at each one of the 68 frames are collected.

We restrict our attention to the case $p = 2$, where two type of events are considered, respectively representing scans hot stimulus (with a height h_h) and scans warm stimulus (with a height h_w). The default parameters, chosen by Glover [1999], to generate the hrf as the difference of two gamma densities is the row vector $r = [5.4, 5.2, 10.8, 7.35, 0.35]$, where the first and third parameters represent the time to peak of the first and second gamma densities (Γ_1 and Γ_2), respectively; the second and fourth parameters represent the approximate full width at half maximum (FWHM) of the first and second gamma densities, respectively; and the fifth parameter (called also DIP) denotes the coefficient of the second gamma density, for more details, see Glover [1999], about modelling the hrf as the difference of two gamma density functions, in the following way:

$$hrf = \frac{\Gamma_1}{max(\Gamma_1)} - DIP\left(\frac{\Gamma_2}{max(\Gamma_2)}\right).$$

Considering $TR_t = 5$ seconds as the temporal step between each frame $Fr_h,\ h = 1, \ldots, 68$, in which all slices are scanned, frame times will be $Fr_{times} = (0, 5, 10, \ldots, 330, 335)$ (see Figure A4.5.1). Remark that, for any of the 68 scans, separated by $TR_t = 5$ seconds, keeping in mind that the first 4 frames are removed, 16 slices $\{S_i,\ i = 1, \ldots, 16\}$, are interleaved every 0.3125 seconds, approximately.

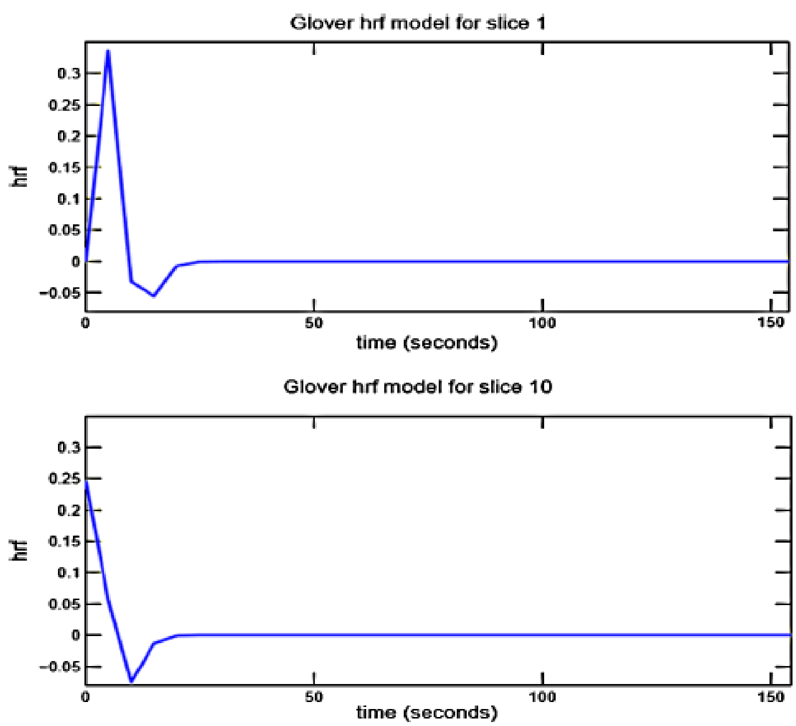


Figure A4.5.1: hrf model in Glover [1999] (without convoluting) obtained by *fmridesign.m* Matlab function, for slices S_i, with $i = 1$ (top) and $i = 10$ (bottom), until frame time $Fr_{times} = 150$ (i.e., the Glover's hrf continues to be zero).

The events matrix $\boldsymbol{E}$, which will be convoluted with the hrf, is a matrix whose rows are the events, and whose columns are the identifier of the event type, the starting event time, the duration of the event, and the height of the response for the event, respectively. In our example, we have considered a block design of 4 scans rest, 4 scans hot stimulus, 4 scans rest, 4 scans warm stimulus, repeating 4 times this block with 4 last scans rest (68 scans total). As noted before, we remove the first 4 frames. The hot event is identified by 1 and the warm event by 2, such that $h_h = 0.5$ and $h_w = 1$. Event times, for hot and warm stimulus, will be $[20, 60, \dots, 260, 300]$, since there are 8 frames between the beginning of events (4 frames for the previous event and 4 frames rest). Then, our events matrix $\boldsymbol{E}$ considered is

$$\boldsymbol{E} = \begin{pmatrix} 1 & 20 & 5 & 0.5 \\ 2 & 60 & 5 & 1 \\ 1 & 100 & 5 & 0.5 \\ 2 & 140 & 5 & 1 \\ 1 & 180 & 5 & 0.5 \\ 2 & 220 & 5 & 1 \\ 1 & 260 & 5 & 0.5 \\ 2 & 300 & 5 & 1 \end{pmatrix}. \tag{A4.35}$$

Convolution of matrix $\boldsymbol{E}$, in (A4.35), with the hrf leads to the set of real–valued 64×2 design matrices

$$\{\boldsymbol{X}_i,\ i = 1, \dots, 16\}, \quad \boldsymbol{X}_i \in \mathbb{R}^{64 \times 2},$$

implemented by *fmridesgin.m* Matlab function (see Figure A4.5.2).

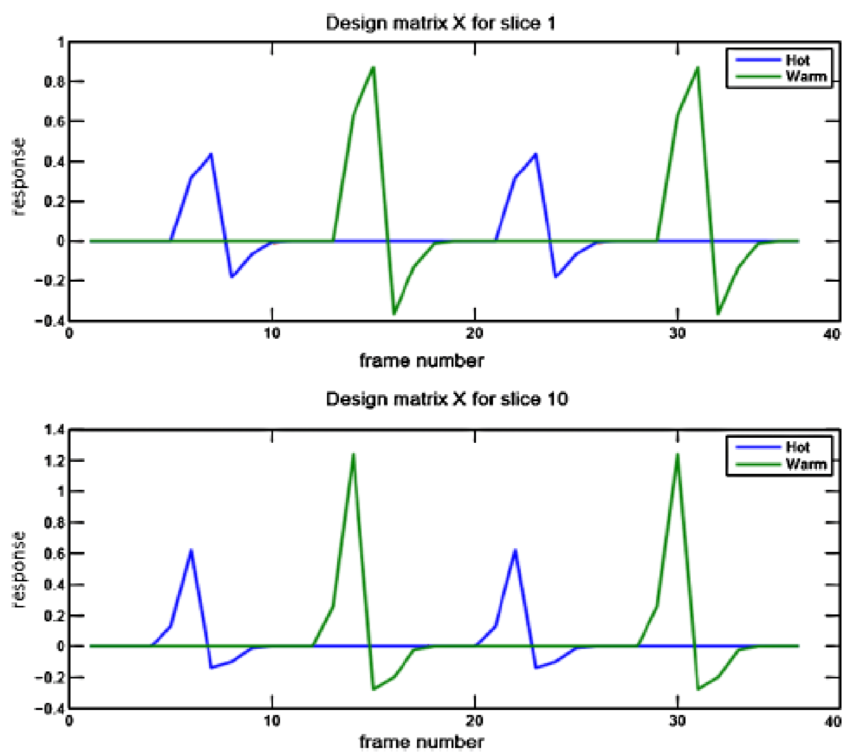


Figure A4.5.2: Design matrix $\boldsymbol{X}_i$ for the first 40 frames, and slices S_i, with $i = 1$ (top) and $i = 10$ (bottom), obtained by *fmridesign.m* Matlab function through the convolution of our events matrix with the hrf model in Glover [1999].

A4.5.2 Hilbert–valued fixed effect model fitting to FMRI data. A comparative study

The estimation results obtained with the implementation of the classical and Hilbert–valued linear model methodology are now compared. Specifically, in the classical case, from the linear model approach presented in Worsley et al. [2002], we consider a fixed–effect model fitting, in the case where the error term is an AR(1) process, ignoring spatial correlation across the voxels. In particular, the MatLab function *fmrilm.m* is implemented to fit model (A4.34) to a single run of fMRI data, allowing for spatially varying temporal correlated errors. The parameters of the spatial varying AR(1) models (from voxel to voxel) are estimated from the sample autocorrelation of the residuals, obtained after estimation of the fixed effect parameter by ordinary least–squares, ignoring temporal correlation of the errors, at each voxel. This procedure could be iterated. That is, the estimated autocorrelation coefficient can be used to pre–whitening the data at each voxel. Hence, the fixed effect parameter is estimated by ordinary least-squares, from such data. This iterative estimation procedure can be repeated several times. However, as pointed out in Worsley et al. [2002], such iterations do not lead to a substantial improvement in practice. A variance reduction technique is then applied in Worsley et al. [2002] to the estimated autocorrelation coefficient (reduced bias sample autocorrelation), consisting of spatial smoothing of the sample autocorrelations. This technique reduces variability, although slightly increases the bias.

In this subsection, we also implement the approach introduced in Appendix A4.2, from the fMRI data set described in Appendix A4.5.1. As commented before, our approach presents the advantage of providing a continuous spatial description of the variation of the fixed effect parameters, as well as of the parameters characterizing the temporal correlated error term, with autoregressive dynamics. Furthermore, the spatial correlations are also incorporated to our functional statistical analysis, computed from the spatial autocovariance and cross-covariance kernels, respectively defining the operators R_0 and R_1, characterizing the functional dependence structure of the ARH(1) error term.

Functional fixed effect model fitting is independently performed at each slice S_i, for $i = 1, \dots, 16$. Specifically, for $i = 1, \dots, 16$, as commented before, a real-valued $n \times p$, with $p = 2$, fixed effect design matrix $\mathbf{X}_i$ is considered (see Appendix A4.5.1). The effects of the two different events studied are combined by the vector of functional fixed effect parameters

$$\boldsymbol{\beta}_i(\cdot) = [\beta_{1,i}(\cdot), \beta_{2,i}(\cdot)]^T \in H^2.$$

Here, H^2 is the Hilbert space of 2–dimensional vector functions, whose components are square–integrable over the spatial rectangular grid considered at each slice. Furthermore, for $i = 1, \dots, 16$,

$$\mathbf{Y}_i(\cdot) = [Y_{1,i}(\cdot), \dots, Y_{n,i}(\cdot)]^T$$

is the H^n–valued Gaussian fMRI data response, with n representing the number of frames ($n = 64$, since the first 4 frames are removed because they do not represent steady–state images). In the computation of the generalized least–squares estimate of $\boldsymbol{\beta}$, the empirical matrices $\left\{\widehat{\boldsymbol{\Lambda}}_k,\ k = 1, \dots, TR\right\}$ are computed from the empirical covariance operators (A4.9), where TR is selected according to the required conditions specified, in relation to the sample size n, in Bosq [2000] (see, in particular, [Bosq, 2000, pp. 101–102 and pp. 116–117], and Remark A4.2.3).

In the subsequent developments, in the results obtained by applying the Hilbert–valued multivariate fixed effect approach, we will distinguish between cases A and B, respectively corresponding to the projection into two and five empirical eigenvectors. For each one of the 16 slices, the temporal and spatial averaged empirical quadratic errors, associated with the estimates of the response, computed with the *fmrilm.m* MatLab function, and with the proposed multivariate Hilbert–valued mixed effect approach, respectively denoted as $EFMSE_{\mathbf{Y}_i^{fMRI}}$ and $EFMSE_{\mathbf{Y}_i^{H}}$, are displayed in Tables A4.5.1–A4.5.2.

Table A4.5.1: $EFMSE_{\mathbf{Y}_i^{fMRI}}$ and $EFMSE_{\mathbf{Y}_i^{H}}$ for case A.

Slices S_i	$EFMSE_{\mathbf{Y}_i^{fMRI}}$	$EFMSE_{\mathbf{Y}_i^{H}}$
1	$2.417(10)^{-}3$	$3.492(10)^{-}3$
2	$3.051(10)^{-}3$	$3.119(10)^{-}3$
3	$4.293(10)^{-}3$	$5.523(10)^{-}3$
4	$6.666(10)^{-}3$	$7.690(10)^{-}3$
5	$8.986(10)^{-}3$	$9.961(10)^{-}3$
6	$8.462(10)^{-}3$	$9.434(10)^{-}3$
7	$1.108(10)^{-}2$	$1.920(10)^{-}2$
8	$1.720(10)^{-}2$	$2.720(10)^{-}2$
9	$1.499(10)^{-}2$	$1.914(10)^{-}2$
10	$1.036(10)^{-}2$	$1.851(10)^{-}2$
11	$1.308(10)^{-}2$	$1.634(10)^{-}2$
12	$1.302(10)^{-}2$	$1.300(10)^{-}2$
13	$7.850(10)^{-}3$	$7.939(10)^{-}3$
14	$6.640(10)^{-}3$	$6.730(10)^{-}3$
15	$3.511(10)^{-}3$	$2.832(10)^{-}3$
16	$2.771(10)^{-}3$	$3.540(10)^{-}3$

Table A4.5.2: $EFMSE_{\mathbf{Y}_i^{fMRI}}$ and $EFMSE_{\mathbf{Y}_i^H}$ for case B.

Slices S_i	$EFMSE_{\mathbf{Y}_i^{fMRI}}$	$EFMSE_{\mathbf{Y}_i^H}$
1	$2.417(10)^{-}3$	$2.592(10)^{-}3$
2	$3.051(10)^{-}3$	$3.119(10)^{-}3$
3	$4.293(10)^{-}3$	$4.733(10)^{-}3$
4	$6.666(10)^{-}3$	$7.671(10)^{-}3$
5	$8.986(10)^{-}3$	$9.065(10)^{-}3$
6	$8.462(10)^{-}3$	$8.435(10)^{-}3$
7	$1.108(10)^{-}2$	$1.120(10)^{-}2$
8	$1.720(10)^{-}2$	$1.919(10)^{-}2$
9	$1.499(10)^{-}2$	$1.524(10)^{-}2$
10	$1.036(10)^{-}2$	$1.040(10)^{-}2$
11	$1.308(10)^{-}2$	$1.481(10)^{-}2$
12	$1.302(10)^{-}2$	$1.299(10)^{-}2$
13	$7.849(10)^{-}3$	$7.929(10)^{-}3$
14	$6.640(10)^{-}3$	$6.719(10)^{-}3$
15	$3.511(10)^{-}3$	$2.829(10)^{-}3$
16	$2.771(10)^{-}3$	$3.540(10)^{-}3$

It can be observed, in Tables A4.5.1–A4.5.2, that the performance of the two approaches is very similar. However, the advantage of the presented approach relies on the important dimension reduction it provides, since, as commented before, we have considered the truncations orders $TR = 2$ (Case A) and $TR = 5$ (Case B). Note that, for each slice, the parameter vector has dimension $2 \times \times (64 \times 64)$, in the model fitted by *fmrilm.m* Matlab function. While the presented approach fits the functional projected model, that, for the the cases A and B studied, is defined in terms of a parameter vector $\boldsymbol{\beta}$ with dimension 2×2 and 2×5, respectively. Furthermore, the iterative estimation method implemented in *fmrilm.m* requires several steps, repeated at each one of the 64×64 voxels in the 16 slices (data pre–whitening, ordinary least-squares estimation of $\boldsymbol{\beta}$, and AR(1) correlation coefficient estimation iterations, jointly with the spatial smoothing of the temporal correlation - reduced bias - parameter estimates).

For the slices 1, 5, 10 and 15, the temporal averaged (frames 5–68) estimated values of the response, applying *fmrilm.m* MatLab function, and the fixed effect model with ARH(1) error term, in cases A and B, are respectively displayed in Figures A4.5.3–A4.5.5. The corresponding empirical time-averaged quadratic errors are displayed in Figures A4.5.6–A4.5.8, respectively.

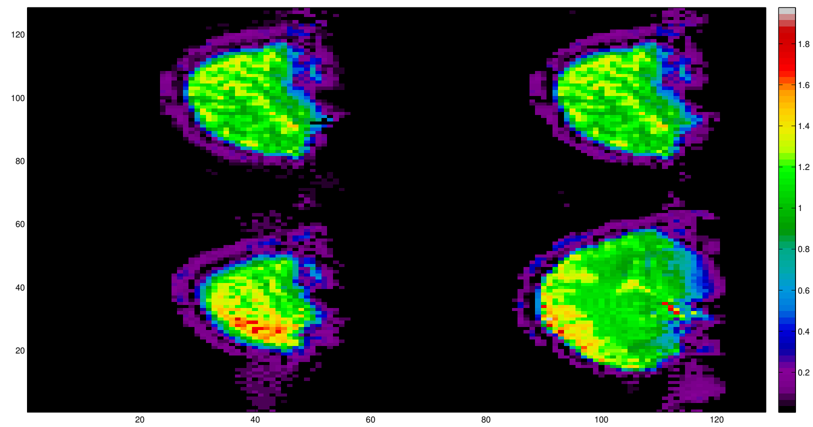


Figure A4.5.3: Averaged in time (frames 5–68) estimated response values for slices $1, 5, 10$ and 15, obtained by applying *fmrilm.m* MatLab function.

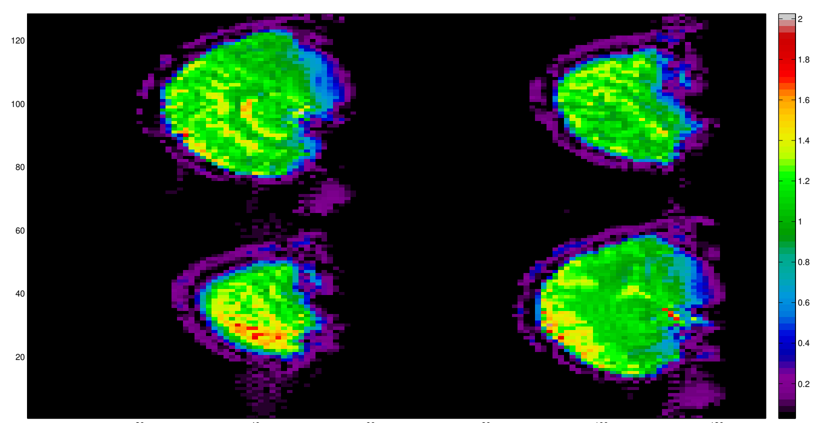


Figure A4.5.4: Averaged in time (frames 5–68) estimated response values for slices $1, 5, 10$ and 15, obtained by applying the fixed effect approach with ARH(1) error term, for case A.

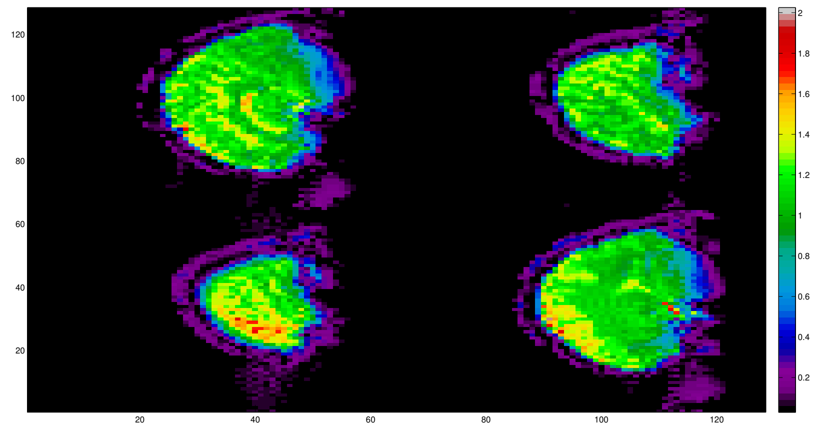


Figure A4.5.5: Averaged in time (frames 5–68) estimated response values for slices $1, 5, 10$ and 15, obtained by applying the fixed effect approach with ARH(1) error term, for case B.

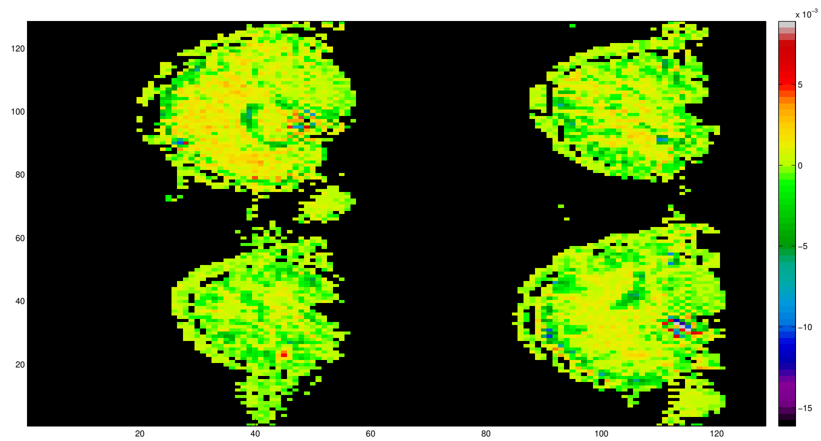


Figure A4.5.6: Averaged in time (frames 5–68) empirical errors for slices $1, 5, 10$ and 15, obtained by applying *fmrilm.m* MatLab function.

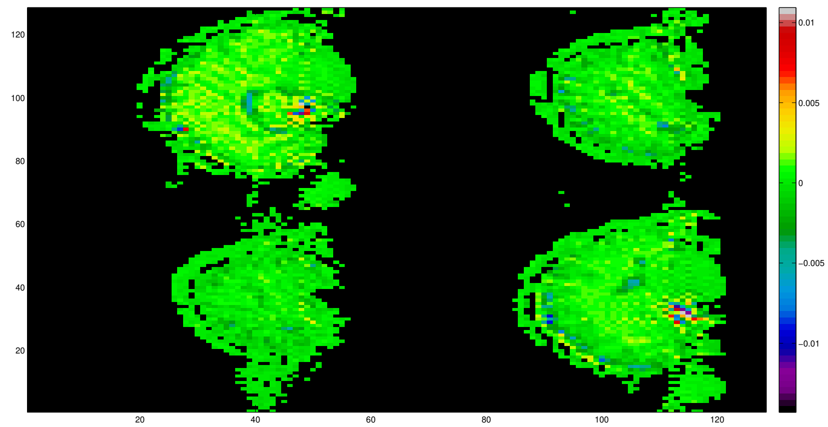


Figure A4.5.7: Averaged in time (frames 5–68) empirical errors for slices $1, 5, 10$ and 15, obtained by applying the fixed effect approach with ARH(1) error term, for case A.

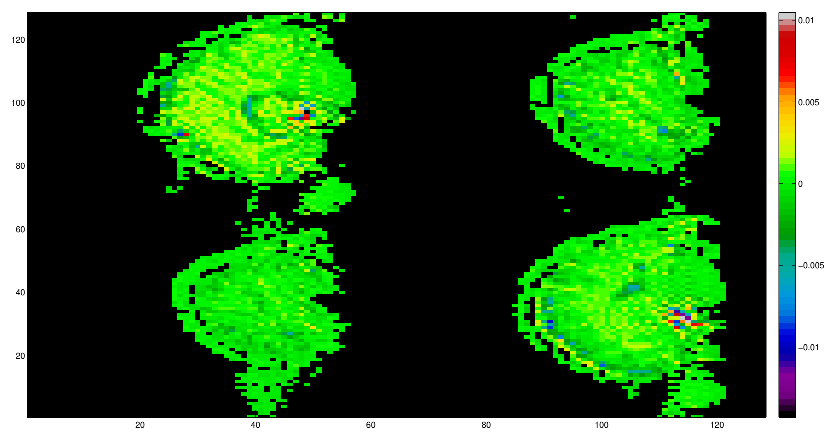


Figure A4.5.8: Averaged in time (frames 5–68) empirical errors for slices $1, 5, 10$ and 15, obtained by applying the fixed effect approach with ARH(1) error term, for case B.

A4.5.3 Significance test

We are interested in contrast the significance of the spatial varying parameter vector $\boldsymbol{\beta}$ combining the effects of the two stimulus considered on the overall brain, in its real-valued, and H^2-valued version. The F statistic in the MatLab function *fmrilm.m* (fMRI linear model), computed, as before, from a single run of fMRI data, leads to the results reflected in Table A4.5.3, on the percentage of brain voxels, where the real–valued fixed effect model with AR(1) term is significative, for each one of the 16 slices considered.

Table A4.5.3: Percentage of brain voxels per slice, where the real-valued fixed effect model with AR(1) error term, fitted by *fmrilm.m* MatLab function, is significative.

S	% voxels with rejection of H_0
1	99.927%
2	99.927%
3	99.707%
4	99.902%
5	99.805%
6	99.951%
7	99.927%
8	99.976%
9	99.805%
10	99.951%
11	99.951%
12	99.902%
13	99.878%
14	99.951%
15	99.951%
16	100%

As described in Appendix A4.3, for each slice, i.e., for $i = 1, \ldots, 16$, the value of the conditional chi–squared test statistics $T_{\boldsymbol{h}}$, in equation (A4.20), is computed, from four realizations of a Gaussian random function $\boldsymbol{h}$, generated from a Gaussian random field ξ, satisfying equation (A4.32) on the rectangle containing each brain slice. As before, the functional response sample size at each slice is 64, since the first four frames are discarded. It can be observed, in the numerical results displayed in Table A4.5.4, for $TR = 16$, and in Table A4.5.5, for $TR = 4$, that the null hypothesis is rejected, in most of the random directions in all the brain slices; i.e., the functional fixed effect model with ARH(1) error term is significative for $\alpha = 0.05$. Note that a very few p–values are slightly larger than $\alpha = 0.05$, with very small difference, that could be produced by the numerical errors accumulated, due to the presence of small values to be inverted. Thus, we can conclude the suitability of our approach, to combine the effects of the scans hot stimulus, and the scans warm stimulus, in a functional spatially continuous framework.

Table A4.5.4: p-values for T_h computed at the 16 slices, considering four random directions, for $TR = 16$.

S	D_1	D_2	D_3	D_4
1	0	0	0.082	0.023
2	$0.590(10)^{-2}$	0	0	0
3	0.018	0.066	0.049	0.030
4	0	0	0	$0.170(10)^{-10}$
5	0	0.026	0	0
6	0	0	0	0
7	$0.710(10)^{-7}$	0	0	0
8	0	0.006	0	0
9	0.049	0	0	0.023
10	$0.390(10)^{-7}$	0.031	0	0
11	0.004	0.006	$0.660(10)^{-6}$	0.052
12	0.046	0	0	0.034
13	$0.340(10)^{-7}$	0.028	0	$0.440(10)^{-3}$
14	0	$0.180(10)^{-6}$	0.021	0.050
15	0	$0.140(10)^{-7}$	0.044	0.052
16	$0.110(10)^{-4}$	$0.230(10)^{-7}$	0	0

Table A4.5.5: p-values for T_h computed at the 16 slices, considering four random directions, for $TR = 4$.

S	D_1	D_2	D_3	D_4
1	0	0.051	0.071	0.011
2	$0.880(10)^{-4}$	0	0	0
3	0.067	0.034	0	0.037
4	0	$0.250(10)^{-4}$	$0.110(10)^{-4}$	0.016
5	$0.370(10)^{-6}$	0	$0.280(10)^{-6}$	0
6	0.001	0	0	$0.220(10)^{-4}$
7	0.064	0.034	0.007	0.044
8	0.072	0.079	0.035	0
9	$0.220(10)^{-5}$	$0.470(10)^{-4}$	0.004	$0.220(10)^{-9}$
10	0	$0.120(10)^{-3}$	$0.370(10)^{-4}$	$0.970(10)^{-7}$
11	0.081	0.058	0	0
12	$0.870(10)^{-4}$	0	0	0.036
13	$0.760(10)^{-3}$	0	0	$0.370(10)^{-3}$
14	$0.210(10)^{-6}$	0	0	0.037
15	0	$0.650(10)^{-4}$	0.032	0
16	$0.540(10)^{-6}$	0	0	$0.520(10)^{-3}$

Comparing results in Tables A4.5.3–A4.5.5, we can conclude that both methodologies, the one presented in Worsley et al. [2002], and the functional approach introduced here, lead to similar results regarding the significance of the models they propose, respectively based on spatial varying real–valued multiplicative coefficients with AR(1) error term, and Hilbert–valued coefficients with ARH(1) error term.

A4.6 Conclusions

As shown in the simulation study, the boundary conditions affect the decay velocity at the boundary of the covariance kernels, defining the functional entries of the matrix covariance operator of the error term. Thus, the dependence range of the error components is directly affected by the boundary conditions. A better performance of the generalized least–squares estimator of the parameter vector $\boldsymbol{\beta}$ is observed, when a fast continuous decay is displayed by the error covariance kernels close to the boundary, as it is observed in the circular domains. Furthermore, in the simulation study undertaken, and in the real–data problem addressed, a good performance of the computed generalized least–squares estimator, and of the test statistics is observed for low truncation orders. Thus, an important dimension reduction is achieved with the presented approach. Summarizing, the proposed approach allows the incorporation of temporal and spatial correlations in the analysis, with an important dimension reduction.

The derivation of similar results under alternative boundary conditions like Neumann and Robin boundary conditions constitutes an open research problem (see, for example, Grebenkov and Nguyen [2013]). Another important research problem is to address the same analysis under a slow decay of the error covariance kernels at the boundary (see, for example, Frías et al. [2017]; Jiang [2012, 2016]; Tong [2011], beyond the Gaussian context).

A4.7 Supplementary Material

The eigenvectors and eigenvalues of the Dirichlet negative Laplacian operator on the regular domains defined by the rectangle, disk and circular sector are described here (see, for example, Grebenkov and Nguyen [2013]). It is well–known that the negative Laplacian operator $(-\boldsymbol{\Delta}_D)$ on a regular bounded open domain $D \subset \mathbb{R}^2$, with Dirichlet boundary conditions, is given by

$$-\boldsymbol{\Delta}_D(f)(x_1,x_2) = -\frac{\partial^2}{\partial x_1^2}f(x_1,x_2) - \frac{\partial^2}{\partial x_2^2}f(x_1,x_2), \quad f(x_1,x_2)=0, \quad (x_1,x_2)\in \partial D, \quad D\subseteq \mathbb{R}^2,$$

where ∂D is the boundary of D. In the subsequent development, we will denote by $\{\phi_k,\ k\geq 1\}$ and $\{\lambda_k(-\boldsymbol{\Delta}_D),\ k\geq 1\}$ the respective eigenvectors and eigenvalues of $(-\boldsymbol{\Delta}_D)$, that satisfy

$$\begin{aligned} -\boldsymbol{\Delta}_D\phi_k(\mathbf{x}) &= \lambda_k(-\boldsymbol{\Delta}_D)\phi_k(\mathbf{x}) \ \left(\mathbf{x}\in D\subseteq \mathbb{R}^2\right), \\ \phi_k(\mathbf{x}) &= 0 \ (\mathbf{x}\in\partial D)\,,\ \forall k\geq 1, \end{aligned}$$

for D being one of the following three domains:

$$D_1 = \prod_{i=1}^{2} [a_i, b_i] \,, \quad D_2 = \left\{ \mathbf{x} \in \mathbb{R}^2 : \ R_0 < \|\mathbf{x}\| < R \right\},$$

and

$$D_3 = \left\{ \mathbf{x} \in \mathbb{R}^2 : \ R_0 < \|\mathbf{x}\| < R, \text{ and } 0 < \varphi < \pi\theta \right\}.$$

A4.7.1 Eigenelements of Dirichlet negative Laplacian operator on rectangles

Let us first consider domain

$$D_1 = \prod_{i=1}^{2} [a_i, b_i] \,.$$

The eigenvectors $\{\phi_{\mathbf{k}}, \ \mathbf{k} \in \mathbb{N}_*^2\}$ and eigenvalues $\{\lambda_{\mathbf{k}}(-\boldsymbol{\Delta}_{D_1}), \ \mathbf{k} \in \mathbb{N}_*^2\}$ of $-\boldsymbol{\Delta}_{D_1}$ are given by (see Grebenkov and Nguyen [2013]):

$$\begin{aligned}
\phi_{\mathbf{k}}(\mathbf{x}) &= \phi_{k_1}^{(1)}(x_1)\, \phi_{k_2}^{(2)}(x_2)\,, \ \lambda_{\mathbf{k}} = \lambda_{k_1}^{(1)} + \lambda_{k_2}^{(2)}, \\
\phi_{k_i}^{(i)}(x_i) &= \sin\left(\frac{\pi k_i x_i}{l_i}\right), \quad x_i \in [a_i, b_i]\,, \ i = 1, 2, \\
\lambda_{k_i}^{(i)} &= \frac{\pi^2 k_i^2}{l_i^2}, \quad k_i \geq 1, \ i = 1, 2,
\end{aligned} \tag{A4.36}$$

where $l_i = b_i - a_i$, for $i = 1, 2$.

A4.7.2 Eigenelements of Dirichlet negative Laplacian operator on disks

In general, for the circular annulus

$$\widetilde{D}_2 = \left\{ \mathbf{x} \in \mathbb{R}^2 : \ R_0 < \|\mathbf{x}\| < R \right\},$$

its rotation symmetry allows us to define $-\boldsymbol{\Delta}_{\widetilde{D}_2}$ in polar coordinates as

$$-\boldsymbol{\Delta}_{\widetilde{D}_2} = -\frac{\partial^2}{\partial r^2} - \frac{1}{r}\frac{\partial}{\partial r} - \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}, \quad x_1 = r\cos\varphi, \ x_2 = r\sin\varphi.$$

The application of variable separation method then leads to the following explicit formula of its eigenfunctions (see, for example, Grebenkov and Nguyen [2013])

$$\phi_{khl}(r, \varphi) = \left[J_k\left(\alpha_{kh} r / R\right) + c_{kh} Y_k\left(\alpha_{kh} r / R\right)\right] \times C_k(l)\,, \tag{A4.37}$$

with
$$C_k(l) = \begin{cases} cos\,(k\varphi) \text{ l=1}, \\ sin\,(k\varphi) \text{ l=2 } (k \neq 0)\,, \end{cases}$$
where $\{J_k\,(z)\}$ and $\{Y_k\,(z)\}$ are the Bessel functions of order k of first and second kind, respectively,
$$\{\lambda_{kh}\left(-\mathbf{\Delta}_{\widetilde{D}_2}\right) = \alpha_{kh}^2/R^2\}$$
are the corresponding eigenvalues, and the sets $\{\alpha_{k,h},\ k \geq 1,\ h = 1, \ldots, M(k)\}$ and $\{c_{k,h},\ k \geq 1,\ h = 1, \ldots, M(k)\}$ are defined from the boundary conditions at $r = R$ and $r = R_0$.

If we focus on domain D_2, the disk, i.e., $R_0 = 0$, the coefficients $\{c_{k,h},\ k \geq 1,\ h = 1, \ldots, M(k)\}$ are set to 0. The eigenfunctions then adopt the following expression:
$$\phi_{khl}\,(r,\varphi) \quad = \quad J_k\,(\alpha_{kh} r/R)\,C_k(l), \quad l = 1, 2, \tag{A4.38}$$
with eigenvalues
$$\lambda_{kh}\,(-\mathbf{\Delta}_{D_2}) = \frac{\alpha_{kh}^2}{R^2}, \quad k \geq 1,\ h = 1, \ldots, M(k),$$
where $\{\alpha_{k,h},\ h = 1, \ldots, M(k)\}$ are the $M(k)$ positive roots of the Bessel function $J_k\,(z)$ of order k. Note that we can also consider truncation at parameter $M(k)$ for $k \geq 1$, since this parameter increases with the increasing of the radius R.

A4.7.3 Eigenelements of Dirichlet negative Laplacian operator on circular sectors

Lastly, we consider domain D_3, the circular sector of radius R and angle $0 < \varphi < \pi\theta$. The eigenvectors and eigenvalues are given by the following expression (see, for example, Grebenkov and Nguyen [2013]):
$$\begin{aligned} \phi_{kh}\,(r,\varphi) \quad &= \quad J_{k/\theta}\,(\alpha_{kh} r/R) \sin\,(k\varphi/\theta)\,, \quad r \in [0, R]\,, \\ \lambda_{kh}\,(-\mathbf{\Delta}_{D_3}) \quad &= \quad \frac{\alpha_{kh}^2}{R^2},\ k \geq 1,\ h = 1, \ldots, M(k), \end{aligned} \tag{A4.39}$$
with $M(k)$ and $\{\alpha_{k,h},\ k \geq 1,\ h = 1, \ldots, M(k)\}$ being given as in the previous section.

A4.7.4 Asymptotic behavior of eigenvalues

A4.7.4.1 The rectangle

The functional data sets generated in Appendix A4.4 must have a covariance matrix operator with functional entries (operators) in the trace class. We then apply the results in Widom [1963] to study the asymptotic order of eigenvalues of the integral equation

$$\int_{\mathbb{R}^2} V^{1/2}(\mathbf{t}) l_{\varepsilon_i}(\mathbf{t}-\mathbf{s}) V^{1/2}(\mathbf{s}) f(\mathbf{s}) d\mathbf{s} = \lambda f(\mathbf{t}).$$

In our case, V is the indicator function on the rectangle, i.e., on domain D_1, and l_{ε_i} is the covariance kernel defining the square root

$$R^{1/2}_{\varepsilon_i\varepsilon_i} = f_i(-\mathbf{\Delta}_{D_1}) = (-\mathbf{\Delta}_{D_1})^{-(d-\gamma_i)}, \quad \gamma_i \in (0, d/2),$$

of the autocovariance operator of the Hilbert-valued error component $\{\varepsilon_i,\ i = 1,\dots,n\}$, with

$$R_{\varepsilon_i\varepsilon_i} = R^{1/2}_{\varepsilon_i\varepsilon_i} R^{1/2}_{\varepsilon_i\varepsilon_i}.$$

Note that with the choice made of functions V and $\{l_{\varepsilon_i},\ i = 1,\dots,n\}$, the conditions assumed in Widom [1963] are satisfied. In particular, the following asymptotic holds:

$$\lambda_k(R^{1/2}_{\varepsilon_i\varepsilon_i}) = \mathcal{O}(k^{-2(d-\gamma_i)/d}), \quad k \longrightarrow \infty, \quad i = 1,\dots,n,$$

(see [Widom, 1963, p. 279, Eq. (2)]). Also, in general, the eigenvalues of the Dirichlet negative Laplacian operator on a regular bounded open domain D satisfy

$$\gamma_k(-\Delta_D) \sim 4\pi \frac{\left(\Gamma\left(1+\frac{d}{2}\right)\right)^{2/d}}{|D|^{2/d}} k^{2/d}, \quad k \longrightarrow \infty.$$

A4.7.4.2 Asymptotic behavior of zeros of Bessel functions.

As before, $J_k(z)$ denotes the Bessel function of the first kind of order k. Let $\{j_{k,h},\ h = 1,\dots,M(k)\}$ be its $M(k)$ roots. In Elbert [2001]; Olver [1951, 1952], it is shown that, for a fixed h and large k, the Olver's expansion holds

$$j_{kh} \simeq k + \delta_h k^{1/3} + \mathcal{O}(k^{-1/3}), \quad k \to \infty.$$

On the other hand, for fixed k and large h, the McMahon's expansion also is satisfied (see, for example, Watson [1966])

$$j_{kh} \simeq \pi\left(h + k/2 - 1/4\right) + \mathcal{O}(h^{-1}), \quad h \to \infty.$$

These results will be applied in Appendix A4.4, in the definition of the eigenvalues of the covariance operators $\{\boldsymbol{R}_{\varepsilon_i\varepsilon_i},\ i = 1,\dots,n\}$, on the disk and circular sector, to ensure their rapid decay to zero, characterizing the trace operator class.

Acknowledgments

This work has been supported in part by project MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

A5

CLASSICAL AND BAYESIAN COMPONENTWISE PREDICTORS FOR NON–ERGODIC ARH(1) PROCESSES

RUIZ-MEDINA, M. D.; ÁLVAREZ-LIÉBANA, J.: Classical and Bayesian componentwise predictors for non–ergodic ARH(1) processes. REVSTAT (2017), in press.

Year	Categ.	Cites	Impact Factor (5 years)	Quartil
2016	Statist. & Probab.	194	1.413	Q4

ABSTRACT

A special class of standard Gaussian autoregressive Hilbertian processes of order one (Gaussian ARH(1) processes), with bounded linear autocorrelation operator, which does not satisfy the usual Hilbert–Schmidt assumption, is considered. To compensate the slow decay of the diagonal coefficients of the autocorrelation operator, a faster decay velocity of the eigenvalues of the trace autocovariance operator of the innovation process is assumed. As usual, the eigenvectors of the autocovariance operator of the ARH(1) process are considered for projection, since, here, they are assumed to be known. Diagonal componentwise classical and Bayesian estimation of the autocorrelation operator is studied for prediction. The asymptotic efficiency and equivalence of both estimators is proved, as well as of their associated componentwise ARH(1) plug–in predictors. A simulation study is undertaken to illustrate the theoretical results derived.

A5.1 INTRODUCTION

Functional time series theory plays a key role in the analysis of high-dimensional data (see, for example, Aue et al. [2015]; Bosq [2000]; Bosq and Blanke [2007]). Inference for stochastic processes can also be addressed from this framework (see Álvarez-Liébana et al. [2016] in relation to functional prediction of the Ornstein–Uhlenbeck process, in an ARH(1) process framework). Bosq [2000] addresses the problem of infinite–dimensional parameter estimation and prediction of ARH(1) processes, in the cases of known and unknown eigenvectors of the autocovariance operator. Alternative projection methodologies have been adopted, for example, in Antoniadis and Sapatinas [2003], in terms of wavelet bases, and Besse and Cardot [1996], in terms of spline bases. The book by Bosq and Blanke [2007] provides a general overview on statistical prediction, including Bayesian predictors, inference by projection and kernel methods, empirical density estimation, and linear processes in high–dimensional spaces (see also Blanke and Bosq [2015] on Bayesian prediction for stochastic processes). Recently, Bosq and Ruiz-Medina [2014] have derived new results on asymptotic efficiency and equivalence of classical and Bayes predictors for l^2–valued Poisson process, where, as usual, l^2 denotes the Hilbert space of square summable sequences. Classical and Bayesian componentwise parameter estimators of the mean function and autocovariance operator, characterizing Gaussian measures in Hilbert spaces, are also compared in terms of their asymptotic efficiency, in that paper.

We first recall that the class of processes studied here could be of interest in applications, for instance, in the context of anomalous physical diffusion processes (see, for example, Gorenflo and Mainardi [2003]; Meerschaert et al. [2002]; Metzler and Klafter [2004], and the references therein). An interesting example of our framework corresponds to the case of spatial fractal diffusion operator, and regular innovations. Specifically, the class of standard Gaussian ARH(1) processes studied have a bounded linear autocorrelation operator, admitting a weak–sense diagonal spectral representation, in terms of the eigenvectors of the autocovariance operator. The sequence of diagonal coefficients, in such a spectral representation, displays an accumulation point at one. The singularity of the autocorrelation kernel is compensated by the regularity of the autocovariance kernel of the innovation process. Namely, the key assumption here is the summability of the quotient between the eigenvalues of the autocovariance operator of the innovation process and of the ARH(1) process. Under suitable conditions, the asymptotic efficiency and equivalence of the studied diagonal componentwise classical and Bayesian estimators of the autocorrelation operator are derived (see Theorem A5.4.1 below). Under the same setting of conditions, the asymptotic efficiency and equivalence of the corresponding classical and Bayesian ARH(1) plug–in predictors are proved as well (see Theorem A5.4.2 below). Although both theorems only refer to the case of known eigenvectors of the autocovariance operator, as illustrated in the simulation study undertaken in Álvarez-Liébana et al. [2017] (see also Álvarez-Liébana [2017]; Ruiz-Medina and Álvarez-Liébana [2018a]), a similar performance is obtained for the case of unknown eigenvectors, in comparison with other componentwise, kernel–based, wavelet-based penalized and nonparametric approaches adopted in the current literature (see Antoniadis and Sapatinas [2003]; Besse and Cardot [1996]; Bosq [2000]; Guillas [2001]; Mas [1999]).

Note that, for θ being the unknown parameter, in order to compute $\mathrm{E}\left\{\theta|X_1,\ldots,X_n\right\}$, with $\{X_1,\ldots,X_n\}$ denoting the functional sample, we suppose that

$$\theta_j \perp \left\{X_{i,j'},\ i\geq 1, j'\neq j\right\},$$

which leads to

$$\langle \mathrm{E}\left\{\theta|X_1,\ldots,X_n\right\}, v_j\rangle_H = \mathrm{E}\left\{\theta_j|X_1,\ldots,X_n\right\} = \mathrm{E}\left\{\theta_j|X_{1,j},\ldots,X_{n,j}\right\}.$$

Here, for each $j\geq 1$, $\theta_j=\langle\theta,v_j\rangle_H$, and $X_{i,j}=\langle X_i,v_j\rangle_H$, for each $i=1,\ldots,n$, with $\langle\cdot,\cdot\rangle_H$ being the inner product in the real separable Hilbert space H. Note that $\{v_j,\ j\geq 1\}$ denotes an orthonormal basis of H, diagonalizing the common autocovariance operator of $(X_1,\ldots,X_n)$. We can then perform an independent computation of the respective posterior distributions of the projections $\{\theta_j,\ j\geq 1\}$, of parameter θ, with respect to the orthonormal basis $\{v_j,\ j\geq 1\}$ of H.

Finally, some numerical examples are considered to illustrate the results derived on asymptotic efficiency and equivalence of moment–based classical and Beta–prior–based Bayes diagonal componentwise parameter estimators, and the associated ARH(1) plug–in predictors.

A5.2 PRELIMINARIES

The preliminary definitions and results needed in the subsequent development are introduced in this section. We first refer to the usual class of standard ARH(1) processes introduced in Bosq [2000].

Definition A5.2.1 *Let H be a real separable Hilbert space. A sequence $Y=\{Y_n,\ n\in\mathbb{Z}\}$ of H–valued random variables on a basic probability space $(\Omega,\mathcal{A},\mathcal{P})$ is called an autoregressive Hilbertian process of order one, associated with (μ,ε,ρ), if it is stationary and satisfies*

$$X_n = Y_n-\mu = \rho(Y_{n-1}-\mu)+\varepsilon_n = \rho(X_{n-1})+\varepsilon_n,\quad n\in\mathbb{Z}, \tag{A5.1}$$

where $\varepsilon=\{\varepsilon_n,\ n\in\mathbb{Z}\}$ is a Hilbert–valued white noise in the strong sense (i.e., a zero–mean stationary sequence of independent $H-$valued random variables with $\mathrm{E}\left\{\|\varepsilon_n\|_H^2\right\}=\sigma^2<\infty$, for every $n\in\mathbb{Z}$), and $\rho\in\mathcal{L}(H)$, with $\mathcal{L}(H)$ being the space of linear bounded operators on H. For each $n\in\mathbb{Z}$, ε_n and X_{n-1} are assumed to be uncorrelated.

If there exists a positive $j_0 \geq 1$ such that $\|\rho^{j_0}\|_{\mathcal{L}(H)} < 1$, then, the ARH$(1)$ process in (A5.1) is standard, and there exists a unique stationary solution to equation (A5.1) admitting a MAH(∞) representation (see [Bosq, 2000, Theorem 3.1, p. 74]).

The autocovariance and cross–covariance operators are given, for each $n \in \mathbb{Z}$, by

$$C = \mathrm{E}\{X_n \otimes X_n\} = \mathrm{E}\{X_0 \otimes X_0\}, \quad D = \mathrm{E}\{X_n \otimes X_{n+1}\} = \mathrm{E}\{X_0 \otimes X_1\}, \quad \text{(A5.2)}$$

where, for $f, g \in H$,

$$f \otimes g(h) = f \langle g, h \rangle_H, \quad \forall h \in H,$$

defines a Hilbert–Schmidt operator on H. The operator C is assumed to be in the trace class. In particular,

$$\mathrm{E}\left\{\|X_n\|_H^2\right\} < \infty, \quad n \in \mathbb{Z}.$$

It is well-known that, from equations (A5.1)–(A5.2), for all $h \in H$, $D(h) = \rho C(h)$ (see, for example, Bosq [2000]). However, since C is a nuclear or trace operator, its inverse operator is an unbounded operator in H. Different methodologies have been adopted to overcome this problem in the current literature on ARH(1) processes. In particular, here, we consider the case where $C(H) = H$, under **Assumption A2** below, since C is assumed to be strictly positive. That is, its eigenvalues are strictly positive and the kernel space of C is trivial. In addition, they are assumed to have multiplicity one. Therefore, for any $f, g \in H$, there exist $\varphi, \phi \in H$ such that $f = C(\varphi)$ and $g = C(\phi)$, and

$$\left\langle C^{-1}(f), C^{-1}(g)\right\rangle_H = \left\langle C^{-1}(C(\varphi)), C^{-1}(C(\phi))\right\rangle_H = \langle \varphi, \phi \rangle_H .$$

In particular,

$$\|C^{-1}(f)\|_H^2 < \infty, \quad \forall f \in H.$$

Assumption A1. The operator ρ in (A5.1) is self–adjoint with $\|\rho\|_{\mathcal{L}(H)} < 1$.

Assumption A2. The operator C is strictly positive, and its positive eigenvalues have multiplicity one. Furthermore, C and ρ admit the following diagonal spectral decompositions, such that for all $f, g \in H$,

$$C(g)(f) = \sum_{k=1}^{\infty} C_k \langle \phi_k, g \rangle_H \langle \phi_k, f \rangle_H \quad \text{(A5.3)}$$

$$\rho(g)(f) = \sum_{k=1}^{\infty} \rho_k \langle \phi_k, g \rangle_H \langle \phi_k, f \rangle_H , \quad \text{(A5.4)}$$

where $\{C_k,\ k \geq 1\}$ and $\{\rho_k,\ k \geq 1\}$ are the respective systems of eigenvalues of C and ρ, and $\{\phi_k,\ k \geq 1\}$ is the common system of orthonormal eigenvectors of the autocovariance operator C.

Remark A5.2.1 *As commented before, we consider here the case where the eigenvectors $\{\phi_k,\ k \geq 1\}$ of the autocovariance operator C are known. Thus, under* ***Assumption A2****, the natural way to formulate a componentwise estimator of the autocorrelation operator ρ is in terms of the respective estimators of its diagonal coefficients*

$\{\rho_k,\ k \geq 1\}$, *computed from the respective projections of the observed functional data,* $(X_0, \ldots, X_T)$, *into* $\{\phi_k,\ k \geq 1\}$. *We adopt here a moment–based classical and Beta–prior–based Bayesian approach in the estimation of such coefficients* $\{\rho_k,\ k \geq 1\}$.

From the Cauchy–Schwarz's inequality, applying the Parseval's identity,

$$\begin{aligned} |\rho(g)(f)|^2 &\leq \sum_{k=1}^{\infty} |\rho_k| \left[\langle \phi_k, g\rangle_H\right]^2 \sum_{k=1}^{\infty} |\rho_k| \left[\langle \phi_k, f\rangle_H\right]^2 \\ &\leq \sum_{k=1}^{\infty} \left[\langle \phi_k, g\rangle_H\right]^2 \sum_{k=1}^{\infty} \left[\langle \phi_k, f\rangle_H\right]^2 = \|g\|_H^2 \|f\|_H^2 < \infty. \end{aligned}$$

Thus, equation (A5.4) holds in the weak sense.

From **Assumption A2**, the projection of X_n into the common eigenvector system $\{\phi_k,\ k \geq 1\}$ leads to the following series expansion in $\mathcal{L}_H^2(\Omega, \mathcal{A}, \mathcal{P})$:

$$X_n = \sum_{k=1}^{\infty} \sqrt{C_k}\eta_k(n)\phi_k, \quad \eta_k(n) = \frac{1}{\sqrt{C_k}} \langle X_n, \phi_k\rangle_H, \tag{A5.5}$$

and, for each $j, p \geq 1$, and $n > 0$,

$$\begin{aligned} \mathrm{E}\{\eta_j(n)\eta_p(n)\} &= \mathrm{E}\left\{\frac{1}{\sqrt{C_j}} \langle X_n, \phi_j\rangle_H \frac{1}{\sqrt{C_p}} \langle X_n, \phi_p\rangle_H\right\} \\ &= \frac{1}{\sqrt{C_j}}\frac{1}{\sqrt{C_p}} C(\phi_j)(\phi_p) \\ &= \frac{1}{\sqrt{C_j}}\frac{1}{\sqrt{C_p}} C_j \langle \phi_j, \phi_p\rangle_H = \delta_{j,p}, \end{aligned}$$

where the last equality is obtained from the orthonormality of the eigenvectors $\{\phi_k,\ k \geq 1\}$. Hence, under **Assumptions A1–A2**, the projection of equation (A5.1) into the elements of the common eigenvector system $\{\phi_k,\ k \geq 1\}$ leads to the following infinite-dimensional system of equations:

$$\sqrt{C_k}\eta_k(n) = \rho_k\sqrt{C_k}\eta_k(n-1) + \varepsilon_k(n), \quad k \geq 1, \tag{A5.6}$$

or equivalently,

$$\eta_k(n) = \rho_k\eta_k(n-1) + \frac{\varepsilon_k(n)}{\sqrt{C_k}}, \quad k \geq 1, \tag{A5.7}$$

where

$$\varepsilon_k(n) = \langle \varepsilon_n, \phi_k\rangle_H, \quad k \geq 1, \quad n \in \mathbb{Z}.$$

Thus, for each $j \geq 1$,

$$\{a_j(n) = \sqrt{C_j}\eta_j(n),\ n \in \mathbb{Z}\}$$

defines a standard AR(1) process. Its moving average representation of infinite order is given by

$$a_j(n) = \sum_{k=0}^{\infty}[\rho_j]^k \varepsilon_j(n-k), \quad n \in \mathbb{Z}. \tag{A5.8}$$

Specifically, under **Assumption A2**,

$$\begin{aligned}
\mathrm{E}\{a_j(n)a_p(n)\} &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}[\rho_j]^k[\rho_p]^l \mathrm{E}\{\varepsilon_j(n-k)\varepsilon_p(n-l)\} \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}[\rho_j]^k[\rho_p]^l \delta_{k,l}\delta_{j,p} = 0, \quad j \neq p, \\
\mathrm{E}\{a_j(n)a_p(n)\} &= \sum_{k=0}^{\infty}\sigma_j^2[\rho_j]^{2k}, \quad j = p,
\end{aligned} \tag{A5.9}$$

where

$$\sigma_j^2 = \mathrm{E}\{\varepsilon_j(n-k)\}^2 = \mathrm{E}\{\varepsilon_j(0)\}^2.$$

From equation (A5.9), under **Assumptions A1–A2**,

$$\begin{aligned}
\mathrm{E}\left\{\|X(n)\|_H^2\right\} &= \sum_{j=1}^{\infty}\mathrm{E}\{a_j(n)\}^2 = \sum_{j=1}^{\infty}\sigma_j^2\sum_{k=0}^{\infty}[\rho_j]^{2k} \\
&= \sum_{j=1}^{\infty}\sigma_j^2\left[\frac{1}{1-[\rho_j]^2}\right] = \sum_{j=1}^{\infty}C_j < \infty,
\end{aligned} \tag{A5.10}$$

with, as before,

$$\sum_{j=1}^{\infty}\sigma_j^2 = \mathrm{E}\left\{\|\varepsilon_n\|_H^2\right\} < \infty.$$

Equation (A5.10) leads to the identity

$$C_j = \left[\frac{\sigma_j^2}{1-\rho_j^2}\right], \quad j \geq 1, \tag{A5.11}$$

from which, we obtain

$$\rho_k = \sqrt{1 - \frac{\sigma_k^2}{\lambda_k(C)}}, \quad \sigma_k^2 = \mathrm{E}\left\{\langle\phi_k, \varepsilon_n\rangle_H\right\}^2, \quad \forall n \in \mathbb{Z}, \quad k \geq 1. \tag{A5.12}$$

Under (A5.11), equation (A5.7) can also be rewritten as

$$\eta_k(n) = \rho_k \eta_k(n-1) + \sqrt{1-\rho_k^2}\frac{\varepsilon_k(n)}{\sigma_k}, \quad k \geq 1,$$

Assumption A2B. The sequences

$$\left\{\sigma_k^2,\ k \geq 1\right\}, \quad \{C_k,\ k \geq 1\}$$

satisfy

$$\begin{aligned} &\frac{\sigma_k^2}{C_k} \leq 1,\ k \geq 1, \quad \lim_{k\to\infty} \frac{\sigma_k^2}{C_k} = 0, \\ &\frac{\sigma_k^2}{C_k} = \mathcal{O}(k^{-1-\gamma}), \quad \gamma > 0, \quad k \to \infty. \end{aligned} \tag{A5.13}$$

Equation (A5.13) means that $\{\sigma_k^2,\ k \geq 1\}$ and $\{C_k,\ k \geq 1\}$ are both summable sequences, with faster decay to zero of the sequence $\{\sigma_k^2,\ k \geq 1\}$ than the sequence $\{C_k,\ k \geq 1\}$, leading, from equations (A5.11)–(A5.12), to the definition of $\{\rho_k^2,\ k \geq 1\}$ as a sequence with accumulation point at one.

Remark A5.2.2 *Under **Assumption A2B** and **A3** below holds.*

For each $k \geq 1$, from equations (A5.6)–(A5.8),

$$\begin{aligned} \sum_{n=1}^{T} [\eta_k(n-1)]^2 &= \frac{1}{C_k}\left[\sum_{n=1}^{T}[\varepsilon_k(n-1)]^2 \right. \\ &\quad \left. + \sum_{n=1}^{T}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}[\rho_k]^l[\rho_k]^p \varepsilon_k(n-1-l)\varepsilon_k(n-1-p)\right] \\ &= \frac{1}{C_k}\left[\sum_{n=1}^{T}[\varepsilon_k(n-1)]^2 + S(T,k)\right], \end{aligned}$$

where

$$S(T,k) = \sum_{n=1}^{T}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}[\rho_k]^l[\rho_k]^p \varepsilon_k(n-1-l)\varepsilon_k(n-1-p).$$

Hence, $\sum_{n=1}^{T}[\varepsilon_k(n-1)]^2 + S(T,k) \geq 0$, for every $T \geq 1$, and $k \geq 1$.

Assumption A3. There exists a sequence of real-valued independent random variables $\left\{\widetilde{M}(k),\ k \geq 1\right\}$ such that

$$\inf_{T\geq 1}\sqrt{\left|\frac{S(T,k)}{T\left(\sum_{n=1}^{T-1}[\varepsilon_k(n)]^2+[\varepsilon_k(0)]^2\right)}\right|} = \inf_{T\geq 1}\sqrt{\left|\frac{\sum_{n=1}^{T}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}[\rho_k]^l[\rho_k]^p\varepsilon_k(n-1-l)\varepsilon_k(n-1-p)}{T\left(\sum_{n=1}^{T-1}[\varepsilon_k(n)]^2+[\varepsilon_k(0)]^2\right)}\right|}$$
$$\geq [\widetilde{M}(k)]^{-1}\ a.s.,$$

with

$$\sum_{k=1}^{\infty}\mathrm{E}\left\{\widetilde{M}(k)\right\}^l < \infty, \quad 1 \leq l \leq 4. \tag{A5.14}$$

Remark A5.2.3 Note that the mean value of

$$\sum_{n=1}^{T}\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}[\rho_k]^l[\rho_k]^p\varepsilon_k(n-1-l)\varepsilon_k(n-1-p)$$

is of order $\frac{T\sigma_k^2}{1-(\rho_k)^2}$*, and the mean value of*

$$T\left(\sum_{n=1}^{T-1}[\varepsilon_k(n)]^2+[\varepsilon_k(0)]^2\right)$$

is of order $T(T-1)\sigma_k^2$*. Hence, for the almost surely boundedness of the inverse of*

$$\left|\frac{S(T,k)}{T\left(\sum_{n=1}^{T-1}[\varepsilon_k(n)]^2+[\varepsilon_k(0)]^2\right)}\right|,$$

by a suitable sequence of random variables with summable l*−moments, for* $l = 1, 2, 3, 4$*, the eigenvalues of operator* ρ *must be close to one but strictly less than one. As commented in Remark A5.2.2, from* ***Assumption A2B****, this condition is satisfied in view of equation (A5.12).*

Assumption A4. $\mathrm{E}\left\{\eta_j(m)\eta_k(n)\right\} = \delta_{j,k}$, with, as before, $\delta_{j,k}$ denoting the Kronecker delta function, for every $m, n \in \mathbb{Z}$, and $j, k \geq 1$.

Remark A5.2.4 ***Assumption A4*** *implies that the cross–covariance operator D admits a diagonal spectral decomposition in terms of the system of eigenvectors $\{\phi_k,\ k \geq 1\}$. Thus, under* ***Assumption A4****, the diagonal spectral decompositions (A5.3)–(A5.4) also hold.*

The classical diagonal componentwise estimator $\widehat{\rho}_T$ of ρ considered here is given by

$$\begin{aligned}
\widehat{\rho}_T &= \sum_{k=1}^{\infty} \widehat{\rho}_{k,T} \phi_k \otimes \phi_k \\
\widehat{\rho}_{k,T} &= \frac{\sum_{n=1}^{T} a_k(n-1) a_k(n)}{\sum_{n=1}^{T} [a_k(n-1)]^2} = \frac{\sum_{n=1}^{T} \langle X_{n-1}, \phi_k \rangle_H \langle X_n, \phi_k \rangle_H}{\sum_{n=1}^{T} [\langle X_{n-1}, \phi_k \rangle_H]^2} \\
&= \frac{\sum_{n=1}^{T} X_{n-1,k} X_{n,k}}{\sum_{n=1}^{T} X_{n-1,k}^2}, \quad k \geq 1.
\end{aligned} \tag{A5.15}$$

From equations (A5.6)–(A5.7) and (A5.11), for each $k \geq 1$,

$$\begin{aligned}
\widehat{\rho}_{k,T} - \rho_k &= \frac{\sum_{n=1}^{T} X_{n-1,k} X_{n,k}}{\sum_{n=1}^{T} [X_{n-1,k}]^2} - \rho_k \\
&= \frac{\sum_{n=1}^{T} \rho_k [\eta_k(n-1)]^2 + (\eta_k(n-1)\varepsilon_k(n))/\sqrt{C_k}}{\sum_{n=1}^{T} [\eta_k(n-1)]^2} - \rho_k \\
&= \rho_k + \frac{\sum_{n=1}^{T} \eta_k(n-1)\varepsilon_k(n)}{\sqrt{C_k} \sum_{n=1}^{T} [\eta_k(n-1)]^2} - \rho_k
\end{aligned}$$

$$= \frac{\sum_{n=1}^{T} \eta_k(n-1)\varepsilon_k(n)}{\sqrt{\sigma_k^2/(1-\rho_k^2)}\sum_{n=1}^{T}[\eta_k(n-1)]^2}$$

$$= \sqrt{1-\rho_k^2}\frac{\sum_{n=1}^{T} \eta_k(n-1)[\varepsilon_k(n)/\sigma_k]}{\sum_{n=1}^{T}[\eta_k(n-1)]^2}. \quad \text{(A5.16)}$$

Remark A5.2.5 It is important to note that, for instance, unconditional bases, like wavelets, provide the spectral diagonalization of an extensive family of operators, including pseudodifferential operators, and in particular, Calderón–Zygmund operators (see Kyriazis and Petrushev [2001]; Meyer and Coifman [1997]). Therefore, the diagonal spectral representations (A5.3)–(A5.4), in ***Assumption A2****, hold for a wide class of autocovariance and cross-covariance operators, for example, in terms of wavelets. When the autocovariance and the cross–covariance operators are related by a continuous function, the diagonal spectral representations (A5.3)–(A5.4) are also satisfied (see [Dautray and Lions, 1990, pp. 119, 126 and 140]).* ***Assumption A2*** *has been considered, for example, in [Bosq, 2000, Theorem 8.5, pp. 215–216; Theorem 8.7, p. 221], to establish strong consistency, although, in this book, a different setting of conditions is assumed. Thus,* ***Assumptions A1–A2*** *already have been used (e.g., in Álvarez-Liébana et al. [2017]; Bosq [2000]; Ruiz-Medina and Álvarez-Liébana [2018a]), and* ***Assumptions A2B, A3*** *and* ***A4*** *appear in Ruiz-Medina et al. [2016].* ***Assumptions A2B*** *is needed since the usual assumption on the Hilbert–Schmidt property of ρ, made by several authors, is not considered here. At the same type, as commented before,* ***Assumptions A2B*** *implies* ***Assumption A3****.*

The following lemmas will be used in the derivation of the main results of this paper, Theorems A5.4.1 and A5.4.2, obtained in the Gaussian ARH(1) context.

Lemma A5.2.1 *Let $\{\mathcal{X}_i,\ i = 1,\dots,n\}$, be the values of a standard zero–mean autoregressive process of order one (AR(1) process) at times $i = 1,\dots,n$, and*

$$\widehat{\rho}_n = \frac{\sum_{i=1}^{n} \mathcal{X}_{i-1}\mathcal{X}_i}{\sum_{i=1}^{n} \mathcal{X}_{i-1}^2},$$

with $\mathcal{X}_1$ representing the random initial condition. Assume that $|\rho| < 1$, and that the innovation process is white noise. Then, as $n \to \infty$,

$$\sqrt{n}\frac{\widehat{\rho}_n - \rho}{\sqrt{1-\rho^2}} \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1).$$

The proof of Lemma A5.2.1 can be found in [Hamilton, 1994, p. 216].

Lemma A5.2.2 *Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two normal distributed random variables having correlation $\rho_{\mathcal{X}_1\mathcal{X}_2}$, and with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2, respectively. Then, the following identities hold:*

$$\begin{aligned} \mathrm{E}\{\mathcal{X}_1\mathcal{X}_2\} &= \mu_1\mu_2 + \rho_{\mathcal{X}_1\mathcal{X}_2}\sigma_1\sigma_2 \\ \mathrm{Var}\{\mathcal{X}_1\mathcal{X}_2\} &= \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2 + 2\rho_{\mathcal{X}_1\mathcal{X}_2}\mu_1\mu_2\sigma_1\sigma_2 + \rho_{\mathcal{X}_1\mathcal{X}_2}^2\sigma_1^2\sigma_2^2 \end{aligned} \tag{A5.17}$$

(see, for example, Aroian [1947]; Ware and Lad [2003]).

Lemma A5.2.3 *For each $k \geq 1$, the following limit is obtained:*

$$\lim_{T\to\infty} T\mathrm{E}\{\widehat{\rho}_{k,T} - \rho_k\}^2 = 1 - \rho_k^2, \quad k \geq 1 \tag{A5.18}$$

(see, for example, Bartlett [1946]).

A5.3 BAYESIAN DIAGONAL COMPONENTWISE ESTIMATION

Now let us denote by R the functional random variable on the basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$, characterized by the prior distribution for ρ. In our case, we assume that R is of the form

$$R(f)(g) = \sum_{k=1}^{\infty} R_k \langle \phi_k, f\rangle_H \langle \phi_k, g\rangle_H \; a.s., \quad \forall f, g \in H,$$

where, for $k \geq 1$, R_k is a real–valued random variable such that $R(\phi_j)(\phi_k) = \delta_{j,k}R_k$, almost surely, for every $j \geq 1$. In the following, R_k is assumed to follow a beta distribution with shape parameters $a_k > 0$ and $b_k > 0$; i.e., $R_k \sim \mathcal{B}(a_k, b_k)$, for every $k \geq 1$. We also assume that R is independent of the functional components of the innovation process $\{\varepsilon_n,\ n \in \mathbb{Z}\}$, and that the random variables $\{R_k,\ k \geq 1\}$, are globally independent. That is, for each $f, g \in H$,

$$\begin{aligned} \varphi_R^{f,g}(t) &= \mathrm{E}\left\{\exp\left(it\sum_{k=1}^{\infty} R_k \langle \phi_k, f\rangle_H \langle \phi_k, g\rangle_H\right)\right\} \\ &= \prod_{k=1}^{\infty} \mathrm{E}\{\exp(itR_k \langle \phi_k, f\rangle_H \langle \phi_k, g\rangle_H)\} = \prod_{k=1}^{\infty} \varphi_{R_k}(t\langle \phi_k, f\rangle_H \langle \phi_k, g\rangle_H). \end{aligned} \tag{A5.19}$$

Thus,

$$\varphi_R(t) = \prod_{k=1}^{\infty} \varphi_{R_k}(t(\phi_k \otimes \phi_k)),$$

where the last identity is understood in the weak–sense; i.e., in the sense of equation (A5.19).

In the definition of R from $\{R_j,\ j \geq 1\}$, we can then apply the Kolmogorov extension Theorem under the condition

$$\sum_{j=1}^{\infty} \frac{a_j b_j}{(a_j + b_j + 1)(a_j + b_j)^2} < \infty$$

(see, for example, Khoshnevisan [2007]).

As in the real–valued case (see Supplementary Material A5.7), considering $b_j > 1$, for each $j \geq 1$, the Bayes estimator of ρ is defined by (see Case 2 in Supplementary Material A5.7)

$$\widetilde{\rho}_T = \sum_{j=1}^{\infty} \widetilde{\rho}_{j,T} \phi_j \otimes \phi_j, \tag{A5.20}$$

with, for every $j \geq 1$,

$$\begin{aligned}
\widetilde{\rho}_{j,T} &= \frac{1}{2\beta_{j,T}} \left[(\alpha_{j,T} + \beta_{j,T}) \pm \sqrt{(\alpha_{j,T} - \beta_{j,T})^2 - 4\beta_{j,T}\sigma_j^2 [2 - (a_j + b_j)]} \right] \\
&= \frac{\left[\sum_{i=1}^{T} x_{i-1,j} x_{i,j} + x_{i-1,j}^2 \right]}{2 \sum_{i=1}^{T} x_{i-1,j}^2} \\
&\pm \frac{\sqrt{\left[\sum_{i=1}^{T} x_{i-1,j} x_{i,j} - x_{i-1,j}^2 \right]^2 - 4\sigma_j^2 \left[\sum_{i=1}^{T} x_{i-1,j}^2 \right] [2 - (a_j + b_j)]}}{2 \sum_{i=1}^{T} x_{i-1,j}^2},
\end{aligned} \tag{A5.21}$$

where

$$\alpha_{j,T} = \sum_{i=1}^{T} x_{i-1,j} x_{i,j}, \quad \beta_{j,T} = \sum_{i=1}^{T} x_{i-1,j}^2,\ j \geq 1,\ n \geq 2. \tag{A5.22}$$

A5.4 ASYMPTOTIC EFFICIENCY AND EQUIVALENCE

In this section, sufficient conditions are derived to ensure the asymptotic efficiency and equivalence of the diagonal componentwise estimators of ρ formulated in the classical (see equation (A5.15)), and in the Bayesian (see equations (A5.20)–(A5.22)) frameworks.

Theorem A5.4.1 *Under Assumptions A1–A2, A2B, A3 and A4, let us assume that the ARH(1) process X satisfies, for each $j \geq 1$, and, for every $T \geq 2$,*

$$\sum_{i=1}^{T} \varepsilon_j(i) X_{i-1,j} \geq 0, \ a.s. \tag{A5.23}$$

That is, $\{\varepsilon_j(i),\ i \geq 1\}$ and $\{X_{i-1,j},\ i \geq 0\}$ are almost surely positive empirically correlated. In addition, for every $j \geq 1$, the hyper–parameters a_j and b_j of the beta prior distribution, $\mathcal{B}(a_j, b_j)$, are such that $a_j + b_j \geq 2$. Then, the following identities are obtained:

$$\lim_{T\to\infty} T\mathrm{E}\left\{\|\widetilde{\rho}_T^{-} - \rho\|^2_{\mathcal{S}(H)}\right\} = \lim_{T\to\infty} T\mathrm{E}\left\{\|\widehat{\rho}_T - \rho\|^2_{\mathcal{S}(H)}\right\} = \sum_{k=1}^{\infty} \frac{\sigma_k^2}{C_k} < \infty, \tag{A5.24}$$

where $\widehat{\rho}_T$ is defined in equation (A5.15), and $\widetilde{\rho}_T^{-}$ is defined from equations (A5.20)–(A5.22), considering

$$\widetilde{\rho}_{j,T}^{-} = \frac{1}{2\beta_{j,T}} \left[(\alpha_{j,T} + \beta_{j,T}) - \sqrt{(\alpha_{j,T} - \beta_{j,T})^2 - 4\beta_{j,T}\sigma_j^2[2 - (a_j + b_j)]}\right], \tag{A5.25}$$

with, as before, for each $j \geq 1$,

$$X_{i,j} = \langle X_i, \phi_j \rangle_H, \quad i = 0, \dots, T,$$

and $\alpha_{j,T}$ and $\beta_{j,T}$ are given in (A5.22), for every $T \geq 2$.

Proof. Under **Assumptions A1–A2**, from Remark A5.8.1 and Corollary A5.8.1 in Supplementary Material A5.8, for each $j \geq 1$, and for T sufficiently large,

$$|\widehat{\rho}_{j,T}| \leq 1, \quad \text{a.s.}$$

Also, under (A5.23),

$$\sum_{i=1}^{T} \rho_j X_{i-1,j}^2 + \varepsilon_j(i) X_{i-1} \geq \sum_{i=1}^{T} \rho_j X_{i-1,j}^2, \quad \text{a.s.},$$

which is equivalent to

$$\widehat{\rho}_{j,T} = \frac{\sum_{i=1}^{T} \rho_j X_{i-1,j}^2 + \varepsilon_j(i) X_{i-1}}{\sum_{i=1}^{T} X_{i-1,j}^2} \geq \rho_j, \quad \text{a.s.}, \tag{A5.26}$$

for every $j \geq 1$.

From (A5.26), to obtain the following a.s. inequality:

$$\begin{aligned} 2|\widetilde{\widehat{\rho}_{j,T}} - \rho_j| &= \left| \widehat{\rho}_{j,T} - \rho_j + 1 - \rho_j - \sqrt{(\widehat{\rho}_{j,T} - 1)^2 - \frac{4\sigma_j^2[2-(a_j+b_j)]}{\beta_{j,T}}} \right| \\ &\leq 2|\widehat{\rho}_{j,T} - \rho_j| \quad \text{a.s}, \quad j \geq 1, \end{aligned} \tag{A5.27}$$

it is sufficient that

$$-\widehat{\rho}_{j,T} + \rho_j \leq 1 - \rho_j - \sqrt{(\widehat{\rho}_{j,T} - 1)^2 - \frac{4\sigma_j^2[2-(a_j+b_j)]}{\beta_{j,T}}} \leq \widehat{\rho}_{j,T} - \rho_j \quad \text{a.s},$$

which is equivalent to

$$0 \leq -\frac{2-(a_j+b_j)}{\beta_{j,T}} \leq 4(\widehat{\rho}_{j,T} - \rho_j)(1-\rho_j)\frac{\beta_{j,T}}{4\sigma_j^2} \quad \text{a.s.}. \tag{A5.28}$$

That is, keeping in mind that

$$\sigma_j^2 = C_j(1-\rho_j^2) = C_j(1+\rho_j)(1-\rho_j),$$

condition (A5.28) can also be expressed as

$$0 \leq -\frac{2-(a_j+b_j)}{\beta_{j,T}} \leq 4(\widehat{\rho}_{j,T} - \rho_j)(1-\rho_j)\frac{\beta_{j,T}}{4C_j(1+\rho_j)(1-\rho_j)}, \quad \text{a.s.}$$

i.e.,

$$0 \leq -\frac{2-(a_j+b_j)}{\beta_{j,T}} \leq (\widehat{\rho}_{j,T} - \rho_j)\frac{\beta_{j,T}}{C_j(1+\rho_j)} \quad \text{a.s},$$

for $j \geq 1$. Since, for each $j \geq 1$,

$$\frac{\beta_{j,T}}{C_j(1+\rho_j)} \geq \frac{\beta_{j,T}}{2C_j},$$

it is sufficient that

$$0 \leq -\frac{2-(a_j+b_j)}{\beta_{j,T}} \leq (\widehat{\rho}_{j,T} - \rho_j)\frac{\beta_{j,T}}{2C_j} \quad \text{a.s.} \tag{A5.29}$$

to hold to ensure that inequality (A5.27) is satisfied. Furthermore, from Remark A5.8.1 and Corollary A5.8.1, in Supplementary Material A5.8, for each $j \geq 1$, $\beta_{j,T} \to \infty$, and

$$\beta_{j,T} = \mathcal{O}(T), \quad T \to \infty, \quad \text{a.s.}, \quad j \geq 1.$$

Also, we have, from such remark and theorem, that

$$(\widehat{\rho}_{j,T} - \rho_j) = \mathcal{O}(1), \quad T \to \infty, \quad \text{a.s.}, \quad j \geq 1.$$

Thus, for each $j \geq 1$, the upper bound, in (A5.29), diverges as $T \to \infty$, which means, that, for T sufficiently large, inequality (A5.27) holds, if $a_j + b_j \geq 2$, for each $j \geq 1$. Now, from (A5.27), under **Assumption A3**, for each $j \geq 1$,

$$T|\widehat{\rho}_{j,T} - \rho_j|^2 \leq \widetilde{M}^2(j)\, a.s., \quad T|\widetilde{\overline{\rho}}_{j,T} - \rho_j|^2 \leq T|\widehat{\rho}_{j,T} - \rho_j|^2 \leq \widetilde{M}^2(j)\, a.s. \tag{A5.30}$$

Furthermore, for each $j \geq 1$, $\beta_{j,T} \to \infty$, and $\beta_{j,T} = \mathcal{O}(T)$, as $T \to \infty$, almost surely. Hence,

$$-\frac{4\sigma_j^2[2-(a_j+b_j)]}{\beta_{j,T}} \longrightarrow 0, \quad T \longrightarrow \infty, \quad \text{a.s.}, \quad \forall j \geq 1.$$

From equation (A5.25), we then have that, for each $j \geq 1$,

$$\begin{aligned}\lim_{T\to\infty} |\widetilde{\overline{\rho}}_{j,T} - \widehat{\rho}_{j,T}| &= \lim_{T\to\infty} \left|\frac{1}{2}\left[(\widehat{\rho}_{j,T}+1) - \left((\widehat{\rho}_{j,T}-1)^2 - \frac{4}{\beta_{j,T}}\sigma_j^2[2-(a_j+b_j)]\right)^{1/2}\right] - \widehat{\rho}_{j,T}\right| \\ &= \lim_{T\to\infty} |\widehat{\rho}_{j,T} - \widehat{\rho}_{j,T}| = 0,\end{aligned} \tag{A5.31}$$

almost surely. Thus, the almost surely convergence, when $T \to \infty$, of $\widetilde{\overline{\rho}}_{j,T}$ and $\widehat{\rho}_{j,T}$ to the same limit is obtained, for every $j \geq 1$.

From equation (A5.30),

$$T[\widetilde{\overline{\rho}}_{j,T} - \widehat{\rho}_{j,T}]^2 \leq 2T\left[\left(\widetilde{\overline{\rho}}_{j,T} - \rho_j\right)^2 + \left(\widehat{\rho}_{j,T} - \rho_j\right)^2\right] \leq 4\widetilde{M}^2(j), \quad \text{a.s.} \tag{A5.32}$$

Since $\mathrm{E}\left\{\widetilde{M}^2(j)\right\} < \infty$, applying the Dominated Convergence Theorem, from equation (A5.32), considering (A5.18) we obtain, for each $j \geq 1$,

$$\lim_{T\to\infty} T\mathrm{E}\left\{\widetilde{\overline{\rho}}_{j,T} - \rho_j\right\}^2 = \lim_{T\to\infty} T\mathrm{E}\left\{\widehat{\rho}_{j,T} - \rho_j\right\}^2 = 1 - \rho_j^2. \tag{A5.33}$$

Under **Assumptions A3**, from (A5.30), for each $j \geq 1$, and for every $T \geq 1$,

$$\mathrm{E}\left\{\widehat{\rho}_{j,T} - \rho_j\right\}^2 \leq \mathrm{E}\left\{\widetilde{M}^2(j)\right\}, \quad T\mathrm{E}\left\{\widetilde{\overline{\rho}}_{j,T} - \rho_j\right\}^2 \leq \mathrm{E}\left\{\widetilde{M}^2(j)\right\}$$

with

$$\sum_{j=1}^{\infty} \mathrm{E}\left\{M^2(j)\right\} < \infty.$$

Applying again the Dominated Convergence Theorem (with integration performed with respect to a counting measure), we obtain from (A5.33), keeping in mind relationship (A5.12),

$$\begin{aligned}\lim_{T\to\infty}\sum_{j=1}^{\infty} T\mathrm{E}\left\{\widetilde{\overline{\rho}}_{j,T}-\rho_j\right\}^2 &= \sum_{j=1}^{\infty}\lim_{T\to\infty} T\mathrm{E}\left\{\widetilde{\overline{\rho}}_{j,T}-\rho_j\right\}^2 = \sum_{j=1}^{\infty}\lim_{T\to\infty} T\mathrm{E}\left\{\widehat{\rho}_{j,T}-\rho_j\right\}^2 \\ &= \sum_{j=1}^{\infty} 1-\rho_j^2 = \sum_{j=1}^{\infty}\frac{\sigma_j^2}{C_j} = \lim_{T\to\infty}\sum_{j=1}^{\infty} T\mathrm{E}\left\{\widehat{\rho}_{j,T}-\rho_j\right\}^2 < \infty,\end{aligned}$$

in view of equation (A5.13) in **Assumption A2B**. That is, equation (A5.24) holds. ■

Theorem A5.4.2 *Under the conditions of Theorem A5.4.1,*

$$\lim_{T\to\infty} T\mathrm{E}\left\{\|\widetilde{\overline{\rho}}_T(X_T)-\rho(X_T)\|_H^2\right\} = \lim_{T\to\infty} T\mathrm{E}\left\{\|\widehat{\rho}_T(X_T)-\rho(X_T)\|_H^2\right\} = \sum_{k=1}^{\infty} C_k(1-\rho_k^2). \tag{A5.34}$$

Here,

$$\begin{aligned}\widetilde{\overline{\rho}}_T(X_T) &= \sum_{j=1}^{\infty}\widetilde{\overline{\rho}}_{j,T}\left\langle X_T,\phi_j\right\rangle_H\phi_j, \\ \widetilde{\overline{\rho}}_{j,T} &= \frac{1}{2\beta_{j,T}}\left[(\alpha_{j,T}+\beta_{j,T})-\sqrt{(\alpha_{j,T}-\beta_{j,T})^2-4\beta_{j,T}\sigma_j^2[2-(a_j+b_j)]}\right],\ j\geq 1 \\ \widehat{\rho}_T(X_T) &= \sum_{j=1}^{\infty}\widehat{\rho}_{j,T}\left\langle X_T,\phi_j\right\rangle_H\phi_j, \quad \widehat{\rho}_{j,T}\frac{\sum_{i=1}^{T}X_{i-1,j}X_{i,j}}{\sum_{i=1}^{T}X_{i-1,j}^2}, \quad j\geq 1 \\ \rho(X_T) &= \sum_{j=1}^{\infty}\rho_j\left\langle X_T,\phi_j\right\rangle_H\phi_j, \quad \rho_j=\rho(\phi_j)(\phi_j), \quad j\geq 1.\end{aligned}$$

Proof.

From equation (A5.31), for every $j,k\geq 1$,

$$\left[\left(\widetilde{\overline{\rho}}_{j,T}-\widehat{\rho}_{j,T}\right)\left(\widetilde{\overline{\rho}}_{k,T}-\widehat{\rho}_{k,T}\right)\right]^2\to 0, \quad \text{a.s.,} \quad T\to\infty. \tag{A5.35}$$

In addition, from equation (A5.32), for every $j,k\geq 1$,

$$\left[\left(\widetilde{\overline{\rho}}_{j,T}-\widehat{\rho}_{j,T}\right)\left(\widetilde{\overline{\rho}}_{k,T}-\widehat{\rho}_{k,T}\right)\right]^2\leq 16\frac{\widetilde{M}^2(k)\widetilde{M}^2(j)}{T^2}\leq 16\widetilde{M}^2(k)\widetilde{M}^2(j), \tag{A5.36}$$

with

$$\mathrm{E}\left\{\widetilde{M}^2(k)\widetilde{M}^2(j)\right\} = \mathrm{E}\left\{\widetilde{M}^2(k)\right\}\mathrm{E}\left\{\widetilde{M}^2(j)\right\} < \infty,$$

under **Assumption A3**. Applying the Dominated Convergence Theorem from (A5.36), the almost surely convergence in (A5.35) implies the convergence in mean to zero, when $T \to \infty$. Furthermore, under **Assumption A3**, for $T \geq 2$,

$$\begin{aligned}
\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} T^2\mathrm{E}\left\{\left(\widetilde{\rho_{j,T}^{-}} - \widehat{\rho}_{j,T}\right)\left(\widetilde{\rho_{k,T}^{-}} - \widehat{\rho}_{k,T}\right)\right]^2 &\leq 16\left[\sum_{\substack{j,k=1\\ j\neq k}}^{\infty} \mathrm{E}\left\{\widetilde{M}^2(j)\right\}\mathrm{E}\left\{\widetilde{M}^2(k)\right\}\right] \\
&+ \left[\sum_{k=1}^{\infty}\mathrm{E}\left\{\widetilde{M}^4(k)\right\}\right] < \infty. \qquad \text{(A5.37)}
\end{aligned}$$

From (A5.37), for every $T \geq 2$,

$$\begin{aligned}
T^2\mathrm{E}\left\{\|\widetilde{\rho_T^{-}} - \widehat{\rho}_T\|_{\mathcal{S}(H)}^4\right\} &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} T^2\mathrm{E}\left\{(\widetilde{\rho_{j,T}^{-}} - \widehat{\rho}_{j,T})(\widetilde{\rho_{k,T}^{-}} - \widehat{\rho}_{k,T})\right\}^2 \\
&\leq 16\left[\sum_{\substack{j,k=1\\ j\neq k}}^{\infty} \mathrm{E}\left\{\widetilde{M}^2(j)\right\}\mathrm{E}\left\{\widetilde{M}^2(k)\right\}\right] \\
&+ \left[\sum_{k=1}^{\infty}\mathrm{E}\left\{\widetilde{M}^4(k)\right\}\right] < \infty. \qquad \text{(A5.38)}
\end{aligned}$$

Equation (A5.38) means that the rate of convergence to zero, as $T \to \infty$, of the functional sequence $\left\{\widetilde{\rho_T^{-}} - \widehat{\rho}_T,\ T \geq 2\right\}$ in the space $\mathcal{L}^4_{\mathcal{S}(H)}(\Omega, \mathcal{A}, P)$ is of order T^{-2}.

From definition of the norm in the space bounded linear operators, applying the Cauchy–Schwarz's inequality, we obtain

$$\begin{aligned}
\mathrm{E}\left\{\|\widetilde{\rho_T^{-}}(X_T) - \widehat{\rho}_T(X_T)\|_H^2\right\} &\leq \mathrm{E}\left\{\|\widetilde{\rho_T^{-}} - \widehat{\rho}_T\|_{\mathcal{L}(H)}^2\|X_T\|_H^2\right\} \\
&\leq \sqrt{\mathrm{E}\left\{\|\widetilde{\rho_T^{-}} - \widehat{\rho}_T\|_{\mathcal{L}(H)}^4\right\}}\sqrt{\mathrm{E}\left\{\|X_T\|_H^4\right\}} \\
&\leq \sqrt{\mathrm{E}\left\{\|\widetilde{\rho_T^{-}} - \widehat{\rho}_T\|_{\mathcal{S}(H)}^4\right\}}\sqrt{\mathrm{E}\left\{\|X_T\|_H^4\right\}}. \qquad \text{(A5.39)}
\end{aligned}$$

From the orthogonal expansion (A5.5) of X_T, in terms of the independent real–valued standard Gaussian random variables $\{\eta_k(T),\ k \geq 1\}$, we have

$$\begin{aligned} \mathrm{E}\left\{\|X_T\|_H^4\right\} &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} C_j C_k \mathrm{E}\left\{\eta_j(T)\eta_k(T)\right\}^2 = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} C_j C_k 3\delta_{j,k} \\ &= 3\sum_{k=1}^{\infty} C_j^2 < \infty. \end{aligned} \tag{A5.40}$$

From equations (A5.38)–(A5.40),

$$\mathrm{E}\left\{\|\widetilde{\rho_T^{-}}(X_T) - \widehat{\rho}_T(X_T)\|_H^2\right\} = \mathcal{O}\left(\frac{1}{T}\right), \quad T \to \infty.$$

Thus, $\widetilde{\rho_T^{-}}(X_T)$ and $\widehat{\rho}_T(X_T)$ have the same limit in the space $\mathcal{L}_H^2(\Omega, \mathcal{A}, \mathcal{P})$.

We now prove the approximation by $\mathrm{Tr}\left(C\left(I - \rho^2\right)\right)$ of the limit, in equation (A5.34). Consider

$$\mathrm{E}\left\{\|\widehat{\rho}_T(X_T) - \rho(X_T)\|_H^2\right\} - \mathrm{Tr}\left(C(I - \rho^2)\right) = \sum_{k=1}^{\infty} \mathrm{E}\left\{\left(\widehat{\rho}_{k,T} - \rho_k\right)^2 \eta_k^2(T)\right\} C_k - C_k(1 - \rho_k^2), \tag{A5.41}$$

where

$$\mathrm{Tr}\left(C(I - \rho^2)\right) = \sum_{k=1}^{\infty} C_k(1 - \rho_k^2).$$

From Lemmas A5.2.1– A5.2.2 (see the last identity in equation (A5.17)), for each $k \geq 1$, and for T sufficiently large,

$$\mathrm{E}\left\{\left(\widehat{\rho}_{k,T} - \rho_k\right)^2 \eta_k^2(T)\right\} \simeq \mathrm{Var}\left\{\widehat{\rho}_{k,T} - \rho_k\right\} \mathrm{Var}\left\{\eta_k\right\} \times \left(1 + 2\left[\mathrm{Corr}\left(\widehat{\rho}_{k,T} - \rho_k, \eta_k(T)\right)\right]^2\right). \tag{A5.42}$$

Under **Assumption A3**, from equations (A5.14)–(A5.16), for every $k \geq 1$,

$$T\mathrm{Var}\left\{\widehat{\rho}_{k,T} - \rho_k\right\} \leq \left(1 - \rho_k^2\right) \mathrm{E}\left\{\widetilde{M}^2(k)\right\}. \tag{A5.43}$$

From equations (A5.41)–(A5.43),

$$
\begin{aligned}
T\mathrm{E}\left\{\|\widehat{\rho}_T(X_T)-\rho(X_T)\|_H^2\right\}-\mathrm{Tr}\left(C(I-\rho^2)\right) &\leq \sum_{k=1}^{\infty}C_k(1-\rho_k^2)\mathrm{E}\left\{\widetilde{M}^2(k)\right\}\\
&\times \left[1+2\left[\mathrm{Corr}\left(\widehat{\rho}_{k,T}-\rho_k,\eta_k(T)\right)\right]^2\right]\\
&- C_k(1-\rho_k^2)\leq \sum_{k=1}^{\infty}3C_k\mathrm{E}\left\{\widetilde{M}^2(k)\right\}\\
&- \sum_{k=1}^{\infty}C_k(1-\rho_k^2)<\infty, \qquad \text{(A5.44)}
\end{aligned}
$$

since

$$
\sum_{k=1}^{\infty}C_k(1-\rho_k^2)\leq\sum_{k=1}^{\infty}C_k<\infty,
$$

by the trace property of C. Here, we have applied the Cauchy–Schwarz's inequality to obtain, for a certain constant $L>0$,

$$
\begin{aligned}
\sum_{k=1}^{\infty}3C_k\mathrm{E}\left\{\widetilde{M}^2(k)\right\} &\leq 3\sqrt{\sum_{k=1}^{\infty}C_k^2\sum_{k=1}^{\infty}\left[\mathrm{E}\left\{\widetilde{M}^2(k)\right\}\right]^2}\\
&\leq 3L\sqrt{\sum_{k=1}^{\infty}C_k\sum_{k=1}^{\infty}\mathrm{E}\left\{\widetilde{M}^2(k)\right\}}<\infty,
\end{aligned}
$$

from the trace property of C, and since

$$
\sum_{k=1}^{\infty}\mathrm{E}\left\{\widetilde{M}^2(k)\right\}<\infty,
$$

under **Assumption A3**.

From equations (A5.18) and (A5.44), one can get, applying the Dominated Convergence Theorem,

$$
\begin{aligned}
\lim_{T\to\infty}T\mathrm{E}\left\{\|\widehat{\rho}_T(X_T)-\rho(X_T)\|_H^2\right\} &= \sum_{k=1}^{\infty}C_k\lim_{T\to\infty}T\mathrm{E}\left\{\widehat{\rho}_{k,T}-\rho_k\right\}^2\\
&\times \lim_{T\to\infty}\left[1+\left[\mathrm{Corr}\left(\widehat{\rho}_{k,T}-\rho_k,\eta_k(T)\right)\right]^2\right]\\
&= \sum_{k=1}^{\infty}C_k\lim_{T\to\infty}T\mathrm{E}\left\{\widehat{\rho}_{k,T}-\rho_k\right\}^2\\
&= \sum_{k=1}^{\infty}C_k(1-\rho_k^2),
\end{aligned}
$$

where we have considered that

$$\lim_{T\to\infty} |\mathrm{Cov}\left(\widehat{\rho}_{k,T} - \rho_k, \eta_k(T)\right)|^2 \leq \lim_{T\to\infty} \mathrm{E}\left\{\widehat{\rho}_{k,T} - \rho_k\right\}^2 \mathrm{E}\left\{\eta_k(T)\right\}^2 = \lim_{T\to\infty} \frac{1-\rho_k^2}{T} = 0.$$

A5.5 NUMERICAL EXAMPLES

This section illustrates the theoretical results derived on asymptotic efficiency and equivalence of the proposed classical and Bayesian diagonal componentwise estimators of the autocorrelation operator, as well as of the associated ARH(1) plug–in predictors. Under the conditions assumed in Theorem A5.4.1, three examples of standard zero–mean Gaussian ARH(1) processes are generated, respectively corresponding to consider different rates of convergence to zero of the eigenvalues of the autocovariance operator. The truncation order k_T in Examples 1–2 (see Sections A5.5.1–A5.5.2) is fixed; i.e., it does not depend on the sample size T (see equations (A5.46)–(A5.47) below). While in Example 3 (see Section A5.5.3), k_T is selected such that

$$\lim_{T\to\infty} C_{k_T}\sqrt{T} = \infty. \tag{A5.45}$$

Specifically, in the first two examples, the choice of k_T is driven looking for a compromise between the sample size and the number of parameters to be estimated. With this aim the value $k_T = 5$ is fixed, independently of T. This is the number of parameters that can be estimated in an efficient way, from most of the values of the sample size T studied. In Example 3, the truncation parameter k_T is defined as a fractional power of the sample size. Note that Example 3 corresponds to the fastest decay velocity of the eigenvalues of the autocovariance operator. Hence, the lowest truncation order for a given sample size must be selected according to the truncation rule (A5.45).

The generation of $N = 1000$ realizations of the functional values $\{X_t,\ t = 0, 1, \dots, T\}$, for

$$T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000],$$

denoting as before the sample size, is performed, for each one of the ARH(1) processes, defined in the three examples below. Based on those generations, and on the sample sizes studied, the truncated empirical functional mean-square errors of the classical and Bayes diagonal componentwise parameter estimators of the autocorrelation operator ρ are computed as follows:

$$\mathrm{EFMSE}_{\overline{\rho}_T} = \frac{1}{N}\sum_{\omega=1}^{N}\sum_{j=1}^{k_n}\left(\overline{\rho}_{j,T}^{\omega} - \rho_j\right)^2, \tag{A5.46}$$

$$\mathrm{EFMSE}_{\overline{\rho}_T(X_T)} = \frac{1}{N}\sum_{\omega=1}^{N}\sum_{j=1}^{k_n}\left(\overline{\rho}_{j,T}^{\omega} - \rho_j\right)^2 X_{T,j}^2, \tag{A5.47}$$

where $\overline{\rho}_{j,T}^{\omega}$ can be the classical $\widehat{\rho}_{j,T}$ or the Bayes $\widetilde{\rho}_{j,T}$ diagonal componentwise estimator of the autocorrelation operator, and $^{\omega}$ denotes the sample point $\omega \in \Omega$ associated with each one of the $N = 1000$ realizations generated of each functional value of the ARH(1) process X.

On the other hand, as assumed in the previous section,

$$\rho_k \sim \mathcal{B}(a_k, b_k), \quad a_k + b_k \geq 2, \quad a_k > 0, \quad b_k > 1,$$

for each $k \geq 1$. Thus, parameters (a_k, b_k) are defined as follows:

$$b_k = 1 + 1/100, \quad a_k = 2^k, \quad k \geq 1, \tag{A5.48}$$

where

$$\mathrm{E}\{\rho_k\} = \frac{a_k}{a_k + b_k} \to 1, \quad \mathrm{Var}\{\rho_k\} = \frac{a_k b_k}{(a_k + b_k + 1)(a_k + b_k)^2} = \mathcal{O}\left(\frac{1}{2^{2k}}\right), \quad k \to \infty, \tag{A5.49}$$

with $\{\rho_k^2,\ k \geq 1\}$ being a random sequence such that its elements tend to be concentrated around point one, when $k \to \infty$. From (A5.49), since

$$\sigma_k^2 = C_k\left(1 - \rho_k^2\right), \quad k \geq 1, \tag{A5.50}$$

Assumption A2B is satisfied. In addition, condition (A5.23) is verified in the generations performed in the Gaussian framework.

A5.5.1 Example 1

Let us assume that the eigenvalues of the autocovariance operator of the ARH(1) process X are given by

$$C_k = \frac{1}{k^{3/2}}, \quad k \geq 1.$$

Thus, C is a strictly positive and trace operator, where

$$\left\{\rho_k^2,\ k \geq 1\right\}, \quad \left\{\sigma_k^2,\ k \geq 1\right\},$$

are generated from (A5.48)–(A5.50).

Tables A5.5.1–A5.5.2 display the values of the empirical functional mean–square errors, given in (A5.46)–(A5.47), associated with $\widehat{\rho}_T$ and $\widetilde{\rho}_T^{-}$, and with the corresponding ARH(1) plug–in predictors, with, as before,

$$T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000], \tag{A5.51}$$

considering $k_T = 5$. The respective graphical representations are displayed in Figures A5.5.1–A5.5.2, where, for comparative purposes, the values of the curve $1/T$ are also drawn for the finite sample sizes (A5.51).

Table A5.5.1: Example 1. Empirical functional mean-square errors $EFMSE_{\widehat{\rho}_T}$.

Sample size	Classical estimator $\widehat{\rho}_T$	Bayes estimator $\widetilde{\rho_T}$
250	$2.13\,(10)^{-3}$	$2.23\,(10)^{-3}$
500	$1.24\,(10)^{-3}$	$1.04\,(10)^{-3}$
750	$8.44\,(10)^{-4}$	$7.13\,(10)^{-4}$
1000	$6.91\,(10)^{-4}$	$5.84\,(10)^{-4}$
1250	$5.97\,(10)^{-4}$	$4.72\,(10)^{-4}$
1500	$4.89\,(10)^{-4}$	$3.98\,(10)^{-4}$
1750	$4.13\,(10)^{-4}$	$3.06\,(10)^{-4}$
2000	$3.61\,(10)^{-4}$	$2.59\,(10)^{-4}$

Table A5.5.2: Example 1. Empirical functional mean-square errors $EFMSE_{\widehat{\rho}_T(X_T)}$.

Sample size	Classical predictor $\widehat{\rho}_T\,(X_T)$	Bayes predictor $\widetilde{\rho_T}\,(X_T)$
250	$1.22\,(10)^{-3}$	$1.42\,(10)^{-3}$
500	$6.08\,(10)^{-4}$	$6.36\,(10)^{-4}$
750	$3.24\,(10)^{-4}$	$4.06\,(10)^{-4}$
1000	$3.05\,(10)^{-4}$	$2.77\,(10)^{-4}$
1250	$2.74\,(10)^{-4}$	$2.39\,(10)^{-4}$
1500	$2.07\,(10)^{-4}$	$1.78\,(10)^{-4}$
1750	$1.71\,(10)^{-4}$	$1.48\,(10)^{-4}$
2000	$1.64\,(10)^{-4}$	$1.42\,(10)^{-4}$

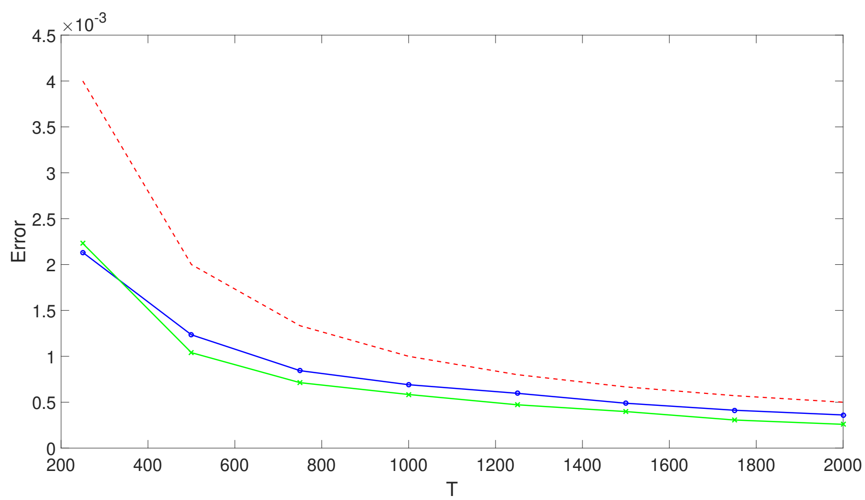


Figure A5.5.1: Example 1. Empirical functional mean-square estimation errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) parameter estimators, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000]$.

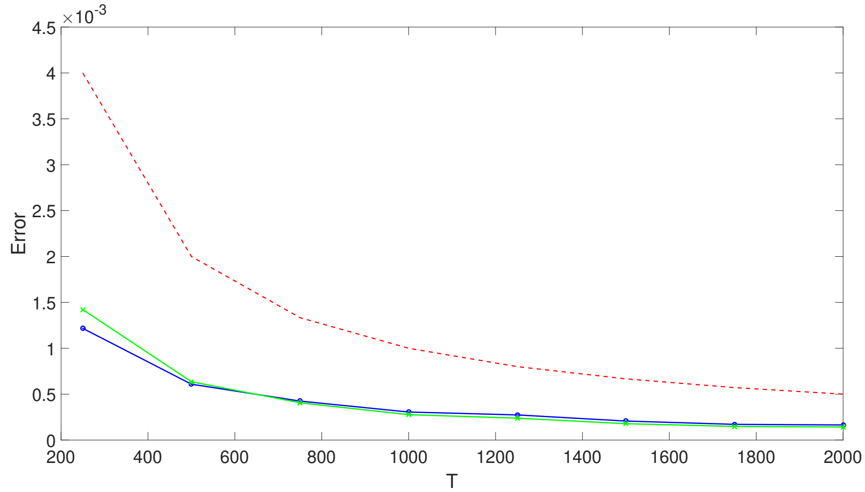


Figure A5.5.2: Example 1. Empirical functional mean-square prediction errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) plug-in predictors, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000]$.

A5.5.2 Example 2

In this example, a bit slower decay velocity, than in Example 1, of the eigenvalues of the autocovariance operator of the ARH(1) process is considered. Specifically,

$$C_k = \frac{1}{k^{1+1/10}}, \quad k \geq 1.$$

Thus, C is a strictly positive self-adjoint trace operator, where $\{\rho_k^2, \ k \geq 1\}$ and $\{\sigma_k^2, \ k \geq 1\}$ are generated, as before, from (A5.48)-(A5.50).

Tables A5.5.3–A5.5.4 show the values of the empirical functional mean–square errors, associated with $\widehat{\rho}_T$ and $\widetilde{\rho}_T^{-}$, and with the corresponding ARH(1) plug–in predictors, respectively. Figures A5.5.3–A5.5.4 provide the graphical representations in comparison with the values of the curve $1/T$ for T given in (A5.51), with, as before, $k_T = 5$.

Table A5.5.3: Example 2. Empirical functional mean–square errors $EFMSE_{\widetilde{\rho}_T}$.

Sample size	Classical estimator $\widehat{\rho}_T$	Bayes estimator $\widetilde{\rho_T}$
250	$4.18\,(10)^{-3}$	$6.09\,(10)^{-3}$
500	$2.20\,(10)^{-3}$	$2.30\,(10)^{-3}$
750	$1.52\,(10)^{-3}$	$1.39\,(10)^{-3}$
1000	$1.14\,(10)^{-3}$	$1.00\,(10)^{-3}$
1250	$9.55\,(10)^{-4}$	$7.97\,(10)^{-4}$
1500	$7.97\,(10)^{-4}$	$6.64\,(10)^{-4}$
1750	$7.01\,(10)^{-4}$	$5.37\,(10)^{-4}$
2000	$6.22\,(10)^{-4}$	$5.00\,(10)^{-4}$

Table A5.5.4: Example 2. Empirical functional mean–square errors $EFMSE_{\widetilde{\rho}_T(X_T)}$.

Sample size	Classical predictor $\widehat{\rho}_T\,(X_T)$	Bayes predictor $\widetilde{\rho_T}\,(X_T)$
250	$3.25\,(10)^{-3}$	$3.18\,(10)^{-4}$
500	$1.59\,(10)^{-3}$	$1.40\,(10)^{-4}$
750	$9.47\,(10)^{-4}$	$8.19\,(10)^{-4}$
1000	$7.89\,(10)^{-4}$	$6.88\,(10)^{-4}$
1250	$7.24\,(10)^{-4}$	$6.10\,(10)^{-4}$
1500	$5.53\,(10)^{-4}$	$4.77\,(10)^{-4}$
1750	$5.31\,(10)^{-4}$	$4.49\,(10)^{-4}$
2000	$4.61\,(10)^{-4}$	$4.00\,(10)^{-4}$

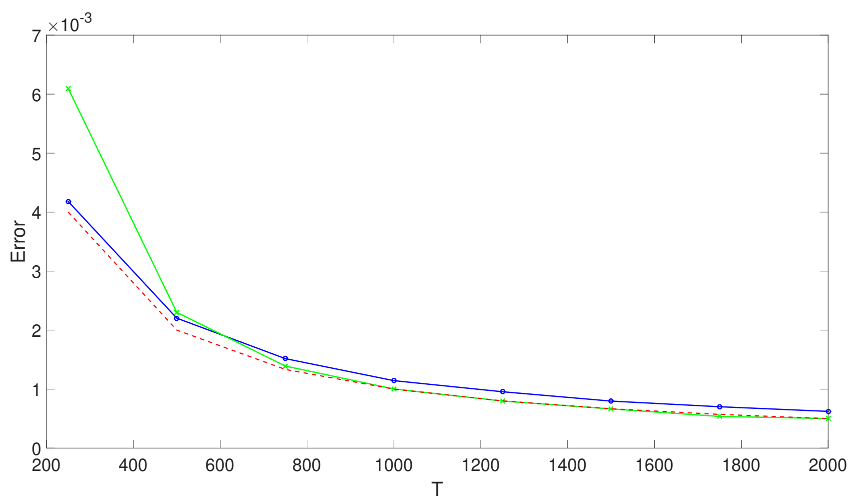


Figure A5.5.3: Example 2. Empirical functional mean–square estimation errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) parameter estimators, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000]$.

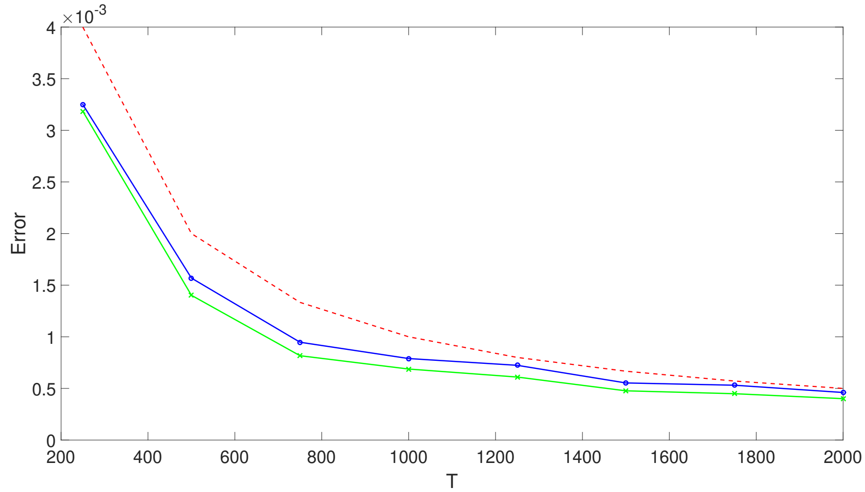


Figure A5.5.4: Example 2. Empirical functional mean–square prediction errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) plug-in predictors, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000]$.

A5.5.3 Example 3

It is well–known that the singularity of the inverse of the autocovariance operator C increases, when the rate of convergence to zero of the eigenvalues of C indicates a faster decay velocity, as in this example. Specifically, here,

$$C_k = \frac{1}{k^2}, \quad k \geq 1.$$

As before, $\{\rho_k^2,\ k \geq 1\}$ and $\{\sigma_k^2,\ k \geq 1\}$ are generated from (A5.48)-(A5.50). The truncation order k_T satisfies

$$k_T = \lceil T^{1/\alpha} \rceil, \quad \lim_{T\to\infty} k_T = \infty, \quad \lim_{T\to\infty} \sqrt{T} C_{k_T} = \infty \tag{A5.52}$$

(see also the simulation study undertaken in Álvarez-Liébana et al. [2017], for the case of ρ being a Hilbert–Schmidt operator). In particular, (A5.52) holds for $\frac{1}{2} - \frac{2}{\alpha} > 0$. Thus, $\alpha > 4$, and we consider $\alpha = 4.1$, i.e., $k_T = \lceil T^{1/4.1} \rceil$.

Tables A5.5.5–A5.5.6 show the empirical functional mean–square errors associated with $\widehat{\rho}_T$ and $\widetilde{\rho}_T^-$, and with the corresponding ARH(1) plug–in predictors, respectively. As before, Figures A5.5.5–A5.5.6 provide the graphical representations, and the values of the curve $1/T$, for T in (A5.51), with the aim of illustrating the rate of convergence to zero of the truncated empirical functional mean quadratic errors.

Table A5.5.5: Example 3. Empirical functional mean-square errors $EFMSE_{\widehat{\rho}_T}$.

Sample size	k_T	Classical estimator $\widehat{\rho}_T$	Bayes estimator $\widetilde{\rho_T}$
250	3	$1.73\,(10)^{-3}$	$1.52\,(10)^{-3}$
500	4	$9.72\,(10)^{-4}$	$1.01\,(10)^{-3}$
750	5	$6.98\,(10)^{-4}$	$7.10\,(10)^{-4}$
1000	5	$5.63\,(10)^{-4}$	$4.35\,(10)^{-4}$
1250	5	$4.49\,(10)^{-4}$	$2.84\,(10)^{-4}$
1500	5	$3.94\,(10)^{-4}$	$2.24\,(10)^{-4}$
1750	6	$3.31\,(10)^{-4}$	$1.84\,(10)^{-4}$
2000	7	$3.05\,(10)^{-4}$	$1.70\,(10)^{-4}$

Table A5.5.6: Example 3. Empirical functional mean–square errors $EFMSE_{\widehat{\rho}_T(X_T)}$.

Sample size	k_T	Classical predictor $\widehat{\rho}_T\,(X_T)$	Bayes predictor $\widetilde{\rho_T}\,(X_T)$
250	3	$1.92\,(10)^{-3}$	$1.31\,(10)^{-3}$
500	4	$8.24\,(10)^{-4}$	$5.75\,(10)^{-4}$
750	5	$5.60\,(10)^{-4}$	$4.08\,(10)^{-4}$
1000	5	$3.52\,(10)^{-4}$	$2.54\,(10)^{-4}$
1250	5	$2.62\,(10)^{-4}$	$1.45\,(10)^{-4}$
1500	5	$2.00\,(10)^{-4}$	$1.02\,(10)^{-4}$
1750	6	$1.37\,(10)^{-4}$	$9.57\,(10)^{-5}$
2000	6	$1.13\,(10)^{-4}$	$8.55\,(10)^{-5}$

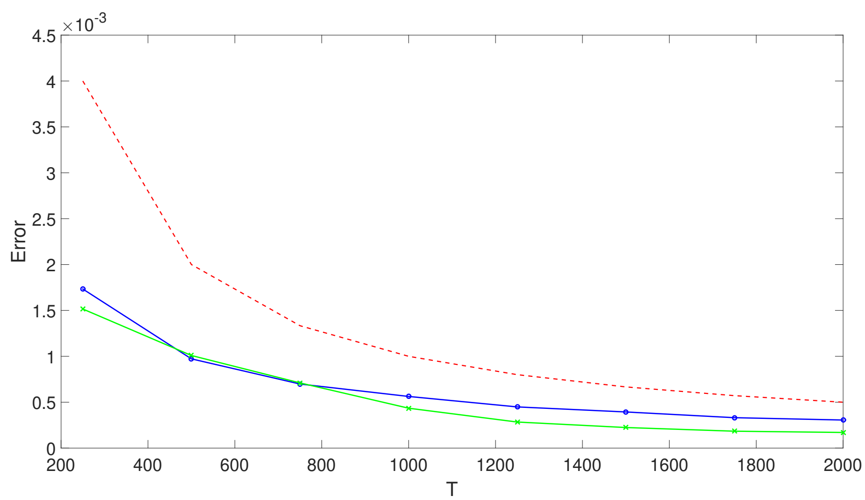


Figure A5.5.5: Example 3. Empirical functional mean–square estimation errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) parameters estimators, with $k_T = \lceil T^{1/\alpha} \rceil$, $\alpha = 4.1$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000]$.

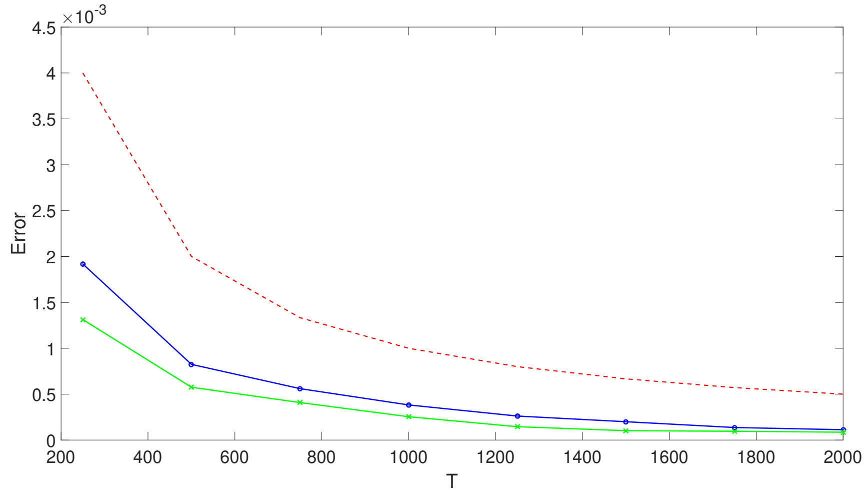


Figure A5.5.6: Example 3. Empirical functional mean–square prediction errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) plug-in predictors, with $k_T = \lceil T^{1/\alpha} \rceil$, $\alpha = 4.1$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = [250, 500, 750, 1000, 1250, 1500, 1750, 2000]$.

In Examples 1–2 in Sections A5.5.1–A5.5.2, where a common fixed truncation order is considered, we can observe that the biggest values of the empirical functional mean–square errors are located at the smallest sample sizes, for which the number $k_T = 5$ of parameters to be estimated is too large, with a slightly worse performance for those sample sizes, in Example 3 in Seciton A5.5.2, where a slower decay velocity, than in Example 1, of the eigenvalues of the autocovariance operator C is considered. Note that, on the other hand, when a slower decay velocity of the eigenvalues of C is given, a larger truncation order is required to explain a given percentage of the functional variance. For the fastest rate of convergence to zero of the eigenvalues of the autocovariance operator C, in Example 3, to compensate the singularity of the inverse covariance operator C^{-1}, a suitable truncation order k_T is fitted, depending on the sample size T, obtaining a slightly better performance than in the previous cases, where a fixed truncation order is studied.

A5.6 FINAL COMMENTS

This paper addresses the case where the eigenvectors of C are known, in relation to the asymptotic efficiency and equivalence of $\widehat{\rho}_{j,T}$ and $\widetilde{\rho}_{j,T}$, and the associated plug-in predictors. However, as shown in the simulation study undertaken in Álvarez-Liébana et al. [2017], a similar performance is obtained in the case where the eigenvectors of C are unknown (see also Bosq [2000] in relation to the asymptotic properties of the empirical eigenvectors of C).

In the cited references in the ARH(1) framework, the autocorrelation operator is usually assumed to belong to the Hilbert–Schmidt class. Here, in the absence of the compactness assumption (in particular, of the Hilbert–Schmidt assumption) on the autocorrelation operator ρ, singular autocorrelation kernels can be considered. As commented in the Section A5.1, the singularity of ρ is compensated by the regularity of the autocovariance kernel of the innovation process, as reflected in **Assumption A2B**.

Theorem A5.4.1 establishes sufficient conditions for the asymptotic efficiency and equivalence of the proposed classical and Bayes diagonal componentwise parameter estimators of ρ, as well as of the associated ARH(1) plug-in predictors (see Theorem A5.4.2). The simulation study illustrates the fact that the truncation order k_T should be selected according to the rate of convergence to zero of the eigenvalues of the autocovariance operator, and depending on the sample size T. Although, a fixed truncation order, independently of T, has also been tested in Examples 1–2, where a compromise between the rate of convergence to zero of the eigenvalues, and the rate of increasing of the sample sizes is found.

A5.7 Supplementary Material: Bayesian estimation of real–valued autoregressive processes of order one

In this section, we consider the Beta–prior–based Bayesian estimation of the autocorrelation coefficient ρ in a standard AR(1) process. Namely, the generalized maximum likelihood estimator of such a parameter is computed, when a beta prior is assumed for ρ. In the ARH(1) framework, we have adopted this estimation procedure in the approximation of the diagonal coefficients $\{\rho_k,\ k \geq 1\}$ of operator ρ with respect to $\{\phi_k \otimes \phi_k,\ k \geq 1\}$, in a Bayesian componentwise context. Note that we also denote by ρ the autocorrelation coefficient of an AR(1) process, since there is no place for confusion here.

Let $\{X_n,\ n \in \mathbb{Z}\}$ be an AR(1) process satisfying

$$X_n = \rho X_{n-1} + \varepsilon_n, \quad n \in \mathbb{Z},$$

where $0 < \rho < 1$, and $\{\varepsilon_n,\ n \in \mathbb{Z}\}$ is a real–valued Gaussian white noise; i.e., $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$, $n \in \mathbb{Z}$, are independent Gaussian random variables, with $\sigma > 0$. Here, we will use the conditional likelihood, and assume that $(x_1, \ldots, x_n)$ are observed for n sufficiently large to ensure that the effect of the random initial condition is negligible. A beta distribution with shape parameters $a > 0$ and $b > 0$ is considered as a-priori distribution on ρ, i.e., $\rho \sim \mathcal{B}(a, b)$. Hence, the distribution of $(x_1, \ldots, x_n, \rho)$ has density

$$\widetilde{L} = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \rho x_{i-1})^2\right) \rho^{a-1}(1-\rho)^{b-1}\frac{\mathbf{1}_{\{0<\rho<1\}}}{\mathbb{B}(a,b)},$$

where

$$\mathbb{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

is the beta function.

We first compute the solution to the equation

$$\begin{aligned}
0 = \frac{\partial \ln \widetilde{L}}{\partial \rho} &= \frac{\partial}{\partial \rho}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \rho x_{i-1})^2 + (a-1)\ln\rho + (b-1)\ln(1-\rho)\right] \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(-2x_{i-1}(x_i - \rho x_{i-1})\right) + \frac{a-1}{\rho} - \frac{b-1}{1-\rho} \\
&= \frac{\alpha_n}{\sigma^2} - \frac{\rho}{\sigma^2}\beta_n + \frac{a-1}{\rho} - \frac{b-1}{1-\rho},
\end{aligned}$$

where

$$\alpha_n = \sum_{i=1}^{n} x_{i-1}x_i, \quad \beta_n = \sum_{i=1}^{n} x_{i-1}^2.$$

Thus, the following equation must be solved:

$$\begin{aligned}
0 &= \frac{\rho(1-\rho)\alpha_n}{\sigma^2} - \frac{\rho^2(1-\rho)}{\sigma^2}\beta_n + (a-1)(1-\rho) - \rho(b-1) \\
0 &= \frac{\beta_n}{\sigma^2}\rho^3 - \frac{\alpha_n + \beta_n}{\sigma^2}\rho^2 + \left(\frac{\alpha_n}{\sigma^2} - [a+b] + 2\right)\rho + (a-1).
\end{aligned}$$

Case 1 Considering $a = b = 1$, and $\sigma^2 = 1$, we obtain the solution

$$\widetilde{\rho}_n = \frac{\sum_{i=1}^{n} x_{i-1}x_i}{\sum_{i=1}^{n} x_{i-1}^2}.$$

Case 2 The general case where $b > 1$ is more intricate, since the solutions are $\widetilde{\rho}_n = 0$, and

$$\begin{aligned}
\widetilde{\rho}_n &= \frac{1}{2\beta_n}\left[(\alpha_n+\beta_n) \pm \sqrt{(\alpha_n-\beta_n)^2 - 4\beta_n\sigma^2[2-(a+b)]}\right] \\
&= \frac{\displaystyle\sum_{i=1}^{n} x_{i-1}x_i + x_{i-1}^2}{\displaystyle 2\sum_{i=1}^{n} x_{i-1}^2} \\
&\pm \frac{\sqrt{\left[\displaystyle\sum_{i=1}^{n} x_{i-1}x_i - x_{i-1}^2\right]^2 - 4\sigma^2\left[\displaystyle\sum_{i=1}^{n} x_{i-1}^2\right][2-(a+b)]}}{\displaystyle 2\sum_{i=1}^{n} x_{i-1}^2}.
\end{aligned}$$

Case 3 For $\sigma^2 = a = 1$, we have

$$\begin{aligned}
\widetilde{\rho}_n &= \frac{1}{2\beta_n}\left[(\alpha_n+\beta_n) \pm \sqrt{(\alpha_n-\beta_n)^2 - 4\beta_n(1-b)}\right] \\
&= \frac{1}{\displaystyle 2\sum_{i=1}^{n} x_{i-1}^2}\left[\sum_{i=1}^{n} x_{i-1}x_i + x_{i-1}^2\right] \\
&\pm \sqrt{\left[\sum_{i=1}^{n} x_{i-1}x_i - x_{i-1}^2\right]^2 - 4\left[\sum_{i=1}^{n} x_{i-1}^2\right](1-b)}.
\end{aligned}$$

A5.8 Supplementary Material 2: strong–ergodic AR(1) processes

This section collects some strong–ergodicity results applied in this paper, for real–valued weak–dependent random sequences. In particular, their application to the AR(1) case is considered.

A real–valued stationary process $\{Y_n,\ n \in \mathbb{Z}\}$ is strongly–ergodic (or ergodic in an almost surely sense), with respect to $\mathrm{E}\{f(Y_0,\ldots,Y_{n-1})\}$ if, as $n \to \infty$,

$$\frac{1}{n-k}\sum_{i=0}^{n-1-k} f(Y_i,\ldots,Y_{i+k}) \longrightarrow^{a.s.} \mathrm{E}\{f(Y_0,\ldots,Y_{n-1})\}, \quad k \geq 0.$$

In particular, the following lemma provides sufficient condition to get the strong–ergodicity for all second–order moments (see, for example, [Stout, 1974, Theorem 3.5.8] and [Billingsley, 1995, p. 495]).

Lemma A5.8.1 *Let $\{\widetilde{\varepsilon}_n,\ n \in \mathbb{Z}\}$ be an i.i.d. sequence of real–valued random variables. If $f : \mathbb{R}^\infty \longrightarrow \mathbb{R}$ is a measurable function, then*

$$Y_n = f\left(\widetilde{\varepsilon}_n, \widetilde{\varepsilon}_{n-1}, \dots\right), \quad n \in \mathbb{Z},$$

is a stationary and strongly–ergodic process for all second–order moments.

Lemma A5.8.1 is now applied to the invertible AR(1) case, when the innovation process is white noise.

Remark A5.8.1 *If $\{Y_n,\ n \in \mathbb{Z}\}$ is a real–valued zero–mean stationary AR(1) process*

$$Y_n = \rho Y_{n-1} + \widetilde{\varepsilon}_n, \quad \rho \in \mathbb{R}, \quad |\rho| < 1, \quad n \in \mathbb{Z},$$

where $\{\widetilde{\varepsilon}_n,\ n \in \mathbb{Z}\}$ is strong white noise, we can define the measurable (even continuous) function

$$f\left(a_0, a_1, \dots\right) = \sum_{k=0}^{\infty} \rho^k a_k,$$

such that, from Lemma A5.8.1 and for each $n \in \mathbb{Z}$,

$$Y_n = \sum_{k=0}^{\infty} \rho^k \widetilde{\varepsilon}_{n-k} = f\left(\widetilde{\varepsilon}_n, \widetilde{\varepsilon}_{n-1}, \dots\right),$$

is a stationary and strongly–ergodic process for all second–order moments.

In the results derived in this paper, Remark A5.8.1 is applied, for each $j \geq 1$, to the real–valued zero–mean stationary AR(1) processes

$$\left\{X_{n,j} = \langle X_n, \phi_j \rangle_H,\ n \in \mathbb{Z}\right\},$$

with $\{X_n,\ n \in \mathbb{Z}\}$ now representing an ARH(1) process.

Corollary A5.8.1 *Under **Assumptions A1–A2**, for each $j \geq 1$, let us consider the real–valued zero–mean stationary AR(1) process $\left\{X_{n,j} = \langle X_n, \phi_j \rangle_H,\ n \in \mathbb{Z}\right\}$, such that, for each $n \in \mathbb{Z}$*

$$X_{n,j} = \rho_j X_{n-1,j} + \varepsilon_{n,j}, \quad \rho_j \in \mathbb{R}, \quad |\rho_j| < 1,$$

Here, $\{\varepsilon_{n,j},\ n \in \mathbb{Z}\}$ is a real-valued strong white noise, for any $j \geq 1$. Thus, for each $j \geq 1$, $\{X_{n,j},\ n \in \mathbb{Z}\}$ is a stationary and strongly-ergodic process for all second-order moments. In particular, for any $j \geq 1$, as $n \to \infty$,

$$\widehat{C}_{n,j} = \frac{1}{n} \sum_{i=1}^{n} X_{i-1,j}^2 \quad \longrightarrow^{a.s.} \quad C_j = \mathrm{E}\left\{X_{i-1,j}^2\right\},\ i \geq 1,$$

$$\widehat{D}_{n,j} = \frac{1}{n-1} \sum_{i=1}^{n} X_{i-1,j} X_{i,j} \quad \longrightarrow^{a.s.} \quad D_j = \mathrm{E}\left\{X_{i-1,j} X_{i,j}\right\},\ i \geq 1.$$

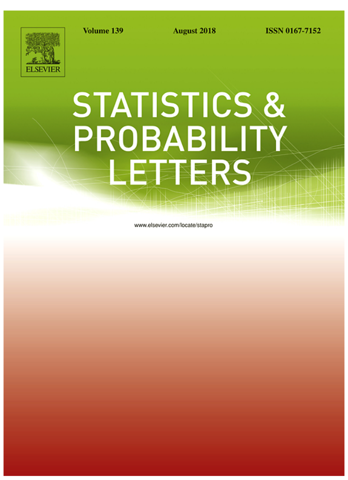
Acknowledgments

This work has been supported in part by projects MTM2012–32674 and MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

A6

A NOTE ON STRONG-CONSISTENCY OF COMPONENTWISE ARH(1) PREDICTORS

RUIZ-MEDINA, M. D.; ÁLVAREZ-LIÉBANA, J.: A note on strong–consistency of componentwise ARH(1) predictors. Statist. Probab. Letters (under minor revision, 2018)



Year	Categ.	Cites	Impact Factor (5 years)	Quartil
2016	Statist. & Probab.	3769	0.636	Q4

ABSTRACT

New results on strong–consistency, in the Hilbert–Schmidt and trace operator norms, are obtained, in the parameter estimation of an autoregressive Hilbertian process of order one (ARH(1) process). In particular, a strongly–consistent diagonal componentwise estimator of the autocorrelation operator is derived, based on its empirical singular value decomposition.

A6.1 INTRODUCTION

There exists an extensive literature on FDA techniques. In the past few years, the primary focus of FDA was mainly on i.i.d. functional observations. The classical book by Ramsay and Silverman [2005] provides a wide overview on FDA techniques (e.g., regression, principal components analysis, linear modelling, canonical correlation analysis, curve registration, and principal differential analysis). In Ferraty and Vieu [2006], an introduction to nonparametric statistics techniques for FDA was provided. We also refer to the recent monograph by Hsing and Eubank [2015], where the usual functional analysis tools in FDA are introduced, addressing several statistical and estimation problems for random elements in function spaces. The monograph by Horváth and Kokoszka [2012] is mainly concerned with inference based on second order statistics. Its most significant feature is an in depth coverage of dependent functional data structures in time and space (including functional time series, and spatially indexed functions).

We also refer to the reader to the methodological survey papers by Cuevas [2014], on the state of the art in FDA, covering nonparametric techniques and discussing central topics in FDA. The Special Issue edited by Goia and Vieu [2016] collects recent advances in the statistical analysis of high–dimensional data from the parametric, semi–parametric and nonparametric FDA frameworks, covering, in particular, functional autoregressive time series modelling, and statistical analysis techniques for spatial FDA.

A central issue in FDA is to take into account the temporal dependence of the observations. Although the literature on scalar and vector time series is huge, there are relatively few contributions dealing with functional time series. This fact is also reflected in [Rao et al., 2012, Chapter7], by Hörmann and Kokoszka [2010], where a sort review of functional time series approaches is provided. The moment–based notion of weak dependence introduced in Hörmann and Kokoszka [2010] is also accommodated to the statistical analysis of functional time series, in this chapter. Indeed, this notion does not require the specification of a data model, and can be used to study the properties of many non–linear sequences (see e.g., Berkes et al. [2011] and Hörmann [2008] for recent applications).

Except for the linear model by Bosq [2000], for functional time series, no general framework has been available in this context. The referred monograph by Bosq [2000] studies the theory of linear functional time series, both in Hilbert and Banach spaces, focusing on the functional autoregressive model. Several authors have studied the asymptotic properties of componentwise estimators of the autocorrelation operator of an ARH(1) process, and of the associated plug-in predictors. We refer to Guillas [2001]; Mas [1999, 2004, 2007], where the efficiency, consistency and asymptotic normality of these estimators are addressed, in a parametric framework (see also Álvarez-Liébana et al. [2016], on estimation of the Ornstein–Uhlenbeck processes in Banach spaces, and Álvarez-Liébana et al. [2017], on weak consistency in the Hilbert–Schmidt operator norm of componentwise estimators). Strong–consistency is derived in the norm of the space of bounded linear operators, in the monograph of Bosq [2000]. In the derivation of these results, the autocorrelation operator is usually assumed to be a Hilbert–Schmidt operator, when the eigenvectors of the autocovariance operator are unknown. This paper proves that, under basically the same setting of conditions as in the cited papers, the componentwise estimator of the autocorrelation operator proposed in Bosq [2000], based on the empirical eigenvectors of the autocovariance operator, is also strongly–consistent in the Hilbert–Schmidt and trace operator norms.

The dimension reduction problem constitutes also a central topic in the parametric, nonparametric and semi–parametric statistics for FDA. Special attention to this topic has been paid, for instance, in the context of functional regression with functional response and functional predictors (see, for example, Ferraty et al. [2012], where asymptotic normality is derived, and Ferraty et al. [2002], in the functional time series framework). In the semi–parametric and nonparametric estimation techniques, a kernel–based formulation is usually adopted. Real–valued covariates were incorporated in the novel semiparametric kernel–based proposal by Aneiros-Pérez and Vieu [2008], providing an extension to the functional partial linear time series framework (see also Aneiros-Pérez and Vieu [2006]). Motivated by spectrometry applications, a two–terms Partitioned Functional Single Index Model is introduced in Goia and Vieu [2015], in a semi–parametric framework. In the ARH(1) process framework, the present paper provides a new diagonal componentwise estimator of the autocorrelation operator, based on its empirical singular value decomposition. Its strong–consistency is proved as well. The diagonal design leads to an important dimension reduction, going beyond the usual isotropic restriction on the kernels involved in the approximation of the regression operator (respectively, autocorrelation operator), in the nonparametric framework.

The outline of the paper is the following. Appendix A6.2 introduces basic definitions and preliminary results. Appendix A6.3 derives strong–consistency of the estimator introduced in Bosq [2000], in the Hilbert–Schmidt and trace operator norms. Appendix A6.4 formulates a strongly–consistent diagonal componentwise estimator. The proofs of the results derived are given in the Supplementary Material in Appendix A6.5.

A6.2 PRELIMINARIES

Let H be a real separable Hilbert space, and let $X = \{X_n,\ n \in \mathbb{Z}\}$ be a zero-mean ARH(1) process on the basic probability space $(\Omega, \mathcal{A}, P)$ satisfying the following equation:

$$X_n(t) = \rho\left(X_{n-1}\right)(t) + \varepsilon_n(t),\ n \in \mathbb{Z}, \tag{A6.1}$$

where ρ is a bounded linear autocorrelation operator; i.e., $\rho \in \mathcal{L}(H)$, satisfying $\|\rho^k\|_{\mathcal{L}(H)} < 1$, for $k \geq k_0$, and for some k_0. The H-valued innovation process $\varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}$ is assumed to be a strong white noise, and to be uncorrelated with the random initial condition. X then admits the MAH(∞) representation

$$X_n = \sum_{k=0}^{\infty} \rho^k\left(\varepsilon_{n-k}\right), \quad n \in \mathbb{Z},$$

providing the unique stationary solution to equation (A6.1) (see Bosq [2000]).

The trace autocovariance operator of the ARH(1) process X is given by

$$C = \mathrm{E}\left\{X_n \otimes X_n\right\} = \mathrm{E}\left\{X_0 \otimes X_0\right\},$$

for $n \in \mathbb{Z}$, and the empirical autocovariance operator C_n is defined as

$$C_n = \frac{1}{n}\sum_{i=0}^{n-1} X_i \otimes X_i, \quad n \geq 2,$$

from a functional sample, $X_0, X_1, \ldots, X_{n-1}$, of the ARH$(1)$ process X.

In the following, we denote by $\{C_j,\ j \geq 1\}$ the sequence of eigenvalues of the autocovariance operator C, satisfying

$$C(\phi_j) = C_j \phi_j, \quad j \geq 1,$$

being $\{\phi_j,\ j \geq 1\}$ the associated system of orthonormal eigenvectors. For n sufficiently large, we denote by $\{C_{n,j},\ j \geq 1\}$ the empirical eigenvalues, and by $\{\phi_{n,j},\ j \geq 1\}$ the empirical eigenvectors of C_n (see [Bosq, 2000, pp. 102–103]), such that

$$C_n \phi_{n,j} = C_{n,j}\phi_{n,j},\ j \geq 1, \quad C_{n,1} \geq \cdots \geq C_{n,n} \geq 0 = C_{n,n+1} = C_{n,n+2} \ldots$$

Consider now the nuclear cross–covariance operator and its empirical version

$$D = \mathrm{E}\left\{X_n \otimes X_{n+1}\right\} = \mathrm{E}\left\{X_0 \otimes X_1\right\}, \quad D_n = \frac{1}{n-1}\sum_{i=0}^{n-2} X_i \otimes X_{i+1}, \quad n \geq 2.$$

The following assumption will appear in the subsequent developments.

Assumption A1. The random initial condition X_0 of the ARH(1) process in (A6.1) satisfies

$$\|X_0\|_H < \infty, \quad a.s.$$

Theorem A6.2.1 *(See [Bosq, 2000, Theorem 4.1, Corollary 4.1 and Theorem 4.8]). If* $\mathrm{E}\left\{\|X_0\|_H^4\right\} < \infty$, *for any* $\beta > \frac{1}{2}$, *as* $n \to \infty$,

$$\frac{n^{1/4}}{(\ln(n))^{\beta}}\|C_n - C\|_{\mathcal{S}(H)} \to^{a.s.} 0, \quad \frac{n^{1/4}}{(\ln(n))^{\beta}}\|D_n - D\|_{\mathcal{S}(H)} \to^{a.s.} 0.$$

Under ***Assumption A1****, and denoting* almost surely *identity by a.s.*

$$\|C_n - C\|_{\mathcal{S}(H)} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) \ a.s., \quad \|D_n - D\|_{\mathcal{S}(H)} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) \ a.s.,$$

where $\|\cdot\|_{\mathcal{S}(H)}$ *is the Hilbert–Schmidt operator norm.*

We will also use the notation, for a truncation parameter k_n,

$$\Lambda_{k_n} = \sup_{1 \leq j \leq k_n} \left\{ (C_j - C_{j+1})^{-1} \right\}, \quad \lim_{n \to \infty} k_n = \infty, \quad \frac{k_n}{n} < 1, \quad k_n \geq 1. \tag{A6.2}$$

A6.3 STRONG–CONSISTENCY IN THE TRACE OPERATOR NORM

This section derives the strong–consistency of the componentwise estimator $\widetilde{\rho}_{k_n}$, given in equation (A6.3) below, in the trace and Hilbert–Schmidt operator norms. In Theorem A6.3.1 below, the following lemma will be applied:

Lemma A6.3.1 *Under **Assumption A1**, if*

$$k_n \Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}} \right),$$

then, for every $x \in H$, such that $\|x\|_H \leq 1$, the following a.s. limit holds:

$$\left\| \rho(x) - \sum_{j=1}^{k_n} \langle \rho(x), \phi_{n,j} \rangle_H \phi_{n,j} \right\|_H \to_{a.s.} 0, \quad n \to \infty.$$

The proof of this lemma is given in the Supplementary Material A6.5 provided below. The following condition is assumed in the remainder of this section:

Assumption A2. The empirical eigenvalue $C_{n,k_n} > 0 \; a.s$, where k_n is the truncation parameter satisfying the conditions established in (A6.2).

Under **Assumption A2**, from a functional sample $X_0, \dots, X_{n-1}$, let us consider the componentwise estimator $\widetilde{\rho}_{k_n}$ of ρ (see [Bosq, 2000, Eq. (8.59), p. 218])

$$\begin{aligned} \widetilde{\rho}_{k_n}(x) &\underset{H}{=} \widetilde{\pi}^{k_n} D_n C_n^{-1} [\widetilde{\pi}^{k_n}]^{\star}(x) \\ &\underset{H}{=} \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \left\langle D_n C_n^{-1}(\phi_{n,j}), \phi_{n,p} \right\rangle_H \phi_{n,p} \left\langle \phi_{n,j}, x \right\rangle_H, \quad \forall x \in H, \end{aligned} \tag{A6.3}$$

where C_n^{-1} is a bounded operator on $\overline{\text{span}}^{\|\cdot\|_H} \{\phi_{n,j}, \; j = 1, \dots k_n\}$.

Theorem A6.3.1 *The following assertions hold: **(i)** Under $\mathrm{E}\left\{ \|X_0\|_H^4 \right\} < \infty$ and **Assumption A2**, consider $\rho \in \mathcal{L}(H)$, and assume Λ_{k_n} in (A6.2) satisfies*

$$\sqrt{k_n} \Lambda_{k_n} = o\left(\frac{n^{1/4}}{(\ln(n))^{\beta}} \right), \quad \beta > 1/2, \quad n \to \infty.$$

Then,

$$\|\widetilde{\rho}_{k_n} - \widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^{\star}\|_1 \to^{a.s.} 0, \quad n \to \infty, \tag{A6.4}$$

where $\widetilde{\rho}_{k_n}$ *is given in (A6.3),* $[\widetilde{\pi}^{k_n}]^{\star}$ *denotes the projection operator into the subspace*

$$\overline{\text{span}}^{\|\cdot\|_H}\{\phi_{n,j},\ j = 1, \dots k_n\} \subseteq H,$$

and $\widetilde{\pi}^{k_n}$ *is its adjoint (the inverse). Here, for a trace operator* $\mathcal{K}$ *on* H, $\|\mathcal{K}\|_1$ *represents the trace norm of* $\mathcal{K}$, *defined, for an orthonormal basis* $\{\varphi_j,\ j \geq 1\}$ *of* H, *as*

$$\|\mathcal{K}\|_1 = \sum_{j=1}^{\infty} \left\langle \sqrt{\mathcal{K}^{\star}\mathcal{K}}(\varphi_j), \varphi_j \right\rangle_H .$$

(ii) *Under **Assumptions A1–A2**, let us consider that* $k_n \Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right)$, *if* ρ *is a trace operator, then,*

$$\|\widetilde{\rho}_{k_n} - \rho\|_1 \to^{a.s.} 0, \quad n \to \infty.$$

The proof of this result is given in the Supplementary Material A6.5.

Remark A6.3.1 *Under **Assumptions A1–A2**, in the case where* ρ *is a Hilbert-Schmidt operator, and*

$$k_n \Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right), \quad \beta > 1/2, \quad n \to \infty,$$

then

$$\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{S}(H)} \to^{a.s.} 0, \quad n \to \infty,$$

since, from (A6.4) and Lemma A6.3.1, applying the Dominated Convergence Theorem,

$$\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{S}(H)} \leq \|\widetilde{\rho}_{k_n} - \widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^{\star}\|_{\mathcal{S}(H)} + \|\widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^{\star} - \rho\|_{\mathcal{S}(H)} \to_{a.s.} 0, \quad n \to \infty.$$

The strong consistency in H of the associated ARH(1) plug-in predictor $\widetilde{\rho}_{k_n}(X_{n-1})$ of X_n then follows (see also Bosq [2000] and the Supplementary Material A6.5).

A6.4 A STRONGLY–CONSISTENT DIAGONAL COMPONENTWISE ESTIMATOR

In this section, we consider the following assumption:

Assumption A3. Assume that C is strictly positive, i.e., $C_j > 0$, for every $j \geq 1$, and D is a nuclear operator such that $\rho = DC^{-1}$ is compact.

Under **Assumption A3**, ρ admits the singular value decomposition

$$\rho(x) \underset{H}{=} \sum_{j=1}^{\infty} \rho_j \langle x, \psi_j \rangle_H \widetilde{\psi}_j, \quad \forall x \in H, \tag{A6.5}$$

where, for every $j \geq 1$,

$$\rho(\psi_j) = \rho_j \widetilde{\psi}_j, \quad \rho_j \in \mathbb{C},$$

being the singular value associated with the right and left eigenvectors ψ_j and $\widetilde{\psi}_j$, respectively. Note that, since D is a nuclear operator, it also admits the singular value decomposition

$$D(x) \underset{H}{=} \sum_{j=1}^{\infty} d_j \langle x, \varphi_j \rangle_H \widetilde{\varphi}_j, \quad x \in H,$$

where $\{\varphi_j,\ j \geq 1\}$ and $\{\widetilde{\varphi}_j,\ j \geq 1\}$ are the right and left eigenvectors of D, and $\{d_j,\ j \geq 1\}$ are its singular values. Under conditions of Theorem A6.2.1, for n sufficiently large, D_n is also a nuclear operator, admitting the singular value decomposition

$$D_n(x) \underset{H}{=} \sum_{j=1}^{\infty} d_{n,j} \langle x, \varphi_{n,j} \rangle_H \widetilde{\varphi}_{n,j}, \quad x \in H,$$

in terms of the right and left eigenvectors, $\{\varphi_{n,j},\ j \geq 1\}$ and $\{\widetilde{\varphi}_{n,j},\ j \geq 1\}$, and the singular values $\{d_{n,j},\ j \geq 1\}$. Applying [Bosq, 2000, Lemma 4.2, p. 103],

$$\begin{aligned}
&\sup_{j\geq 1} |C_j - C_{n,j}| \leq \|C - C_n\|_{\mathcal{L}(H)} \leq \|C - C_n\|_{\mathcal{S}(H)} \to_{a.s.} 0,\ n \to \infty, \\
&\sup_{j\geq 1} |d_j - d_{n,j}| \leq \|D - D_n\|_{\mathcal{S}(H)} \to_{a.s.} 0,\ n \to \infty.
\end{aligned} \tag{A6.6}$$

From (A6.6), $D_n C_n^{-1}$ is compact, for n sufficiently large, admitting the singular value decomposition

$$D_n C_n^{-1}(x) = \sum_{j=1}^{n} \widehat{\rho}_{n,j} \widetilde{\psi}_{n,j} \langle x, \psi_{n,j} \rangle_H, \quad x \in H, \tag{A6.7}$$

where $D_n C_n^{-1}(\psi_{n,j}) = \widehat{\rho}_{n,j} \widetilde{\psi}_{n,j}$, for $j = 1, \dots, n$, with $\{\psi_{n,j},\ j \geq 1\}$ and $\{\widetilde{\psi}_{n,j},\ j \geq 1\}$ being the empirical right and left eigenvectors of ρ.

*Remark A6.4.1 Under **Assumption A3**, and the conditions in Theorem A6.3.1 **(ii)** (respectively, in Remark A6.3.1), for n sufficiently large, we have*

$$\begin{aligned}\sup_{x\in H:\|x\|_H\leq 1}\|D_nC_n^{-1}(x)-DC^{-1}(x)\|_H &\leq 2\|D_nC_n^{-1}\|_{\mathcal{L}(H)}\left[\sum_{j=1}^{k_n}\|\phi'_{n,j}-\phi_{n,j}\|_H+\sum_{j=k_n+1}^{\infty}\|\phi'_{n,j}\|_H\right]\\ &+ \|\widetilde{\rho}_{k_n}-DC^{-1}\|_{\mathcal{L}(H)}\to_{a.s.} 0,\quad n\to\infty.\end{aligned}$$

Thus,

$$\|D_nC_n^{-1}-DC^{-1}\|_{\mathcal{L}(H)}\to_{a.s.} 0,\quad n\to\infty. \tag{A6.8}$$

Indeed, under **Assumption A3**, equation (A6.8) holds, if the conditions assumed in Bosq [2000] for the strong–consistency of $\widetilde{\rho}_{k_n}$ in $\mathcal{L}(H)$ are satisfied. From Remark A6.4.1, and equations (A6.5) and (A6.7), applying [Bosq, 2000, Lemma 4.2, p. 103],

$$\sup_{j\geq 1}|\widehat{\rho}_{n,j}-\rho_j|\leq \|D_nC_n^{-1}-DC^{-1}\|_{\mathcal{L}(H)}\to^{a.s.} 0,\ n\to\infty.$$

Let us define the following quantity:

$$\Lambda_{k_n}^{\rho}=\sup_{1\leq j\leq k_n}\left\{\left(|\rho_j|^2-|\rho_{j+1}|^2\right)^{-1}\right\}, \tag{A6.9}$$

where k_n is a truncation parameter satisfying (A6.2). We now apply the methodology of the proof of [Bosq, 2000, Lemma 4.3, p. 104; Corollary 4.3, p. 107], to obtain the strong–consistency of the empirical right and left eigenvectors, $\{\psi_{n,j},\ j\geq 1\}$ and $\{\widetilde{\psi}_{n,j},\ j\geq 1\}$ of ρ, under the following additional assumption:

Assumption A4. Let us consider

$$\left[\sup_{j\geq 1}|\rho_j|+\sup_{j\geq 1}|\widehat{\rho}_{n,j}|\right]\leq 1.$$

Lemma A6.4.1 *Under **Assumptions A3–A4**, and the conditions of Theorem A6.3.1 **(ii)** (respectively the conditions assumed in Remark A6.3.1), if $\Lambda_{k_n}^{\rho}$ in (A6.9) is such that*

$$\Lambda_{k_n}^{\rho}=o\left(\frac{1}{\|D_nC_n^{-1}-DC^{-1}\|_{\mathcal{L}(H)}}\right),$$

then,

$$\sup_{1\leq j\leq k_n}\|\psi_{n,j}-\psi'_{n,j}\|_H\to_{a.s.} 0,\qquad \sup_{1\leq j\leq k_n}\|\widetilde{\psi}_{n,j}-\widetilde{\psi}'_{n,j}\|_H\to_{a.s.} 0,$$

where, for $j \geq 1,\ n \geq 2,$

$$\psi'_{n,j} = sgn \left\langle \psi_{n,j}, \psi_j \right\rangle_H \psi_j \quad \widetilde{\psi}'_{n,j} = sgn \left\langle \widetilde{\psi}_{n,j}, \widetilde{\psi}_j \right\rangle_H \widetilde{\psi}_j,$$

with $sgn\langle \psi_{n,j}, \psi_j \rangle_H = \mathbf{1}_{\langle \psi_{n,j}, \psi_j \rangle_H \geq 0} - \mathbf{1}_{\langle \psi_{n,j}, \psi_j \rangle_H < 0}$ *and* $sgn\langle \widetilde{\psi}_{n,j}, \widetilde{\psi}_j \rangle_H = \mathbf{1}_{\langle \widetilde{\psi}_{n,j}, \widetilde{\psi}_j \rangle_H \geq 0} - \mathbf{1}_{\langle \widetilde{\psi}_{n,j}, \widetilde{\psi}_j \rangle_H < 0}$.

The proof of this lemma is given in the Supplementary Material A6.5. The following diagonal componentwise estimator $\widehat{\rho}_{k_n}$ of ρ is formulated:

$$\widehat{\rho}_{k_n}(x) = \sum_{j=1}^{k_n} \widehat{\rho}_{n,j} \left\langle x, \psi_{n,j} \right\rangle_H \widetilde{\psi}_{n,j}, \quad x \in H,$$

where, as before, k_n is a truncation parameter satisfying the conditions assumed in Lemma A6.4.1. The next result derives the strong–consistency of $\widehat{\rho}_{k_n}$.

Theorem A6.4.1 *Under the conditions of Lemma A6.4.1, if*

$$k_n \Lambda_{k_n}^{\rho} = o\left(\frac{1}{\|D_n C_n^{-1} - DC^{-1}\|_{\mathcal{L}(H)}} \right),$$

then,

$$\|\widehat{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} \to_{a.s.} 0, \quad n \to \infty.$$

The proof of this result is given in the Supplementary Material A6.5.

A6.5 SUPPLEMENTARY MATERIAL

The proofs of the results derived above are given in this supplementary material.

Proof of Lemma A6.3.1.

Proof. Let us denote $\phi'_{n,j} = sgn\langle \phi_j, \phi_{n,j} \rangle_H \phi_j$, where $sgn\langle \phi_j, \phi_{n,j} \rangle_H = \mathbf{1}_{\langle \phi_j, \phi_{n,j} \rangle_H \geq 0} - \mathbf{1}_{\langle \phi_j, \phi_{n,j} \rangle_H < 0}$. For every $x \in H$, such that $\|x\|_H \leq 1$, applying the triangle and Cauchy–Schwarz's inequalities, we obtain, as $n \to \infty$,

$$
\begin{aligned}
&\left\| \sum_{j=1}^{k_n} \langle \rho(x), \phi_{n,j} \rangle_H \phi_{n,j} - \rho(x) \right\|_H \leq \left\| \sum_{j=1}^{k_n} \langle \rho(x), \phi_{n,j} \rangle_H \phi_{n,j} - \left\langle \rho(x), \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H \\
+ \quad & \left\| \sum_{j=k_n+1}^{\infty} \left\langle \rho(x), \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H \\
= \quad & \left\| \sum_{j=1}^{k_n} \langle \rho(x), \phi_{n,j} \rangle_H (\phi_{n,j} - \phi'_{n,j}) + \left\langle \rho(x), \phi_{n,j} - \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H \\
&+ \left\| \sum_{j=k_n+1}^{\infty} \left\langle \rho(x), \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H \leq \sum_{j=1}^{k_n} \left| \langle \rho(x), \phi_{n,j} \rangle_H \right| \|\phi_{n,j} - \phi'_{n,j}\|_H \\
&+ \left| \left\langle \rho(x), \phi_{n,j} - \phi'_{n,j} \right\rangle_H \right| \|\phi'_{n,j}\|_H + \left\| \sum_{j=k_n+1}^{\infty} \left\langle \rho(x), \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H \\
&\leq \sum_{j=1}^{k_n} \|\rho\|_{\mathcal{L}(H)} \|\phi_{n,j} - \phi'_{n,j}\|_H + \|\rho\|_{\mathcal{L}(H)} \|\phi_{n,j} - \phi'_{n,j}\|_H \\
&+ \left\| \sum_{j=k_n+1}^{\infty} \left\langle \rho(x), \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H = 2 \sum_{j=1}^{k_n} \|\rho\|_{\mathcal{L}(H)} \|\phi_{n,j} - \phi'_{n,j}\|_H \\
&+ \left\| \sum_{j=k_n+1}^{\infty} \left\langle \rho(x), \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H \\
&\leq 4\sqrt{2} \|\rho\|_{\mathcal{L}(H)} k_n \Lambda_{k_n} \|C_n - C\|_{\mathcal{S}(H)} + \left\| \sum_{j=k_n+1}^{\infty} \left\langle \rho(x), \phi'_{n,j} \right\rangle_H \phi'_{n,j} \right\|_H ,
\end{aligned}
\tag{A6.10}
$$

since, from [Bosq, 2000, Corollary 4.3, p.107],

$$
\sup_{1 \leq j \leq k_n} \|\phi_{n,j} - \phi'_{n,j}\|_H \leq 2\sqrt{2} \Lambda_{k_n} \|C_n - C\|_{\mathcal{S}(H)}.
$$

From (A6.10), under the condition

$$
k_n \Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}} \right),
$$

applying Theorem A6.2.1, we obtain

$$
\left\| \rho(x) - \sum_{j=1}^{k_n} \langle \rho(x), \phi_{n,j} \rangle_H \phi_{n,j} \right\|_H \to_{a.s.} 0, \quad n \to \infty.
$$

■

Proof of Theorem A6.3.1.

Proof.
(i) Applying the Hölder and triangle inequalities, since $\rho = DC^{-1}$ is bounded, from Theorem A6.2.1, under $\sqrt{k_n}\Lambda_{k_n} = o\left(\frac{n^{1/4}}{(\ln(n))^{\beta}}\right)$, $\beta > 1/2$,

$$\begin{aligned}
&\|\widetilde{\pi}^{k_n} D_n C_n^{-1}[\widetilde{\pi}^{k_n}]^\star - \widetilde{\pi}^{k_n} DC^{-1}[\widetilde{\pi}^{k_n}]^\star\|_1 \\
&\leq \sqrt{k_n}\|\widetilde{\pi}^{k_n} D_n C_n^{-1}[\widetilde{\pi}^{k_n}]^\star - \widetilde{\pi}_{k_n} DC^{-1}[\widetilde{\pi}_{k_n}]^\star\|_{\mathcal{S}(H)} \\
&\leq \sqrt{k_n}\|\widetilde{\pi}^{k_n} D_n C_n^{-1}[\widetilde{\pi}^{k_n}]^\star - \widetilde{\pi}^{k_n} DC_n^{-1}[\widetilde{\pi}^{k_n}]^\star\|_{\mathcal{S}(H)} \\
&+\sqrt{k_n}\|\widetilde{\pi}^{k_n} DC_n^{-1}[\widetilde{\pi}^{k_n}]^\star - \widetilde{\pi}_{k_n} DC^{-1}[\widetilde{\pi}_{k_n}]^\star\|_{\mathcal{S}(H)} \\
&= \sqrt{k_n}\|\widetilde{\pi}^{k_n}(D_n - D)C_n^{-1}[\widetilde{\pi}^{k_n}]^\star\|_{\mathcal{S}(H)} \\
&+\sqrt{k_n}\|\widetilde{\pi}^{k_n} DC^{-1}\left[CC_n^{-1}C_n - CC^{-1}C_n\right] C_n^{-1}[\widetilde{\pi}_{k_n}]^\star\|_{\mathcal{S}(H)} \\
&\leq \sqrt{k_n}C_{k_n}^{-1}\left[\|D - D_n\|_{\mathcal{S}(H)} + \|DC^{-1}\|_{\mathcal{L}(H)}\|C - C_n\|_{\mathcal{S}(H)}\right] \\
&\leq \sqrt{k_n}\Lambda_{k_n}\left[\|D - D_n\|_{\mathcal{S}(H)} + \|DC^{-1}\|_{\mathcal{L}(H)}\|C - C_n\|_{\mathcal{S}(H)}\right] \\
&\leq K\sqrt{k_n}\Lambda_{k_n}\left[\|C - C_n\|_{\mathcal{S}(H)} + \|D - D_n\|_{\mathcal{S}(H)}\right] \to_{a.s.} 0, \quad n \to \infty, \qquad \text{(A6.11)}
\end{aligned}$$

for $\|\rho\|_{\mathcal{L}(H)} \leq K$ and $K \geq 1$. Then,

$$\|\widetilde{\rho}_{k_n} - \widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^\star\|_1 \to^{a.s.} 0, \quad n \to \infty.$$

(ii) Under Assumptions A1–A2, from Theorem A6.2.1,

$$\|C_n - C\|_{\mathcal{S}(H)} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) \ a.s.,$$
$$\|D_n - D\|_{\mathcal{S}(H)} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) \ a.s..$$

Additionally, under $k_n\Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right)$, from Theorem A6.2.1, it can then be proved, in a similar way to the derivation of equation (A6.11), as $n \to \infty$,

$$\|\widetilde{\pi}^{k_n} D_n C_n^{-1}[\widetilde{\pi}^{k_n}]^\star - \widetilde{\pi}^{k_n} DC^{-1}[\widetilde{\pi}^{k_n}]^\star\|_1 \to_{a.s.} 0. \qquad \text{(A6.12)}$$

Let us now consider

$$\|\widetilde{\rho}_{k_n} - \rho\|_1 \leq \|\widetilde{\rho}_{k_n} - \widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^\star\|_1 + \|\widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^\star - \rho\|_1. \qquad \text{(A6.13)}$$

From equation (A6.12), the first term at the right-hand side of inequality (A6.13) converges a.s. to zero. From Lemma A6.3.1, $\widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^{\star}$ converges a.s. to ρ, in $\mathcal{L}(H)$, as $n \to \infty$. Since ρ is trace operator, the Dominated Convergence Theorem leads to

$$\|\widetilde{\pi}^{k_n}\rho[\widetilde{\pi}^{k_n}]^{\star} - \rho\|_1 \to_{a.s.} 0, \quad n \to \infty,$$

and

$$\|\widetilde{\rho}_{k_n} - \rho\|_1 \to^{a.s.} 0, \quad n \to \infty.$$

■

Strong-consistency of the plug-in predictor

Corollary A6.5.1 *Under the conditions of Theorem A6.3.1 **(ii)**,*

$$\|\widetilde{\rho}_{k_n}(X_{n-1}) - \rho(X_{n-1})\|_H \to_{a.s.} 0, \quad n \to \infty.$$

Proof.

Let

$$\|X_0\|_{\infty,H} = \inf\{c;\ P(\|X_0\|_H > c) = 0\} < \infty,$$

under **Assumption A1**. From Theorem A6.3.1 **(ii)** (see also Remark A6.3.1), we then have

$$\begin{aligned}
&\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} \to^{a.s.} 0, \quad n \to \infty, \\
&\|\widetilde{\rho}_{k_n}(X_{n-1}) - \rho(X_{n-1})\|_H \leq \|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)}\|X_0\|_{\infty,H} \to^{a.s.} 0,\ n \to \infty.
\end{aligned}$$

■

Proof of Lemma A6.4.1.

Proof.

Under **Assumption A3**, $\rho^{\star}\rho$, $[D_n C_n^{-1}]^{\star}[D_n C_n^{-1}]$, $\rho\rho^{\star}$ and $[D_n C_n^{-1}][D_n C_n^{-1}]^{\star}$ are self–adjoint compact operators, admitting the following diagonal spectral series representations in H :

$$\begin{aligned}
\rho^{\star}\rho &\underset{H}{=} \sum_{j=1}^{\infty} |\rho_j|^2 \psi_j \otimes \psi_j \quad [D_n C_n^{-1}]^{\star}[D_n C_n^{-1}] \underset{H}{=} \sum_{j=1}^{n} |\widehat{\rho}_{n,j}|^2 \psi_{n,j} \otimes \psi_{n,j}, \\
\rho\rho^{\star} &\underset{H}{=} \sum_{j=1}^{\infty} |\rho_j|^2 \widetilde{\psi}_j \otimes \widetilde{\psi}_j \quad D_n C_n^{-1}[D_n C_n^{-1}]^{\star} \underset{H}{=} \sum_{j=1}^{n} |\widehat{\rho}_{n,j}|^2 \widetilde{\psi}_{n,j} \otimes \widetilde{\psi}_{n,j}.
\end{aligned} \tag{A6.14}$$

From (A6.14), applying the triangle inequality,

$$\begin{aligned}
&\|\rho^\star\rho(\psi_{n,j}) - |\rho_j|^2\psi_{n,j}\|_H \le \|\rho^\star\rho(\psi_{n,j}) - [D_nC_n^{-1}]^\star[D_nC_n^{-1}](\psi_{n,j})\|_H \\
&+\|[D_nC_n^{-1}]^\star[D_nC_n^{-1}](\psi_{n,j}) - |\rho_j|^2\psi_{n,j}\|_H \\
&\le 2\|\rho^\star\rho - [D_nC_n^{-1}]^\star[D_nC_n^{-1}]\|_{\mathcal{L}(H)}.
\end{aligned} \tag{A6.15}$$

On the other hand,

$$\begin{aligned}
&\|\psi_{n,j} - \psi'_{n,j}\|_H^2 = \sum_{l=1}^{\infty}\left[\langle\psi_{n,j},\psi_l\rangle_H - \text{sgn}\,\langle\psi_{n,j},\psi_l\rangle_H\,\langle\psi_j,\psi_l\rangle_H\right]^2 \\
&= \sum_{l\neq j}\left[\langle\psi_{n,j},\psi_l\rangle_H\right]^2 + \left[\langle\psi_{n,j},\psi_j\rangle_H - \text{sgn}\,\langle\psi_{n,j},\psi_j\rangle_H\right]^2 \\
&= \sum_{l\neq j}\left[\langle\psi_{n,j},\psi_l\rangle_H\right]^2 + \left[1 - \left|\langle\psi_{n,j},\psi_j\rangle_H\right|\right]^2 \\
&= \sum_{l\neq j}\left[\langle\psi_{n,j},\psi_l\rangle_H\right]^2 + \sum_{l=1}^{\infty}\left[\langle\psi_{n,j},\psi_l\rangle_H\right]^2 - 2\left|\langle\psi_{n,j},\psi_j\rangle_H\right| + \left|\langle\psi_{n,j},\psi_j\rangle_H\right|^2 \\
&\le 2\sum_{l\neq j}\left[\langle\psi_{n,j},\psi_l\rangle_H\right]^2.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\|\rho^\star\rho(\psi_{n,j}) - |\rho_j|^2\psi_{n,j}\|_H^2 = \sum_{l=1}^{\infty}\left[\langle\psi_{n,j},|\rho_l|^2\psi_l\rangle_H - \langle\psi_{n,j},|\rho_j|^2\psi_l\rangle_H\right]^2 \\
&\ge \min_{l\neq j}\left||\rho_l|^2 - |\rho_j|^2\right|^2\sum_{l\neq j}\left[\langle\psi_{n,j},\psi_l\rangle_H\right]^2 \\
&\ge \min_{l\neq j}\left||\rho_l|^2 - |\rho_j|^2\right|^2\frac{1}{2}\|\psi_{n,j} - \psi'_{n,j}\|_H^2 \\
&\ge \alpha_j^2\frac{1}{2}\|\psi_{n,j} - \psi'_{n,j}\|_H^2,
\end{aligned} \tag{A6.16}$$

where $\alpha_1 = (|\rho_1|^2 - |\rho_2|^2)$, and

$$\alpha_j = \min\left\{|\rho_{j-1}|^2 - |\rho_j|^2, |\rho_j|^2 - |\rho_{j+1}|^2\right\}, \quad j \ge 2.$$

From equations (A6.15) and (A6.16), we have

$$\|\psi_{n,j} - \psi'_{n,j}\|_H \le a_j\|\rho^\star\rho - [D_nC_n^{-1}]^\star[D_nC_n^{-1}]\|_{\mathcal{L}(H)}, \tag{A6.17}$$

where $a_1 = 2\sqrt{2}(|\rho_1|^2 - |\rho_2|^2)^{-1}$, and

$$a_j = 2\sqrt{2}\max\left\{\left(|\rho_{j-1}|^2 - |\rho_j|^2\right)^{-1}, \left(|\rho_j|^2 - |\rho_{j+1}|^2\right)^{-1}\right\}.$$

In a similar way, considering the operators $\rho\rho^\star$ and $\widehat{\rho}_{k_n}\widehat{\rho}^\star_{k_n}$ instead of $\rho^\star\rho$ and $\widehat{\rho}^\star_{k_n}\widehat{\rho}_{k_n}$, respectively, we can obtain

$$\|\widetilde{\psi}_{n,j} - \widetilde{\psi}'_{n,j}\|_H \leq a_j\|\rho\rho^\star - [D_nC_n^{-1}][D_nC_n^{-1}]^\star\|_{\mathcal{L}(H)}. \tag{A6.18}$$

From equations (A6.17)–(A6.18),

$$\begin{aligned}\sup_{1\leq j\leq k_n}\|\psi_{n,j} - \psi'_{n,j}\|_H &\leq 2\sqrt{2}\Lambda^\rho_{k_n}\|\rho^\star\rho - [D_nC_n^{-1}]^\star[D_nC_n^{-1}]\|_{\mathcal{L}(H)}\\ \sup_{1\leq j\leq k_n}\|\widetilde{\psi}_{n,j} - \widetilde{\psi}'_{n,j}\|_H &\leq 2\sqrt{2}\Lambda^\rho_{k_n}\|\rho\rho^\star - [D_nC_n^{-1}][D_nC_n^{-1}]^\star\|_{\mathcal{L}(H)}.\end{aligned} \tag{A6.19}$$

Since, under **Assumption A4**,

$$\begin{aligned}\|\rho^\star\rho - [D_nC_n^{-1}]^\star[D_nC_n^{-1}]\|_{\mathcal{L}(H)} &\leq \|D_nC_n^{-1} - DC^{-1}\|_{\mathcal{L}(H)}\\ \|\rho\rho^\star - [D_nC_n^{-1}][D_nC_n^{-1}]^\star\|_{\mathcal{L}(H)} &\leq \|D_nC_n^{-1} - DC^{-1}\|_{\mathcal{L}(H)},\end{aligned} \tag{A6.20}$$

we obtain from Remark A6.4.1, and (A6.19)–(A6.20), keeping in mind that

$$\Lambda^\rho_{k_n} = o\left(\frac{1}{\|D_nC_n^{-1} - DC^{-1}\|_{\mathcal{L}(H)}}\right),$$

then

$$\sup_{1\leq j\leq k_n}\|\psi_{n,j} - \psi'_{n,j}\|_H \to_{a.s.} 0, \quad \sup_{1\leq j\leq k_n}\|\widetilde{\psi}_{n,j} - \widetilde{\psi}'_{n,j}\|_H \to_{a.s.} 0, \quad n\to\infty.$$

■

PROOF OF THEOREM A6.4.1.

Proof. For every $x \in H$, such that $\|x\|_H \leq 1$, we obtain

$$\begin{aligned}\|\widehat{\rho}_{k_n}(x) - \rho(x)\|_H &\leq \|\widehat{\rho}_{k_n}\widetilde{\Pi}^{k_n}(x) - \rho\Pi^{k_n}(x)\|_H + \|\rho\Pi^{k_n}(x) - \rho\widetilde{\Pi}^{k_n}(x)\|_H\\ &+ \|\rho\widetilde{\Pi}^{k_n}(x) - \rho(x)\|_H = a_n(x) + b_n(x) + c_n(x),\end{aligned}$$

where $\widetilde{\Pi}^{k_n}$ denotes the projection operator into the subspace of H generated by $\{\psi_{n,j},\ j\geq 1\}$, and Π^{k_n} is the projection operator into the subspace of H generated by $\{\psi_j,\ j\geq 1\}$.

Consider first the term $a_n(x)$. As given in Appendix A6.4 of the paper, under **Assumption A3**, from [Bosq, 2000, Lemma 4.2, p. 103] (see also Remark A6.4.1),

$$\sup_{j\geq 1}|\widehat{\rho}_{n,j}-\rho_j| \quad\leq\quad \|D_nC_n^{-1}-DC^{-1}\|_{\mathcal{L}(H)} \to^{a.s.} 0,\ n\to\infty. \tag{A6.21}$$

Applying now the triangle and the Cauchy–Schwarz's inequalities, from equations (A6.19) and (A6.21), as $n\to\infty$,

$$\begin{aligned}
&a_n(x) = \|\widehat{\rho}_{k_n}\widetilde{\Pi}^{k_n}(x)-\rho\Pi^{k_n}(x)\|_H \leq \sum_{j=1}^{k_n}|\widehat{\rho}_{n,j}-\rho_j|\left\langle x,\psi_{n,j}\right\rangle_H\|\widetilde{\psi}_{n,j}\|_H\\
&+|\rho_j|\left\langle x,\psi_{n,j}-\psi'_{n,j}\right\rangle_H\|\widetilde{\psi}_{n,j}\|_H+|\rho_j|\left\langle x,\psi'_{n,j}\right\rangle_H\|\widetilde{\psi}_{n,j}-\widetilde{\psi}'_{n,j}\|_H\\
&\leq \sum_{j=1}^{k_n}|\widehat{\rho}_{n,j}-\rho_j|+|\rho_j|\left[\|\psi_{n,j}-\psi'_{n,j}\|_H+\|\widetilde{\psi}_{n,j}-\widetilde{\psi}'_{n,j}\|_H\right]\\
&\leq k_n\|D_nC_n^{-1}-DC^{-1}\|_{\mathcal{L}(H)}\\
&+k_n\|\rho\|_{\mathcal{L}(H)}\left[2\sqrt{2}\Lambda^{\rho}_{k_n}\|\rho^{\star}\rho-[D_nC_n^{-1}]^{\star}[D_nC_n^{-1}]\|_{\mathcal{L}(H)}\right]\\
&+k_n\|\rho\|_{\mathcal{L}(H)}\left[2\sqrt{2}\Lambda^{\rho}_{k_n}\|\rho\rho^{\star}-[D_nC_n^{-1}][D_nC_n^{-1}]^{\star}\|_{\mathcal{L}(H)}\right],
\end{aligned}\tag{A6.22}$$

which converges a.s. to zero, under **Assumption A4** (see also equation (A6.20)), since

$$k_n\Lambda^{\rho}_{k_n} = o\left(\frac{1}{\|D_nC_n^{-1}-DC^{-1}\|_{\mathcal{L}(H)}}\right). \tag{A6.23}$$

Applying the triangle and Cauchy–Schwarz's inequalities, from equation (A6.19), in a similar way to (A6.22), under **Assumption A4** and (A6.23), we obtain

$$\begin{aligned}
&b_n(x) = \|\rho\Pi^{k_n}(x)-\rho\widetilde{\Pi}^{k_n}(x)\|_H \leq \sum_{j=1}^{k_n}\|x\|_H\|\psi'_{n,j}-\psi_{n,j}\|_H|\rho_j|\|\widetilde{\psi}'_{n,j}\|_H\\
&+\|x\|_H\|\|\psi_{n,j}\|_H\left\|\rho\left(\widetilde{\psi}'_{n,j}-\widetilde{\psi}_{n,j}\right)\right\|_H\\
&\leq \sum_{j=1}^{k_n}\|\psi'_{n,j}-\psi_{n,j}\|_H|\rho_j|+\|\rho\|_{\mathcal{L}(H)}\|\widetilde{\psi}'_{n,j}-\widetilde{\psi}_{n,j}\|_H\\
&\leq \|\rho\|_{\mathcal{L}(H)}\sum_{j=1}^{k_n}\|\psi'_{n,j}-\psi_{n,j}\|_H+\|\widetilde{\psi}'_{n,j}-\widetilde{\psi}_{n,j}\|_H \to_{a.s.} 0,\ n\to\infty.
\end{aligned}$$

In a similar way to the proof of Lemma A6.3.1, from (A6.19), under **Assumption A4** and (A6.23), as

$n \to \infty$,

$$c_n(x) = \|\rho\widetilde{\Pi}^{k_n}(x) - \rho(x)\|_H \leq 2\,\|\rho\|_{\mathcal{L}(H)} \sum_{j=1}^{k_n} \|\psi_{n,j} - \psi'_{n,j}\|_H$$
$$+ \left\| \sum_{j=k_n+1}^{\infty} \left\langle \rho(x), \psi'_{n,j} \right\rangle_H \psi'_{n,j} \right\|_H \to_{a.s} 0. \tag{A6.24}$$

Taking supremum in $x \in H$, such that $\|x\|_H \leq 1$, from equations (A6.22)–(A6.24), we obtain the desired result.

■

Acknowledgments

This work has been supported in part by project MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

A7

FUNCTIONAL TIME SERIES: A REVIEW AND COMPARATIVE STUDY

ÁLVAREZ-LIÉBANA, J.: Functional time series: a review and comparative study. Stat. Science (under revision, 2018)

Year	Categ.	Cites	Impact Factor (5 years)	Quartil
2016	Statist. & Probab.	5095	3.441	Q1

ABSTRACT

Since the beginning, time series framework has played a key role in the analysis of temporally correlated data. Due to the huge computing advances, data began to be gathered with an increasingly temporal resolution level, in a manner that time series, valued in a function spaces, have become crucial. This paper intends to provide to the reader a comprehensive overview about the main theoretically and computational aspects concerning the estimation and prediction of functional time series. Particularly, we pay attention to the estimation and prediction results, derived in both parametric and nonparametric frameworks, in the context of Hilbert–valued autoregressive processes of order one (ARH(1) processes, with H being a Hilbert space). A componentwise estimator of the autocorrelation operator of an ARH(1) process is also here formulated, such that its strong–consistency is proved. A comparative study between different ARH(1) prediction approaches is developed in the simulation study undertaken, aimed at illustrating to the beginners the behaviour and numerical aspects of the most used methodologies.

A7.1 INTRODUCTION

This paper presents an overview of the main estimation and prediction approaches, formulated in the context of functional time series. Henceforth, we consider as function space a real separable Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$, being $\langle \cdot, \cdot \rangle_H$ its associated inner product. Our interest particularly relies on forecasting continuous–time stochastic processes

$$\xi = \{\xi_t, \, t \geq 0\} \, ,$$

which are turned into a set of zero–mean H–valued random variables

$$X = \{X_n(t) := \xi_{n\delta+t}, \quad 0 \leq t \leq \delta, \quad n \in \mathbb{Z}\} \, ,$$

defined over a real interval $[0, \delta]$. Namely, we will focus on the estimation and prediction of the zero–mean Hilbertian autoregressive process of order one (ARH(1) process)

$$X_n \, (t) = \rho \, (X_{n-1}) \, (t) + \varepsilon_n \, (t) \, , \quad X_n, \, \varepsilon_n \in H, \quad \rho: \, H \longrightarrow H, \quad t \in [0, \delta] \, , \quad n \in \mathbb{Z}, \tag{A7.1}$$

as a functional extension of the classical zero–mean AR(1) process

$$x_n = a x_{n-1} + \epsilon_n, \quad x_n, \, \epsilon_n, \, a \in \mathbb{R}, \quad n \in \mathbb{Z}. \tag{A7.2}$$

ARH(1) framework also arises as natural extension of the multivariate time series framework, going beyond the finite–dimensional structure of the state space, to the infinite–dimensional state space, since usually separability is a required assumption. In general, a complete and separable normed space is considered, as state space of the infinite dimensional random variable. In this case, the parametric space is usually a product of suitable operator spaces, such that bounded linear operator space (correlation structure), Hilbert–Schmidt operators or trace operators (covariance structure). In equation (A7.1), ρ denotes the bounded and linear autocorrelation operator and $\varepsilon = \{\varepsilon_n, \, n \in \mathbb{Z}\}$ its H–valued innovation process. The well–known moment–based estimator of a in (A7.2), when $\{\epsilon_n, \, n \in \mathbb{Z}\}$ is assumed to be strong white noise, has usually been formulated as

$$\widehat{a} = \frac{\frac{1}{n-1}\sum_{i=0}^{n-2} x_i x_{i+1}}{\frac{1}{n}\sum_{i=0}^{n-1} x_i^2}, \quad n \geq 2. \tag{A7.3}$$

We address the approaches existing in the current literature concerning the functional estimation of ρ in (A7.1). In fact, several functional extensions of the moment–based estimation achieved in (A7.3) has already been proposed.

A7.1.1 Motivating the estimation and prediction of ARH(1) processes

Because of its high flexibility, the forecasting of functional time series framework is being crucial in a wide range of applications. As an early precedent, prediction of electricity consumption in Bologna (Italy) was studied in Cavallini et al. [1994], by applying the ARH(1) estimation methodology firstly proposed in Bosq [1991] (see Section A7.1.2 below). In the context of medical data, the analysis of electrocardiograms, by using an ARH(1) model, was performed in Bernard [1997]. Functional data depth framework was dealt in Fraiman and Muniz [2001], and applied to Nasdaq 100 Index stock prices. Forecasting of sulfur dioxide levels was addresed in de Castro et al. [2005], where the ARH(1) model was tested in comparison with functional kernel techniques. Hyndman and Ullah [2007] formulated a robust forecasting approach of the mortality and fertility rates. The stability of credit card transactions, issued by Vilnius Bank, and modelled by an ARH(1) model, was tested in Horváth et al. [2010]. An application to the analysis of biomedical data, such as white matter structures, was addressed in Sorensen et al. [2013]. A functional testing was proposed in Reimherr and Nicolae [2014], to check the nullity of covariates such as treatment effects in medicine. We also refer to Burfield et al. [2015]; Shang et al. [2016], where applications to chemical data and forecasting population in UK have been developed, respectively. The analysis of European call options, in the framework of functional autoregressive models, was developed in Liu et al. [2016]. Recently, Fischer et al. [2017] have modelled how affects the engine idling of school buses on the amount of toxic particles, by using a Bayesian functional time series framework. In addition, Functional Analysis of Variance (FANOVA), from H–valued correlated data, with spatial rectangular or circular supports, was developed in Álvarez-Liébana and Ruiz-Medina [2017], where a fixed effect model, with an ARH(1) error term, is adopted. These developments were applied to the analysis of Functional Magnetic Resonance Imaging (fMRI) techniques, such that at each voxel (3–dimensional pixel), fMRI response depends on the external stimulus by convolution with a hemodynamic response function.

A7.1.2 Background

As commented, the model displayed in (A7.1) was firstly introduced by Bosq [1991]. The functional estimation problem was addressed by a moment–based estimation, as a functional extension of (A7.3), of the linear bounded autocorrelation operator, providing the least–squares functional predictor (ARH(1) predictor). As detailed below in Section A7.2, the projection into the theoretical and empirical eigenvectors of the autocovariance operator is considered. Central limit theorems, formulated in Merlevède [1996a];

Merlevède et al. [1997], have been applied to derive the asymptotic properties of ARH(1) parameter estimators and predictors. Close graph Theorem allowed Mas [1999] to derive a truncated componentwise estimator of the adjoint of the autocorrelation operator. Enhancements to the model firstly established in Bosq [1991], under the Hilbert–Schmidt assumption over the autocorrelation operator, were presented in the comprehensive monograph by Bosq [2000]. Specifically, the asymptotic properties of the formulated truncated componentwise parameter estimator of the autocorrelation operator, and of their associated plug–in predictors, were derived. Improvements of the above–referred results were also provided in Mas [2000], where an extra regularity condition on the inverse of the autocovariance operator was imposed, to obtain the so–called resolvent class estimators (see Section A7.2 below). Efficiency of the componentwise estimator of the autocorrelation operator proposed in Bosq [2000] was studied in Guillas [2001]. Asymptotic behaviour of the ARH(1) estimators was analysed in Mas [2004, 2007] under weaker assumptions, such as the compactness of the autocorrelation operator. Weak–consistency results, in the norm of Hilbert–Schmidt operators, have recently been proposed in Álvarez-Liébana et al. [2017], when the covariance and autocorrelation operators admit a diagonal spectral decomposition, in terms of a common eigenvectors system, under the Hilbert–Schmidt assumption of the autocorrelation operator. Alternative ARH(1) estimation techniques were presented in Besse and Cardot [1996], where a spline–smoothed–penalized functional principal component analysis (spline–smoothed–penalized FPCA), with rank constraint, was performed (see also Cardot [1998]). A robust spline–smoothed–penalized FPCA estimator of the autocorrelation operator was also discussed in Besse et al. [2000]. Statistical tests for the lack of dependence, in the context of linear processes in function spaces, were derived in Kokoszka et al. [2008]. Change point analysis was applied to test the stability and stationarity of an ARH(1) process, in Horváth et al. [2010, 2014], respectively. The case of the autocorrelation operator depending on an unknown real–valued parameter has also been considered (see Kara-Terki and Mourid [2016]). This scenario can be applied to the prediction of an Ornstein–Uhlenbeck process, as performed in Álvarez-Liébana et al. [2016]. Ruiz-Medina and Álvarez-Liébana [2018a] recently provide sufficient conditions for the strong–consistency, in the trace norm, of the autocorrelation operator of an ARH(1) process, when it does not admit a diagonal spectral decomposition.

An extension of the classical ARH(1) to ARH(p) processes, with p greater than one, was proposed in Bosq [2000]. An extended class of ARH(1) processes, known as ARHX(1) processes, by including exogenous variables in their dependence structure, was firstly formulated in Guillas [2000], and subsequently dealt by Damon and Guillas [2002, 2005]. The first derivatives of an ARH(1) process were considered in Marion and Pumo [2004], as the exogenous variables to be included in the model, by introducing the so–called ARHD(1) process. Conditional autoregressive Hilbertian processes of order one (CARH(1) processes) were introduced in Guillas [2002], aimed at including exogenous information in a non–additive way. Mourid [2004] considers the randomness of the autocorrelation operator by conditioning to each element of the sample space. Weakly dependent processes were analysed in Hörmann and Kokoszka [2010]. Hilbertian periodically correlated autoregressive processes of order one (PCARH(1) processes) have been defined by Soltani and Hashemi [2011], and later extended to the Banach–valued context by Parvardeh et al. [2017]. Spatial extension of the classical ARH(1) models was firstly proposed in Ruiz-Medina [2011]. Their moment–based estimation was detailed in Ruiz-Medina [2012]. Recently, Ruiz-Medina and Álvarez-Liébana [2017a] have derived the asymptotic efficiency and equivalence of both, classical and Beta–prior–based Bayesian diagonal componentwise ARH(1) parameter estimators and predictors, when the autocorrelation operator is not assumed to be a compact operator.

Concerning alternative bases, Grenander's theory of sieves was adapted by Bensmain and Mourid [2001], for the estimation of the autocorrelation operator of an ARH(1) process, from a Fourier–basis–based decomposition in a finite dimensional subspace. Antoniadis and Sapatinas [2003] suggested three linear wavelet–based predictors, two of them are built from the componentwise and resolvent estimators of the autocorrelation operator, already established in Bosq [2000]; Mas [2000], respectively. Focusing on the predictor, the idea of replacing the FPC with the directions more relevant to forecasting, by searching a reduced–rank approximation, was firstly exhibited in Kargin and Onatski [2008]. As an extension of the work by Hyndman and Ullah [2007], a weighted version of the FPLSR and FPCA approaches was established in Hyndman and Shang [2009], with a decreasing weighting over time, as often, e.g., in demography. For the purpose of taking into account the information coming from the dynamic dependence, which is usually ignored in the FPCA literature, a dynamic functional principal components approach was simultaneously introduced by Hörmann et al. [2015]; Panaretos and Tavakoli [2013a].

Moving–average Hilbertian processes (MAH processes), and autoregressive and moving–average Hilbertian processes (ARMAH processes), can be defined as a particular case of Hilbertian general linear processes (LPH). From the previous above–referred work by Bosq [1991], sufficient conditions for the invertibility of LPH were provided in Merlevède [1995, 1996b]. A Markovian representation of a stationary and invertible LPH, as well as a consistent plug–in predictor, was derived in Merlevède [1997]. The conditional central limit theorem was extended to functional stochastic processes in Dedecker and Merlevède [2003], allowing its application to LPH. The weak convergence for the empirical autocovariance and cross–covariance operators of LPH was proved in Mas [2002]. Further results, that those one formulated in Bosq [2000] for LPH, were obtained by Bosq [2007]; Bosq and Blanke [2007], where the study of a consistent predictor of MAH processes was also addressed. Componentwise estimation of a MAH(1) process was studied in Turbillon et al. [2008], under properly assumptions. Wang [2008] proposed a real–valued non–linear ARMA model, in which functional MA coefficients were considered. An extensive literature has also been developed concerning the Banach–valued time series framework. Nevertheless, Banach–valued context is out of the scope of this paper, but new results developed in Ruiz-Medina and Álvarez-Liébana [2018b], on the strong–consistency estimation and prediction of an ARB(1) process in the norm of bounded linear operators on a Banach space, are strongly recommended.

A great amount of authors have been developed alternative nonparametric prediction techniques, in both functional time series and functional regression frameworks, where the main goal is to forecast the predictable part of the paths. Besse et al. [2000] formulated a functional nonparametric kernel–based predictor of an ARH(1) process. nonparametric methods were proposed in Cuevas et al. [2002], in the estimation of the underlying linear operator of a functional linear regression, where both explanatory and response variables are valued in a function space. A two–steps prediction approach, based on a nonparametric kernel–based prediction of the scaling coefficients, with respect to a wavelet basis decomposition, was firstly exhibited in Antoniadis et al. [2006]. This method, also so–called kernel wavelet functional (KWF) method, was improved in Cugliari [2011]. In the case where the response is a Hilbert–valued variable, and the explanatory variable takes its values in a general function space, equipped with a semi–metric, Ferraty et al. [2012] obtained a nonparametric kernel estimator of the underlying regression operator.

A7.1.3 Outline

The outline of this paper is as follows. References detailed in Sections A7.2–A7.6 are divided by thematic areas in a chronicle. Section A7.2 is devoted to the study of the different ARH(1) componentwise estimation frameworks, based on the projections into the theoretical and empirical eigenvectors of the autocovariance operator. Section A7.3 deals improvements of the classical ARH(1) framework. Parametric forecasting of functional time series, based on the projection into alternative bases, such as Fourier, B–spline or wavelet bases, will be presented in Section A7.4. Section A7.5 studies MAH processes, as a particular case of LPH. nonparametric techniques are described in Section A7.6. We formulate in Section A7.7 new results on the strong–consistency of a truncated componentwise estimation, under a diagonal framework. In Section A7.8, a comparative study is undertaken to illustrate the performance of some of the most used methodologies. Specifically, the approaches presented in Section A7.7.1, as well as those ones in Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001] are compared. Auxiliary results are provided in the Supplementary Material (see Sections A7.9.1–A7.9.3), where the numerical results here obtained are outlined as well (see Sections A7.9.3–A7.9.4 in the Supplementary Material provided).

A7.2 ARH(1) componentwise estimation, based on the eigenvectors of the autocovariance operator

ARH(1) process introduced by Bosq [1991] seeks to extend the classical AR(1) model in (A7.2) to functional data. In the sequel, let us consider zero–mean stationary processes. The ARH(1) process was defined by

$$X_n(t) = \rho\left(X_{n-1}\right)(t) + \varepsilon_n(t), \quad X_n,\, \varepsilon_n \in H, \quad n \in \mathbb{Z}, \quad \rho \in \mathcal{L}(H),$$

where $\mathcal{L}(H)$ is the space of bounded linear operators on H. If $\varepsilon = \{\varepsilon_n,\, n \in \mathbb{Z}\}$ is assumed to be a H–valued strong white noise and uncorrelated with the initial condition,

$$\sum_{j=0}^{\infty} \left\|\rho^j\right\|^2_{\mathcal{L}(H)} < \infty$$

is required to the stationarity condition. From the central limit theorems formulated in Merlevède [1996a]; Merlevède et al. [1997], the following asymptotic results for the autocovariance operator $C = \mathrm{E}\left\{X_n \otimes X_n\right\}$, for $n \in \mathbb{Z}$, under

$$\mathrm{E}\left\{\|X_0\|_H^4\right\} < \infty, \quad \|X_0\|_H < \infty \quad \text{(so–called **Assumption A3**)},$$

were obtained in Bosq [1999a,b]:

$$\|C_n - C\|_{\mathcal{S}(H)} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) \; a.s., \quad C_n = \frac{1}{n}\sum_{i=0}^{n-1} X_i \otimes X_i, \quad n \geq 2,$$

being $\mathcal{S}(H)$ the class of Hilbert–Schmidt operators on H. Since C is compact, from the close graph Theorem, the adjoint of the autocorrelation operator $\rho^* = C^{-1}D$ is bounded of the domain of C^{-1}, which is a dense subspace in H. Then, the autocorrelation operator can be built as $\rho = (DC^{-1})^{**}$. Mas [1999] provided the asymptotic normality of the formulated estimator of ρ^*, under

$$\mathrm{E}\left\{\left\|C^{-1}\varepsilon_0\right\|_H^2\right\} < \infty,$$

by projecting into

$$H_{k_n} = sp\left(\phi_1, \dots, \phi_{k_n}\right),$$

when the eigenvectors $\{\phi_j,\ j \geq 1\}$ of C are assumed to be known. Assumption

$$C_1 > C_2 > \ldots > C_j > \ldots > 0 \quad (\textbf{Assumption A1})$$

was imposed, where $\{C_j,\ j \geq 1\}$ denote eigenvalues of C.

The asymptotic results formulated in Merlevède [1996a]; Merlevède et al. [1997] were also crucial in the derivation of some extra asymptotic properties for C and $D = \mathrm{E}\{X_n \otimes X_{n+1}\}$ by Bosq [2000]. In particular, under **Assumption A3**, for any $\beta > 1/2$,

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} \|C_n - C\|_{\mathcal{S}(H)} \longrightarrow 0\, a.s., \qquad \frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} \|D_n - D\|_{\mathcal{S}(H)} \longrightarrow 0\, a.s.,$$

was proved, being

$$D_n = \frac{1}{n-1}\sum_{i=0}^{n-2} X_i \otimes X_{i+1}$$

the empirical estimator of the cross–covariance operator, for each $n \geq 2$. Under **Assumptions A1** and **A3**, as well as the Hilbert–Schmidt assumption over ρ, when a spectral decomposition of C_n is achieved in terms of $\{C_{n,j},\ j \geq 1\}$ and $\{\phi_{n,j},\ j \geq 1\}$, the strong–consistency of the following non–diagonal estimator $\widetilde{\rho}_n(x) = \left(\widetilde{\pi}^{k_n} D_n C_n^{-1} \widetilde{\pi}^{k_n}\right)(x)$ was derived in the above–referred work:

$$\widetilde{\rho}_n(x) = \sum_{l=1}^{k_n} \widetilde{\rho}_{n,l}(x)\phi_{n,l}, \quad \widetilde{\rho}_{n,l}(x) = \sum_{j=1}^{k_n} C_{n,j}^{-1}\left(\frac{1}{n-1}\sum_{i=0}^{n-2} \widetilde{X}_{i,n,j}\widetilde{X}_{i+1,n,l}\right)\langle x, \phi_{n,j}\rangle_H, \tag{A7.4}$$

assuming that $\{\phi_j,\ j \geq 1\}$ are unknown, with $\widetilde{X}_{i,n,j} = \langle X_i, \phi_{n,j}\rangle_H$, for any $j \geq 1$ and $i \in \mathbb{Z}$, being $\widetilde{\pi}^{k_n}$ the orthogonal projector into $\widetilde{H}_{k_n} = sp\left(\phi_{n,1}, \dots, \phi_{n,k_n}\right)$, for a suitable truncation parameter k_n, such that

$$\lim_{n\to\infty} k_n = \infty, \quad \frac{k_n}{n} < 1.$$

In the estimation approach formulated in equation (A7.4), the non–diagonal autocorrelation operator and covariance operator of the error term are defined as follows:

$$
\begin{aligned}
\rho\left(X\right)(t) &= \int_a^b \psi\left(t,s\right)X(s)ds,\ \psi\left(t,s\right)=\sum_{j=1}^{\infty}\sum_{h=1}^{\infty}\rho_{j,h}\phi_j(t)\phi_h(s), && \text{(A7.5)}\\
C_{\varepsilon} &= \sum_{j=1}^{\infty}\sum_{h=1}^{\infty}\sigma_{j,h}^2\phi_j\otimes\phi_h. && \text{(A7.6)}
\end{aligned}
$$

Besides the componentwise estimator of ρ^*, Mas [2000] proposed to approximate C by a linear operator smoothed by a family of functions

$$
\left\{b_{n,p}(x)=\frac{x^p}{(x+b_n)^{p+1}},\ p\geq 0,\ n\in\mathbb{N}\right\},
$$

which converge to $1/x$ point–wise, being $\{b_n,\ n\in\mathbb{N}\}$ a strictly positive sequence decreasing to zero. The formulated estimators (so–called resolvent class estimators) of ρ^* were given by Mas [2000] in the following way:

$$
\widehat{\rho}_{n,p}^{*}=b_{n,p}\left(C_n\right)D_n^{*},\quad b_{n,p}\left(C_n\right)=\left(C_n+b_nI_H\right)^{-(p+1)}C_n^p,\quad p\geq 0,\quad n\in\mathbb{N},
$$

being I_H the identity operator on H, in a manner that $b_{n,p}\left(C_n\right)$ is a compact operator, for each $p\geq 1$ and $n\in\mathbb{N}$, with deterministic norm equal to b_n^{-1}. Under the non–diagonal scenario in equations (A7.5)–(A7.6), a similar philosophy was adopted by Guillas [2001], in the derivation of the efficiency of the componentwise estimator of ρ formulated in Bosq (2000), in ways that C_n was regularized by a sequence

$$
u=\{u_n,\ n\geq 1\},\quad 0<u_n\leq\beta C_{k_n},\quad 0<\beta<1.
$$

Hence, let us defined

$$
C_{n,j,u}^{-1}=\frac{1}{\max\left(C_{n,j},u_n\right)},\quad j\geq 1,\ n\geq 2.
$$

An efficient estimator, when $\{\phi_j,\ j\geq 1\}$ are unknown, and under **Assumptions A1** and **A3**, was stated in Guillas [2001] by

$$
\widetilde{\rho}_{n,u}(x)=\sum_{l=1}^{k_n}\left(\sum_{j=1}^{k_n}C_{n,j,u}^{-1}\left(\frac{1}{n-1}\sum_{i=0}^{n-2}\widetilde{X}_{i,n,j}\widetilde{X}_{i+1,n,l}\right)\langle x,\phi_{n,j}\rangle_H\right)\phi_{n,l}, \tag{A7.7}
$$

for a well–suited truncation parameter, providing the mean–square convergence. Remark that in equation (A7.7), Hilbert–Schmidt condition over ρ is not needed. We may also cite Mas [2004], where the asymptotic properties, in the norm of $\mathcal{L}(H)$, of the estimator of ρ^* formulated in Mas [1999], were derived, such that the weaker condition of compactness of ρ was assumed. **Assumptions A1** and **A3**, and conditions in Mas [1999], were also required. This compactness condition, jointly with

$$
\left\|C^{-1/2}\rho\right\|_{\mathcal{L}(H)}<\infty,
$$

i.e., ρ should be, at least, as smooth as $C^{1/2}$, was also imposed in Mas [2007], where the weak–convergence of the estimator of ρ^* was addressed, under the convexity of the spectrum of C, when $k_n = o\left(\frac{n^{1/4}}{\ln(n)}\right)$. Álvarez-Liébana et al. [2017] recently established a weakly consistent diagonal componentwise estimator of ρ, in the norm of $\mathcal{S}(H)$, when C and ρ admit a diagonal spectral decomposition in terms of $\{\phi_j,\ j \geq 1\}$. The mean–square convergence of the following estimator of ρ, when eigenvectors of C are assumed to be known and ρ is a symmetric operator, was proved, for a well–suited truncation parameter k_n and $n \geq 2$:

$$\widehat{\rho}_{k_n} = \sum_{j=1}^{k_n} \widehat{\rho}_{n,j} \phi_j \otimes \phi_j, \quad \widehat{\rho}_{n,j} = \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}} = \frac{n}{n-1} \frac{\sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\sum_{i=0}^{n-1} X_{i,j}^2}, \quad \widehat{C}_{n,j} \neq 0 \, a.s.,$$

under the strictly positiveness of C and extra mild assumptions. A diagonal strongly–consistent estimator is formulated in Section A7.7, as well as a strongly–consistent plug–in predictor, when eigenvectors of C are unknown (see Sections A7.9.1–A7.9.2 in the Supplementary Material provided, concerning the case when eigenvectors of C are known).

Alternative ARH(1) estimation parametric techniques, based on a modified version of the functional principal component analysis (FPCA) framework above–referred, have been developed. A spline–smoothed–penalized FPCA, with rank constraint, was presented in Besse and Cardot [1996] (and later applied by Besse et al. [2000], on the forecasting of climatic variations). In that work, the paths were previously smoothed solving the following nonparametric optimization problem:

$$\min_{\widehat{X}_i^{q,\ell} \in H_q} \left\{ \frac{1}{np} \sum_{i=0}^{n-1} \sum_{j=1}^{p} \left(X_i(t_j) - \widehat{X}_i^{q,\ell}(t_j) \right)^2 + \ell \left\| D^2 \widehat{X}_i^{q,\ell} \right\|_{L^2([0,1])}^2 \right\}, \tag{A7.8}$$

being ℓ the penalized parameter and $\{t_j,\ j = 1, \ldots, p\}$ the set of knots. The q–dimensional subspace

$$H_q \subset \left\{ f : \ \left\| D^2 f \right\|_{L^2([0,1])}^2 < \infty \right\}$$

is spanned by

$$\{ A(\ell) v_j,\ j = 1, \ldots, q \},$$

being $A(\ell)$ the smoothing hat–matrix and $\{v_j,\ j = 1, \ldots, q\}$ the eigenvectors associated to the first q–largest eigenvalues of the matrix

$$S = \frac{1}{n} A(\ell)^{1/2} X' I_n X A(\ell)^{1/2}.$$

Estimator of ρ was then built in Besse and Cardot [1996] by $\widehat{\rho}_{q,\ell} = \widehat{D}_{q,\ell} \widehat{C}_{q,\ell}^{-1}$, with

$$\widehat{C}_{q,\ell} = \frac{1}{n} \sum_{i=0}^{n-1} \widehat{X}_i^{q,\ell} \otimes \widehat{X}_i^{q,\ell}, \quad \widehat{D}_{q,\ell} = \frac{1}{n-1} \sum_{i=0}^{n-2} \widehat{X}_i^{q,\ell} \otimes \widehat{X}_{i+1}^{q,\ell}.$$

See also Cardot [1998], where a spline–smoothed–penalized FPCA was achieved into the Sobolev space $H_2^2([0,1])$, providing a consistent componentwise truncated estimator of ρ of an ARH(p) process.

Based on the perturbation theory, Mas and Menneteau [2003b] proved how the asymptotic behaviour of a self–adjoint random operator is equivalent to that of its associated eigenvectors and eigenvalues. The results derived in Mas and Menneteau [2003a] are completed by Menneteau [2005], focusing on the law of the iterated logarithm, under the above–referred ARH(1) framework. In a more general framework, the lack of dependence of a functional linear model was tested in Kokoszka et al. [2008], under **Assumptions A1**, **A3** and the asymptotic properties of C_n derived in Bosq [2000]. As discussed in Kokoszka et al. [2008], their approach can be adapted to the ARH(1) framework, and therefore, the nullity of the autocorrelation operator can be tested. In the above–referred works, the null hypotheses of the constancy of ρ and the stationarity condition have implicitly been assumed. Horváth et al. [2014] derived testing methods on the stationarity of functional time series (against change point alternative and the so–called two alternatives integrated and deterministic trend). The asymptotic normality of the empirical principal components of a wide class of functional stochastic processes (even non–linear weakly dependent functional time series) was derived in Kokoszka and Reimherr [2013a].

A7.3 Extensions of the classical ARH(1) model

Enhancements to the classical ARH(1) model have been developed during the last decades. A great amount of them will be detailed in this Section, arranging the references in chronicle by blocks.

From the previous asymptotic results, the natural extension of ARH(1) to ARH(p) processes, with p greater than one, was presented in Bosq [2000] as

$$X_n = \sum_{k=1}^{p} \rho_k \left(X_{n-k}\right) + \varepsilon_n, \quad n \in \mathbb{Z},$$

and $\rho_k \in \mathcal{L}(H)$, for any $k = 1, \dots, p$, being ρ_p a non–null operator on H. By its Markovian properties, ARH(p) model was rewritten by Bosq [2000] as the H^p–valued ARH(1) process

$$Y_n = \rho'\left(Y_{n-1}\right) + \varepsilon'_n, \quad Y_n = \left(X_n, \dots, X_{n-p+1}\right) \in H^p, \quad \varepsilon'_n = \left(\varepsilon_n, 0, \dots, 0\right) \in H^p$$

and

$$\rho' = \begin{pmatrix} \rho_1 & \rho_2 & \dots & \rho_{p-1} & \rho_p \\ I_H & 0_H & \dots & 0_H & 0_H \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_H & 0_H & \dots & I_H & 0_H \end{pmatrix} \begin{array}{l} \\ \\ \in \mathcal{L}(H^p), \\ \longrightarrow \textbf{p-th row} \end{array}$$

where H^p denotes the cartesian product of p copies of H, being a Hilbert space endowed with $\langle \cdot, \cdot \rangle_p$. In the equation above, I_H and 0_H denote the identity and null operators on H, respectively. The crucial choice of the lag order p was discussed in Kokoszka and Reimherr [2013b], when $\rho_k \in \mathcal{S}(H)$, for any $k = 1, \dots, p$, and $\|\rho'\|_{\mathcal{L}(H^p)} < 1$. The following multistage testing procedure was proposed in the mentioned work, based on the estimation of the operators ρ_k, for each $k = 1, \ldots, p$:

$$\begin{aligned} H_0 &: X \text{ is an i.i.d. sequence} \quad vs \quad H_{p-1} : X \text{ is an ARH}(1) \text{ process}, \\ H_{p-1} &: X \text{ is an ARH(p-1) process} \quad vs \quad H_p : X \text{ is an ARH(p) process}, \end{aligned}$$

in a manner that the method continues while a null hypothesis is not be rejected.

Aimed at including exogenous information in the dependence structure, ARH(1) processes with exogenous variables (ARHX(1) processes) were introduced in Damon and Guillas [2002]; Guillas [2000], as follows:

$$X_n = \rho\left(X_{n-1}\right) + \sum_{k=1}^{p} a_k\left(Z_{n,k}\right) + \varepsilon_n, \quad n \in \mathbb{Z}, \quad a_k, \rho \in \mathcal{L}(H),\ k = 1, \ldots, p, \tag{A7.9}$$

being $Z = \{Z_{n,k},\ n \in \mathbb{Z},\ k = 1, \ldots, p\}$ the exogenous variables. Guillas [2000] initally proposed an autoregressive of order 1 inner dependence structure on Z (i.e., $a_k = 0_H$, for any $k = 2, \ldots, p$), while the ARH(p) structure displayed in (A7.9) was subsequently established in Damon and Guillas [2002, 2005].

The first derivatives of the random paths of an ARH(1) process were included by Marion and Pumo [2004] as the exogenous variables (so–called ARHD(1) process), when the trajectories belong to the Sobolev space $H_2^1\left([0,1]\right)$. The ARH(1) process was given by

$$X_n = \rho\left(X_{n-1}\right) + \Psi\left(X_{n-1}'\right) + \varepsilon_n, \quad n \in \mathbb{Z}, \quad \rho,\ \Psi \in \mathcal{K}(H)$$

being $\mathcal{K}(H)$ the set of compact operator on H, and was reformulated by Mas and Pumo [2007] as the ARH(1) process:

$$X_n = A\left(X_{n-1}\right) + \varepsilon_n, \quad A = \Phi + \Psi D \in \mathcal{K}(H),\ \|A\|_{\mathcal{L}(H)} < 1, \quad D(f) = f',$$

with

$$\langle f, g\rangle_W = \int_0^1 f(t)g(t)dt + \int_0^1 f'(t)g'(t)dt, \quad f, g \in W^{2,1}\left([0,1]\right).$$

After pointing out some extensions, where exogenous information has additively been incorporated, Guillas [2002] proposed an i.i.d. sequence of Bernoulli variables

$$I = \{I_n,\ n \in \mathbb{Z}\}$$

to condition an ARH(1) process, in a non–additive way. A conditional autoregressive Hilbertian process of order one (CARH(1) process, also known as doubly stochastic Hilbertian process of order one) was then formulated, for any $n \in \mathbb{Z}$:

$$X_n = \rho_{I_n}\left(X_{n-1}\right) + \varepsilon_n = \begin{cases} \rho_0\left(X_{n-1}\right) + \varepsilon_n, & \text{if } I_n = 0 \\ \rho_1\left(X_{n-1}\right) + \varepsilon_n, & \text{otherwise} \end{cases}, \quad \rho_0,\ \rho_1 \in \mathcal{L}(H). \tag{A7.10}$$

An extension of (A7.10), where the latent process was considered as a continuous multivariate process $V = \{V_n,\ n \in \mathbb{Z}\}$, was established in Cugliari [2013]. Mourid [2004] proposed to consider the randomness of ρ by defining it from a basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$ into $\mathcal{L}(H)$; i.e., $\rho^{\omega} \in \mathcal{L}(H)$, for each $\omega \in \Omega$. ARH(p) processes with random coefficients (RARH(p) processes) were then introduced (see also Allam and Mourid [2014]).

A new branch in the field of functional time series, when the data is gathered on a grid assuming a spatial interaction, was firstly introduced by Ruiz-Medina [2011]. In that work, a novel family of spatial stochastic processes (SARH(1) processes), which can be seen as the Hilbert–valued extension of spatial autoregressive processes of order one (SAR(1) processes), was defined as follows:

$$X_{i,j} = R + \rho_1 (X_{i-1,j}) + \rho_2 (X_{i,j-1}) + \rho_3 (X_{i-1,j-1}) + \varepsilon_{i,j},\ (i,j) \in \mathbb{Z}^2, \quad R \in H, \tag{A7.11}$$

being $\rho_h \in \mathcal{L}(H),\ h = 1,2,3$, and based on the so–called Markov property of the three points for a spatial stochastic process. In (A7.11), ρ_h is assumed to be decomposed in terms of the eigenvalues $\{\lambda_{k,h},\ k \geq 1\}$ and the biorthonormal systems of left and right eigenvectors, $\{\psi_k,\ k \geq 1\}$ and $\{\phi_k,\ k \geq 1\}$, respectively, for each $h = 1,2,3$. The spatial innovation process $\{\varepsilon_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ is imposed to be a two–parameter martingale difference sequence, with $\mathrm{E}\{\varepsilon_{i,j} \otimes \varepsilon_{i,j}\}$ not depending on the coordinates $(i,j) \in \mathbb{Z}^2$. Ruiz-Medina [2011] derived an unique stationary solution to the SARH(1) state equation (A7.11), providing its inversion. Extended classes of models of functional spatial time series are also formulated in that paper. Moment–based estimators of the functional parameters involved in the SARH(1) equation were proposed in Ruiz-Medina [2012], where their performance is illustrated with a real data application, for spatial functional prediction of ocean surface temperature.

A new set of sufficient conditions was provided in Ruiz-Medina and Álvarez-Liébana [2017a] for the asymptotic efficiency of diagonal componentwise estimators of the autocorrelation operator of a stationary ARH(1) process, under both classical and Beta–prior–based Bayesian scenarios. In particular, under **Assumption A1**, ρ is assumed to be linear bounded and self–adjoint operator, while the usual Hilbert–Schmidt condition is not imposed. Stronger assumptions for the eigenvalues $\{\sigma_j^2,\ j \geq 1\}$ of $C_{\varepsilon} = \mathrm{E}\{\varepsilon_n \otimes \varepsilon_n\}$ were considered, to offset the slower decay rate of the eigenvalues $\{\rho_j,\ j \geq 1\}$ of ρ. Specifically, if $\rho = \sum_{j=1}^{\infty} \rho_j \phi_j \otimes \phi_j$, conditions

$$\rho_j = \sqrt{1 - \frac{\sigma_j^2}{C_j}}, \quad \frac{\sigma_j^2}{C_j} \leq 1, \quad \frac{\sigma_j^2}{C_j} = \mathcal{O}\left(j^{-1-\gamma}\right), \quad \gamma > 0, \quad \sigma_j^2 = \mathrm{E}\left\{\langle \varepsilon_n, \phi_j \rangle_H^2\right\},\ j \geq 1,$$

were assumed. The asymptotic equivalence of the estimators was also provided, as well as of the their associated plug–in predictors. The Beta–prior–based Bayesian estimator of $\rho = \sum_{j=1}^{\infty} \widetilde{\rho}_{n,j} \phi_j \otimes \phi_j$ was then derived in Ruiz-Medina and Álvarez-Liébana [2017a] as follows:

$$\widetilde{\rho}_{n,j} = \frac{1}{2\beta_{n,j}} \left((\alpha_{n,j} + \beta_{n,j}) - \sqrt{(\alpha_{n,j} - \beta_{n,j})^2 - 4\beta_{n,j}\sigma_j^2 \left(2 - (a_j + b_j)\right)} \right),$$

with

$$\alpha_{n,j} = \sum_{i=0}^{n-1} X_{i,j} X_{i+1,j}, \quad \beta_{n,j} = \sum_{i=0}^{n-1} X_{i,j}^2, \quad j \geq 1, \quad n \in \mathbb{Z},$$

being (a_j, b_j) the Beta parameters such that $\rho_j \sim \mathcal{B}(a_j, b_j)$, for any $j \geq 1$. We may also cite Ruiz-Medina and Álvarez-Liébana [2018a], where sufficient conditions for the strong–consistency, in the trace norm, of the above–formulated diagonal componentwise estimator of the autocorrelation operator of an ARH(1) process, are provided. Note that, in that paper, ρ is not assumed to admit a diagonal spectral decomposition with respect to the eigenvectors of the autocovariance operator C. See also Kowal et al. [2017], where a two–level hierarchical model has recently been proposed on the prediction of an ARH(p) process, applied to the forecasting of the U.S. Treasury nominal yield curve.

A7.4 ARH estimation approaches based on alternative bases

In this section, we pay attention to the ARH(1) estimation, based on the projection into alternative bases to the eigenvectors of C. The sieves method was adapted by Bensmain and Mourid [2001] for the estimation of the autocorrelation operator of an ARH(1) process. A novel consistent estimator was derived under both scenarios, when ρ is a bounded linear operator, and under the Hilbert–Schmidt condition. Specifically, ρ was estimated considering different subsets (so–called sieves) $\{\Theta_m,\ m \in \mathbb{N}\}$ of the parametric space Θ, where ρ takes its values, equipped with a metric d, such that Θ_m is a compact set, with $\Theta_m \subset \Theta_{m+1}$ and $\bigcup_{m\in\mathbb{N}} \Theta_m$ is dense in Θ.

In particular, in the former case, when ρ is assumed to be a bounded linear operator

$$\rho(f)(t) = \int_0^1 K(t-x) f(x) dx,$$

depending on a kernel $K(\cdot)$, then

$$X_n(t) = \int_0^1 K(t-s) X_{n-1}(s) ds + \varepsilon_n(t).$$

The Fourier basis

$$\left\{\phi_{2k}(t) = \sqrt{2}\cos(2\pi k t),\ \phi_{2k+1}(t) = \sqrt{2}\sin(2\pi k t),\ k \geq 1\right\}$$

and $\phi_0(t) = I_{[0,1]}$ was considered, being $I_{[0,1]}$ the identity function over the interval $[0,1]$. The ARH(1) state equation was then developed as

$$\begin{cases} a_0(X_n) = a_0(K) a_0(X_{n-1}) + a_0(\varepsilon_n), \\ a_k(X_n) = (a_k(K) a_k(X_{n-1}) - b_k(K) b_k(X_{n-1}))/2 + a_k(\varepsilon_n) \\ b_k(X_n) = (a_k(K) b_k(X_{n-1}) + b_k(K) a_k(X_{n-1}))/2 + b_k(\varepsilon_n) \end{cases}$$

for each $n \in \mathbb{Z}$ and $k \geq 1$, being

$$\{a_k\left(X_n\right),\, a_k\left(\varepsilon_n\right),\, a_k\left(K\right),\, k \geq 1\}, \quad \{b_k\left(X_n\right),\, b_k\left(\varepsilon_n\right),\, b_k\left(K\right),\, k \geq 1\}$$

the Fourier coefficients respect to cosine and sine functions, respectively. Bensmain and Mourid [2001] assumed that $b_k(t) = 0$, for each $t \in [0, 1]$ and $k \geq 0$, in a manner that estimation of ρ was then reached by forecasting the Fourier coefficients $\{c_k = a_k\left(K\right),\, k \geq 0\}$ in the sieve

$$\Theta_{m_n} = \left\{ K(t) = c_0 I_{[0,1]} + \sum_{k=1}^{m_n} c_k \sqrt{2} \cos\left(2\pi k t\right), \quad \sum_{k=1}^{m_n \to \infty} k^2 c_k^2 \leq m_n \right\}.$$

The non–diagonal componentwise estimator formulated in Bosq [2000] was used in Laukaitis and Rackauskas [2002], by considering regularized paths in terms of a B–spline basis. In that work, the forecasting of the intensity of both cash withdrawal in cash machines (so–called automatic teller machines or ATM) was achieved. Antoniadis and Sapatinas [2003] discussed how the prediction of functional stochastic processes can be seen as a linear ill–posed inverse problem, providing a few approaches about the regularization techniques required. In the context of 1–year–ahead forecasting of the climatological ENSO time series, they also proposed three linear wavelet–basis–based ARH(1) predictors, one of which is based on the resolvent estimators of ρ formulated in Mas [2000]. From the componentwise estimation framework developed in Bosq [2000], they derived regularized wavelet estimators, by means of a previously wavelet–basis–based smoothing method:

$$\widetilde{Y}_{i,\widehat{\lambda}^M} = \widetilde{X}_{i,\widehat{\lambda}^M} - \frac{1}{n}\sum_{i=0}^{n-1} \widetilde{X}_{i,\widehat{\lambda}^M}, \quad \widetilde{X}_{i,\widehat{\lambda}^M} = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha_{j_0 k}^{i}} \phi_{j_0 k} + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \widehat{\beta_{jk}^{i}} \psi_{jk}, \tag{A7.12}$$

for any $i \in \mathbb{Z}$, with smoothing parameter $\widehat{\lambda}^M = \left(\sum_{j=1}^{M} \sigma_j^2\right)\left(\sum_{j=1}^{M} C_j\right)/N$. The plug–in predictor was given by

$$\widetilde{\rho}_{n,\widehat{\lambda}^M}\left(X_{n-1}\right) = \sum_{j=1}^{k_n} \left(\frac{1}{n-1} \sum_{k=1}^{k_n}\sum_{i=0}^{n-2} \frac{1}{\widetilde{C}_{n,\widehat{\lambda}^M,k}} \widetilde{X}_{n-1,\widehat{\lambda}^M,k} \widetilde{Y}_{i,\widehat{\lambda}^M,k} \widetilde{Y}_{i+1,\widehat{\lambda}^M,j} \right) \widetilde{\phi}_j^M, \tag{A7.13}$$

with

$$\widetilde{X}_{n-1,\widehat{\lambda}^M,j} = \langle \widetilde{\phi}_j^M, X_{n-1} \rangle_H$$

and

$$\widetilde{Y}_{i+1,\widehat{\lambda}^M,j} = \langle \widetilde{\phi}_j^M, \widetilde{Y}_{i+1,\widehat{\lambda}^M} \rangle_H,$$

for each $j = 1, \ldots, k_n$ and $i = 0, \ldots, n-1$, where

$$\left\{ \widetilde{C}_{n,\widehat{\lambda}^M,j},\, j \geq 1 \right\}$$

and

$$\left\{\widetilde{\phi}_j^M,\ j \geq 1\right\}$$

denote the eigenvalues and eigenvectors, respectively, of the empirical estimator

$$\widetilde{C}_{n,\widehat{\lambda}^M} = \frac{1}{n}\sum_{i=0}^{n-1} \widetilde{Y}_{i,\widehat{\lambda}^M} \otimes \widetilde{Y}_{i,\widehat{\lambda}^M}.$$

Values

$$\left\{\widehat{\alpha_{j_0k}^i},\ \phi_{j_0k},\ k = 0, \ldots, 2^{j_0} - 1\right\}$$

and

$$\left\{\widehat{\beta_{jk}^i},\ \psi_{jk},\ j \geq j_0,\ k = 0, \ldots, 2^j - 1\right\},$$

for $i = 0, \ldots, n-1$, in equation (A7.12), denote the scaling coefficients, at $J - j_0$ resolutions levels, for a primary resolution level $j_0 < J$. **Assumptions A1** and **A3** were imposed, along with

$$nC_{k_n}^4 \to \infty, \quad \frac{1}{n}\sum_{j=1}^{k_n} \frac{b_j}{C_j^2} \to 0, \quad b_j = \max\left((C_{j-1} - C_j)^{-1}, (C_j - C_{j+1})^{-1}\right). \tag{A7.14}$$

Hyndman and Ullah [2007] detailed an alternative robust version of FPCA, avoiding the instability induced by outlying observations. Forecasting of mortality and fertility rates was there performed. A weighted version of the approach presented in the mentioned work by Hyndman and Ullah [2007], considering the largest weights for the most recent data (required in fields such as demography), was developed in Hyndman and Shang [2009]. Instead of the curve–by–curve forecasting established in Hyndman and Shang [2009]; Hyndman and Ullah [2007], a multivariate VARMA model was applied by Aue et al. [2015], to avoid the loss of information invoked by the uncorrelated assumption of FPC scores.

In addition, Kargin and Onatski [2008] focused on the ARH(1) predictor, instead of on the operators ρ and C themselves. They proposed to replace the FPC with directions more relevant to forecasting, by searching a reduced–rank approximation (see also Didericksen and Kokoszka [2012], where a comparative study, between approaches in Bosq [2000] and Kargin and Onatski [2008], was carried out). Their method, so–called predictive factor decomposition, is built by searching of a minimal operator $\rho \in R_p$, aimed to minimize

$$\mathrm{E}\left\{\|X_n - \rho(X_{n-1})\|_H^2\right\},$$

being R_p the set of p–rank operator. The predictor was then given by

$$\widehat{X}_n = \sum_{l=1}^{p} \langle X_{n-1}, \widehat{b}_l^\alpha\rangle_H D_n \widehat{b}_l^\alpha, \quad \widehat{b}_l^\alpha = \alpha\widehat{x}_l^\alpha + \sum_{j=1}^{K} \frac{\langle \widehat{x}_l^\alpha, \phi_{n,j}\rangle_H}{C_{n,j}^{1/2}}\phi_{n,j}, \quad l = 1, \ldots, p,$$

being $\{\widehat{x}_l^\alpha,\ l = 1, \ldots, p\}$ a linear combination of the eigenvectors $\{\phi_{n,j},\ j = 1, \ldots, p\}$ of the empirical autocovariance operator.

A7.5 Hilbert–valued moving–average and general linear processes

This section is devoted to describe the main contributions in the field of Hilbertian moving–average processes (MAH processes), including the general case of Hilbertian general linear processes (LPH). The case of ARMAH processes is considered as well. From the Wold decomposition of a LPH

$$X_n = \varepsilon_n + \sum_{k=1}^{\infty} a_k \left(\varepsilon_{n-k}\right), \quad n \in \mathbb{Z}, \quad a_k \in \mathcal{L}(H),\ k \geq 1,$$

the stationarity is held as long as $\varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}$ is a H–valued SWN and

$$\sum_{k=1}^{\infty} \|a_k\|^2_{\mathcal{L}(H)} < \infty.$$

Building on the early works by Bosq [1991], the invertibility of LPH was proved in Merlevède [1995] if and only if

$$1 - \sum_{j=1}^{\infty} z^j \left\|a_j\right\|_{\mathcal{L}(H)} \neq 0, \quad |z| < 1.$$

Merlevède [1997] provided a Markovian representation of stationary and invertible LPH in a subspace

$$H_\beta = \left\{ X :\ \|X\|_{H_\beta} = \sum_{k=1}^{\infty} \beta_k \left\|X_k\right\|^2_H < \infty \right\},\ \beta = \left\{\beta_k > 0,\ k \geq 1\right\},\ \sum_{k=1}^{\infty} \beta_k < \infty,$$

being β a decreasing sequence. Let us define the H_β–random variables $Y_n = (X_n, X_{n-1}, \ldots, X_{n-p}, \ldots)'$ and $e_n = (\varepsilon_n, 0, 0, \ldots)'$, for each $n \in \mathbb{Z}$. A strongly–consistent plug–in predictor was derived in Merlevède [1997], by estimating

$$R = \begin{pmatrix} \rho_1 & \rho_2 & \cdots & \rho_p & \cdots \\ I_H & 0_H & \ldots & 0_H & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ 0_H & 0_H & \ldots & I_H & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix} \begin{matrix} \\ \\ \\ \longrightarrow \textbf{p–th row} \\ \\ \end{matrix}$$

being $R \in \mathcal{L}\left(H_\beta\right)$, under

$$\|R\|_{\mathcal{L}(H_\beta)} < 1, \quad \mathrm{E}\left\{\|Y_0\|^4_{H_\beta}\right\} < \infty.$$

Mas [2002] studied weak–convergence for the empirical autocovariance and cross–covariance operators of LPH.

MAH(q) and ARMAH(p,q) processes, with p and q greater than one, as a particular case of LPH, were defined in Bosq and Blanke [2007] as

$$X_n = \varepsilon_n + \sum_{k=1}^{q} l_k\left(\varepsilon_{n-k}\right), \quad l_k \in \mathcal{L}(H), \quad \|l_k\|_{\mathcal{L}(H)} < 1,$$

and

$$X_n = \varepsilon_n + \sum_{j=1}^{p} \rho_j\left(X_{n-j}\right) + \sum_{k=1}^{q} l_k\left(\varepsilon_{n-k}\right), \quad l_k,\ \rho_j \in \mathcal{L}(H),$$

respectively, for each $n \in \mathbb{Z}$ and $k = 1, \ldots, q,\ j = 1, \ldots, p$, with $\|l_k\|_{\mathcal{L}(H)} < 1,\ \|\rho_j\|_{\mathcal{L}(H)} < 1$. LPH in a wide sense, when $\{a_j,\ j \geq 1\}$ are allowed to be unbounded, were studied in Bosq [2007]; Bosq and Blanke [2007]. Unlike the estimation of an ARH(1) process, troubles in the estimation of the operator l of a MAH(1) process arise from the non–linear behaviour of the moment equation. We may cite Turbillon et al. [2008], where the estimation of the MAH(1) model

$$X_n = \varepsilon_n + l\left(\varepsilon_{n-1}\right), \quad l \in \mathcal{K}(H)$$

under

$$\left\|DC^{-1}\right\|_{\mathcal{L}(H)} < 1/2, \quad \left\|D^{*}C^{-1}\right\|_{\mathcal{L}(H)} < 1/2,$$

was reached. A special framework was introduced in Wang [2008], where a real–valued non–linear ARIMA(p,d,q) model was modified, in a manner that functional MA coefficients were included:

$$X_n + \sum_{j=1}^{p} \rho_j X_{n-j} = \varepsilon_n + \sum_{k=1}^{q} f_k\left(X_{n-k-d}\right)\varepsilon_{n-k}, \quad n \in \mathbb{Z}, \tag{A7.15}$$

being $\{f_k,\ k \geq 1\}$ a set of arbitrary univariate functions. Forecasting of the Chinese Consumer Price Index, which monthly collects prices paid by middle–class consumers for a standard basket of goods and services, was achieved in Chen et al. [2016], adopting smooth functions as functional MA coefficients in equation (A7.15). Furthermore, a survey about the asymptotic properties of LPH, derived in the above–referred works by Merlevède [1995, 1996a, 1997], was achieved in Bosq [2000]; Bosq and Blanke [2007]. Useful tools proposed by Hyndman and Shang [2008], such as visualization and outlier detection, can be applied to observed ARMAH processes, obeying a functional linear model. Outlier detection in French male age–specific mortality data was also achieved in that work.

A7.6 NONPARAMETRIC FUNCTIONAL TIME SERIES FRAMEWORK

Lastly, let us see the main references in the context of nonparametric functional time series and functional linear regression, when both explanatory and response variables, take values in a space of functions.

As a functional extension of the work by Poggi [1994], a nonparametric kernel–based predictor was formulated in Besse et al. [2000]

$$\widehat{X}_n^{h_n} = \frac{\sum_{i=0}^{n-2} \widehat{X}_{i+1} K\left(\frac{\left\|\widehat{X}_i - X_{n-1}\right\|^2_{L^2([0,\delta])}}{h_n}\right)}{\sum_{i=0}^{n-2} K\left(\frac{\left\|\widehat{X}_i - X_{n-1}\right\|^2_{L^2([0,\delta])}}{h_n}\right)}, \tag{A7.16}$$

being

$$\widehat{X}_i = argmin \left\|D\widehat{X}_i\right\|^2_{L^2([0,\delta_1])}, \quad i \in \mathbb{N},$$

K the usual Gaussian kernel and D a d–th order differential operator. Cuevas et al. [2002] addressed the strong–consistency estimation of the underlying linear operator of a linear regression, when both explanatory and response variables are assumed to be H–valued random variables, with $H = L^2([0,\delta_1])$. In particular, the design is given by the triangular array

$$\{X_{i,n}(t), \ 1 \leq i \leq n\},$$

providing the model

$$Y_{i,n} = \Psi(X_{i,n}) + \varepsilon_{i,n}, \quad X_{i,n} \in L^2([0,\delta_1]), \quad Y_{i,n} \in L^2([0,\delta_2]),$$

under $\Psi \in \mathcal{L}(L^2([0,\delta_1]), L^2([0,\delta_2]))$.

Antoniadis et al. [2006] introduced (see also Antoniadis et al. [2012]), the two–steps prediction approach so–called kernel wavelet functional (KWF) method, where strongly–mixing conditions are imposed. An expansion of stationary functional time series into a discrete wavelet basis $\left\{\psi_k^J, \ k = 0, \dots, 2^J - 1\right\}$, at scale J, is achieved, and the forecasting of $\widehat{X}_n = \mathrm{E}\{X_n | X_{n-1}, \dots, X_0\}$, for each $n \in \mathbb{Z}$, was then performed by

$$\widehat{X}_n^J(\cdot) = \sum_{k=0}^{2^J-1} \widehat{\xi}_{n,k}^J \psi_k^J(\cdot), \quad \widehat{\Xi}_n = \frac{\sum_{i=0}^{n-2} K\left(D\left(P(\Xi_n), D\left(P(\Xi_i)\right)\right)/h_n\right)\Xi_{i+1}}{1/n + \sum_{i=0}^{n-2} K\left(D\left(P(\Xi_n), D\left(P(\Xi_i)\right)\right)/h_n\right)}, \tag{A7.17}$$

where

$$\widehat{\Xi}_n = \left\{\widehat{\xi}_{n,k}^J : \ k = 0, 1, \dots, 2^J - 1\right\}$$

denotes, for each $n \in \mathbb{Z}$, the set of predicted scaling coefficients, at scale J, being $P(\Xi_i)$ the set of wavelet coefficients derived by the so–called pyramid algorithm (more details can be found in Mallat [1989]), for any $i = 0, 1, \ldots, n-1$. Distance $D(\cdot, \cdot)$ in (A7.17), for a two set of discrete wavelet coefficients $\{\theta^i_{j,k},\ i = 1, 2\}$, at scale $j = j_0, \ldots, J-1$ and location $k = 0, \ldots, 2^j - 1$, is given by

$$D\left(\theta^1, \theta^2\right) = \sum_{j=j_0}^{J-1} 2^{-j/2} d_j\left(\theta^1, \theta^2\right), \quad d_j\left(\theta^1, \theta^2\right) = \left(\sum_{k=0}^{2^j-1} \left(\theta^1_{j,k} - \theta^2_{j,k}\right)^2\right)^{1/2}.$$

Functional versions of partial least–squares regression and principal component regression (denoted as FPLSR and FPCR, respectively) were formulated in Reiss and Ogden [2007]. In this work, a functional smoothing–based approach to signal regression was adopted, where decompositions in terms of B–spline bases and roughness penalties are involved. Let us now consider a general functional linear regression model, when Hilbert–valued response and $\mathcal{F}$–valued explanatory variables are considered, when $\mathcal{F}$ is defined as a general function space, equipped with a semi–metric d and its associated topology $\mathcal{T}_{\mathcal{F}}(X, t) = \{X_1 \in \mathcal{F} :\ d(X_1, X) \leq t\}$. In this framework, a nonparametric kernel–based estimator of the underlying regression operator was derived in Ferraty et al. [2012] as follows, for each $i = 0, \ldots, n-1$:

$$Y_i = \Psi(X_i) + \varepsilon, \quad \widehat{Y_n} = \widehat{\Psi}_{h_n}(X_n), \quad \widehat{\Psi}_{h_n}(X_n) = \frac{\displaystyle\sum_{i=0}^{n-2} X_{i+1} K\left(\frac{d(X_i, X_{n-1})}{h_n}\right)}{\displaystyle\sum_{i=0}^{n-2} K\left(\frac{d(X_i, X_{n-1})}{h_n}\right)},$$

being K a Gaussian kernel (see also Ferraty and Vieu [2006], concerning the choice of a suitable semi–metric d).

A7.7 ARH(1) strongly–consistent diagonal componentwise estimator

We here derive the conditions required on the strong–consistency of a componentwise estimator of ρ, when it admits a diagonal spectral decomposition in terms of the common eigenvectors system $\{\phi_j,\ j \geq 1\}$. In that case, an important dimension reduction is achieved. This spectral diagonalization can be reached under a wide range of scenarios, leading to a sparse representation of kernels of the associated integral operators (see more details in Ruiz-Medina and Álvarez-Liébana [2018a]. We assume that $\{\phi_j,\ j \geq 1\}$ are unknown (see Sections A7.9.1–A7.9.2 in the Supplementary Material provided, when $\{\phi_j,\ j \geq 1\}$ are known).

A7.7.1 ARH(1) model: diagonal framework

As before, let $X = \{X_n,\ n \in \mathbb{Z}\}$ be a zero–mean stationary ARH(1) process on the basic probability space $(\Omega, \mathcal{A}, P)$, satisfying:

$$X_n(t) = \rho(X_{n-1})(t) + \varepsilon_n(t), \quad \rho \in \mathcal{L}(H), \quad \|\rho\|_{\mathcal{L}(H)} < 1, \quad n \in \mathbb{Z}, \tag{A7.18}$$

when the H–valued innovation process $\varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}$ is assumed to be strong white noise, and to be uncorrelated with X_0, with $\sigma^2_\varepsilon = \mathrm{E}\{\|\varepsilon_n\|^2_H\} < \infty$, for all $n \in \mathbb{Z}$. In addition, let us consider the following assumptions:

Assumption A1. The autocovariance operator $C = \mathrm{E}\{X_n \otimes X_n\}$, for every $n \in \mathbb{Z}$, is a strictly positive and self–adjoint operator, in the trace class. Its eigenvalues $\{C_j,\ j \geq 1\}$ then satisfy

$$C_1 > C_2 > \ldots > C_j > \ldots > 0$$

and

$$\sum_{j=1}^{\infty} C_j < \infty, \quad C(f)(g) = \sum_{j=1}^{\infty} C_j \langle \phi_j, f \rangle_H \langle \phi_j, g \rangle_H, \ \forall f, g \in H.$$

Assumption A2. The autocorrelation operator is a self–adjoint and Hilbert–Schmidt operator, admitting the following diagonal spectral decomposition:

$$\rho(f)(g) = \sum_{j=1}^{\infty} \rho_j \langle \phi_j, f \rangle_H \langle \phi_j, g \rangle_H, \quad \sum_{j=1}^{\infty} \rho_j^2 < \infty, \quad \forall f, g \in H,$$

where the set $\{\rho_j,\ j \geq 1\}$ denotes the eigenvalues of ρ, with respect to $\{\phi_j,\ j \geq 1\}$.

Under **Assumptions A1–A2**, the cross–covariance operator can be also diagonally decomposed, with regard to the eigenvectors of C, providing a set of eigenvalues $\{D_j = \rho_j C_j,\ j \geq 1\}$. Projections of (A7.18) into $\{\phi_j,\ j \geq 1\}$ lead to the stationary zero–mean AR(1) representation, under $\|\rho\|_{\mathcal{L}(H)} = \sup_{j \geq 1} |\rho_j| < 1$:

$$X_{n,j} = \rho_j X_{n-1,j} + \varepsilon_{n,j}, \quad X_{n,j} = \langle X_n, \phi_j \rangle_H, \ \varepsilon_{n,j} = \langle \varepsilon_n, \phi_j \rangle_H, \ j \geq 1, \ n \in \mathbb{Z}.$$

A7.7.2 Diagonal strongly–consistent estimator: eigenvectors of C are unknown

From model proposed in (A7.18), we can formally defined the autocorrelation operator as $\rho(x) = DC^{-1}(x)$, for any $x \in H$. Nevertheless, it is well known that the operator C cannot be inverted in the whole domain. That is, an empirical estimator of C must be computed. In the case of $\{\phi_j,\ j \geq 1\}$ are unknown, we can define an empirical estimator C_n, admitting a diagonal spectral decomposition in terms of $\{C_{n,j},\ j \geq 1\}$ and $\{\phi_{n,j},\ j \geq 1\}$, satisfying, for each $n \geq 2$:

$$C_{n,1} \geq \cdots \geq C_{n,n} > 0 = C_{n,n+1} = C_{n,n+2} = \ldots, \tag{A7.19}$$

$$C_n = \frac{1}{n} \sum_{i=0}^{n-1} X_i \otimes X_i = \sum_{j=1}^{\infty} C_{n,j} \phi_{n,j} \otimes \phi_{n,j}, \tag{A7.20}$$

$$C_{n,j} = \frac{1}{n} \sum_{i=0}^{n-1} \widetilde{X}_{i,n,j}^2, \ j \geq 1. \tag{A7.21}$$

In the following,

$$\widetilde{X}_{i,n,j} = \langle X_i, \phi_{n,j} \rangle_H, \quad \phi'_{n,j} = \mathrm{sgn} \langle \phi_{n,j}, \phi_j \rangle_H \phi_j,$$

for each $i \in \mathbb{Z}$, $j \geq 1$ and $n \geq 2$, where $\operatorname{sgn}\langle \phi_{n,j}, \phi_j \rangle_H = \mathbf{1}_{\langle \phi_{n,j}, \phi_j \rangle_H \geq 0} - \mathbf{1}_{\langle \phi_{n,j}, \phi_j \rangle_H < 0}$. Since $\{\phi_{n,j},\ j \geq 1\}$ is a complete orthonormal, for each $n \geq 2$, an empirical estimator D_n can be also formulated, leading to the following non–diagonal representation:

$$D_n = \sum_{j,l=1}^{\infty} D^*_{n,j,l} \phi_{n,j} \otimes \phi_{n,l}, \quad D^*_{n,j,l} = \langle D_n(\phi_{n,j}), \phi_{n,l} \rangle_H = \frac{1}{n-1} \sum_{i=0}^{n-2} \widetilde{X}_{i,n,j} \widetilde{X}_{i+1,n,l},$$

for each $j, l \geq 1$ and $n \geq 2$. Henceforth, we denote as $D_{n,j} = \langle D_n(\phi_{n,j}), \phi_{n,j} \rangle_H$. The following assumption is here deemed, jointly with **Assumption A3**:

Assumption A4. $C_{n,k_n} > 0\ a.s$, where k_n is a suitable truncation parameter $k_n < n$, with $\lim_{n \to \infty} k_n = \infty$.

From **Assumption A4**, a diagonal componentwise estimator is defined as

$$\widetilde{\rho}_{k_n} = \sum_{j=1}^{k_n} \widetilde{\rho}_{n,j} \phi_{n,j} \otimes \phi_{n,j}, \quad \widetilde{\rho}_{n,j} = \frac{D_{n,j}}{C_{n,j}} = \frac{n}{n-1} \frac{\sum_{i=0}^{n-2} \widetilde{X}_{i,n,j} \widetilde{X}_{i+1,n,j}}{\sum_{i=0}^{n-1} \widetilde{X}^2_{i,n,j}}, \ j \geq 1,\ n \geq 2. \tag{A7.22}$$

Under **Assumptions A1–A4**, the strong–consistency of the diagonal estimator $\widetilde{\rho}_{k_n}$ of ρ is proved in Proposition A7.7.1 below. The large–sample behaviour of (A7.22) is numerically illustrated in Section A7.9.4 of the Supplementary Material.

Proposition A7.7.1 *Let k_n be a truncation parameter, given under conditions mentioned in **Assumption A4**, such that, for any $\beta > \frac{1}{2}$,*

$$\Lambda_{k_n} = o\left(n^{1/4} (\ln(n))^{\beta - 1/2}\right), \quad \frac{1}{C_{k_n}} \sum_{j=1}^{k_n} a_j = \mathcal{O}\left(n^{1/4} (\ln(n))^{-\beta}\right), \tag{A7.23}$$

where

$$\Lambda_{k_n} = \sup_{1 \leq j \leq k_n} (C_j - C_{j+1})^{-1},$$

and

$$a_1 = 2\sqrt{2} \frac{1}{C_1 - C_2}, \quad a_j = 2\sqrt{2} \max\left(\frac{1}{C_{j-1} - C_j}, \frac{1}{C_j - C_{j+1}} \right), \ 2 \leq j \leq k_n.$$

*Then, under **Assumptions A1–A4**,*

$$\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} \longrightarrow^{a.s.} 0, \quad \|(\widetilde{\rho}_{k_n} - \rho)(X_{n-1})\|_H \longrightarrow^{a.s.} 0, \quad n \to \infty.$$

In particular, the following upper bound can be derived:

$$\begin{aligned}\|\widetilde{\rho}_{k_n}-\rho\|_{\mathcal{L}(H)} &\leq \sup_{1\leq j\leq k_n}\left|\widetilde{\rho}_{n,j}-\frac{D_{n,j}}{C_j}\right|+\sup_{1\leq j\leq k_n}\left|\frac{D_{n,j}}{C_j}-\rho_j\right| \\ &+ 2\sum_{j=1}^{k_n}\frac{|D_{n,j}|}{C_j}\left\|\phi_{n,j}-\phi'_{n,j}\right\|_H+\sup_{j>k_n}|\rho_j|\,. \qquad \text{(A7.24)}\end{aligned}$$

Proof. Under **Assumptions A1–A2** and equation (A7.22), for any $x\in H$,

$$\begin{aligned}\|\widetilde{\rho}_{k_n}(x)-\rho(x)\|_H &\leq \left\|\sum_{j=1}^{k_n}\widetilde{\rho}_{n,j}\langle\phi_{n,j},x\rangle_H\phi_{n,j}-\sum_{j=1}^{k_n}\rho_j\langle\phi_j,x\rangle_H\phi_j\right\|_H \\ &+ \left\|\sum_{j=1}^{k_n}\rho_j\langle\phi_j,x\rangle_H\phi_j-\sum_{j=1}^{\infty}\rho_j\langle\phi_j,x\rangle_H\phi_j\right\|_H \\ &= a_{k_n}(x)+b_{k_n}(x). \qquad \text{(A7.25)}\end{aligned}$$

Clearly, under **Assumption A2**, $\lim_{n\to\infty} b_{k_n}(x)=0$. Let us now study the behaviour of term $a_{k_n}(x)$. From equations (A7.19)–(A7.22), under **Assumption A4**,

$$\begin{aligned}a_{k_n}(x) &\leq \left\|\sum_{j=1}^{k_n}\left(\frac{D_{n,j}}{C_{n,j}}-\frac{D_{n,j}}{C_j}\right)\langle\phi_{n,j},x\rangle_H\phi_{n,j}\right\|_H \\ &+ \left\|\sum_{j=1}^{k_n}\frac{D_{n,j}}{C_j}\langle\phi_{n,j},x\rangle_H\phi_{n,j}-\sum_{j=1}^{k_n}\rho_j\langle\phi'_{n,j},x\rangle_H\phi'_{n,j}\right\|_H \\ &= a_{k_n,1}(x)+a_{k_n,2}(x), \qquad \text{(A7.26)}\end{aligned}$$

where $\langle\phi_j,x\rangle_H\phi_j=\langle\phi'_{n,j},x\rangle_H\phi'_{n,j}$, with, as before,

$$\phi'_{n,j}=\mathrm{sgn}\langle\phi_{n,j},\phi_j\rangle_H\phi_j,\quad \mathrm{sgn}\langle\phi_{n,j},\phi_j\rangle_H=\mathbf{1}_{\langle\phi_{n,j},\phi_j\rangle_H\geq 0}-\mathbf{1}_{\langle\phi_{n,j},\phi_j\rangle_H<0},\quad j\geq 1,\quad n\geq 2.$$

From equation (A7.26),

$$\begin{aligned}a_{k_n,1}(x) &\leq \sum_{j=1}^{k_n}|D_{n,j}|\frac{|C_j-C_{n,j}|}{C_{n,j}C_j}\left|\langle\phi_{n,j},x\rangle_H\right|\|\phi_{n,j}\|_H \\ &\leq \|C-C_n\|_{\mathcal{L}(H)}\frac{1}{C_{k_n}}\sum_{j=1}^{k_n}\left|\frac{D_{n,j}}{C_{n,j}}\right|\left|\langle\phi_{n,j},x\rangle_H\right|.\end{aligned}$$

Thus, from Cauchy–Schwarz's inequality (see also Remark A7.9.2 provided in the Supplementary Material),

$$a_{k_n,1}(x) \leq \|C - C_n\|_{\mathcal{L}(H)} \frac{1}{C_{k_n}} \left(\sum_{j=1}^{k_n} \frac{D_{n,j}^2}{C_{n,j}^2} \right)^{1/2} \left(\sum_{j=1}^{\infty} \langle \phi_{n,j}, x \rangle_H^2 \right)^{1/2}$$
$$\leq 2 \|C - C_n\|_{\mathcal{L}(H)} \frac{1}{C_{k_n}} k_n^{1/2} \|x\|_H \ a.s. \tag{A7.27}$$

Under **Assumption A1**, for $n \geq \widetilde{n}_0$, $k_n < C_{k_n}^{-1}$, which implies that, from the definition of $\{a_j,\ j \geq 1\}$ (see Remark A7.9.4 provided in the Supplementary Material),

$$a_{k_n,1}(x) \leq 2 \|C - C_n\|_{\mathcal{L}(H)} C_{k_n}^{-3/2} \|x\|_H < 2 \|C - C_n\|_{\mathcal{L}(H)} \|x\|_H C_{k_n}^{-1/2} \sum_{j=1}^{k_n} a_j \ a.s. \tag{A7.28}$$

From condition (A7.23), there also exists a positive real number $M < \infty$ and an integer n_0 such that, for certain $\beta > \frac{1}{2}$ and $n \geq n_0$, with n_0 large enough,

$$C_{k_n}^{-1/2} \sum_{j=1}^{k_n} a_j < C_{k_n}^{-1} \sum_{j=1}^{k_n} a_j \leq M n^{1/4} (\ln(n))^{-\beta}. \tag{A7.29}$$

From equations (A7.28)–(A7.29), for any $n \geq \max(\widetilde{n}_0, n_0)$, since $x \in H$ and under **Assumption A3** (see Theorem A7.9.1 provided in the Supplementary Material),

$$a_{k_n,1}(x) < 2M \frac{n^{1/4}}{(\ln(n))^{\beta}} \|C - C_n\|_{\mathcal{L}(H)} \|x\|_H \to^{a.s.} 0, \quad n \to \infty.$$

Concerning $a_{k_n,2}(x)$ in (A7.26), under **Assumptions A1–A2** and Cauchy–Schwarz's inequality (see also Remark A7.9.2 provided in the Supplementary Material provided),

$$a_{k_n,2}(x) \leq \left\| \sum_{j=1}^{k_n} \frac{D_{n,j}}{C_j} \left(\langle \phi_{n,j}, x \rangle_H - \langle \phi_{n,j}^{'}, x \rangle_H \right) \phi_{n,j} \right\|_H$$
$$+ \left\| \sum_{j=1}^{k_n} \frac{D_{n,j}}{C_j} \langle \phi_{n,j}^{'}, x \rangle_H \left(\phi_{n,j} - \phi_{n,j}^{'} \right) \right\|_H$$
$$+ \left\| \sum_{j=1}^{k_n} \left(\frac{D_{n,j}}{C_j} - \rho_j \right) \langle \phi_{n,j}^{'}, x \rangle_H \phi_{n,j}^{'} \right\|_H$$
$$\leq 2 \sup_{j \geq 1} |C_{n,j}| C_{k_n}^{-1} \sum_{j=1}^{k_n} \left| \langle \phi_{n,j} - \phi_{n,j}^{'}, x \rangle_H \right| \|\phi_{n,j}\|_H$$

$$
\begin{aligned}
&+ \quad 2 \sup_{j \geq 1} |C_{n,j}|\, C_{k_n}^{-1} \sum_{j=1}^{k_n} \left| \langle \phi_{n,j}^{'}, x \rangle_H \right| \left\| \phi_{n,j} - \phi_{n,j}^{'} \right\|_H \\
&+ \quad \sup_{j \geq 1} |D_{n,j} - D_j|\, C_{k_n}^{-1} \left\| \sum_{j=1}^{k_n} \langle \phi_{n,j}^{'}, x \rangle_H \phi_{n,j}^{'} \right\|_H \quad a.s. \qquad (A7.30)
\end{aligned}
$$

Hence, from Cauchy–Schwarz's inequality,

$$
\begin{aligned}
a_{k_n,2}(x) \quad \leq \quad & 2 \sup_{j \geq 1} |C_{n,j}|\, \|x\|_H\, C_{k_n}^{-1} \sum_{j=1}^{k_n} \left\| \phi_{n,j} - \phi_{n,j}^{'} \right\|_H \\
+ \quad & 2 \sup_{j \geq 1} |C_{n,j}|\, \|x\|_H\, C_{k_n}^{-1} \sum_{j=1}^{k_n} \left\| \phi_{n,j}^{'} \right\|_H \left\| \phi_{n,j} - \phi_{n,j}^{'} \right\|_H \\
+ \quad & \sup_{j \geq 1} |D_{n,j} - D_j|\, \|x\|_H\, C_{k_n}^{-1}. \qquad (A7.31)
\end{aligned}
$$

Since, for n sufficiently large, from Theorem A7.9.1 included in the Supplementary Material provided, and under **Assumption A3**, C_n admits a diagonal decomposition in terms of $\{C_{n,j},\ j \geq 1\}$,

$$
a_{k_n,2}(x) \leq 4 \|C_n\|_{\mathcal{L}(H)} \|x\|_H\, C_{k_n}^{-1} \sum_{j=1}^{k_n} \left\| \phi_{n,j} - \phi_{n,j}^{'} \right\|_H + \sup_{j \geq 1} |D_{n,j} - D_j|\, \|x\|_H\, C_{k_n}^{-1}\ a.s. \quad (A7.32)
$$

From results in [Bosq, 2000, Lemma 4.3],

$$
a_{k_n,2}(x) \leq 4 \|C_n\|_{\mathcal{L}(H)} \|x\|_H \|C_n - C\|_{\mathcal{L}(H)}\, C_{k_n}^{-1} \sum_{j=1}^{k_n} a_j + \sup_{j \geq 1} |D_{n,j} - D_j|\, \|x\|_H\, C_{k_n}^{-1}\ a.s. \quad (A7.33)
$$

On the one hand, from equation (A7.23), there exists a positive real number $M < \infty$ and an integer n_0 large enough such that, for certain $\beta > \frac{1}{2}$ and $n \geq n_0$,

$$
a_{k_n,2}(x) \leq 4M \|C_n\|_{\mathcal{L}(H)} \|x\|_H \|C_n - C\|_{\mathcal{L}(H)} \frac{n^{1/4}}{(\ln(n))^{\beta}} + \sup_{j \geq 1} |D_{n,j} - D_j|\, \|x\|_H\, C_{k_n}^{-1}\ a.s. \quad (A7.34)
$$

On the other hand, for n large enough and any $\beta > \frac{1}{2}$,

$$
a_{k_n,2}(x) < M \|x\|_H \left(4 \|C_n\|_{\mathcal{L}(H)} \|C_n - C\|_{\mathcal{L}(H)} + \sup_{j \geq 1} |D_{n,j} - D_j| \right) \frac{n^{1/4}}{(\ln(n))^{\beta}}\ a.s. \qquad (A7.35)
$$

Hence, since $\|C_n\|_{\mathcal{L}(H)} < \infty$ and $\|x\|_H < \infty$, from Theorem A7.9.1 and Corollary A7.9.2 both included in the Supplementary Material provided, from conditions (A7.23) and under **Assumptions A1–**

A3,

$$a_{k_n,2}(x) \to^{a.s.} 0, \quad n \to \infty. \tag{A7.36}$$

Taking supremum in $x \in H$, with $\|x\|_H = 1$, at the left–hand side of equation (A7.25), from equations (A7.26)–(A7.36), we obtain the desired result. Strong-consistency of the associated plug-in predictor is directly obtained, under **Assumption A3**. The upper-bound in (A7.24) can be directly obtained from $b_{k_n}(x)$, $a_{k_n,1}(x)$ and $a_{k_n,2}(x)$, reflected in equations (A7.25)-(A7.26) and (A7.30).

■

When ρ does not admit a diagonal spectral representation, an almost sure upper bound for the error $\|\widetilde{\rho}_{k_n} - \rho\|^2_{\mathcal{S}(H)}$ is provided in Section A7.9.3 in the Supplementary Material provided, being $\|\cdot\|^2_{\mathcal{S}(H)}$ the norm of the Hilbert-Schmidt operators on H.

A7.8 Comparative study: an evaluation of the performance

A comparative study is undertaken to illustrate the performance of the ARH(1) predictor formulated in Section A7.7, and those ones given by Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001], under different diagonal, pseudo-diagonal and non-diagonal scenarios, when $\{\phi_j,\ j \geq 1\}$ are unknown. Additionally to the figures displayed in this section, more numerical results can also be found in the tables included in Section A7.9.5 of the Supplementary Material provided.

In all of the scenarios considered, the autocovariance operator C is defined as the inverse of a power of the Dirichlet negative Laplacian operator on $[0, \delta]$. Namely, the spectral decomposition of C is determined, in all of the scenarios, by

$$C(f)(g) = \sum_{j=1}^{\infty} C_j \langle \phi_j, f\rangle_H \langle \phi_j, g\rangle_H, \quad \phi_j(x) = \sqrt{\frac{2}{\delta}} \sin\left(\frac{\pi j x}{\delta}\right), \quad C_j = c_1 j^{-\beta_C}, \tag{A7.37}$$

for $f, g \in H = L^2((0,\delta))$ and $x \in (0,\delta)$, being c_1 a positive constant. In the remaining, we fix $(0,\delta) = (0,4)$. Concerning $\{C_j,\ j \geq 1\}$, different rates β_C will be regarded, such that **Assumption A1** is directly held; i.e., $\beta_C > 1$.

On the other hand, the coefficients corresponding to the spectral decomposition of ρ and C_ε (see equations (A7.5)-(A7.6) above), related to the tensorial product $\{\phi_j \otimes \phi_h,\ j, h \geq 1\}$, are given by

$$\rho_{j,j} = c_2 j^{-\beta_\rho}, \quad \sigma^2_{j,j} = C_j \left(1 - \rho^2_{j,j}\right), \quad j \geq 1,$$

with $\beta_\rho = 11/10$, and, for any $j \neq h,\ j, h \geq 1$,

$$\rho_{j,h} = \begin{cases} 0, & \textbf{scenario D} \\ e^{-|j-h|/W} & \textbf{scenario PD} \\ \frac{1}{K}\frac{1}{|j-h|^2+1} & \textbf{scenario ND} \end{cases}, \quad \sigma^2_{j,h} = \begin{cases} 0, & \textbf{scenario D} \\ e^{-|j-h|^2/W} & \textbf{scenario PD} \\ e^{-|j-h|^2/W} & \textbf{scenario ND} \end{cases}$$

for diagonal (D), pseudodiagonal (PD) and non-diagonal (ND) scenarios, being $\frac{1}{K} = 0.275$. Henceforth, c_2 is a constant in $(0,1)$, verifying **Assumption A2**.

A7.8.1 Large-sample behaviour of the ARH(1) plug-in predictors

Large-sample behaviour of the ARH(1) plug-in predictor formulated in Section A7.7.2, as well as those ones in Bosq [2000]; Guillas [2001] (see equations (A7.4) and (A7.7) above, respectively), will be illustrated. ARH(1) plug-in predictors established in Section A7.7 will be only considered under diagonal scenarios.

As commented earlier (see equation (A7.4) above), **Assumptions A1** and **A3–A5**, and the Hilbert-Schmidt assumption of ρ, are required in the strong-consistency results by Bosq [2000]. Condition (A7.23) was also imposed. From [Bosq, 2000, Example 8.6] conditions therein considered are held under any scenario in which the truncation parameter $k_n = \lceil \log(n) \rceil$ is adopted, under **Assumptions A1–A4** (it can be proved as condition (A7.23) is also verified when $k_n = \lceil \log(n) \rceil$). In the formulation of mean-square convergence, Guillas also considered **Assumptions A1**, **A3** and **A5**. From [Guillas, 2001, Theorem 2 and Example 4], if the regularization sequence above-referred (see equation (A7.7)) verifies

$$\alpha \frac{C_{k_n}^{\gamma}}{n^{\epsilon}} \leq u_n \leq \beta C_{k_n}, \quad 0 < \beta < 1, \quad \alpha > 0, \quad \gamma = 1, \quad \epsilon = 0,$$

then the mean-square consistency is achieved. Namely, if

$$k_n = \lceil e' n^{1/(8\delta_C + 2)} \rceil, \quad e' = 17/10,$$

the rate of convergence in quadratic mean is of order of

$$n^{-\delta_C/(4\delta_C+1)}.$$

Since

$$\lceil (17/10)\, n^{1/(8\delta_C+2)} \rceil < \lceil \ln(n) \rceil$$

for n large enough, condition (A7.23) is also verified when

$$k_n = \lceil e' n^{1/(8\delta_C+2)} \rceil.$$

For sample sizes $n_t = 35000 + 40000\,(t-1)\,,\ t = 1, \dots, 10$, the error measure

$$F\left(k_n, n_t, \beta\right) = \left(\sum_{l=1}^{N} \mathbf{1}_{\left(\xi_{n_t,\beta}, \infty\right)} \left(\left\| \left(\rho - \overline{\rho}_{k_n}^{l}\right) \left(X_{n-1}^{l}\right) \right\|_H^{k_n} \right) \right) / N, \qquad \text{(A7.38)}$$

will be displayed (see Figures A7.8.1-A7.8.3 below), being $\mathbf{1}_{\left(\xi_{n_t,\beta}, \infty\right)}$ the indicator function over the interval $(\xi_{n_t,\beta}, \infty)$, where $\xi_{n_t,\beta}$ numerically fits the almost sure rate of convergence of $\left\| \left(\rho - \overline{\rho}_{k_n}^{l}\right) \left(X_{n-1}^{l}\right) \right\|_H^{k_n}$. The following diagonal subscenarios will be considered (see Figure A7.8.1), when the diagonal data generation is assumed, for

$$\delta_\rho = 11/10, \quad n_t = 35000 + 40000(t-1),\ t = 1,\dots,10, \quad \xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/2}}, \quad \beta = 65/100$$

$$\delta_C = \begin{cases} 3/2 & \textbf{scenarios } D_1,\ D_3 \\ 24/10 & \textbf{scenarios } D_2,\ D_4 \end{cases}, \quad k_n = \begin{cases} \lceil \ln(n) \rceil & \textbf{scenarios } D_1,\ D_2 \\ \lceil e' n^{1/(8\delta_C+2)} \rceil & \textbf{scenarios } D_3,\ D_4 \end{cases},$$

being $e' = 17/10$. As discussed, conditions formulated in Bosq [2000] and Proposition A7.7.1 of the current paper are held for scenarios D_1-D_2, while in scenarios D_3-D_4, the conditions assumed in Proposition A7.7.1, Bosq [2000]; Guillas [2001] are verified. In the subscenarios D_1-D_4,

$$\left\| \left(\rho - \overline{\rho}_{k_n}^l\right)\left(X_{n-1}^l\right)\right\|_H^{k_n} = \sqrt{\int_a^b \left(\sum_{j=1}^{k_n} \rho_j X_{n-1,n,j}^l \phi_j(t) - \sum_{j=1}^{k_n} \overline{\rho}_{n,j}^l \left(X_{n-1}^l\right)\phi_{n,j}^l(t)\right)^2 dt}, \quad \text{(A7.39)}$$

is computed, being $\overline{\rho}_{k_n}^l\left(X_{n-1}^l\right)$ the predictors defined in (A7.19)-(A7.22), (A7.4) and (A7.7), respectively, for any $j = 1,\dots,k_n$, and based on the l–th generation of the values $\widetilde{X}_{i,n,j}^l = \langle X_i^l, \phi_{n,j}^l \rangle_H$, for $l = 1,\dots,N$, with $N = 500$ simulations.

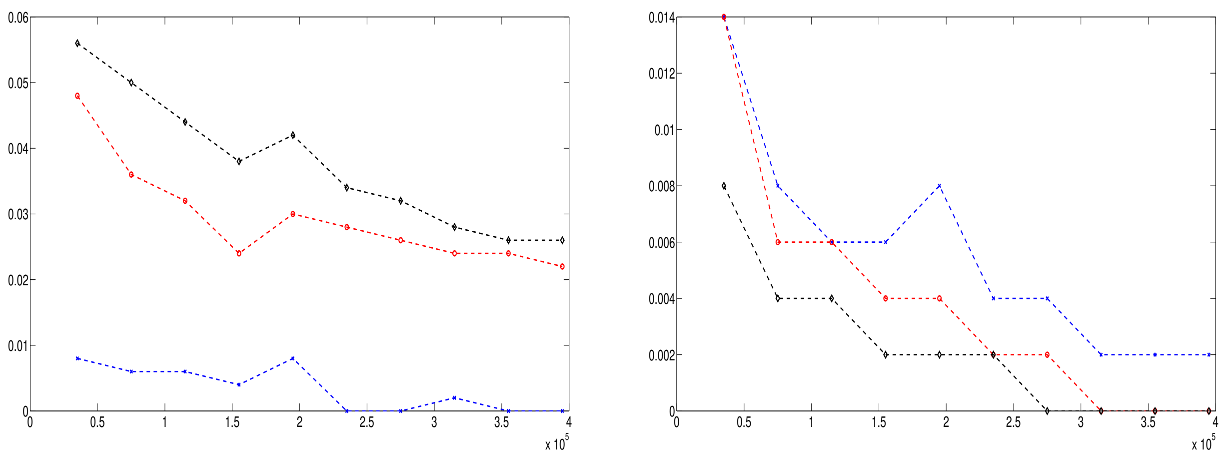


Figure A7.8.1: $F(k_n, n_t, \beta)$ values, for scenarios D_2 (on left) and D_4 (on right), for our approach (blue star dotted line) and those one presented in Bosq [2000] (red circle line) and Guillas [2001] (black diamond line). The curve $\xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/2}}$, with $\beta = 65/100$, is adopted.

Parameters δ_C and k_n for pseudodiagonal scenarios (scenarios PD_1-PD_4) and non-diagonal scenarios (scenarios ND_1-ND_4) are fixed as done above for scenarios D_1-D_4, being

$$\delta_2 = 11/10, \quad n_t = 35000 + 40000(t-1),\ t = 1,\dots,10, \quad \xi_{n_t,\beta} = (\ln(n_t))^\beta n_t^{-1/3}.$$

Values of $\beta = 3/10$ and $\beta = 125/100$ are distinguished for pseudodiagonal and non-diagonal scenarios (see Figure A7.8.3), respectively. Note that, as discussed above, different values of $\left\{\rho_{j,h},\ \sigma_{j,h}^2, \quad j,\ h \geq 1\right\}$ are adopted for these cases. In fact, under pseudodiagonal and non-diagonal frameworks, the following truncated norm is then computed, instead of (A7.70):

$$\sqrt{\int_a^b \left(\int_a^b \left(\sum_{j,k=1}^{k_n} \rho_{j,k}\phi_j(t)\phi_k(s)\right) ds - \sum_{j=1}^{k_n} \overline{\rho}_{n,j}^l \left(X_{n-1}^l\right) \phi_{n,j}^l(t)\right)^2 dt}. \tag{A7.40}$$

Remark that PD_1-PD_2 and ND_1-ND_2 scenarios verify conditions required in Bosq [2000], while scenarios PD_3-PD_4 and ND_3-ND_4 are included in both setting of conditions.

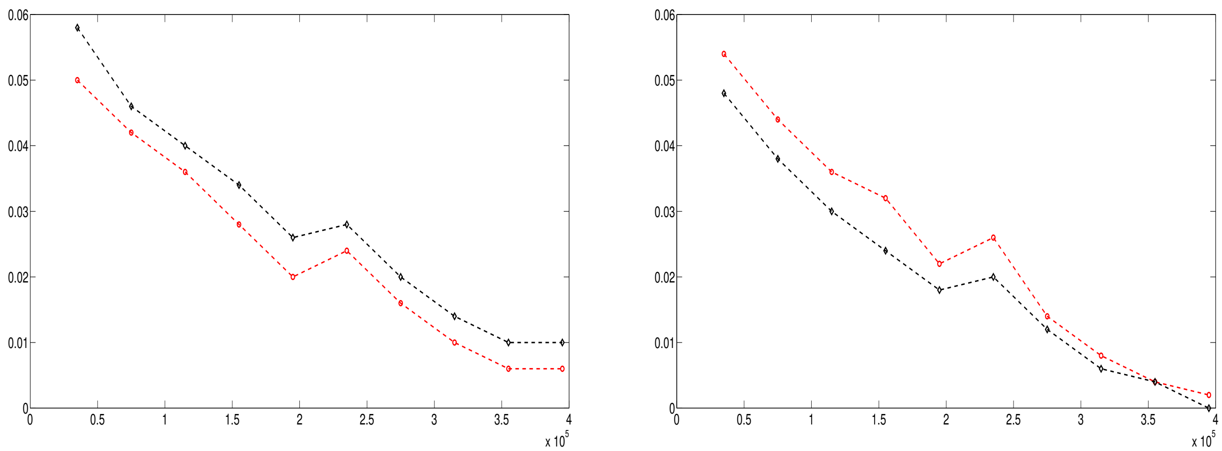


Figure A7.8.2: $F\left(k_n, n_t, \beta\right)$ values, for scenario PD_2 (on left) and scenario PD_4 (on right), for approaches presented in Bosq [2000] (red circle dotted line) and Guillas [2001] (black diamond dotted line). The curve $\xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/3}}$, with $\beta = 3/10$, is adopted.

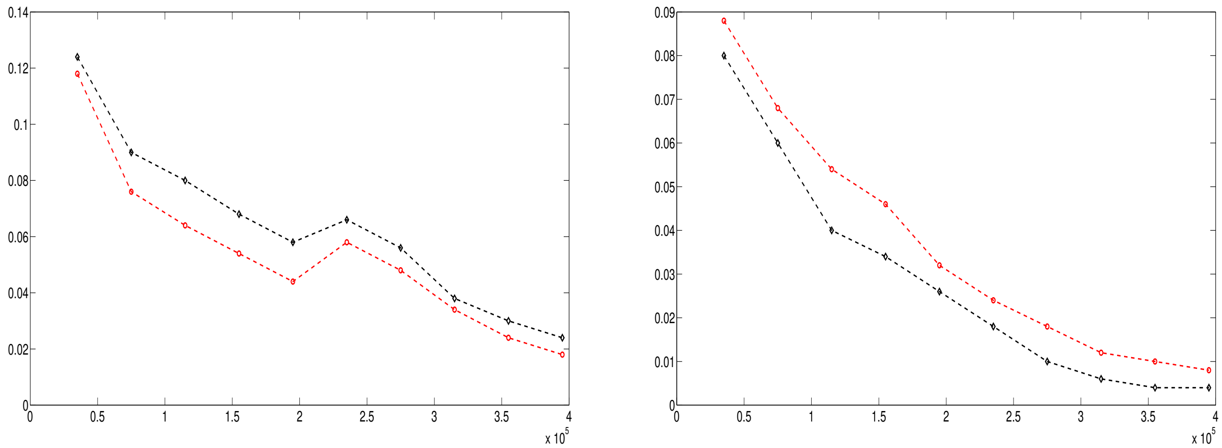


Figure A7.8.3: $F(k_n, n_t, \beta)$ values, for scenario ND_2 (on left) and scenario ND_4 (on right), for approaches presented in Bosq [2000] (red circle line) and Guillas [2001] (black diamond line), with $\xi_{n_t,\beta} = \frac{(\ln(n_t))^{\beta}}{n_t^{1/3}}$ and $\beta = 125/100$.

Diagonal scenarios $D_1 - D_4$ have been applied to the three componentwise plug-in predictors. As expected, the amount of values $\left\|\left(\rho - \overline{\rho}_{k_n}^{l}\right)\left(X_{n-1}^{l}\right)\right\|_H^{k_n}$, which lie within the band $[0, \xi_{n_t,\beta})$, is greater as long as the decay rate of the eigenvalues of C is faster. Since a diagonal framework is considered in scenarios $D_1 - D_2$, a better performance of the approach here proposed can be noticed, in comparison with those ones by Bosq [2000]; Guillas [2001], where errors appear, when sample sizes are not sufficiently large, in the estimation of the non-diagonal componentwise coefficients of ρ. This possible effect of the non-diagonal design, under a diagonal scenario, is not observed, for truncation rules selecting a very small number of terms, in relation to the sample size. This fact occurs in the truncation rule adopted in scenarios $D_3 - D_4$. In the pseudo-diagonal and non-diagonal scenarios, methodologies in Bosq [2000]; Guillas [2001] are compared, such that curves $\xi_{n_t,\beta} = \frac{(\ln(n_t))^{\beta}}{n_t^{1/3}}$, for $\beta = 3/10$ and $\beta = 125/100$, numerically fit their almost sure rate of convergence. As observed (see also Tables 2-4 in the Supplementary Material provided), the sample-size dependent truncation rule, according to the rate of convergence to zero of the eigenvalues of C, plays a crucial role in the observed performance of both approaches.

A7.8.2 Small-sample behaviour of the ARH(1) predictors

Smaller sample sizes must be adopted in this subsection, since computational limitations arise when the approaches formulated in Antoniadis and Sapatinas [2003] (see equations (A7.12)-(A7.14) above), as well as penalized predictor and non-parametric kernel-based predictor in Besse et al. [2000] (see equations (A7.8) and (A7.16), respectively), are included in the comparative study. See also Section A7.9.5 in the Supplementary Material, where extra numerical results are provided.

On the one hand, **Assumptions A1** and **A3**, and conditions in (A7.14), are required when regularized-wavelet-based prediction approach is applied. In particular, since $C_j = c_1 j^{-\delta_C}$, for any $j \geq 1$, if

$k_n = \lceil n^{1/\alpha} \rceil$ is adopted, then

$$1 - \frac{4\delta_C}{\alpha} > 0 \Rightarrow \alpha > 4\delta_C.$$

Additionally to $k_n = \lceil \ln(n) \rceil$, the truncation parameter $k_n = \lceil n^{1/\alpha} \rceil$ will be adopted, with $\alpha = 6.5$ and $\alpha = 10$, for $\delta_C = 3/2$ and $\delta_C = 24/10$, respectively. Furthermore, $F(k_n, n_t, \beta)$ values defined in (A7.69)-(A7.71) are computed for the wavelet-based approach just replacing $\{\phi_{n,j},\ j \geq 1\}$ by $\left\{\widetilde{\phi}_j^M,\ j \geq 1\right\}$ (see equations (A7.12)-(A7.14)). As before, since

$$\lceil n^{1/\alpha} \rceil < \lceil \ln(n) \rceil, \quad \lceil n^{1/\alpha} \rceil < \lceil (17/10) n^{1/(8\delta_C+2)} \rceil, \quad \alpha = 6.5, \quad \alpha = 10,$$

conditions formulated in Section A7.7, as well as in Bosq [2000]; Guillas [2001], are verified when $k_n = \lceil n^{1/\alpha} \rceil$, with $\alpha = 6.5$ and $\alpha = 10$, is studied.

On the other hand, the referred methodologies in Besse et al. [2000] are implemented, the following alternative norm replaces the norm reflected in (A7.70)-(A7.71), respectively, for values $F(k_n, n_t, \beta)$:

$$\left\| \left(\rho - \overline{\rho}_{k_n}^l\right)\left(X_{n-1}^l\right) \right\|_H = \sqrt{\int_a^b \left(\rho\left(X_{n-1}^l\right)(t) - \overline{\rho}_{k_n}^l\left(X_{n-1}^l\right)(t)\right)^2 dt}, \quad l = 1, \dots, N. \tag{A7.41}$$

In this small-sample size context, the following diagonal subscenarios will be considered (see Figure A7.8.4), when the diagonal data generation is assumed, for

$$\delta_\rho = 11/10, \quad n_t = 750 + 500(t-1),\ t = 1, \dots, 13, \quad \xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/2}}, \quad \beta = 65/100$$

$$\delta_C = \begin{cases} 3/2 & \textbf{scenarios } D_5,\ D_7 \\ 24/10 & \textbf{scenarios } D_6,\ D_8 \end{cases}, \quad k_n = \begin{cases} \lceil \ln(n) \rceil & \textbf{scenarios } D_5,\ D_6 \\ \lceil n^{1/\alpha} \rceil,\ \alpha = 6.5 & \textbf{scenarios } D_7,\ D_8 \end{cases},$$

being $q = 10$ the dimension of the subspace H_q involved in the penalized estimation proposed in Besse et al. [2000] (see also equation (A7.8)). Remark that, since approaches formulated in Besse et al. [2000] not depend on the truncation parameter k_n adopted, we only perform them for scenarios D_5-D_6, where conditions imposed in that paper are verified. In the case of kernel-based predictor is used, two bandwidths $h_n = 0.15,\ 0.25$ are considered in both scenarios. Conditions formulated in Bosq [2000] and Proposition A7.7.1 of the current paper are held for all scenarios, while the conditions assumed in Antoniadis and Sapatinas [2003]; Guillas [2001] are only verified under scenarios D_7-D_8.

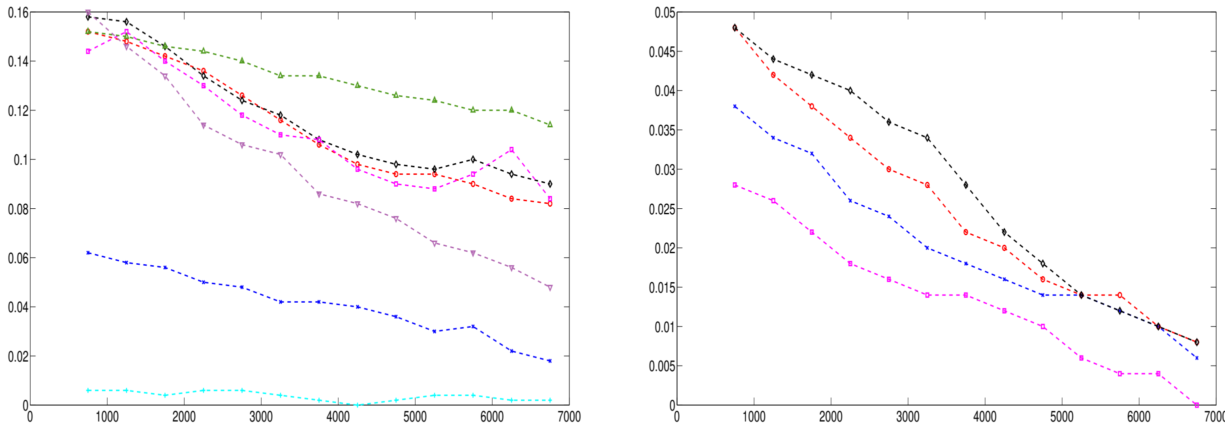
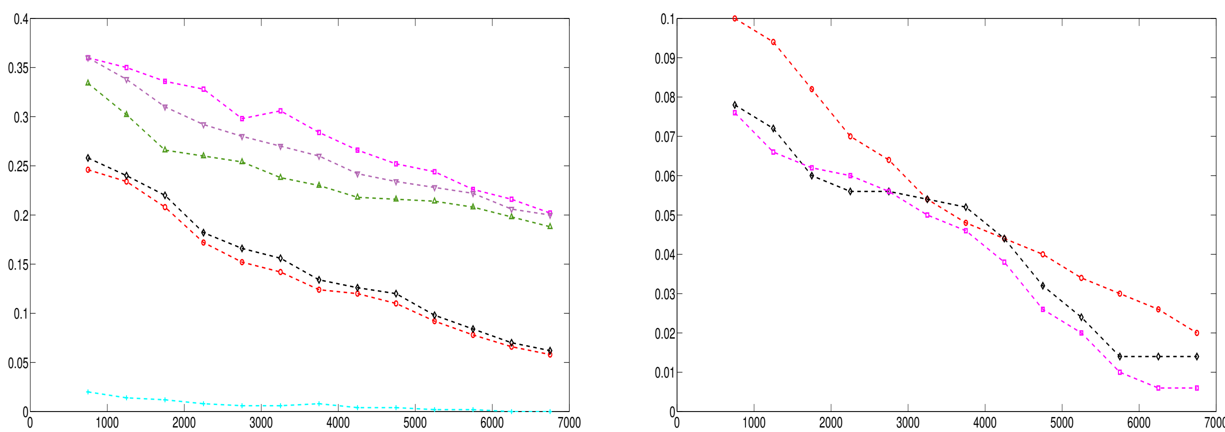


Figure A7.8.4: $F(k_n, n_t, \beta)$ values, for scenario D_6 (on left) and scenario D_8 (on right), for our approach (blue star line) and those one presented in Antoniadis and Sapatinas [2003] (pink square line), Besse et al. [2000] (cyan blue plus line for penalized prediction; dark green upward-pointing triangle and purple downward-pointing triangle lines, for kernel-based prediction, for $h_n = 0.15$ and $h_n = 0.25$, respectively), Bosq [2000] (red circle line) and Guillas [2001] (black diamond line). The curve $\xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/2}}$, with $\beta = 65/100$, is drawn (light green line).



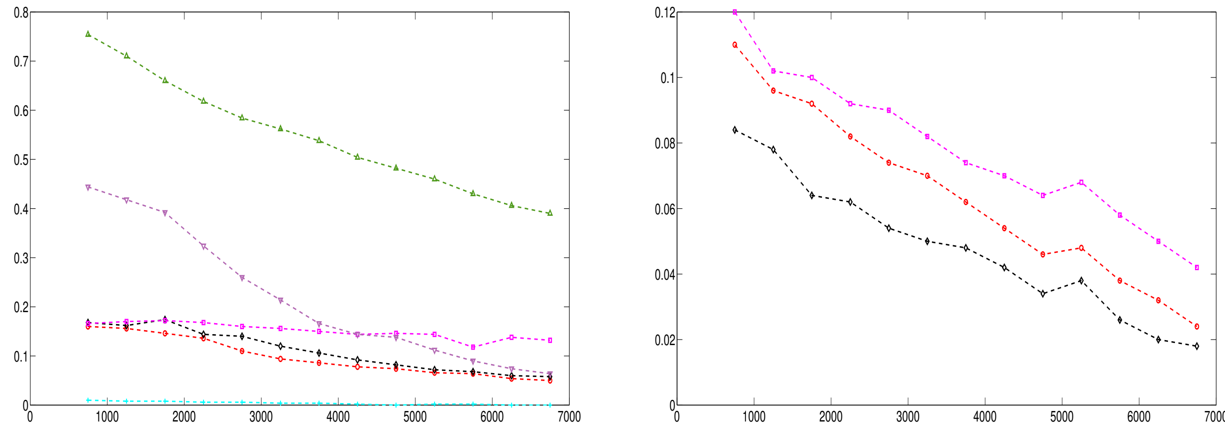


Figure A7.8.5: $F(k_n, n_t, \beta)$ values, for scenarios PD_6 (at the top, on left) and PD_8 (at the top, on right), and scenarios ND_6 (at the bottom, on left) and ND_8 (at the bottom, on right), for approaches presented in Antoniadis and Sapatinas [2003] (pink square line), Besse et al. [2000] (cyan blue plus line for penalized prediction; dark green upward-pointing triangle and purple downward-pointing triangle lines, for kernel-based prediction, for $h_n = 1.2$ and $h_n = 1.7$, respectively), Bosq [2000] (red circle line) and Guillas [2001] (black diamond line). The curve $\xi_{n_t,\beta} = \frac{(\ln(n_t))^{\beta}}{n_t^{1/3}}$, with $\beta = 3/10$ (at the top) and $\beta = 125/100$ (at the bottom), is drawn (light green line).

The same values of δ_C and k_n are adopted when pseudodiagonal scenarios (scenarios PD_5-PD_8) and non-diagonal scenarios (scenarios ND_5-ND_8) are analysed (see Figure A7.8.5). As before, the curve $\xi_{n_t,\beta} = \frac{(\ln(n_t))^{\beta}}{n_t^{1/3}}$ is regarded, for pseudodiagonal and non-diagonal scenarios, with $\beta = 3/10$ and $\beta = 125/100$, respectively. While conditions in Bosq [2000] are verified for all scenarios, scenarios developed by Antoniadis and Sapatinas [2003]; Guillas [2001] are only held when the truncation parameter proposed in Antoniadis and Sapatinas [2003] is adopted. When smaller sample sizes are adopted, and approaches formulated in Antoniadis and Sapatinas [2003]; Besse et al. [2000] are included in the comparative study, new scenarios have been considered. Note that even when small sample sizes are studied, a good performance of the ARH(1) plug-in predictor given in equations (A7.19)-(A7.22) is observed. As well as the regularized wavelet-based approach detailed in Antoniadis and Sapatinas [2003] becomes the best methodology for small sample sizes, in comparision with the componentwise techniques above mentioned. Note that the good performance observed corresponds to the truncation rule proposed by these authors, with a small number of terms. While, when a larger number of terms is considered, according to the alternative truncation rules tested, the observed outperformance does not hold. While the penalized prediction approach proposed in Besse et al. [2000] has been shown as the more accurate, is, however, less affected by the regularity conditions imposed on the autocovariance kernel (see Tables 5-10 included in the Supplementary Material). Furthermore, a drawback of both approaches in Antoniadis and Sapatinas [2003]; Besse et al. [2000] is that they require large computational times. The underlying dependence structure cannot be provided in those approaches.

A7.9 Supplementary Material

This section provides, as Supplementary Material, the auxiliary results required (see Sections A7.9.1–A7.9.2 below). Under a non-diagonal framework, Section A7.9.3 provides a theoretical almost sure upper bound for the error in the norm of $\mathcal{S}(H)$ associated with the diagonal componentwise estimator of the autocorrelation operator considered in Section A7.7.2, when the eigenvectors of the autocovariance operator are unknown. A simulation study is undertaken in Section A7.9.4 to illustrate the large sample behaviour of the formulated estimator. Tables displaying more detailed numerical results are provided in Section A7.9.5

A7.9.1 Diagonal strongly-consistent estimator when the eigenvectors of C are known

From model proposed in (A7.18), we can formally defined the autocorrelation operator as $\rho(x) = DC^{-1}(x)$, for any $x \in H$. Nevertheless, it is well known that the operator C cannot be inverted in the whole domain. That is, an empirical estimator of C must be computed. In the case of $\{\phi_j,\ j \geq 1\}$ are known, we can define an empirical estimator $\widehat{C}_n$, as well as of $\widehat{D}_n$, admitting the following diagonal spectral decomposition, for each $n \geq 2$:

$$\widehat{C}_n = \sum_{j=1}^{\infty} \widehat{C}_{n,j}\phi_j \otimes \phi_j, \quad \widehat{C}_{n,j} = \frac{1}{n}\sum_{i=0}^{n-1} X_{i,j}^2, \quad j \geq 1, \tag{A7.42}$$

$$\widehat{D}_n = \sum_{j=1}^{\infty} \widehat{D}_{n,j}\phi_j \otimes \phi_j, \quad \widehat{D}_{n,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,j}X_{i+1,j}, \quad j \geq 1, \tag{A7.43}$$

where $\{\phi_j,\ j \geq 1\}$ is the eigenvectors system of C, with $\left\{\widehat{C}_{n,j},\ j \geq 1\right\}$ and $\left\{\widehat{D}_{n,j},\ j \geq 1\right\}$ being the eigenvalues of $\widehat{C}_n$ and $\widehat{D}_n$, respectively, for any $n \geq 2$.

Remark A7.9.1 *Under definitions in equations (A7.42)-(A7.43), the diagonal componentwise estimator, introduced in equation (A7.44) below, for the autocorrelation operator ρ, naturally arises, which is different from the componentwise estimator approaches based on the projection of the natural empirical covariance operators C_n and D_n, given by*

$$C_n = \frac{1}{n}\sum_{i=0}^{n-1} X_i \otimes X_i, \quad D_n = \frac{1}{n-1}\sum_{i=0}^{n-2} X_i \otimes X_{i+1}, \quad n \geq 2,$$

into the empirical eigenvectors.

Henceforth, the following assumption will be also required in this subsection:

Assumption A5. $X_{0,j}^2 = \langle X_0, \phi_j\rangle_H^2 > 0$, a.s., for every $j \geq 1$.

Remark A7.9.2 *From Cauchy-Schwarz's inequality, for any $j \geq 1$ and $n \geq 2$, under* **Assumption A5**,

$$\left| \frac{\frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2} \right| \leq \frac{2 \left(\frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2 \right)}{\frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2} = 2 \, a.s.$$

From **Assumption A5**, let us consider the diagonal componentwise estimator of ρ,

$$\widehat{\rho}_{k_n} = \sum_{j=1}^{k_n} \widehat{\rho}_{n,j} \phi_j \otimes \phi_j, \quad \widehat{\rho}_{n,j} = \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}} = \frac{n}{n-1} \frac{\sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\sum_{i=0}^{n-1} X_{i,j}^2}, \quad j \geq 1, \; n \geq 2. \tag{A7.44}$$

Remark A7.9.3 *Note that, under* **Assumption A1**, *the eigenvalues of C are strictly positive, with multiplicity one, and $C(H) = H$, where $C(H)$ denotes the range of C. For $f, \; g \in C(H)$, there exist $\varphi, \psi \in H$ such that $f = C(\varphi)$ and $g = C(\psi)$, and the following identities hold:*

$$\begin{aligned} \langle f, g \rangle_{C(H)} &= \langle C^{-1} C(\varphi), C^{-1} C(\psi) \rangle_H = \langle \varphi, \psi \rangle_H < \infty, \\ \|f\|_{C(H)}^2 &= \langle C^{-1} C(\varphi), C^{-1} C(\varphi) \rangle_H = \|\varphi\|_H^2 < \infty. \end{aligned} \tag{A7.45}$$

From Parseval's identity, $\|x\|_{C(H)}^2 = \sum_{j=1}^{\infty} \frac{[\langle x, \phi_j \rangle_H]^2}{C_j^2} < \infty$, *for any $x \in C(H)$. Thus, the range of C can be also defined by*

$$C(H) = \left\{ x \in H : \quad \sum_{j=1}^{\infty} \frac{\langle x, \phi_j \rangle_H^2}{C_j^2} < \infty \right\}. \tag{A7.46}$$

Under **Assumptions A1–A3** and **A5**, the following proposition provides the strong–consistency, in the norm of $\mathcal{L}(H)$, of the estimator (A7.44) of the autocorrelation operator, as well as of its associated ARH(1) plug–in predictor, in the underlying Hilbert space. Asymptotic properties derived in Section A7.9.2 below are required.

Proposition A7.9.1 *Under* **Assumptions A1–A3** *and* **A5**, *for a truncation parameter $k_n < n$, with $\lim_{n \to \infty} k_n = \infty$,*

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \|\widehat{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} \longrightarrow^{a.s.} 0, \quad \|(\widehat{\rho}_{k_n} - \rho)(X_{n-1})\|_H \longrightarrow^{a.s.} 0, \quad n \to \infty.$$

Proof.

Under **Assumption A1**, $C(H) = H$, as a set of functions. Then, for every $x \in C(H) = H$, under **Assumptions A2** and **A5**, from Parseval's identity and Remark A7.9.2,

$$\begin{aligned}
\|(\widehat{\rho}_{k_n} - \rho)(x)\|_H^2 &= \sum_{j=1}^{k_n} \left[(\widehat{\rho}_{n,j} - \rho_j) \langle x, \phi_j \rangle_H \right]^2 + \sum_{j=k_n}^{\infty} \left[\rho_j \langle x, \phi_j \rangle_H \right]^2 \\
&\leq \sum_{j=1}^{k_n} \left[\frac{D_j - \widehat{D}_{n,j}}{C_j} \langle x, \phi_j \rangle_H \right]^2 \\
&+ \sum_{j=1}^{k_n} \left[\frac{C_j - \widehat{C}_{n,j}}{C_j} \widehat{\rho}_{n,j} \langle x, \phi_j \rangle_H \right]^2 + \sum_{j=k_n}^{\infty} \left[\rho_j \langle x, \phi_j \rangle_H \right]^2 \\
&\leq \left[\sup_{1 \leq j \leq k_n} \left| D_j - \widehat{D}_{n,j} \right|^2 + 2 \sup_{1 \leq j \leq k_n} \left| C_j - \widehat{C}_{n,j} \right|^2 \right] \\
&\times \sum_{j=1}^{k_n} \left[\frac{\langle x, \phi_j \rangle_H}{C_j} \right]^2 + \sum_{j=k_n}^{\infty} \left[\rho_j \langle x, \phi_j \rangle_H \right]^2, \quad \text{a.s.} \qquad \text{(A7.47)}
\end{aligned}$$

Thus, taking the square root in booth sides of (A7.47), and the supremum in $x \in H = C(H)$, with $\|x\|_H = 1$, at the left–hand side, we obtain

$$\begin{aligned}
\|\widehat{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} &\leq \sup_{x \in H,\ \|x\|_H = 1} \left(\left[\sup_{1 \leq j \leq k_n} \left| D_j - \widehat{D}_{n,j} \right|^2 + 2 \sup_{1 \leq j \leq k_n} \left| C_j - \widehat{C}_{n,j} \right|^2 \right] \right. \\
&\left. \times \sum_{j=1}^{k_n} \left[\frac{\langle x, \phi_j \rangle_H}{C_j} \right]^2 + \sum_{j=k_n}^{\infty} \left[\rho_j \langle x, \phi_j \rangle_H \right]^2 \right)^{1/2} \quad \text{a.s.} \qquad \text{(A7.48)}
\end{aligned}$$

Furthermore, from **Assumptions A1–A2** and Remark A7.9.3, for every $x \in C(H) = H$,

$$\lim_{n \to \infty} \sum_{j=1}^{k_n} \left[\frac{\langle x, \phi_j \rangle_H}{C_j} \right]^2 = \|x\|_{C(H)}^2 < \infty, \quad \lim_{n \to \infty} \sum_{j=k_n}^{\infty} \left[\rho_j \langle x, \phi_j \rangle_H \right]^2 = 0. \qquad \text{(A7.49)}$$

Under **Assumptions A1–A3**, from Corollary A7.9.2 (see equations (A7.52)–(A7.53) in the Section A7.9.2 below), as $n \to \infty$,

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{1 \leq j \leq k_n} \left| C_j - \widehat{C}_{n,j} \right| \to^{a.s.} 0, \quad \frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{1 \leq j \leq k_n} \left| D_j - \widehat{D}_{n,j} \right| \to^{a.s.} 0. \qquad \text{(A7.50)}$$

Finally, from equations (A7.48)–(A7.50), as $n \to \infty$,

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \|\widehat{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} \to^{a.s.} 0.$$

Strong–consistency of the associated plug–in predictor is directly derived keeping in mind that

$$\|(\widehat{\rho}_{k_n} - \rho)(X_{n-1})\|_H \leq \|\widehat{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)} \|X_{n-1}\|_H, \quad \|X_{n-1}\|_H < \infty \, a.s.$$

■

A7.9.2 Asymptotic properties of the empirical eigenvalues and eigenvectors

This section presents the auxiliary results needed on the formulation of the theoretical results derived in Section A7.7. The asymptotic properties of the eigenvalues involved in the spectral decomposition of $\widehat{C}_n, \widehat{D}_n, C_n$ and D_n will be obtained in Corollary A7.9.1 below. Corollary A7.9.2 provides the asymptotic properties of the diagonal coefficients of D_n, with respect to the eigenvectors of C_n. In the derivation of these results, the following theorem plays a crucial role (see [Bosq, 2000, Theorem 4.1, Corollary 4.1 and Theorem 4.8]).

Theorem A7.9.1 *Under Assumption A3, for any $\beta > \frac{1}{2}$, as $n \to \infty$,*

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \|C_n - C\|_{\mathcal{S}(H)} \to^{a.s.} 0, \quad \frac{n^{1/4}}{(\ln(n))^{\beta}} \|D_n - D\|_{\mathcal{S}(H)} \to^{a.s.} 0,$$

and, if $\|X_0\|_H$ is bounded, being $\|\cdot\|_{\mathcal{S}(H)}$ the norm of Hilbert–Schmidt operators,

$$\|C_n - C\|_{\mathcal{S}(H)} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) a.s., , \|D_n - D\|_{\mathcal{S}(H)} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) a.s.$$

From Theorem A7.9.1, we obtain the following corollary on the asymptotic properties of the eigenvalues $\left\{\widehat{C}_{n,j},\ j \geq 1\right\}$ and $\left\{\widehat{D}_{n,j},\ j \geq 1\right\}$ of $\widehat{C}_n$ and $\widehat{D}_n$, respectively, as well as of the eigenvalues $\{C_{n,j},\ j \geq 1\}$ of the empirical estimator C_n and the diagonal coefficients $\widetilde{D}_{n,j} = D_n(\widetilde{\phi}_{n,j})(\widetilde{\phi}_{n,j}),\ j \geq 1$, with, for n sufficiently large,

$$D_n(\widetilde{\phi}_{n,j}) = \widetilde{D}_{n,j}\widetilde{\phi}_{n,j}, \quad j \geq 1. \tag{A7.51}$$

Corollary A7.9.1 *Under* **Assumptions A1–A3**, *the following identities hold, for any* $\beta > \frac{1}{2}$:

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} \left|\widehat{C}_{n,j} - C_j\right| \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|C_n - C\|_{\mathcal{S}(H)} \to^{a.s.} 0, \tag{A7.52}$$

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} \left|\widehat{D}_{n,j} - D_j\right| \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|D_n - D\|_{\mathcal{S}(H)} \to^{a.s.} 0, \tag{A7.53}$$

where, as before, $\{C_j,\ j \geq 1\}$ *and* $\{D_j,\ j\geq 1\}$ *are the systems of eigenvalues of* C *and* D*, respectively;* $\left\{\widehat{C}_{n,j},\ j \geq 1\right\}$ *and* $\left\{\widehat{D}_{n,j},\ j\geq 1\right\}$ *are given in (A7.43). In addition, for* n *sufficiently large,*

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} |C_{n,j} - C_j| \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|C_n - C\|_{\mathcal{S}(H)} \to^{a.s.} 0, \tag{A7.54}$$

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} \left|D_n(\widetilde{\phi}_{n,j})(\widetilde{\phi}_{n,j}) - D_j\right| \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|D_n - D\|_{\mathcal{S}(H)} \to^{a.s.} 0, \tag{A7.55}$$

where $\{C_{n,j},\ j\geq 1\}$ *are the empirical eigenvalues of* $C_n = \frac{1}{n}\sum_{i=0}^{n-1} X_i \otimes X_i$*, and* $\left\{\widetilde{D}_{n,j},\ j\geq 1\right\}$ *are given in (A7.51).*

Proof. Since $\widehat{C}_n$, with

$$\sum_{j=1}^{\infty} \widehat{C}_{n,j} = \frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=1}^{\infty} X_{i,j}^2 = \frac{1}{n}\sum_{i=0}^{n-1} \|X_i\|_H^2,$$

is in the trace class, then, under **Assumptions A1–A2**,

$$\begin{aligned}
\frac{n^{1/4}}{(\ln(n))^{\beta}} \|\widehat{C}_n - C\|_{\mathcal{S}(H)} &= \frac{n^{1/4}}{(\ln(n))^{\beta}} \sqrt{\sum_{j\geq 1} |\widehat{C}_{n,j} - C_j|^2} \\
&= \frac{n^{1/4}}{(\ln(n))^{\beta}} \sqrt{\sum_{j\geq 1} |C_n(\phi_j)(\phi_j) - C_j|^2} \\
&\leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \sqrt{\sum_{j,l\geq 1} |C_n(\phi_k)(\phi_l) - \delta_{j,l}C_j|^2} \\
&= \frac{n^{1/4}}{(\ln(n))^{\beta}} \|C_n - C\|_{\mathcal{S}(H)},
\end{aligned} \tag{A7.56}$$

where $\delta_{j,l}$ denotes the Kronecker delta function. From (A7.56), applying Theorem A7.9.1 under **Assumption A3**,

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} |\widehat{C}_{n,j} - C_j| \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|\widehat{C}_n - C\|_{\mathcal{S}(H)} \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|C_n - C\|_{\mathcal{S}(H)} \to^{a.s.} 0,$$

as we wanted to prove. Equation (A7.53) is obtained in a similar way to equation (A7.52), under **Assumptions A2–A3**, and keeping in mind that $\widehat{D}_n$ is, a.s., in the trace class, with

$$\left|\sum_{j=1}^{\infty} \widehat{D}_{n,j}\right| \leq 2 \sum_{j=1}^{\infty} \widehat{C}_{n,j} \; a.s.,$$

then

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} |\widehat{D}_{n,j} - D_j| \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|D_n - D\|_{\mathcal{S}(H)} \to^{a.s.} 0.$$

From Theorem A7.9.1 and under **Assumption A3** for n sufficiently large, C_n is a Hilbert–Schmidt operator, and in particular, it is a compact operator. Thus, applying [Bosq, 2000, Lemma 4.2] and Theorem A7.9.1, for $n \geq n_0$, with n_0 sufficiently large, we obtain

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{k\geq 1} |C_{n,k} - C_k| \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|C_n - C\|_{\mathcal{L}(H)} \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|C_n - C\|_{\mathcal{S}(H)} \to^{a.s.} 0.$$

Finally, as done in the derivation of (A7.54), equation (A7.55) is obtained, under **Assumptions A2–A3**, from Theorem A7.9.1 and applying [Bosq, 2000, Lemma 4.2]. ■

The following lemma, which contains some assertions from [Bosq, 2000, Corollary 4.3], provides information on the asymptotic properties of the empirical eigenvectors.

Lemma A7.9.1 *Assume that $\|X_0\|_H$ is bounded, and if $\{k_n\}$ is a sequence of integers such that*

$$\Lambda_{k_n} = o\left(\left(\frac{n}{\log n}\right)^{1/2}\right),$$

as $n \to \infty$, with

$$\Lambda_{k_n} = \sup_{1\leq j\leq k_n} (C_j - C_{j+1})^{-1}, \quad 1 \leq j \leq k_n, \tag{A7.57}$$

*then, under **Assumption A1**,*

$$\sup_{1\leq j\leq k_n} \|\phi'_{n,j} - \phi_{n,j}\|_H \to^{a.s.} 0, \quad n \to \infty,$$

being $\phi'_{n,j} = sgn\,\langle \phi_{n,j}, \phi_j\rangle_H\, \phi_j$, for each $i \in \mathbb{Z}$, $j \geq 1$ and $n \geq 2$, where $sgn\langle \phi_{n,j}, \phi_j\rangle_H = \mathbf{1}_{\langle \phi_{n,j},\phi_j\rangle_H \geq 0} - \mathbf{1}_{\langle \phi_{n,j},\phi_j\rangle_H < 0}$ and $\{\phi_{n,j},\ j \geq 1\}$ denoting the empirical eigenvalues of the empirical estimator C_n.

Let us now consider the following lemma to obtain the strong–consistency of $\{D_{n,j},\ j \geq 1\}$ (see Corollary A7.9.2 below).

Lemma A7.9.2 *Assume that $\|X_0\|_H$ is bounded, and if $\{k_n\}$ is a sequence of integers such that*

$$\Lambda_{k_n} = o\left(n^{1/4}(\ln(n))^{\beta-1/2}\right),$$

*as $n \to \infty$, where Λ_{k_n} is defined in equation (A7.57) under **Assumptions A1** and A3. The following limit then holds, for any $\beta > 1/2$,*

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{1\leq j\leq k_n} \|\phi'_{n,j} - \phi_{n,j}\|_H \to^{a.s.} 0, \quad n \to \infty, \tag{A7.58}$$

for any $\beta > 1/2$, where $\{\phi'_{n,j},\ j \geq 1\}$ denote the empirical eigenvalues of the empirical estimator C_n.

Proof. From [Bosq, 2000, Lemma 4.3], for any $n \geq 2$ and $1 \leq j \leq k_n$,

$$\left\|\phi'_{n,j} - \phi_{n,j}\right\|_H \leq a_j \left\|C_n - C\right\|_{\mathcal{L}(H)} \leq 2\sqrt{2}\Lambda_{k_n} \left\|C_n - C\right\|_{\mathcal{S}(H)},$$

which implies that

$$\mathcal{P}\left(\sup_{1\leq j\leq k_n} \|\phi'_{n,j} - \phi_{n,j}\|_H \geq \eta\right) \leq \mathcal{P}\left(\|C_n - C\|_{\mathcal{S}(H)} \geq \frac{\eta}{2\sqrt{2}\Lambda_{k_n}}\right).$$

Thus, since $\|X_0\|_H$ is bounded, from [Bosq, 2000, Theorem 4.2], and under **Assumption A3**, for any $\eta > 0$, and $\beta > 1/2$,

$$\begin{aligned}
&\mathcal{P}\left(\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{1\leq j\leq k_n} \|\phi'_{n,j} - \phi_{n,j}\|_H \geq \eta\right) \\
\leq\ &\mathcal{P}\left(\|C_n - C\|_{\mathcal{S}(H)} \geq \frac{\eta}{2\sqrt{2}\Lambda_{k_n}} \frac{(\ln(n))^{\beta}}{n^{1/4}}\right) \\
\leq\ &4\exp\left(-\frac{n\dfrac{\eta^2}{8\Lambda_{k_n}^2}\dfrac{(\ln(n))^{2\beta}}{n^{1/2}}}{\gamma_1 + \delta_1 \dfrac{\eta}{2\sqrt{2}\Lambda_{k_n}}\dfrac{(\ln(n))^{\beta}}{n^{1/4}}}\right) \\
=\ &\mathcal{O}\left(n^{-\frac{\eta^2}{\gamma_1 + \eta\delta_1\left(\frac{\ln(n)}{n}\right)^{1/2}}}\right), \quad n \to \infty.
\end{aligned} \tag{A7.59}$$

Thus, taking $\eta^2 > \gamma_1 + \delta_1\eta$, sequence (A7.59) is summable, and applying the Borel–Cantelli Lemma we arrive to the desired result.

■

Corollary A7.9.2 *Under the conditions of Lemma A7.9.2, considering now* ***Assumptions A1–A3****, for $\beta > \frac{1}{2}$, and n sufficiently large,*

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{j\geq 1} |D_{n,j} - D_j| \to^{a.s.} 0, \quad n \to \infty, \tag{A7.60}$$

where $\{D_{n,j},\ j \geq 1\}$ are defined as $D_{n,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,n,j}X_{i+1,n,j}$, for each $n \geq 2$ and $j \geq 1$.

Proof. From Theorem A7.9.1, under **Assumption A3**, there exists an n_0 such that for $n \geq n_0$, D_n is a Hilbert-Schmidt operator. Then, for $n \geq n_0$, for every $j \geq 1$, applying orthonormality of the empirical eigenvectors $\{\phi_{n,j},\ j \geq 1\}$, under **Assumptions A1–A2**,

$$\begin{aligned}
\frac{n^{1/4}}{(\ln(n))^{\beta}} |D_{n,j} - D_j| &= \frac{n^{1/4}}{(\ln(n))^{\beta}} |D_n(\phi_{n,j})(\phi_{n,j}) - D_n(\phi_{n,j})(\phi_j) + D_n(\phi_{n,j})(\phi_j) \\
&\quad - D(\phi_{n,j})(\phi_j) + D(\phi_{n,j})(\phi_j) - D(\phi_j)(\phi_j)| \\
&\leq \frac{n^{1/4}}{(\ln(n))^{\beta}} [\|D_n(\phi_{n,j})\|_H \|\phi_{n,j} - \phi_j\|_H \\
&\quad + \|(D_n - D)(\phi_{n,j})\|_H \|\phi_j\|_H + \|D(\phi_{n,j} - \phi_j)\|_H \|\phi_j\|_H] \\
&\leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \left[\|D_n\|_{\mathcal{L}(H)} \|\phi_{n,j} - \phi_j\|_H + \|D_n - D\|_{\mathcal{L}(H)} \right. \\
&\quad \left. + \|D\|_{\mathcal{L}(H)} \|\phi_{n,j} - \phi_j\|_H \right].
\end{aligned} \tag{A7.61}$$

From Theorem A7.9.1, under **Assumption A3**,

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \|D_n - D\|_{\mathcal{L}(H)} \leq \frac{n^{1/4}}{(\ln(n))^{\beta}} \|D_n - D\|_{\mathcal{S}(H)} \to^{a.s.} 0, \tag{A7.62}$$

and, for n sufficiently large, $\|D_n\|_{\mathcal{L}(H)} < \infty$. Furthermore, from Lemma A7.9.2 (see equation (A7.58)),

$$\frac{n^{1/4}}{(\ln(n))^{\beta}} \sup_{1\leq j\leq k_n} \|\phi_{n,j} - \phi_j\|_H \to^{a.s.} 0. \tag{A7.63}$$

Hence, from equations (A7.62)-(A7.63), taking the supremum in j at the left-hand side of equation (A7.61), we obtain equation (A7.60). ■

Remark A7.9.4 *Under* ***Assumption A1****, let us now consider the sequence $\{a_j,\ j \geq 1\}$ given by*

$$a_1 = 2\sqrt{2}\frac{1}{C_1 - C_2}, \quad a_j = 2\sqrt{2}\max\left(\frac{1}{C_{j-1} - C_j}, \frac{1}{C_j - C_{j+1}}\right),\ j \geq 2,$$

If $C_j > C_{j+1}$, when $1 \leq j \leq k_n$, hence $a_j > 0$ for any $1 \leq j \leq k_n$, and then $a_{k_n} < \sum_{j=1}^{k_n} a_j$, for a truncation parameter $\lim_{n\to\infty} k_n = \infty$, with $k_n < n$. Moreover, there exists an integer j_0 large enough such that, for any $j \geq j_0$, $a_j > 1$. In particular, if k_n is large enough,

$$\frac{1}{C_{k_n}} < \frac{1}{C_{k_n} - C_{k_n+1}} < a_{k_n} < \sum_{j=1}^{k_n} a_j, \quad \sum_{j=1}^{k_n} a_j > 1.$$

A7.9.3 One–sided upper a.s. asymptotic estimate of the $\mathcal{S}(H)$ norm of the error associated with $\widetilde{\rho}_{k_n}$

In this section, ρ does not admit a diagonal spectral decomposition in terms of the eigenvectors of C, being ρ not positive, nor trace operator, but it is a Hilbert-Schmidt operator. In this more general framework, an asymptotically almost surely one-sided upper estimate of the $\mathcal{S}(H)$ norm of the error associated with $\widehat{\rho}_{k_n}$ is derived. See Ruiz-Medina and Álvarez-Liébana [2018a], where sufficient conditions for the strong-consistency, in the trace norm, of the autocorrelation operator of an ARH(1) process, when it is a positive trace operator which does not admit a diagonal spectral decomposition, are provided.

Proposition A7.9.2 *Let us assume that ρ is a Hilbert-Schmidt, but not positive nor trace operator. Under **Assumption A5**, and conditions imposed in Lemma A7.9.2,*

$$\|\widetilde{\rho}_{k_n} - \rho\|^2_{\mathcal{S}(H)} \leq \|\rho\|^2_{\mathcal{S}(H)} - \sum_{j=1}^{\infty} \left(\rho\left(\phi_j\right)\left(\phi_j\right)\right)^2 = \sum_{j\neq k}^{\infty} \left(\frac{D\left(\phi_j\right)\left(\phi_k\right)}{C_j}\right)^2 < \infty.$$

In particular, for n sufficiently large,

$$\|\widetilde{\rho}_{k_n} - \rho\|^2_{\mathcal{S}(H)} \leq \sum_{j\neq k}^{\infty} \left[\frac{D_n(\phi_{n,j})(\phi_{n,k})}{C_{n,j}}\right]^2 \quad a.s.$$

Proof.

Let us consider the eigenvectors $\{\phi_{n,j},\ j \geq 1\}$ of C_n. Applying Parseval's identity, we obtain

$$\begin{aligned}
\|\widetilde{\rho}_{k_n} - \rho\|^2_{\mathcal{S}(H)} &= \|(\widetilde{\rho}_{k_n} - \rho)^*(\widetilde{\rho}_{k_n} - \rho)\|_1 \\
&= \sum_{j=1}^{\infty} \langle(\widetilde{\rho}_{k_n} - \rho)(\phi_{n,j}), (\widetilde{\rho}_{k_n} - \rho)(\phi_{n,j})\rangle_H \\
&= \sum_{j=1}^{\infty} \|(\widetilde{\rho}_{k_n} - \rho)(\phi_{n,j})\|^2_H
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\left[\langle\widetilde{\rho}_{k_n}(\phi_{n,j}),\phi_{n,k}\rangle_H - \langle\rho(\phi_{n,j}),\phi_{n,k}\rangle_H\right]^2 \\
&= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\left[\langle\widetilde{\rho}_{k_n}(\phi_{n,j}),\phi_{n,k}\rangle_H\right]^2 + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\left[\langle\rho(\phi_{n,j}),\phi_{n,k}\rangle_H\right]^2 \\
&- \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} 2\,\langle\widetilde{\rho}_{k_n}(\phi_{n,j}),\phi_{n,k}\rangle_H\,\langle\rho(\phi_{n,j}),\phi_{n,k}\rangle_H \\
&\leq \sum_{j=1}^{\infty}\sum_{k=1}^{k_n}\delta_{j,k}[D_{n,j}C_{n,j}^{-1}]^2 + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}[\left\langle DC^{-1}(\phi_{n,j}),\phi_{n,k}\right\rangle_H]^2 \\
&- 2\sum_{j=1}^{\infty}\sum_{k=1}^{k_n}\delta_{j,k}D_{n,j}C_{n,j}^{-1}\left\langle C^{-1}(\phi_{n,j}),D^{*}(\phi_{n,k})\right\rangle_H \\
&= \sum_{j=1}^{\infty}[D_{n,j}C_{n,j}^{-1}]^2 - 2D_{n,j}C_{n,j}^{-1}\left\langle DC^{-1}(\phi_{n,j}),\phi_{n,j}\right\rangle_H \\
&+ \sum_{j=1}^{\infty}\left[\left\langle DC^{-1}(\phi_{n,j}),\phi_{n,j}\right\rangle_H\right]^2 \\
&+ \sum_{j\neq k}^{\infty}\left\langle\left[DC^{-1}(\phi_{n,j}),\phi_{n,k}\right\rangle_H\right]^2 \\
&= \sum_{j=1}^{\infty}[D_{n,j}C_{n,j}^{-1} - DC^{-1}(\phi_{n,j})(\phi_{n,j})]^2 \\
&+ \sum_{j\neq k}^{\infty}\left[\left\langle DC^{-1}(\phi_{n,j}),\phi_{n,k}\right\rangle_H\right]^2,
\end{aligned}
\tag{A7.64}
$$

where $\delta_{j,k}$ denotes the Kronecker delta function, and $\|\cdot\|_1$ represents the trace operator norm. From Theorem A7.9.1, under **Assumption A3**,

$$
\begin{aligned}
\|D_nC_n^{-1} - DC^{-1}\|_{\mathcal{S}(H)} &= \|D_nC_n^{-1} - DC_n^{-1} \\
&+ DC_n^{-1} - DC^{-1}\|_{\mathcal{S}(H)} \\
&\leq \|D_nC_n^{-1} - DC_n^{-1}\|_{\mathcal{S}(H)} \\
&+ \|DC_n^{-1} - DC^{-1}\|_{\mathcal{S}(H)} = \|(D_n - D)C_n^{-1}\|_{\mathcal{S}(H)} \\
&+ \|D(C_n^{-1} - C^{-1})\|_{\mathcal{S}(H)},
\end{aligned}
$$

leading to

$$
\lim_{n\to\infty}\sum_{j=1}^{\infty}[D_{n,j}C_{n,j}^{-1} - DC^{-1}(\phi_{n,j})(\phi_{n,j})]^2 = 0 \quad a.s. \tag{A7.65}
$$

From equations (A7.64)–(A7.65), and from Lemma A7.9.2,

$$\begin{aligned}
\lim_{n\to\infty} \|\widetilde{\rho}_{k_n} - \rho\|^2_{\mathcal{S}(H)} &= \lim_{n\to\infty} \|(\widetilde{\rho}_{k_n} - \rho)^*(\widetilde{\rho}_{k_n} - \rho)\|_1 \\
&= \lim_{n\to\infty} \sum_{j\neq k}^{\infty} \left[\left\langle DC^{-1}(\phi_{n,j}), \phi_{n,k}\right\rangle_H\right]^2 \\
&= \lim_{n\to\infty} \sum_{j\neq k}^{\infty} \Big[DC^{-1}(\phi_{n,j})(\phi_{n,k}) - DC^{-1}(\phi_{n,j})(\phi_k) \\
&\quad +DC^{-1}(\phi_{n,j})(\phi_k) - \sum_{j\neq k}^{\infty} DC^{-1}(\phi_j)(\phi_k) \\
&\quad +DC^{-1}(\phi_j)(\phi_k)\Big]^2 \\
&\leq \lim_{n\to\infty} \sum_{j\neq k}^{\infty} \Big[\|DC^{-1}(\phi_{n,j})\|_H \|\phi_{n,k} - \phi_k\|_H \\
&\quad +\|DC^{-1}\|_{\mathcal{L}(H)} \|\phi_{n,j} - \phi_j\|_H + DC^{-1}(\phi_j)(\phi_k)\Big]^2 \\
&= \sum_{j\neq k}^{\infty} [DC^{-1}(\phi_j)(\phi_k)]^2 \leq \|\rho\|^2_{\mathcal{S}(H)} \quad \text{a.s}
\end{aligned}$$

Therefore, when ρ is not positive, nor trace operator, but it is Hilbert-Schmidt operator, the norm of the error associated with $\widetilde{\rho}_{k_n}$, in the space of Hilbert-Schmidt operators, is a.s. asymptotically upper bounded by the following quantity:

$$\|\rho\|^2_{\mathcal{S}(H)} - \sum_{j=1}^{\infty} [\rho(\phi_j)(\phi_j)]^2 = \sum_{j\neq k}^{\infty} \left[\frac{D(\phi_j)(\phi_k)}{C_j}\right]^2 < \infty. \tag{A7.66}$$

Equation (A7.66) can be approximated by the empirical quantity:

$$\sum_{j\neq k}^{\infty} \left[\frac{D_n(\phi_{n,j})(\phi_{n,k})}{C_{n,j}}\right]^2 .$$

Thus, for n sufficiently large,

$$\|\widetilde{\rho}_{k_n} - \rho\|^2_{\mathcal{S}(H)} \leq \sum_{j\neq k}^{\infty} \left[\frac{D_n(\phi_{n,j})(\phi_{n,k})}{C_{n,j}}\right]^2 \quad \text{a.s.}$$

■

A7.9.4 Simulation study: large-sample behaviour of the componentwise estimator of ρ, when eigenvectors of C are unknown

A brief simulation study is undertaken to illustrate the theoretical results on the strong-consistency of the formulated diagonal componentwise estimator of ρ, when $\{\phi_j,\ j \geq 1\}$ are unknown and a Gaussian diagonal data generation is achieved. An almost sure rate of convergence is fitted as well.

In all of the scenarios considered, the autocovariance operator C is defined as the inverse of a power of the Dirichlet negative Laplacian operator on $[0,\delta]$. Namely, the spectral decomposition of C is determined, in all of the scenarios, by

$$C(f)(g) = \sum_{j=1}^{\infty} C_j \langle \phi_j, f\rangle_H \langle \phi_j, g\rangle_H, \quad \phi_j\,(x) = \sqrt{\frac{2}{\delta}} \sin\left(\frac{\pi j x}{\delta}\right), \quad C_j = c_1 j^{-\beta_C},$$

for $f, g \in H = L^2((0,\delta))$ and $x \in (0,\delta)$, being c_1 a positive constant. In the remaining, we fix $(0,\delta) = (0,4)$. Concerning $\{C_j,\ j \geq 1\}$, different rates β_C will be regarded, such that **Assumption A1** is directly held; i.e., $\beta_C > 1$.

In a diagonal context, autocorrelation operator and covariance operator of the error term are approximated as follows, with $M = 50$:

$$\rho\,(X)\,(t) \simeq \sum_{j=1}^{M} \rho_{j,j} \langle \phi_j, X\rangle_H \phi_j(t), \quad C_{\varepsilon}\,(X)\,(t) \simeq \sum_{j=1}^{M} \sigma_{j,j}^2 \langle \phi_j, X\rangle_H \phi_j(t),$$

where

$$\rho_{j,j} = c_2 j^{-\delta_\rho}, \quad \sigma_{j,j}^2 = C_j\,(1 - \rho_{j,j})\,,$$

for any $j \geq 1$, being $\delta_\rho > 1/2$, and c_2 a constant which belongs to $(0,1)$. Thus, ρ is a diagonal self-adjoint Hilbert-Schmidt operator, with $\|\rho\|_{\mathcal{L}(H)} = \sup_{j\geq 1} |\rho_j| < 1$, under **Assumption A2**.

Simulations are then performed under **Assumptions A1–A4**, and the empirical version of the upper-bound derived in (A7.24) in the main paper is considered:

$$\begin{aligned} UB\,(k_n, l) \quad &= \quad \sup_{1\leq j\leq k_n} \left| \widetilde{\rho}_{n,j}^{l} - \frac{D_{n,j}^{l}}{C_j} \right| + \sup_{1\leq j\leq k_n} \left| \frac{D_{n,j}^{l}}{C_j} - \rho_j \right| \\ &+ \quad 2 \sum_{j=1}^{k_n} \frac{\left|D_{n,j}^{l}\right|}{C_j} \left\| \phi_{n,j}^{l} - \phi_{n,j}^{\prime,l} \right\|_H + \sup_{j>k_n} |\rho_j|\,, \end{aligned} \tag{A7.67}$$

being $k_n = \lceil \ln(n) \rceil$, in a manner that conditions imposed in Proposition A7.7.1 are held (see [Bosq, 2000, Example 8.6]. In equation (A7.67), superscript l denotes the estimator computed based on the lth generation of the values

$$\widetilde{X}_{i,n,j}^{l} = \langle X_i^l, \phi_{n,j}^l \rangle, \quad l = 1,\dots,N, \quad j = 1,\dots,k_n, \quad i = 0,\dots,n-1.$$

Here, $N = 500$ realizations have been generated, with shape parameters

$$\delta_C = 61/60,\ 3/2,\ 9/5, \quad \delta_\rho = 11/10.$$

Discretization step $\Delta t = 0.06$ has been adopted. For sample sizes $n_t = 35000 + 40000\,(t-1)$, for each $t = 1, \dots, 10$,

$$E\left(k_n, n_t, \beta\right) = \left(\sum_{l=1}^{N} \mathbf{1}_{\left(\xi_{n_t,\beta}, \infty\right)} \left(UB\left(k_n, l\right)\right) \right) / N, \quad \xi_{n_t,\beta} = \frac{\left(\ln(n_t)\right)^{\beta}}{n_t^{1/3}}, \tag{A7.68}$$

values are reflected in Table A7.9.1, in which the curve $\xi_{n_t,\beta}$ is fitted as the almost sure rate of convergence, with $\beta = 95/100$. In equation (A7.68), $\mathbf{1}_{\left(\xi_{n_t,\beta}, \infty\right)}$ denotes the indicator function over the interval $(\xi_{n_t,\beta}, \infty)$.

Table A7.9.1: $E\left(k_n, n_t, \beta\right)$ values defined in (A7.68), for $\beta = 95/100$ and $N = 500$ realizations, with $\delta_\rho = 11/10$, and $\delta_C = 61/60,\ 3/2,\ 9/5$, considering $n_t = 35000 + 40000(t-1)$, $t = 1, \dots, 10$, and $k_n = \lceil \ln(n) \rceil$.

n_t	k_n	$\delta_C = 61/60$	$\delta_C = 3/2$	$\delta_C = 9/5$
35000	10	$\frac{33}{500}$	$\frac{28}{500}$	$\frac{20}{500}$
75000	11	$\frac{19}{500}$	$\frac{15}{500}$	$\frac{11}{500}$
115000	11	$\frac{10}{500}$	$\frac{8}{500}$	$\frac{5}{500}$
155000	11	$\frac{4}{500}$	$\frac{3}{500}$	$\frac{2}{500}$
195000	12	$\frac{5}{500}$	$\frac{4}{500}$	$\frac{2}{500}$
235000	12	$\frac{3}{500}$	$\frac{1}{500}$	$\frac{1}{500}$
275000	12	$\frac{3}{500}$	0	0
315000	12	0	$\frac{1}{500}$	0
355000	12	$\frac{1}{500}$	0	0
395000	12	0	0	0

The convergence to zero of the empirical mean of

$$\|\widehat{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)}$$

is numerically illustrated in Figure A7.9.1 below, which displays the empirical mean of values $UB\left(k_n, l\right)$, against the curve $\xi_{n_t,\beta}$, for each $l = 1, \dots, N$ realizations, with $N = 500$, $\beta = 95/100$ and $k_n = \lceil \ln(n) \rceil$.

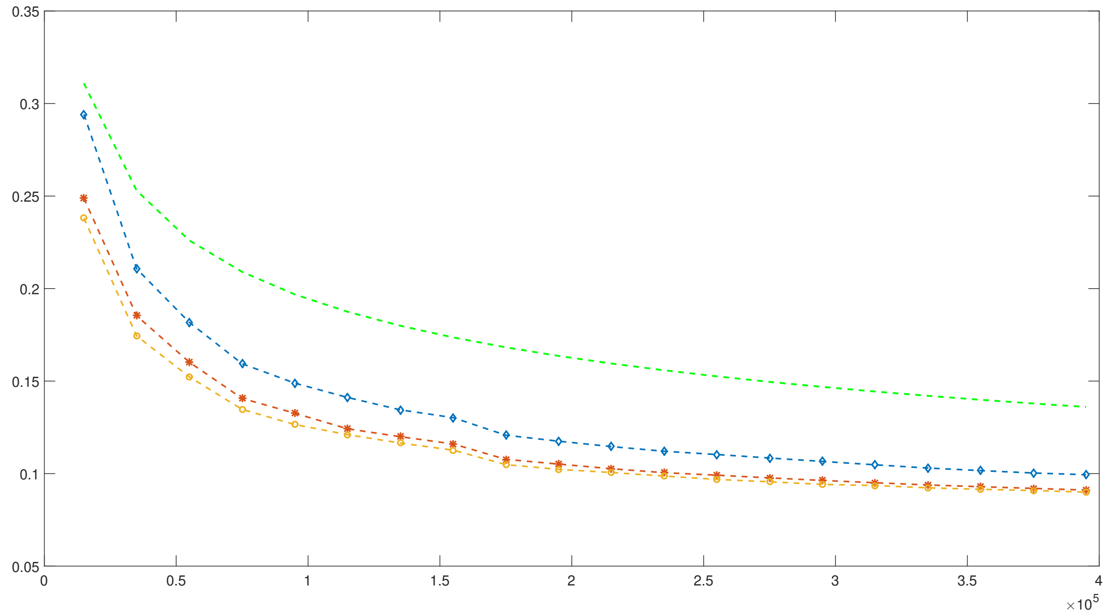


Figure A7.9.1: Empirical mean of $UB(k_n, l)$, for $l = 1, \ldots, N$, with $N = 500$, $\delta_\rho = 11/10$, $n_t = 15000 + 20000(t-1)$, $t = 1, \ldots, 20$, and $k_n = \lceil \ln(n) \rceil$. Shape parameters $\delta_C = 61/60,\ 3/2,\ 9/5$ are considered (blue diamond, red star and yellow circle dotted lines, respectively). The curve $\xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/3}}$, with $\beta = 95/100$, is also drawn (green dotted line).

A theoretical almost sure rate of convergence for the diagonal componentwise estimator $\widetilde{\rho}_{k_n}$ has not been derived in Proposition A7.7.1. However, when a diagonal data generation is performed, under different rates of convergence to zero of the eigenvalues of C, the curve

$$\xi_{n_t,95/100} = \frac{(\ln(n))^{95/100}}{n^{1/3}}$$

is numerically fitted, when large samples sizes are considered. As expected, for the largest shape parameter value δ_C, corresponding to the fastest decay velocity of the eigenvalues of the autocovariance operator, we obtain the fastest convergence to zero of $\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(H)}$. From results displayed in Figure A7.9.1, the empirical mean of the upper bound in (A7.67), computed from $N = 500$ realizations, is showed that can be upper bounded by the curve $\xi_{n_t,95/100} = \frac{(\ln(n))^{95/100}}{n^{1/3}}$, for the parameters adopted.

A7.9.5 Comparative study: numerical results

Tables A7.9.2-A7.9.4 and A7.9.5-A7.9.10 of this section reflect, in more detail, the numerical results obtained in the comparative study performed in Section A7.8. Details about the comparative study, and about which conditions are held under any scenario for each approach considered, can be found in the referred section. As remarked at the beginning of that section, the ARH(1) diagonal componentwise plug-in predictor already established will be only considered under diagonal scenarios. Strong-consistency results for the estimator $\widetilde{\rho}_{k_n}$, in the trace norm, when ρ is a positive and trace operator, which does not admit a diagonalization in terms of the eigenvectors of C, have recently been provided in Ruiz-Medina and Álvarez-Liébana [2018a].

The coefficients corresponding to the spectral decomposition of ρ and C_{ε}, related to the tensorial product $\{\phi_j \otimes \phi_h,\ j,h \geq 1\}$, are given by

$$\rho_{j,j} = c_2 j^{-\beta_\rho}, \quad \sigma^2_{j,j} = C_j \left(1 - \rho^2_{j,j}\right),$$

for each $j \geq 1$ and $\beta_\rho = 11/10$, and, for any $j \neq h,\ j,h \geq 1$,

$$\rho_{j,h} = \begin{cases} 0, & \textbf{scenario D} \\ e^{-|j-h|/W} & \textbf{scenario PD} \\ \frac{1}{K}\frac{1}{|j-h|^2+1} & \textbf{scenario ND} \end{cases}, \quad \sigma^2_{j,h} = \begin{cases} 0, & \textbf{scenario D} \\ e^{-|j-h|^2/W} & \textbf{scenario PD} \\ e^{-|j-h|^2/W} & \textbf{scenario ND} \end{cases}$$

for diagonal (D), pseudodiagonal (PD) and non-diagonal (ND) scenarios, being $\frac{1}{K} = 0.275$. As before, c_2 is a constant in $(0,1)$, verifying **Assumption A2**.

Thus, the error measure

$$F\left(k_n, n_t, \beta\right) = \left(\sum_{l=1}^{N} \mathbf{1}_{\left(\xi_{n_t,\beta},\infty\right)} \left(\left\|\left(\rho - \overline{\rho}^l_{k_n}\right)\left(X^l_{n-1}\right)\right\|^{k_n}_H\right)\right)/N, \tag{A7.69}$$

will be displayed (see Figures A7.8.1-A7.8.3 below), being $\mathbf{1}_{\left(\xi_{n_t,\beta},\infty\right)}$ the indicator function over the interval $(\xi_{n_t,\beta},\infty)$, where $\xi_{n_t,\beta}$ numerically fits the almost sure rate of convergence of

$$\left\|\left(\rho - \overline{\rho}^l_{k_n}\right)\left(X^l_{n-1}\right)\right\|^{k_n}_H.$$

The following diagonal subscenarios will be considered (see Figure A7.8.1), when the diagonal data generation is assumed, for

$$\delta_\rho = 11/10, \quad n_t = 35000 + 40000(t-1),\ t = 1,\dots,10, \quad \xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/2}},$$

with $\beta = 65/100$:

$$\delta_C = \begin{cases} 3/2 & \textbf{scenarios } D_1,\ D_3 \\ 24/10 & \textbf{scenarios } D_2,\ D_4 \end{cases}, \quad k_n = \begin{cases} \lceil \ln(n) \rceil & \textbf{scenarios } D_1,\ D_2 \\ \lceil e' n^{1/(8\delta_C+2)} \rceil & \textbf{scenarios } D_3,\ D_4 \end{cases},$$

being $e' = 17/10$. As discussed, conditions formulated in Bosq [2000] and Proposition A7.7.1 of the current paper are held for scenarios D_1-D_2, while in scenarios D_3-D_4, the conditions assumed in Proposition A7.7.1, Bosq [2000]; Guillas [2001] are verified. In the subscenarios D_1-D_4,

$$\left\|\left(\rho - \overline{\rho}_{k_n}^l\right)\left(X_{n-1}^l\right)\right\|_H^{k_n} = \sqrt{\int_a^b \left(\sum_{j=1}^{k_n} \rho_j X_{n-1,n,j}^l \phi_j(t) - \sum_{j=1}^{k_n} \overline{\rho}_{n,j}^l \left(X_{n-1}^l\right) \phi_{n,j}^l(t)\right)^2 dt}, \quad \text{(A7.70)}$$

is computed, being $\overline{\rho}_{k_n}^l \left(X_{n-1}^l\right)$ the predictors defined in (A7.19)-(A7.22), Bosq [2000]; Guillas [2001], respectively, for any $j = 1, \dots, k_n$, and based on the lth generation of the values $\widetilde{X}_{i,n,j}^l = \langle X_i^l, \phi_{n,j}^l \rangle_H$, for $l = 1, \dots, N$, with $N = 500$ simulations. See more details in Section A7.8.

Table A7.9.2: $F(k_n, n_t, \beta)$ values in (A7.69)-(A7.70), for scenarios $D_1 - D_4$. O.A. denotes the approach here detailed; B denotes the approach in Bosq [2000]; G denotes the approach in Guillas [2001].

		Scenario D_1			*Scenario* D_2			*Scenario* D_3				*Scenario* D_4			
n_t	k_n	O.A.	B	G	O.A.	B	G	k_n	O.A.	B	G	k_n	O.A.	B	G
35000	10	$\frac{11}{500}$	$\frac{68}{500}$	$\frac{70}{500}$	$\frac{4}{500}$	$\frac{24}{500}$	$\frac{28}{500}$	3	$\frac{13}{500}$	$\frac{12}{500}$	$\frac{10}{500}$	2	$\frac{7}{500}$	$\frac{7}{500}$	$\frac{4}{500}$
75000	11	$\frac{9}{500}$	$\frac{62}{500}$	$\frac{66}{500}$	$\frac{3}{500}$	$\frac{18}{500}$	$\frac{25}{500}$	3	$\frac{9}{500}$	$\frac{9}{500}$	$\frac{6}{500}$	2	$\frac{4}{500}$	$\frac{3}{500}$	$\frac{2}{500}$
115000	11	$\frac{6}{500}$	$\frac{59}{500}$	$\frac{62}{500}$	$\frac{3}{500}$	$\frac{16}{500}$	$\frac{22}{500}$	3	$\frac{6}{500}$	$\frac{5}{500}$	$\frac{5}{500}$	2	$\frac{3}{500}$	$\frac{3}{500}$	$\frac{2}{500}$
155000	11	$\frac{4}{500}$	$\frac{57}{500}$	$\frac{60}{500}$	$\frac{2}{500}$	$\frac{12}{500}$	$\frac{19}{500}$	3	$\frac{5}{500}$	$\frac{4}{500}$	$\frac{4}{500}$	2	$\frac{3}{500}$	$\frac{2}{500}$	$\frac{1}{500}$
195000	12	$\frac{6}{500}$	$\frac{60}{500}$	$\frac{64}{500}$	$\frac{4}{500}$	$\frac{15}{500}$	$\frac{21}{500}$	4	$\frac{6}{500}$	$\frac{4}{500}$	$\frac{3}{500}$	3	$\frac{4}{500}$	$\frac{2}{500}$	$\frac{1}{500}$
235000	12	$\frac{4}{500}$	$\frac{58}{500}$	$\frac{61}{500}$	0	$\frac{14}{500}$	$\frac{17}{500}$	4	$\frac{4}{500}$	$\frac{3}{500}$	$\frac{2}{500}$	3	$\frac{2}{500}$	$\frac{1}{500}$	$\frac{1}{500}$
275000	12	$\frac{3}{500}$	$\frac{51}{500}$	$\frac{58}{500}$	0	$\frac{13}{500}$	$\frac{16}{500}$	4	$\frac{3}{500}$	$\frac{2}{500}$	$\frac{1}{500}$	3	$\frac{2}{500}$	$\frac{1}{500}$	0
315000	12	$\frac{3}{500}$	$\frac{50}{500}$	$\frac{55}{500}$	$\frac{1}{500}$	$\frac{12}{500}$	$\frac{14}{500}$	4	$\frac{2}{500}$	$\frac{1}{500}$	$\frac{1}{500}$	3	$\frac{1}{500}$	0	0
355000	12	$\frac{2}{500}$	$\frac{47}{500}$	$\frac{53}{500}$	0	$\frac{12}{500}$	$\frac{13}{500}$	4	$\frac{2}{500}$	$\frac{1}{500}$	0	3	$\frac{1}{500}$	0	0
395000	12	$\frac{2}{500}$	$\frac{44}{500}$	$\frac{51}{500}$	0	$\frac{11}{500}$	$\frac{13}{500}$	4	$\frac{2}{500}$	0	0	3	$\frac{1}{500}$	0	0

Parameters δ_C and k_n for pseudodiagonal scenarios (scenarios PD_1-PD_4) and non-diagonal scenarios (scenarios ND_1-ND_4) are fixed as done above for scenarios D_1-D_4, being

$$\delta_2 = 11/10, \quad n_t = 35000 + 40000(t-1),\ t = 1, \dots, 10, \quad \xi_{n_t,\beta} = (\ln(n_t))^\beta n_t^{-1/3}.$$

Values of $\beta = 3/10$ and $\beta = 125/100$ are distinguished for pseudodiagonal and non-diagonal scenarios, respectively. Note that, as discussed above, different values of $\left\{\rho_{j,h},\ \sigma_{j,h}^2,\quad j,\ h \geq 1\right\}$ are adopted for these cases. In fact, under pseudodiagonal and non-diagonal frameworks, the following truncated norm is then computed, instead of (A7.70):

$$\sqrt{\int_a^b \left(\int_a^b \left(\sum_{j,k=1}^{k_n} \rho_{j,k}\phi_j(t)\phi_k(s)\right) ds - \sum_{j=1}^{k_n} \overline{\rho}_{n,j}^l \left(X_{n-1}^l\right) \phi_{n,j}^l(t)\right)^2 dt}. \quad \text{(A7.71)}$$

Remark that PD_1-PD_2 and ND_1-ND_2 scenarios verify conditions required in Bosq [2000], while scenarios PD_3-PD_4 and ND_3-ND_4 are included in both setting of conditions.

Table A7.9.3: $F(k_n, n_t, \beta)$ values in (A7.69) and (A7.71), for scenarios $PD_1 - PD_4$. B denotes the approach in Bosq [2000]; G denotes the approach in Guillas [2001].

	Scenario PD_1			*Scenario* PD_2			*Scenario* PD_3			*Scenario* PD_4		
n_t	k_n	B	G	k_n	B	G	k_n	B	G	k_n	B	G
35000	10	$\frac{32}{500}$	$\frac{33}{500}$	10	$\frac{25}{500}$	$\frac{29}{500}$	3	$\frac{28}{500}$	$\frac{26}{500}$	2	$\frac{27}{500}$	$\frac{24}{500}$
75000	11	$\frac{29}{500}$	$\frac{31}{500}$	11	$\frac{21}{500}$	$\frac{23}{500}$	3	$\frac{26}{500}$	$\frac{24}{500}$	2	$\frac{22}{500}$	$\frac{19}{500}$
115000	11	$\frac{26}{500}$	$\frac{28}{500}$	11	$\frac{18}{500}$	$\frac{20}{500}$	3	$\frac{23}{500}$	$\frac{21}{500}$	2	$\frac{18}{500}$	$\frac{15}{500}$
155000	11	$\frac{24}{500}$	$\frac{26}{500}$	11	$\frac{14}{500}$	$\frac{17}{500}$	3	$\frac{19}{500}$	$\frac{17}{500}$	2	$\frac{16}{500}$	$\frac{12}{500}$
195000	12	$\frac{19}{500}$	$\frac{21}{500}$	12	$\frac{10}{500}$	$\frac{13}{500}$	4	$\frac{14}{500}$	$\frac{12}{500}$	3	$\frac{11}{500}$	$\frac{9}{500}$
235000	12	$\frac{16}{500}$	$\frac{16}{500}$	12	$\frac{12}{500}$	$\frac{14}{500}$	4	$\frac{15}{500}$	$\frac{10}{500}$	3	$\frac{13}{500}$	$\frac{10}{500}$
275000	12	$\frac{12}{500}$	$\frac{13}{500}$	10	$\frac{8}{500}$	$\frac{10}{500}$	4	$\frac{9}{500}$	$\frac{7}{500}$	3	$\frac{7}{500}$	$\frac{6}{500}$
315000	12	$\frac{9}{500}$	$\frac{15}{500}$	12	$\frac{5}{500}$	$\frac{7}{500}$	4	$\frac{5}{500}$	$\frac{4}{500}$	3	$\frac{4}{500}$	$\frac{3}{500}$
355000	12	$\frac{8}{500}$	$\frac{11}{500}$	12	$\frac{3}{500}$	$\frac{5}{500}$	4	$\frac{3}{500}$	$\frac{3}{500}$	3	$\frac{2}{500}$	$\frac{2}{500}$
395000	12	$\frac{6}{500}$	$\frac{9}{500}$	12	$\frac{3}{500}$	$\frac{5}{500}$	4	$\frac{2}{500}$	$\frac{1}{500}$	3	$\frac{1}{500}$	0

Table A7.9.4: $F(k_n, n_t, \beta)$ values in (A7.69) and (A7.71), for scenarios $ND_1 - ND_4$. B denotes the approach in Bosq [2000]; G denotes the approach in Guillas [2001].

	Scenario ND_1			*Scenario* ND_2			*Scenario* ND_3			*Scenario* ND_4		
n_t	k_n	B	G	k_n	B	G	k_n	B	G	k_n	B	G
35000	10	$\frac{67}{500}$	$\frac{71}{500}$	10	$\frac{59}{500}$	$\frac{62}{500}$	3	$\frac{55}{500}$	$\frac{47}{500}$	2	$\frac{44}{500}$	$\frac{40}{500}$
75000	11	$\frac{44}{500}$	$\frac{50}{500}$	11	$\frac{38}{500}$	$\frac{45}{500}$	3	$\frac{36}{500}$	$\frac{31}{500}$	2	$\frac{34}{500}$	$\frac{30}{500}$
115000	11	$\frac{47}{500}$	$\frac{52}{500}$	11	$\frac{32}{500}$	$\frac{40}{500}$	3	$\frac{30}{500}$	$\frac{21}{500}$	2	$\frac{27}{500}$	$\frac{20}{500}$
155000	11	$\frac{51}{500}$	$\frac{55}{500}$	11	$\frac{27}{500}$	$\frac{34}{500}$	3	$\frac{27}{500}$	$\frac{25}{500}$	2	$\frac{23}{500}$	$\frac{17}{500}$
195000	12	$\frac{39}{500}$	$\frac{44}{500}$	12	$\frac{22}{500}$	$\frac{29}{500}$	4	$\frac{21}{500}$	$\frac{14}{500}$	3	$\frac{16}{500}$	$\frac{13}{500}$
235000	12	$\frac{40}{500}$	$\frac{42}{500}$	12	$\frac{29}{500}$	$\frac{33}{500}$	4	$\frac{18}{500}$	$\frac{16}{500}$	3	$\frac{12}{500}$	$\frac{9}{500}$
275000	12	$\frac{35}{500}$	$\frac{37}{500}$	12	$\frac{24}{500}$	$\frac{28}{500}$	4	$\frac{19}{500}$	$\frac{13}{500}$	3	$\frac{9}{500}$	$\frac{5}{500}$
315000	12	$\frac{24}{500}$	$\frac{28}{500}$	12	$\frac{17}{500}$	$\frac{19}{500}$	4	$\frac{11}{500}$	$\frac{8}{500}$	3	$\frac{6}{500}$	$\frac{3}{500}$
355000	12	$\frac{21}{500}$	$\frac{25}{500}$	12	$\frac{12}{500}$	$\frac{15}{500}$	4	$\frac{7}{500}$	$\frac{4}{500}$	3	$\frac{5}{500}$	$\frac{2}{500}$
395000	12	$\frac{18}{500}$	$\frac{21}{500}$	12	$\frac{9}{500}$	$\frac{12}{500}$	4	$\frac{6}{500}$	$\frac{3}{500}$	3	$\frac{4}{500}$	$\frac{2}{500}$

When approaches formulated in Antoniadis and Sapatinas [2003]; Besse et al. [2000] are included, smaller sample sizes must be considered due to computational limitations. Hence, a small-sample comparative study is shown in Tables A7.9.5-A7.9.10. When the referred methodologies in Besse et al. [2000] are implemented, the following alternative norm replaces the norm reflected in (A7.70)-(A7.71), respectively, for values $F(k_n, n_t, \beta)$:

$$\left\| \left(\rho - \overline{\rho}_{k_n}^{l}\right) \left(X_{n-1}^{l}\right) \right\|_H = \sqrt{\int_a^b \left(\rho\left(X_{n-1}^{l}\right)(t) - \overline{\rho}_{k_n}^{l}\left(X_{n-1}^{l}\right)(t)\right)^2 dt}, \quad l = 1, \dots, N. \tag{A7.72}$$

In this small-sample size context, the following diagonal subscenarios will be considered, when the diagonal data generation is assumed, for $\delta_\rho = 11/10$, $n_t = 750 + 500(t-1)$, $t = 1, \dots, 13$, and $\xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/2}}$, with $\beta = 65/100$:

$$\delta_C = \begin{cases} 3/2 & \textbf{scenarios } D_5,\ D_7 \\ 24/10 & \textbf{scenarios } D_6,\ D_8 \end{cases}, \quad k_n = \begin{cases} \lceil \ln(n) \rceil & \textbf{scenarios } D_5,\ D_6 \\ \lceil n^{1/\alpha} \rceil,\ \alpha = 6.5 & \textbf{scenarios } D_7,\ D_8 \end{cases},$$

being $q = 10$ the dimension of the subspace H_q involved in the penalized estimation proposed in Besse et al. [2000]. Remark that, since approaches formulated in Besse et al. [2000] not depend on the truncation parameter k_n adopted, we only perform them for scenarios D_5-D_6, where conditions imposed in that paper are verified. In the case of kernel-based predictor is used, two bandwidths $h_n = 0.15,\ 0.25$ are considered in both scenarios. Conditions formulated in Bosq [2000] and Proposition A7.7.1 of the current paper are held for all scenarios, while the conditions assumed in Antoniadis and Sapatinas [2003]; Guillas [2001] are only verified under scenarios D_7-D_8.

The same values of δ_C and k_n are adopted when pseudodiagonal scenarios (scenarios PD_5-PD_8) and non-diagonal scenarios (scenarios ND_5-ND_8) are analysed. As before, the curve $\xi_{n_t,\beta} = \frac{(\ln(n_t))^\beta}{n_t^{1/3}}$ is regarded, for pseudodiagonal and non-diagonal scenarios, with $\beta = 3/10$ and $\beta = 125/100$, respectively. While conditions in Bosq [2000] are verified for all scenarios, scenarios developed by Antoniadis and Sapatinas [2003]; Guillas [2001] are only held when the truncation parameter proposed in Antoniadis and Sapatinas [2003] is adopted. When smaller sample sizes are adopted, and approaches formulated in Antoniadis and Sapatinas [2003]; Besse et al. [2000] are included in the comparative study, new scenarios have been considered. Note that even when small sample sizes are studied, a good performance of the ARH(1) plug-in predictor given in equations (A7.19)-(A7.22) is observed. As well as the regularized wavelet-based approach detailed in Antoniadis and Sapatinas [2003] becomes the best methodology for small sample sizes, in comparision with the componentwise techniques above mentioned. Note that the good performance observed corresponds to the truncation rule proposed by these authors, with a small number of terms. While, when a larger number of terms is considered, according to the alternative truncation rules tested, the observed outperformance does not hold. While the penalized prediction approach proposed in Besse et al. [2000] has been shown as the more accurate, is, however, less affected by the regularity conditions imposed on the autocovariance kernel. Furthermore, a drawback of both approaches in Antoniadis and Sapatinas [2003]; Besse et al. [2000] is that they require large computational times. The underlying dependence structure cannot be provided in those approaches.

Table A7.9.5: $F(k_n, n_t, \beta)$ values in (A7.69)-(A7.70), for scenarios 13-16. O.A., B, G and AS denote the approaches in Antoniadis and Sapatinas [2003]; Bosq [2000]; Guillas [2001], respectively.

		Scenario 13				*Scenario 14*					*Scenario 15*					*Scenario 16*			
n_t	k_n	O.A.	B	G	AS	O.A.	B	G	AS	k_n	O.A.	B	G	AS	k_n	O.A.	B	G	AS
750	6	$\frac{42}{500}$	$\frac{83}{500}$	$\frac{90}{500}$	$\frac{84}{500}$	$\frac{31}{500}$	$\frac{76}{500}$	$\frac{79}{500}$	$\frac{72}{500}$	2	$\frac{30}{500}$	$\frac{33}{500}$	$\frac{34}{500}$	$\frac{21}{500}$	1	$\frac{19}{500}$	$\frac{24}{500}$	$\frac{24}{500}$	$\frac{14}{500}$
1250	7	$\frac{28}{500}$	$\frac{74}{500}$	$\frac{88}{500}$	$\frac{76}{500}$	$\frac{29}{500}$	$\frac{74}{500}$	$\frac{78}{500}$	$\frac{76}{500}$	2	$\frac{27}{500}$	$\frac{29}{500}$	$\frac{30}{500}$	$\frac{20}{500}$	2	$\frac{17}{500}$	$\frac{21}{500}$	$\frac{22}{500}$	$\frac{13}{500}$
1750	7	$\frac{27}{500}$	$\frac{70}{500}$	$\frac{84}{500}$	$\frac{75}{500}$	$\frac{28}{500}$	$\frac{71}{500}$	$\frac{73}{500}$	$\frac{70}{500}$	2	$\frac{25}{500}$	$\frac{25}{500}$	$\frac{27}{500}$	$\frac{17}{500}$	2	$\frac{16}{500}$	$\frac{19}{500}$	$\frac{21}{500}$	$\frac{11}{500}$
2250	7	$\frac{26}{500}$	$\frac{66}{500}$	$\frac{81}{500}$	$\frac{71}{500}$	$\frac{25}{500}$	$\frac{68}{500}$	$\frac{67}{500}$	$\frac{65}{500}$	3	$\frac{24}{500}$	$\frac{23}{500}$	$\frac{26}{500}$	$\frac{15}{500}$	2	$\frac{13}{500}$	$\frac{17}{500}$	$\frac{20}{500}$	$\frac{9}{500}$
2750	7	$\frac{28}{500}$	$\frac{68}{500}$	$\frac{82}{500}$	$\frac{70}{500}$	$\frac{24}{500}$	$\frac{63}{500}$	$\frac{62}{500}$	$\frac{59}{500}$	3	$\frac{21}{500}$	$\frac{21}{500}$	$\frac{26}{500}$	$\frac{12}{500}$	2	$\frac{12}{500}$	$\frac{15}{500}$	$\frac{18}{500}$	$\frac{8}{500}$
3250	8	$\frac{25}{500}$	$\frac{66}{500}$	$\frac{76}{500}$	$\frac{72}{500}$	$\frac{21}{500}$	$\frac{58}{500}$	$\frac{59}{500}$	$\frac{55}{500}$	3	$\frac{20}{500}$	$\frac{20}{500}$	$\frac{25}{500}$	$\frac{10}{500}$	2	$\frac{10}{500}$	$\frac{14}{500}$	$\frac{17}{500}$	$\frac{7}{500}$
3750	8	$\frac{23}{500}$	$\frac{60}{500}$	$\frac{72}{500}$	$\frac{72}{500}$	$\frac{21}{500}$	$\frac{53}{500}$	$\frac{54}{500}$	$\frac{54}{500}$	3	$\frac{17}{500}$	$\frac{18}{500}$	$\frac{24}{500}$	$\frac{9}{500}$	2	$\frac{9}{500}$	$\frac{11}{500}$	$\frac{14}{500}$	$\frac{7}{500}$
4250	8	$\frac{23}{500}$	$\frac{59}{500}$	$\frac{70}{500}$	$\frac{71}{500}$	$\frac{20}{500}$	$\frac{49}{500}$	$\frac{51}{500}$	$\frac{48}{500}$	3	$\frac{14}{500}$	$\frac{16}{500}$	$\frac{18}{500}$	$\frac{9}{500}$	2	$\frac{8}{500}$	$\frac{10}{500}$	$\frac{11}{500}$	$\frac{6}{500}$
4750	8	$\frac{21}{500}$	$\frac{56}{500}$	$\frac{67}{500}$	$\frac{69}{500}$	$\frac{18}{500}$	$\frac{47}{500}$	$\frac{49}{500}$	$\frac{45}{500}$	3	$\frac{13}{500}$	$\frac{13}{500}$	$\frac{15}{500}$	$\frac{8}{500}$	2	$\frac{7}{500}$	$\frac{8}{500}$	$\frac{9}{500}$	$\frac{5}{500}$
5250	8	$\frac{18}{500}$	$\frac{55}{500}$	$\frac{65}{500}$	$\frac{68}{500}$	$\frac{15}{500}$	$\frac{47}{500}$	$\frac{48}{500}$	$\frac{44}{500}$	3	$\frac{12}{500}$	$\frac{10}{500}$	$\frac{13}{500}$	$\frac{8}{500}$	2	$\frac{7}{500}$	$\frac{7}{500}$	$\frac{7}{500}$	$\frac{3}{500}$
5750	8	$\frac{20}{500}$	$\frac{58}{500}$	$\frac{66}{500}$	$\frac{68}{500}$	$\frac{16}{500}$	$\frac{45}{500}$	$\frac{50}{500}$	$\frac{47}{500}$	3	$\frac{11}{500}$	$\frac{9}{500}$	$\frac{11}{500}$	$\frac{7}{500}$	2	$\frac{6}{500}$	$\frac{7}{500}$	$\frac{6}{500}$	$\frac{2}{500}$
6250	8	$\frac{16}{500}$	$\frac{57}{500}$	$\frac{62}{500}$	$\frac{67}{500}$	$\frac{11}{500}$	$\frac{42}{500}$	$\frac{47}{500}$	$\frac{52}{500}$	3	$\frac{9}{500}$	$\frac{8}{500}$	$\frac{10}{500}$	$\frac{7}{500}$	2	$\frac{5}{500}$	$\frac{5}{500}$	$\frac{5}{500}$	$\frac{2}{500}$
6750	8	$\frac{14}{500}$	$\frac{54}{500}$	$\frac{59}{500}$	$\frac{67}{500}$	$\frac{9}{500}$	$\frac{41}{500}$	$\frac{45}{500}$	$\frac{42}{500}$	3	$\frac{7}{500}$	$\frac{8}{500}$	$\frac{8}{500}$	$\frac{6}{500}$	2	$\frac{3}{500}$	$\frac{4}{500}$	$\frac{4}{500}$	0

Table A7.9.6: $F\left(k_n, n_t, \beta\right)$ values in (A7.69) and (A7.72), for scenarios 13-14. $B_{0.15}$ and $B_{0.25}$ denotes the kernel-based approach in Besse et al. [2000], for $h_n = 0.15,\ 0.25$, respectively. B_q denotes its penalized prediction approach.

	Scenario 13			*Scenario 14*		
n_t	$B_{0.15}$	$B_{0.25}$	B_q	$B_{0.15}$	$B_{0.25}$	B_q
750	$\frac{85}{500}$	$\frac{88}{500}$	$\frac{6}{500}$	$\frac{76}{500}$	$\frac{80}{500}$	$\frac{3}{500}$
1250	$\frac{80}{500}$	$\frac{79}{500}$	$\frac{6}{500}$	$\frac{75}{500}$	$\frac{73}{500}$	$\frac{3}{500}$
1750	$\frac{76}{500}$	$\frac{71}{500}$	$\frac{5}{500}$	$\frac{73}{500}$	$\frac{67}{500}$	$\frac{2}{500}$
2250	$\frac{78}{500}$	$\frac{60}{500}$	$\frac{4}{500}$	$\frac{72}{500}$	$\frac{57}{500}$	$\frac{3}{500}$
2750	$\frac{73}{500}$	$\frac{57}{500}$	$\frac{4}{500}$	$\frac{70}{500}$	$\frac{53}{500}$	$\frac{3}{500}$
3250	$\frac{75}{500}$	$\frac{53}{500}$	$\frac{2}{500}$	$\frac{67}{500}$	$\frac{51}{500}$	$\frac{2}{500}$
3750	$\frac{70}{500}$	$\frac{49}{500}$	$\frac{2}{500}$	$\frac{67}{500}$	$\frac{43}{500}$	$\frac{1}{500}$
4250	$\frac{72}{500}$	$\frac{44}{500}$	$\frac{1}{500}$	$\frac{65}{500}$	$\frac{41}{500}$	0
4750	$\frac{68}{500}$	$\frac{39}{500}$	$\frac{3}{500}$	$\frac{63}{500}$	$\frac{38}{500}$	$\frac{1}{500}$
5250	$\frac{65}{500}$	$\frac{496}{500}$	$\frac{3}{500}$	$\frac{62}{500}$	$\frac{33}{500}$	$\frac{2}{500}$
5750	$\frac{62}{500}$	$\frac{34}{500}$	$\frac{2}{500}$	$\frac{60}{500}$	$\frac{31}{500}$	$\frac{2}{500}$
6250	$\frac{60}{500}$	$\frac{33}{500}$	$\frac{3}{500}$	$\frac{60}{500}$	$\frac{28}{500}$	$\frac{1}{500}$
6750	$\frac{59}{500}$	$\frac{33}{500}$	$\frac{3}{500}$	$\frac{57}{500}$	$\frac{24}{500}$	$\frac{1}{500}$

Table A7.9.7: $F\left(k_n, n_t, \beta\right)$ values in (A7.69) and (A7.71), for scenarios $PD_1 - PD_4$. B and G denote the approaches in Bosq [2000]; Guillas [2001], respectively; AS denotes the approach in Antoniadis and Sapatinas [2003].

	Scenario PD_1				*Scenario* PD_2				*Scenario* PD_3				*Scenario* PD_4			
n_t	k_n	B	G	AS	k_n	B	G	AS	k_n	B	G	AS	k_n	B	G	AS
750	6	$\frac{135}{500}$	$\frac{146}{500}$	$\frac{176}{500}$	6	$\frac{123}{500}$	$\frac{129}{500}$	$\frac{180}{500}$	2	$\frac{62}{500}$	$\frac{48}{500}$	$\frac{41}{500}$	1	$\frac{50}{500}$	$\frac{39}{500}$	$\frac{38}{500}$
1250	7	$\frac{124}{500}$	$\frac{130}{500}$	$\frac{166}{500}$	7	$\frac{117}{500}$	$\frac{120}{500}$	$\frac{175}{500}$	2	$\frac{60}{500}$	$\frac{42}{500}$	$\frac{37}{500}$	2	$\frac{47}{500}$	$\frac{36}{500}$	$\frac{33}{500}$
1750	7	$\frac{113}{500}$	$\frac{122}{500}$	$\frac{159}{500}$	7	$\frac{104}{500}$	$\frac{110}{500}$	$\frac{168}{500}$	2	$\frac{53}{500}$	$\frac{36}{500}$	$\frac{34}{500}$	2	$\frac{41}{500}$	$\frac{30}{500}$	$\frac{31}{500}$
2250	7	$\frac{89}{500}$	$\frac{115}{500}$	$\frac{153}{500}$	7	$\frac{86}{500}$	$\frac{91}{500}$	$\frac{164}{500}$	3	$\frac{49}{500}$	$\frac{31}{500}$	$\frac{29}{500}$	2	$\frac{35}{500}$	$\frac{28}{500}$	$\frac{30}{500}$
2750	7	$\frac{80}{500}$	$\frac{100}{500}$	$\frac{133}{500}$	7	$\frac{76}{500}$	$\frac{83}{500}$	$\frac{149}{500}$	3	$\frac{44}{500}$	$\frac{28}{500}$	$\frac{27}{500}$	2	$\frac{32}{500}$	$\frac{28}{500}$	$\frac{28}{500}$
3250	8	$\frac{99}{500}$	$\frac{104}{500}$	$\frac{139}{500}$	8	$\frac{71}{500}$	$\frac{78}{500}$	$\frac{153}{500}$	3	$\frac{40}{500}$	$\frac{26}{500}$	$\frac{26}{500}$	2	$\frac{27}{500}$	$\frac{27}{500}$	$\frac{25}{500}$
3750	8	$\frac{67}{500}$	$\frac{78}{500}$	$\frac{136}{500}$	8	$\frac{62}{500}$	$\frac{67}{500}$	$\frac{142}{500}$	3	$\frac{35}{500}$	$\frac{24}{500}$	$\frac{25}{500}$	2	$\frac{24}{500}$	$\frac{26}{500}$	$\frac{23}{500}$
4250	8	$\frac{65}{500}$	$\frac{74}{500}$	$\frac{129}{500}$	8	$\frac{60}{500}$	$\frac{63}{500}$	$\frac{133}{500}$	3	$\frac{30}{500}$	$\frac{23}{500}$	$\frac{22}{500}$	2	$\frac{22}{500}$	$\frac{22}{500}$	$\frac{19}{500}$
4750	8	$\frac{61}{500}$	$\frac{63}{500}$	$\frac{127}{500}$	8	$\frac{55}{500}$	$\frac{60}{500}$	$\frac{126}{500}$	3	$\frac{28}{500}$	$\frac{19}{500}$	$\frac{20}{500}$	2	$\frac{20}{500}$	$\frac{16}{500}$	$\frac{13}{500}$
5250	8	$\frac{48}{500}$	$\frac{51}{500}$	$\frac{125}{500}$	8	$\frac{46}{500}$	$\frac{49}{500}$	$\frac{122}{500}$	3	$\frac{25}{500}$	$\frac{17}{500}$	$\frac{16}{500}$	2	$\frac{17}{500}$	$\frac{12}{500}$	$\frac{10}{500}$
5750	8	$\frac{4}{500}$	$\frac{49}{500}$	$\frac{122}{500}$	8	$\frac{39}{500}$	$\frac{42}{500}$	$\frac{113}{500}$	3	$\frac{20}{500}$	$\frac{14}{500}$	$\frac{13}{500}$	2	$\frac{15}{500}$	$\frac{7}{500}$	$\frac{5}{500}$
6250	8	$\frac{38}{500}$	$\frac{45}{500}$	$\frac{118}{500}$	8	$\frac{33}{500}$	$\frac{35}{500}$	$\frac{108}{500}$	3	$\frac{19}{500}$	$\frac{13}{500}$	$\frac{10}{500}$	2	$\frac{13}{500}$	$\frac{7}{500}$	$\frac{3}{500}$
6750	8	$\frac{36}{500}$	$\frac{40}{500}$	$\frac{114}{500}$	8	$\frac{29}{500}$	$\frac{31}{500}$	$\frac{101}{500}$	3	$\frac{13}{500}$	$\frac{12}{500}$	$\frac{9}{500}$	2	$\frac{10}{500}$	$\frac{8}{500}$	$\frac{3}{500}$

Table A7.9.8: $F(k_n, n_t, \beta)$ values in (A7.69) and (A7.72), for scenarios $PD_1 - PD_2$. $B_{1.2}$ and $B_{1.7}$ denotes the kernel-based approach in Besse et al. [2000], for $h_n = 1.2,\ 1.7$, respectively. B_q denotes its penalized prediction approach.

	Scenario PD_1			*Scenario* PD_2		
n_t	$B_{1.2}$	$B_{1.7}$	B_q	$B_{1.2}$	$B_{1.7}$	B_q
750	$\frac{174}{500}$	$\frac{233}{500}$	$\frac{18}{500}$	$\frac{167}{500}$	$\frac{180}{500}$	$\frac{10}{500}$
1250	$\frac{158}{500}$	$\frac{214}{500}$	$\frac{10}{500}$	$\frac{151}{500}$	$\frac{169}{500}$	$\frac{7}{500}$
1750	$\frac{149}{500}$	$\frac{199}{500}$	$\frac{9}{500}$	$\frac{133}{500}$	$\frac{155}{500}$	$\frac{6}{500}$
2250	$\frac{146}{500}$	$\frac{185}{500}$	$\frac{7}{500}$	$\frac{130}{500}$	$\frac{146}{500}$	$\frac{4}{500}$
2750	$\frac{131}{500}$	$\frac{190}{500}$	$\frac{6}{500}$	$\frac{127}{500}$	$\frac{140}{500}$	$\frac{3}{500}$
3250	$\frac{129}{500}$	$\frac{193}{500}$	$\frac{5}{500}$	$\frac{119}{500}$	$\frac{135}{500}$	$\frac{3}{500}$
3750	$\frac{125}{500}$	$\frac{162}{500}$	$\frac{6}{500}$	$\frac{115}{500}$	$\frac{130}{500}$	$\frac{4}{500}$
4250	$\frac{138}{500}$	$\frac{160}{500}$	$\frac{4}{500}$	$\frac{109}{500}$	$\frac{121}{500}$	$\frac{2}{500}$
4750	$\frac{133}{500}$	$\frac{162}{500}$	$\frac{2}{500}$	$\frac{108}{500}$	$\frac{117}{500}$	$\frac{2}{500}$
5250	$\frac{120}{500}$	$\frac{154}{500}$	$\frac{1}{500}$	$\frac{107}{500}$	$\frac{114}{500}$	$\frac{1}{500}$
5750	$\frac{118}{500}$	$\frac{156}{500}$	$\frac{2}{500}$	$\frac{104}{500}$	$\frac{111}{500}$	$\frac{1}{500}$
6250	$\frac{116}{500}$	$\frac{144}{500}$	$\frac{1}{500}$	$\frac{99}{500}$	$\frac{103}{500}$	0
6750	$\frac{111}{500}$	$\frac{135}{500}$	0	$\frac{94}{500}$	$\frac{100}{500}$	0

Table A7.9.9: $F(k_n, n_t, \beta)$ values in (A7.69) and (A7.71), for scenarios $ND_1 - ND_4$. B and G denote the approaches in Bosq [2000]; Guillas [2001], respectively; AS denotes the approach in Antoniadis and Sapatinas [2003].

	Scenario ND_1				*Scenario* ND_2				*Scenario* ND_3				*Scenario* ND_4			
n_t	k_n	B	G	AS	k_n	B	G	AS	k_n	B	G	AS	k_n	B	G	AS
750	6	$\frac{86}{500}$	$\frac{90}{500}$	$\frac{88}{500}$	6	$\frac{80}{500}$	$\frac{84}{500}$	$\frac{83}{500}$	2	$\frac{73}{500}$	$\frac{66}{500}$	$\frac{75}{500}$	1	$\frac{55}{500}$	$\frac{42}{500}$	$\frac{60}{500}$
1250	7	$\frac{81}{500}$	$\frac{84}{500}$	$\frac{86}{500}$	7	$\frac{78}{500}$	$\frac{81}{500}$	$\frac{85}{500}$	2	$\frac{69}{500}$	$\frac{64}{500}$	$\frac{71}{500}$	2	$\frac{48}{500}$	$\frac{39}{500}$	$\frac{51}{500}$
1750	7	$\frac{77}{500}$	$\frac{80}{500}$	$\frac{85}{500}$	7	$\frac{73}{500}$	$\frac{87}{500}$	$\frac{86}{500}$	2	$\frac{64}{500}$	$\frac{60}{500}$	$\frac{70}{500}$	2	$\frac{46}{500}$	$\frac{32}{500}$	$\frac{50}{500}$
2250	7	$\frac{73}{500}$	$\frac{77}{500}$	$\frac{86}{500}$	7	$\frac{68}{500}$	$\frac{72}{500}$	$\frac{84}{500}$	3	$\frac{59}{500}$	$\frac{56}{500}$	$\frac{63}{500}$	2	$\frac{41}{500}$	$\frac{31}{500}$	$\frac{46}{500}$
2750	7	$\frac{70}{500}$	$\frac{73}{500}$	$\frac{83}{500}$	7	$\frac{55}{500}$	$\frac{70}{500}$	$\frac{80}{500}$	3	$\frac{50}{500}$	$\frac{54}{500}$	$\frac{55}{500}$	2	$\frac{37}{500}$	$\frac{27}{500}$	$\frac{45}{500}$
3250	8	$\frac{65}{500}$	$\frac{68}{500}$	$\frac{82}{500}$	8	$\frac{47}{500}$	$\frac{60}{500}$	$\frac{78}{500}$	3	$\frac{47}{500}$	$\frac{50}{500}$	$\frac{51}{500}$	2	$\frac{35}{500}$	$\frac{25}{500}$	$\frac{41}{500}$
3750	8	$\frac{54}{500}$	$\frac{59}{500}$	$\frac{80}{500}$	8	$\frac{43}{500}$	$\frac{53}{500}$	$\frac{75}{500}$	3	$\frac{45}{500}$	$\frac{43}{500}$	$\frac{48}{500}$	2	$\frac{31}{500}$	$\frac{24}{500}$	$\frac{37}{500}$
4250	8	$\frac{51}{500}$	$\frac{57}{500}$	$\frac{77}{500}$	8	$\frac{39}{500}$	$\frac{46}{500}$	$\frac{72}{500}$	3	$\frac{42}{500}$	$\frac{38}{500}$	$\frac{40}{500}$	2	$\frac{27}{500}$	$\frac{21}{500}$	$\frac{35}{500}$
4750	8	$\frac{45}{500}$	$\frac{51}{500}$	$\frac{79}{500}$	8	$\frac{37}{500}$	$\frac{41}{500}$	$\frac{73}{500}$	3	$\frac{35}{500}$	$\frac{33}{500}$	$\frac{38}{500}$	2	$\frac{23}{500}$	$\frac{17}{500}$	$\frac{32}{500}$
5250	8	$\frac{40}{500}$	$\frac{49}{500}$	$\frac{73}{500}$	8	$\frac{33}{500}$	$\frac{36}{500}$	$\frac{72}{500}$	3	$\frac{37}{500}$	$\frac{35}{500}$	$\frac{41}{500}$	2	$\frac{24}{500}$	$\frac{19}{500}$	$\frac{34}{500}$
5750	8	$\frac{38}{500}$	$\frac{43}{500}$	$\frac{74}{500}$	8	$\frac{32}{500}$	$\frac{34}{500}$	$\frac{59}{500}$	3	$\frac{33}{500}$	$\frac{32}{500}$	$\frac{37}{500}$	2	$\frac{19}{500}$	$\frac{13}{500}$	$\frac{29}{500}$
6250	8	$\frac{34}{500}$	$\frac{37}{500}$	$\frac{70}{500}$	8	$\frac{27}{500}$	$\frac{30}{500}$	$\frac{69}{500}$	3	$\frac{30}{500}$	$\frac{30}{500}$	$\frac{36}{500}$	2	$\frac{16}{500}$	$\frac{10}{500}$	$\frac{25}{500}$
6750	8	$\frac{30}{500}$	$\frac{33}{500}$	$\frac{68}{500}$	8	$\frac{25}{500}$	$\frac{29}{500}$	$\frac{66}{500}$	3	$\frac{29}{500}$	$\frac{25}{500}$	$\frac{35}{500}$	2	$\frac{12}{500}$	$\frac{9}{500}$	$\frac{21}{500}$

Table A7.9.10: $F(k_n, n_t, \beta)$ values in (A7.69) and (A7.72), for scenarios $ND_1 - ND_2$. $B_{1.2}$ and $B_{1.7}$ denotes the kernel-based approach in Besse et al. [2000], for $h_n = 1.2,\ 1.7$, respectively. B_q denotes its penalized prediction approach.

	Scenario ND_1			*Scenario* ND_2		
n_t	$B_{1.2}$	$B_{1.7}$	B_q	$B_{1.2}$	$B_{1.7}$	B_q
750	$\frac{449}{500}$	$\frac{281}{500}$	$\frac{7}{500}$	$\frac{377}{500}$	$\frac{222}{500}$	$\frac{5}{500}$
1250	$\frac{434}{500}$	$\frac{225}{500}$	$\frac{5}{500}$	$\frac{355}{500}$	$\frac{209}{500}$	$\frac{4}{500}$
1750	$\frac{436}{500}$	$\frac{164}{500}$	$\frac{5}{500}$	$\frac{330}{500}$	$\frac{196}{500}$	$\frac{4}{500}$
2250	$\frac{426}{500}$	$\frac{142}{500}$	$\frac{4}{500}$	$\frac{309}{500}$	$\frac{162}{500}$	$\frac{3}{500}$
2750	$\frac{422}{500}$	$\frac{123}{500}$	$\frac{3}{500}$	$\frac{292}{500}$	$\frac{130}{500}$	$\frac{3}{500}$
3250	$\frac{417}{500}$	$\frac{105}{500}$	$\frac{3}{500}$	$\frac{281}{500}$	$\frac{107}{500}$	$\frac{2}{500}$
3750	$\frac{376}{500}$	$\frac{97}{500}$	$\frac{3}{500}$	$\frac{269}{500}$	$\frac{83}{500}$	$\frac{2}{500}$
4250	$\frac{358}{500}$	$\frac{80}{500}$	$\frac{2}{500}$	$\frac{252}{500}$	$\frac{72}{500}$	$\frac{1}{500}$
4750	$\frac{345}{500}$	$\frac{71}{500}$	$\frac{1}{500}$	$\frac{241}{500}$	$\frac{69}{500}$	0
5250	$\frac{313}{500}$	$\frac{61}{500}$	0	$\frac{230}{500}$	$\frac{56}{500}$	$\frac{1}{500}$
5750	$\frac{262}{500}$	$\frac{55}{500}$	$\frac{1}{500}$	$\frac{215}{500}$	$\frac{45}{500}$	$\frac{1}{500}$
6250	$\frac{240}{500}$	$\frac{52}{500}$	$\frac{1}{500}$	$\frac{203}{500}$	$\frac{37}{500}$	0
6750	$\frac{230}{500}$	$\frac{46}{500}$	0	$\frac{195}{500}$	$\frac{32}{500}$	0

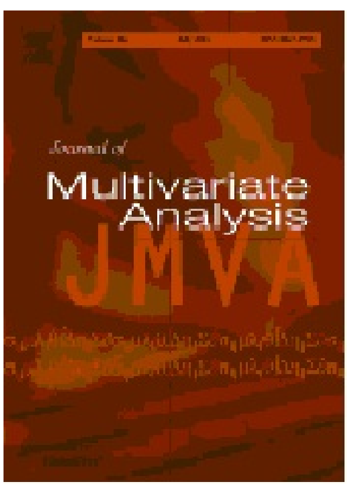
Acknowledgments

This work has been supported in part by project MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

A8

STRONGLY-CONSISTENT AUTOREGRESSIVE PREDICTORS IN ABSTRACT BANACH SPACES

RUIZ-MEDINA, M. D.; ÁLVAREZ-LIÉBANA, J.: Strong–consistent autoregressive predictors in abstract Banach spaces. J. Multivariate Anal. (under minor revision, 2018)



Year	Categ.	Cites	Impact Factor (5 years)	Quartil
2016	Statist. & Probab.	3938	1.229	Q3

ABSTRACT

This work derives new results on strongly–consistent estimation and prediction, for autoregressive processes of order one in a Banach separable space B (ARB(1) processes). The consistency results are obtained, in the norm of the space $\mathcal{L}(B)$ of bounded linear operators on B, for the componentwise estimator of the autocorrelation operator. The strong–consistency of the associated plug–in predictor then follows in the B–norm. A Gelfand triple is defined, involving the Hilbert space constructed in the Kuelbs's Lemma in Kuelbs [1970]. A nuclear embedding introduces the Reproducing Kernel Hilbert Space (RKHS), generated by the autocovariance operator, into the Hilbert space conforming the Rigged–Hilbert–Space structure. This paper extends Bosq [2000]; Labbas and Mourid [2002].

A8.1 INTRODUCTION

In the last few decades, there exists a growing interest on the statistical analysis of high–dimensional data, from the Functional Data Analysis (FDA) perspective. The book by Ramsay and Silverman [2005] provides an overview on FDA techniques, extended from the multivariate data context, or specifically formulated for the FDA framework. The monograph by Hsing and Eubank [2015] introduces functional analytical tools usually applied in the estimation of random elements in function spaces. The book by Horváth and Kokoszka [2012] is mainly concerned with inference based on second order statistics. A central topic in this book is the analysis of functional data, displaying dependent structures in time and space. The methodological survey paper by Cuevas [2014], on the state of the art in FDA, discusses central topics in FDA. Recent advances in the statistical analysis of high–dimensional data, from the parametric, semiparametric and nonparametric FDA frameworks, are collected in the Special Issue by Goia and Vieu [2016].

Linear time series models traditionally arise for processing temporal linear correlated data. In the FDA context, the monograph by Bosq [2000] introduces linear functional time series theory. The RKHS, generated by the autocovariance operator, plays a crucial role in the estimation approach presented in this monograph. In particular, the eigenvectors of the autocovariance operator are considered for projection (see also Álvarez-Liébana [2017]). Its empirical version is computed, when they are unknown. The resulting plug–in predictor is obtained as a linear functional of the observations, based on the empirical approximation of the autocorrelation operator. This approach exploits the Hilbert space structure, and its extension to the metric space context, and, in particular, to the Banach space context, requires to deriving a relationship (continuous embeddings) between the Banach space norm, and the RKHS norm, induced by the autocovariance operator, in contrast with the nonparametric regression approach for functional prediction (see, for instance, Ferraty et al. [2012], where asymptotic normality is derived). Specifically, in the nonparametric approach, a linear combination of the observed response values is usually considered. That is the case of the nonparametric local–weighting–based approach, involving weights defined from an isotropic kernel, depending on the metric or semi–metric of the space, where the regressors take their values (see, for example, Ferraty and Vieu [2006]; see also Ferraty et al. [2002], in the functional time series framework). The nonparametric approach is then more flexible regarding the structure of the space where the functional values of the regressors lie (usually a semi–metric space is considered). However, some computational drawbacks are present in its implementation, requiring the resolution of several selection problems. For instance, a choice of the smoothing parameter, and the kernel involved, in the definition of the weights, should be performed. Real–valued covariates were incorporated in the novel semiparametric kernel–based proposal by Aneiros-Pérez and Vieu [2008], involving an extension to the functional partial linear time series framework (see also Aneiros-Pérez and Vieu [2006]). Goia and Vieu [2015] also adopt a semi–parametric approach in their formulation of a two–terms Partitioned Functional Single Index Model. Geenens [2011] exploits the alternative provided by semi–metrics to avoid the curse of infinite dimensionality of some functional estimators.

On the other hand, in a parametric linear framework, Mas and Pumo [2010] introduced functional time series models in Banach spaces. In particular, strong mixing conditions and the absolute regularity of Banach–valued autoregressive processes have been studied in Allam and Mourid [2001]. Empirical estimators for Banach–valued autoregressive processes are studied in Bosq [2002], where, under some regularity conditions, and for the case of orthogonal innovations, the empirical mean is proved to be asymptotically optimal, with respect to almost surely (a.s.) convergence, and convergence of order two. The empirical autocovariance operator was also interpreted as a sample mean of an autoregressive process in a suitable space of linear operators. The extension of these results to the case of weakly dependent innovations is obtained in Dehling and Sharipov [2005]. A strongly–consistent sieve estimator of the autocorrelation operator of a Banach–valued autoregressive process is considered in Rachedi and Mourid [2003]. Limit theorems for a seasonality estimator, in the case of Banach autoregressive perturbations, are formulated in Mourid [2002]. Confidence regions for the periodic seasonality function, in the Banach space of continuous functions, is obtained as well. An approximation of Parzen's optimal predictor, in the RKHS framework, is applied in Mokhtari and Mourid [2003], for prediction of temporal stochastic process in Banach spaces. The existence and uniqueness of an almost surely strictly periodically correlated solution, to the first order autoregressive model in Banach spaces, is derived in Parvardeh et al. [2017]. Under some regularity conditions, limit results are obtained for AR$\mathcal{D}$(1) processes in Hajj [2011], where $\mathcal{D} = \mathcal{D}([0,1])$ denotes the Skorokhod space of right–continuous functions on $[0,1]$, having limit to the left at each $t \in [0,1]$. Conditions for the existence of strictly stationary solutions of ARMA equations in Banach spaces, with independent and identically distributed noise innovations, are derived in Spangenberg [2013].

In the derivation of strong–consistency results for ARB(1) componentwise estimators and predictors, Bosq [2000] restricts his attention to the case of the Banach space $\mathcal{C}([0,1])$ of continuous functions on $[0,1]$, with the supremum norm. Labbas and Mourid [2002] considers an ARB(1) context, for B being an arbitrary real separable Banach space, under the construction of a Hilbert space $\widetilde{H}$, where B is continuously embedded, as given in the Kuelbs's Lemma in [Kuelbs, 1970, Lemma 2.1]. Under the existence of a continuous extension to $\widetilde{H}$ of the autocorrelation operator $\rho \in \mathcal{L}(B)$, Labbas and Mourid [2002] obtain the strong-consistency of the formulated componentwise estimator of ρ, and of its associated plug–in predictor, in the norms of $\mathcal{L}(\widetilde{H})$, and $\widetilde{H}$, respectively.

functional data in nuclear spaces, arising, for example, in the observation of the solution to stochastic fractional and multifractional linear pseudodifferential equations (see, for example, Anh et al. [2016a,b]). The scales of Banach spaces constituted by fractional Sobolev and Besov spaces play a central role in the context of nuclear spaces. Continuous (nuclear) embeddings usually connect the elements of these scales (see, for example, Triebel [1983]). In this paper, a Rigged–Hilbert–Space structure is defined, involving the separable Hilbert space $\widetilde{H}$, appearing in the construction of the Kuelbs's Lemma in [Kuelbs, 1970, Lemma 2.1]. A key assumption, here, is the existence of a continuous (Hilbert–Schmidt) embedding introducing the RKHS, associated with the autocovariance operator of the ARB(1) process, into the Hilbert space generating the Gelfand triple, equipped with a finer topology than the B–topology. Under this scenario, strong–consistency results are derived, in the space $\mathcal{L}(B)$ of bounded linear operators on B, considering an abstract separable Banach space framework.

The outline of this paper is as follows. Notation and preliminaries are fixed in Appendix A8.2. Fundamental assumptions and some key lemmas are formulated in Appendix A8.3, and proved in Appendix A8.4. The main result of this paper on strong–consistency is derived in Appendix A8.5. Appendix A8.6 provides some examples. Final comments on our approach can be found in Appendix A8.7. The Supplementary Material provides in Appendix A8.8 illustrates numerically the results derived in Appendix A8.5, under the scenario described in Appendix A8.6, in a simulation study.

A8.2 PRELIMINARIES

Let $(B, \|\cdot\|_B)$ be a real separable Banach space, with the norm $\|\cdot\|_B$, and let $\mathcal{L}^2_B(\Omega, \mathcal{A}, \mathcal{P})$, the space of zero-mean B–valued random variables X such that

$$\sqrt{\int_B \|X\|^2_B d\mathcal{P}} < \infty.$$

Consider $X = \{X_n,\ n \in \mathbb{Z}\}$ to be a zero–mean B–valued stochastic process on the basic probability space $(\Omega, \mathcal{A}, \mathcal{P})$ satisfying (see Bosq [2000]):

$$X_n = \rho\left(X_{n-1}\right) + \varepsilon_n, \quad n \in \mathbb{Z}, \quad \rho \in \mathcal{L}(B), \tag{A8.1}$$

where ρ denotes the autocorrelation operator of X. In equation (A8.1), the B–valued innovation process $\varepsilon = \{\varepsilon_n,\ n \in \mathbb{Z}\}$ on $(\Omega, \mathcal{A}, \mathcal{P})$ is assumed to be strong white noise, uncorrelated with the random initial condition. Thus, ε is a zero–mean Banach–valued stationary process, with independent and identically distributed components, and with $\sigma^2_\varepsilon = \mathrm{E}\left\{\|\varepsilon_n\|^2_B\right\} < \infty$, for each $n \in \mathbb{Z}$. Assume that there exists an integer $j_0 \geq 1$ such that

$$\left\|\rho^{j_0}\right\|_{\mathcal{L}(B)} < 1. \tag{A8.2}$$

Then, equation (A8.1) admits an unique strictly stationary solution with $\sigma^2_X = \mathrm{E}\left\{\|X_n\|^2_B\right\} < \infty$; i.e., belonging to $\mathcal{L}^2_B(\Omega, \mathcal{A}, \mathcal{P})$, given by $X_n = \displaystyle\sum_{j=0}^{\infty} \rho^j\left(\varepsilon_{n-j}\right)$, for each $n \in \mathbb{Z}$ (see Bosq [2000]). Under (A8.2), the autocovariance operator C of an ARB(1) process X is defined from the autocovariance operator of $X_0 \in \mathcal{L}^2_B(\Omega, \mathcal{A}, \mathcal{P})$, as

$$C\left(x^*\right) = \mathrm{E}\left\{x^*(X_0)X_0\right\}, \quad x^* \in B^*.$$

The cross–covariance operator D is given by

$$D\left(x^*\right) = \mathrm{E}\left\{x^*(X_0)X_1\right\}, \quad x^* \in B^*.$$

Since C is assumed to be a nuclear operator, there exists a sequence $\{x_j,\ j \geq 1\} \subset B$ such that, for every $x^* \in B^*$ (see [Bosq, 2000, Eq. (6.24), p. 156]):

$$C(x^*) = \sum_{j=1}^{\infty} x^*(x_j)\, x_j, \quad \sum_{j=1}^{\infty} \|x_j\|_B^2 < \infty.$$

D is also assumed to be a nuclear operator. Then, there exist sequences $\{y_j,\ j \geq 1\} \subset B$ and $\{x_j^{**},\ j \geq 1\} \subset B^{**}$ such that, for every $x^* \in B^*$,

$$D(x^*) = \sum_{j=1}^{\infty} x_j^{**}(x^*) y_j, \quad \sum_{j=1}^{\infty} \left\|x_j^{**}\right\|_{B^{**}} \|y_j\| < \infty,$$

(see [Bosq, 2000, Eq. (6.23), p. 156]). Empirical estimators of C and D are respectively given by (see [Bosq, 2000, Eqs. (6.45) and (6.58), pp. 164–168]), for $n \geq 2$,

$$C_n(x^*) = \frac{1}{n} \sum_{i=0}^{n-1} x^*(X_i)(X_i), \quad D_n(x^*) = \frac{1}{n-1} \sum_{i=0}^{n-2} x^*(X_i)(X_{i+1}), \quad x^* \in B^*.$$

[Kuelbs, 1970, Lemma 2.1], now formulated, plays a key role in our approach.

Lemma A8.2.1 *If B is a real separable Banach space with norm $\|\cdot\|_B$, then, there exists an inner product $\langle\cdot,\cdot\rangle_{\widetilde{H}}$ on B such that the norm $\|\cdot\|_{\widetilde{H}}$, generated by $\langle\cdot,\cdot\rangle_{\widetilde{H}}$, is weaker than $\|\cdot\|_B$. The completion of B under the norm $\|\cdot\|_{\widetilde{H}}$ defines the Hilbert space $\widetilde{H}$, where B is continuously embedded.*

Denote by $\{x_n,\ n \in \mathbb{N}\} \subset B$, a dense sequence in B, and by $\{F_n,\ n \in \mathbb{N}\} \subset B^*$ a sequence of bounded linear functionals on B, satisfying

$$F_n(x_n) = \|x_n\|_B, \quad \|F_n\| = 1, \tag{A8.3}$$

such that

$$\|x\|_B = \sup_{n \in \mathbb{N}} |F_n(x)|, \quad x \in B. \tag{A8.4}$$

The inner product $\langle\cdot,\cdot\rangle_{\widetilde{H}}$, and its associated norm, in Lemma A8.2.1, is defined by

$$\begin{aligned} \langle x, y\rangle_{\widetilde{H}} &= \sum_{n=1}^{\infty} t_n F_n(x) F_n(y), \quad x, y \in \widetilde{H}, \\ \|x\|_{\widetilde{H}}^2 &= \sum_{n=1}^{\infty} t_n \{F_n(x)\}^2 \leq \|x\|_B^2, \quad x \in B, \end{aligned} \tag{A8.5}$$

where $\{t_n,\ n \in \mathbb{N}\}$ is a sequence of positive numbers such that $\sum_{n=1}^{\infty} t_n = 1$.

A8.3 MAIN ASSUMPTIONS AND PRELIMINARY RESULTS

In view of Lemma A8.2.1, for every $n \in \mathbb{Z}$, $X_n \in B \hookrightarrow \widetilde{H}$ satisfies a.s.

$$X_n \underset{\widetilde{H}}{=} \sum_{j=1}^{\infty} \langle X_n, v_j \rangle_{\widetilde{H}} v_j, \quad n \in \mathbb{Z},$$

for any orthonormal basis $\{v_j,\ j \geq 1\}$ of $\widetilde{H}$. The trace autocovariance operator

$$C = \mathrm{E}\left\{ \left(\sum_{j=1}^{\infty} \langle X_n, v_j \rangle_{\widetilde{H}} v_j \right) \otimes \left(\sum_{j=1}^{\infty} \langle X_n, v_j \rangle_{\widetilde{H}} v_j \right) \right\}$$

of the extended ARB(1) process is a trace operator in $\widetilde{H}$, admitting a diagonal spectral representation, in terms of its eigenvalues $\{C_j,\ j \geq 1\}$ and eigenvectors $\{\phi_j,\ j \geq 1\}$, that provide an orthonormal system in $\widetilde{H}$. Summarizing, in the subsequent developments, the following identities in $\widetilde{H}$ will be considered, for the extended version of ARB(1) process X. For each $f, h \in \widetilde{H}$,

$$C(f) \underset{\widetilde{H}}{=} \sum_{j=1}^{\infty} C_j \langle f, \phi_j \rangle_{\widetilde{H}} \phi_j \tag{A8.6}$$

$$D(h) \underset{\widetilde{H}}{=} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \langle D(\phi_j), \phi_k \rangle_{\widetilde{H}} \langle h, \phi_j \rangle_{\widetilde{H}} \phi_k$$

$$C_n(f) \underset{\widetilde{H}\ a.s.}{=} \sum_{j=1}^{n} C_{n,j} \langle f, \phi_{n,j} \rangle_{\widetilde{H}} \phi_{n,j} \tag{A8.7}$$

$$C_{n,j} \underset{a.s.}{=} \frac{1}{n} \sum_{i=0}^{n-1} X_{i,n,j}^2,\ X_{i,n,j} = \langle X_i, \phi_{n,j} \rangle_{\widetilde{H}},\ C_n(\phi_{n,j}) \underset{\widetilde{H}\ a.s.}{=} C_{n,j}\phi_{n,j}$$

$$D_n(h) \underset{\widetilde{H}\ a.s.}{=} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \langle D_n(\phi_{n,j}), \phi_{n,k} \rangle_{\widetilde{H}} \langle h, \phi_{n,j} \rangle_{\widetilde{H}} \phi_{n,k}, \tag{A8.8}$$

where, for $n \geq 2$, $\{\phi_{n,j},\ j \geq 1\}$ is a complete orthonormal system in $\widetilde{H}$, and

$$C_{n,1} \geq C_{n,2} \geq \cdots \geq C_{n,n} \geq 0 = C_{n,n+1} = C_{n,n+2} = \ldots.$$

The following assumption plays a crucial role in the derivation of the main results in this paper.

Assumption A1. $\|X_0\|_B$ is a.s. bounded, and the eigenspace V_j, associated with $C_j > 0$ in (A8.6) is one-dimensional for every $j \geq 1$.

Under **Assumption A1**, we can define the following quantities:

$$a_1 = 2\sqrt{2}\frac{1}{C_1 - C_2}, \quad a_j = 2\sqrt{2}\max\left(\frac{1}{C_{j-1} - C_j}, \frac{1}{C_j - C_{j+1}}\right), \quad j \geq 2. \tag{A8.9}$$

Remark A8.3.1 This assumption can be relaxed to considering multidimensional eigenspaces by redefining the quantities a_j, for each $j \geq 1$, as the quantities c_j, for each $j \geq 1$, given in [Bosq, 2000, Lemma 4.4].

Assumption A2. Let k_n such that

$$C_{n,k_n} > 0, \quad \text{(a.s.)} \quad k_n \to \infty, \quad \frac{k_n}{n} \to 0,\ n \to \infty.$$

Remark A8.3.2 Consider

$$\Lambda_{k_n} = \sup_{1 \leq j \leq k_n} (C_j - C_{j+1})^{-1}. \tag{A8.10}$$

For n sufficiently large,

$$k_n < C_{k_n}^{-1} < \frac{1}{C_{k_n} - C_{k_n+1}} < a_{k_n} < \Lambda_{k_n} < \sum_{j=1}^{k_n} a_j.$$

Assumption A3. The following limit holds:

$$\sup_{x \in B;\ \|x\|_B \leq 1} \left\| \rho(x) - \sum_{j=1}^{k} \langle \rho(x), \phi_j \rangle_{\widetilde{H}} \phi_j \right\|_B \to 0, \quad k \to \infty. \tag{A8.11}$$

Assumption A4. $\{C_j,\ j \geq 1\}$ are such that the inclusion of $\mathcal{H}(X)$ into $\widetilde{H}^*$ is continuous; i.e.,

$$\mathcal{H}(X) \hookrightarrow \widetilde{H}^*,$$

where $\hookrightarrow$ denotes, as usual, the continuous embedding, $\widetilde{H}^*$ the dual space of $\widetilde{H}$ and $\mathcal{H}(X)$ the Reproducing Kernel Hilbert Space associated with C.

Let us consider the closed subspace H of B with the norm induced by the inner product $\langle \cdot, \cdot \rangle_H$ defined as follows:

$$H = \left\{ x \in B;\ \sum_{n=1}^{\infty} \{F_n(x)\}^2 < \infty \right\}, \quad \langle f, g \rangle_H = \sum_{n=1}^{\infty} F_n(f) F_n(g), \quad f, g \in H. \tag{A8.12}$$

Then, H is continuously embedded into B, and the following remark provides the isometric isomorphism established by the Riesz Representation Theorem between the spaces $\widetilde{H}$ and its dual $\widetilde{H}^*$.

Remark A8.3.3 *Let $f^*, g^* \in \widetilde{H}^*$, and $f, g \in \widetilde{H}$, such that, for every $n \geq 1$, consider $F_n(f^*) = \sqrt{t_n} F_n(\widetilde{f})$, $F_n(g^*) = \sqrt{t_n} F_n(\widetilde{g})$, and $F_n(\widetilde{f}) = \sqrt{t_n} F_n(f)$, $F_n(\widetilde{g}) = \sqrt{t_n} F_n(g)$, for certain $\widetilde{f}, \widetilde{g} \in H$. Then, the following identities hold:*

$$\begin{aligned}\langle f^*, g^* \rangle_{\widetilde{H}^*} &= \sum_{n=1}^{\infty} \frac{1}{t_n} F_n(f^*) F_n(g^*) = \sum_{n=1}^{\infty} \frac{1}{t_n} \sqrt{t_n}\sqrt{t_n} F_n(\widetilde{f}) F_n(\widetilde{g}) = \left\langle \widetilde{f}, \widetilde{g} \right\rangle_H \\ &= \sum_{n=1}^{\infty} t_n F_n(f) F_n(g) = \langle f, g \rangle_{\widetilde{H}} .\end{aligned}$$

Lemma A8.3.1 *Under Assumption A4, the following continuous embeddings hold:*

$$\mathcal{H}(X) \hookrightarrow \widetilde{H}^* \hookrightarrow B^* \hookrightarrow H \hookrightarrow B \hookrightarrow \widetilde{H} \hookrightarrow [\mathcal{H}(X)]^*, \tag{A8.13}$$

where

$$\begin{aligned}\widetilde{H} &= \left\{ x \in B;\ \sum_{n=1}^{\infty} t_n \left\{F_n(x)\right\}^2 < \infty \right\}, \quad \langle f, g \rangle_{\widetilde{H}} = \sum_{n=1}^{\infty} t_n F_n(f) F_n(g),\ f, g \in \widetilde{H} \\ H &= \left\{ x \in B;\ \sum_{n=1}^{\infty} \left\{F_n(x)\right\}^2 < \infty \right\}, \quad \langle f, g \rangle_H = \sum_{n=1}^{\infty} F_n(f) F_n(g),\ f, g \in H \\ \widetilde{H}^* &= \left\{ x \in B;\ \sum_{n=1}^{\infty} \frac{1}{t_n} \left\{F_n(x)\right\}^2 < \infty \right\}, \quad \langle f, g \rangle_{\widetilde{H}^*} = \sum_{n=1}^{\infty} \frac{1}{t_n} F_n(f) F_n(g),\ f, g \in \widetilde{H}^* \\ \mathcal{H}(X) &= \left\{ x \in \widetilde{H};\ \left\langle C^{-1}(x), x \right\rangle_{\widetilde{H}} < \infty \right\}, \\ \langle f, g \rangle_{\mathcal{H}(X)} &= \left\langle C^{-1}(f), g \right\rangle_{\widetilde{H}},\ f, g \in C^{1/2}(\widetilde{H}) \\ [\mathcal{H}(X)]^* &= \left\{ x \in \widetilde{H};\ \langle C(x), x \rangle_{\widetilde{H}} < \infty \right\} \\ \langle f, g \rangle_{[\mathcal{H}(X)]^*} &= \langle C(f), g \rangle_{\widetilde{H}}\ f, g \in C^{-1/2}(\widetilde{H}).\end{aligned}$$

Proof. Let us consider the following inequalitites, for each $x \in B$,:

$$\begin{aligned}\|x\|_{\widetilde{H}} &= \sqrt{\sum_{j=1}^{\infty} t_n \left\{F_n(x)\right\}^2} \leq \|x\|_B = \sup_{n \geq 1} |F_n(x)|, \\ \|x\|_B &= \sup_{n \geq 1} |F_n(x)| \leq \sqrt{\sum_{n=1}^{\infty} \left\{F_n(x)\right\}^2} = \|x\|_H \leq \sum_{n=1}^{\infty} |F_n(x)| = \|x\|_{B^*}, \\ \|x\|_{B^*} &= \sum_{n=1}^{\infty} |F_n(x)| \leq \sqrt{\sum_{n=1}^{\infty} \frac{1}{t_n} \left\{F_n(x)\right\}^2} = \|x\|_{\widetilde{H}^*}.\end{aligned} \tag{A8.14}$$

Under **Assumption A4** (see also Remark A8.3.3), for every $f \in C^{1/2}(\widetilde{H}) = \mathcal{H}(X)$,

$$\|f\|_{\mathcal{H}(X)} = \sqrt{\langle C^{-1}(f), f\rangle_{\widetilde{H}}} \geq \|f\|_{\widetilde{H}^*} = \sqrt{\sum_{n=1}^{\infty} \frac{1}{t_n} \{F_n(x)\}^2}. \quad \text{(A8.15)}$$

From equations (A8.14)–(A8.15), the inclusions in (A8.13) are continuous.

■

It is well–known that $\{\phi_j,\ j \geq 1\}$ is also an orthogonal system in $\mathcal{H}(X)$. Futhermore, under **Assumption A4**, from Lemma A8.3.1,

$$\{\phi_j,\ j \geq 1\} \subset \mathcal{H}(X) \hookrightarrow \widetilde{H}^* \hookrightarrow B^* \hookrightarrow H.$$

Therefore, from equation (A8.12), for every $j \geq 1$,

$$\|\phi_j\|_H^2 = \sum_{m=1}^{\infty} \{F_m(\phi_j)\}^2 < \infty. \quad \text{(A8.16)}$$

The following assumption is now considered on the norm (A8.16):

Assumption A5. The continuous embedding $i_{\mathcal{H}(X),H} : \mathcal{H}(X) \hookrightarrow H$ belongs to the trace class. That is,

$$\sum_{j=1}^{\infty} \|\phi_j\|_H^2 < \infty.$$

Let $\{F_m,\ m \geq 1\}$ be defined as in Lemma A8.2.1. **Assumption A5** leads to

$$\sum_{j=1}^{\infty} \left\langle i_{\mathcal{H}(X),H}(\phi_j), \phi_j \right\rangle_H = \sum_{j=1}^{\infty} \sum_{m=1}^{\infty} \{F_m(\phi_j)\}^2 = \sum_{m=1}^{\infty} N_m < \infty, \quad \text{(A8.17)}$$

where, in particular, from equation (A8.17),

$$N_m = \sum_{j=1}^{\infty} \{F_m(\phi_j)\}^2 < \infty, \quad \sup_{m \geq 1} N_m = N < \infty \quad \text{(A8.18)}$$

$$V = \sup_{j \geq 1} \|\phi_j\|_B \leq \sum_{j=1}^{\infty} \sum_{m=1}^{\infty} \{F_m(\phi_j)\}^2 < \infty. \quad \text{(A8.19)}$$

The following preliminary results are considered from [Bosq, 2000, Theorem 4.1, pp. 98–99; Corollary 4.1, pp. 100–101; Theorem 4.8, pp. 116–117]).

Lemma A8.3.2 *Under* **Assumption A1***, the following identities hold, for any standard AR$\widetilde{H}(1)$ process (e.g., the extension to $\widetilde{H}$ of ARB(1) process X satisfying equation (A8.1)),*

$$\|C_n - C\|_{\mathcal{S}(\widetilde{H})} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) \, a.s., \quad \|D_n - D\|_{\mathcal{S}(\widetilde{H})} = \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{1/2}\right) \, a.s., \tag{A8.20}$$

where $\|\cdot\|_{\mathcal{S}(\widetilde{H})}$ is the norm in the Hilbert space $\mathcal{S}(\widetilde{H})$ of Hilbert–Schmidt operators on $\widetilde{H}$; i.e., the subspace of compact operators $\mathcal{A}$ such that

$$\sum_{j=1}^{\infty} \langle \mathcal{A}^* \mathcal{A}(\varphi_j), \varphi_j \rangle_{\widetilde{H}} < \infty,$$

for any orthonormal basis $\{\varphi_j, \ j \geq 1\}$ of $\widetilde{H}$.

Lemma A8.3.3 *Under* **Assumption A1***, let $\{C_j, \ j \geq 1\}$ and $\{C_{n,j}, \ j \geq 1\}$ in (A8.6)– (A8.7), respectively. Then,*

$$\left(\frac{n}{\ln(n)}\right)^{1/2} \sup_{j \geq 1} |C_{n,j} - C_j| \longrightarrow 0 \, a.s., \quad n \to \infty.$$

Lemma A8.3.4 *(See details in [Bosq, 2000, Corollary 4.3, p. 107]) Under* **Assumption A1***, consider Λ_{k_n} in equation (A8.10) satisfying*

$$\Lambda_{k_n} = o\left(\left(\frac{n}{\ln(n)}\right)^{1/2}\right), \quad n \to \infty.$$

Then,

$$\sup_{1 \leq j \leq k_n} \|\phi'_{n,j} - \phi_{n,j}\|_{\widetilde{H}} \longrightarrow 0 \, a.s., \quad n \to \infty,$$

where, for $j \geq 1$, and $n \geq 2$,

$$\phi'_{n,j} = \mathrm{sgn}\langle \phi_{n,j}, \phi_j \rangle_{\widetilde{H}} \phi_j, \quad \mathrm{sgn}\langle \phi_{n,j}, \phi_j \rangle_{\widetilde{H}} = \mathbf{1}_{\langle \phi_{n,j}, \phi_j \rangle_{\widetilde{H}} \geq 0} - \mathbf{1}_{\langle \phi_{n,j}, \phi_j \rangle_{\widetilde{H}} < 0},$$

with $\mathbf{1}_{\cdot}$ being the indicator function.

An upper bound for $\|c\|_{B \times B} = \left\| \sum_{j=1}^{\infty} C_j \phi_j \otimes \phi_j \right\|_{B \times B}$ is now obtained.

Lemma A8.3.5 *Under* ***Assumption A5****, the following inequality holds:*

$$\|c\|_{B\times B} = \sup_{n,m\geq 1} |C\left(F_n\right)\left(F_m\right)| \leq N \|C\|_{\mathcal{L}(\widetilde{H})},$$

where N has been introduced in equation (A8.18), $\mathcal{L}(\widetilde{H})$ denotes the space of bounded linear operators on $\widetilde{H}$, and $\|\cdot\|_{\mathcal{L}(\widetilde{H})}$ the usual uniform norm on such a space.

Let us consider the following notation.

$$\begin{aligned}
c &\underset{\widetilde{H}\otimes\widetilde{H}}{=} \sum_{j=1}^{\infty} C_j \phi'_{n,j} \otimes \phi'_{n,j} \underset{\widetilde{H}\otimes\widetilde{H}}{=} \sum_{j=1}^{\infty} C_j \phi_j \otimes \phi_j, \quad c_n \underset{\widetilde{H}\otimes\widetilde{H}}{=} \sum_{j=1}^{\infty} C_{n,j} \phi_{n,j} \otimes \phi_{n,j}. \\
c - c_n &\underset{\widetilde{H}\otimes\widetilde{H}}{=} \sum_{j=1}^{\infty} C_j \phi'_{n,j} \otimes \phi'_{n,j} - \sum_{j=1}^{\infty} C_{n,j} \phi_{n,j} \otimes \phi_{n,j} \qquad \text{(A8.21)}
\end{aligned}$$

Remark A8.3.4 *From Lemma A8.3.2, for n sufficiently large, there exist positive constants K_1 and K_2 such that*

$$K_1 \langle C(\varphi), \varphi\rangle_{\widetilde{H}} \leq \langle C_n(\varphi), \varphi\rangle_{\widetilde{H}} \leq K_2 \langle C(\varphi), \varphi\rangle_{\widetilde{H}}, \quad \forall \varphi \in \widetilde{H}.$$

In particular, for every $x \in \mathcal{H}(X) = C^{1/2}(\widetilde{H})$, considering n sufficiently large,

$$\begin{aligned}
&\frac{1}{K_1} \left\langle C^{-1}(x), x\right\rangle_{\widetilde{H}} \geq \left\langle C_n^{-1}(x), x\right\rangle_{\widetilde{H}} \geq \frac{1}{K_2} \left\langle C^{-1}(x), x\right\rangle_{\widetilde{H}} \\
\Leftrightarrow\ &\frac{1}{K_1} \|x\|^2_{\mathcal{H}(X)} \geq \left\langle C_n^{-1}(x), x\right\rangle_{\widetilde{H}} \geq \frac{1}{K_2} \|x\|^2_{\mathcal{H}(X)}. \qquad \text{(A8.22)}
\end{aligned}$$

Equation (A8.22) means that, for n sufficiently large, the norm of the RKHS $\mathcal{H}(X)$ of X is equivalent to the norm of the RKHS generated by C_n, with spectral kernel c_n given in (A8.21).

Lemma A8.3.6 *Under* ***Assumptions A1*** *and* ***A4–A5****, let us consider Λ_{k_n} in (A8.10) satisfying*

$$\sqrt{k_n} \Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right), \quad n \to \infty, \qquad \text{(A8.23)}$$

where k_n has been introduced in ***Assumption A2***. *The following a.s. inequality then holds:*

$$\|c - c_n\|_{B \times B} \leq \max(N, \sqrt{N}) \left[\|C - C_n\|_{\mathcal{L}(\widetilde{H})} \right.$$
$$+ 2 \max\left(\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}}, \sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}} \right) \left[\sup_{l \geq 1} \sup_{m \geq 1} \left| F_l(\phi'_{n,m}) \right| \right]$$
$$\left. \times \sqrt{k_n 8 \Lambda_{k_n}^2 \|C_n - C\|_{\mathcal{L}(\widetilde{H})}^2 + \sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2} \right.$$

Therefore, $\|c - c_n\|_{B \times B} \to_{a.s.} 0$, *as* $n \to \infty$.

Lemma A8.3.7 *For a standard ARB(1) process satisfying equation (A8.1), under* ***Assumptions A1*** *and* ***A3–A5***, *for n sufficiently large,*

$$\sup_{1 \leq j \leq k_n} \left\| \phi_{n,j} - \phi'_{n,j} \right\|_B$$
$$\leq \frac{2}{C_{k_n}} \left[\max(N, \sqrt{N}) \left[\|C - C_n\|_{\mathcal{L}(\widetilde{H})} \right. \right.$$
$$+ 2 \max\left(\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}}, \sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}} \right) \left(\sup_{l \geq 1} \sup_{m \geq 1} \left| F_l(\phi'_{n,m}) \right| \right)$$
$$\left. \times \sqrt{k_n 8 \Lambda_{k_n}^2 \|C_n - C\|_{\mathcal{L}(\widetilde{H})}^2 + \sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2} \right]$$
$$\left. + \sup_{1 \leq j \leq k_n} \|\phi_{n,j} - \phi'_{n,j}\|_{\widetilde{H}} N \|C\|_{\mathcal{S}(\widetilde{H})} + V \|C - C_n\|_{\mathcal{S}(\widetilde{H})} \right] \quad a.s. \tag{A8.24}$$

Under (A8.23),

$$\sup_{1 \leq j \leq k_n} \left\| \phi_{n,j} - \phi'_{n,j} \right\|_B \longrightarrow 0 \, a.s., \quad n \to \infty.$$

Lemma A8.3.8 *Under* ***Assumption A3***, *if*

$$\sum_{j=1}^{k_n} \|\phi_{n,j} - \phi'_{n,j}\|_B \to_{a.s.} 0, , \quad n \to \infty,$$

then

$$\sup_{x \in B;\, \|x\|_B \leq 1} \left\| \rho(x) - \sum_{j=1}^{k_n} \langle \rho(x), \phi_{n,j} \rangle_{\widetilde{H}} \, \phi_{n,j} \right\|_B \longrightarrow 0 \, a.s., \quad n \to \infty. \tag{A8.25}$$

Remark A8.3.5 *Under the conditions of Lemma A8.3.7, if*

$$k_n^{3/2}\Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}}\right), \qquad \sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 = o\left(\frac{1}{k_n}\right), \; n \to \infty,$$

then, equation (A8.25) holds.

Let us know consider the projection operators

$$\widetilde{\Pi}^{k_n}(x) = \sum_{j=1}^{k_n} \langle x, \phi_{n,j}\rangle_{\widetilde{H}} \phi_{n,j}, \quad \Pi^{k_n}(x) = \sum_{j=1}^{k_n} \langle x, \phi'_{n,j}\rangle_{\widetilde{H}} \phi'_{n,j}, \quad x \in B \subset \widetilde{H}. \quad \text{(A8.26)}$$

Remark A8.3.6 *Under the conditions of Remark A8.3.5, let*

$$\widetilde{\Pi}^{k_n} \rho \widetilde{\Pi}^{k_n} = \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \langle \rho(\phi_{n,j}), \phi_{n,p}\rangle_{\widetilde{H}} \, \phi_{n,j} \otimes \phi_{n,p},$$

then

$$\sup_{x \in B;\, \|x\|_B \leq 1} \left\| \rho(x) - \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \langle x, \phi_{n,j}\rangle_{\widetilde{H}} \langle \rho(\phi_{n,j}), \phi_{n,p}\rangle_{\widetilde{H}} \, \phi_{n,p} \right\|_B \longrightarrow 0 \; a.s., \; n \to \infty.$$

A8.4 PROOFS OF LEMMAS

PROOF OF LEMMA A8.3.5

Proof. Applying the Cauchy–Schwarz's inequality, for every $k, l \geq 1$,

$$\begin{aligned} |C(F_k, F_l)| &= \left| \sum_{j=1}^{\infty} C_j F_k(\phi_j) F_l(\phi_j) \right| \leq \sqrt{\sum_{j=1}^{\infty} C_j [F_k(\phi_j)]^2 \sum_{p=1}^{\infty} C_p [F_l(\phi_p)]^2} \\ &\leq \sup_{j \geq 1} |C_j| \sqrt{\sum_{j=1}^{\infty} [F_k(\phi_j)]^2 \sum_{p=1}^{\infty} [F_l(\phi_p)]^2} = \sup_{j \geq 1} |C_j| \sqrt{N_k N_l}, \end{aligned}$$

where $\{F_n,\ n \geq 1\}$ have been introduced in equation (A8.3), and satisfy (A8.4)–(A8.5). Under **Assumption A5**, from equation (A8.18),

$$\|c\|_{B \times B} = \sup_{k,l \geq 1} |C(F_k, F_l)| \leq \sup_{k,l \geq 1} \sup_{j \geq 1} |C_j| \sqrt{N_k N_l} = N \sup_{j \geq 1} |C_j| = N \|C\|_{\mathcal{L}(\widetilde{H})}.$$

Proof of Lemma A8.3.6

Proof. Let us first consider the following identities and inequalities:

$$
\begin{aligned}
|C - C_n(F_k)(F_l)| &= \left| \sum_{j=1}^{\infty} C_j F_k(\phi'_{n,j}) F_l(\phi'_{n,j}) - C_{n,j} F_k(\phi_{n,j}) F_l(\phi_{n,j}) \right| \\
&\leq \sum_{j=1}^{\infty} |C_j| |F_k(\phi'_{n,j})| |F_l(\phi'_{n,j}) - F_l(\phi_{n,j})| \\
&\qquad + \sup_j |C_j - C_{n,j}| |F_k(\phi'_{n,j}) F_l(\phi_{n,j})| \\
&\qquad + |C_{n,j} F_l(\phi_{n,j})| |F_k(\phi'_{n,j}) - F_k(\phi_{n,j})| \\
&\leq \sqrt{\sum_{j=1}^{\infty} C_j \left\{F_k(\phi'_{n,j})\right\}^2 \sum_{j=1}^{\infty} C_j \left\{F_l(\phi'_{n,j}) - F_l(\phi_{n,j})\right\}^2} \\
&+ \sup_{j \geq 1} |C_j - C_{n,j}| \sqrt{\sum_{j=1}^{\infty} \left\{F_k(\phi'_{n,j})\right\}^2 \sum_{j=1}^{\infty} \left\{F_l(\phi_{n,j})\right\}^2} \\
&+ \sqrt{\sum_{j=1}^{\infty} C_{n,j} \left\{F_l(\phi_{n,j})\right\}^2 \sum_{j=1}^{\infty} C_{n,j} \left\{F_k(\phi'_{n,j}) - F_k(\phi_{n,j})\right\}^2} \\
&\leq \sqrt{N_k} \sqrt{\sum_{j=1}^{\infty} C_j \left\{F_l(\phi'_{n,j}) - F_l(\phi_{n,j})\right\}^2} \\
&+ \sup_{j \geq 1} |C_j - C_{n,j}| \sqrt{N_k} \sqrt{N_l} \\
&+ \sqrt{N_l} \sqrt{\sum_{j=1}^{\infty} C_{n,j} \left\{F_k(\phi'_{n,j}) - F_k(\phi_{n,j})\right\}^2} \\
&\leq \max(N, \sqrt{N}) \left[\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})} \sum_{j=1}^{\infty} \left\{F_l(\phi'_{n,j} - \phi_{n,j})\right\}^2} \right. \\
&\qquad + \|C - C_n\|_{\mathcal{L}(\widetilde{H})} \\
&\qquad \left. + \sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})} \sum_{j=1}^{\infty} \left\{F_k(\phi'_{n,j} - \phi_{n,j})\right\}^2} \right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq \max(N,\sqrt{N})\Bigg[\|C-C_n\|_{\mathcal{L}(\widetilde{H})} \\
&\quad+\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}\sum_{j=1}^{\infty}\sum_{m=1}^{\infty}\left\{F_l(\phi'_{n,m})\right\}^2\left\{\left\langle\phi'_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}-\left\langle\phi_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}\right\}^2} \\
&\quad+\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}\sum_{j=1}^{\infty}\sum_{m=1}^{\infty}\left\{F_k(\phi'_{n,m})\right\}^2\left\{\left\langle\phi'_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}-\left\langle\phi_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}\right\}^2}\,\Bigg] \\
&= \max(N,\sqrt{N})\Bigg[\|C-C_n\|_{\mathcal{L}(\widetilde{H})} \\
&\quad+\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\left\{F_l(\phi'_{n,m})\right\}^2\sum_{j=1}^{\infty}\left\{\left\langle\phi'_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}-\left\langle\phi_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}\right\}^2} \\
&\quad+\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\left\{F_k(\phi'_{n,m})\right\}^2\sum_{j=1}^{\infty}\left\{\left\langle\phi'_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}-\left\langle\phi_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}\right\}^2}\,\Bigg] \\
&= \max(N,\sqrt{N})\Bigg[\|C-C_n\|_{\mathcal{L}(\widetilde{H})} \\
&\quad+\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\left\{F_l(\phi'_{n,m})\right\}^2\sum_{j=1}^{\infty}\left\{\left\langle\phi_{n,j},\phi_{n,m}\right\rangle_{\widetilde{H}}-\left\langle\phi_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}\right\}^2} \\
&\quad+\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\left\{F_k(\phi'_{n,m})\right\}^2\sum_{j=1}^{\infty}\left\{\left\langle\phi_{n,j},\phi_{n,m}\right\rangle_{\widetilde{H}}-\left\langle\phi_{n,j},\phi'_{n,m}\right\rangle_{\widetilde{H}}\right\}^2}\,\Bigg] \\
&= \max(N,\sqrt{N})\Bigg[\|C-C_n\|_{\mathcal{L}(\widetilde{H})} \\
&\quad+\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\left\{F_l(\phi'_{n,m})\right\}^2\|\phi_{n,m}-\phi'_{n,m}\|_{\widetilde{H}}^2} \\
&\quad+\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\left\{F_k(\phi'_{n,m})\right\}^2\|\phi_{n,m}-\phi'_{n,m}\|_{\widetilde{H}}^2}\,\Bigg] \\
&\leq \max(N,\sqrt{N})\Bigg[\|C-C_n\|_{\mathcal{L}(\widetilde{H})} \\
&\quad+\sup_{m\geq 1}\left|F_l(\phi'_{n,m})\right|\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\|\phi_{n,m}-\phi'_{n,m}\|_{\widetilde{H}}^2} \\
&\quad+\sup_{m\geq 1}\left|F_k(\phi'_{n,m})\right|\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}\sum_{m=1}^{\infty}\|\phi_{n,m}-\phi'_{n,m}\|_{\widetilde{H}}^2}\,\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\leq \quad & \max(N, \sqrt{N}) \left[\|C - C_n\|_{\mathcal{L}(\widetilde{H})} \right. \\
& + \max\left(\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}}, \sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}} \right) \\
& \left. \left[\sup_{m \geq 1} \left| F_l(\phi'_{n,m}) \right| + \sup_{m \geq 1} \left| F_k(\phi'_{n,m}) \right| \right] \sqrt{\sum_{m=1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2} \right]. \qquad \text{(A8.27)}
\end{aligned}
$$

Under **Assumption A5**, from equation (A8.17),

$$
\sum_{m=1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 < \infty, \quad \sup_{m \geq 1} \left| F_k(\phi'_{n,m}) \right| < \infty, \quad k \geq 1.
$$

Thus, considering k_n, as given in **Assumption A2**,

$$
\begin{aligned}
& \sum_{m=1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 = \sum_{m=1}^{k_n} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 + \sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 \\
& \leq k_n \sup_{1 \leq m \leq k_n} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 + \sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 \\
& \leq k_n 8 \Lambda_{k_n}^2 \|C_n - C\|_{\mathcal{L}(\widetilde{H})}^2 + \sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 \qquad \text{(A8.28)}
\end{aligned}
$$

From equation (A8.20), under $\Lambda_{k_n} = o\left(\sqrt{\frac{n}{\ln(n)}} \right)$,

$$
k_n 8 \Lambda_{k_n}^2 \|C_n - C\|_{\mathcal{L}(\widetilde{H})}^2 \leq k_n 8 \Lambda_{k_n}^2 \|C_n - C\|_{\mathcal{S}(\widetilde{H})}^2 \to_{a.s.} 0, \ n \to \infty. \qquad \text{(A8.29)}
$$

Under **Assumption A5**,

$$
\sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2 \to_{a.s.} 0, \quad n \to \infty. \qquad \text{(A8.30)}
$$

From equations (A8.27)–(A8.30), since, under **Assumption A5**,

$$
\sup_{k \geq 1} \sup_{m \geq 1} \left| F_k(\phi'_{n,m}) \right| < \infty,
$$

we have $\|c - c_n\|_{B \times B} = \sup_{k,l} |(C - C_n)(F_k)(F_l)| \to_{a.s.} 0$, as $n \to \infty$.

■

Proof of Lemma A8.3.7

Proof. Let us first consider the following a.s. equalities

$$\begin{aligned} C_{n,j}\left(\phi_{n,j}-\phi'_{n,j}\right) &= C_n\left(\phi_{n,j}\right)-C_{n,j}\phi'_{n,j}=\left(C_n-C\right)\left(\phi_{n,j}\right) \\ &+ C\left(\phi_{n,j}-\phi'_{n,j}\right)+\left(C_j-C_{n,j}\right)\phi'_{n,j}. \end{aligned} \tag{A8.31}$$

From equation (A8.31), keeping in mind **Assumption A2**,

$$\begin{aligned} \left\|\phi_{n,j}-\phi'_{n,j}\right\|_B &\leq \frac{1}{C_{n,j}}\left\|\left(C_n-C\right)\left(\phi_{n,j}\right)\right\|_B+\frac{1}{C_{n,j}}\left\|C\left(\phi_{n,j}-\phi'_{n,j}\right)\right\|_B \\ &+ \frac{1}{C_{n,j}}\left\|\left(C_j-C_{n,j}\right)\phi'_{n,j}\right\|_B=\frac{1}{C_{n,j}}\left[S_1+S_2+S_3\right], \quad \text{a.s..} \end{aligned} \tag{A8.32}$$

For n sufficiently large, from Lemmas A8.3.5 and A8.3.6, applying the Cauchy–Schwarz's inequality, for every $j \geq 1$,

$$\begin{aligned} S_1 &= \left\|\left(C_n-C\right)\left(\phi_{n,j}\right)\right\|_B \\ &= \sup_m\left|\sum_{k=1}^{\infty} C_{n,k}F_m(\phi_{n,k})\left\langle\phi_{n,k},\phi_{n,j}\right\rangle_{\widetilde{H}}-\sum_{k=1}^{\infty} C_kF_m(\phi'_{n,k})\left\langle\phi'_{n,k},\phi_{n,j}\right\rangle_{\widetilde{H}}\right| \\ &= \sup_m\left|\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} t_lF_l(\phi_{n,j})\left\{C_{n,k}F_m(\phi_{n,k})F_l(\phi_{n,k})-C_kF_m(\phi'_{n,k})F_l(\phi'_{n,k})\right\}\right| \\ &= \sup_m\left|\sum_{l=1}^{\infty} t_lF_l(\phi_{n,j})\sum_{k=1}^{\infty} C_{n,k}F_m(\phi_{n,k})F_l(\phi_{n,k})-C_kF_m(\phi'_{n,k})F_l(\phi'_{n,k})\right| \\ &\leq \sup_m\sqrt{\sum_{l=1}^{\infty} t_l[F_l(\phi_{n,j})]^2} \\ &\quad\times\sqrt{\sum_{l=1}^{\infty} t_l\left\{\sum_{k=1}^{\infty} C_{n,k}F_m(\phi_{n,k})F_l(\phi_{n,k})-C_kF_m(\phi'_{n,k})F_l(\phi'_{n,k})\right\}^2} \\ &\leq \|\phi_{n,j}\|_{\widetilde{H}}\sqrt{\sum_{l=1}^{\infty} t_l\sup_{m,l}\left|\sum_{k=1}^{\infty} C_{n,k}F_m(\phi_{n,k})F_l(\phi_{n,k})-C_kF_m(\phi'_{n,k})F_l(\phi'_{n,k})\right|} \\ &\quad= \|c_n-c\|_{B\times B} \\ &\quad\leq \max(N,\sqrt{N})\left[\|C-C_n\|_{\mathcal{L}(\widetilde{H})}\right. \\ &\quad\left.+2\max\left(\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}},\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}}\right)\left[\sup_{l\geq 1}\sup_{m\geq 1}\left|F_l(\phi'_{n,m})\right|\right]\right] \end{aligned}$$

$$\times \sqrt{k_n 8\Lambda_{k_n}^2 \|C_n - C\|_{\mathcal{L}(\widetilde{H})}^2 + \sum_{m=k_n+1}^{\infty} \|\phi_{n,m} - \phi'_{n,m}\|_{\widetilde{H}}^2} \tag{A8.33}$$

$$\begin{aligned} S_2 &= \left\|C\left(\phi_{n,j} - \phi'_{n,j}\right)\right\|_B = \sup_m \left|\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} t_l C_k F_m(\phi'_{n,k}) F_l(\phi'_{n,k}) F_l\left(\phi_{n,j} - \phi'_{n,j}\right)\right| \\ &\leq \sup_m \sqrt{\sum_{l=1}^{\infty} t_l \left\{F_l\left(\phi_{n,j} - \phi'_{n,j}\right)\right\}^2} \sqrt{\sum_{l=1}^{\infty} t_l \left\{\sum_{k=1}^{\infty} C_k F_m(\phi'_{n,k}) F_l(\phi'_{n,k})\right\}^2} \\ &\leq \|\phi_{n,j} - \phi'_{n,j}\|_{\widetilde{H}} \sup_{m,l} \left|\sum_{k=1}^{\infty} C_k F_m(\phi'_{n,k}) F_l(\phi'_{n,k})\right| \\ &= \|\phi_{n,j} - \phi'_{n,j}\|_{\widetilde{H}} \|c\|_{B\times B} \leq \|\phi_{n,j} - \phi'_{n,j}\|_{\widetilde{H}} N \|C\|_{\mathcal{S}(\widetilde{H})}, \quad \text{a.s.} \end{aligned} \tag{A8.34}$$

Under Assumption A3,

$$S_3 \leq \sup_{j\geq 1} |C_j - C_{n,j}| \left\|\phi'_{n,j}\right\|_B \leq V \|C - C_n\|_{\mathcal{L}(\widetilde{H})} \leq V \|C - C_n\|_{\mathcal{S}(\widetilde{H})}, \text{ a.s.} \tag{A8.35}$$

In addition, from Lemma A8.3.2,

$$\|C_n - C\|_{\mathcal{S}(\widetilde{H})} \to_{a.s.} 0, \quad n \to \infty,$$

and

$$C_{n,j} \to_{a.s.} C_j, \quad n \to \infty.$$

For $\varepsilon = C_{k_n}/2$, we can find n_0 such that for $n \geq n_0$,

$$\begin{aligned} &\|C_n - C\|_{\mathcal{L}(\widetilde{H})} \leq \varepsilon = C_{k_n}/2, \quad \text{a.s.} \\ &|C_{n,k_n} - C_{k_n}| \leq \widetilde{\varepsilon} \leq \|C_n - C\|_{\mathcal{L}(\widetilde{H})} \\ &C_{n,k_n} \geq C_{k_n} - \widetilde{\varepsilon} \geq C_{k_n} - \|C_n - C\|_{\mathcal{L}(\widetilde{H})} \geq C_{k_n} - C_{k_n}/2 \geq C_{k_n}/2. \end{aligned} \tag{A8.36}$$

From equations (A8.32)–(A8.35), for n large enough such that equation (A8.36) holds, the following almost surely inequalities are satisfied. For $1 \leq j \leq k_n$,

$$
\begin{aligned}
&\sup_{1\leq j\leq k_n} \left\|\phi_{n,j}-\phi'_{n,j}\right\|_B \\
&\leq \frac{1}{C_{n,k_n}}\left[\max(N,\sqrt{N})\left[\|C-C_n\|_{\mathcal{L}(\widetilde{H})}\right.\right. \\
&+2\max\left(\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}},\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}}\right)\left\{\sup_{l\geq 1}\sup_{m\geq 1}\left|F_l(\phi'_{n,m})\right|\right\} \\
&\left.\times\sqrt{k_n 8\Lambda_{k_n}^2\|C_n-C\|^2_{\mathcal{L}(\widetilde{H})}+\sum_{m=k_n+1}^{\infty}\|\phi_{n,m}-\phi'_{n,m}\|^2_{\widetilde{H}}}\right] \\
&\left.+\sup_{1\leq j\leq k_n}\|\phi_{n,j}-\phi'_{n,j}\|_{\widetilde{H}}N\|C\|_{\mathcal{S}(\widetilde{H})}+V\|C-C_n\|_{\mathcal{S}(\widetilde{H})}\right] \\
&\leq \frac{2}{C_{k_n}}\left[\max(N,\sqrt{N})\left[\|C-C_n\|_{\mathcal{L}(\widetilde{H})}\right.\right. \\
&+2\max\left(\sqrt{\|C\|_{\mathcal{L}(\widetilde{H})}},\sqrt{\|C_n\|_{\mathcal{L}(\widetilde{H})}}\right)\left\{\sup_{l\geq 1}\sup_{m\geq 1}\left|F_l(\phi'_{n,m})\right|\right\} \\
&\left.\times\sqrt{k_n 8\Lambda_{k_n}^2\|C_n-C\|^2_{\mathcal{L}(\widetilde{H})}+\sum_{m=k_n+1}^{\infty}\|\phi_{n,m}-\phi'_{n,m}\|^2_{\widetilde{H}}}\right] \\
&\left.+\sup_{1\leq j\leq k_n}\|\phi_{n,j}-\phi'_{n,j}\|_{\widetilde{H}}N\|C\|_{\mathcal{S}(\widetilde{H})}+V\|C-C_n\|_{\mathcal{S}(\widetilde{H})}\right] \quad a.s.
\end{aligned}
$$

Hence, equation (A8.24) holds. The a.s. convergence to zero directly follows from Lemma A8.3.2, under (A8.23).

■

Proof of Lemma A8.3.8

Proof. The following identities are considered:

$$
\begin{aligned}
&\sum_{j=1}^{k_n}\left\langle\rho(x),\phi_{n,j}\right\rangle_{\widetilde{H}}\phi_{n,j}-\sum_{j=1}^{k_n}\left\langle\rho(x),\phi'_{n,j}\right\rangle_{\widetilde{H}}\phi'_{n,j} \\
&=\sum_{j=1}^{k_n}\left\langle\rho(x),\phi_{n,j}\right\rangle_{\widetilde{H}}(\phi_{n,j}-\phi'_{n,j})+\sum_{j=1}^{k_n}\left\langle\rho(x),\phi_{n,j}-\phi'_{n,j}\right\rangle_{\widetilde{H}}\phi'_{n,j}. \qquad \text{(A8.37)}
\end{aligned}
$$

From equation (A8.37), applying the Cauchy–Schwarz's inequality, under **Assumption A3**,

$$
\begin{aligned}
&\sup_{x\in B;\ \|x\|_B\leq 1}\left\|\sum_{j=1}^{k_n}\langle \rho(x),\phi_{n,j}\rangle_{\widetilde{H}}\,\phi_{n,j}-\sum_{j=1}^{\infty}\langle \rho(x),\phi'_{n,j}\rangle_{\widetilde{H}}\,\phi'_{n,j}\right\|_B\\
&\leq \sup_{x\in B;\ \|x\|_B\leq 1}\sum_{j=1}^{k_n}\|\rho(x)\|_{\widetilde{H}}\|\phi_{n,j}\|_{\widetilde{H}}\|\phi_{n,j}-\phi'_{n,j}\|_B\\
&\qquad\qquad +\|\rho(x)\|_{\widetilde{H}}\|\phi_{n,j}-\phi'_{n,j}\|_{\widetilde{H}}\|\phi'_{n,j}\|_B\\
&+\sup_{x\in B;\ \|x\|_B\leq 1}\left\|\sum_{j=k_n+1}^{\infty}\left\langle \rho(x),\phi'_{n,j}\right\rangle_{\widetilde{H}}\phi'_{n,j}\right\|_B\\
&\leq \sup_{x\in B;\ \|x\|_B\leq 1}\|\rho(x)\|_{\widetilde{H}}\left(\sum_{j=1}^{k_n}\|\phi_{n,j}-\phi'_{n,j}\|_B+\|\phi_{n,j}-\phi'_{n,j}\|_B\sup_j\|\phi'_{n,j}\|_B\right)\\
&+\sup_{x\in B;\ \|x\|_B\leq 1}\left\|\sum_{j=k_n+1}^{\infty}\left\langle \rho(x),\phi'_{n,j}\right\rangle_{\widetilde{H}}\phi'_{n,j}\right\|_B\\
&\leq \sup_{x\in B;\ \|x\|_B\leq 1}\|\rho\|_{\mathcal{L}(\widetilde{H})}\|x\|_{\widetilde{H}}(1+V)\sum_{j=1}^{k_n}\|\phi_{n,j}-\phi'_{n,j}\|_B\\
&+\sup_{x\in B;\ \|x\|_B\leq 1}\left\|\sum_{j=k_n+1}^{\infty}\left\langle \rho(x),\phi'_{n,j}\right\rangle_{\widetilde{H}}\phi'_{n,j}\right\|_B\\
&\leq \|\rho\|_{\mathcal{L}(\widetilde{H})}(1+V)\sum_{j=1}^{k_n}\|\phi_{n,j}-\phi'_{n,j}\|_B\\
&+\sup_{x\in B;\ \|x\|_B\leq 1}\left\|\sum_{j=k_n+1}^{\infty}\left\langle \rho(x),\phi'_{n,j}\right\rangle_{\widetilde{H}}\phi'_{n,j}\right\|_B\to 0,\quad n\to_{a.s.}\infty.
\end{aligned}
$$

■

A8.5 ARB(1) ESTIMATION AND PREDICTION. STRONG CONSISTENCY RESULTS

For every $x\in B\subset\widetilde{H}$, the following componentwise estimator $\widetilde{\rho}_{k_n}$ of ρ will be considered:

$$
\widetilde{\rho}_{k_n}(x)=\left(\widetilde{\Pi}^{k_n}D_nC_n^{-1}\widetilde{\Pi}^{k_n}\right)(x)=\left(\sum_{j=1}^{k_n}\frac{1}{C_{n,j}}\langle x,\phi_{n,j}\rangle_{\widetilde{H}}\widetilde{\Pi}^{k_n}D_n(\phi_{n,j})\right),
$$

where $\widetilde{\Pi}^{k_n}$ has been introduced in equation (A8.26), and C_n, $C_{n,j}$, $\phi_{n,j}$ and D_n have been defined in equations (A8.7)–(A8.8), respectively.

Theorem A8.5.1 *Let X be, as before, a standard ARB(1) process. Under the conditions of Lemmas A8.3.7 and A8.3.8 (see Remark A8.3.5), for all $\eta > 0$,*

$$\mathcal{P}\left(\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(B)} \geq \eta\right) \leq \mathcal{K}\exp\left(-\frac{n\eta^2}{Q_n}\right),$$

where

$$Q_n = \mathcal{O}\left(\left\{C_{k_n}^{-1} k_n \sum_{j=1}^{k_n} a_j\right\}^2\right), \quad n \to \infty.$$

Therefore, if

$$k_n C_{k_n}^{-1} \sum_{j=1}^{k_n} a_j = o\left(\sqrt{\frac{n}{\ln(n)}}\right), \quad n \to \infty, \tag{A8.38}$$

then,

$$\|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(B)} \to_{a.s} 0, \quad n \to \infty.$$

Proof. For every $x \in B$, such that $\|x\|_B \leq 1$, applying the triangle inequality, under **Assumptions A1–A2**,

$$\begin{aligned} \|\widetilde{\Pi}^{k_n} D_n C_n^{-1} \widetilde{\Pi}^{k_n}(x) - \widetilde{\Pi}^{k_n} \rho \widetilde{\Pi}^{k_n}(x)\|_B &\leq \|\widetilde{\Pi}^{k_n}(D_n - D) C_n^{-1} \widetilde{\Pi}^{k_n}(x)\|_B \\ &+ \|\widetilde{\Pi}^{k_n}(D C_n^{-1} - \rho)\widetilde{\Pi}^{k_n}(x)\|_B \\ &= S_1(x) + S_2(x). \end{aligned} \tag{A8.39}$$

Under **Assumption A3**, considering inequality (A8.36),

$$\begin{aligned} S_1(x) &= \|\widetilde{\Pi}^{k_n}(D_n - D) C_n^{-1} \widetilde{\Pi}^{k_n}(x)\|_B \\ &\leq \left\| C_{n,k_n}^{-1} \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \langle x, \phi_{n,j}\rangle_{\widetilde{H}} \langle (D_n - D)(\phi_{n,j}), \phi_{n,p}\rangle_{\widetilde{H}}\, \phi_{n,p} \right\|_B \\ &\leq \left|C_{n,k_n}^{-1}\right| \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \left|\langle x, \phi_{n,j}\rangle_{\widetilde{H}}\right| \left|\langle (D_n - D)(\phi_{n,j}), \phi_{n,p}\rangle_{\widetilde{H}}\right| \|\phi_{n,p}\|_B \\ &\leq 2 C_{k_n}^{-1} k_n \|D_n - D\|_{\mathcal{L}(\widetilde{H})} \sum_{p=1}^{k_n} \|\phi_{n,p}\|_B \\ &\leq 2V C_{k_n}^{-1} k_n^2 \|D_n - D\|_{\mathcal{S}(\widetilde{H})}. \end{aligned} \tag{A8.40}$$

Furthermore, applying the triangle inequality,

$$\begin{aligned} S_2(x) &= \|\widetilde{\Pi}^{k_n}(D C_n^{-1} - \rho)\widetilde{\Pi}^{k_n}(x)\|_B \\ &\leq \|\widetilde{\Pi}^{k_n} D C_n^{-1} \widetilde{\Pi}^{k_n}(x) - \widetilde{\Pi}^{k_n} D C^{-1} \Pi^{k_n}(x)\|_B \\ &+ \|\widetilde{\Pi}^{k_n} D C^{-1} \Pi^{k_n}(x) - \widetilde{\Pi}^{k_n} \rho \widetilde{\Pi}^{k_n}(x)\|_B = S_{21}(x) + S_{22}(x). \end{aligned} \tag{A8.41}$$

Under **Assumptions A1–A2**, C^{-1} and C_n^{-1} are bounded on the subspaces generated by $\{\phi_j,\ j=1,\ldots,k_n\}$ and $\{\phi_{n,j},\ j=1,\ldots,k_n\}$, respectively. Consider now

$$\begin{aligned}
S_{21}(x) &= \|\widetilde{\Pi}^{k_n} D C_n^{-1} \widetilde{\Pi}^{k_n}(x) - \widetilde{\Pi}^{k_n} D C^{-1} \Pi^{k_n}(x)\|_B \\
&= \left\| \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \frac{1}{C_{n,j}} \left\langle x, \phi_{n,j} - \phi'_{n,j} \right\rangle_{\widetilde{H}} \left\langle D(\phi_{n,j}), \phi_{n,p} \right\rangle_{\widetilde{H}} \phi_{n,p} \right. \\
&\quad + \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \left(\frac{1}{C_{n,j}} - \frac{1}{C_j} \right) \left\langle x, \phi'_{n,j} \right\rangle_{\widetilde{H}} \left\langle D(\phi_{n,j}), \phi_{n,p} \right\rangle_{\widetilde{H}} \phi_{n,p} \\
&\quad \left. + \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \frac{1}{C_j} \left\langle x, \phi'_{n,j} \right\rangle_{\widetilde{H}} \left\langle D(\phi_{n,j} - \phi'_{n,j}), \phi_{n,p} \right\rangle_{\widetilde{H}} \phi_{n,p} \right\|_B \\
&\leq \sum_{j=1}^{k_n} \sum_{p=1}^{k_n} \left| \frac{1}{C_{n,k_n}} \right| \|\phi_{n,j} - \phi'_{n,j}\|_{\widetilde{H}} \|D\|_{\mathcal{L}(\widetilde{H})} \|\phi_{n,p}\|_B \\
&\quad + \left| \frac{1}{C_{n,j}} - \frac{1}{C_j} \right| \|D\|_{\mathcal{L}(\widetilde{H})} \|\phi_{n,p}\|_B \\
&\quad + \left| \frac{1}{C_j} \right| \|D\|_{\mathcal{L}(\widetilde{H})} \|\phi_{n,j} - \phi'_{n,j}\|_{\widetilde{H}} \|\phi_{n,p}\|_B .
\end{aligned} \tag{A8.42}$$

From [Bosq, 2000, Lemma 4.3, p. 104], for every $j \geq 1$, under **Assumption A1**,

$$\|\phi_{n,j} - \phi'_{n,j}\|_{\widetilde{H}} \leq a_j \|C_n - C\|_{\mathcal{L}(\widetilde{H})}, \tag{A8.43}$$

where $\{a_j,\ j \geq 1\}$ has been introduced in (A8.9), for $j \geq 1$. Then, in equation (A8.42), considering again inequality (A8.36), keeping in mind that $C_j^{-1} \leq a_j$, we obtain

$$\begin{aligned}
S_{21}(x) &\leq 4 C_{k_n}^{-1} \sum_{p=1}^{k_n} \|\phi_{n,p}\|_B \|D\|_{\mathcal{L}(\widetilde{H})} \|C_n - C\|_{\mathcal{L}(\widetilde{H})} \sum_{j=1}^{k_n} a_j \\
&\leq 4 V k_n C_{k_n}^{-1} \|D\|_{\mathcal{L}(\widetilde{H})} \|C_n - C\|_{\mathcal{S}(\widetilde{H})} \sum_{j=1}^{k_n} a_j .
\end{aligned} \tag{A8.44}$$

Applying again the triangle and the Cauchy–Schwarz inequalities, from (A8.43),

$$\begin{aligned}
S_{22} &= \|\widetilde{\Pi}^{k_n} DC^{-1}\Pi^{k_n}(x) - \widetilde{\Pi}^{k_n}\rho\widetilde{\Pi}^{k_n}(x)\|_B \\
&= \left\| \sum_{j=1}^{k_n}\sum_{p=1}^{k_n} \left\langle x, \phi'_{n,j} - \phi_{n,j}\right\rangle_{\widetilde{H}} \left\langle \rho(\phi'_{n,j}), \phi_{n,p}\right\rangle_{\widetilde{H}} \phi_{n,p} \right. \\
&\quad \left. + \left\langle x, \phi_{n,j}\right\rangle_{\widetilde{H}} \left\langle \rho(\phi'_{n,j} - \phi_{n,j}), \phi_{n,p}\right\rangle_{\widetilde{H}} \phi_{n,p} \right\| \\
&\leq \sum_{j=1}^{k_n}\sum_{p=1}^{k_n} \|x\|_{\widetilde{H}} \|\phi'_{n,j} - \phi_{n,j}\|_{\widetilde{H}} \|\rho\|_{\mathcal{L}(\widetilde{H})} \|\phi'_{n,j}\|_{\widetilde{H}} \|\phi_{n,p}\|_{\widetilde{H}} \|\phi_{n,p}\|_B \\
&\quad + \|x\|_{\widetilde{H}} \|\phi_{n,j}\|_{\widetilde{H}} \|\rho\|_{\mathcal{L}(\widetilde{H})} \|\phi'_{n,j} - \phi_{n,j}\|_{\widetilde{H}} \|\phi_{n,p}\|_{\widetilde{H}} \|\phi_{n,p}\|_B \\
&\leq 2\|\rho\|_{\mathcal{L}(\widetilde{H})} \|C_n - C\|_{\mathcal{S}(\widetilde{H})} \left(\sum_{p=1}^{k_n} \|\phi_{n,p}\|_B\right) \left(\sum_{j=1}^{k_n} a_j\right) \\
&\leq 2V\|\rho\|_{\mathcal{L}(\widetilde{H})} \|C_n - C\|_{\mathcal{S}(\widetilde{H})} k_n \sum_{j=1}^{k_n} a_j. \qquad \text{(A8.45)}
\end{aligned}$$

From equations (A8.39)–(A8.45),

$$\begin{aligned}
&\sup_{x\in B;\ \|x\|_B \leq 1} \|\widetilde{\Pi}^{k_n} D_n C_n^{-1}\widetilde{\Pi}^{k_n}(x) - \widetilde{\Pi}^{k_n}\rho\widetilde{\Pi}^{k_n}(x)\|_B \\
&\leq 2VC_{k_n}^{-1}k_n^2 \|D_n - D\|_{\mathcal{S}(\widetilde{H})} \\
&+ \|C_n - C\|_{\mathcal{S}(\widetilde{H})} 2Vk_n \sum_{j=1}^{k_n} a_j \left(2C_{k_n}^{-1}\|D\|_{\mathcal{L}(\widetilde{H})} + \|\rho\|_{\mathcal{L}(\widetilde{H})}\right). \qquad \text{(A8.46)}
\end{aligned}$$

From equation (A8.46), applying now [Bosq, 2000, Theorem 4.2, p. 99; Theorem 4.8, p. 116], one can get, for $\eta > 0$,

$$
\begin{aligned}
&\mathcal{P}\left(\sup_{x\in B;\ \|x\|_B\leq 1}\|\widetilde{\Pi}^{k_n}D_nC_n^{-1}\widetilde{\Pi}^{k_n}(x)-\widetilde{\Pi}^{k_n}\rho\widetilde{\Pi}^{k_n}(x)\|_B>\eta\right)\\
&\leq \mathcal{P}\left(\sup_{x\in B;\ \|x\|_B\leq 1}S_1(x)>\eta\right)+\mathcal{P}\left(\sup_{x\in B;\ \|x\|_B\leq 1}S_{21}(x)+S_{22}(x)>\eta\right)\\
&\leq \mathcal{P}\left(\|D_n-D\|_{\mathcal{S}(\widetilde{H})}>\frac{\eta}{2VC_{k_n}^{-1}k_n^2}\right)\\
&+\mathcal{P}\left(\|C_n-C\|_{\mathcal{S}(\widetilde{H})}>\frac{\eta}{2Vk_n\sum_{j=1}^{k_n}a_j\left[2C_{k_n}^{-1}\|D\|_{\mathcal{L}(\widetilde{H})}+\|\rho\|_{\mathcal{L}(\widetilde{H})}\right]}\right)\\
&\leq 8\exp\left(-\frac{n\eta^2}{\left(2VC_{k_n}^{-1}k_n^2\right)^2\left(\gamma+\delta\left(\frac{\eta}{2VC_{k_n}^{-1}k_n^2}\right)\right)}\right)+4\exp\left(-\frac{n\eta^2}{Q_n}\right),
\end{aligned}
\tag{A8.47}
$$

with γ and δ being positive numbers, depending on ρ and $\mathcal{P}_{\varepsilon_0}$, respectively, introduced in [Bosq, 2000, Theorems 4.2 and 4.8]. Here,

$$
\begin{aligned}
Q_n &= 4V^2k_n^2\left(\sum_{j=1}^{k_n}a_j\right)^2\left[2C_{k_n}^{-1}\|D\|_{\mathcal{L}(\widetilde{H})}+\|\rho\|_{\mathcal{L}(\widetilde{H})}\right]^2\\
&\times\left[\alpha_1+\beta_1\frac{\eta}{2Vk_n\sum_{j=1}^{k_n}a_j\left[2C_{k_n}^{-1}\|D\|_{\mathcal{L}(\widetilde{H})}+\|\rho\|_{\mathcal{L}(\widetilde{H})}\right]}\right],
\end{aligned}
\tag{A8.48}
$$

where again α_1 and β_1 are positive constants depending on ρ and $\mathcal{P}_{\varepsilon_0}$, respectively. From equations (A8.47) and (A8.48), if

$$
k_nC_{k_n}^{-1}\sum_{j=1}^{k_n}a_j=o\left(\sqrt{\frac{n}{\ln(n)}}\right),\quad n\to\infty,
$$

then, the Borel–Cantelli lemma, and Lemma A8.3.8 and Remarks A8.3.5– A8.3.6 lead to the desired a.s. convergence to zero.

■

Corollary A8.5.1 *Under the conditions of Theorem A8.5.1,*

$$\|\widetilde{\rho}_{k_n}(X_n) - \rho(X_n)\|_B \to_{a.s.} 0, \quad n \to \infty.$$

The proof is straightforward from Theorem A8.5.1, since

$$\|\widetilde{\rho}_{k_n}(X_n) - \rho(X_n)\|_B \leq \|\widetilde{\rho}_{k_n} - \rho\|_{\mathcal{L}(B)}\|X_0\|_B \to_{a.s} 0, \quad n \to \infty,$$

under **Assumption A1**.

A8.6 EXAMPLES: WAVELETS IN BESOV AND SOBOLEV SPACES

It is well–known that wavelets provide orthonormal bases of $L^2(\mathbb{R})$, and unconditional bases for several function spaces including Besov spaces,

$$\left\{B^s_{p,q}, \quad s \in \mathbb{R}, \quad 1 \leq p, q \leq \infty\right\}.$$

Sobolev or Hölder spaces constitute interesting particular cases of Besov spaces (see, for example, Triebel [1983]). Consider now orthogonal wavelets on the interval $[0,1]$. Adapting wavelets to a finite interval requires some modifications as described in Cohen et al. [1993]. Let $s > 0$, for an $[s]+1$-regular Multiresolution Analysis (MRA) of $L^2([0,1])$, where $[\cdot]$ stands for the integer part, the father φ and the mother ψ wavelets are such that $\varphi, \psi \in \mathcal{C}^{[s]+1}([0,1])$. Also φ and its derivatives, up to order $[s]+1$, have a fast decay (see [Daubechies, 1988, Corollary 5.2]). Let $2^J \geq 2([s]+1)$, the construction in Cohen et al. [1993] starts from a finite set of 2^J scaling functions $\left\{\varphi_{J,k},\ k = 0, 1, \ldots, 2^J - 1\right\}$. For each $j \geq J$, a set 2^j wavelet functions $\{\psi_{j,k},\ k = 0, 1, \ldots, 2^j - 1\}$ are also considered. The collection of these functions,

$$\left\{\varphi_{J,k},\ k = 0, 1, \ldots, 2^J - 1\right\}, \quad \left\{\psi_{j,k},\ k = 0, 1, \ldots, 2^j - 1\right\}, \quad j \geq J,$$

form a complete orthonormal system of $L^2\left([0,1]\right)$. The associated reconstruction formula is given by:

$$f(t) = \sum_{k=0}^{2^J-1} \alpha^f_{J,k}\varphi_{J,k}(t) + \sum_{j\geq J}\sum_{k=0}^{2^j-1} \beta^f_{j,k}\psi_{j,k}(t), \quad \forall t \in [0,1], \quad \forall f \in L^2\left([0,1]\right), \tag{A8.49}$$

where

$$\begin{aligned}
\alpha^f_{J,k} &= \int_0^1 f(t)\overline{\varphi_{J,k}(t)}dt, \quad k = 0, \ldots, 2^J - 1, \\
\beta^f_{j,k} &= \int_0^1 f(t)\overline{\psi_{j,k}(t)}dt, \quad k = 0, \ldots, 2^j - 1,\ j \geq J.
\end{aligned}$$

The Besov spaces $B^s_{p,q}([0,1])$ can be characterized in terms of wavelets coefficients. Specifically, denote by $\mathcal{S}'$ the dual of $\mathcal{S}$, the Schwarz space, $f \in \mathcal{S}'$ belongs to $B^s_{p,q}([0,1])$, $s \in \mathbb{R}$, $1 \leq p, q \leq \infty$, if and only if

$$\|f\|_{p,q}^{s} \equiv \|\varphi * f\|_p + \left(\sum_{j=1}^{\infty} \left(2^{js}\|\psi_j * f\|_p\right)^q\right)^{1/q} < \infty. \tag{A8.50}$$

For $\beta > 1/2$, consider $\mathcal{T} : H_2^{-\beta}([0,1]) \longrightarrow H_2^{\beta}([0,1])$ be a self–adjoint positive operator on $L^2([0,1])$, belonging to the unit ball of trace operators on $L^2([0,1])$. Assume that

$$\mathcal{T} : H_2^{-\beta}([0,1]) \longrightarrow H_2^{\beta}([0,1]), \quad \mathcal{T}^{-1} : H_2^{\beta}([0,1]) \longrightarrow H_2^{-\beta}([0,1])$$

are bounded linear operators. In particular, there exists an orthonormal basis $\{v_k,\ k \geq 1\}$ of $L^2([0,1])$ such that, for every $l \geq 1$, $\mathcal{T}(v_l) = t_l v_l$, with $\sum_{l\geq 1} t_l = 1$. In what follows, consider $\{v_l,\ l \geq 1\}$ to be a wavelet basis, and define the kernel t of $\mathcal{T}$ as, for $s,t \in [0,1]$,

$$t(s,t) = \frac{1}{2^J}\sum_{k=0}^{2^{J-1}} \varphi_{J,k}(s)\varphi_{J,k}(t) + \frac{2^{2\beta}-1}{2^{2\beta(1-J)}}\sum_{j\geq J}\sum_{k=0}^{2^j-1} 2^{-2j\beta}\psi_{j,k}(s)\psi_{j,k}(t). \tag{A8.51}$$

In Lemma A8.2.1,

$$\{F_{\mathbf{m}}\} = \{F_{J,k}^{\varphi},\ k = 0,\dots,2^J-1\} \cup \{F_{j,k}^{\psi},\ k=0,\dots,2^j-1,\ j \geq J\}$$

are then defined as follows:

$$\begin{aligned} F_{J,k}^{\varphi} &= \varphi_{J,k}, \quad k = 0,\dots,2^J-1 \\ F_{j,k}^{\psi} &= \psi_{j,k}, \quad k=0,\dots,2^j-1, \quad j \geq J. \end{aligned} \tag{A8.52}$$

Furthermore, the sequence

$$\{t_{\mathbf{m}}\} = \{t_{J,k}^{\varphi},\ k=0,\dots,2^J-1\} \cup \{t_{j,k}^{\psi},\ k=0,\dots,2^j-1,\ j\geq J\},$$

involved in the definition of the inner product in $\widetilde{H}$, is given by:

$$\begin{aligned} t_{J,k}^{\varphi} &= \frac{1}{2^J}, \quad k = 0,\dots,2^{J-1}. \\ t_{j,k}^{\psi} &= \frac{2^{2\beta}-1}{2^{2\beta(1-J)}}2^{-2j\beta}, \quad k=0,\dots,2^{j-1}, \quad j \geq J. \end{aligned} \tag{A8.53}$$

In view of [Angelini et al., 2003, Proposition 2.1], the choice (A8.52)–(A8.53) of $\{F_{\mathbf{m}}\}$ and $\{t_{\mathbf{m}}\}$ leads to the definition of

$$\widetilde{H} = [H_2^{\beta}([0,1])]^* = H_2^{-\beta}([0,1]),$$

constituted by the restriction to $[0,1]$ of the tempered distributions $g \in \mathcal{S}'(\mathbb{R})$, such that $(I-\Delta)^{-\beta/2}g \in L^2(\mathbb{R})$, with $(I-\Delta)^{-\beta/2}$ denoting the Bessel potential of order β (see Triebel [1983]). Let now define $B = B_{\infty,\infty}^0([0,1],)$ and $B^* = B_{1,1}^0([0,1])$. From equation (A8.50), the corresponding norms, in term of the discrete wavelet transform introduced in equation (A8.49), are given by, for every $f \in B$,

$$\|f\|_B = \sup\left\{\left|\alpha_{J,k}^f\right|,\ k=0,\ldots,2^{J-1};\ \left|\beta_{j,k}^f\right|,\ k=0,\ldots,2^j-1;\ j\geq J\right\} \quad \text{(A8.54)}$$

$$\|g\|_{B^*} = \sum_{k=0}^{2^J-1}\left|\alpha_{J,k}^g\right| + \sum_{j=J}^{\infty}\sum_{k=0}^{2^j-1}\left|\beta_{j,k}^g\right|, \quad \forall g \in B^*. \quad \text{(A8.55)}$$

Therefore,

$$B^* = B_{1,1}^0([0,1]) \hookrightarrow H = L^2([0,1]) \hookrightarrow B = B_{\infty,\infty}^0 \hookrightarrow \widetilde{H} = H_2^{-\beta}([0,1]). \quad \text{(A8.56)}$$

Also, for $\beta > 1/2$,

$$\widetilde{H}^* = H^{\beta}([0,1]) \hookrightarrow B^* = B_{1,1}^0([0,1]).$$

For $\gamma > 2\beta$, consider the operator $C = (I-\Delta)^{-\gamma}$; i.e., given by the $2\gamma/\beta$ power of the Bessel potential of order β, restricted to $L^2([0,1])$. From spectral theorems on spectral calculus (see Triebel [1983]), for every $g \in C^{1/2}\left(H^{-\beta}([0,1])\right)$,

$$\begin{aligned}\|g\|_{\mathcal{H}(X)}^2 &= \left\langle C^{-1}(f), f\right\rangle_{H^{-\beta}([0,1])} = \left\langle (I-\Delta)^{-\beta/2}\left(C^{-1}(f)\right), (I-\Delta)^{-\beta/2}(f)\right\rangle_{L^2([0,1])} \\ &= \sum_{j=1}^{\infty} f_j^2\lambda_j\left((I-\Delta)^{(\gamma-\beta)}\right) \geq \sum_{j=1}^{\infty} f_j^2\lambda_j\left((I-\Delta)^{\beta}\right) \\ &= \|f\|_{H^{\beta}([0,1])}^2 = \|f\|_{\widetilde{H}^*}^2,\end{aligned} \quad \text{(A8.57)}$$

where

$$f_j = \int_0^1 (I-\Delta)^{-\beta/2}(f)(s)(I-\Delta)^{-\beta/2}(\phi_j)(s)ds,$$

with $\{\phi_j,\ j\geq 1\}$ denoting the eigenvectors of the Bessel potential $(I-\Delta)^{-\beta/2}$ of order β, restricted to $L^2([0,1])$, and $\{\lambda_j\left((I-\Delta)^{\gamma-\beta}\right),\ j\geq 1\}$ being the eigenvalues of $(I-\Delta)^{-\beta}C^{-1}$ on $L^2([0,1])$. Thus, **Assumption A4** holds. Furthermore, from embedding theorems between fractional Sobolev spaces (see Triebel [1983]), **Assumption A5** also holds, under the condition $\gamma > 2\beta > 1$, considering $H = L^2([0,1])$.

A8.7 FINAL COMMENTS

Appendix A8.6 illustrates the motivation of the presented approach in relation to functional prediction in nuclear spaces. Specifically, the current literature on ARB(1) prediction has been developed for $B = \mathcal{C}[0,1]$, the space of continuous functions on $[0,1]$, with the supremum norm (see, for instance, Álvarez-Liébana et al. [2016]; Bosq [2000]), and $B = \mathcal{D}([0,1])$, constituted by the right–continuous functions on $[0,1]$, having limit to the left at each $t \in [0,1]$, with the Skorokhod topology (see, for example, Hajj [2011]). This paper provides a more flexible framework, where functional prediction can be performed, in a consistent way, for instance, in nuclear spaces, as follows from the continuous inclusions showed in Appendix A8.6.

Note that the two above–referred usual Banach spaces, $\mathcal{C}[0,1]$ and $\mathcal{D}([0,1])$, are included in the Banach space B considered in Appendix A8.6 (see Supplementary Material in Appendix A8.8 about the simulation study undertaken).

A8.8 SUPPLEMENTARY MATERIAL

This document provides the Supplementary Material to the current paper. Specifically, a simulation study is undertaken to illustrate the results derived, on strong consistency of functional predictors, in abstract Banach spaces, from the ARB(1) framework. The results are also illustrated in the case of discretely observed functional data.

A8.8.1 Simulation study

$$\|f\|_B = \sup\left\{\left|\alpha^f_{J,k}\right|,\ k=0,\dots,2^{J-1};\ \left|\beta^f_{j,k}\right|,\ k=0,\dots,2^j-1;\ j=J,\dots,M\right\} \tag{A8.58}$$

where

$$\begin{aligned}
\alpha^f_{J,k} &= \int_0^1 f(t)\overline{\varphi_{J,k}(t)}dt, \quad k=0,\dots,2^J-1,\\
\beta^f_{j,k} &= \int_0^1 f(t)\overline{\psi_{j,k}(t)}dt, \quad k=0,\dots,2^j-1, \quad j\geq J.
\end{aligned}$$

Thus, equation (A8.58) corresponds to the choice $B = B^0_{\infty,\infty}([0,1])$, when resolution level M is fixed for truncation. Therefore, $B^* = B^0_{1,1}([0,1])$ is considered with the truncated norm

$$\|g\|_{B^*} = \sum_{k=0}^{2^J-1}\left|\alpha^g_{J,k}\right| + \sum_{j=J}^{M}\sum_{k=0}^{2^j-1}\left|\beta^g_{j,k}\right|, \quad g\in B^*, \tag{A8.59}$$

where $\{\alpha^g_{J,k}\}$ and $\{\beta^g_{j,k}\}$ are the respective father and mother wavelet coefficients of function g. Furthermore, as given in Appendix A8.6 of the manuscript,

$$\widetilde{H}^* = H_2^{\beta}([0,1]) = B^{\beta}_{2,2}([0,1]), \quad \widetilde{H} = H_2^{-\beta}([0,1]) = B^{-\beta}_{2,2}([0,1]),$$

for $\beta > 1/2$. Since Daubechies wavelets of order $N = 10$ are selected as orthogonal wavelet basis, with

$N = 10$ vanishing moments, according to [Angelini et al., 2003, p. 271 and Lemma 2.1], and [Antoniadis and Sapatinas, 2003, p. 153], we have considered $J = 2$, and $M = \lceil \log_2(L/2) \rceil = 10$, for $L = 2^{11}$ nodes, in the discrete wavelet transform applied. In addition, value $\beta = 6/10 > 1/2$ has been tested, with $\gamma = 2\beta + \epsilon$, $\epsilon = 0.01$ (see definition above of the extended version of operator C on $\widetilde{H} = H^{-\beta}([0,1])$). The covariance kernel is now displayed in Figure A8.8.1 (see [Dautray and Lions, 1990, pp. 119–140] and [Grebenkov and Nguyen, 2013, p. 6]).

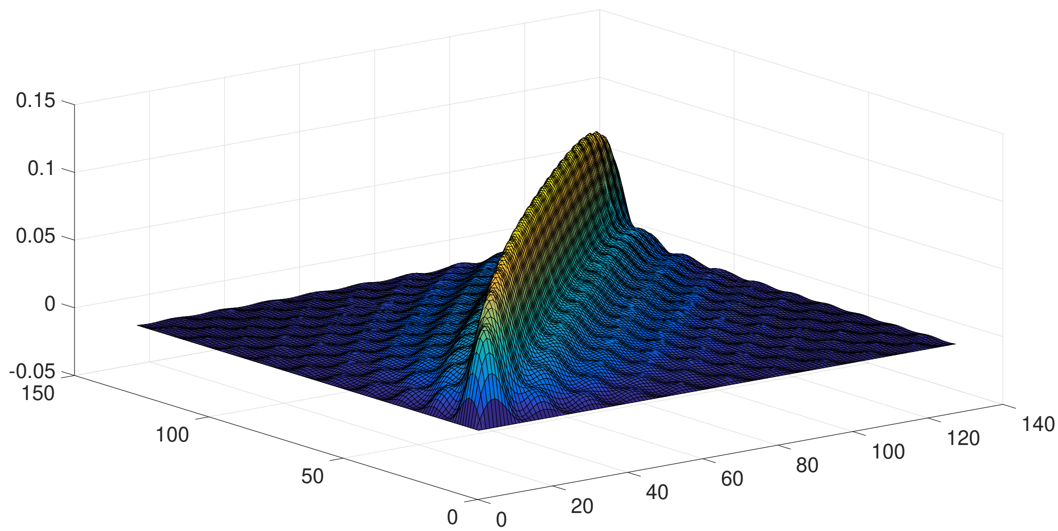


Figure A8.8.1: Covariance kernel defining C, generated with discretization step size $\Delta h = 0.0372$.

Under **Assumption A3**, operator ρ admits the following extended representation in $\widetilde{H} = H^{-\beta}([0,1])$, and in B :

$$\langle \rho(\phi_j), \phi_h \rangle_{H^{-\beta}([0,1])} = \begin{cases} (1+j)^{-1.5} & j = h \\ e^{-|j-h|/W} & j \neq h \end{cases},$$

Operator C_ε also admits, in this case, the following extended version in $\widetilde{H} = H^{-\beta}([0,1])$:

$$\langle C_\varepsilon(\phi_j), \phi_h \rangle_{H^{-\beta}([0,1])} = \begin{cases} C_j \left(1 - \rho_{j,j}^2\right) & j = h \\ e^{-|j-h|^2/W^2} & j \neq h \end{cases},$$

being $W = 0.4$.

A8.8.1.1 Large-sample behaviour of the ARB(1) plug-in predictor

The ARB(1) process is generated with discretization step size $\Delta h = 0.0372$. The resulting functional values of ARB(1) process X are showed in Figure A8.8.2 for sample sizes

$$n_t = [2500, 5000, 15000, 25000, 40000, 55000, 80000, 100000, 130000, 165000]\,.$$

In this section (but not in the next one), the generated discrete values are interpolated and smoothed, applying the *'cubicspline'* option in *'fit.m'* MatLab function, with, as commented before, the number of nodes $L = 2^{11} = 2048$, then $M = 10$, and $\Delta\widetilde{h} = 0.0093$. In the following computations, $N = 250$ replications are generated for each functional sample size, and $k_n = \ln(n)$ has been tested.

The random initial condition X_0 has been generated from a truncated zero–mean Gaussian distribution. Figure A8.8.3 illustrates the fact that **Assumption A1** holds, and Figure A8.8.4 is displayed to check **Assumption A2**.

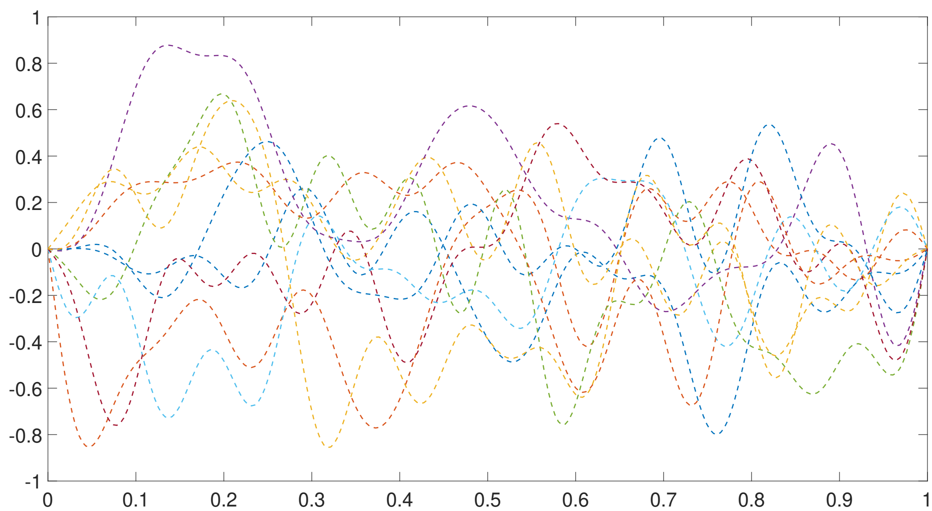


Figure A8.8.2: Functional values X_t, for sample sizes $[2.5, 5, 15, 25, 40, 55, 80, 100, 130, 165] \times 10^3$ and discretization step size $\Delta h = 0.0372$.

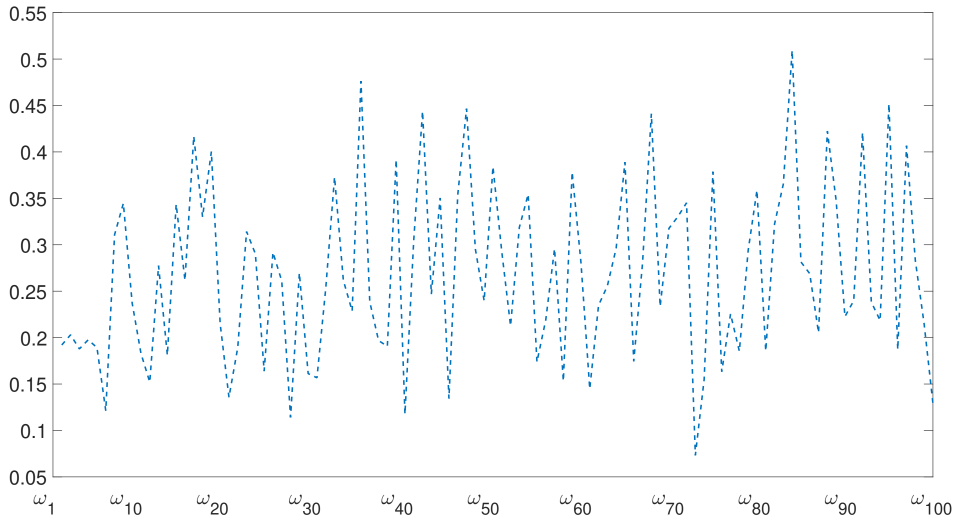


Figure A8.8.3: A set of 100 values of $\|X_0(\omega_l)\|_B$, $l = 1, \ldots, 100$, (blue dotted line) are generated, for discretization step $\Delta h = 0.0372$.

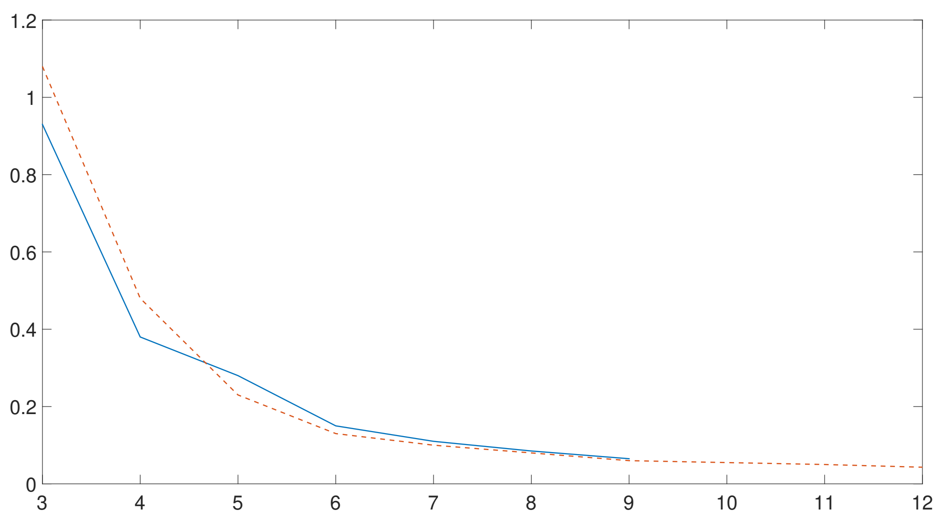


Figure A8.8.4: Assumption A2 is checked for sample sizes $n_t = 35000$ (blue line) and $n_t = 395000$ (orange dotted line), displaying the decay rate of empirical eigenvalues $\{C_{n,j},\ j = 3, \ldots, k_n\}$, being $k_n = \lceil \ln(n) \rceil$.

Condition (A8.38) in Theorem A8.5.1 has been checked as well (see Figure A8.8.5).

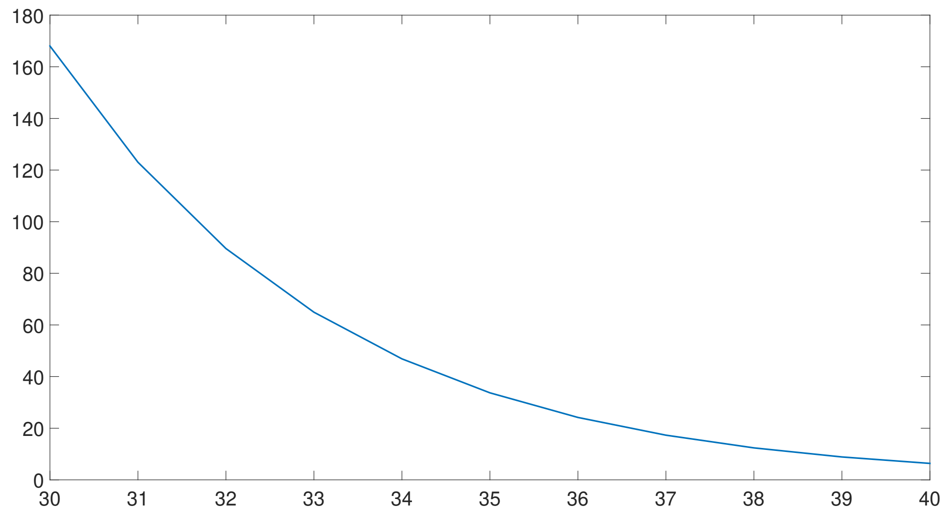


Figure A8.8.5: Values for $\left(k_n C_{k_n}^{-1} \sum_{j=1}^{k_n} a_j\right) \left(n^{1/2} (\ln(n))^{-1/2}\right)^{-1}$, tested for truncation parameters $k_n = 30, \ldots, 40$, linked to sample sizes by the truncation rule $k_n = \ln(n)$.

To illustrate Theorem A8.5.1 and Corollary A8.5.1, Table A8.8.1 displays the proportion of values of the random variable $\left\| \rho(X_{n_t}) - \widehat{X}_{n_t+1} \right\|_B$ that are larger than the upper bound

$$\xi_{n_t} = \exp\left(\frac{-n_t}{C_{k_{n_t}}^{-2} k_{n_t}^2 \left(\sum_{j=1}^{k_{n_t}} a_j \right)^2} \right), \quad t = 1, \ldots, 10, \tag{A8.60}$$

from the 250 values generated, for each functional sample size n_t, $t = 1, \ldots, 10$, reflected below.

Table A8.8.1: Proportion of simulations whose error B-norm is larger than the upper bound in equation (A8.60). Truncation parameter $k_n = \ln(n)$ and $N = 250$ realizations have been considered, for each functional sample size.

n_t	
$n_1 = 2500$	$\frac{13}{250}$
$n_2 = 5000$	$\frac{11}{250}$
$n_3 = 15000$	$\frac{7}{250}$
$n_4 = 25000$	$\frac{4}{250}$
$n_5 = 40000$	$\frac{2}{250}$
$n_6 = 55000$	$\frac{1}{250}$
$n_7 = 80000$	0
$n_8 = 100000$	$\frac{1}{250}$
$n_9 = 130000$	0
$n_{10} = 165000$	0

Figure A8.8.6 below illustrates the asymptotic efficiency. The curve $n^{-1/4}$ is also displayed (red dotted line).

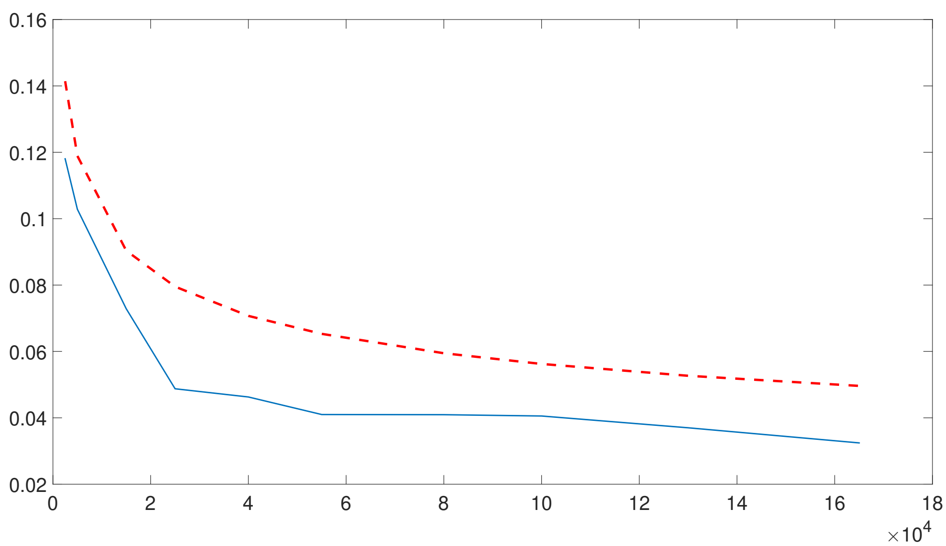


Figure A8.8.6: Asymptotic efficiency. Empirical mean-square error (blue solid line) $\mathrm{E}\left\{\left\|\rho\left(X_{n_t}\right) - \widehat{X}_{n_t+1}\right\|_B^2\right\}$, based on $N = 250$ simulations. The curve $n^{-1/4}$ is also drawn (red dotted line).

A8.8.1.2 Asymptotic behaviour of discretely observed ARB(1) processes

The results in Theorem A8.5.1 and Corollary A8.5.1 are now tested for different discretization step sizes:

$$\left\{\Delta h_r = \left(2^{8+r} - 1\right)^{-1}, \, r = 1, \dots, 7\right\}, \quad \Delta h_r \longrightarrow^{r \to \infty} 0,$$

that is,

$$\begin{aligned} \Delta h_1 &= 1.96 \left(10^{-3}\right), \quad \Delta h_2 = 9.78 \left(10^{-4}\right), \\ \Delta h_3 &= 4.89 \left(10^{-4}\right), \quad \Delta h_4 = 2.44 \left(10^{-4}\right), \\ \Delta h_5 &= 1.22 \left(10^{-4}\right), \quad \Delta h_6 = 6.10 \left(10^{-5}\right), \\ \Delta h_7 &= 3.06 \left(10^{-5}\right). \end{aligned}$$

Due to computational limitations involved in the smallest discretization step sizes, we restrict our attention here to the sample sizes

$$\{n_t = 5000 + 10000 \, (t - 1) \, , \, t = 1, 2, 3\} \, ,$$

and $N = 120$ realizations have been generated, for each functional sample size. The same nodes are considered as in the previous section, in the implementation of the discrete wavelet transform, without previous smoothing of the discretely generated data.

Table A8.8.2 displays the results obtained on the proportion of values, from the 120 generated values,

$$\left\| \rho \left(X_{n_t}^{h,r}\right) - \widehat{X}_{n_t+1}^{h,r} \right\|_B, \quad h = 1, \dots, 120,$$

that are larger than the upper bound (A8.60), considering different discretization step sizes, for each sample size

$$\{n_t = 5000 + 10000 \, (t - 1) \, , \, t = 1, 2, 3\} \, ,$$

and for the corresponding truncation orders $\{k_{n_t} = \ln(n_t), \, t = 1, 2, 3\}$.

Table A8.8.2: Proportions of simulations whose error B-norms are larger than the upper bound in (A8.60), for sample sizes $n = [5000, 15000, 35000]$. Truncation parameter $k_n = \ln(n)$ has been considered. For each one of the functional sample sizes, the results displayed correspond to discretization step sizes $\left\{\Delta h_r = \left(2^{8+r} - 1\right)^{-1}, \ r = 1, \ldots, 7\right\}$. We have generated $N = 120$ simulations, for each sample and discretization step size.

	$n_1 = 5000$	$n_2 = 15000$	$n_3 = 35000$
$\Delta h_1 = 1.96\left(10^{-3}\right)$	$\frac{12}{120}$	$\frac{7}{120}$	$\frac{6}{120}$
$\Delta h_2 = 9.78\left(10^{-4}\right)$	$\frac{8}{120}$	$\frac{4}{120}$	$\frac{4}{120}$
$\Delta h_3 = 4.89\left(10^{-4}\right)$	$\frac{4}{120}$	$\frac{2}{120}$	$\frac{2}{120}$
$\Delta h_4 = 2.44\left(10^{-4}\right)$	$\frac{2}{120}$	$\frac{1}{120}$	$\frac{1}{120}$
$\Delta h_5 = 1.22\left(10^{-4}\right)$	$\frac{2}{120}$	$\frac{1}{120}$	0
$\Delta h_6 = 6.10\left(10^{-5}\right)$	$\frac{1}{120}$	0	0
$\Delta h_7 = 3.06\left(10^{-5}\right)$	$\frac{1}{120}$	0	0

Acknowledgments

This work has been supported in part by project MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

References

[Aach and Church 2001] Aach, J. ; Church, G. M.: Alignment gene expression time series with time warping algorithms. *Bioinformatics* 17 (2001), pp. 495–508. – URL https://www.ncbi.nlm.nih.gov/pubmed/11395426

[Abdel-Aal and Al-Garni 1997] Abdel-Aal, R. E. ; Al-Garni, A. Z.: Forecasting monthly electric energy consumption in Eastern Saudi Arabia using univariate time–series analysis. *Energy* 22 (1997), pp. 1059–1069. – DOI: doi.org/10.1016/S0360-5442(97)00032-7

[Abraham et al. 2003] Abraham, C. ; Cornillon, P. A. ; Matzner-Løber, E. ; Molinari, N.: Unsupervised curve clustering using B-splines. *Scand. J. Statist.* 30 (2003), pp. 581–595. – DOI: doi.org/10.1111/1467-9469.00350

[Abramovich and Angelini 2006] Abramovich, F. ; Angelini, C.: Testing in mixed-effects FANOVA models. *J. Statist. Plann. Inference* 136 (2006), pp. 4326–4348. – DOI: doi.org/10.1016/j.jspi.2005.06.002

[Abramovich et al. 2004] Abramovich, F. ; Antoniadis, A. ; Sapatinas, T. ; Vidakovic, B.: Optimal testing in functional analysis of variance models. *Int. J. Wavelets Multiresolut. Inf. Process.* 2 (2004), pp. 323–349. – URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.140.6784&rep=rep1&type=pdf

[Albaqshi 2017] Albaqshi, A. M. H.: *Generalized Partial Least Squares Approach for Nominal Multinomial Logit Regression Models with a Functional Covariate.* University of Northern Colorado, College of Education and Behavioral Sciences, PhD. Thesis, 2017. – URL https://digscholarship.unco.edu/cgi/viewcontent.cgi?article=1432&context=dissertations

[Allam and Mourid 2001] Allam, A. ; Mourid, T.: Propriétés de mélanges des processus autorégressifs Banachiques. *C. R. Acad. Sci. Paris Sér. I* 333 (2001), pp. 363–368. – DOI: doi.org/10.1016/S0764-4442(01)02031-6

[Allam and Mourid 2014] Allam, A. ; Mourid, T.: Covariance operator estimation of a functional autoregressive process with random coefficients. *Statist. Probab. Lett.* 84 (2014), pp. 1–8. – DOI: doi.org/10.1016/j.spl.2013.09.018

[Álvarez-Liébana 2017] Álvarez-Liébana, J.: Functional time series: a review and comparative study. *Submitted* (2017)

[Álvarez-Liébana et al. 2016] Álvarez-Liébana, J. ; Bosq, D. ; Ruiz-Medina, M. D.: Consistency of the plug-in functional predictor of the Ornstein-Uhlenbeck in Hilbert and Banach spaces. *Statist. Probab. Lett.* 117 (2016), pp. 12–22. – DOI: doi.org/10.1016/j.spl.2016.04.023

[Álvarez-Liébana et al. 2017] Álvarez-Liébana, J. ; Bosq, D. ; Ruiz-Medina, M. D.: Asymptotic properties of a componentwise ARH(1) plug-in predictor. *J. Multivariate Anal.* 155 (2017), pp. 12–34. – DOI: doi.org/10.1016/j.jmva.2016.11.009

[Álvarez-Liébana and Ruiz-Medina 2015] Álvarez-Liébana, J. ; Ruiz-Medina, M. D.: Functional statistical classification of non–linear dynamics and random surfaces roughness in control systems. *Int. J. Math. Models Methods Appl. Sciences* 9 (2015), pp. 1–20. – URL http://www.naun.org/main/NAUN/ijmmas/2015/a022001-354.pdf

[Álvarez-Liébana and Ruiz-Medina 2017] Álvarez-Liébana, J. ; Ruiz-Medina, M. D.: The effect of the spatial domain in FANOVA models with ARH(1) term. *Stat. Interface* 10 (2017), pp. 607–628. – DOI: doi.org/10.4310/SII.2017.v10.n4.a7

[Anderson 2003] Anderson, T. W.: *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 2003. – ISBN 9780471360919

[Aneiros and Vieu 2017] Aneiros, G. ; Vieu, P.: Sparse nonparametric model for regression with functional covariate. *J. Nonparametr. Stat.* 28 (2017), pp. 839–859. – DOI: doi.org/10.1080/10485252.2016.1234050

[Aneiros-Pérez et al. 2011] Aneiros-Pérez, G. ; Cao, R. ; Vilar-Fernández, J. M.: Functional methods for time series prediction: a nonparametric approach. *J. Forecasting* 30 (2011), pp. 377–392. – DOI: doi.org/10.1002/for.1169

[Aneiros-Pérez and Vieu 2006] Aneiros-Pérez, G. ; Vieu, P.: Semi-functional partial linear regression. *Statist. Probab. Lett.* 76 (2006), pp. 1102–1110. – DOI: doi.org/10.1016/j.spl.2005.12.007

[Aneiros-Pérez and Vieu 2008] Aneiros-Pérez, G. ; Vieu, P.: Nonparametric time series prediction: a semifunctional partial linear modeling. *J. Multivariate Anal.* 99 (2008), pp. 834–857. – DOI: doi.org/10.1016/j.jmva.2007.04.010

[Angelini et al. 2003] Angelini, C. ; Canditiis, D. D. ; Leblanc, F.: Wavelet regression estimation in nonparametric mixed effect models. *J. Multivariate Anal.* 85 (2003), pp. 267–291. – DOI: doi.org/10.1016/S0047-259X(02)00055-6

[Anh et al. 2016a] Anh, V. V. ; Leonenko, N. N. ; Ruiz-Medina, M. D.: Space-time fractional stochastic equations on regular bounded open domains. *Fract. Calc. Appl. Anal.* 19 (2016a), pp. 1161–1199. – DOI: doi.org/10.1515/fca-2016-0061

[Anh et al. 2016b] Anh, V. V. ; Leonenko, N. N. ; Ruiz-Medina, M. D.: Fractional-in-time and multifractional-in-space stochastic partial differential equations. *Fract. Calc. Appl. Anal.* 19 (2016b), pp. 1434–1459. – DOI: doi.org/10.1515/fca-2016-0074

[Antoniadis et al. 2012] Antoniadis, A. ; Brosat, X. ; Cugliari, J. ; Poggi, J. M.: Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. *J. SFdS* 153 (2012), pp. 52–78. – URL https://tel.archives-ouvertes.fr/tel-00647334

[Antoniadis et al. 2006] Antoniadis, A. ; Paparoditis, E. ; Sapatinas, T.: A functional wavelet-kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 68 (2006), pp. 837–857. – DOI: doi.org/10.1111/j.1467-9868.2006.00569.x

[Antoniadis and Sapatinas 2003] Antoniadis, A. ; Sapatinas, T.: Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* 87 (2003), pp. 133–158. – DOI: doi.org/10.1016/S0047-259X(03)00028-9

[Antoniadis and Sapatinas 2007] Antoniadis, A. ; Sapatinas, T.: Estimation and inference in functional mixed-effects models. *Comput. Statist. Data Anal.* 51 (2007), pp. 4793–4813. – DOI: doi.org/10.1016/j.csda.2006.09.038

[Aroian 1947] Aroian, L. A.: The probability function of a product of two normal distributed variables. *Ann. Math. Statist.* 18 (1947), pp. 256–271. – DOI: doi.org/10.1214/aoms/1177730442

[Ash and Gardner 1975] Ash, R. B. ; Gardner, M. F.: Topics in stochastic processes. *Bull. Amer. Math. Soc.* 82 (1975), pp. 817–820. – URL https://www.sciencedirect.com/science/book/9780120652709

[Aston and Kirch 2011a] Aston, J. A. D. ; Kirch, C.: Detecting and estimating epidemic changes in dependent functional data. *CRiSM Research Report, University of Warwick* (2011a)

[Aston and Kirch 2011b] Aston, J. A. D. ; Kirch, C.: Estimation of the distribution of change–points swith application to fMRI data. *CRiSM Research Reports, University of Warwick* (2011b)

[Aue et al. 2015] Aue, A. ; Norinho, D. ; Hörmann, S.: On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* 110 (2015), pp. 378–392. – DOI: doi.org/10.1080/01621459.2014.909317

[Baíllo et al. 2011] Baíllo, A. ; Cuesta-Albertos, J. A. ; Cuevas, A.: Supervised classification for a family of Gaussian functional models. *Scand. J. Statist.* 38 (2011), pp. 480–498. – DOI: doi.org/10.1111/j.1467-9469.2011.00734.x

[Banerjee et al. 2004] Banerjee, S. ; Carlin, B. P. ; Gelfand, A. E.: *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall, London, 2004. – ISBN 9781584884101

[Bartlett 1946] Bartlett, M. S.: On the theoretical specification and sampling properties of autocorrelated time series. *Supplement to J. Roy. Stat. Soc.* 8 (1946), pp. 27–41. – URL http://www.jstor.org/stable/2983611

[Bensmain and Mourid 2001] Bensmain, N. ; Mourid, T.: Estimateur "sieve" de l'opérateur d'un processus ARH(1). *C. R. Acad. Sci. Paris Sér. I Math.* 332 (2001), pp. 1015–1018. – DOI: doi.org/10.1016/S0764-4442(01)01954-1

[Benyelles and Mourid 2001] Benyelles, W. ; Mourid, T.: Estimation de la période d'un processus à temps continu à représentation autorégressive. *C. R. Acad. Sci. Paris Sér. I Math.* 333 (2001), pp. 245–248. – DOI: doi.org/10.1016/S0764-4442(01)02030-4

[Berkes et al. 2009] Berkes, I. ; Gabrys, R. ; Horváth, L. ; Kokoszka, P.: Detecting changes in the mean of functional observations. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* 71 (2009), pp. 927–946. – DOI: doi.org/10.1111/j.1467-9868.2009.00713.x

[Berkes et al. 2011] Berkes, I. ; Hörmann, S. ; Schauer, J.: Split invariance principles for stationary processes. *Ann. Probab.* 39 (2011), pp. 2441–2473. – DOI: doi.org/10.1214/10-AOP603

[Berlinet et al. 2008] Berlinet, A. ; Biau, G. ; Rouvière, L.: Functional supervised classification with wavelets. *Ann. I. S. U. P.* 52 (2008), pp. 61–80. – URL https://hal.archives-ouvertes.fr/hal-00459437

[Bernard 1997] Bernard, P.: Analyse de signaux physiologiques. *Mémorie Univ. Cathol. Angers.* (1997)

[Berrendero et al. 2018a] Berrendero, J. R. ; Bueno-Larraz, B. ; Cuevas, A.: A RKHS model for variable selection in functional linear regression. *J. Multivariate Anal. (in press)* (2018a). – DOI: doi.org/10.1016/j.jmva.2018.04.008

[Berrendero et al. 2018b] Berrendero, J. R. ; Bueno-Larraz, B. ; Cuevas, A.: On Mahalanobis distance in functional settings (submitted). *arXiv:1803.06550* (2018b). – URL https://arxiv.org/abs/1803.06550

[Besag 1986] Besag, J.: On the statistical analysis of dirty pictures. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 48 (1986), pp. 259–302. – URL www.jstor.org/stable/2345426

[Besse and Cardot 1996] Besse, P. C. ; Cardot, H.: Approximation spline de la prévision d'un processu fonctionnel autoregréssif d'ordre 1. *Canad. J. Statist.* 24 (1996), pp. 467–487. – DOI: doi.org/10.2307/3315328

[Besse et al. 2000] Besse, P. C. ; Cardot, H. ; Stephenson, D. B.: Autoregressive forecasting of some functional climatic variations. *Scand. J. Statist.* 27 (2000), pp. 673–687. – DOI: doi.org/10.1111/1467-9469.00215

[Biau et al. 2003] Biau, G. ; Bunea, F. ; Wegkamp, M. H.: Functional classification in Hilbert spaces. *IEEE Trans. Inform. Theory* 1 (2003), pp. 1–8. – URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1435658

[Biau et al. 2010] Biau, G. ; Cérou, F. ; Guyader, A.: Rates of convergence of the functional k–nearest neighbor estimate. *IEEE Trans. Inform. Theory* 56 (2010), pp. 2034–2040. – URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5437442

[Billingsley 1995] Billingsley, P.: *Probability and Measure, 3nd edition*. John Wiley & Sons, Chicago, 1995. – ISBN 9780471007104

[Blanke and Bosq 2014] Blanke, D. ; Bosq, D.: Exponential bounds for intensity of jumps. *Math. Methods Statist.* 23 (2014), pp. 239–255. – DOI: doi.org/10.3103/S1066530714040012

[Blanke and Bosq 2015] Blanke, D. ; Bosq, D.: Bayesian prediction for stochastic processes: Theory and applications. *Sankhya A* 77 (2015), pp. 79–105. – DOI: doi.org/10.1007/s13171-014-0059-y

[Blanke and Bosq 2016] Blanke, D. ; Bosq, D.: Detecting and estimating intensity of jumps for discretely observed ARMAD(1,1) processes. *J. Multivariate Anal.* 146 (2016), pp. 119–137. – DOI: 10.1016/j.jmva.2015.08.014

[Boente and Fraiman 2000] Boente, G. ; Fraiman, R.: Kernel–based functional principal components. *Statist. Prob. Lett.* 48 (2000), pp. 335–345. – DOI: doi.org/10.1016/S0167-7152(00)00014-6

[Bohorquez et al. 2017] Bohorquez, M. ; Giraldo, R. ; Mateu, J.: Multivariate functional random fields: prediction and optimal sampling. *Stoch. Environ. Res. Risk Assess.* 31 (2017), pp. 53–70. – DOI: doi.org/10.1007/s00477-016-1266-y

[Bongiorno et al. 2014] Bongiorno, G. ; Goia, A. ; Salinelli, E. ; Vieu, P.: *Contributions in infinite–dimensional statistics and related topics*. In: Soc. Editrice Esculapio, Bologna, 2014. – ISBN 9788874887637

[de Boor 1978] Boor, C. de: *A practical guide to splines*. Springer, New York, 1978. – ISBN 9780387953663

[Bosq 1991] Bosq, D.: Modelization, non-parametric estimation and prediction for continuous time processes. *Nonparametric functional estimation and related topics, NATO, ASI Series* 335 (1991), pp. 509–529

[Bosq 1996] Bosq, D.: Limit theorems for Banach-valued autoregressive processes. Applications to real continuous time processes. *Bull. Belg. Math. Soc. Simon Stevin* 3 (1996), pp. 537–555. – URL https://projecteuclid.org/download/pdf_1/euclid.bbms/1105652783

[Bosq 1999a] Bosq, D.: Autoregressive representation for the empirical covariance operator of an ARH(1). *C. R. Acad. Sci. Paris Sér. I Math.* 329 (1999a), pp. 531–534

[Bosq 1999b] Bosq, D.: Autoregressive Hilbertian processes. *Ann. I. S. U. P.* 43 (1999b), pp. 25–55

[Bosq 2000] Bosq, D.: *Linear Processes in Function Spaces*. Springer, New York, 2000. – ISBN 9781461211549

[Bosq 2002] Bosq, D.: Estimation of mean and covariance operator of autoregressive processes in Banach spaces. *Stat. Inference Stoch. Process.* 5 (2002), pp. 287–306. – DOI: doi.org/10.1023/A:1021279131053

[Bosq 2004] Bosq, D.: Standard Hilbert moving averages. *Ann. I. S. U. P.* 48 (2004), pp. 17–28

[Bosq 2007] Bosq, D.: General linear processes in Hilbert spaces and prediction. *J. Statist. Plann. Inference* 137 (2007), pp. 879–894. – DOI: doi.org/10.1016/j.jspi.2006.06.014

[Bosq and Blanke 2007] Bosq, D. ; Blanke, D.: *Inference and predictions in large dimensions*. Wiley, 2007. – ISBN 9780470017616

[Bosq and Ruiz-Medina 2014] Bosq, D. ; Ruiz-Medina, M. D.: Bayesian estimation in a high dimensional parameter framework. *Electron. J. Stat.* 8 (2014), pp. 1604–1640. – DOI: doi.org/10.1214/14-EJS935

[Bouveyron 2004] Bouveyron, C.: Statistique en grande dimension: problématiques et enjeux. *J. SFdS* 155 (2004), pp. 36–37. – URL http://publications-sfds.math.cnrs.fr/index.php/J-SFdS/article/view/275/255

[Brillinger 1981] Brillinger, D. R.: *Time Series Analysis: Data Analysis and Theory*. Holt, Rinehart & Winston, New York, 1981. – ISBN 9780898715019

[Brockwell and Davis 1987] Brockwell, P. J. ; Davis, R. A.: *Time Series: Theory and Methods, 2nd ed.* Springer, New York, 1987. – ISBN 9781441903198

[Bueno-Larraz and Klepsch 2018] Bueno-Larraz, B. ; Klepsch, J.: Variable selection for the prediction of $C[0,1]$–valued AR processes using RKHS. *Technometrics (under minor revision)* (2018)

[Bühlmann and de Geer 2011] Bühlmann, P. ; Geer, S. V. de: *Statistics for high–dimensional data*. Methods, Theory and Applications, in: Springer Series in Statistics, Springer, Heildeberg, 2011. – ISBN 9783642201929

[Bulaevskaya and Oehlert 2007] Bulaevskaya, V. L. ; Oehlert, G. W.: A penalized likelihood approach to magnetic resonance image reconstruction. *Stat. Med.* 26 (2007), pp. 352–374. – DOI: doi.org/10.1002/sim.2545

[Bunea et al. 2006] Bunea, F. ; Wegkamp, M. ; Auguste, A.: Consistent variable selection in high dimensional regression via multiple testing. *J. Statist. Plann. Inference* 136 (2006), pp. 4349–4364. – DOI: doi.org/10.1016/j.jspi.2005.03.011

[Burba et al. 2009] Burba, F. ; Ferraty, F. ; Vieu, P.: k–Nearest Neighbour method in functional nonparametric regression. *J. Nonparametr. Statist.* 21 (2009), pp. 453–469. – DOI: doi.org/10.1080/10485250802668909

[Burfield et al. 2015] Burfield, R. ; Neumann, C. ; Saunders, C. P.: Review and application of functional data analysis to chemical data. *Chem. Intell. Lab. Systems* 149 (2015), pp. 97–106. – DOI: doi.org/10.1016/j.chemolab.2015.07.006

[Burkardt 2011] Burkardt, J.: MATLAB Source Codes. (2011). – URL http://people.sc.fsu.edu/jburkardt

[Cai and Hall 2006] Cai, T. ; Hall, P.: Prediction in functional linear regression. *Ann. Statist.* 34 (2006), pp. 2159–2179. – DOI: doi.org/10.1214/009053606000000830

[Canale and Ruggiero 2016] Canale, A. ; Ruggiero, M.: Bayesian nonparametric forecasting of monotonic functional time series. *Electron. J. Statist.* 10 (2016), pp. 3265–3286. – DOI: doi.org/10.1214/16-EJS1190

[Cardot 1998] Cardot, H.: Convergence du lissage spline de la prévision des processus autorégressifs fonctionnels. *C. R. Acad. Sci. Paris Sér. I Math.* 326 (1998), pp. 755–758

[Cardot et al. 2007] Cardot, H. ; Crambes, C. ; Kneip, A. ; Sarda, P.: Smoothing splines estimators in functional linear regression with errors–in–variables. *Comput. Statist. Data Anal.* 51 (2007), pp. 4832–4848. – DOI: doi.org/10.1016/j.csda.2006.07.029

[Cardot et al. 2003] Cardot, H. ; Ferraty, F. ; Mas, A. ; Sarda, P.: Testing hypotheses in the functional linear model. *Scand. J. Statist.* 30 (2003), pp. 241–255. – URL www.jstor.org/stable/4616760

[Cardot et al. 1999] Cardot, H. ; Ferraty, F. ; Sarda, P.: Functional linear model. *Statist. Prob. Lett.* 45 (1999), pp. 11–22. – URL http://cardot.perso.math.cnrs.fr/CFS99.pdf

[Cardot and Sarda 2011] Cardot, H. ; Sarda, P.: Functional linear regression. *In: Ferraty, F., Romain, Y. (Eds.) The Oxford Handbook of Functional Data Analysis. Oxford University Press, Oxford* (2011), pp. 21–46

[de Castro et al. 2005] Castro, B. M. F. de ; González-Manteiga, W. ; Guillas, S.: Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics* 2 (2005), pp. 212–222. – URL http://www.jstor.org/stable/25470983

[Cavallini et al. 1994] Cavallini, A. ; Montanari, G. C. ; Logginiand, M. ; Lessi, O. ; Cacciari, M.: Nonparametric prediction of harmonic levels in electrical networks. *Proc. IEEE* (1994), pp. 165–171. – DOI: doi.org/10.1109/59.575735

[Chaouch et al. 2017] Chaouch, M. ; Laib, N. ; Louani, D.: Rate of uniform consistency for a class of mode regression on functional stationary ergodic data. *Stat. Methods Appl.* 26 (2017), pp. 19–47. – DOI: doi.org/10.1007/s10260-016-0356-9

[Chen et al. 2011] Chen, D. ; Hall, P. ; Müller, H. G.: Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* 39 (2011), pp. 1720–1747. – DOI: doi.org/10.1214/11-AOS882

[Chen and Müller 2012] Chen, D. ; Müller, H. G.: Nonlinear manifold representations for functional data. *Ann. Statist.* 40 (2012), pp. 1–29. – DOI: doi.org/10.1214/11-AOS936

[Chen et al. 2012] Chen, J. ; Lin, D. ; Hochner, H.: Semiparametric maximum likelihood methods for analyzing genetic and environmental effects with case-control motherchild pair data. *Biometrics* 68 (2012), pp. 869–877. – DOI: doi.org/10.1111/j.1541-0420.2011.01728.x

[Chen et al. 2016] Chen, S. X. ; Lei, L. ; Tu, Y.: Functional coefficient moving average model with applications to forecasting chinese CPI. *Statist. Sinica* 26 (2016), pp. 1649–1672. – URL https://mpra.ub.uni-muenchen.de/67074/

[Chiou et al. 2004] Chiou, J. M. ; Müller, H. G. ; Wang, J. L.: Functional response models. *Statist. Sinica* 14 (2004), pp. 675–693. – URL http://anson.ucdavis.edu/~mueller/fdr9-15.pdf

[Chiou et al. 2003] Chiou, J. M. ; Müller, H. G. ; Wang, J. L. ; Carey, J. R.: A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statist. Sinica* 13 (2003), pp. 1119–1133. – URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2597815/

[Christensen and Yetkin 2005] Christensen, W. F. ; Yetkin, F. Z.: Spatio-temporal analysis of auditory cortex activation as detected with silent event related fMRI. *Stat. Med.* 24 (2005), pp. 2539–2556. – DOI: doi.org/10.1002/sim.2111

[Chung et al. 2005] Chung, K. K. ; Robbins, S. M. ; Dalton, K. M. ; Davidson, R. J. ; Alexander, A. L. ; Evans, A. C.: Cortical thickness analysis in autism with heat. *Neuroimage* 25 (2005), pp. 1256–1265. – DOI: doi.org/10.1016/j.neuroimage.2004.12.052

[Cohen et al. 1993] Cohen, A. ; Daubechies, I. ; Vial, P.: Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harm. Anal.* 1 (1993), pp. 54–81. – DOI: doi.org/10.1006/acha.1993.1005

[Collazos et al. 2016] Collazos, J. A. A. ; Dias, R. ; Zambomb, A. Z.: Consistent variable selection for functional regression models. *J. Multivariate Anal.* 146 (2016), pp. 63–71. – DOI: doi.org/10.1016/j.jmva.2015.06.007

[Crambes et al. 2009] Crambes, C. ; Kneip, A. ; Sarda, P.: Smoothing splines estimators for functional linear regression. *Ann. Statist.* 37 (2009), pp. 35–72. – DOI: doi.org/10.1214/07-AOS563

[Cuesta-Albertos and Febrero-Bande 2010] Cuesta-Albertos, J. A. ; Febrero-Bande, M. F.: Multiway ANOVA for functional data. *Test* 19 (2010), pp. 537–557. – DOI: doi.org/10.1007/s11749-010-0185-3

[Cuesta-Albertos et al. 2007] Cuesta-Albertos, J. A. ; Fraiman, R. ; Ransford, T.: A sharp form of the Cramér–Wold theorem. *J. Theoret. Probab.* 20 (2007), pp. 201–209. – DOI: doi.org/10.1007/s10959-007-0060-7

[Cuesta-Albertos and Nieto-Reyes 2008] Cuesta-Albertos, J. A. ; Nieto-Reyes, A.: The random Tukey depth. *Comput. Statist. Data Anal.* 52 (2008), pp. 4979–4988. – DOI: doi.org/10.1016/j.csda.2008.04.021

[Cuevas 2014] Cuevas, A.: A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference* 147 (2014), pp. 1–23. – DOI: doi.org/10.1016/j.jspi.2013.04.002

[Cuevas et al. 2002] Cuevas, A. ; Febrero, M. ; Fraiman, R.: Linear functional regression: the case of fixed design and functional response. *Canad. J. Statist.* 30 (2002), pp. 285–300. – DOI: doi.org/10.2307/3315952

[Cuevas et al. 2004] Cuevas, A. ; Febrero-Bande, M. F. ; Fraiman, R.: An ANOVA test for functional data. *Comput. Statist. Data Anal.* 47 (2004), pp. 111–122. – DOI: doi.org/10.1016/j.csda.2003.10.021

[Cuevas and Fraiman 2009] Cuevas, A. ; Fraiman, R.: On depth measures and dual statistics. A methodology for dealing with general data. *J. Multvariate Anal.* 100 (2009), pp. 753–766. – DOI: doi.org/10.1016/j.jmva.2008.08.002

[Cugliari 2011] Cugliari, J.: *Prévision non paramétrique de processus à valeurs fonctionnelles. Application à la consommation d'électricité*, University of Paris-Sud 11, PhD. Thesis, 2011. – URL https://tel.archives-ouvertes.fr/tel-00647334

[Cugliari 2013] Cugliari, J.: Conditional autoregressive Hilbertian processes. *arXiv:1302.3488.* (2013)

[Damon and Guillas 2002] Damon, J. ; Guillas, S.: The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* 13 (2002), pp. 759–774. – DOI: doi.org/10.1002/env.527

[Damon and Guillas 2005] Damon, J. ; Guillas, S.: Estimation and simulation of autoregressie Hilbertian processes with exogenous variables. *Stat. Inference Stoch. Process.* 8 (2005), pp. 185–204. – DOI: doi.org/10.1007/s11203-004-1031-6

[Daubechies 1988] Daubechies, I.: Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.* 41 (1988), pp. 909–996. – DOI: doi.org/10.1002/cpa.3160410705

[Dautray and Lions 1990] Dautray, R. ; Lions, J.-L.: *Mathematical Analysis and Numerical Methods for Science and Technology Volume 3: Spectral Theory and Applications*. Springer, New York, 1990. – ISBN 9783642615290

[Davis et al. 2008] Davis, R. A. ; Lee, T. C. M. ; Rodríguez-Yam, G. A.: Break detection for a class of non–linear time series models. *J. Time Series Anal.* 29 (2008), pp. 834–867. – DOI: doi.org/10.1111/j.1467-9892.2008.00585.x

[Davis and Mikoschb 2008] Davis, R. A. ; Mikoschb, T.: Extreme value theory for space–time processes with heavy-tailed distributions. *Stochastic Process. Appl.* 118 (2008), pp. 560–584. – DOI: doi.org/10.1016/j.spa.2007.06.001

[Dedecker and Merlevède 2003] Dedecker, J. ; Merlevède, F.: The conditional central limit theorem in Hilbert spaces. *Stochastic Process. Appl.* 108 (2003), pp. 229–262. – DOI: doi.org/10.1016/j.spa.2003.07.004

[Dehling and Sharipov 2005] Dehling, H. ; Sharipov, O. S.: Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations. *Stat. Inference Stoch. Process.* 8 (2005), pp. 137–149. – DOI: doi.org/10.1007/s11203-003-0382-8

[Delaigle and Hall 2012] Delaigle, A. ; Hall, P.: Methodology and theory for partial least squares applied to functional data. *Ann. Statist.* 40 (2012), pp. 322–352. – DOI: doi.org/10.1214/11-AOS958

[Delicado et al. 2010] Delicado, P. ; Giraldo, R. ; Comas, C. ; Mateu, J.: Statistics for spatial functional data: some recent contributions. *Environmetrics* 21 (2010), pp. 224–239. – DOI: doi.org/10.1002/env.1003

[Delzell et al. 2012] Delzell, D. A. P. ; Gunst, R. F. ; Schucany, W. R. ; Carmack, P. S. ; Lin, Q. ; Spence, J. S. ; Haley, R. W.: Key properties of D-optimal designs for event-related functional MRI experiments with applications to nonlinear models. *Stat. Med.* 31 (2012), pp. 3907–3920. – DOI: doi.org/10.1002/sim.5449

[Dette and Derbort 2001] Dette, H. ; Derbort, S.: Analysis of variance in nonparametric regression models. *J. Multivariate Anal.* 76 (2001), pp. 110–137. – DOI: doi.org/10.1006/jmva.2000.1913

[Diaconis and Freeman 1999] Diaconis, P. ; Freeman, D.: Iterated random functions. *SIAM Rev.* 41 (1999), pp. 45–76. – DOI: doi.org/10.1137/S0036144598338446

[Didericksen and Kokoszka 2012] Didericksen, D. ; Kokoszka, P.: Empirical properties of forecast with the functional autoregressive model. *Comput. Statist.* 27 (2012), pp. 285–298. – DOI: doi.org/10.1007/s00180-011-0256-2

[Diebold and Li 2006] Diebold, F. X. ; Li, C.: Forecasting the term structure of government bond yields. *J. Econometrics* 130 (2006), pp. 337–364. – DOI: doi.org/10.1016/j.jeconom.2005.03.005

[Doob 1942] Doob, J. L.: The Brownian movement and stochastic equations. *Ann. Math.* 43 (1942), pp. 319–337. – URL https://www.jstor.org/stable/pdf/1968873.pdf

[Elbert 2001] Elbert, A.: Some recent results on the zeros of Bessel functions and orthogonal polynomials. *J. Comput. Appl. Math.* 133 (2001), pp. 65–83. – DOI: doi.org/10.1016/S0377-0427(00)00635-X

[Escabias et al. 2014] Escabias, M. ; Aguilera, A. M. ; Aguilera-Morillo, M. C.: Functional PCA and Base–Line Logit Models. *J. Classification* 31 (2014), pp. 296–324. – DOI: doi.org/10.1007/s00357-014-9162-y

[Ezzahrioui and Ould-Saïd 2008] Ezzahrioui, M. ; Ould-Saïd, E.: Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.* 20 (2008), pp. 3–18. – DOI: doi.org/10.1080/10485250701541454

[Ezzahrioui and Ould-Saïd 2010] Ezzahrioui, M. ; Ould-Saïd, E.: Some asymptotic results of a non–parametric conditional mode estimator for functional time–series data. *Statist. Neerlandica* 64 (2010), pp. 171–201. – DOI: doi.org/10.1111/j.1467-9574.2010.00449.x

[Fahrmeir and Kneib 2011] Fahrmeir, L. ; Kneib, T.: *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford University Press, New York, 2011. – ISBN 9780199533022

[Fan 1996] Fan, J.: Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* 91 (1996), pp. 674–688. – DOI: doi.org/10.2307/2291663

[Fan and Li 2001] Fan, J. ; Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96 (2001), pp. 1348–1360. – URL http://www.jstor.org/stable/3085904

[Febrero-Bande et al. 2008] Febrero-Bande, M. ; Galeano, P. ; González-Manteiga, W.: Outlier detection in functional data by depth measures with application to identify abnormal NO_x levels. *Environmetrics* 19 (2008), pp. 331–345. – DOI: doi.org/10.1002/env.878

[Febrero-Bande et al. 2017] Febrero-Bande, M. ; Galeano, P. ; González-Manteiga, W.: Functional Principal Component Regression and Functional Partial Least Squares Regression: an overview and a comparative study. *Int. Statist. Review* 85 (2017), pp. 61–83. – DOI: doi.org/10.1111/insr.12116

[Febrero-Bande and González-Manteiga 2013] Febrero-Bande, M. ; González-Manteiga, W.: Generalized additive models for functional data. *Test* 22 (2013), pp. 278–292. – DOI: doi.org/10.1007/s11749-012-0308-0

[Ferraty et al. 2013a] Ferraty, F. ; Goia, A. ; Salinelli, E. ; Vieu, P.: Functional projection pursuit regression. *Test* 22 (2013), pp. 293–320. – DOI: doi.org/10.1007/s11749-012-0306-2

[Ferraty et al. 2002] Ferraty, F. ; Goia, A. ; Vieu, P.: Functional nonparametric model for time series: a fractal approach for dimension reduction. *Test* 11 (2002), pp. 317–344. – DOI: doi.org/10.1007/BF02595710

[Ferraty et al. 2012] Ferraty, F. ; Keilegom, I. V. ; Vieu, P.: Regression when both response and predictor are functions. *J. Multivariate Anal.* 109 (2012), pp. 10–28. – DOI: doi.org/10.1016/j.jmva.2012.02.008

[Ferraty et al. 2010] Ferraty, F. ; Laksaci, A. ; Tadj, A. ; Vieu, P.: Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference* 140 (2010), pp. 335–352. – DOI: doi.org/10.1016/j.jspi.2009.07.019

[Ferraty et al. 2011] Ferraty, F. ; Laksaci, A. ; Tadj, A. ; Vieu, P.: Kernel regression with functional response. *Electron. J. Statist.* 5 (2011), pp. 159–171. – DOI: doi.org/10.1214/11-EJS600

[Ferraty et al. 2003] Ferraty, F. ; Peuch, A. ; Vieu, P.: Modélé á indice fonctionnel simple. *C. R. Acad. Sci. Paris Sér. I* 336 (2003), pp. 1025–1028. – DOI: doi.org/10.1016/S1631-073X(03)00239-5

[Ferraty et al. 2013b] Ferraty, F. ; Sued, M. ; Vieu, P.: Mean estimation with data missing at random for functional covariables. *Statistics* 47 (2013), pp. 688–706. – DOI: doi.org/10.1080/02331888.2011.650172

[Ferraty and Vieu 2006] Ferraty, F. ; Vieu, P.: *Nonparametric functional data analysis: theory and practice*. Springer, 2006. – ISBN 9780387303697

[Ferraty and Vieu 2009] Ferraty, F. ; Vieu, P.: Additive prediction and boosting for functional data. *Comput. Statist. Data Anal.* 53 (2009), pp. 1400–1413. – DOI: doi.org/10.1016/j.csda.2008.11.023

[Ferraty and Vieu 2011] Ferraty, F. ; Vieu, P.: Kernel regression estimation for functional data. *In: Ferraty, F., Romain, Y.(Eds.) The Oxford Handbook of Functional Data Analysis. Oxford University Press, Oxford* (2011), pp. 72–129. – URL https://hal.archives-ouvertes.fr/hal-00794727

[Fischer et al. 2017] Fischer, J. H. ; Zhang, Q. ; Zhu, Y. ; Weiss, R. E.: Functional time series models for ultrafine particle distributions. *Ann. Appl. Stat.* 11 (2017), pp. 297–319. – DOI: doi.org/10.1214/16-AOAS1004

[Fraiman and Muniz 2001] Fraiman, R. ; Muniz, G.: Trimmed means for functional data. *Test* 10 (2001), pp. 419–440. – DOI: doi.org/10.1007/BF02595706

[Frías et al. 2017] Frías, M. P. ; Ivanov, A. V. ; Leonenko, N. N. ; Martínez, F. ; Ruiz-Medina, M. D.: Detecting hidden periodicities for models with cyclical errors. *Stat. Interface* 10 (2017), pp. 107–118. – DOI: doi.org/10.4310/SII.2017.v10.n1.a10

[Friston et al. 1998] Friston, K. J. ; Fletcher, P. ; Josephs, O. ; Holmes, A. P. ; Rugg, M. D. ; Turner, R.: Event-related fMRI: Characterising differential responses. *Neuroimage* 7 (1998), pp. 30–40. – DOI: doi.org/10.1006/nimg.1997.0306

[Friston et al. 1995] Friston, K. J. ; Holmes, A. P. ; Worsley, K. J. ; Poline, J. B. ; Frith, C. D. ; Frackowiak, R. S. J.: Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* 2 (1995), pp. 189––210. – DOI: doi.org/10.1002/hbm.460020402

[Fryba 1999] Fryba, L.: *Vibration of solids and structures under moving loads*. Thomas Telford Ltd., London, 1999. – ISBN 9789401196857

[Gabrys et al. 2010] Gabrys, R. ; Horváth, L. ; Kokoszka, P.: Tests for error correlation in the functional linear model. *J. Am. Stat. Assoc.* 105 (2010), pp. 1113–1125. – DOI: doi.org/10.1198/jasa.2010.tm09794

[Gabrys and Kokoszka 2007] Gabrys, R. ; Kokoszka, P.: Portmanteau test of independence for functional observations. *J. Am. Stat. Assoc.* 102 (2007), pp. 1338–1348. – DOI: doi.org/10.1198/016214507000001111

[Ge and Fan 2013] Ge, M. ; Fan, L.: Learning Optimal Kernel for Pattern Classification. *WSEAS Transactions on Mathematics* 5 (2013), pp. 491–500. – URL http://www.wseas.org/multimedia/journals/mathematics/2013/56-367.pdf

[Geenens 2011] Geenens, G.: Curse of dimensionality and related issues in nonparametric functional regression. *Statistics Surveys* 5 (2011), pp. 30–43. – DOI: doi.org/10.1214/09-SS049

[Gerstner 2007] Gerstner, T.: *Sparse Grid Quadrature Methods for Comp. Finance*. University of Bonn, lecture notes, 2007. – URL http://wissrech.iam.uni-bonn.de/research/pub/gerstner/gerstner_habil.pdf

[Gerstner and Griebel 1998] Gerstner, T. ; Griebel, M.: Numerical Integration using Sparse Grids. *Numer. Algorithms* 18 (1998), pp. 209–232. – DOI: doi.org/10.1023/A:1019129717644

[Ghosh and Kaabouch 2014] Ghosh, D. ; Kaabouch, N.: A Survey on Remote Sensing Scene Classification Algorithms. *WSEAS Transactions on Signal Processing* 10 (2014), pp. 504–519. – URL https://www.researchgate.net/publication/268388228_A_Survey_on_Remote_Sensing_Scene_Classification_Algorithms

[Giraldo et al. 2010] Giraldo, R. ; Delicado, P. ; Mateu, J.: Geostatistics for functional data: an ordinary kriging approach. *Environ. Ecol. Stat.* 18 (2010), pp. 411–426. – URL https://upcommons.upc.edu/bitstream/handle/2117/1099/mateu-EES.pdf

[Glendinning and Fleet 2007] Glendinning, R. ; Fleet, S.: Classifying functional time series. *Signal Process.* 87 (2007), pp. 79–100. – DOI: doi.org/10.1016/j.sigpro.2006.04.006

[Glover 1999] Glover, G. H.: Deconvolution of impulse response in event-related bold fMRI. *Neuroimage* 9 (1999), pp. 416–429. – DOI: doi.org/10.1006/nimg.1998.0419

[Goia and Vieu 2014] Goia, A. ; Vieu, P.: Some advances in semiparametric functional data modelling. *In: Contributions in infinite–dimensional statistics and related topics, Esculapio, Bologna* (2014), pp. 135–141. – URL https://www.researchgate.net/publication/303112040_Some_advances_in_semiparametric_functional_data_modelling

[Goia and Vieu 2015] Goia, A. ; Vieu, P.: A Partitioned Single Functional Index model. *Comput. Statist.* 30 (2015), pp. 673–692. – DOI: doi.org/10.1007/s00180-014-0530-1

[Goia and Vieu 2016] Goia, A. ; Vieu, P.: An introduction to recent advances in high/infinite dimensional statistics. *J. Multivariate Anal.* 146 (2016), pp. 1–6. – DOI: doi.org/10.1016/j.jmva.2015.12.001

[Gorenflo and Mainardi 2003] Gorenflo, R. ; Mainardi, F.: Fractional diffusion processes: probability distribution and continuous time random walk. *Lecture Notes in Physics* 621 (2003), pp. 148–166. – DOI: doi.org/10.1007/3-540-44832-2_8

[Graczyk and Jakubowski 2006] Graczyk, P. ; Jakubowski, T.: Analysis of Ornstein-Uhlenbeck and Laguerre stochastic processes. *École CIMPA Familles orthogonales et semigroupes en analyse et probabilités* (2006). – URL http://math.univ-angers.fr/publications/prepub/fichiers/00229.pdf

[Grebenkov and Nguyen 2013] Grebenkov, D. S. ; Nguyen, B. T.: Geometrical structure of Laplacian eigenfunctions. *SIAM Rev.* 55 (2013), pp. 601–667. – DOI: doi.org/10.1137/120880173

[Grivas and Chaloulakou 2006] Grivas, G. ; Chaloulakou, A.: Artificial neural network models for prediction of PM10 hourly concentrations in the greater area of Athens, Greece. *Atmospheric Environment* 40 (2006), pp. 1216–1229. – DOI: doi.org/10.1016/j.atmosenv.2005.10.036

[Grollemund et al. 2018] Grollemund, P. M. ; Abraham, C. ; Baragatti, M. ; Pudlo, P.: Bayesian Functional Linear Regression with Sparse Step Functions. *Bayesian Anal. (to appear)* (2018). – DOI: doi.org/10.1214/18-BA1095

[Gruen et al. 2017] Gruen, M. E. ; Alfaro-Córdoba, M. ; Thomson, A. E. ; Worth, A.C. ; Staicu, A. M. ; Lascelles, B. D. X.: The Use of Functional Data Analysis to Evaluate Activity in a Spontaneous Model of Degenerative Joint Disease Associated Pain in Cats. *PLoS ONE* 12 (2017). – DOI: doi.org/10.1371/journal.pone.0169576

[Gu 2002] Gu, C.: *Smoothing Spline ANOVA Models.* Springer, New York, 2002. – ISBN 9781461453697

[Guillas 2000] Guillas, S.: Non-causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1). *C. R. Acad. Sci. Paris Sér. I* 331 (2000), pp. 91–94

[Guillas 2001] Guillas, S.: Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. *Statist. Probab. Lett.* 55 (2001), pp. 281–291. – DOI: doi.org/10.1016/S0167-7152(01)00151-1

[Guillas 2002] Guillas, S.: Doubly stochastic Hilbertian processes. *J. Appl. Probab* 39 (2002), pp. 566–580. – DOI: doi.org/10.1239/jap/1034082128

[Guo 2002] Guo, W.: Inference in smoothing spline analysis of variance. *J. R. Stat. Soc. Ser. B Stat. Methodol* 64 (2002), pp. 887–898. – DOI: doi.org/10.1111/1467-9868.00367

[Gurland 1956] Gurland, J.: Quadratic forms in normally distributed random variables. *Sankhya A* 17 (1956), pp. 37–50. – URL http://www.jstor.org/stable/25048285

[Haan and Lin 2001] Haan, L. D. ; Lin, T.: On convergence toward an extreme value distribution in $\mathcal{C}([0,1])$. *Ann. Probab.* 29 (2001), pp. 467–483. – DOI: doi.org/10.1214/aop/1008956340

[Hajj 2011] Hajj, L. E.: Limit theorems for $D[0,1]$-valued autoregressive processes. *C. R. Acad. Sci. Paris Sér. I Math.* 349 (2011), pp. 821–825

[Hajj 2013] Hajj, L. E.: *Inférence statistique pour des variables fonctionnelles á sauts*. Paris, University Paris VI, PhD. Thesis, 2013

[Hall and Horowitz 2007] Hall, P. ; Horowitz, J. L.: Methodology and convergence rates for functional linear regression. *Ann. Statist.* 35 (2007), pp. 70–91. – DOI: doi.org/10.1214/009053606000000957

[Hall and Hosseini-Nasab 2006] Hall, P. ; Hosseini-Nasab, M.: On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 68 (2006), pp. 109–126. – DOI: doi.org/10.1111/j.1467-9868.2005.00535.x

[Hall et al. 2001] Hall, P. ; Poskitt, D. ; Presnell, B.: A functional data–analytic approach to signal discrimination. *Technometrics* 43 (2001), pp. 1–9. – DOI: doi.org/10.1198/00401700152404273

[Hamilton 1994] Hamilton, J. D.: *Time Series Analysis*. Princeton University Press, Princenton, 1994. – ISBN 9780691042893

[Härdle et al. 2004] Härdle, W. ; Müller, M. ; Sperlich, S. ; Werwatz, A.: *Nonparametric and semiparametric models*. Springer–Verlag, Berlin, 2004

[Hidalgo and Ruiz-Medina 2012] Hidalgo, M. M. ; Ruiz-Medina, M. D.: Local wavelet–vaguelette–based functional classification of gene expression data. *Biom. J.* 54 (2012), pp. 75–93. – DOI: doi.org/10.1002/bimj.201000135

[Hörmann 2008] Hörmann, S.: Augmented GARCH sequences: dependence structure and asymptotics. *Bernoulli* 14 (2008), pp. 543–561. – DOI: doi.org/10.3150/07-BEJ120

[Hörmann et al. 2013] Hörmann, S. ; Horváth, L. ; Reeder, R.: A functional version of the ARCH model. *Econometric Theory.* 29 (2013), pp. 267–288. – DOI: doi.org/10.1017/S0266466612000345

[Hörmann and Kidziński 2015] Hörmann, S. ; Kidziński, L.: A note on estimation in Hilbertian linear models. *Scand. J. Statist.* 42 (2015), pp. 43–62. – DOI: doi.org/10.1111/sjos.12094

[Hörmann et al. 2015] Hörmann, S. ; Kidziński, L. ; Hallin, M.: Dynamic functional principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 77 (2015), pp. 319–348. – DOI: doi.org/10.1111/rssb.12076

[Hörmann and Kokoszka 2010] Hörmann, S. ; Kokoszka, P.: Weakly dependent functional data. *Ann. Statist.* 38 (2010), pp. 1845–1884. – DOI: doi.org/10.1214/09-AOS768

[Hörmann and Kokoszka 2011] Hörmann, S. ; Kokoszka, P.: Consistency of the mean and the principal components of spatially distributed functional data. *In: Recent advances in functional data analysis and related topics. Contrib. Statist. Physica-Verlag/Springer, Heidelberg* (2011), pp. 169–175. – DOI: doi.org/10.1016/j.jmaa.2016.12.037

[Hörmann et al. 2016] Hörmann, S. ; Kokoszka, P. ; Nisol, G.: Detection of periodicity in functional time series. *Ann. Statist. (submitted)* (2016). – URL https://arxiv.org/abs/1607.02017

[Horváth et al. 2010] Horváth, L. ; Huskova, M. ; Kokoszka, P.: Testing the stability of the functional autoregressive process. *J. Multivariate Anal.* 101 (2010), pp. 352–367. – DOI: doi.org/10.1016/j.jmva.2008.12.008

[Horváth and Kokoszka 2012] Horváth, L. ; Kokoszka, P.: *Inference for functional data with applications*. Springer, New York, 2012. – ISBN 9781461436553

[Horváth et al. 2014] Horváth, L. ; Kokoszka, P. ; Rice, G.: Testing stationarity of functional time series. *J. Econometrics* 179 (2014), pp. 66–82. – DOI: doi.org/10.1016/j.jeconom.2013.11.002

[Horváth and Reeder 2012] Horváth, L. ; Reeder, R.: Detecting changes in functional linear models. *J. Multivariate Anal.* 111 (2012), pp. 310–334. – DOI: doi.org/10.1016/j.jmva.2012.04.007

[Hsing and Eubank 2015] Hsing, T. ; Eubank, R.: *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, 2015. – ISBN 9781118762547

[Huang 1998] Huang, J. Z.: Projection estimation in multiple regression with application to functional ANOVA models. *Ann. Statist.* 26 (1998), pp. 242–272. – DOI: doi.org/10.1214/aos/1030563984

[Hyndman and Shang 2009] Hyndman, R. H. ; Shang, H. L.: Forecasting functional time series. *J. Korean Statist. Soc.* 38 (2009), pp. 199–211

[Hyndman and Shang 2008] Hyndman, R. J. ; Shang, H. L.: Bagplots, Boxplots and Outlier Detection for Functional Data. *In: Functional and Operatorial Statistics, Physica-Verlag* (2008)

[Hyndman and Ullah 2007] Hyndman, R. J. ; Ullah, Md. S.: Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Statist. Data Anal.* 51 (2007), pp. 4942–4956

[Ignaccolo et al. 2014] Ignaccolo, R. ; Mateu, J. ; Giraldo, R.: Kriging with external drift for functional data for air quality monitoring. *Stoch. Environ. Res. Risk. Assess.* 28 (2014), pp. 1171–1186. – DOI: doi.org/10.1007/s00477-013-0806-y

[Izenman 2008] Izenman, A.: *Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning*. Springer, New York, 2008. – ISBN 9780387781891

[Jackson 2004] Jackson, J. E.: *A user's guide to Principal Components*. Wiley, New York, 2004. – ISBN 9780471622673

[James and Hastie 2001] James, G. M. ; Hastie, T. J.: Functional linear discriminant analysis for irregular sampled curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63 (2001), pp. 533–550. – DOI: doi.org/10.1111/1467-9868.00297

[James and Sugar 2003] James, G. M. ; Sugar, C. A.: Clustering for sparsely sampled functional data. *J. Amer. Statist. Soc.* 98 (2003), pp. 397–408. – URL http://www.jstor.org/stable/30045249

[Jiang 2012] Jiang, H.: Berry-Essen bounds and the law of the iterated logarithm for estimators for parameters in an Ornstein-Uhlenbeck process with linear drift. *J. Appl. Probab.* 49 (2012), pp. 978–989. – DOI: doi.org/10.1239/jap/1354716652

[Jiang 2016] Jiang, Y.: An exponential-squared estimator in the autoregressive model with heavy-tailed errors. *Stat. Interface* 9 (2016), pp. 233–238. – DOI: doi.org/10.4310/SII.2016.v9.n2.a10

[Kaarnioja 2013] Kaarnioja, V.: *Smolyak Quadrature (Master's Thesis)*. Helsinki, Department of Mathematics and Statistics. University of Helsinki, PhD. Thesis, 2013. – URL https://helda.helsinki.fi/bitstream/handle/10138/40159/thesis.pdf

[Kadanoff 2000] Kadanoff, L. P.: *Statistical Physics: Statics, Dynamics and Renormalization*. World Scientific, Singapur, 2000. – ISBN 9789810237585

[Kara et al. 2017a] Kara, L. Z. ; Laksaci, A. ; Rachdi, M. ; Vieu, P.: Data–driven kNN estimation in nonparametric functional data analysis. *J. Multivariate Anal.* 153 (2017a), pp. 176–188. – DOI: doi.org/10.1016/j.jmva.2016.09.016

[Kara et al. 2017b] Kara, L. Z. ; Laksaci, A. ; Rachdi, M. ; Vieu, P.: Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. Nonparametr. Stat.* 29 (2017b), pp. 85–107. – DOI: doi.org/10.1080/10485252.2016.1254780

[Kara-Terki and Mourid 2016] Kara-Terki, N. ; Mourid, T.: Local asymptotic normality of Hilbertian autoregressive processes. *C. R. Acad. Sci. Paris Sér. I* 354 (2016), pp. 634–638. – DOI: doi.org/10.1016/j.crma.2016.03.006

[Kargin and Onatski 2008] Kargin, V. ; Onatski, A.: Curve forecasting by functional autoregression. *J. Multivariate Anal.* 99 (2008), pp. 2508–2526. – DOI: doi.org/10.1016/j.jmva.2008.03.001

[Kaufman and Sain 2010] Kaufman, C. G. ; Sain, S. R.: Bayesian functional ANOVA modeling using Gaussian processes prior distributions. *Bayesian Anal.* 5 (2010), pp. 123–150. – DOI: doi.org/10.1214/10-BA505

[Kaziska 2011] Kaziska, D. M.: Functional analysis of variance, discriminant analysis, and clustering in a manifold of elastic curves. *Communications in Statistics Theory and Methods* 40 (2011), pp. 2487–2499. – DOI: doi.org/10.1080/03610926.2010.484159

[Khoshnevisan 2007] Khoshnevisan, D.: *Probability*. American Mathematical Society, UE, 2007. – ISBN 9780821842157

[Kidziński et al. 2016] Kidziński, L. ; Kokoszka, P. ; Jouzdani, N. M.: Principal component analysis of periodically correlated functional time series (submitted). (2016). – URL https://arxiv.org/abs/1612.00040

[Kleptsyna and Breton 2002] Kleptsyna, M. L. ; Breton, A. L.: Statistical analysis of the fractional Ornstein–Uhlenbeck type process. *Stat. Inference Stoch. Process.* 5 (2002), pp. 229–248. – DOI: doi.org/10.1023/A:1021220818545

[Kloeden and Platen 1992] Kloeden, P. E. ; Platen, E.: *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992. – ISBN 9783662126165

[Kokoszka et al. 2008] Kokoszka, P. ; Maslova, I. ; Sojka, J. ; Zhu, L.: Testing for lack of dependence in the functional linear model. *Canad. J. Statist.* 36 (2008), pp. 207–222. – DOI: doi.org/10.1002/cjs.5550360203

[Kokoszka and Reimherr 2013a] Kokoszka, P. ; Reimherr, M.: Asymptotic normality of the principal components of functional time series. *Stochastic Process. Appl.* 123 (2013a), pp. 1546–1562. – DOI: doi.org/10.1016/j.spa.2012.12.011

[Kokoszka and Reimherr 2013b] Kokoszka, P. ; Reimherr, M.: Determining the order of the functional autoregressive model. *J. Time Ser. Anal.* 34 (2013b), pp. 116–129. – DOI: doi.org/10.1111/j.1467-9892.2012.00816.x

[Kowal 2017] Kowal, D. R.: *Bayesian methods for functional and time series data*, Faculty of the Graduate School, PhD. Thesis, 2017. – URL https://static1.squarespace.com/static/5805533c03596e738b64cc5d/t/5ab90152575d1fb8a0bead5c/1522073958415/Thesis_Kowal.pdf

[Kowal et al. 2017] Kowal, D. R. ; Matteson, D. S. ; Ruppert, D.: Functional autoregression for sparsely sample data. *J. Bus. Econom. Statist.* (2017), pp. 1–13. – DOI: doi.org/10.1080/07350015.2017.1279058

[Kuelbs 1970] Kuelbs, J.: Gaussian measures on a Banach spaces. *J. Funct. Anal.* 5 (1970), pp. 354–367. – DOI: doi.org/10.1016/0022-1236(70)90014-5

[Kuhnt and Rehage 2016] Kuhnt, S. ; Rehage, A.: An angle–based multivariate functional pseudo–depth for shape outlier detection. *J. Multivariate Anal.* 146 (2016), pp. 325–340. – DOI: doi.org/10.1016/j.jmva.2015.10.016

[Kutoyants 2004] Kutoyants, Y.: *Statistical Inference for Ergodic Diffusion Processes*. Springer Series in Statistics, London, 2004. – ISBN 9781447138662

[Kyriazis and Petrushev 2001] Kyriazis, G. ; Petrushev, P.: New bases for Triebel-Lizorkin and Besov spaces. *Transactions of the American Mathematical Society* 354 (2001), pp. 749–776. – URL http://www.ams.org/journals/tran/2002-354-02/S0002-9947-01-02916-6/S0002-9947-01-02916-6.pdf

[Labbas and Mourid 2002] Labbas, A. ; Mourid, T.: Estimation et prévision d'un processus autorégressif Banach. *C. R. Acad. Sci. Paris Sér. I* 335 (2002), pp. 767–772. – DOI: doi.org/10.1016/S1631-073X(02)02544-X

[Laïb 2005] Laïb, N.: Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. *J. Statist. Plann. Inference* 134 (2005), pp. 116–139. – DOI: doi.org/10.1016/j.jspi.2004.01.019

[Laïb and Louani 2010] Laïb, N. ; Louani, D.: Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J. Multivariate Anal.* 101 (2010), pp. 2266–2281. – DOI: doi.org/10.1016/j.jmva.2010.05.010

[Laïb and Louani 2011] Laïb, N. ; Louani, D.: Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. *J. Statist. Plann. Inference* 141 (2011), pp. 116–139. – DOI: doi.org/10.1016/j.jspi.2004.01.019

[Laloë 2008] Laloë, T.: A k–nearest neighbor approach for functional regression. *Statist. Prob. Lett.* 78 (2008), pp. 1189–1193. – DOI: doi.org/10.1016/j.spl.2007.11.014

[Lange and Zeger 1997] Lange, N. ; Zeger, S. L.: Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with Discussion). *J. Appl. Statist.* 46 (1997), pp. 1––29. – DOI: doi.org/10.1111/1467-9876.00046

[Laukaitis 2008] Laukaitis, A.: Functional data analysis for cash flow and transactions intensity continuous-time prediction using Hilbert-valued autoregressive processes. *European J. Oper. Res.* 185 (2008), pp. 1607–1614. – DOI: doi.org/10.1016/j.ejor.2006.08.030

[Laukaitis and Rackauskas 2002] Laukaitis, A. ; Rackauskas, A.: Functional data analysis of payment systems. *Nonlinear Anal. Model. Control* 7 (2002), pp. 53–68. – URL https://www.mii.lt/na/issues/NA_0702/NA07205.pdf

[Laukaitis and Vasilecas 2009] Laukaitis, A. ; Vasilecas, O.: Estimation of the autoregressive operator by wavelet packets. *Statist. Probab. Lett.* 79 (2009), pp. 38–43. – DOI: doi.org/10.1016/j.spl.2008.07.011

[Ledoux and Talagrand 2011] Ledoux, M. ; Talagrand, M.: *Probability in Banach spaces*. Springer-Verlag, Berlin, 2011. – ISBN 9783642202124

[Lee et al. 2017] Lee, J. S. ; Zakeri, I. F. ; Butte, N. F.: Functional data analysis of sleeping energy expenditure. *PLoS ONE* 12 (2017). – DOI: doi.org/10.1371/journal.pone.0177286

[Leng and Müller 2006] Leng, X. ; Müller, H. G.: Time ordering of gene coexpression. *Biostatistics* 7 (2006), pp. 569–584. – DOI: doi.org/10.1093/biostatistics/kxj026

[Lerch and Evans 2005] Lerch, J. ; Evans, A. C.: Cortical thickness analysis examined through power analysis and a population simulation. *Neuroimage* 24 (2005), pp. 163–173. – DOI: doi.org/10.1016/j.neuroimage.2004.07.045

[Li and Hsing 2007] Li, Y. ; Hsing, T.: On rates of convergence in functional linear regression. *J. Multivariate Anal.* 98 (2007), pp. 1782–1804. – DOI: doi.org/10.1016/j.jmva.2006.10.004

[Li et al. 2010] Li, Y. M. ; Wang, N. ; Carroll, R. J.: Generalized functional linear models with semiparametric single–index interactions. *J. Amer. Statist. Assoc.* 105 (2010), pp. 621–633. – DOI: doi.org/10.1198/jasa.2010.tm09313

[Li et al. 2011] Li, Y. M. ; Zhu, H. T. ; Shen, D. G. ; Lin, W. L. ; Gilmore, J. ; Ibrahim, J. G.: Multiscale adaptive regression models for neuroimaging data. *J. R. Stat. Soc. Ser. B Stat. Methodol* 73 (2011), pp. 559–578. – DOI: doi.org/10.1111/j.1467-9868.2010.00767.x

[Lian et al. 2016] Lian, H. ; Choi, T. ; Meng, J. ; Jo, S.: Posterior convergence for Bayesian functional linear regression. *J. Multivariate Anal.* 150 (2016), pp. 27–41. – DOI: doi.org/10.1016/j.jmva.2016.04.008

[Liao et al. 2012] Liao, C. ; Worsley, K. J. ; Poline, J. B. ; Aston, J. A. D. ; Duncan, G. H. ; Evans, A. C.: Estimating the delay of the fMRI response. *Neuroimage* 16 (2012), pp. 593–606. – DOI: doi.org/10.1006/nimg.2002.1096

[Lin 2000] Lin, Y.: Tensor product space ANOVA models. *Ann. Statist.* 28 (2000), pp. 734–755. – DOI: doi.org/10.1214/aos/1015951996

[Ling et al. 2017] Ling, N. ; Liu, Y. ; Vieu, P.: On asymptotic properties of functional conditional mode estimation with both stationary ergodic and responses MAR. *In: Aneiros G., G. Bongiorno E., Cao R., Vieu P. (eds) Functional Statistics and Related Fields. Contributions to Statistics* (2017), pp. 173–178. – DOI: doi.org/10.1007/978-3-319-55846-2_23

[Liptser and Shiraev 2001] Liptser, R. S. ; Shiraev, A. N.: *Statistics of Random Processes I, II*. Springer, New York, 2001. – ISBN 9783662100288

[Liu et al. 2016] Liu, X. ; Xiao, H. ; Chen, R.: Convolutional autoregressive models for functional time series. *J. Econometrics* 194 (2016), pp. 263–282. – DOI: doi.org/10.1016/j.jeconom.2016.05.006

[Liu and Müller 2003] Liu, X. L. ; Müller, H. G.: Modes and clustering for time–warped gene expression profile data. *Bioinformatics* 19 (2003), pp. 1937–1944. – URL https://www.ncbi.nlm.nih.gov/pubmed/14555627

[López-Pintado and Romo 2006] López-Pintado, S. ; Romo, J.: Depth based classification for functional data. *In: DIMACS Series in Discrete Mathematics* 72 (2006), pp. 103–130. – URL https://e-archivo.uc3m.es/bitstream/handle/10016/231/ws055611.pdf?sequence=1

[López-Pintado and Romo 2009] López-Pintado, S. ; Romo, J.: On the concept of depth for functional data. *J. Amer. Statist. Assoc.* 104 (2009), pp. 486–503. – DOI: doi.org/10.1198/jasa.2009.0108

[Mallat 1989] Mallat, S. G.: A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Inteligence* 11 (1989), pp. 674–693. – URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=192463

[Mallat 2009] Mallat, S. G.: *A Wavelet Tour of Signal Processing: The Sparse Way, 3rd ed.* Academic Press, Burlington, MA, 2009. – ISBN 9780123743701

[Marion and Pumo 2004] Marion, J. M. ; Pumo, B.: Comparison of ARH(1) and ARHD(1) models on physiological data. *Ann. I.S.U.P.* 48 (2004), pp. 29–38. – URL https://www.researchgate.net/publication/288849889_Comparaison_des_modeles_ARH1_et_ARHD1_sur_des_donnees_physiologiques

[Mas 1999] Mas, A.: Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1). *C. R. Acad. Sci. Paris Sér. I Math.* 329 (1999), pp. 899–902. – DOI: doi.org/10.1016/S0764-4442(00)87496-0

[Mas 2000] Mas, A.: *Estimation d'opérateurs de corrélation de processus fonctionnels: lois limites, tests, déviations modérées*. Paris, Université de Paris 6, PhD. Thesis, 2000

[Mas 2002] Mas, A.: Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Process. Appl.* 99 (2002), pp. 117–135. – DOI: doi.org/10.1016/S0304-4149(02)00087-X

[Mas 2004] Mas, A.: Consistance du prédicteur dans le modéle ARH(1): le cas compact. *Ann. I.S.U.P.* 48 (2004), pp. 39–48. – URL http://www.math.univ-montp2.fr/~mas/JIsup2.pdf

[Mas 2007] Mas, A.: Weak-convergence in the functional autoregressive model. *J. Multivariate Anal.* 98 (2007), pp. 1231–1261. – DOI: doi.org/10.1016/j.jmva.2006.05.010

[Mas and Menneteau 2003a] Mas, A. ; Menneteau, L.: Large and moderate deviations for infinite dimensional autoregressive processes. *J. Multivariate Anal.* 87 (2003a), pp. 241–260. – DOI: doi.org/10.1016/S0047-259X(03)00053-8

[Mas and Menneteau 2003b] Mas, A. ; Menneteau, L.: Perturbation approach applied to the asymptotic study of random operators. *Progress in Probability* 55 (2003b), pp. 127–134. – DOI: doi.org/10.1007/978-3-0348-8059-6_8

[Mas and Pumo 2007] Mas, A. ; Pumo, B.: The ARHD model. *J. Statist. Plann. Inference* 137 (2007), pp. 538–553. – DOI: doi.org/10.1016/j.jspi.2005.12.006

[Mas and Pumo 2010] Mas, A. ; Pumo, B.: *Linear processes for functional data. The Oxford Handbook of functional data*. Ferraty and Romain Eds., Oxford, 2010. – URL https://arxiv.org/abs/0901.2503

[Maslova et al. 2010] Maslova, I. ; Kokoszka, P. ; Sojka, J. ; Zhu, L.: Statistical significance testing for the association of magnetometer records at high-, mid- and low latitudes during substorms days. *Planetary and Space Science* 58 (2010), pp. 437–445. – DOI: doi.org/10.1016/j.pss.2009.11.004

[Masry 2005] Masry, E.: Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.* 115 (2005), pp. 155–1772. – DOI: doi.org/10.1016/j.spa.2004.07.006

[Meerschaert et al. 2002] Meerschaert, M. M. ; Benson, D. A. ; Scheffler, H. P. ; Baeumer, B.: Stochastic solution of space-time fractional diffusion equations. *Phys. Rev. E* 65 (2002), pp. 1103–1106. – DOI: doi.org/10.1103/PhysRevE.65.041103

[Menneteau 2005] Menneteau, L.: Some laws of the iterated logarithm in Hilbertian autoregressive models. *J. Multivariate Anal.* 92 (2005), pp. 405–425. – DOI: doi.org/10.1016/j.jmva.2003.07.001

[Merlevède 1995] Merlevède, F.: Sur l'inversibilité des processus linéaires à valeurs dans un espace de Hilbert. *C. R. Acad. Sci. Paris Sér. I* 321 (1995), pp. 477–480

[Merlevède 1996a] Merlevède, F.: Central limit theorem for linear processes with values in a Hilbert space. *Stochastic Process. Appl.* 65 (1996a), pp. 103–114. – DOI: doi.org/10.1016/S0304-4149(96)00099-3

[Merlevède 1996b] Merlevède, F.: Lois des grands nombres et loi du logarithme itéré compacte pour des processus linéaires à valeurs dans un espace de Banach de type 2. *C. R. Acad. Sci. Paris Sér. I Math* 323 (1996b), pp. 521–524

[Merlevède 1997] Merlevède, F.: Résultats de convergence presque sûre pour l'estimation et la prévision des processus linéaires Hilbertiens. *C. R. Acad. Sci. Paris Sér. I* 324 (1997), pp. 573–576

[Merlevède et al. 1997] Merlevède, F. ; Peligrad, M. ; Utev, S.: Sharp conditions for the CLT of linear processes in a Hilbert space. *J. Theoret. Probab.* 10 (1997), pp. 681–693. – DOI: doi.org/10.1023/A:1022653728014

[Metzler and Klafter 2004] Metzler, R. ; Klafter, J.: The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *J. Physics A* 37 (2004), pp. R161–R208. – DOI: doi.org/10.1088/0305-4470/37/31/R01

[Meyer and Coifman 1997] Meyer, Y. ; Coifman, R.: *Wavelets, Calderón-Zygmund and multilinear operators*. Cambridge University Press, 1997. – ISBN 9780521794732

[Mills 2013] Mills, T. C.: Time series modelling of temperatures: an example from Kefalonia. *Met. Apps.* 21 (2013), pp. 578–584. – DOI: doi.org/10.1002/met.1379

[Mohammadzadeh et al. 2013] Mohammadzadeh, S. ; Mosayebi, S. A. ; Moosapoor, R.: Investigating on the Effects of Random Irregularities of Railway Track by Half-Bogie Model. *Int. J. Adv. Railway Engineering* 1 (2013), pp. 61–75. – URL http://www.ijare.ir/?_action=articleInfo&article=3701

[Mokhtari and Mourid 2003] Mokhtari, F. ; Mourid, T.: Prediction of continuous time autoregressive processes via the Reproducing Kernel Spaces. *Stat. Inference Stoch. Process.* 6 (2003), pp. 247–266. – DOI: doi.org/10.1023/A:1025852517084

[Mörters and Peres 2010] Mörters, P. ; Peres, Y.: *Brownian Motion*. Cambridge University Press, Cambridge, 2010. – ISBN 9780521760188

[Mourid 1993] Mourid, T.: Processus autorégressifs Banachiques d'ordre supérieur. *C. R. Acad. Sci. Paris Sér. I Math.* 317 (1993), pp. 1167–1172

[Mourid 1996] Mourid, T.: Représentation autorégressive dans un espace de Banach de processus réels à temps continu et équivalence des lois. *C. R. Acad. Sci. Paris Sér. I Math.* 322 (1996), pp. 1219–1224

[Mourid 2002] Mourid, T.: Statistiques d'une saisonnalité perturbée par un processus a représentation autorégressive. *C. R. Acad. Sci. Paris S.ér. I* 334 (2002), pp. 909–912. – URL https://ac.els-cdn.com/S1631073X0202352X/1-s2.0-S1631073X0202352X-main.pdf?_tid=ef83e95e-88ee-48ab-a807-6ffe093c47db&acdnat=1520249666_40eb91e83146cfe681a1e1e66a1679c5

[Mourid 2004] Mourid, T.: Processus autorégressifs Hilbertiens à coefficients aléatoires. *Ann. I. S. U. P.* 48 (2004), pp. 79–85. – URL http://www.lsta.lab.upmc.fr/modules/resources/download/labsta/Annales_ISUP/Couv_ISUP_48-3.pdf

[Müller and Stadtmüller 2005] Müller, H. G. ; Stadtmüller, U.: Generalized functional linear models. *Ann. Statist.* 33 (2005), pp. 774–805. – DOI: doi.org/10.1214/009053604000001156

[Nadaraya 1964] Nadaraya, E. A.: On Estimating Regression. *Theory Probab. Appl.* 9 (1964), pp. 141–142. – DOI: doi.org/10.1137/1109020

[Nason 2008] Nason, G. P.: *Wavelet Methods in Statistics with R*. Springer, Berlin, 2008. – ISBN 9780387759616

[Nerini et al. 2010] Nerini, D. ; Monestiez, P. ; Mantea, C.: Cokriging for spatial functional data. *J. Multivariate Anal.* 101 (2010), pp. 409–418. – DOI: doi.org/10.1016/j.jmva.2009.03.005

[Novak and Ritter 1998] Novak, E. ; Ritter, K.: The curse of dimension and a universal method for numerical integration. *Multivariate Approx. Splines* 125 (1998), pp. 177–188. – DOI: doi.org/10.1007/978-3-0348-8871-4_15

[Ogden 1997] Ogden, R. T.: *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhauser, Boston, MA, 1997. – ISBN 9781461207092

[Oladunni 2013] Oladunni, O. O.: Regularized Least Squares Piecewise Multi-classification Machine. *WSEAS Transactions on Computers* 12 (2013), pp. 18–27. – URL http://www.wseas.org/multimedia/journals/computers/2013/56-476.pdf

[Olver 1951] Olver, F. W. J.: A further method for the evaluation of zeros of Bessel functions and some new asymptotic expansions for zeros of functions of large order. *Math. Proc. Cambridge Philos. Soc.* 47 (1951), pp. 699–712. – DOI: doi.org/10.1017/S0305004100027158

[Olver 1952] Olver, F. W. J.: Some new asymptotic expansions for Bessel functions of large orders. *Math. Proc. Cambridge Philos. Soc.* 48 (1952), pp. 414–427. – DOI: doi.org/10.1017/S030500410002781X

[Panaretos and Tavakoli 2013a] Panaretos, V. M. ; Tavakoli, S.: Cramér-Karhunen-Loève representation and harmonic PCA of functional time series. *Stochastic Process. Appl.* 123 (2013), pp. 2779–2807. – DOI: doi.org/10.1016/j.spa.2013.03.015

[Panaretos and Tavakoli 2013b] Panaretos, V. M. ; Tavakoli, S.: Fourier analysis of stationary time series in function space. *Ann. Statist.* 41 (2013), pp. 568–603. – DOI: doi.org/10.1214/13-AOS1086

[Parvardeh et al. 2017] Parvardeh, A. ; Jouzdani, N. M. ; Mahmoodi, S. ; Soltani, A. R.: First order autoregressive periodically correlated model in Banach spaces: existence and central limit theorem. *J. Multivariate Anal.* 449 (2017), pp. 756–768. – DOI: doi.org/10.1016/j.jmaa.2016.12.037

[Parzen 1961] Parzen, E.: An approach to time series analysis. *Ann. Math. Statist.* 32 (1961), pp. 951–989. – DOI: doi.org/10.1214/aoms/1177704840

[Petris 2013] Petris, G.: A Bayesian framework for functional time series analysis. *arXiv:1311.0098v2* (2013). – URL https://arxiv.org/abs/1311.0098

[Pezzulli and Silverman 1993] Pezzulli, S. ; Silverman, B. W.: Some properties of smoothed principal components analysis for functional data. *Comput. Statist.* 8 (1993), pp. 1–1

[Poggi 1994] Poggi, J. M.: Prévision non paramétrique de la consommation électrique. *Revue de Statistique Appliquée* 42 (1994), pp. 93–98. – URL http://www.numdam.org/article/RSA_1994__42_4_83_0.pdf

[Poggi and Portier 2011] Poggi, J. M. ; Portier, B.: PM_{10} forecasting using clusterwise regression. *Atmospheric Environment* 45 (2011), pp. 7005–7014. – DOI: doi.org/10.1016/j.atmosenv.2011.09.016

[Prato and Zabczyk 2002] Prato, G. D. ; Zabczyk, J.: *Second Order Partial Differential Equations in Hilbert spaces*. Cambridge University Press, New York, 2002. – ISBN 9780511543210

[Preda and Saporta 2005] Preda, C. ; Saporta, G.: PLS regression on a stochastic process. *Comput. Stat. Data Anal.* 48 (2005), pp. 149–158. – DOI: doi.org/10.1016/j.csda.2003.10.003

[Pumo 1992] Pumo, B.: *Estimation et prévision de processus autorégressifs fonctionnels*. Paris, Université de Paris 6, PhD. Thesis, 1992

[Pumo 1998] Pumo, B.: Prediction of continuous time processes by $C_{[0,1]}$-valued autoregressive process. *Stat. Inference Stoch. Process.* 1 (1998), pp. 297–309

[Rachedi 2004] Rachedi, F.: Vitesse de convergence de l'estimateur crible d'un ARB(1). *Ann. I. S. U. P.* 48 (2004), pp. 87–96. – URL http://www.agro-montpellier.fr/sfds/CD/textes/rachedi1.pdf

[Rachedi 2005] Rachedi, F.: Vitesse de convergence en norme p-intégrale et normalité asymptotique de l'estimateur crible de l'opérateur d'un ARB(1). *C. R. Math. Acad. Sci. Paris Sér. I* 341 (2005), pp. 369–374. – DOI: doi.org/10.1016/j.crma.2005.05.009

[Rachedi and Mourid 2003] Rachedi, F. ; Mourid, T.: Estimateur crible de l'opérateur d'un processus ARB(1). Sieve estimator of the operator in ARB(1) process. *C. R. Acad. Sci. Paris Sér. I* 336 (2003), pp. 605–610. – DOI: doi.org/10.1016/S1631-073X(03)00061-X

[Ramsay 1982] Ramsay, J. O.: When the data are functions. *Psychometrika* 47 (1982), pp. 379–396. – DOI: doi.org/10.1007/BF02293704

[Ramsay and Dalzell 1991] Ramsay, J. O. ; Dalzell, C.: Some tools for functional data analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 53 (1991), pp. 539–572. – URL http://www.jstor.org/stable/pdf/2345586.pdf

[Ramsay and Silverman 2005] Ramsay, J. O. ; Silverman, B. W.: *Functional data analysis, 2nd ed.* Springer, New York, 2005. – ISBN 978038740080

[Rao et al. 2012] Rao, T. S. ; Rao, S. S. ; Rao, C. R.: *Time Series Analysis: Methods and Applications*. Elsevier, 2012. – ISBN 9780444538581

[Reimherr and Nicolae 2014] Reimherr, M. ; Nicolae, D.: A functional data analysis approach for genetic association studies. *Ann. Appl. Stat.* 8 (2014), pp. 406–429. – DOI: doi.org/10.1214/13-AOAS692

[Reiss and Ogden 2007] Reiss, P. T. ; Ogden, T.: Functional principal component regression and functional partial least-squares. *J. Amer. Statist. Assoc.* 102 (2007), pp. 984–996. – DOI: doi.org/10.1198/016214507000000527

[Ruiz-Medina 2011] Ruiz-Medina, M. D.: Spatial autoregressive and moving average Hilbertian processes. *J. Multivariate Anal.* 102 (2011), pp. 292–305. – DOI: doi.org/10.1016/j.jmva.2010.09.005

[Ruiz-Medina 2012] Ruiz-Medina, M. D.: Spatial functional prediction from spatial autoregressive Hilbertian processes. *Environmetrics* 23 (2012), pp. 119–128. – DOI: doi.org/10.1002/env.1143

[Ruiz-Medina 2016] Ruiz-Medina, M. D.: Functional analysis of variance for Hilbert–valued multivariate fixed effect models. *Statistics* 50 (2016), pp. 689–715. – DOI: doi.org/10.1080/02331888.2015.1094069

[Ruiz-Medina and Álvarez-Liébana 2017] Ruiz-Medina, M. D. ; Álvarez-Liébana, J.: Classical and Bayesian componentwise predictors for non–ergodic ARH(1) processes. *REVSTAT (in press)* (2017)

[Ruiz-Medina and Álvarez-Liébana 2017a] Ruiz-Medina, M. D. ; Álvarez-Liébana, J.: Classical and Bayesian componentwise predictors for non-ergodic ARH(1) processes. *REVSTAT (in press)* (2017a)

[Ruiz-Medina and Álvarez-Liébana 2018a] Ruiz-Medina, M. D. ; Álvarez-Liébana, J.: A note on strong-consistency of componentwise ARH(1) predictors. *Stat. Probab. Lett. (under minor revision)* (2018a)

[Ruiz-Medina and Álvarez-Liébana 2018b] Ruiz-Medina, M. D. ; Álvarez-Liébana, J.: Strongly–consistent autoregressive predictors in abstract Banach spaces. *J. Multivariate Anal. (under minor revision)* (2018b)

[Ruiz-Medina et al. 2014] Ruiz-Medina, M. D. ; Espejo, R. M. ; Ugarte, M. D. ; Militino, A. F.: Functional time series analysis of spatio–temporal epidemiological data. *Stoch. Environ. Res. Risk. Assess.* 28 (2014), pp. 943–954. – DOI: doi.org/10.1007/s00477-013-0794-y

[Ruiz-Medina et al. 2016] Ruiz-Medina, M. D. ; Romano, E. ; Fernández-Pascual, R.: Plug-in prediction intervals for a special class of standard ARH(1) processes. *J. Multivariate Anal.* 146 (2016), pp. 138–150. – DOI: doi.org/10.1016/j.jmva.2015.09.001

[Ruiz-Medina and Salmerón 2009] Ruiz-Medina, M. D. ; Salmerón, R.: Functional maximum-likelihood estimation of ARH(p) models. *Stoch. Environ. Res. Risk. Assess.* 24 (2009), pp. 131–146. – DOI: doi.org/10.1007/s00477-009-0306-2

[Sathyanarayana et al. 2016] Sathyanarayana, A. ; Joty, S. ; Fernández-Luque, L. ; Ofli, F. ; Srivastava, J. ; Elmagarmid, A. ; Taheri, S.: Sleep Quality Prediction From Wearable Data Using Deep Learning. *JMIR mHealth and uHealth* 4 (2016). – DOI: doi.org/10.2196/mhealth.6562

[Schneider 2014] Schneider, P.: *NURB Curves: A Guide for the Uninitiated*. MACTECH, 2014. – URL http://www.mactech.com/articles/develop/issue_25/schneider.html

[Schoenberg 2012] Schoenberg, I. J.: Spline Functions and the Problem of Graduation. *Proc. Natl. Acad. Sci. U. S. A.* 52 (2012), pp. 947–950. – URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC300377/

[Scholkopf and Smola 2002] Scholkopf, B. ; Smola, A.: *Learning with Kernels*. MIT Press, Cambridge, 2002. – URL https://www.cs.utah.edu/~piyush/teaching/learning-with-kernels.pdf

[Schumaker 1981] Schumaker, M.: *Splines functions: basic theory*. Wiley, New York, 1981. – ISBN 9780511618994

[Shang 2013] Shang, H. L.: Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density. *Comput. Statist. Data Anal.* 67 (2013), pp. 185–198. – DOI: doi.org/10.1016/j.csda.2013.05.006

[Shang 2018] Shang, H. L.: Bootstrap methods for stationary functional time series. *Stat Comput* 28 (2018), pp. 1–10. – DOI: doi.org/10.1007/s11222-016-9712-8

[Shang et al. 2016] Shang, H. L. ; Smith, P. W. F. ; Bijak, J. ; Wiśniowski, A.: A multilevel functional data method for forecasting population, with an application to the United Kingdom. *International J. Forecasting* 32 (2016), pp. 629–649. – DOI: doi.org/10.1016/j.ijforecast.2015.10.002

[Shaw et al. 2006] Shaw, P. ; Greenstein, D. ; Lerch, J. ; Ciasen, L. ; Lenroot, R. ; Gogtay, N. ; Evans, A. ; Rapoport, J. ; Giedd, J.: Intellectual ability and cortical development in children and adolescents. *Nature* 440 (2006), pp. 676–679. – DOI: doi.org/10.1038/nature04513

[Shen and Faraway 2004] Shen, Q. ; Faraway, J.: An F test for linear models with functional responses. *Statist. Sinica* 14 (2004), pp. 1239–1257. – URL http://www.jstor.org/stable/24307230

[Shishebor et al. 2011] Shishebor, Z. ; Soltani, A. R. ; Zamani, A.: Asymptotic distribution for periodograms of infinite dimensional discrete time periodically correlated processes. *J. Multivariate Anal.* 102 (2011), pp. 1118–1125. – DOI: doi.org/10.1016/j.jmva.2011.03.005

[Shojaie and Michailidis 2009] Shojaie, A. ; Michailidis, G.: Analysis of gene sets based on the underlying regulatory network. *J. Comp. Biol.* 16 (2009). – DOI: doi.org/10.1089/cmb.2008.0081

[Skorokhod 1956] Skorokhod, A. V.: Limit theorems for stochastic processes. *Theor. Probab. Appl.* 1 (1956), pp. 261–290. – DOI: doi.org/10.1137/1101022

[Slini et al. 2006] Slini, T. ; Kaprara, A. ; Karatzas, K. ; Mousiopoulos, N.: PM10 forecasting por Thessaloniki, Greece. *Enviromental Modelling & Software* 21 (2006), pp. 559–565. – DOI: doi.org/10.1016/j.envsoft.2004.06.011

[Sobczyk 1991] Sobczyk, K.: *Stochastic Differential Equations, with Applications to Physics and Engineering*. Kluwer Academic Publishers, Dordrecht, 1991. – ISBN 9789401137126

[Soltani and Hashemi 2011] Soltani, A. R. ; Hashemi, M.: Periodically correlated autoregressive Hilbertian processes. *Stat. Inference Stoch. Process.* 14 (2011), pp. 177–188. – DOI: doi.org/10.1007/s11203-011-9056-0

[Soltani and Shishebor 2007] Soltani, A. R. ; Shishebor, Z.: On infinite dimensional discrete time periodically correlated processes. *Rocky Mountain J. Math.* 37 (2007), pp. 1043–1058. – URL http://www.jstor.org/stable/44239382

[Soltani et al. 2010] Soltani, A. R. ; Shishebor, Z. ; Zamani, A.: Inference on periodograms of infinite dimensional discrete time periodically correlated processes. *J. Multivariate Anal.* 101 (2010), pp. 368–373. – DOI: doi.org/10.1016/j.jmva.2009.01.004

[Sorensen et al. 2013] Sorensen, H. ; Goldsmith, J. ; Sangalli, L. M.: An introduction with medical applications to functional data analysis. *Stat. Med.* 32 (2013), pp. 5222–5240. – DOI: doi.org/10.1002/sim.5989

[Spangenberg 2013] Spangenberg, F.: Strictly stationary solutions of ARMA equations in Banach spaces. *J. Multivariate Anal.* 121 (2013), pp. 127–138. – DOI: doi.org/10.1016/j.jmva.2013.06.007

[Spitzner et al. 2003] Spitzner, D. J. ; Marron, J. S. ; Essick, G. K.: Mixed-model functional ANOVA for studying human tactile perception. *J. Amer. Statist. Assoc.* 98 (2003), pp. 263–272. – URL http://www.jstor.org/stable/30045234

[Stadlober et al. 2008] Stadlober, E. ; Hörmann, S. ; Pfeiler, B.: Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment* 42 (2008), pp. 1098–1109. – DOI: doi.org/10.1016/j.atmosenv.2007.10.073

[Stone et al. 1997] Stone, C. J. ; Hansen, M. ; Kooperberg, C. ; Truong, Y. K.: Polynomial splines and their tensor products in extended linear modeling. *Ann. Statist.* 25 (1997), pp. 1371–1470. – URL https://projecteuclid.org/download/pdf_1/euclid.aos/1031594728

[Stout 1974] Stout, W. F.: *Almost sure convergence*. Academic, New York-London, 1974. – ISBN 9780126727500

[Taylor and Worsley 2007] Taylor, J. E. ; Worsley, K. J.: Detecting sparse signals in random fields, with an application to brain mapping. *J. Amer. Statist. Assoc.* 102 (2007), pp. 913–928. – URL http://www.jstor.org/stable/27639934

[Tibshirani 1996] Tibshirani, R. J.: Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58 (1996), pp. 267–288. – URL http://www.jstor.org/stable/2346178

[Tong 2011] Tong, W.: Threshold models in time series analysis–30 years on. *Stat. Interface* 4 (2011), pp. 107–118. – DOI: doi.org/10.1.1.457.9000

[Touloumis 2015] Touloumis, A.: Nonparametric Stein–type shrinkage covariance matrix estimators in high–dimensional settings. *Comput. Statist. Data Anal.* 83 (2015), pp. 251–261. – DOI: doi.org/10.1016/j.csda.2014.10.018

[Triebel 1983] Triebel, H.: *Theory of function spaces II*. Birkhauser, Basel, 1983. – ISBN 978303460411

[Turbillon et al. 2008] Turbillon, C. ; Bosq, D. ; Marion, J. M. ; Pumo, B.: Estimation du paramètre des moyennes mobiles Hilbertiennes. *C. R. Acad. Sci. Paris Sér. I Math.* 346 (2008), pp. 347–350. – DOI: doi.org/10.1016/j.crma.2008.01.008

[Turbillon et al. 2007] Turbillon, C. ; Marion, J. M. ; Pumo, B.: Estimation of the moving–average operator in a Hilbert space. *In: Recent advances in stochastic modeling and data analysis. World Sci. Publ., Hackensack, NJ* (2007), pp. 597–604. – DOI: doi.org/10.1142/9789812709691_0070

[Uhlenbeck and Ornstein 1930] Uhlenbeck, G. E. ; Ornstein, L. S.: On the theory of Brownian motion. *Phys. Rev.* 36 (1930), pp. 823–841. – DOI: doi.org/10.1103/PhysRev.36.823

[Vasicek 1977] Vasicek, O.: An equilibrium characterization of the term structure. *J. Financial Economics* 5 (1977), pp. 177–188. – DOI: doi.org/10.1016/0304-405X(77)90016-2

[Vidakovic 1998] Vidakovic, B.: Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Statist. Assoc.* 93 (1998), pp. 173–179. – DOI: doi.org/10.2307/2669614

[Vieu 2018] Vieu, P.: On dimension reduction models for functional data. *Statist. Prob. Lett.* 136 (2018), pp. 134–138. – DOI: doi.org/10.1016/j.spl.2018.02.032

[Wahba et al. 1995] Wahba, G. ; Wang, Y. ; Gu, C. ; Klein, R. ; Klein, B.: Smoothing splines ANOVA for exponential families, with application to the Winsconsin epidemiological study of diabetic retinopathy. *Ann. Statist.* 23 (1995), pp. 1865–1895. – URL http://www.jstor.org/stable/2242776

[Wand and Ormerod 2011] Wand, M. P. ; Ormerod, J. T.: Penalized wavelets: embedding wavelets into semiparametric regression. *Electron. J. Stat.* 5 (2011), pp. 1654–1717. – DOI: doi.org/10.1214/11-EJS652

[Wang 2008] Wang, H. B.: Nonlinear ARMA models with functional MA coefficients. *J. Time Ser. Anal.* 29 (2008), pp. 1032–1056. – DOI: doi.org/10.1111/j.1467-9892.2008.00594.x

[Wang et al. 2016] Wang, J. L. ; Chiou, J. M. ; Müller, H. G.: Review of functional data analysis. *Annu. Rev. Statist.* 3 (2016), pp. 257–295. – DOI: doi.org/10.1146/annurev-statistics-041715-033624

[Wang and Uhlenbeck 1945] Wang, M. C. ; Uhlenbeck, G. E.: On the theory of Brownian motion II. *Rev. Modern Phys.* 17 (1945), pp. 323–342. – DOI: doi.org/10.1103/RevModPhys.17.323

[Wang et al. 2010] Wang, S. ; Qian, L. ; Carroll, R. J.: Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika* 97 (2010), pp. 79–93. – DOI: doi.org/10.1093/biomet/asp073

[Ware and Lad 2003] Ware, R. ; Lad, F.: Approximating the distribution for sums of product of normal variables. *Research-Paper 2003-15. Department of Mathematics and Statistics (University of Canterbury - New Zealand)* (2003). – URL http://www.math.canterbury.ac.nz/research/ucdms2003n15.pdf

[Wasilkowski and Wozniakowski 1995] Wasilkowski, G. ; Wozniakowski, H.: Explicit Cost Bounds of Algorithms for Multivariate Tensor Product Problems. *J. Complexity* 11 (1995), pp. 1–56. – DOI: doi.org/10.1006/jcom.1995.1001

[Watson 1966] Watson, G. N.: *A Treatise on the Theory of Bessel Functions (Second Edition)*. Cambridge University Press, Cambridge, 1966. – ISBN 052106743X

[Watson 1964] Watson, G. S.: Smooth regression analysis. *Sankhya A* 26 (1964), pp. 359–372. – URL http://www.jstor.org/stable/25049340

[Widom 1963] Widom, H.: Asymptotic behavior of the eigenvalues of certain integral equations. *Trans. Amer. Math. Soc.* 109 (1963), pp. 278–295. – DOI: doi.org/10.2307/1993907

[Worsley et al. 2002] Worsley, K. J. ; Liao, C. ; Aston, J. ; Petre, V. ; Duncan, G. H. ; Morales, F. ; Evans, A. C.: A general statistical analysis for fMRI data. *Neuroimage* 15 (2002), pp. 1–15. – DOI: doi.org/10.1006/nimg.2001.0933

[Wu and Shao 2004] Wu, W. B. ; Shao, X.: Limit theorems for iterated random functions. *J. Appl. Probab.* 41 (2004), pp. 425–436. – URL http://www.jstor.org/stable/3216026

[Yang et al. 2009a] Yang, J. Y. ; Peng, Z. L. ; Yu, Z. G. ; Zhang, R. J. ; Anh, V. ; Wang, D.: Prediction of protein structural classes by recurrence quantification analysis based on chaos game representation. *J. Theor. Biol.* 257 (2009), pp. 618–626. – DOI: doi.org/10.1016/j.jtbi.2008.12.027

[Yang et al. 2009b] Yang, J. Y. ; Yu, Z. G. ; Anh, V.: Clustering structures of large proteins using multifractal analyses based on a 6–letter model and hydrophobicity scale of amino acids. *Chaos Solitons Fractals* 40 (2009), pp. 607–620. – DOI: doi.org/10.1016/j.chaos.2007.08.014

[Youcef et al. 2013] Youcef, K. ; Sabiha, T. ; Mostafa, D. E. ; Ali, D. ; Bachir, M.: Dynamic analysis of train–bridge system and riding comfort of trains with rail irregularities. *J. Mech. Sci. Technol.* 27 (2013), pp. 951–962. – DOI: doi.org/10.1007/s12206-013-0206-8

[Zhang 2013] Zhang, J. T.: *Analysis of Variance for Functional Data*. Champman and Hall, London, 2013. – ISBN 9781439862735

[Zhang et al. 2011] Zhang, X. ; Shao, X. ; Hayhoe, K. ; Wuebbles, D.: Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Stat.* 5 (2011), pp. 1765–1796. – DOI: doi.org/10.1214/11-EJS655

[Zhao et al. 2015] Zhao, Y. ; Chen, H. ; Ogden, R. T.: Wavelet-based weighted LASSO and screening approaches in functional linear regression. *J. Comput. Graph. Statist.* 24 (2015), pp. 655–675. – DOI: doi.org/10.1080/10618600.2014.925458

[Zhao et al. 2012] Zhao, Y. ; Ogden, R. T. ; Reiss, P. T.: Wavelet-based LASSO in functional linear regression. *J. Comput. Graph. Statist.* 21 (2012), pp. 600–617. – DOI: doi.org/10.1080/10618600.2012.679241

[Zhu et al. 2014] Zhu, H. T. ; Fan, J. Q. ; Kong, L. L.: Spatially varying coefficient models with applications in neuroimaging data with jumping discontinuity. *J. Amer. Statist. Assoc.* 109 (2014), pp. 977–990. – DOI: doi.org/10.1214/12-AOS1045

[Zhu et al. 2012] Zhu, H. T. ; Li, R. ; Kong, L. L.: Multivariate varying coefficient model and its application in neuroimaging data. *Ann. Statist.* 40 (2012), pp. 2634–2666. – DOI: doi.org/10.1214/12-AOS1045

[Zoglat 2008] Zoglat, A.: Functional analysis of variance. *App. Math. Sci.* 2 (2008), pp. 1115–1129. – URL http://www.m-hikari.com/ams/ams-password-2008/ams-password21-24-2008/zoglatAMS21-24-2008.pdf

[Zou 2006] Zou, H.: The adaptive LASSO and its oracle properties. *J. Amer. Statist. Assoc.* 101 (2006), pp. 1418–1429. – DOI: doi.org/10.1198/016214506000000735

Colophon

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Arno Pro, designed by Robert Slimbach in the style of book types from the Aldine Press in Venice, and issued by Adobe in 2007. A template, which can be used to format a PhD thesis with this look and feel, has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/ or from the author at suchow@post.harvard.edu.