EPFL

Thèse n° 8251

Sparsely Observed Functional Time Series: Theory and Applications

Présentée le 5 mars 2021

Faculté des sciences de base
Chaire de statistique mathématique
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Tomáš RUBÍN

Acceptée sur proposition du jury

Prof. S. C. Olhede, présidente du jury
Prof. V. Panaretos, directeur de thèse
Dr S. Tavakoli, rapporteur
Dr Ł. Kidziński, rapporteur
Prof. S. Morgenthaler, rapporteur

École
polytechnique
fédérale
de Lausanne

2021

Acknowledgements

It feels like yesterday when I arrived at EPFL to start my PhD studies and what a journey it was. I want to use these two pages to thank all of you who helped me make the most of my *séjour* in Lausanne.

First and foremost I would like to express my gratitude to my thesis advisor, Victor Panaretos, whose support spanned over many levels throughout the years I worked on my thesis. His optimism and enthusiasm kept me going during the first year of my PhD when I struggled to narrow down a research project, later on his ability to look at the global picture combined with attention to details helped me to get my key papers published, and his storytelling and narrative driven presentation skills that I tried to learn from are going to be invaluable in my future career. Furthermore, I enjoyed travelling and presenting at conferences and I want to thank Victor for providing me with funding for numerous events: I had the opportunity to travel to Les Diablerets, Tübingen, Iaşi, Copenhagen, Limassol, Palermo, and London.

I would also like to thank my jury members: Sofia Olhede for chairing the oral exam and ensuring the overall pleasant experience during the defence, Łukasz Kidziński and Shahin Tavakoli for accepting to be the external examiners and for their constructive and very detailed feedback I received, and Stephan Morgenthaler for accepting to be the internal examiner and I wish him all the best in his retirement.

The administrative help of Maroussia Schaffner-Portillo, Anna Dietler, and Nadia Kaiser was invaluable during my years at EPFL. It was pleasure to be office neighbours with Nadia and I thank her for giving me tips on integration in Switzerland and on my alpine excursions.

The friendly environment within the Chair of Mathematical Statistics helped me shape my research progress. I see how advantageous it was to just knock on my colleagues' doors and ask seemingly trivial research questions. I want to thank for these discussions as well as the informal chats to all the current and previous chair members: Anirvan, Guillaume, Kate, Kartik, Laya, Marie-Hélène, Neda, Pavol, Soham, Tomáš, Valentina, and Yoav. I had also chance to interact with the members of the other statistical and probability theory chairs both at EPFL and at the events such as the CUSO summer and winter schools: Anthony, Daria, Erwan, Hao, Hélène, Jacques, Jonathan, Léo, Linda, Max, Raphaël, Rémy, Sonia, Soumaya, Stefano, Thomas, Tim, Yousra, and Zhiwen: thank you for creating a nice environment! In particular, I want to thank Sonia and Yoav for helping me with the French translation of the abstract.

What would be the point of living in Switzerland without the Alps? I may now admit that the proximity to the mountains was one of the deciding factors to do my PhD in Lausanne and,

indeed, hiking, climbing, mountaineering or skiing were my plans for almost every weekend. I have met many friendly people during these activities: Arwa, Elodie, Hao, Jan, Jan, Justin, Kiara, Kornelia, Léo, Leonid, Maria, Mariella, Michael, Nadine, Radek, Simon, Sylvain, and the others from *Club Montagne*. I want to thank you not only for being the companions and support during my adventures but also for teaching me a great deal of climbing and mountaineering techniques.

A big part of Switzerland are the languages. When I arrived in Lausanne, my French skills were essentially non-existent but thanks to being practically thrown in at the deep end of the francophone swimming pool I am now leaving Lausanne with decent knowledge of French. I want to thank Simon and Thomas for organising *Stammtisch* and the other language exchange events which facilitated my learning of French and helped me maintain my German proficiency. Furthermore, I thank Martyna for teaching me to understand Polish pretty well and Yoav for picking up a few phrases in Hebrew from him.

I want to thank all my friends in Hradec Králové and Prague who helped me to stay connected with my home town and country. Beer tastes always the best over there. In particular, I enjoyed playing numerous escape rooms, eating way too big burgers, and chatting about both important and unimportant topics with Jan, Jirka, and Martin.

Finally, I want to express my gratitude to my parents, Lenka and Jaroslav, for encouraging me to learn mathematics since my early childhood and supporting me throughout my entire studies. The fact that I am defending my PhD degree today is only because of your continuous support, *děkuji*. A special thanks goes to my brother, Jan, who helped me in difficult times.

Lausanne, October 9, 2020 Tomáš Rubín

Abstract

Functional time series is a temporally ordered sequence of not necessarily independent random curves. While the statistical analysis of such data has been traditionally carried out under the assumption of completely observed functional data, it may well happen that the statistician only has access to a relatively low number of sparse measurements for each random curve. These discrete measurements may be moreover irregularly scattered in each curve's domain, missing altogether for some curves, and be contaminated by measurement noise. This sparse sampling protocol escapes from the reach of established estimators in functional time series analysis and therefore requires development of a novel methodology.

The core objective of this thesis is development of a non-parametric statistical toolbox for analysis of sparsely observed functional time series data. Assuming smoothness of the latent curves, we construct a local-polynomial-smoother based estimator of the *spectral density operator* producing a consistent estimator of the complete second order structure of the data. Moreover, the spectral domain recovery approach allows for prediction of latent curve data at a given time by borrowing strength from the estimated dynamic correlations in the entire time series across time. Further to predicting the latent curves from their noisy point samples, the method fills in gaps in the sequence (curves nowhere sampled), denoises the data, and serves as a basis for forecasting.

A classical non-parametric apparatus for encoding the dependence between a pair of or among a multiple functional time series, whether sparsely or fully observed, is the *functional lagged regression model*. This consists of a linear filter between the regressors time series and the response. We show how to tailor the smoother based estimators for the estimation of the *cross-spectral density operators* and the *cross-covariance operators* and, by means of spectral truncation and Tikhonov regularisation techniques, how to estimate the lagged regression filter and predict the response process.

The simulation studies revealed the following findings: (i) if one has freedom to design a sampling scheme with a fixed number of measurements, it is advantageous to sparsely distribute these measurements in a longer time horizon rather than concentrating over a shorter time horizon to achieve dense measurements in order to diminish the spectral density estimation error, (ii) the developed functional recovery predictor surpasses the *static* predictor not exploiting the temporal dependence, (iii) neither of the two considered regularisation techniques can, in general, dominate the other for the estimation in functional lagged regression models. The new methodologies are illustrated by applications to real data: the meteorological data

revolving around the *fair-weather atmospheric electricity* measured in *Tashkent, Uzbekistan*, and at *Wank mountain, Germany*; and a case study analysing the dependence of the *US Treasury yield curve* on macroeconomic variables.

As a secondary contribution, we present a novel simulation method for general stationary functional time series defined through their spectral properties. A simulation study shows universality of such approach and superiority of the spectral domain simulation over the temporal domain in some situations.

Keywords: functional data analysis, spectral density operator, autocovariance operator, non-parametric regression, functional lagged regression, spectral domain simulation, fair-weather electricity, US Treasury yield curve.

Résumé

Les *séries temporelles fonctionnelles* sont des séquences de courbes aléatoires ordonnées dans le temps et non nécessairement indépendantes. Traditionnellement, l'analyse statistique de telles séries se fait sous l'hypothèse que les données fonctionnelles sont complètement observées. Cependant, il arrive que le statisticien n'ait accès qu'à un nombre relativement restreint de mesures éparses pour chaque courbe aléatoire. De plus, ces mesures discrètes peuvent être dispersée de façon irrégulière sur le domaine de la courbe ou complètement absentes pour certaines courbes, ou bien contaminées par des erreurs de mesures. Les estimateurs établis pour l'analyse des séries temporelles fonctionnelles ne tiennent pas compte de tels protocoles d'échantillonnage et nécessite donc le développement d'une nouvelle méthodologie.

L'objectif principal de cette thèse est le développement d'outils statistiques non-paramétriques pour analyser des séries temporelles fonctionnelle observées de façon éparse. En supposant que les courbes latentes sont régulières, nous construisons un estimateur de l'*opérateur de densité spectrale* basé sur la régression locale. Nous obtenons ainsi un estimateur consistent de la structure de second ordre des données. En outre, nous faisons appel à l'approche de la prédiction dans le domaine spectrale pour estimer les courbes latentes à un temps donnée basées les corrélations dynamiques estimées le long de séries entière. Ainsi, notre méthode permet d'estimer la courbe latente à partir d'observations empiriques, mais aussi de retrouver des parties non-observées de la série, réduire le bruit, et servir de base pour des prévisions dans le future.

Le *modèle de régression fonctionnelle décalée* est une technique non-paramétrique classique pour décrire la dépendance entre deux ou plusieurs séries temporelles fonctionnelles. Il s'agit d'un filtre linéaire entre les séries temporelles et la réponse. Nous adaptons des approches basées sur le lissage à l'estimation des *opérateurs de densité spectrale croisée* et les *opérateurs de covariance croisée* et nous utilisons *troncature spectrale* et la *régularisation de Tikhonov* afin d'estimer le filtre de régression décalée et prédire la variable de réponse.

Les simulations montrent les résultats suivants : (i) Dans le cas où on peut définir librement un plan d'échantillonnage avec un nombre fixé de mesures, il est plus avantageux de distribuer ces mesures le long d'une longue période plutôt que de les condenser sur une période plus courte en obtenant des mesures plus denses pour minimiser l'erreur d'estimation de la densité spectrale. (ii) Le prédicateur proposé est plus performant que le prédicateur *statique* qui ne tient pas compte de la dépendance temporelle. (iii) D'une façon générale, aucune des deux techniques de régularisation considérées ne peut dominer l'autre pour l'estimation

dans le modèle de régression fonctionnelle décalée. La nouvelle méthodologie est illustrée à travers des applications à des données réelles : Notre première étude concerne des données météorologiques et particulièrement l'*électricité atmosphérique par beau temps* mesurée en *Tachkent, Ouzbékistan*, et au *Wank en Allemagne*. Dans un deuxième exemple, nous analysons la dépendance entre la courbe des taux du trésor des États-Unis et certaines variables macro-économiques.

Une contribution secondaire de cette thèse est d'introduire une nouvelle méthode pour simuler des séries temporelles fonctionnelles stationnaires à partir de leur propriétés spectrales. Nous démontrons que cette approche est universelle et que dans certaines situations, une simulation basée sur le domaine spectral est supérieure à celle basée sur le domaine temporel.

Mots-clés : données fonctionnelles, opérateur de densité spectral, opérateur de covariance, régression non-paramétrique, régression fonctionnelle décalée, simulation dans le domaine spectrale, électricité atmosphérique par beau temps, courbe des taux du trésor des États-Unis.

Contents

List of figures

List of tables

Introduction

Functional data analysis is a statistical discipline consisting of approaches and methodologies to analyse data ensembles of random functions (Ramsay and Silverman, 2013; Ferraty and Vieu, 2006; Hsing and Eubank, 2015; Wang et al., 2016) where each function is viewed as one data object, an atom, or one sample element. Examples of such data sets include growth curves, handwriting data, weather records, or speech recordings (Ramsay and Silverman, 2007). From the statistical viewpoint, inferences from such data can be seen as the inference problem over repeated independent realisations of a stochastic process. The infinite dimensional probabilistic nature of stochastic processes brings over many challenges where the functional data analysis methodologies need to deviate from multivariate analysis methods: the analysis of infinite dimensional problems requires tools from functional analysis while many standard inference problems may become ill-posed. Another challenge comes from the fact that the random curves, as infinite dimensional objects, can never be observed directly in real data problems. Indeed, they can be recorded only by a finite number of measurements and the random curves themselves must be considered as latent objects. If this sampling is sufficiently dense, however, a pre-smoothing step can be applied and the data are then treated as genuinely functional while provably retaining the same asymptotic performance (Hall et al., 2006). This pre-smoothing approach was promoted and popularised by Ramsay and Silverman (2013, 2007).

In some cases, however, the pure "platonic" functional data are recorded only at a small number of locations in their domain, in which case we speak about *sparsely observed functional data*. Figure 1 demonstrates this phenomenon. Data sets admitting this sparse sampling regime often include longitudinal studies (Yao et al., 2005a) where the measurements of a certain subject are recorded at regular or irregular intervals, while different subjects are considered to be independent but coming from the same population. In their seminal paper, Yao et al. (2005a) demonstrated how to overcome the sparsity challenge and statistically infer the probabilistic properties of the underlying latent stochastic process up to second order. Specifically they showed how to employ local-polynomial smoothing techniques for the estimation of the mean function and the covariance kernel. These two objects are quintessential for statistical applications such us the dimensionality reduction, latent trajectory prediction (Yao et al., 2005a), or regression (Yao et al., 2005b). The local-polynomial smoothing techniques have been later improved by Hall et al. (2006) and Li and Hsing (2010) who strengthened the

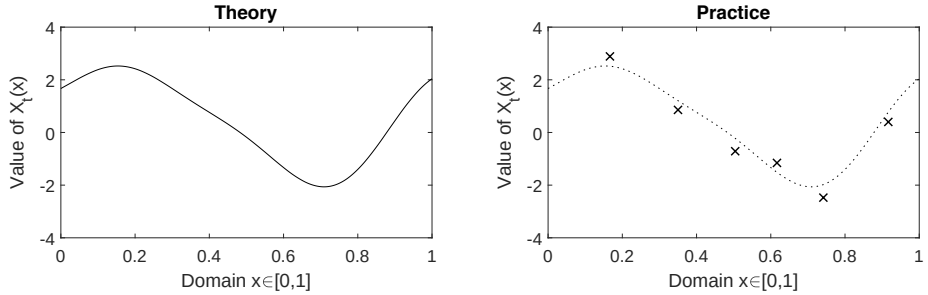

Figure 1 – **Left:** The “Theory" picture, where $X_t(x), x \in [0,1]$, is a fully observed functional datum (solid line). **Right:** The “Practice" picture, where the functional datum $X_t(x), x \in [0,1]$, is latent, and one can either observe dense noiseless observations (dotted line) or sparse noisy observations (crosses) with additive noise $Y_{tj} = X_t(x_{tj}) + \epsilon_{tj}$ observed at locations $x_{tj}, j = 1,\ldots,6$. In the dense case, one can typically behave as if the true latent function were observed. The sparse case, however, needs new tools.

non-parametric convergence rate and allowed for mixed regimes where some functional data are observed sparsely while some densely. Other notable approaches on how to deal with sparsity include fitting by minimizing a convex criterion function and expressing the estimator within a reproducing kernel Hilbert space (Cai and Yuan, 2010; Wong and Zhang, 2019), and expressing the latent functional data in a pre-specified basis and estimating the dynamics using the expectation maximisation algorithm (Rice and Wu, 2001; James et al., 2000).

Another challenge in functional data may comes in the form of dependence. Until now, we have considered functional data being manifested as an independent samples from the same population, hence the problem falls within the classical independent identically distributed setting. In many applications, however, the data, whether multivariate or functional, are recorded sequentially, forming a temporal sequence of measurements. In this sense, a temporally ordered sequence indexed by integers consisting of functional data is called *functional time series*. Historically, the development of statistical methods analysing functional time series data has relied on generalisations of univariate and multivariate linear time series models into function spaces (Bosq, 2000; Blanke and Bosq, 2007). This is a non-trivial task thanks to the inherent ill-posedness of many inverse problems in infinite dimensions. More recently, the field has moved away from the realm of linear processes. Hörmann and Kokoszka (2010) considered weakly dependent functional time series, showed the effect of this weak temporal dependence on principal component analysis, and studied the problem of estimating the long-run covariance operator. Moreover, they introduced the concept of L^p-m-approximability which proved to be useful in proving many asymptotic results in functional time series literature. Horváth et al. (2013) provided additional contributions for functional time series by deriving a central limit theorem for the mean function of a stationary weakly dependent time series and also studied the long-run covariance operator estimation problem.

Further advancement in functional time series research moved on to estimation of complete second order structure by fully non-parametric techniques while especially fruitful became

the analysis and estimation in the spectral domain. Panaretos and Tavakoli (2013a) generalised the definitions of the spectral density and the spectral density matrix from the univariate or multivariate time series analysis respectively, and defined their infinite dimensional counterpart: the *spectral density operators* and the *spectral density kernels*, and showed how these object can be used for optimal dimensionality reduction by the means of the harmonic principal components using the Cramér-Karhunen-Loève representation. Panaretos and Tavakoli (2013b) showed how to estimate the said spectral density operators and kernels by the means of periodogram smoothing while their asymptotic theory generalises the cumulant-type mixing conditions, à la Brillinger (1981). In independent parallel work, Hörmann et al. (2015a) introduced the dynamic principal components using similar ideas as the harmonic principal components of Panaretos and Tavakoli (2013a). Moreover, Hörmann et al. (2015a) provided with an alternative estimator of the spectral density operator by the functional version of Bartlett's estimate (Bartlett, 1950) while relying on L^p-m-approximability for their asymptotic theory.

Despite recent progress in functional time series research, the focus has been almost exclusively on fully observed functional time series. These methods can be successfully applied only in the dense observation regime where the functional data can be constructed by a pre-smoothing step. According to our knowledge, there are only a handful of exceptions considering discrete measurements on functional time series: Panaretos and Tavakoli (2013b) showed that their asymptotic results are stable under dense discrete observations regimes with measurement errors of decaying magnitude, and Kowal et al. (2017) studied the functional autoregressive processes by the means of Bayesian hierarchical Gaussian models. A related problem was studied by Paul and Peng (2011), who considered correlated sparsely observed functional data with separable covariance structure, but the focus was not on dynamics.

The primary objective of this thesis is to address this important gap in literature and provide a framework of non-parametric estimation, prediction, and regression for functional time series dynamics observed sparsely, irregularly and with measurement errors. Therefore we speak about genuinely sparse data in the same vein as the sparse observation framework due to Yao et al. (2005a) and as visualised on Figure 1. We consider the problems of estimating the complete first and second order dynamics of such data, estimation of the cross-dependence between multiple sparsely observed functional time series, and recovering the latent functional data. Furthermore, we analyse the lagged regression link between two sparsely observed functional time series.

A secondary objective of this thesis is the problem of simulation of functional time series with given spectral properties. During the development of the results on sparsely observed functional time series it became apparent that functional time series defined directly through their spectrum could become useful for Monte Carlo assessments of proposed methods. In the past, most simulation studies relied on functional linear processes, most often only functional autoregressive processes, to asses the novel methods even if their applicability went beyond the linear case. Our spectral simulation method can generate samples from arbitrary

stationary functional time series defined through their spectral density operators and thus extending assessment opportunities well beyond the functional linear processes. Moreover, some functional time series analysis methods depend heavily (Hörmann et al., 2015b; Zhang, 2016; Tavakoli and Panaretos, 2016) on the spectral properties of the data, rather than on their specification in the temporal domain. Therefore simulating functional time series samples with "custom" defined spectral density operators can be useful to assess the performance of such methods.

Detailed overview and the novel contributions of the thesis

Chapter 1 presents the background theory of *functional data analysis* that the novel results of this thesis are built on. Firstly, in Section 1.1 we review the basics of operator theory with special attention to trace class & Hilbert-Schmidt operators, and the integral operators defined by a kernel function. We rigorously define *Hilbert space valued random elements* which serve as the probabilistic model for functional data and discuss the subtle differences between these random elements and stochastic processes, and provide conditions where this dual perspective intersects. Moreover we establish basic results in the statistical inference for samples of independent identically distributed random data. Secondly, in Section 1.2 we present the concept of *sparsely observed functional data* where the observations are seen as a sampled version of a latent functional datum. We explain the difference between dense and sparse sampling schemes and introduce the smoother-based estimators for sparsely observed independent identically distributed data.

Thirdly, Section 1.3 of Chapter 1 introduces the temporal dependence structure among functional data by defining the notion of *functional time series*. We establish the basic definitions such as *stationarity*, the *mean function*, and the *lagged autocovariance operators*. Attention is given to various notions of *weak dependence*: the *cumulant-mixing conditions*, L^p*-m-approximability*, and the *strong mixing* conditions, which turn out to be useful for establishing the asymptotic behaviour of estimators. The spectral domain approach for capturing and encoding the full second order structure is considered: we recall the definitions of the *spectral density operator* and the *weak spectral density operator*; and how they can be used for dimensionality reduction using the *Cramér-Karhunen-Loève decomposition*. Furthermore, the concept of a *linear filter* helps to construct another functional time series from the original one while providing with simple spectral analysis of the new time series and the cross-dependence relation of the two. This is done through the notions of the *frequency response function* and the *cross-spectral density operator*. These tools are particularly useful for the spectral analysis of functional linear processes. Subsection 1.3.8 includes the derivation of the spectral density operators of the FARMA(p,q) and the possibly long-range dependent FARFIMA(p,d,q) processes which are novel contributions in spectral theory of functional time series that have been presented in Rubín and Panaretos (2020c). We end the section with some inferential results in the case of fully observed functional time series data: we consider estimators of the mean function, the lagged autocovariance operators, and the spectral density operators.

Chapter 2 presents the core results of the thesis: the toolbox for estimation and prediction for sparsely observed functional time series. Firstly we rigorously establish the sparsely sampled functional time series framework for a single functional time series. We introduce the novel smoother based estimators of the mean function, the lagged autocovariance operators, and the spectral density operators. In particular, the estimated spectral density operator structure is useful in providing a consistent estimator of the entire second-order dynamics of data which can be used for functional data recovery: the prediction of the latent functional data while allowing for uncertainty quantification by means of confidence bands. Section 2.3 establishes the asymptotic behaviour justifications of the above mentioned estimators. Notably we prove consistency and convergence rates under two different sets of conditions. Firstly, we prove (suboptimal) convergence rates under the cumulant-mixing conditions by generalising the proof techniques from the independent case. Secondly, by imposing stronger conditions, the α-mixing conditions, we succeeded in obtaining the convergence rates that match the known minimax optimal convergence rates in the non-parametric literature. At the end of the section we demonstrate how to estimate the cross-dependence manifested by the cross-covariance operators and the cross-spectral density operators for pairs of time series consisting of sparsely or fully observed functional time series, and multivariate time series. The results of this chapter are primarily based on Rubín and Panaretos (2020b) and partly on Rubín and Panaretos (2020a), while extending some of the results further.

The functional lagged regression problem, the model constituting the general linear coupling between two functional time series by the means of a filter, is studied in Chapter 3. The spectral analysis of the problem reveals that the model becomes quite simple in the spectral domain, where the cross-spectral density between the response process and the regressor time series is simply a composition of the frequency response function of the filter and the spectral density operator of the regressor time series. In order to estimate this frequency response function, and hence the regression filter itself, one needs to solve an ill-posed inverse problem. We propose two methods to overcome this challenge: *spectral truncation* essentially projecting the inversion problem into finite dimensions, and *Tikhonov regularisation* adding a scaled identity operator to the uninvertible spectral density operator, making the inversion possible. We prove the consistency of the filter coefficient estimator by the both regularisation techniques when the data come from the sparsely observed functional time series and discuss, in general, the advantages and the disadvantages of both. Furthermore, we extend the functional lagged regression model for designs with multiple input regression time series including a combination of sparsely or fully observed functional time series, or multivariate time series. The contents of this chapter are based on Rubín and Panaretos (2020a) while some of the results are further developed.

Chapter 4 provides empirical demonstrations by means of simulations studies and data analyses:

- The simulation study of Section 4.1 probes the finite sample performance of the spectral density operator estimator and the functional recovery predictor of Chapter 2. In par-

ticular the study answers the design question whether the spectral density estimation reduces estimation error when the functional time series is sampled more densely but over shorter period of time, or if one should sacrifice the spatial resolution and observe the time series over a longer time-horizon. It turns out that the latter is true. Moreover the simulation study reveals that the latent functional data prediction achieves up to 60 % prediction error reduction by incorporating the temporal dependence over the estimator treating the data as independent.

- The second simulation study, presented in Section 4.2, probes the finite sample performance of the functional lagged regression estimators introduced in Chapter 3. Our simulation study reveals that neither of the two considered regularisation methods, the spectral truncation and the Tikhonov regularisation, can dominate the other as their performance depends on the spectral density operators' eigenfunction alignment and spacing.

- Section 4.3 illustrates the core methodologies introduced in Chapter 2 on the analysis of the fair-weather atmospheric electricity data measured in *Tashkent, Uzbekistan*, representing a genuine sparsely observed functional time series. Our analysis discovered and estimated the intra-day variation, the yearly-periodicity, and the intra-day covariance structure as well as dependence across days. Moreover, we illustrate the functional recovery predictor of Section 2.2.

- In Section 4.4 we analyse the dependence of the reported visibility at the *Wank mountain, Germany*, on the sparsely observed functional time series of fair-weather atmospheric electricity and the fully observed functional time series of intra-day temperatures. The comparison of the two considered regularisation techniques suggest that Tikhonov regularisation outperforms spectral truncation. Moreover, Figure 4.15 provides a transparent visualisation schema of a lagged regression model with multiple regressors.

- Finally, in Section 4.5 we analyse the dependence of the *US Treasury yield curve*, treated as a sparsely observed functional time series, on some macroeconomic variables. Our non-parametric estimation procedure confirms previous findings in the econometrics literature conducted by parametric approaches, in particular we found that the US federal funds target rate strongly impacts the short end of the yield curve.

The first simulation study (Section 4.1) and the Tashkent data analysis (Section 4.3) appeared in Rubín and Panaretos (2020b), while the second simulation study (Section 4.2) and the Wank data analysis (Section 4.4) were presented in Rubín and Panaretos (2020a). The US Treasury yield curve (Section 4.5) was analysed in Rubín (2020).

Chapter 5 presents a novel simulation method for generating functional time series samples. We introduce a general framework on how to simulate stationary functional time series in the spectral domain and how to implement the simulation algorithm in a few concrete setups. The spectral domain simulation methods enjoy often faster computational speed than

the discretisation of the spatial domain and simulation in the temporal domain approach, while constituting a universal simulation method for processes defined through their spectral properties. The results presented in this chapter are based on Rubín and Panaretos (2020c).

Finally, Chapter 6 concludes the thesis by listing some possible directions of future research development in the domain of sparsely observed functional time series.

1 Functional data analysis

This chapter provides an overview of the fundamental theory upon which the results of this thesis are built. We start from the basics of the functional data analysis, a branch of statistics considering inference problems with infinite dimensional data. Later we present the challenges of sparse data and the state-of-the art tools for serially dependent functional data, the so-called functional time series.

Section 1.1 reviews the foundations of functional data analysis (Ramsay and Silverman, 2013; Hsing and Eubank, 2015). In order to rigorously analyse functional data, a dataset composed of curves, we firstly present the theory of random variables in function spaces, typically $L^2([0,1],\mathbb{R})$, which serves as the probabilistic model for random curves. This concept requires some tools from functional analysis and operator theory while special attention is given to trace-class operators as they play an important role in the behaviour of random elements.

In Section 1.2 we revisit two main approaches how to deal with sparsely observed functional data. If the data are measured discretely and possibly with error but there is enough signal within the measurements of each curve, one may resort to a pre-processing step and smooth the discrete data to construct curves (Ramsay and Silverman, 2013). If the discrete measurements are very sparse, however, the development of a new method is needed. We review the seminal article by Yao et al. (2005a) who introduced non-parametric estimators using local-polynomial smoothers.

Section 1.3 presents the deviation from the setting of independent identically distributed random data to data sets where the curves are collected serially and dependence might be present. Such data set can arise when, for example, an intra-day development of a certain variable is recorded day-by-day, constructing a series of intra-day curves. We present the basic probabilistic and statistical concepts of such objects, called functional time series. In particular we present the spectral approach developed in Panaretos and Tavakoli (2013a,b) and Hörmann et al. (2015a,b), and review the basic results on estimation assuming a fully observed functional data.

1.1 The basis for functional data analysis

1.1.1 Operator theory background

Throughout this thesis we work with a separable Hilbert space $\mathscr{H}$ equipped with the inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\| \cdot \|$. Typically we consider $\mathscr{H} = L^2([0,1], \mathbb{R})$ or $\mathscr{H} = L^2([0,1], \mathbb{C})$ with its usual inner product $\langle f, g \rangle = \int_0^1 f(x)\overline{g(x)}\,\mathrm{d}x$ for $f, g \in L^2([0,1], \mathbb{C})$ where $\bar{x}$ denotes the complex conjugate of a complex number $x \in \mathbb{C}$. The theoretical results derived within Subsection 1.1.1 are, in general, valid for any separable Hilbert space. The results on sparsely observed functional data necessitate the concepts of continuity and smoothness and we shall model them as smooth functions in $L^2([0,1], \mathbb{R})$. Therefore we the results in the following subsections, sections, and chapters are predominantly formulated for $L^2([0,1], \mathbb{R})$.

A linear mapping $T : \mathscr{H} \to \mathscr{H}$ is called a *bounded linear operator* on $\mathscr{H}$ if

$$\|T\|_{\mathscr{L}(\mathscr{H})} \overset{def}{=} \sup\{\|Tv\|_{\mathscr{H}} : v \in \mathscr{H}, \|v\| \leq 1\} < \infty.$$

The space of bounded operators on $\mathscr{H}$, denoted as $\mathscr{L}(\mathscr{H})$, equipped with the *operator norm* $\| \cdot \|_{\mathscr{L}(\mathscr{H})}$ is a Banach space.

The *adjoint operator* to a bounded operator T is defined as the unique operator $T^* \in \mathscr{L}(\mathscr{H})$ such that

$$\langle Tu, v \rangle = \langle u, T^* v \rangle, \qquad u, v \in \mathscr{H}.$$

The operator $T \in \mathscr{L}(\mathscr{H})$ is called

- *compact* if for any bounded sequence $(v_n)_{n=1}^{\infty}$ in $\mathscr{H}$, the sequence $(Tv_n)_{n=1}^{\infty}$ contains a convergent subsequence,
- *self-adjoint* if $T = T^*$,
- *non-negative definite* if

$$\langle Tu, u \rangle \geq 0, \qquad u \in \mathscr{H}.$$

In *complex* Hilbert spaces, non-negative definite operators are necessarily self-adjoint.

Compact operators admit the following decompositions as sums of rank one operators defined by the notion of the *tensor product*. For vectors/functions $f, g \in \mathscr{H}$ defined their tensor product as the operator $f \otimes g : \mathscr{H} \to \mathscr{H}$, $v \mapsto \langle v, g \rangle f$.

Proposition 1.1.1. *A compact operator $T \in \mathscr{L}(\mathscr{H})$ admits a singular value decomposition. Specifically there exist a sequence of non-negative real numbers $\{\sigma_j\}_{j=1}^{\infty}$ decreasing to zero called the singular values of T, and orthonormal sequences $\{\varphi_j\}_{j=1}^{\infty}$, $\{\psi_j\}_{j=1}^{\infty}$ in $\mathscr{H}$ such that*

$$T = \sum_{j=1}^{\infty} \sigma_j \left(\varphi_j \otimes \psi_j\right) = \sum_{j=1}^{\infty} \sigma_j \langle \cdot, \psi_j \rangle \varphi_j$$

where the sum is understood in the operator norm. If $T \in \mathscr{L}(\mathscr{H})$ is moreover self-adjoint then it admits the spectral decomposition, i.e. there exists a sequence of real numbers $\{\lambda_j\}_{j=1}^{\infty}$ converging to zero called the eigenvalues of T, and an orthonormal sequence $\{\varphi_j\}_{j=1}^{\infty}$ in $\mathscr{H}$ called the eigenfunctions of T such that

$$T = \sum_{j=1}^{\infty} \lambda_j \left(\varphi_j \otimes \varphi_j\right) = \sum_{j=1}^{\infty} \lambda_j \langle \cdot, \varphi_j \rangle \varphi_j$$

where the sum is understood in the operator norm. If $T \in \mathscr{L}(\mathscr{H})$ is furthermore non-negative definite, the eigenvalues of T are real and non-negative.

The class of compact operators can be further refined to the two following classes of operators.

- An operator $T \in \mathscr{L}(\mathscr{H})$ is called *Hilbert-Schmidt* if for any orthonormal basis $(e_j)_{j=1}^{\infty}$ in $\mathscr{H}$,

  $$\|T\|_2^2 \overset{def}{=} \sum_{j=1}^{\infty} \left\|Te_j\right\|^2 < \infty.$$

  In that case, $\|T\|_2$ is the *Hilbert-Schmidt norm* of T and its definition is independent of the choice of $(e_j)_{j=1}^{\infty}$. The space of Hilbert-Schmidt operators, denoted as $\mathscr{L}_2(\mathscr{H})$, is itself a separable Hilbert space with the inner product

  $$\langle T_1, T_2 \rangle_2 = \sum_{j=1}^{\infty} \left\langle T_1 e_j, T_2 e_j \right\rangle$$

  for $T_1, T_2 \in \mathscr{H}$ where the definition is again independent of the choice of the orthonormal basis $(e_j)_{j=1}^{\infty}$. Moreover, if $T \in \mathscr{L}_2(\mathscr{H})$ is furthermore self-adjoint then we may choose $e_j \equiv \varphi_j$ to be the eigenfunction basis and therefore $\|T\|_2 = \left(\sum_{j=1}^{\infty} \lambda_j^2\right)^{1/2}$.

- An operator $T \in \mathscr{L}(\mathscr{H})$ is called *trace-class* (or *nuclear*) if for any orthonormal basis $(e_j)_{j=1}^{\infty}$ in $\mathscr{H}$

  $$\|T\|_1 \overset{def}{=} \sum_{j=1}^{\infty} \left\langle \left(T^*T\right)^{1/2} e_j, e_j \right\rangle < \infty$$

  where $(T^*T)^{1/2}$ is the so-called *square root operator* of the non-negative self-adjoint operator T^*T, i.e. $(T^*T)^{1/2}(T^*T)^{1/2} = T^*T$. Such operator exists uniquely.

  In that case, $\|T\|_1$ is the *trace norm* of T and we may define the *trace* of T as

  $$\operatorname{Tr} T = \sum_{j=1}^{\infty} \left\langle Te_j, e_j \right\rangle.$$

  The definitions are independent of the choice of the orthonormal basis $(e_j)_{j=1}^{\infty}$. If T is furthermore self-adjoint we may set $e_j \equiv \varphi_j$ to be the eigenfunction basis and then $\|T\|_1 = \sum_{j=1}^{\infty} |\lambda_j|$.

The space of trace-class operators is denoted $\mathscr{L}_1(\mathscr{H})$ and, being equipped with the norm $\|T\|_1$, is a Banach space.

The above defined classes of operators satisfy the following chain of inclusions

$$\mathscr{L}_1(\mathscr{H}) \subset \mathscr{L}_2(\mathscr{H}) \subset \mathscr{L}_{compact}(\mathscr{H}) \subset \mathscr{L}(\mathscr{H})$$

where $\mathscr{L}_{compact}(\mathscr{H})$ is the set of compact operators on $\mathscr{H}$.

1.1.2 Kernel functions, integral operators, and Mercer's theorem

The key objects of interest of this thesis are covariance operators and spectral density operators (to be defined later in this chapter) and the crucial ingredient for their definition are the notions of a bivariate *kernel* and the *integral operator*, and their properties. For the purpose of this subsection we consider $\mathscr{H} = L^2([0,1],\mathbb{R})$, i.e. the square integrable function with the domain $[0,1]$ but any other compact interval is appropriate.

A *kernel* function is a bivariate function $K : [0,1] \times [0,1] \to \mathbb{C}$. The kernel function K induces by right integration a linear operator $\mathcal{K}$:

$$\left(\mathcal{K}f\right)(s) = \int_0^1 K(s,t) f(t)\,\mathrm{d}t, \qquad f \in L^2([0,1],\mathbb{C}), \quad s \in [0,1]. \tag{1.1}$$

A sufficient condition for the operator defined by (1.1) is, for example, continuity of the kernel K. In that case, the operator $\mathcal{K}$ is moreover compact.

The kernel function K is called non-negative definite if $\int_0^1 \int_0^1 K(s,t) f(s) f(t)\,\mathrm{d}s\,\mathrm{d}t \geq 0$ for all $f \in L^2([0,1],\mathbb{C})$. In fact, the kernel function K is non-negative definite if and only if the operator $\mathcal{K}$ is non-negative definite. Moreover, the kernel function K is symmetric (i.e. $K(s,t) = \overline{K(t,s)}$ for $s, t \in [0,1]$) if and only if $\mathcal{K}$ is self-adjoint.

In fact, the integral operators with square integrable kernels correspond to Hilbert-Schmidt operators by Hilbert-Schmidt kernel theorem.

Theorem 1.1.2 (Heil (2018)[Thm 8.4.8])**.**

(i) Let $K(\cdot,\cdot)$ be a kernel function such that $\int_0^1 \int_0^1 |K(s,t)|^2\,\mathrm{d}s\,\mathrm{d}t < \infty$. Then the integral operator $\mathcal{K}$ defined by (1.1) *is Hilbert-Schmidt and $\int_0^1 \int_0^1 |K(s,t)|^2\,\mathrm{d}s\,\mathrm{d}t = \|\mathcal{K}\|_2^2$.*

(ii) Let $T \in \mathscr{L}_2(L^2([0,1],\mathbb{C}))$, i.e. let T be a Hilbert-Schmidt operator on $L^2([0,1],\mathbb{C})$. Then there exist a kernel function $K : [0,1] \times [0,1] \to \mathbb{C}$ satisfying $\int_0^1 \int_0^1 |K(s,t)|^2\,\mathrm{d}s\,\mathrm{d}t < \infty$ such that $K(\cdot,\cdot)$ induces the operator T. Formally, for its induced integral operator $\mathcal{K}$ defined by (1.1) *holds: $\mathcal{K} = T$.*

We give a special attention to symmetric and non-negative definite kernels because, as we later see, they represent key components in functional data analysis (Subsection 1.1.3) and

the spectral analysis of functional time series (Subsection 1.3.4). Moreover, the symmetric and non-negative definite kernel K can be represented by the eigenvalues and eigenfunctions of $\mathcal{K}$ by Mercer's theorem:

Theorem 1.1.3 (Hsing and Eubank (2015)[Thm 4.6.5])**.** *Let $K(\cdot,\cdot)$ be a continuous, symmetric, non-negative kernel and $\mathcal{K}$ its corresponding integral operator. Denote $\{\lambda_j\}_{j=1}^{\infty}$ the eigenvalues and $\{\varphi_j\}_{j=1}^{\infty}$ the eigenfunctions of $\mathcal{K}$. Then $\{\varphi_j\}_{j=1}^{\infty}$ are continuous functions in $L^2([0,1],\mathbb{C})$ and the kernel K admits the representation*

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(s)\varphi_j(t), \qquad s,t \in [0,1],$$

where the sum converges absolutely and uniformly.

Corollary 1.1.4 (Hsing and Eubank (2015)[Thm 4.6.7])**.** *Under the conditions of Theorem 1.1.3, the operator $\mathcal{K}$ is trace-class and*

$$\operatorname{Tr}\mathcal{K} = \|\mathcal{K}\|_1 = \int_0^1 K(t,t)\,\mathrm{d}t,$$
$$\|\mathcal{K}\|_2^2 = \int_0^1 \int_0^1 K^2(s,t)\,\mathrm{d}s\,\mathrm{d}t.$$

1.1.3 Random elements in infinite dimensional spaces

Throughout the this section and the whole thesis we assume that there is a given complete probability space $(\Omega,\mathcal{F},\mathbb{P})$ large enough to allow for the existence of the all considered random objects. In functional data analysis it is common to consider a random curve as a stochastic process, i.e. a collection of random variables $X \equiv \{X(x) : x \in [0,1]\}$ where $X(x)$ is a $\mathcal{F}$-measurable random variable for each $x \in [0,1]$. However, this fact alone, together with the existence of the second moments, does not automatically guarantee that the process X is a $\mathcal{F}$-measurable random element in $L^2([0,1],\mathbb{R})$. There are two techniques how to overcome this issue: either one assumes that the stochastic process takes values in a reproducing kernel Hilbert space (RKHS) (Berlinet and Thomas-Agnan, 2011) or requiring the stochastic process to be mean-square continuous (Hsing and Eubank, 2015). In the following we outline the second approach starting with some basic definitions for stochastic processes.

The *mean function* μ of the stochastic process X is defined as

$$\mu(x) = \mathbb{E}[X(x)], \qquad x \in [0,1], \tag{1.2}$$

and its *covariance kernel* R^X as

$$R^X(x,y) = \mathbb{E}\left[\big(X(x)-\mu(x)\big)\big(X(y)-\mu(y)\big)\right], \qquad x,y \in [0,1]. \tag{1.3}$$

The *covariance operator* $\mathscr{R}^X$ is induced by the right-hand integration (1.1) from the covariance kernel.

Throughout the thesis we assume that the random process X satisfies

$$\lim_{n\to\infty} \mathbb{E}\left[\left| X(x_n) - X(x) \right|^2\right] = 0$$

for all sequences $\{x_n\}_{n=1}^{\infty} \subset [0,1]$ such that $x_n \to x \in [0,1]$. Such process is called *mean-square continuous*.

Proposition 1.1.5 (Hsing and Eubank (2015)[Thm 7.3.2])**.** *Let $X \equiv \{X(x) : x \in [0,1]\}$ be a stochastic process with finite second moments. Then the following statements are equivalent.*

(i) The process X is mean-square continuous.

(ii) Its mean function is continuous on $[0,1]$ and its covariance kernel is continuous on $[0,1]^2$.

(iii) Its mean function is continuous on $[0,1]$ and its covariance kernel is continuous on the diagonal of $[0,1]^2$.

Consider now the stochastic process X as a mapping jointly from the curves' domain $[0,1]$ and the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. as a function $(x,\omega) \mapsto X(x,\omega)$. The following theorem provides with a sufficient condition for X to be well-defined as a random element in $L^2([0,1],\mathbb{R})$.

Theorem 1.1.6 (Hsing and Eubank (2015)[Theorems 7.4.1 and 7.4.2])**.** *Let X be a mean-square continuous stochastic process. Moreover assume that the mappings $x \mapsto X(\omega,x)$, $x \in [0,1]$, are continuous for each $\omega \in \Omega$. Then the mapping $(x,\omega) \mapsto X(x,\omega)$ is jointly measurable and the mapping $\omega \mapsto X(\omega,\cdot)$ is an $\mathcal{F}$-measurable mapping to $L^2([0,1],\mathbb{R})$, i.e. X is a random element in $L^2([0,1])$.*

From now on we shall adopt this dual perspective and view the object X as both a stochastic process with continuous trajectories with well-defined evaluations $X(x)$ and as a random element in $L^2([0,1],\mathbb{R})$. Consequently, the mean function μ defined in (1.2) is an element of $L^2([0,1],\mathbb{R})$ and the covariance operator $\mathscr{R}^X$ associated with the covariance kernel (1.3) can be written as

$$\mathscr{R}^X = \mathbb{E}\left[\left(X-\mu\right) \otimes \left(X-\mu\right)\right]$$

where $\otimes$ is the tensor product defined as $f \otimes g = \langle \cdot, f \rangle g$ as an operator on $L^2([0,1],\mathbb{R})$. Furthermore $\mathbb{E}\left[\|X\|^2\right] < \infty$.

The covariance operator is the central object of interest in functional data analysis because it is the key to optimal dimensionality reduction. The celebrated Karhunen-Loève expansion (Karhunen, 1946; Loève, 1946; Ash and Gardner, 2014; Grenander, 1981) provides with optimal finite rank approximation scheme for stochastic processes.

Theorem 1.1.7 (Karhunen-Loève expansion)**.** *Let $X \in L^2([0,1],\mathbb{R})$ be a mean-square continuous stochastic process with continuous sample trajectories with mean function μ, covariance*

kernel having decomposition by Mercer's theorem 1.1.3, $R^X(x,y) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(x)\varphi_j(y)$ *where* $\{\lambda_j, \varphi_j\}_{j=1}^{\infty}$ *is the eigendecomposition of the covariance operator* $\mathscr{R}^X$*. Then the process* X *can be approximated by the sums*

$$X_n(x) = \mu(x) + \sum_{j=1}^{n} \xi_j \varphi_j(x), \qquad x \in [0,1]. \tag{1.4}$$

where $\xi_j = \langle X - \mu, \varphi_j \rangle$, $j \in \mathbb{N}$ *in the two following regimes of convergence*

$$\sup_{x\in[0,1]} \mathbb{E}\left[|X(x) - X_n(x)|^2\right] \to 0, \qquad as \quad n \to \infty, \tag{1.5}$$

$$\mathbb{E}\left[\|X - X_n\|^2\right] \to 0, \qquad as \quad n \to \infty. \tag{1.6}$$

For $j,k \in \mathbb{N}$, $\mathbb{E}\xi_j = 0$ *and* $\mathbb{E}\xi_j\xi_k = \lambda_j \delta_{j,k}$ *with* $\delta_{j,k} = 1$ *if* $j = k$ *and zero otherwise. Moreover, the representation* (1.4) *is optimal. Fix* $n \in \mathbb{N}$ *and take an arbitrary orthonormal sequence* $\{e_j\}_{j=1}^{\infty}$*. Then*

$$\mathbb{E}\left[\left\| X - \mu - \sum_{j=1}^{n} \left\langle X - \mu, e_j \right\rangle e_j \right\|^2\right] \geq \mathbb{E}\left[\|X - X_n\|^2\right]. \tag{1.7}$$

The optimality (1.7) of the Karhunen-Loève expansion (1.4) is crucial for functional principal component analysis (Grenander, 1950; Dauxois et al., 1982) as a dimensionality reduction technique. Moreover, it constitutes a vehicle for functional linear models (Müller, 2005; Morris, 2015) and clustering algorithms (Chiou and Li, 2008; Leng and Müller, 2006; Peng and Müller, 2008), the serves as a basis for regularisation of ill-posed inverse problems (Panaretos et al., 2010; Wang et al., 2016)

The above stated results can be straightforwardly extended to complex-valued stochastic processes and random elements in $L^2([0,1],\mathbb{C})$.

1.1.4 Inference for fully observed functional data

In this section we review the basic inference results for the first and the second order structure of functional data. The setting presented here can be seen as abstract because we shall assume that we fully observed n curves, specifically we observe independent identically distributed $L^2([0,1],\mathbb{R})$-valued random elements $X_1,\ldots,X_n$ with $\mathbb{E}\|X_1\|^2 < \infty$.

The natural estimators of the mean function μ and the covariance kernel R^X of X_1 are given by

$$\hat{\mu}(x) = \frac{1}{n}\sum_{j=1}^{n} X_j(x), \qquad x \in [0,1],$$

$$\hat{R}^X(x,y) = \frac{1}{n}\sum_{j=1}^{n} \left(X_j(x) - \hat{\mu}(x)\right)\left(X_j(y) - \hat{\mu}(y)\right), \qquad x,y \in [0,1].$$

Denote $\hat{\mathscr{R}}^X$ the integral operator induced by the kernel $\hat{R}^X$.

The analysis of the consistency and the asymptotic distribution of the estimators are direct consequences of the functional versions of the strong law of large numbers and the central limit theorem. Just as in the univariate and the multivariate case, the central limit theorem involves convergence to the normal distribution whose definition in infinite dimensional spaces is reviewed bellow.

Definition 1.1.8. The $\mathscr{H}$-valued random element X is said to follow a normal distribution with the mean $\mu^X \in \mathscr{H}$ and the covariance operator $\mathscr{R}^X \in \mathscr{L}_1(\mathscr{H})$, denoting $X \sim N(\mu^X, \mathscr{R}^X)$, if

$$\langle X - \mu^X, f \rangle \sim N\big(0, \langle \mathscr{R}^X f, f \rangle\big)$$

for all $f \in \mathscr{H}$.

Proposition 1.1.9 (Bosq (2000)[Thms 2.4, 2.7])**.** *Let $\{Y_j\}_{j=1}^{\infty}$ be a sequence of independent identically distributed random elements in $\mathscr{H}$.*

- *Assume that $\mathbb{E}Y_1 = \mu_Y$ is well defined. Then*

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{a.s.} \mu_Y, \qquad as \quad n \to \infty.$$

- *Assume that $\mathbb{E}\|Y_1\|^2 < \infty$ and denote $\mu_Y = \mathbb{E}Y_1$ and $\mathscr{R}^Y = \mathbb{E}[(Y_1 - \mu_Y) \otimes (Y_1 - \mu_Y)]$. Then*

$$n^{-1/2}\sum_{i=1}^{n} \big(Y_i - \mu_Y\big) \xrightarrow{d} Z, \qquad as \quad n \to \infty,$$

  where $Z \sim N(0, \mathscr{R}^Y)$.

Corollary 1.1.10. *Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of independent identically distributed functional data in $L^2([0,1],\mathbb{R})$ with $\mathbb{E}\|X_1\|^2 < \infty$, and denoting the mean $\mu = \mathbb{E}X_1$ and the covariance operator $\mathscr{R}^X = \mathbb{E}[(X_1 - \mu) \otimes (X_1 - \mu)]$.*

(i) Then the empirical mean is a consistent estimator

$$\hat{\mu} \xrightarrow{a.s.} \mu, \qquad as \quad n \to \infty,$$

and asymptotically normal

$$n^{1/2}\big(\hat{\mu} - \mu\big) \xrightarrow{d} Z_1, \qquad as \quad n \to \infty,$$

where $Z_1 \sim N(0, \mathscr{R}^X)$ and the above convergences are understood in the $L^2([0,1],\mathbb{R})$-norm.

(ii) The empirical covariance operator is consistent

$$\hat{\mathscr{R}}^X \xrightarrow{a.s.} \mathscr{R}^X, \qquad as \quad n \to \infty,$$

and, assuming further $\mathbb{E}\|X_1\|^4 < \infty$*, it is also asymptotically normal*

$$n^{1/2}\left(\hat{\mathscr{R}}^X - \mathscr{R}^X\right) \xrightarrow{d} Z_2, \qquad as \quad n \to \infty,$$

where Z_2 *is a* $\mathscr{L}_2(L^2([0,1],\mathbb{R}))$*-valued random element with the covariance operator* $\mathscr{R}^{Z_2}$ *given by*

$$\mathscr{R}^{Z_2} = \mathbb{E}[Z_2 \otimes Z_2] = \mathbb{E}\left[\left\{(X-\mu)\otimes(X-\mu) - \mathscr{R}^X\right\} \tilde{\otimes} \left\{(X-\mu)\otimes(X-\mu) - \mathscr{R}^X\right\}\right]$$

where $\tilde{\otimes}$ *is the tensor product on* $\mathscr{L}_2(L^2([0,1],\mathbb{R}))$ *defined for* $\mathscr{A},\mathscr{B} \in \mathscr{L}_2(L^2([0,1],\mathbb{R}))$ *as an operator given by* $(\mathscr{A}\tilde{\otimes}\mathscr{B})\mathscr{C} = \langle \mathscr{C}, \mathscr{A}\rangle_{\mathscr{L}_2(L^2([0,1],\mathbb{R}))}\mathscr{B}$ *for* $\mathscr{C} \in \mathscr{L}_2(L^2([0,1],\mathbb{R}))$*. The convergences above are understood in the* $\mathscr{L}_2(L^2([0,1],\mathbb{R}))$*-norm.*

Proof. The part (i) is a direct application of Proposition 1.1.9, setting $Y_i = X_i, i \in \mathbb{N}$, and $\mathscr{H} = L^2([0,1],\mathbb{R})$. The part (ii) is again an application of Proposition 1.1.9 but this time to the random elements $Y_i = (X_i - \mu)\otimes(X_i-\mu)$, $i \in \mathbb{N}$, and the Hilbert space $\mathscr{H} = \mathscr{L}_2(L^2([0,1],\mathbb{R}))$. □

1.2 Sparsely observed functional data

1.2.1 Different regimes of sparsity

The theoretical setting of fully observed functional data as introduced in Subsection 1.1.4 cannot be realistic in any application. Due to the impossibility of storing infinite dimensional data, the functional data are recorder only by means of finite number of measurements. Throughout this thesis we consider the sampling regime manifested by point-wise evaluation of the functions on a finite grid which are further corrupted by an additive noise, often interpreted as a measurement error. We consider the usual model (Shi et al., 1996; Staniswalis and Lee, 1998; Rice and Wu, 2001; Yao et al., 2005a) for functional data:

$$U_{ij} = X_i(x_{ij}) + \epsilon_{ij}, \qquad i = 1,\dots,n, \quad j = 1,\dots,N_i, \tag{1.8}$$

where x_{ij} is the measurement location for the j-th observation on the i-th curve for $i = 1,\dots,n$, $j = 1,\dots,N_i$ with N_i denoting the number of observations available on the i-th curve. The measurement noise contamination $\{\epsilon_{ij}\}$ is assumed to be an ensemble of independent identically distributed zero-mean random variables with variance $\sigma^2 > 0$ which are independent on the underlying functional data $X_1,\dots,X_n$.

Suitable estimation procedures and their performances obviously depend on the number of measurements available per curve. Generally speaking, we may split the discussion into two regimes:

- *Dense sampling.* When each curve is sampled with high frequency, we have enough information to well estimate the curve X_i from the observations realised on this curve

U_{ij}, $j = 1,\dots,N_i$. The estimation of the functional data X_i, done separately for each $i = 1,\dots,n$, relies on smoothing methods (Ramsay and Silverman, 2013, 2007) some of which are reviewed in Subsection 1.2.2. This procedure is the reason why we speak about "smooth-first-then-estimate" approach because once the pre-smoothing step is done we work with the estimates as if we were given fully observed functional data.

The asymptotic theory derived under dense sampling regime provides with an interesting observation. If the resolution of the discrete sampling is sufficiently dense, for example if $\min_{i=1,\dots,n} N_i \gg n^{1/4}$ an the gaps between measurements are uniform, then the convergence rates for the estimation of the mean function μ and the covariance operator $\mathscr{R}^X$ are still of order $\sqrt{n}$ (Hall et al., 2006; Li and Hsing, 2010), just like in the case of fully observed functional data (Corollary 1.1.10). This essentially means that the estimator from such densely sampled discretely measured data are as good as if we had access to the full functional data instead, at least asymptotically speaking.

- *Sparse sampling.* In many applications, however, the functional data are not sampled densely. The number of observation locations can be quite low in comparison with the sample size n, maybe even bounded $\max_{i=1,\dots,n} N_i \le C < \infty$. Under such genuinely sparse sampling regime, a development of new methods under the name "estimate-first-then-smooth" is required. These methods use the information across the entire data sample when constructing the estimates for individual curves, thus "borrow strength" from the other observations. In Subsection 1.2.3, we review the smoother based estimators introduced in the article by Yao et al. (2005a).

1.2.2 "Smooth-first-then-estimate" approach

The "smooth-first-then-estimate" approach views the transition from the discretely recorded date into continuously defined functional data simply a preprocessing step. This approach was popularised by Ramsay and Silverman (2013) who also explains the methodology by applications on a handful of case studies in (Ramsay and Silverman, 2007).

Considering the sampling regime (1.8) with a regular grid, i.e. $N_1 = \cdots = N_n \equiv N$ and $x_{1j} = \cdots = x_{nj} \equiv x_j$, $j = 1,\dots,N$. The classical path is to express the latent functional data $X_1,\dots,X_n$ as a linear combination of a fixed and finite number p of basis functions $\varphi_1,\dots,\varphi_K$:

$$X_i(x) = \sum_{k=1}^{K} c_{ik}\varphi_k(x), \qquad x \in [0,1], \quad i = 1,\dots,n, \tag{1.9}$$

where $\{c_{ij}\}$ are unknown coefficients. The common choice of the basis functions $\varphi_1,\dots,\varphi_K$ include the B-splines, which are efficient in expressing smooth curves (Ramsay and Silverman, 2007), the Fourier basis for periodic data (Ramsay and Silverman, 2007), and the wavelets, which are capable to capture rapid changes even with a low number of basis elements (Mallat, 1999).

The usual method to estimate the unknown coefficients $c_i = (c_{i1}, \ldots, c_{ik})$ by minimising the ordinary least squares

$$\sum_{j=1}^{N} \left(U_{ij} - \sum_{k=1}^{K} c_{ik} \varphi_k(x_j) \right)^2 = (U_i - \Phi c_i)^\top (U_i - \Phi c_i) = \| U_i - \Phi c_i \|^2 \tag{1.10}$$

where the $U_i = (U_{i1}, \ldots, U_{iN_i})$ and Φ is a $N \times K$ matrix defined as $\{\varphi_k(x_j)\}$. The ordinary least squares estimator of c_i is given by

$$c_i = \left(\Phi^\top \Phi \right)^{-1} \Phi^\top U_i.$$

Besides the basis choice, which is usually done by domain knowledge and expected properties of the smooth curves as discussed above, the goodness of fit of the representation (1.9) depends on the specification of the number of basis functions K. The selection of K come down to a typical bias-variance trade-off. A small value K could mean that the representation (1.9) cannot fit well the underlying functional data, while a large value of K results in over-fitting meaning the interpolation starts to pick up the noise in the measurements and the visualised fit would feature to many abrupt changes.

Because the selection of the discrete parameter K might be challenging, an alternative to the ordinary least squares fit (1.10) is to impose roughness penalisation into the loss function (1.10). This approach essentially shifts the selection problem into a continuous tuning parameter and usually provides a better fit to real data (Ramsay and Silverman, 2007). The penalised least squares estimator f of the functional datum X_i becomes

$$f = \underset{f \in C^2([0,1])}{\arg\min} \left\{ \left(U_{ij} - f(x_j) \right)^2 + \lambda \int_0^1 \left[\frac{\partial^2}{\partial x^2} f(x) \right]^2 \mathrm{d}x \right\} \tag{1.11}$$

where the minimisation runs through $C^2([0,1])$, i.e. the space of twice continuously differentiable functions on $[0,1]$, and where $\lambda > 0$ is a tuning parameter called the smoothing parameter.

Remarkably, the minimum of the infinite dimensional optimisation problem (1.11) lays in a finite dimensional vector space. Using the framework of reproducing kernel Hilbert spaces it can be shown that the solutions is in fact a cubic spline (de Boor, 1978; Heckman, 2012). The penalisation in (1.11) is based on the second derivative of the functional data, thus it is trying to push the fitted curves towards linear functions and it penalises quick deviations from the linearity. For a more general discussion of the use of roughness penalty in statistics see, e.g. Green and Silverman (1993). It should be noted that the loss function in (1.11) penalising the roughness is only one of the possibilities and any penalty given by a reproducing kernel Hilbert space norm is applicable and the conclusions of the discussion in this paragraph apply.

Alternative smoothing methods to obtain the estimates of functional data include kernel

smoothing, localised basis or polynomial expansions (Wand and Jones, 1994; Fan and Gijbels, 1996; Efromovich, 2008).

1.2.3 Kernel smoothing based estimators

Estimation of model dynamics by kernel smoothers

In this section we are going to present in more detail the kernel smoothing methods (Fan and Gijbels, 1996) (also known as local polynomial regression) for the estimation of the first and the second order structure of the data distribution. Not only do these objects produce insights for the data dynamics interpretation, but they can also be used to estimate the latent functional data themselves.

We are again considering the sampling scheme (1.8) but, this time, we consider the more difficult case with the numbers of observations N_i quite small for each curve and where the observation grid $\{x_{tj}\}$ can be irregular. This challenging setting is also commonly called *longitudinal data*.

The foundational article considering the connections between longitudinal data analysis and the functional data analysis is due to Yao et al. (2005a) who proposed to estimate the mean function μ and the covariance kernel $R^X(\cdot,\cdot)$ by the means of the kernel smoothers and established a method to estimate the principal scores through calculating the conditional expectation (PACE), which we are going to review here.

Kernel smoothing is a non-parametric technique to estimate the regression function by performing a polynomial regression at each input point (evaluation point), incorporating always only the neighbouring points. Moreover, the observations are weighted in terms of the distance from the each evaluation point by the means of a kernel function. Throughout this presentation, we work with the Epanechnikov kernel defined as $K(v) = (3/4)(1-v^2)$ for $v \in [-1,1]$ and zero otherwise, but any other usual choice of kernels is appropriate. For other kernels and their comparison, see Fan and Gijbels (1996)[§3.2.6].

First of all, let us estimate the mean function μ using the local-liner smoother. Yao et al. (2005a) propose to pool all the data and look at the scatter plot $\{(x_{ij}, U_{ij}) : j = 1,\dots,N_i; i = 1,\dots,n\}$. Because the additive measurement errors $\{\epsilon_{ij}\}$ in the scheme (1.8) have zero mean, it immediately follows that $\mathbb{E}\left[U_{ij}|x_{ij}\right] = \mu(x_{ij})$ and therefore the mean function can be estimated by smoothing the said scatter plot. In particular, Yao et al. (2005a) propose to use the local-linear smoother to estimate the mean function at point $x \in [0,1]$ as $\hat{\mu}(x) = \hat{c}_0^{(1)}$ by minimising the following sum of squares

$$\left(\hat{c}_0^{(1)}, \hat{c}_1^{(1)}\right) = \underset{c_0^{(1)}, c_1^{(1)}}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K\left(\frac{x - x_{ij}}{B_\mu}\right) \left\{U_{ij} - c_0^{(1)} - c_1^{(1)}(x - x_{ij})\right\}^2$$

where $B_\mu > 0$ is the bandwidth parameter for the smoothing.

Next, we now focus on the second order structure of the data and want to estimate the covariance kernel $R^X(\cdot,\cdot)$. Yao et al. (2005a) define the "raw" covariances as

$$G_i(x_{ij},x_{ik}) = (U_{ij} - \hat{\mu}(x_{ij}))(U_{ik} - \hat{\mu}(x_{ik})).$$

We anticipate that

$$\mathbb{E}\left[G_i(x_{ij},x_{ik})|x_{ij},x_{ik}\right] \approx R^X(x_{ij},x_{ik}) + \sigma^2\delta_{jk} \tag{1.12}$$

where $\delta_{jk} = 1$ if $i = k$ and zero otherwise. The deviation of the expectation on the left-hand side of (1.12) from the expression on the right-hand side stems only from the estimator $\hat{\mu}$. If we knew the true value of μ, (1.12) would be an equality. Moreover, (1.12) reveals that the additive measurement errors $\{\epsilon_{ij}\}$ contribute only to the diagonal of the covariance kernel, which was already observed by Staniswalis and Lee (1998). Thus, Yao et al. (2005a) propose to estimate the covariance kernel $R^X(\cdot,\cdot)$ by smoothing the three dimensional scatter plot $\{(x_{ij},x_{ik},G_i(x_{ij},x_{ik})) : j,k = 1,\dots,N_i\ j \neq k; i = 1,\dots,n\}$ by a local-linear surface smoother. The covariance kernel at point $(x,y) \in [0,1]^2$ is estimated by $\hat{R}^X(x,y) = \hat{c}_0^{(2)}$ where $\hat{c}_0^{(2)}$ is obtained by solving the following least squares problem

$$\left(\hat{c}_0^{(2)},\hat{c}_1^{(2)},\hat{c}_2^{(2)}\right) = \underset{c_0^{(2)},c_1^{(2)},c_2^{(2)}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j\neq k} K\left(\frac{x-x_{ij}}{B_R}\right)K\left(\frac{x-x_{ik}}{B_R}\right) \times \left\{G_i(x_{ij},x_{ik}) - c_0^{(2)} - c_1^{(2)}(x-x_{ij}) - c_2^{(2)}(x-x_{ik})\right\}^2$$

where $B_R > 0$ is the bandwidth parameter.

Finally we aim to estimate the measurement error variance σ^2 by the approach suggested by (Yao et al., 2003, 2005a) for which we need the two following ingredients: the estimator of the diagonal of the covariance kernel of $\{X_t\}$ with and without the measurement noise contamination. Firstly, we estimate the diagonal of $R^X(\cdot,\cdot)$ without the measurement noise contamination by the local-quadratic smoother along the direction perpendicular to the diagonal. For $x \in [0,1]$ we set $\bar{R}^X(x) = \hat{c}_0^{(3)}$ where

$$\left(\hat{c}_0^{(3)},\hat{c}_1^{(3)},\hat{c}_2^{(3)}\right) = \underset{c_0^{(3)},c_1^{(3)},c_2^{(3)}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq k}}^{N_i}\sum_{k=1}^{N_i} K\left(\frac{x_{ij}-x}{B_R}\right)K\left(\frac{x_{ik}-x}{B_R}\right) \times \left\{G_i(x_{ij},x_{ik}) - \hat{c}_0^{(3)} - \hat{c}_1^{(3)}\Delta(x_{ij},x_{ik}) - \hat{c}_2^{(3)}\Delta(x_{ij},x_{ik})^2\right\}^2$$

where $\Delta(x_{ij},x_{ik})$ is the distance of the point (x_{ij},x_{ik}) from the diagonal equipped with the positive sign if the point (x_{ij},x_{ik}) is above the diagonal, and with the negative sign if below. Formally

$$\Delta(x_{ij},x_{ik}) = \operatorname{sign}\left(x_{ik}-x_{ij}\right)\sqrt{\left(P(x_{ij},x_{ik})-x_{ij}\right)^2 + \left(P(x_{ij},x_{ik})-x_{ik}\right)^2} \tag{1.13}$$

where $\operatorname{sign}(\cdot) \in \{-1,0,1\}$ is the sign function and $P(x_{ij},x_{ik})$ is the first coordinate of the point

(x_{ij}, x_{ik}) projected onto the diagonal of $[0,1]^2$.

Secondly, we estimate the function $x \mapsto R^X(x,x) + \sigma^2$, $x \in [0,1]$, i.e. the noise contaminated diagonal of the covariance kernel $R^X(\cdot,\cdot)$. For $x \in [0,1]$ and a bandwidth parameter $B_V > 0$, we use the local-linear line smoother and set $\hat{V}(x) = \hat{c}_0^{(4)}$ where

$$\left(\hat{c}_0^{(4)}, \hat{c}_1^{(4)}\right) = \underset{c_0^{(4)}, c_1^{(4)}}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{N_t} K\left(\frac{x_{ij} - x}{B_V}\right) \left\{G_i(x_{ij}, x_{ij}) - c_0^{(3)} - c_1^{(3)}(x - x_{ij})\right\}^2.$$

Having the estimates $\bar{R}(\cdot)$ and $\hat{V}(\cdot)$, the measurement error variance σ^2 is estimated by integrating the difference

$$\hat{\sigma}^2 = \int_0^1 \left(\hat{V}(x) - \bar{R}^X(x)\right) \mathrm{d}x. \tag{1.14}$$

In case the right-hand side of (1.14) is negative, we recommend to replace it by a small positive number following the methodology of Yao et al. (2005a).

Principal components analysis through conditional expectation (PACE)

Once the above presented estimators are established, Yao et al. (2005a) explained how to estimate the Principal Component scores by Conditional Expectation (PACE). We start the review of their methodology by the description of the case with known model dynamics, i.e. with the known parameters μ, R^X, σ^2, and later we plug-in their estimated values. The covariance operator $\mathscr{R}^X$ corresponding to the covariance kernel R^X is self-adjoint and non-negative definite, therefore it features the spectral decomposition

$$\mathscr{R}^X = \sum_{k=1}^{\infty} \lambda_k \varphi_k \otimes \varphi_k$$

where $\{\lambda_k\}_{k=1}^{\infty}$ are its eigenvalues and $\{\varphi_k\}_{k=1}^{\infty}$ the eigenfunctions.

In the setting of the densely observed functional data, the principal component scores are traditionally estimated by approximating the integral

$$\xi_{ik} = \int_0^1 \left(X_i(x) - \mu(x)\right) \varphi_k(x) \, dt$$

by the approximating sum

$$\hat{\xi}_{ik} = \sum_{j=2}^{N_i} \left(U_{ij} - \hat{\mu}(x_{ij})\right) \varphi_k(x_{ij})(x_{ij} - x_{i,j-1}).$$

The limitation of this approach is that the summation cannot approximate ξ_{ik} well in the sparse design. Moreover, the values of X_i are not observed and one has to substitute U_{ij} for $X_i(x_{ij})$ regardless of the presence of measurement errors and additional shrinkage would be

required (Yao et al. (2003)).

To overcome this challenge Yao et al. (2005a) propose to predict the score ξ_{ik} by the conditional expectation

$$\mathbb{E}\left[\xi_{ik}|U_{i1},\ldots,U_{iN_i}\right]. \tag{1.15}$$

This conditional expectation is easy to calculate under the assumption that X is a Gaussian process. Denote

$$\begin{aligned}\tilde{U}_i &= \left(U_{i1},\ldots,U_{iN_i}\right)^\top, & \tilde{X}_i &= \left(X_i(x_{i1}),\ldots,X_i(x_{iN_i})\right)^\top,\\ \boldsymbol{\mu}_i &= \left(\mu(x_{1i}),\ldots,\mu(x_{1N_1})\right)^\top, & \boldsymbol{\varphi}_{ik} &= \left(\varphi_k(x_{i1}),\ldots,\varphi_k(x_{iN_i})\right)^\top.\end{aligned}$$

Then the conditional expectation (1.15) can be written as

$$\mathbb{E}\left[\xi_{ik}|\tilde{U}_i\right] = \mathbb{E}\left[\xi_{ik}\right] + \operatorname{cov}(\xi_{ik},\tilde{U}_i)\operatorname{var}(\tilde{U}_i)^{-1}\left(\tilde{U}_i-\boldsymbol{\mu}_i\right),$$

where

$$\boldsymbol{\Sigma}_{\tilde{U}_i} = \operatorname{var}(\tilde{U}_i) = \operatorname{var}(\tilde{X}_i) + \sigma^2 I_{N_i} = \left(\sum_{k=1}^{\infty}\lambda_k\varphi_k(x_{ij})\varphi_k(x_{il})\right)_{j,l=1}^{N_i} + \sigma^2 I_{N_i}. \tag{1.16}$$

where we have used

$$\begin{aligned}&\operatorname{cov}(X_i(x_{ij}),X_i(x_{il})) = \operatorname{cov}\left(\sum_{k=1}^{\infty}\xi_{ik}\varphi_k(x_{ij}),\sum_{m=1}^{\infty}\xi_{im}\varphi_m(x_{il})\right) =\\ &= \sum_{k,m=1}^{\infty}\operatorname{cov}\left(\xi_{ik},\xi_{im}\right)\varphi_k(x_{ij})\varphi_m(x_{il}) = \sum_{k=1}^{\infty}\operatorname{var}\left(\xi_{ik}\right)\varphi_k(x_{ij})\varphi_k(x_{il}) = \sum_{k=1}^{\infty}\lambda_k\varphi_k(x_{ij})\varphi_k(x_{il}).\end{aligned}$$

Now, thanks to the lack of correlation among ξ_{ik}'s and the fact that ξ_{ik} and ϵ_{ij} are independent,

$$\operatorname{cov}(\xi_{ik},U_{ij}) = \operatorname{cov}\left(\xi_{ik},\mu(x_{ij})+\sum_{l=1}^{\infty}\xi_{il}\varphi_l(x_{il})+\epsilon_{il}\right) = \sum_{l=1}^{\infty}\operatorname{cov}(\xi_{ik},\xi_{il})\phi_l(x_{il}) = \lambda_k\varphi_k(x_{ik}),$$

Composing the above stated results together yields the PACE estimates the principal components

$$\tilde{\xi}_{ik}^P = \lambda_k\boldsymbol{\varphi}_{ik}^\top\boldsymbol{\Sigma}_{U_i}^{-1}\left(\tilde{U}_i-\boldsymbol{\mu}_i\right). \tag{1.17}$$

Now, the model dynamics parameters $\mu,\mathscr{R}^X,\sigma^2$ are in general unknown, hence need to be estimated for example by the methods described in Subsection 1.2.3. Denote the spectral decomposition of the estimator

$$\hat{\mathscr{R}}^X = \sum_{k=1}^{\infty}\hat{\lambda}_k\hat{\varphi}_k\otimes\hat{\varphi}_k. \tag{1.18}$$

Because the kernel based methods do not guarantee that the resulting covariance operator is non-negative definite, the spectral decomposition (1.18) may feature some negative tail

eigenvalues $\hat{\lambda}_k$. It is recommended to discard these eigenvalues altogether (Yao et al. (2005a) speaks about the "fitted" covariance operator). Moreover, Yao et al. (2005a) suggest to truncate the expansion (1.18) to a number K of principal components determined, as a model selection problem, by one of the following ways:

- leave-one-curve-out cross-validation,
- AIC and BIC information criteria based on Gaussianity assumption,
- subjective truncation at certain threshold $0 < \tau < 1$ such that the percentage of the total variance explained by the leading principal components not exceeding τ, say $\tau = 0.9, 0.95$.

Having truncated the negative eigenvalues and possibly truncated the spectral representation of the empirical covariance operator (1.18), the PACE estimator is then constructed from its theoretical counterpart (1.17) by substituting the unknown quantities by their empirical estimators:

$$\hat{\xi}_{ik}^P = \hat{\lambda}_k \hat{\boldsymbol{\varphi}}_{ik}^\top \hat{\boldsymbol{\Sigma}}_{U_i}^{-1} \left(\tilde{U}_i - \hat{\boldsymbol{\mu}}_i \right). \tag{1.19}$$

It can be shown Yao et al. (2005a)[Thm 3] that under certain conditions, the PACE estimator (1.19) is consistent for its theoretical counterpart (1.17)

$$\hat{\xi}_{ik}^P \xrightarrow{p} \tilde{\xi}_{ik}^P, \qquad \text{as} \quad n \to \infty,$$

in the sparse regime.

Once the estimators of the principal components have been constructed, the latent functional data X_i can be recovered by

$$\hat{X}_i(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\xi}_{ik}^P \hat{\varphi}_k(x). \tag{1.20}$$

Since the predictor (1.20) involves the quantities $\hat{\Sigma}$ and $\hat{\mu}$, which have been constructed from the entire data set, the recovery (1.20) is said that it is "borrowing strength" from the other observations, using the language by Yao et al. (2005a).

1.3 Functional time series

1.3.1 History and state-of-the art of functional time series research

The research of functional time series, a temporally ordered sequence of dependent functional data, is historically linked to the study of functional linear processes, primarily the functional autoregressive process (Bosq, 1983, 1991, 1996, 1999; Mas, 2007). The functional autoregressive process is particularly important because it allows for an explicit formula for the functional best linear predictor (Bosq, 2000; Kargin and Onatski, 2008) that has been used extensively

in literature (Besse and Cardot, 1996; Besse et al., 2000) together with some variants of the forecasting algorithm (Damon and Guillas, 2002; Antoniadis and Sapatinas, 2003). Further to the functional autoregressive processes, the prediction problem was studied also for the functional moving average processes (Chen et al., 2016; Aue and Klepsch, 2017) and the functional autoregressive moving average processes (Klepsch et al., 2017). It is perhaps fair to say that the family of linear processes is well understood in the context of functional time series, which is underlined by overview publications by Bosq (2000); Bosq and Blanke (2008).

Further development in functional time series domain was due to abandoning the linear structure and investigating sequences of stationary processes. Hörmann and Kokoszka (2010) established the notion of L^p-m-approximability for weakly dependent functional time series and studied the estimation of the long-run covariance operator. Horváth et al. (2013) provided with a central limit theorem for weakly dependent functional data. On top of that, additional univariate or multivariate methods have been adapted for the functional time series setting that serve for prediction (Aue et al., 2015, 2017; Klepsch and Klüppelberg, 2017; Laurini, 2014; Hörmann et al., 2013) and testing (Aue and van Delft, 2020; Górecki et al., 2018; Gao et al., 2019).

Parallel to the time domain approaches, the statistical analysis of functional time series has been fruitful also in the spectral domain. The foundational work is due to Panaretos and Tavakoli (2013b) who established the underlying spectral theory for functional time series and Panaretos and Tavakoli (2013a) established the Cramér-Karhunen-Loeve representation of a functional time series. Moreover, Panaretos and Tavakoli (2013a); Hörmann et al. (2015a); Kidziński et al. (2018) established estimators of the spectral density operators and studied spectral domain dimensionality reduction techniques based on harmonic/dynamic principal components analysis. Furthermore, the spectral domain tools have been successfully used to solve other problems, such as functional lagged regression (Hörmann et al., 2015b; Pham and Panaretos, 2018), stationarity testing (Horváth et al., 2014), periodicity detection (Hörmann et al., 2018), two-sample testing problem (Tavakoli and Panaretos, 2016; Leucht et al., 2018), or white noise testing (Zhang, 2016). Further theoretical contributions in the functional time series spectral theory include van Delft and Eichler (2020) and Kokoszka and Jouzdani (2020)

Another frontier that was tackled in the functional time series literature is the concept of long-range dependence. Li et al. (2019) generalised this phenomenon known from univariate and multivariate time series literature (Granger and Joyeux, 1980; Hosking, 1981) into function spaces and defined the functional autoregressive fractionally integrated moving average process. Shang (2020) provided with quantitative comparison of long-range dependence estimators. In order to allow the spectral analysis of long-range dependent functional time series, the concept of spectral density needs to be generalised into the notion of the *weak* spectral density operator (Tavakoli, 2014). Some spectral domain results for possibly long-range dependent Gaussian processes are established by Ruiz-Medina (2019).

Other contributions in functional time series literature include locally stationary functional

time series (van Delft et al., 2017; van Delft and Eichler, 2018; van Delft and Dette, 2018; Barigozzi et al., 2019) or functional factor models (Kokoszka et al., 2015; Hays et al., 2012; Gao et al., 2019; Tavakoli et al., 2020).

1.3.2 Framework and notation

As we established in Subsection 1.1.3, the classical approach in functional data analysis is to view the data as random objects in the Hilbert space $\mathscr{H}$, the usual choice being $L^2([0,1],\mathbb{R})$. While the setup of the independent identically distributed functional data reviewed in Subsection 1.1.4 and Section 1.2 is relevant in many applications, the functional data may also come ordered in time and exhibit temporal dependence. Ignoring this temporal dependence could lead to wrong conclusions and, therefore, a development of time series models in functional settings is needed.

A functional time series is a time ordered sequence of random elements in $L^2([0,1],\mathbb{R})$, denoted as $\{X_t\}_{t\in\mathbb{Z}}$. Throughout this thesis we work exclusively with time series with finite second moments, i.e. $\mathbb{E}\|X_t\|^2 < \infty$, $t \in \mathbb{Z}$, and which are second order stationary in the time variable t (see the definition bellow). These basic assumptions are imposed implicitly without explicitly repeating them in the stated definitions, results or theorems.

Definition 1.3.1. A functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ is called *second-order stationary* in the time variable t, if the expectations $\mathbb{E}X_t$ and $\mathbb{E}[X_{t+h} \otimes X_t]$ are independent of $t \in \mathbb{Z}$ for all values $h \in \mathbb{Z}$. Moreover, X is called *strictly stationary* if for any tuple $(t_1,\dots,t_k) \in \mathbb{Z}^k$ with $k \in \mathbb{N}$, the random elements $(X_{t_1+h},\dots,X_{t_k+h})$ and $(X_{t_1},\dots,X_{t_k})$ share the same law for any $h \in \mathbb{Z}$.

Definition 1.3.1 allows to define the *mean function* $\mu(x) = \mathbb{E}X_0(x)$, $x \in [0,1]$, and the *lag-h autocovariance kernel* $R_h(x,y) = \mathrm{cov}(X_h(x), X_0(y)) = \mathbb{E}[(X_h(x)-\mu(x))(X_0(y)-\mu(y))]$, $x, y \in [0,1]$ for $h \in \mathbb{Z}$. The corresponding integral operator to R_h is called the *lag-h autocovariance operator*, is denoted as $\mathscr{R}_h^X$ and satisfies $\mathscr{R}_h^X = \mathbb{E}[(X_h - \mu) \otimes (X_0 - \mu)]$.

The framework of stationary functional time series is broad, therefore more structure is required to deliver results useful for a statistical inference. There are essentially two approaches:

(i) *The "parametric approach"*. The first approach to probabilistic modelling of functional time series relies on defining dynamical models for the data, usually as functional linear processes, the generalisation of scalar or multivariate linear processes into function spaces, which are reviewed in Subsection 1.3.8. The statistical inference is then concentrated on estimation of these models' infinite dimensional parameters and performing forecasts. We call this approach as "parametric" despite the parameters involved in such models being infinite dimensional which corresponds strictly speaking to non-parametric statistics. This fact is in contrast with the following approach.

(ii) *The fully non-parametric approach.* The second way is to analyse the second order structure of the functional time series directly without imposing a specific dynamics of

the process. Especially fruitful turns out to be the spectral domain statistical inference reviewed in Subsection 1.3.4. However, the non-parametric statistical analysis from a single realisation of $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ requires to limit the amount of temporal dependence among the values of X_t for varying t. Some concepts of the weak dependence conditions are reviewed in Subsection 1.3.3.

1.3.3 Weak dependence structure and mixing conditions

In order to probe the statistical properties from a single realisation of a functional time series, we need to assume that temporal dependence sufficiently diminishes with the increasing separation in the temporal domain. Essentially, if we split the functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ into two parts, $\{X_t\}_{t=-\infty}^{s_1}$ and $\{X_t\}_{t=s_2}^{\infty}$ for $s_1 < s_2$, we want to define a notion of the vanishing dependence as the distance $s_2 - s_1$ grows. Generally speaking, we call any such condition limiting the dependence of the above listed partitions a *weak dependence condition*. In the following we review the most commonly used instances of weak dependence.

Cumulant mixing conditions

Perhaps the most intuitive notion of the weak dependence relies on the second moments of functional time series. The second order structure of the stationary functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ is governed by the autocovariance kernels $R_h^X(\cdot,\cdot)$, and equivalently the autocovariance operators $\mathscr{R}_h^X$, for $h \in \mathbb{Z}$. In order to limit the temporal dependence we would like to impose the conditions on the speed of decay of $R_h^X(\cdot,\cdot)$ and $\mathscr{R}_h^X$. A priori it is not clear what norm one should consider for these objects and, as it turns out, various norms have been considered in different articles.

Relying on the second-order structure, the weak dependence condition can be manifested in one of the following ways

$$\sum_{h\in\mathbb{Z}} \left\|\mathscr{R}_h^X\right\|_2 < \infty, \tag{1.21}$$

$$\sum_{h\in\mathbb{Z}} \left\|\mathscr{R}_h^X\right\|_1 < \infty, \tag{1.22}$$

$$\sum_{h\in\mathbb{Z}} \left\|\mathscr{R}_h^X\right\|_{\mathscr{L}(\mathscr{H})} < \infty, \tag{1.23}$$

$$\sum_{h\in\mathbb{Z}} \left|\mathrm{Tr}(\mathscr{R}_h^X)\right| < \infty, \tag{1.24}$$

$$\sum_{h\in\mathbb{Z}} \left|\!\left|\!\left|R_h^X\right|\!\right|\!\right|_\infty < \infty, \tag{1.25}$$

$$\sum_{h\in\mathbb{Z}} \left|\!\left|\!\left|R_h^X\right|\!\right|\!\right|_2 < \infty \tag{1.26}$$

where $|||\cdot|||_\infty$ denotes the supremum norm defined as $\left|\!\left|\!\left|R_h^X\right|\!\right|\!\right|_\infty = \sup_{x,y\in[0,1]} |R_h^X(x,y)|$ and $|||\cdot|||_2$ denotes the Frobenius norm $\left|\!\left|\!\left|R_h^X\right|\!\right|\!\right|_2 = \left(\iint |R_h^X(x,y)|^2 \,\mathrm{d}x\,\mathrm{d}y\right)^{1/2}$.

We can go one step further and consider the higher order moment dependence structures, relying on the notion of the cumulants kernels (Panaretos and Tavakoli, 2013b). Assuming $\mathbb{E}\|X_t\|^k < \infty$ for all $t \in \mathbb{Z}$ allows us to define the k-th order cumulant kernel of X_t as $\mathrm{Cum}(X_{t_1}, \ldots, X_{t_k})$ for $t_1, \ldots, t_k \in \mathbb{Z}$ by

$$\mathrm{Cum}(X_{t_1}, \ldots, X_{t_k})(x_1, \ldots, x_k) = \mathrm{cum}(X_{t_1}(x_1), \ldots, X_{t_k}(x_k)), \quad x_1, \ldots, x_k \in [0,1],$$

where $\mathrm{cum}(Y_1, \ldots, Y_k)$ is the cumulant of the scalar random variables $Y_1, \ldots, Y_k$ defined as (Rosenblatt, 2012)

$$\mathrm{cum}(Y_1, \ldots, Y_k) = \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)! \prod_{\ell=1}^{p} \mathbb{E}\left[\prod_{j\in\nu_\ell} Y_j\right]$$

where the sum is over all unordered partitions of $1, \ldots, k$.

The functional time series $\{X_t\}$ is called k-th order stationary, if for all $j = 1, \ldots, k-1$

$$\mathrm{Cum}(X_{t_1+h}, \ldots, X_{t_j+h})(x_1, \ldots, x_j) = \mathrm{Cum}(X_{t_1}, \ldots, X_{t_j})(x_1, \ldots, x_j)$$

for any $t_1, \ldots, t_j \in \mathbb{Z}$, $h \in \mathbb{Z}$ and $x_1, \ldots, x_j \in [0,1]$.

Panaretos and Tavakoli (2013b) consider two sets of conditions based on cumulants.

Definition 1.3.2. Let $l \in \mathbb{N}_0$ and $k \in \mathbb{N}$.

- The functional time series $\{X_t\}$ satisfies the $\mathbf{C}(\mathbf{l}, \mathbf{k})$ condition if it is k-th order stationary and
$$\sum_{t_1,\ldots,t_{k-1}} (1 + |t_j|^l) \left|\!\left|\!\left| \mathrm{Cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0) \right|\!\right|\!\right|_2 < \infty$$
for each $j = 1, \ldots, k-1$. The Frobenius norm of a cumulant kernel of order j is defined as $\left|\!\left|\!\left| \mathrm{Cum}(Z_1, \ldots, Z_j) \right|\!\right|\!\right|_2 = \int_{[0,1]^j} |\mathrm{Cum}(Z_1, \ldots, Z_j)(x_1, \ldots, x_j)|^2 \, \mathrm{d}x_1 \ldots \mathrm{d}x_j$.
- The functional time series $\{X_t\}$ satisfies the $\mathbf{C}'(\mathbf{l}, \mathbf{k})$ condition if it is k-th order stationary and
$$\sum_{t_1,\ldots,t_{k-1}} (1 + |t_j|^l) \left|\!\left|\!\left| \mathrm{Cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0) \right|\!\right|\!\right|_\infty < \infty$$
for each $j = 1, \ldots, k-1$ where $|||\cdot|||_\infty$ denotes the supremum norm.

In particular, the 4-th order cumulant conditions are required for the proofs of the asymptotic behaviour of spectral density operators estimators in Chapter 2.

α-mixing conditions

One of the most famous dependence conditions, at least in the realm of scalar time series, is called α*-mixing*, or *strong mixing* (Rosenblatt, 2012; Doukhan, 2012).

Definition 1.3.3. The strictly stationary functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ is α*-mixing*, or *strong mixing*, if

$$\alpha(m) = \sup_{A\in\mathcal{F}_{-\infty}^{0},\, B\in\mathcal{F}_{m}^{+\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

satisfies $\alpha(m) \to 0$ as $m \to \infty$ where $\mathcal{F}_{-\infty}^{0}$ is the sigma algebra generated by $\{\ldots, X_{-1}, X_0\}$ and $\mathcal{F}_{m}^{+\infty}$ is the sigma algebra generated by $\{X_m, X_{m+1}, \ldots\}$.

Usually, assuming $\alpha(m) \to 0$ as $m \to \infty$ alone is insufficient for the proofs and one needs to further impose a certain rate of the decay.

α-mixing conditions have been successfully applied in the functional time series literature. Antoniadis et al. (2009) used α-mixing for the prediction problem and Aston and Kirch (2012b,a) considered the change-point detection problem for fMRI data. In this thesis we consider α-mixing conditions in Theorems 2.3.4 and 2.3.5 for the proof of optimal convergence rates of the non-parametric estimators established in Chapter 2 as opposed to the suboptimal convergence rates demonstrated in Theorems 2.3.1 and 2.3.2 proof under weaker cumulant mixing conditions.

L^p-m-approximability

Another weak dependence concept is that of L^p-m-approximability, introduced by Hörmann and Kokoszka (2010). The heuristic behind their definition is to approximate the functional time series by m-dependent processes where a process is called *m-dependent* if the sigma algebras (defined in the previous paragraph) $\mathcal{F}_{-\infty}^{m}$ and $\mathcal{F}_{-\infty}^{0}$ are independent while the approximation is valid in the stochastic L^p norm defined as $(\mathbb{E}\|X\|^p)^{1/p}$ for a random element X.

Definition 1.3.4. A functional time series $\{X_t\}$ is called *L^p-m-approximable* with $p \geq 1$ if $\mathbb{E}\|X_t\|^p < \infty$, $t \in \mathbb{Z}$, and it can be represented as $X_t = f(\epsilon_t, \epsilon_{t-1}, \ldots)$ where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent identically distributed random elements taking values and a measurable space S and $f : S^{\infty} \to L^2([0,1],\mathbb{R})$ is a measurable function. Moreover, denote $\{\epsilon'_t\}_{t\in\mathbb{Z}}$ an independent copy of $\{\epsilon_t\}_{t\in\mathbb{Z}}$ and let

$$X_t^{(m)} = f(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-m+1}, \epsilon'_{t-m}, \epsilon'_{t-m-1}, \ldots)$$

then we require

$$\sum_{m=1}^{\infty} \left(\mathbb{E}\left\|X_m - X_m^{(m)}\right\|^p\right)^{1/p} < \infty.$$

Hörmann and Kokoszka (2010) show that L^p-m-approximability is not directly comparable with α-mixing conditions but seems to be easier to verify. Moreover, they give sufficient conditions for a functional linear process to be L^p-m-approximable and show an example when a simple autoregressive process of order 1 may fail to satisfy α-mixing.

Horváth et al. (2013) derived some inference results for L^2-m-approximable functional time series, such as the consistency of the mean function and the estimator of the long-run covariance operator $\sum_{h\in\mathbb{Z}} \mathscr{R}_h^X$. Further articles that successfully used the assumptions of L^p-m-approximability include Kokoszka and Reimherr (2013); Hörmann et al. (2015b); Aue et al. (2014, 2015); Hörmann et al. (2013, 2015a,b, 2018).

1.3.4 Spectral analysis of functional time series

Spectral analysis comprises an established collection of methods in univariate and multivariate time series analysis (Brockwell and Davis, 2009; Priestley, 1981a,b; Brillinger, 1983) and signal processing (Stoica and Moses, 1997). The principal idea of this approach is to express a stochastic process or a signal as a Fourier transform of a process whose components are independent (or at least uncorrelated) for distinct frequencies. As a consequence, the probabilistic analysis of some problems, e.g. the lagged regression (Chapter 3) or simulation of stochastic processes (Chapter 5), simplifies greatly resulting into the popularity of the spectral analysis.

In this subsection we are going to present the spectral approach to functional time series analysis. The generalisation of the key objects in the spectral domain analysis have been first published by Panaretos and Tavakoli (2013a) who defined the spectral density operators and spectral density kernels under various *weak* dependence conditions. Later, the notion of spectral density was extended to the concept of weak spectral density operator by Tavakoli (2014) which does not require weak dependence and allows for the spectral analysis of time series with long-range dependence, such as the analysis of the FARFIMA process in Subsection 1.3.8.

Spectral density operators under weak dependence

Analogous to the notion of the spectral density of a univariate time series (Priestley, 1981a), the spectral density operator and the spectral density kernels are defines as a Fourier transform of the autocovariance operators $\mathscr{R}_h^X$ and the autocovariance kernels R_h^X respectively. The *spectral density operators* $\{\mathscr{F}_\omega^X\}_{\omega\in[-\pi,\pi]}$ and the *spectral density kernels* $\{f_\omega^X(x,y) : (x,y) \in [0,1]^2\}_{\omega\in[-\pi,\pi]}$ are therefore defined by the formula

$$\mathscr{F}_\omega^X = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \mathscr{R}_h^X e^{-\mathrm{i}h\omega}, \qquad \omega \in [-\pi,\pi], \tag{1.27}$$

$$f_\omega^X(x,y) = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} R_h^X(x,y) e^{-\mathrm{i}h\omega}, \qquad \omega \in [-\pi,\pi], \tag{1.28}$$

provided that the formulae (1.27) and (1.28) converge in an appropriate sense.

First, we present the definition by Panaretos and Tavakoli (2013b) of the spectral density operator under the strongest considered norm, the trace norm. Moreover, Panaretos and Tavakoli (2013b) defined also the spectral density kernel in the pointwise sense.

Proposition 1.3.5 (Panaretos and Tavakoli (2013b))**.** *Assume the weak dependence condition* (1.22)*, then* (1.27) *converges in the trace norm and defines the spectral density operator as an self-adjoint, non-negative, trace-class operator. Moreover, the spectral density operators are bounded in the trace norm*

$$\left\|\mathscr{F}_{\omega}^{X}\right\|_{1} \leq \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \left\|\mathscr{R}_{h}^{X}\right\|_{1}, \qquad \omega \in [-\pi,\pi]$$

and we have the inverse formula

$$\mathscr{R}_{h}^{X} = \int_{-\pi}^{\pi} \mathscr{F}_{\omega}^{X} \,\mathrm{d}\omega, \qquad h \in \mathbb{Z} \tag{1.29}$$

in the trace norm.

Assuming further either the condition (1.25) *(set $p = 2$) or* (1.26) *(set $p = \infty$), the spectral density kernel* (1.28) *is well defined and the sum converges in the Frobenius norm $|||\cdot|||_2$ or in the supremum norm $|||\cdot|||_\infty$ respectively. Moreover, the spectral density kernels are uniformly continuous in $|||\cdot|||_p$ for $p \in \{2,\infty\}$, and the inversion formula holds in $|||\cdot|||_p$*

$$R_{h}^{X}(x,y) = \int_{-\pi}^{\pi} f_{\omega}^{X}(x,y) \,\mathrm{d}\omega, \qquad h \in \mathbb{Z}. \tag{1.30}$$

Hörmann et al. (2015a) defined the spectral density operator assuming the sumability of the autocovariance operators in the Hilbert-Schmidt norm resulting in a slightly weaker assumption.

Proposition 1.3.6 (Hörmann et al. (2015a)[Appendix A.2])**.** *Assume the condition* (1.21)*. Then the spectral density operator* (1.27) *is well-defined and the sum converges in the Hilbert-Schmidt norm and the inversion formula* (1.29) *hold in the Hilbert-Schmidt norm. Moreover, the spectral density operator is self-adjoint, non-negative definite and Hilbert-Schmidt.*

The advantage of the slightly weaker assumption (1.21) is that it is implied by the concept of L^2-m-approximability (Subsection 1.3.3).

Corollary 1.3.7 (Hörmann et al. (2015a)[Proposition 6])**.** *Assume that the functional time series $\{X_t\}$ is L^2-m-approximable. Then the condition* (1.21) *is satisfied and therefore the conclusions of Proposition 1.3.6 hold. On top of that, the spectral density operators are in fact trace-class.*

In fact, Tavakoli (2014) has given the weakest weak dependence condition, and demonstrated that the sumability in the operator norm (1.23) is sufficient for the definition of the spectral density operator.

Proposition 1.3.8 (Tavakoli (2014)[Definition 2.3.5])**.** *Assume the condition* (1.23)*. Then the spectral density operator* (1.27) *is well-defined and the sum converges in the operator norm. Moreover, the spectral density operators are continuous in ω with respect to the operator norm, are non-negative and compact, and the inversion formula* (1.29) *holds in the operator norm.*

Assuming further (1.24)*, the spectral density operators are bounded uniformly in the trace norm*

$$\left\|\mathscr{F}_{\omega}^{X}\right\|_{1} \leq \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \left|\mathrm{Tr}\mathscr{R}_{h}^{X}\right| < \infty, \qquad \omega \in [-\pi,\pi].$$

Weak spectral density operator

Finally, we present here the concept of the *weak* spectral density operator as presented by Tavakoli (2014) which states the definition from the inversion formula perspective. Consider the functional time series $\{X_t\}$ with values in the real separable Hilbert space $\mathscr{H}$ and denote the complexification of this space as $\mathscr{H}^{\mathbb{C}}$.

Definition 1.3.9 (Tavakoli (2014)[Definition 2.3.1])**.** Let $\{X_t\}_{t\in\mathbb{Z}}$ be an $\mathscr{H}$-valued second order stationary functional time series with finite second moments. If there exists a mapping $\mathscr{F}^X \in L^1([-\pi,\pi],\mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$ (i.e. the Bochner space of the integrable mappings from $[-\pi,\pi]$ into the space of trace-class operators on $\mathscr{H}^{\mathbb{C}}$ with the trace norm) such that

$$\mathscr{R}_{h}^{X} = \int_{-\pi}^{\pi} \mathscr{F}_{\omega}^{X} e^{\mathrm{i}h\omega}\,\mathrm{d}\omega \tag{1.31}$$

for all $h \in \mathbb{Z}$, then $\mathscr{F}^X$ is called the *weak spectral density operator*.

Since the weak spectral density operator is defined only as an element of the aforementioned Bochner space, it is defined uniquely only almost everywhere, and the pointwise evaluations $\omega \to \mathscr{F}_{\omega}^{X}$ are not defined for fixed $\omega \in [-\pi,\pi]$.

Obviously, if the spectral density operator is defined under the conditions of Proposition 1.3.5, the spectral density operators is also the *weak* spectral density operator.

The advantage of the definition of the weak spectral density operator is that it allows, for example, for functional time series with long-range dependence structure whose spectral density is unbounded at the neighbourhood of zero, see the spectral analysis of the long-range dependent FARFIMA process in Subsection 1.3.8.

1.3.5 The Cramér-Karhunen-Loève decomposition

While the Karhunen-Loève expansion (Theorem 1.1.7) provides with optimal dimensionality reduction for the set-up of independent identically distributed sequence $\{X_t\}$, it is not the best approach when the temporal dependence among $\{X_t\}$ is present. In this section we shall gather some results on dimensionality reduction for functional time series in the spectral domain which are important for the simulation of functional time series with given spectrum (Chapter 5).

In order to take into account the temporal dependence one begins by decomposing the time series into distinct frequencies, a step made rigorous by means of the functional Cramér

representation, due to Panaretos and Tavakoli (2013a)[Thm 2.1] and Tavakoli (2014)[Thm 2.4.3]. We combine the two statements into a single statement, to be used for our purposes, below.

Consider the functional time series $\{X_t\}_{t\in\mathbb{Z}}$ with values in the real separable Hilbert space $\mathscr{H}$ and denote its complexification as $\mathscr{H}^{\mathbb{C}}$.

Proposition 1.3.10 (Functional Cramér representation)**.** *Let the functional time series $X \equiv \{X_t\}_{t\in\mathbb{Z}}$ admit the weak spectral density operator $\mathscr{F}^X \in L^p([-\pi,\pi],\mathscr{L}_1(\mathscr{H}^{\mathbb{C}})$ for some $p\in(1,\infty]$. Then X permits the functional Cramér representation*

$$X_t = \int_{-\pi}^{\pi} e^{\mathrm{i}\,t\omega}\,\mathrm{d}Z_\omega, \qquad \textit{almost surely.} \tag{1.32}$$

where stochastic integral (1.32) *can be understood in Riemann–Stieltjes limit sense*

$$\mathbb{E}\left[\left\| X_t - \sum_{k=1}^{K} e^{\mathrm{i}\,t\omega_k}\left(Z_{\omega_{k+1}} - Z_{\omega_k}\right)\right\|^2\right] \to \infty, \qquad \textit{as} \quad K\to\infty, \tag{1.33}$$

where $-\pi=\omega_1<\cdots<\omega_{k+1}=\pi$ and $\max|\omega_{k+1}-\omega_k|\to 0$ as $K\to\infty$. For each $\omega\in[-\pi,\pi]$, Z_ω is a random element in $\mathscr{H}^{\mathbb{C}}$ defined by

$$Z_\omega = \lim_{T\to\infty}\sum_{|t|<T}\left(1+\frac{|t|}{T}\right) g_\omega(t) X_{-t} \tag{1.34}$$

where the limit holds with respect to $\mathbb{E}\|\cdot\|^2$ and

$$g_\omega(t) = \frac{1}{2\pi}\int_{-\pi}^{\omega} e^{-\mathrm{i}\,t\alpha}\,\mathrm{d}\alpha, \qquad \omega\in[-\pi,\pi].$$

Moreover, the process $\{Z_\omega\}_{\omega\in[-\pi,\pi]}$ satisfies $\mathbb{E}[\|Z_\omega\|_2^2] = \int_{-\pi}^{\omega}\|\mathscr{F}_\alpha^X\|_1\,\mathrm{d}\alpha$ for $\omega\in[-\pi,\pi]$ and $\mathbb{E}[Z_\omega\otimes Z_{\omega'}] = \int_{-\pi}^{\min(\omega,\omega')}\mathscr{F}_\alpha^X\,\mathrm{d}\alpha$ for $\omega,\omega'\in[-\pi,\pi]$, and has orthogonal increments

$$\mathbb{E}\left\langle Z_{\omega_1}-Z_{\omega_2}, Z_{\omega_3}-Z_{\omega_4}\right\rangle = 0, \qquad \omega_1>\omega_2\geq\omega_3>\omega_4.$$

The Cramér representation (1.32) provides a scheme for decomposing X into distinct frequencies. For $0=\omega_1<\cdots<\omega_{k+1}=2\pi$ we have an approximation by (1.33)

$$X_t \approx \sum_{k=1}^{K} e^{\mathrm{i}\,t\omega_k}\left(Z_{\omega_{k+1}} - Z_{\omega_k}\right). \tag{1.35}$$

The approximation (1.35) essentially decomposes the functional time series $\{X_t\}_{t\in\mathbb{Z}}$ into uncorrelated components $Z_{\omega_{k+1}}-Z_{\omega_k}$, $k=1,\ldots,K$. Heuristically, the covariance operator of the increment $Z_{\omega_{k+1}}-Z_{\omega_k}$ is expected to be close to $\mathscr{F}_{\omega_k}^X(\omega_{k+1}-\omega_k)$. By virtue of being a non-negative definite operator, the spectral density operator $\mathscr{F}_\omega^X$, admits a spectral decomposition

of its own at each frequency ω,

$$\mathscr{F}_\omega^X = \sum_{n=1}^{\infty} \lambda_n(\omega)\varphi_n(\omega) \otimes \varphi_n(\omega) \tag{1.36}$$

where $\{\lambda_n(\omega)\}_{n=1}^{\infty}$ are the eigenvalues of $\mathscr{F}_\omega^X$, called the *harmonic eigenvalues*, and their associate eigenfunctions $\{\varphi_n(\omega)\}_{n=1}^{\infty}$, called the *harmonic eigenfunctions*. This suggests a second level of approximation, namely using the Karhunen-Loève expansion to write

$$X_t \approx \sum_{k=1}^{K} e^{\mathrm{i}\, t\omega_k} \sum_{n=1}^{\infty} \xi_n^{(k)} \varphi_n(\omega_k)$$

with $\xi_n^{(k)} = \langle Z_{\omega_{k+1}} - Z_{\omega_k}, \varphi_n(\omega_k)\rangle / \sqrt{\lambda_n(\omega_k)}$ and then truncating at $N \in \mathbb{N}$

$$X_t \approx \sum_{k=1}^{K} e^{\mathrm{i}\, t\omega_k} \sum_{n=1}^{N} \xi_n^{(k)} \varphi_n(\omega_k). \tag{1.37}$$

The approximation (1.37) consists of finite number of uncorrelated random variables $\xi_n^{(k)}$, $k = 1\ldots, K$, $n = 1, \ldots, N$ and will serve as the basis for our simulation method described in Section 5.1.1. To rigorously define this approach, and show its optimality, we must consider the stochastic integral

$$\int_{-\pi}^{\pi} e^{\mathrm{i}\, t\omega} C(\omega)\, \mathrm{d}Z_\omega \tag{1.38}$$

which can be defined by the means similar to the Itô stochastic integral, rigorously proved in Panaretos and Tavakoli (2013a) and Tavakoli (2014). If $\mathscr{F} \in L([-\pi,\pi], \mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$ for $p \in (1,\infty]$, then (1.38) is well defined for $C \in \mathscr{M}$ where $\mathscr{M}$ is the completion of $L^{2q}([-\pi,\pi], \mathscr{L}(\mathscr{H}^{\mathbb{C}}))$, where $p^{-1} + q^{-1} = 1$, with respect to the norm $\|\cdot\|_{\mathscr{M}} = \sqrt{\langle\cdot,\cdot\rangle_{\mathscr{M}}}$ where

$$\langle A, B\rangle_{\mathscr{M}} = \int_{-\pi}^{\pi} \mathrm{Tr}\left(A(\omega)\mathscr{F}_\omega^X B(\omega)^*\right) \mathrm{d}\omega, \qquad A, B \in \mathscr{M}.$$

In this notation, one has (Panaretos and Tavakoli (2013a)[Thm 3.7], Tavakoli (2014)[Thm 2.8.2]):

Proposition 1.3.11 (Optimality of Cramér-Karhunen-Loève representation)**.** *Let the functional time series $X \equiv \{X_t\}_{t\in\mathbb{Z}}$, satisfying the functional Cramér representation* (1.32)*, admit the weak spectral density operator $\mathscr{F}^X \in L^1([-\pi,\pi], \mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$ such that the function $\omega \in [-\pi,\pi] \mapsto \mathscr{F}_\omega^X$ is continuous on $[-\pi,\pi]$ with respect to the operator norm $\|\cdot\|_{\mathscr{L}(\mathscr{H}^{\mathbb{C}})}$ and all the non-zero harmonic eigenvalues of $\mathscr{F}_\omega^X$ are distinct, $\omega \in [-\pi,\pi]$. Let*

$$X_t^* = \int_{-\pi}^{\pi} e^{\mathrm{i}\, t\omega} C(\omega)\, \mathrm{d}Z_\omega$$

with $C \in \mathscr{M}$. Let $N : [-\pi,\pi] \to \mathbb{N}$ be a càdlàg function. Then, the solution to

$$\min \mathbb{E}\left[\left\|X_t - X_t^*\right\|^2\right] \qquad \textit{subject to} \quad \mathrm{rank}(C(\omega)) \le N(\omega)$$

is given by

$$C(\omega) = \sum_{n=1}^{N(\omega)} \varphi_n(\omega) \otimes \varphi_n(\omega).$$

Moreover, the approximation error is given by

$$\mathbb{E}\left[\left\|X_t - X_t^*\right\|^2\right] = \int_{-\pi}^{\pi} \left\{ \sum_{n=N(\omega)+1}^{\infty} \lambda_n(\omega) \right\} \mathrm{d}\omega.$$

Proposition 1.3.11 justifies that the process

$$X_t^* = \int_{-\pi}^{\pi} \sum_{n=1}^{N} e^{\mathrm{i}\,t\omega} \left(\varphi_n(\omega) \otimes \varphi_n(\omega)\right) \mathrm{d}Z_\omega \tag{1.39}$$

yields optimal dimension reduction when we set the rank requirement $N(\omega) \equiv N \in \mathbb{N}$ uniformly across all frequencies. Although the definition of the finite dimensional reduction (1.39) appears quite abstract, it turns out that one can represent X^* in one-to-one manner as an N-dimensional multivariate time series using a particular choice of the filter of the original time series X. Because the simulation method presented in Chapter 5 is based directly on the approximations (1.37) and (1.39), we do not pursue the multivariate time series representation here and refer the reader to Panaretos and Tavakoli (2013a); Tavakoli (2014).

An equivalent dimension reduction technique and multivariate time series representation result, defined entirely in terms of filtering (without reference to a representation theorem such as Proposition 1.3.11) was introduced by Hörmann et al. (2015a).

1.3.6 Cross-covariance operators and cross-spectral density

Until now, the presented theory has dealt only with one functional time series. It is however common in applications that the dependence between multiple time series is of interest. In this section we introduce the basic definition of analysis of such dependence by means of the cross-covariance operators and the cross-spectral density operator.

Let $X = \{X_t\}_{t\in\mathbb{Z}}$ and $Y = \{Y_t\}_{t\in\mathbb{Z}}$ be stationary functional time series with finite second moments in the space $L^2([0,1],\mathbb{R})$ and denote their mean functions as μ_X and μ_Y. Furthermore consider a univariate time series $Z = \{Z_t\}_{t\in\mathbb{Z}}$ with the mean μ_Z.

For $h \in \mathbb{Z}$, define the lag-h *cross-covariance operators* of X and Y or Z by the formulae

$$\begin{aligned}
\mathscr{R}_h^{XY} &= \mathbb{E}\left[(X_h - \mu_X) \otimes (Y_0 - \mu_Y)\right] \qquad \left(\in \mathscr{L}_2(L^2([0,1],\mathbb{R}))\right), \\
\mathscr{R}_h^{XZ} &= \mathbb{E}\left[(X_h - \mu_X)(Z_0 - \mu_Z)\right] \qquad \left(\in L^2([0,1],\mathbb{R})\right), \\
\mathscr{R}_h^{ZX} &= \langle \cdot, \mathscr{R}_h^{XZ} \rangle \qquad \left(\in L^2([0,1],\mathbb{R})^*\right),
\end{aligned}$$

where $L^2([0,1],\mathbb{R})^*$ denotes the space of linear functionals on $L^2([0,1],\mathbb{R})$, also called the *dual*

space of $L^2([0,1],\mathbb{R})$. The lag-h *cross-covariance kernels* are defined by

$$R_h^{XY}(x,y) = \mathbb{E}\left[\left(X_h(x)-\mu_X(x)\right)\left(Y_0(y)-\mu_Y(y)\right)\right], \qquad x,y\in[0,1],$$
$$R_h^{XZ}(x) = \mathbb{E}\left[\left(X_h(x)-\mu_X(x)\right)\left(Z_0-\mu_Z\right)\right], \qquad x\in[0,1].$$

Furthermore, we define the *cross-spectral density operators* by the formulae

$$\mathscr{F}_\omega^{XY} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \mathscr{R}_h^{XY} e^{-\mathrm{i}h\omega}, \qquad \omega\in[-\pi,\pi], \tag{1.40}$$
$$\mathscr{F}_\omega^{XZ} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \mathscr{R}_h^{XZ} e^{-\mathrm{i}h\omega}, \qquad \omega\in[-\pi,\pi], \tag{1.41}$$
$$\mathscr{F}_\omega^{ZX} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \mathscr{R}_h^{ZX} e^{-\mathrm{i}h\omega}, \qquad \omega\in[-\pi,\pi], \tag{1.42}$$

where (1.40) converges in the Hilbert-Schmidt norm if

$$\sum_{h\in\mathbb{Z}} \left\|\mathscr{R}_h^{XY}\right\|_2 < \infty,$$

and (1.41) converges in the vector norm and (1.42) in the operator norm, provided

$$\sum_{h\in\mathbb{Z}} \left\|\mathscr{R}_h^{XZ}\right\| < \infty.$$

The *cross-spectral density kernels* are defined by the formulae

$$f_\omega^{XY}(x,y) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} R_h^{XY}(x,y)e^{-\mathrm{i}h\omega}, \qquad x,y\in[0,1], \quad \omega\in[-\pi,\pi], \tag{1.43}$$
$$f_\omega^{XZ}(x) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} R_h^{XZ}(x)e^{-\mathrm{i}h\omega}, \qquad x\in[0,1], \quad \omega\in[-\pi,\pi], \tag{1.44}$$

where (1.43) and (1.44) converge in the supremum norm, provided

$$\sum_{h\in\mathbb{Z}} \sup_{x,y\in[0,1]} \left|R_h^{XY}(x,y)\right| < \infty,$$
$$\sum_{h\in\mathbb{Z}} \sup_{x\in[0,1]} \left|R_h^{XZ}(x)\right| < \infty,$$

respectively.

1.3.7 Functional filters and frequency response functions

Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a mean-zero stationary functional time series in the separable real Hilbert space $\mathscr{H}_1$, with the weak spectral density operator

$$\mathscr{F}^X \in L^p([-\pi,\pi],\mathscr{L}_1(\mathscr{H}_1^{\mathbb{C}})) \qquad \text{for some} \quad p\in(1,\infty]. \tag{1.45}$$

Therefore, its lag-h autocovariance operators $\mathscr{R}_h^X$ satisfy

$$\mathscr{R}_h^X = \int_{-\pi}^{\pi} \mathscr{F}_\omega^X e^{\mathrm{i}h\omega}\,\mathrm{d}\omega, \qquad h \in \mathbb{Z}.$$

A *functional filter*, or simply a *filter*, is a sequence of coefficients $\{\theta_s\}_{s\in\mathbb{Z}}$ where $\theta_s \in \mathscr{L}(\mathscr{H}_1, \mathscr{H}_2)$ and $\mathscr{H}_2$ is a possible different separable real Hilbert space and $\mathscr{L}(\mathscr{H}_1, \mathscr{H}_2)$ is the space of bounded linear operator from $\mathscr{H}_1$ to $\mathscr{H}_2$. Formally, define the filtered functional time series $Y = \{Y_t\}_{t\in\mathbb{Z}}$ as

$$Y_t = \sum_{s\in\mathbb{Z}} \theta_s X_{t-s}, \qquad t \in \mathbb{Z}, \tag{1.46}$$

and the *frequency response function* of $\{\theta_s\}$ as

$$\Theta(\omega) = \sum_{s\in\mathbb{Z}} \theta_s e^{-\mathrm{i}s\omega}, \qquad \omega \in [-\pi, \pi], \tag{1.47}$$

provided (1.46) and (1.47) converge in an appropriate sense which is justified by the following two proposition that are direct consequences of Tavakoli (2014)[Thm 2.5.5, Remark 2.5.6].

Proposition 1.3.12. *Assume* (1.45) *and that the filter* $\{\theta_s\}$ *satisfies*

$$\sum_{s\in\mathbb{Z}} \|\theta_s\|_{\mathscr{L}(\mathscr{H}_1,\mathscr{H}_2)} < \infty.$$

Then the sum on the right-hand side of (1.46) *converges with respect to* $\mathbb{E}\|\cdot\|^2$ *and* $Y = \{Y_t\}_{t\in\mathbb{Z}}$ *is a second-order stationary mean-zero functional time series with values in* $\mathscr{H}_2$*. Moreover, the sum on the right-hand side of* (1.47) *converges in* $\mathscr{M}$ *(defined in Subsection 1.3.5) and the weak spectral density operator* $\mathscr{F}^Y \in L^1([-\pi,\pi], \mathscr{L}_1(\mathscr{H}_2^{\mathbb{C}}))$ *of the functional time series* $Y = \{Y_t\}_{t\in\mathbb{Z}}$ *is given by*

$$\mathscr{F}_\omega^Y = \Theta(\omega)\mathscr{F}_\omega^X\Theta(\omega)^*, \qquad \omega \in [-\pi, \pi] \tag{1.48}$$

and the lag-h autocovariance operators of Y *are given by*

$$\mathscr{R}_h^Y = \mathbb{E}[Y_h \otimes Y_0] = \int_{-\pi}^{\pi} \mathscr{F}_\omega^Y e^{\mathrm{i}t\omega}\,\mathrm{d}\omega, \qquad h \in \mathbb{Z}. \tag{1.49}$$

Furthermore, if (1.23) *and* (1.24) *hold for the time series* X*, then* $\sum_{h\in\mathbb{Z}} \|\mathscr{R}_h^Y\|_{\mathscr{L}(\mathscr{H}_2)} < \infty$.

Proposition 1.3.13. *Assume that the functional time series* X *is* m*-correlated for some* $m \in \mathbb{N}$*, i.e.* $\mathscr{R}_h^X = 0$ *for* $|h| > m$*, and the filter* $\{\theta_s\}$ *satisfies*

$$\sum_{s\in\mathbb{Z}} \|\theta_s\|^2_{\mathscr{L}(\mathscr{H}_1,\mathscr{H}_2)} < \infty.$$

Then the sum on the right-hand side of (1.46) *converges with respect to* $\mathbb{E}\|\cdot\|^2$ *and* $Y = \{Y_t\}_{t\in\mathbb{Z}}$ *is a second-order stationary mean-zero functional time series with values in* $\mathscr{H}_2$*, the right-hand side of* (1.47) *converges in* $\mathscr{M}$ *(defined in Subsection 1.3.5), the weak spectral density operator*

$\mathscr{F}^Y \in L^1([-\pi,\pi],\mathscr{L}_1(\mathscr{H}_2^{\mathbb{C}}))$ *is given by* (1.48) *and the inverse formula* (1.49) *holds.*

Hörmann et al. (2015b) provided with one more sufficient condition for the frequency response function to be well-defined and to guarantee the existence of the cross-spectral density operator:

Proposition 1.3.14 (Hörmann et al. (2015b))**.** *Assume the weak-dependence condition* (1.21) *and that the filter* $\{\theta_s\}$ *satisfies*

$$\sum_{s\in\mathbb{Z}} \|\theta_s\|_{\mathscr{L}(\mathscr{H}_1,\mathscr{H}_2)} < \infty.$$

Then in addition to the conclusions of Proposition 1.3.12, the cross-covariance operators satisfy

$$\mathscr{R}_h^{YX} = \sum_{k\in\mathbb{Z}} \theta_k \mathscr{R}_{h-k}^X, \qquad k \in \mathbb{Z}, \tag{1.50}$$

$$\sum_{h\in\mathbb{Z}} \left\|\mathscr{R}_h^{YX}\right\|_2 < \infty, \tag{1.51}$$

and the cross-spectral density operator between Y *and* X *is given by*

$$\mathscr{F}_\omega^{YX} = \Theta(\omega)\mathscr{F}_\omega^X, \qquad \omega \in [-\pi,\pi]. \tag{1.52}$$

Similarly to Proposition 1.3.14 proved by Hörmann et al. (2015b) we can show the corresponding result for the cross-spectral density kernel.

Proposition 1.3.15. *If* $\mathscr{H}_1 = L^2([0,1],\mathbb{R})$ *and* (1.25) *holds, then we have the following results:*

- $\mathscr{H}_2 = L^2([0,1],\mathbb{R})$ *and the filter coefficients* θ_s *are Hilbert-Schmidt operators represented by kernels* $(\theta_s f)(x) = \int_0^1 T_s(x,y) f(y)\,\mathrm{d}y$ *for* $f \in L^2([0,1],\mathbb{R})$ *satisfying*

$$\sum_{s\in\mathbb{Z}} \sup_{x,y\in[0,1]} \left|T_s(x,y)\right| < \infty, \tag{1.53}$$

  then the cross-covariance kernel between Y *and* X *is given by*

$$R_h^{YX}(s,t) = \sum_{k\in\mathbb{Z}} \int_0^1 T_k(s,x) R_{h-k}^X(x,t)\,\mathrm{d}x, \qquad s,t\in[0,1], \quad h\in\mathbb{Z}, \tag{1.54}$$

  satisfies

$$\sum_{h\in\mathbb{Z}} \sup_{s,t\in[0,1]} \left|R_h^{YX}(s,t)\right| < \infty. \tag{1.55}$$

  Therefore the cross-spectral density kernel is given by (1.43).

- $\mathscr{H}_2 = \mathbb{R}$ *and the filter coefficients* θ_s *are functionals on* $\mathscr{H}_1$ *associated with its Riez-representers* $T_s(\cdot)$ *via the equation* $\theta_s g = \langle T_s, g\rangle = \int_0^1 T_s(x)g(x)\,\mathrm{d}x$ *for* $g \in L^2([0,1],\mathbb{R})$*. Assume further that*

$$\sum_{s\in\mathbb{Z}} \sup_{x\in[0,1]} |T_s(x)| < \infty. \tag{1.56}$$

Then the cross-covariance between Y and X is given by

$$R_h^{YX}(x) = \sum_{k\in\mathbb{Z}} \int_0^1 T_k(y) R_{h-k}^X(y,x)\,\mathrm{d}y, \qquad x\in[0,1], \quad h\in\mathbb{Z}, \tag{1.57}$$

and they satisfy

$$\sum_{h\in\mathbb{Z}} \sup_{x\in[0,1]} \left|R_h^{YX}(x)\right| < \infty. \tag{1.58}$$

Consequently, the cross-spectral density is given by (1.44).

Proof. Assume now that $\mathscr{H}_1 = \mathscr{H}_2 = L^2([0,1],\mathbb{R})$. In order to verify (1.54) we start with the right hand side of (1.50) and expressing the operator $\theta_s \mathscr{R}_{h-k}^X$ as an integral operator. For $g\in L^2([0,1],\mathbb{R})$:

$$\left(\theta_s \mathscr{R}_{h-k}^X g\right)(x) = \int_0^1 T_k(x,y)\left(\int_0^1 R_{h-k}^X(y,z) g(z)\,\mathrm{d}z\right)\mathrm{d}y = \int_0^1 \left(\int_0^1 T_k(x,y) R_{h-k}^X(y,z)\,\mathrm{d}y\right) g(z)\,\mathrm{d}z. \tag{1.59}$$

Therefore $\int_0^1 T_k(x,y) R_{h-k}^X(y,z)\,\mathrm{d}y$ is the kernel of the integral operator of $\theta_s \mathscr{R}_{h-k}^X$ and (1.54) holds. Furthermore, (1.25) and (1.53) imply (1.55).

The case with $\mathscr{H}_1 = L^2([0,1],\mathbb{R})$, $\mathscr{H}_2 = \mathbb{R}$ and the derivation of (1.57) are analogous to (1.59) and the claim (1.58) follows from the assumptions (1.25) and (1.56). □

1.3.8 Linear functional time series

The early functional time series research is linked to the study of linear processes in infinite dimensional spaces. In particular, Bosq (1983, 1991, 1996, 1999); Mas (2007) considered linear processes in Hilbert and Banach spaces, provided with the theoretical toolbox for the asymptotic analysis of these probabilistic processes, and studied the estimation and prediction problems. An extensive review of functional linear processes can be found in Bosq (2000).

Introduction: functional autoregressive and moving average processes

The most basic stochastic linear process is the *autoregressive process of order 1* (Bosq, 2000; Mas, 2007), denoted as FAR(1), defined as a $\mathscr{H}$-valued stochastic process satisfying the iterative equation

$$\left(X_t - \mu\right) = \mathscr{A}\left(X_{t-1} - \mu\right) + \epsilon_t, \qquad t\in\mathbb{Z}, \tag{1.60}$$

where μ is the mean function, the autoregressive operator $\mathscr{A}$ is a bounded linear operator, $\mathscr{A}\in\mathscr{L}(\mathscr{H})$, and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of mean-zero independent identically distributed random elements in $\mathscr{H}$ with the covariance operator Σ.

Proposition 1.3.16 (Bosq (2000))**.** *Suppose one of the two following equivalent conditions:*

(i) There exists $j_0\in\mathbb{N}$ such that $\|\mathscr{A}^{j_0}\|_{\mathscr{L}(\mathscr{H})} < 1$.

(ii) There exist $a > 0$ *and* $0 < b < 1$ *such that* $\|\mathscr{A}^j\|_{\mathscr{L}(\mathscr{H})} \leq ab^j$ *for all* $j \in \mathbb{N}_0$.

Then, there exists a unique stationary solution to the iterative equation (1.60) *which is given by*

$$X_t = \mu + \sum_{j=0}^{\infty} \mathscr{A}^j \epsilon_{t-j}$$

and the autocovariance operators of $\{X_t\}$ *are given by*

$$\begin{aligned}
\mathscr{R}_0^X &= \sum_{j=0}^{\infty} \mathscr{A}^j \Sigma \mathscr{A}^{*j}, \\
\mathscr{R}_h^X &= \mathscr{A}^h \mathscr{R}_0^X, \qquad h \in \mathbb{N}_0, \\
\mathscr{R}_{-h}^X &= \mathscr{R}_0^X \mathscr{A}^{*h}, \qquad h \in \mathbb{N}_0.
\end{aligned}$$

The condition (ii) of Proposition 1.3.16 also guarantees that $\sum_{j=0}^{\infty} \|\mathscr{A}^j\|_{\mathscr{L}(\mathscr{H})} < \infty$.

Another functional linear process that has been analysed is the *functional moving average process* (Chen et al., 2016) of order 1, denoted as FMA(1), as defined as

$$X_t = \mu + \epsilon_t + \mathscr{B}\epsilon_{t-1}, \qquad t \in \mathbb{Z},$$

where $\mathscr{B} \in \mathscr{L}(\mathscr{H})$ and $\mu, \{\epsilon_t\}$ are like before. It is trivial to verify that

$$\begin{aligned}
\mathscr{R}_0^X &= \Sigma + \mathscr{B}\Sigma\mathscr{B}^*, \\
\mathscr{R}_1^X &= \mathscr{B}\Sigma, \\
\mathscr{R}_{-1}^X &= \Sigma\mathscr{B}^*, \\
\mathscr{R}_h^X &= 0, \qquad |h| \geq 2.
\end{aligned}$$

Functional moving average autoregressive process

The combination of the functional autoregressive processes and the moving average processes constitutes the class of functional autoregressive moving average processes, denoted as FARMA(p,q), has been presented by Klepsch et al. (2017). In the following text we recall the time domain analysis of FARMA processes and then develop our new results on the frequency domain analysis thereof.

The FARMA(p,q) process, $p, q \in \mathbb{N}_0$, is a sequence $X = \{X_t\}_{t\in\mathbb{Z}}$ of random $\mathscr{H}$-elements, satisfying the equation

$$\left(X_t - \mu\right) = \sum_{j=1}^{p} \mathscr{A}_j \left(X_{t-j} - \mu\right) + \epsilon_t + \sum_{j=1}^{q} \mathscr{B}_j \epsilon_{t-j}, \qquad t \in \mathbb{Z}, \tag{1.61}$$

where $\mathscr{A}_1, \ldots, \mathscr{A}_p$ and $\mathscr{B}_1, \ldots, \mathscr{B}_q$ are bounded linear operators and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of

zero-mean i.i.d. random elements in $\mathscr{H}$ with the covariance operator $\mathscr{S}$.

The time-domain analysis of the FARMA(p,q) process was considered by Klepsch et al. (2017), who in particular established:

Theorem 1.3.17 (Klepsch et al. (2017))**.** *Assume that there exists $j_0 \in \mathbb{N}$ such that the operator*

$$\tilde{\mathscr{A}} = \begin{bmatrix} \mathscr{A}_1 & \cdots & \mathscr{A}_{p-1} & \mathscr{A}_p \\ \mathrm{I} & & & 0 \\ & \ddots & & \vdots \\ & & \mathrm{I} & 0 \end{bmatrix}$$

satisfies

$$\|\tilde{\mathscr{A}}^{j_0}\|_{\mathscr{L}(\mathscr{H}^p)} < 1 \tag{1.62}$$

where I *is the identity operator on $\mathscr{H}$ and $\|\cdot\|_{\mathscr{L}(\mathscr{H}^p)}$ denotes the operator norm on $\mathscr{L}(\mathscr{H}^p)$, the space of bounded linear operators acting on the product space $\mathscr{H}^p = \mathscr{H} \times \cdots \times \mathscr{H}$. Then the FARMA($p,q$) process defined by* (1.61) *is uniquely defined, stationary, and causal.*

We now show that, under the same assumptions as those by Klepsch et al. (2017), we may characterise the FARMA(p,q) process in the spectral domain:

Theorem 1.3.18. *Under the assumptions of Theorem 1.3.17, the process satisfies the weak dependence condition* (1.23) *with $\mathscr{R}_h^X$, and its spectral density operator at frequency $\omega \in [-\pi,\pi]$ is given by*

$$\mathscr{F}_\omega^X = \frac{1}{2\pi}\mathcal{A}(e^{-\mathrm{i}\omega})^{-1}\mathcal{B}(e^{-\mathrm{i}\omega})\mathscr{S}\mathcal{B}(e^{-\mathrm{i}\omega})^*\left[\mathcal{A}(e^{-\mathrm{i}\omega})^*\right]^{-1} \tag{1.63}$$

where

$$\mathcal{A}(z) = \mathrm{I} - \mathscr{A}_1 z - \cdots - \mathscr{A}_p z^p, \tag{1.64}$$

$$\mathcal{B}(z) = \mathrm{I} + \mathscr{B}_1 z + \cdots + \mathscr{B}_p z^q. \tag{1.65}$$

are $\mathscr{H}$-valued polynomials in the variable $z \in \mathbb{C}$.

Proof. Denoting Δ to be the backshift operator, the equation (1.61) can be rewritten as

$$\mathcal{A}(\Delta)X_t = \mathcal{B}(\Delta)\epsilon_t. \tag{1.66}$$

We start with the analysis of the moving average part $\eta_t = \mathcal{B}(\Delta)\epsilon_t$. The spectral density operator of the stochastic innovation process $\{\epsilon_t\}$ is trivially given by $\mathscr{F}_\omega^\epsilon = (2\pi)^{-1}\mathscr{S}$. The filter $\mathcal{B}(\Delta)$, whose filter coefficients are given by $\mathcal{B}(\Delta)_s = \mathscr{B}_s$ for $s = 0,\ldots,q$ and $\mathcal{B}(\Delta)_s = 0$ otherwise, defines the frequency response function $B(\omega) = \mathcal{B}(e^{-\mathrm{i}\omega})$. Thus, the moving average process $\eta = \{\eta_t\}$ admits the spectral density operator

$$\mathscr{F}_\omega^\eta = \frac{1}{2\pi}\mathcal{B}(e^{-\mathrm{i}\omega})\mathscr{S}\mathcal{B}(e^{-\mathrm{i}\omega})^*$$

by Proposition 1.3.12. Obviously, the moving average process $\eta = \{\eta_t\}$ is q-correlated, i.e. $\mathscr{R}_h^\eta = 0$ for $|h| > q$, and therefore satisfies the conditions (1.23) and (1.22), and it is easy to verify that $\mathscr{F}^\eta \in L^\infty([-\pi,\pi], \mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$.

We now wish to invert (1.66) and write the process X as

$$X_t = \mathcal{A}^{-1}(\Delta)\left[\mathcal{B}(\Delta)\epsilon_t\right] = \mathcal{A}^{-1}(\Delta)\eta_t. \tag{1.67}$$

As part of their existence proof, Klepsch et al. (2017)[Thm 3.8] defined a state space process representation of (1.61) as a process in the product space $\mathscr{H}^p$

$$\underbrace{\begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix}}_{\Xi_t} = \underbrace{\begin{bmatrix} \mathscr{A}_1 & \cdots & \mathscr{A}_{p-1} & \mathscr{A}_p \\ \mathrm{I} & & & 0 \\ & \ddots & & \vdots \\ & & \mathrm{I} & 0 \end{bmatrix}}_{\tilde{\mathscr{A}}} \underbrace{\begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \end{bmatrix}}_{\Xi_{t-1}} + \underbrace{\begin{bmatrix} \eta_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\tilde{\eta}_t}, \qquad t \in \mathbb{Z}.$$

They showed that the process Ξ can be written as

$$\Xi_t = \sum_{j=0}^{\infty} \tilde{\mathscr{A}}^j \tilde{\eta}_{t-j}, \qquad t \in \mathbb{Z}, \tag{1.68}$$

where

$$\sum_{j=0}^{\infty} \|\tilde{\mathscr{A}}^j\|_{\mathscr{L}(\mathscr{H}^p)} < \infty \tag{1.69}$$

by the assumption (1.62). Set P_1 to be the projection operator onto the first component:

$$P_1 : \quad \mathscr{H}^p \to \quad \mathscr{H}, \qquad (f_1, \ldots, f_n) \mapsto \quad f_1.$$

Applying P_1 to (1.68) yields $X_t = \sum_{j=0}^{\infty} P_1 \tilde{\mathscr{A}}^j P_1^* \eta_{t-j}$ which essentially means that the filter $\mathcal{A}(\Delta)^{-1}$ is given by $(\mathcal{A}(\Delta)^{-1})_s = P_1 \tilde{\mathscr{A}}^s P_1^*$ for $s \geq 0$ and zero otherwise. Moreover, (1.69) implies

$$\sum_{s \in \mathbb{Z}} \left\| [\mathcal{A}(\Delta)^{-1}]_s \right\|_{\mathscr{L}(\mathscr{H})} < \infty. \tag{1.70}$$

Finally, the application of Proposition 1.3.12 onto the filter $\mathcal{A}(\Delta)^{-1}$ and functional time series η gives us the spectral density of X given by the formula (1.63). Moreover, because η is q-correlated, it trivially satisfies the conditions (1.23) and (1.22) with $\mathscr{R}_h^\eta$, therefore the FARMA(p,q) process X also satisfies the weak dependence conditions (1.23) with $\mathscr{R}_h^X$. □

Functional autoregressive fractionally integrated moving average process

Long range dependence (a.k.a. long memory) is a well known phenomenon in time series analysis, consisting in a time series exhibiting slow decay of its temporal dependence (Hurst,

1951; Mandelbrot and van Ness, 1968; Beran, 1994; Palma, 2007). The need to model and analyse such series has led to the definition of autoregressive fractionally integrated moving average (ARFIMA) processes (Granger and Joyeux, 1980; Hosking, 1981).

Such long-range dependencies have also been detected functional time series, for example in series of daily volatility (Casas and Gao, 2008), and inspired the theoretical framework of long-range dependent functional time series model (Li et al., 2019) and associated estimation methods (Shang, 2020).

Li et al. (2019) defined the functional ARFIMA process, denoted as FARFIMA(p,d,q), with $p,q \in \mathbb{N}_0$ and $d \in (-1/2,1/2)$ models as a sequence $\tilde{X} = \{\tilde{X}_t\}_{t\in\mathbb{Z}}$ of random $\mathscr{H}$-elements via the equation

$$(\mathrm{I}-\Delta)^d \tilde{X}_t = X_t \tag{1.71}$$

where Δ is the back-shift operator and $X = \{X_t\}_{t\in\mathbb{Z}}$ is the FARMA(p,q) process defined via equation (1.61). The operation $(\mathrm{I}-\Delta)^d$ denotes the fractional integration with the power $d \in (-1/2,1/2)$ of the filtration $\mathrm{I}-\Delta$. Refer to Li et al. (2019) or Hosking (1981) for the definition of the fractional powers.

When $d = 0$, the FARFIMA(p,d,q) reduces to the FARMA(p,q) model.

Li et al. (2019) established the existence and uniqueness results of the FARFIMA(p,d,q) process and its time-domain properties:

Theorem 1.3.19 (Li et al. (2019))**.** *The FARFIMA(p,d,q) process $\tilde{X} = \{\tilde{X}_t\}_{t\in\mathbb{Z}}$ with $p,q \in \mathbb{N}_0$ and $d \in (-1/2,1/2)$ defined by the equation* (1.71) *exists and constitutes a uniquely defined stationary causal functional time series provided the autoregressive part satisfies the condition* (1.62)*. Furthermore, if $d \in (0,1/2)$ the FARFIMA(p,d,q) process exhibits the long-memory dependence, according to the definition of Li et al. (2019).*

Under the same assumptions as Li et al. (2019) we now determine the analytical expression of the spectral density operators of the FARFIMA(p,d,q) process:

Theorem 1.3.20. *Under the assumptions of Theorem 1.3.19, the FARFIMA(p,d,q) process admits the weak spectral density $\mathscr{F}^{\tilde{X}} \in L^1([-\pi,\pi],\mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$ satisfying*

$$\mathscr{F}^{\tilde{X}}_{\omega} = \frac{1}{2\pi}\left[2\sin\left(\frac{\omega}{2}\right)\right]^{-2d} \mathcal{A}(e^{-\mathrm{i}\omega})^{-1}\mathcal{B}(e^{-\mathrm{i}\omega})\mathscr{S}\mathcal{B}(e^{-\mathrm{i}\omega})^*\left[\mathcal{A}(e^{-\mathrm{i}\omega})^*\right]^{-1}, \quad \omega \in (-\pi,\pi), \tag{1.72}$$

where $\mathcal{A}$ and $\mathcal{B}$ are given at (1.64) *and* (1.65)*. The lag-h autocovariance operators of $\tilde{X}$ satisfy*

$$\mathscr{R}^{\tilde{X}}_h = \int_{-\pi}^{\pi} \mathscr{F}^{\tilde{X}}_{\omega} e^{\mathrm{i}h\omega}\,\mathrm{d}\omega, \qquad h \in \mathbb{Z}.$$

Proof. Building upon the results of Theorem 1.3.17 we write the FARMA(p,q) process as

$$X_t = \mathcal{A}(\Delta)^{-1}\mathcal{B}(\Delta)\epsilon_t = \mathcal{A}(\Delta)^{-1}\eta_t$$

where $\eta_t = \mathcal{B}(\Delta)\epsilon_t$ is the functional moving average process. Formally inverting the filter (1.71) yields

$$\tilde{X}_t = (\mathrm{I}-\Delta)^{-d} X_t = (\mathrm{I}-\Delta)^{-d}\mathcal{A}(\Delta)^{-1}\eta_t$$

Following the proof of Hosking (1981)[Thm 1], define the function $c(z) = (1-z)^{-d}$, $z \in \mathbb{C}$. Then the power series expansion of c converges for $|z| \leq 1$ as long as $d < 1/2$ and we can write $c(z) = \sum_{k=0}^{\infty} c_k z^k$, $|z| \leq 1$. Moreover, using the binomial expansion for $(1-z)^{-d}$ it can be shown (Hosking, 1981) that the coefficients satisfy

$$c_k \sim \frac{k^{d-1}}{(d-1)!}, \qquad \text{as} \quad k \to \infty. \tag{1.73}$$

Define with the filter $\mathscr{C} = \{\mathscr{C}_k\}_{k\in\mathbb{Z}}$ with filter coefficients $\mathscr{C}_k = c_k \mathrm{I}$ for $k \in \mathbb{N}_0$ where I is the identity operator on $\mathscr{H}^{\mathbb{C}}$, and zero otherwise. Obviously $\mathscr{C} = (\mathrm{I}-\Delta)^{-d}$ in the sense of equality of filters. By the asymptotic relation (1.73), the filter satisfies

$$\sum_{k\in\mathbb{Z}} \|\mathscr{C}_k\|^2_{\mathscr{L}(\mathscr{H}^{\mathbb{C}})} < \infty. \tag{1.74}$$

The convolution of the filters $\mathscr{C}$ and $\mathcal{A}(\Delta)^{-1}$, denoted as $\mathscr{D} = \mathscr{C} * \mathcal{A}(\Delta)^{-1}$, is given by

$$\mathscr{D}_s = \begin{cases} \sum_{k=0}^{s} \mathscr{C}_k \left[\mathcal{A}(\Delta)^{-1}\right]_{s-k}, & s \geq 0, \\ 0, & s < 0. \end{cases}$$

Young's convolution inequality (Hewitt and Ross, 2012)[Thm 20.18], (1.70) and (1.74) imply

$$\sum_{k\in\mathbb{Z}} \|\mathscr{D}_k\|^2_{\mathscr{L}(\mathscr{H}^{\mathbb{C}})} < \infty.$$

Because the moving average process η_t is q-correlated, Proposition 1.3.13 implies existence and stationary of the FARFIMA(p,d,q) process defined by the filter $\tilde{X}_t = \mathscr{D}\eta_t = \mathscr{C}\left[\mathcal{A}(\Delta)^{-1}\eta_t\right]$. Moreover, the process $\tilde{X}$ admits the weak spectral density $\mathscr{F}^{\tilde{X}} \in L^1([-\pi,\pi], \mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$ given by

$$\begin{aligned} \mathscr{F}^{\tilde{X}}_{\omega} &= \frac{1}{2\pi}\mathscr{D}(\omega)\mathscr{F}^{\eta}_{\omega}\mathscr{D}(\omega)^* \\ &= \frac{1}{2\pi} c(e^{-\mathrm{i}\omega})\mathcal{A}(e^{-\mathrm{i}\omega})\mathcal{B}(e^{-\mathrm{i}\omega})\mathcal{S}\mathcal{B}(e^{-\mathrm{i}\omega})^* \left[\mathcal{A}(e^{-\mathrm{i}\omega})^*\right]^{-1} \overline{c(e^{-\mathrm{i}\omega})} \\ &= \frac{1}{2\pi} \left[2\sin(\omega/2)\right]^{-2d} \mathcal{A}(e^{-\mathrm{i}\omega})\mathcal{B}(e^{-\mathrm{i}\omega})\mathcal{S}\mathcal{B}(e^{-\mathrm{i}\omega})^* \left[\mathcal{A}(e^{-\mathrm{i}\omega})^*\right]^{-1}, \end{aligned}$$

for $\omega \in (-\pi,\pi)$, where we have used that $c(e^{-\mathrm{i}\omega})\mathrm{I} = (1-e^{-\mathrm{i}\omega})^{-d}\mathrm{I} = \sum_{k=0}^{\infty} \mathscr{C}_k e^{-\mathrm{i}k\omega}$ is the frequency response function of the filter $\mathscr{C}$ and $c(e^{-\mathrm{i}\omega})\overline{c(e^{-\mathrm{i}\omega})} = |1-e^{-\mathrm{i}\omega}|^{-2d} = [2\sin(\omega/2)]^{-2d}$. □

Note that for $d > 0$, the term $[2\sin(\omega/2)]^{-2d}$ in formula (1.72) is unbounded in the neigh-

bourhood of 0. The spectral density being unbounded in the neighbourhood of zero is quintessential also for the univariate ARFIMA processes (Hosking, 1981).

1.3.9 Inference for fully observed functional time series

In this subsection we list some previous results concerning estimation for fully observed functional time series. We start with the estimation of the mean function and proceed with the second order structure of the data embodied by the lagged autocovariance operators and the spectral density operators.

Estimation of the mean function

The first order structure of the functional time series data is often overlooked in functional time series methodological research literature as it is seen as uncomplicated and hence the mean function is often assumed to be zero. Indeed, the sufficient condition for the empirical mean function $\hat{\mu} = (1/T)\sum_{t=1}^{T} X_t$ of the functional time series $\{X_t\}$ is the weak-dependence condition (1.22) alone. We prove this simple result in the following proposition.

Proposition 1.3.21. *Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a functional time series in $\mathscr{H}$ with mean function μ and lagged autocovariance operators $\mathscr{R}_h^X, h \in \mathbb{Z}$ satisfying* (1.22)*. Then, the empirical mean function satisfies*

$$\begin{aligned} \mathbb{E}\hat{\mu} &= \mu, \\ \operatorname{var}(\hat{\mu}) &= \frac{1}{T}\sum_{|h|\le T}\left(1-\frac{|h|}{T}\right)\mathscr{R}_h^X, \\ \mathbb{E}\left[\left\|\hat{\mu}-\mu\right\|_{\mathscr{H}}^2\right] &\le \frac{1}{T}\sum_{h\in\mathbb{Z}}\left\|\mathscr{R}_h^X\right\|_1, \end{aligned}$$

and hence $\hat{\mu}$ is an unbiased and $\sqrt{T}$-consistent estimator of μ in the $\mathscr{H}$-norm as $T\to\infty$.

Proof. Clearly $\mathbb{E}\hat{\mu} = \mu$, hence the estimator is unbiased. Calculating its variance yields

$$\operatorname{var}(\hat{\mu}) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathscr{R}_{t-t'}^X = \frac{1}{T}\sum_{|h|\le T}\left(1-\frac{|h|}{T}\right)\mathscr{R}_h^X.$$

And finally, taking the trace norm over the above equation proves the last statement

$$\mathbb{E}\left[\left\|\hat{\mu}-\mu\right\|_{\mathscr{H}}^2\right] = \operatorname{var}\left(\|\hat{\mu}\|_{\mathscr{H}}\right) = \left\|\frac{1}{T}\sum_{|h|\le T}\frac{1-|h|}{T}\mathscr{R}_h^X\right\|_1 \le \frac{1}{T}\sum_{|h|\le T}\frac{1-|h|}{T}\left\|\mathscr{R}_h^X\right\|_1 \le \frac{1}{T}\sum_{h\in\mathbb{Z}}\left\|\mathscr{R}_h^X\right\|_1.$$

□

A deeper analysis of the empirical mean function consistency is provided by Bosq (2000) [§2.4]

who studied in fact Banach space valued functional time series and gave weaker conditions for the empirical mean consistency. Other results on empirical mean function estimation can be found in Horváth and Kokoszka (2012); Horváth et al. (2013).

Estimation of lagged autocovariance and cross-covariance operators

The second order structure of functional time series data is quintessential for many functional time series applications. Consider a pair of stationary functional time series $\{X_t\}$ and $\{Y_t\}$ with values in the Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ respectively. We are interested in estimating the lag-h autocovariance operator $\mathscr{R}_h^X = \mathbb{E}\left[(X_h - \mu_X) \otimes (X_0 - \mu_X)\right]$ and the lag-h cross-covariance operator $\mathscr{R}_h^{YX} = \mathbb{E}\left[(Y_h - \mu_Y) \otimes (X_0 - \mu_X)\right]$ for fixed $h > 0$ where $\mu_X = \mathbb{E}X_0$ and $\mu_Y = \mathbb{E}Y_0$ are the mean functions of the respective functional time series.

Having the sample $(Y_1, X_1), \ldots, (Y_T, X_T)$, we estimate the lag-$h$ autocovariance operator and the cross-covariance operator by their empirical counterparts

$$\hat{\mathscr{R}}_h^X = \frac{1}{T}\sum_{t=1}^{T-h}\left(X_{t+h} - \hat{\mu}_X\right) \otimes \left(X_t - \hat{\mu}_X\right), \tag{1.75}$$

$$\hat{\mathscr{R}}_h^{YX} = \frac{1}{T}\sum_{t=1}^{T-h}\left(Y_{t+h} - \hat{\mu}_Y\right) \otimes \left(X_t - \hat{\mu}_X\right). \tag{1.76}$$

The consistency of the above estimators, however, is not guaranteed solely by the weak dependence and finer assumptions are necessary. In the following we provide with the consistency results under L^4-m-approximability.

Proposition 1.3.22 (Hörmann et al. (2015b)[Lemma 5])**.**

- *Let $\{X_t\}$ be an L^4-m-approximable functional time series in $\mathscr{H}_1$. Then for fixed $h \in \mathbb{N}$, the lag-h empirical autocovariance operator* (1.75) *satisfies*

$$\mathbb{E}\left[\left\|\hat{\mathscr{R}}_h^X - \mathscr{R}_h^X\right\|_{\mathscr{L}_2(\mathscr{H}_1)}\right] \leq \frac{\kappa}{T^{1/2}}$$

  where the constant κ is independent of h and T.

- *Let $\{(Y_t, X_t)\}$ be an L^4-m-approximable functional time series in $\mathscr{H}_2 \times \mathscr{H}_1$. Then for fixed $h \in \mathbb{N}$, the lag-h empirical cross-covariance operator* (1.76) *satisfies*

$$\mathbb{E}\left[\left\|\hat{\mathscr{R}}_h^{YX} - \mathscr{R}_h^{YX}\right\|_{\mathscr{L}_2(\mathscr{H}_1,\mathscr{H}_2)}\right] \leq \frac{\kappa}{T^{1/2}}$$

  where the constant κ is independent of h and T.

An alternative treatment of the asymptotic results for the autocovariance operators estimators of autoregressive processes is presented in Bosq (2000)[Chapter 4].

Estimation in the spectral domain

To estimate the spectral density operators $\{\mathscr{F}_\omega^X\}$ one has to resort to smoothing or a different sort of regularization at some point. Panaretos and Tavakoli (2013b) performed kernel smoothing of the periodogram in the spectral domain whereas Hörmann et al. (2015a) made use of Barlett's estimate. Bartlett's estimate involves a weighted average of the lagged autocovariances, with a choice of weights that downweighs higher order lags. From the theoretical perspective, this approach is equivalent to kernel smoothing of the periodogram (Priestley, 1981a, §6.2.3). In fact, Bartlett's weights correspond to the Fourier coefficients of the smoothing kernel, assumed compactly supported. In the following we review the two approaches, starting with the periodogram smoothing.

Let $\{X_t\}_{t\in\mathbb{Z}}$ be an $\mathscr{H}$-valued functional time series where $\mathscr{H}$ is a real separable Hilbert space and denote $\mathscr{H}^{\mathbb{C}}$ its complexification. Panaretos and Tavakoli (2013b) and Tavakoli (2014) consider the discrete Fourier transform of the sample $X_0,\dots,X_{T-1}$ defined by

$$\tilde{X}_\omega = \frac{1}{2\pi T}\sum_{t=0}^{T-1} X_t e^{-\mathrm{i}\omega t}, \qquad \omega\in[-\pi,\pi],$$

where $\tilde{X}_\omega$ takes values in $\mathscr{H}^{\mathbb{C}}$, and the periodogram by

$$\mathscr{P}_\omega^X = \tilde{X}_\omega \otimes \overline{\tilde{X}_\omega}, \qquad \omega\in[-\pi,\pi],$$

where $\overline{\tilde{X}_\omega}\in\mathscr{H}^{\mathbb{C}}$ is the complex conjugate element to $\tilde{X}_\omega$, and the periodogram $\mathscr{P}_\omega^X$ is an operator in $\mathscr{L}_1(\mathscr{H}^{\mathbb{C}})$.

As in the univariate or multivariate case (Brillinger, 1983), the periodogram of the functional time series is an unbiased but inconsistent estimator of the spectral density operator $\mathscr{F}_\omega^X$. Therefore one needs to address this issue by smoothing. Consider a smoothing kernel $K(\cdot)$ as an even probability density function bounded in variation with compact support on $[0,1]$ and the weights

$$W(x) = \sum_{j\in\mathbb{Z}} \frac{1}{B_{\mathscr{P}}} K\left(\frac{x+2\pi j}{B_{\mathscr{P}}}\right)$$

where $B_{\mathscr{P}}>0$ is the smoothing bandwidth. Panaretos and Tavakoli (2013b) then defined the smoothed peridogram as

$$\tilde{\mathscr{F}}_\omega^X = \frac{2\pi}{T}\sum_{s=1}^{T-1} W\left(\omega-\frac{2\pi s}{T}\right)\mathscr{P}^X_{\frac{2\pi s}{T}}, \qquad \omega\in[-\pi,\pi], \tag{1.77}$$

and proved its asymptotic behaviour.

Proposition 1.3.23 (Panaretos and Tavakoli (2013b) or Tavakoli (2014)[§3.5])**.** *Assume that the functional time series* $\{X_t(x) : x\in[0,1]\}_{t\in\mathbb{Z}}$ *satisfies the conditions* $\mathbf{C}(\mathbf{1},\mathbf{2})$ *and* $\mathbf{C}(\mathbf{0},\mathbf{4})$ *(Definition 1.3.2) and that* $B_{\mathscr{P}}\to 0$ *as* $T\to\infty$*. Then the smoother periodogram estimator* (1.77) *is*

consistent and

$$\sup_{\omega\in[-\pi,\pi]} \mathbb{E}\left[\left\|\tilde{\mathscr{F}}_\omega^X - \mathscr{F}_\omega^X\right\|_2^2\right] = O_{\mathbb{P}}(B_{\mathscr{P}}^2 + B_{\mathscr{P}}^{-2}T^{-1}),$$
$$\int_{-\pi}^{\pi} \mathbb{E}\left[\left\|\tilde{\mathscr{F}}_\omega^X - \mathscr{F}_\omega^X\right\|_2^2\right] \mathrm{d}\omega = O_{\mathbb{P}}(B_{\mathscr{P}}^2 + B_{\mathscr{P}}^{-1}T^{-1}),$$

as $T \to \infty$.

The other approach to the smoothing of the periodogram in the spectral domain is the Bartlett's estimator (Hörmann et al., 2015a,b) in the temporal domain. Throughout the entire thesis we restrict attention to triangular window (*Barlett's weights*) defined as $W_h = (1-|h|/L)$ for $|h| < L$ and 0 otherwise for span parameter $L \in \mathbb{N}$ as it seems to be a popular choice (Hörmann and Kokoszka, 2010; Hörmann et al., 2015a).

For the samples $X_1,\dots,X_T$ and $Y_1,\dots,Y_T$ of functional time series in $\mathscr{H}_1$ and $\mathscr{H}_2$, define the Bartlett's estimator

$$\hat{\mathscr{F}}_\omega^X = \frac{1}{2\pi}\sum_{h=-L}^{L} W_h \hat{\mathscr{R}}_h^X e^{-\mathrm{i}h\omega}, \qquad \omega\in[-\pi,\pi], \tag{1.78}$$

$$\hat{\mathscr{F}}_\omega^{YX} = \frac{1}{2\pi}\sum_{h=-L}^{L} W_h \hat{\mathscr{R}}_h^{YX} e^{-\mathrm{i}h\omega}, \qquad \omega\in[-\pi,\pi], \tag{1.79}$$

where $\hat{\mathscr{R}}_h^X$ and $\hat{\mathscr{R}}_h^{YX}$ are the standard empirical autocovariance and cross-covariance operators (1.75) and (1.76).

Proposition 1.3.24 (Hörmann et al. (2015b)[Lemma 1])**.** *Assume that the Bartlett's span parameter* $L \to \infty$ *as* $T \to \infty$.

- *Let* $\{X_t\}$ *be an* L^4*-*m*-approximable functional time series in* $\mathscr{H}_1$*. Then the spectral density estimator* (1.78) *is consistent*

$$\sup_{\omega\in[-\pi,\pi]} \left\|\hat{\mathscr{F}}_\omega^X - \mathscr{F}_\omega^X\right\|_{\mathscr{L}(\mathscr{H}_1)} = O_{\mathbb{P}}\left(\psi_T^X\right), \qquad \text{as} \quad T\to\infty,$$

  where $\{\psi_T^X\}$ *is a null-sequence.*

- *Let* $\{Y_t, X_t\}$ *be an* L^4*-*m*-approximable functional time series in* $\mathscr{H}_2\times\mathscr{H}_1$*. Then the cross-spectral density estimator* (1.79) *is consistent*

$$\sup_{\omega\in[-\pi,\pi]} \left\|\hat{\mathscr{F}}_\omega^{YX} - \mathscr{F}_\omega^{YX}\right\|_{\mathscr{L}(\mathscr{H}_1,\mathscr{H}_2)} = O_{\mathbb{P}}\left(\psi_T^{YX}\right), \qquad \text{as} \quad T\to\infty,$$

  where $\{\psi_T^{YX}\}$ *is a null-sequence.*

The sequences $\{\psi_T^X\}$ *and* $\{\psi_T^{YX}\}$ *are given by complicated formulae which can be found in Hörmann et al. (2015b).*

2 Sparsely observed functional time series

This chapter introduces the core results of the thesis: the estimators of the first and the second order dynamics of sparsely observed functional time series and the cross-dependence between two functional time series. Most of the results presented in this chapter are collected from Rubín and Panaretos (2020b) and some of the cross-dependence estimation results from Rubín and Panaretos (2020a), while some of the results are extended further.

We deal in detail with the estimation when the data come from a single sparsely observed functional time series in Section 2.1. We present the estimators of the model dynamics components in the time domain: the mean function and the autocovariance operators; and the estimator of the spectral density operator. The spectral density operators encode the entire second order structure of the data, which we benefit from in Section 2.2 and show how to predict (recover) the latent functional data themselves. In Section 2.3 we analyse the asymptotic behaviour of the proposed estimators while the proofs being technical and long are postponed until Section 2.7. Section 2.4 introduces the estimators of the dependence between two time series manifested by the estimation of the cross-spectral density operator. We present estimators of the cross-spectral densities for all possible pairs of time series fully or sparsely observed functional time series, or of univariate and multivariate time series. The proposed methodology is tested on a simulation study in Section 4.1 and demonstrated on the application to the fair-weather atmospheric electricity in Section 4.3.

2.1 Estimation of functional time series dynamics from sparse data

2.1.1 Notations and sparse observation regime

We consider the functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ as a time series of continuous and smooth curves in the Hilbert space $\mathscr{H}_1 = L^2([0,1],\mathbb{R})$. The smoothness is essential for our methods and the concrete smoothness assumptions are listed in Section 2.3. Moreover, the functional time series is assumed to be observed only at discrete and sparse measurement

locations with noise contamination, i.e. we assume the model

$$U_{tj} = X_t(x_{tj}) + \epsilon^X_{tj}, \qquad j = 1, \dots, N^X_t, \quad t = 1, \dots, T, \tag{2.1}$$

where $x_{tj} \in [0,1]$ is the position of the j-th measurement on the t-th curve for $j = 1, \dots, N^X_t$ with N^X_t the number of measurements on t-th curve. The additive noise $\{\epsilon^X_{tj}\}$ is assumed to be an ensemble of independent identically distributed random variables with mean zero and variance $\sigma^2_X > 0$.

The sampling regime (2.1) is assumed to satisfy the following probabilistic assumptions:

- The measurement locations $\{x_{tj}\}$ is an ensemble of independent identically distributed random variables with values in the domain $[0,1]$.
- The counts of measurement locations $\{N_t\}$ are independent identically distributed random variables with the $\mathbb{N}_0$-valued law N satisfying $\mathbb{P}(N > 1) > 0$ and $\mathbb{E}N < \infty$. We allow $N = 0$ with positive probability resulting into the fact that some curves $X_1, \dots, X_T$ may be sampled nowhere.
- The measurement locations $\{x_{tj}\}$, their counts $\{N_t\}$, the measurement error $\{\epsilon_{tj}\}$, and the underlying functional time series $\{X_t\}$ are independent.

We assume that the time series X is second-order stationary and denote its lag-h autocovariance operator at autocovariance kernel as $\mathscr{R}^X_h$ and R^X_h respectively, as defined in Subsection 1.3.2. The elemental assumption is the weak dependence

(A1) The stationary functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ satisfies the weak dependence assumptions (1.22) and (1.25).

Therefore, the spectral density operator and the spectral density kernel are well-defined by the formulae (1.27) and (1.28) respectively. Furthermore, the lag-h autocovariance operators and kernels can be recovered by the formulae (1.29) and (1.30) respectively, by Proposition 1.3.5.

2.1.2 Estimation in the time domain

Given the sparsely observed data $\{U_{tj}\}$ generated by the observation scheme (2.1), we wish to estimate the mean function μ_X and the lagged autocovariance kernels $R^X_h(\cdot,\cdot)$. We are proposing to estimate the mean function μ_X, the lag-0 covariance kernel $R^X_0(\cdot,\cdot)$, and the measurement error variance σ^2_X by similar tools as in the independent identically distributed case reviewed in Subsection 1.2.3, therefore we adjust here the notations for the functional time series data and do not repeat the local-polynomial regression background discussed in detail in the aforementioned subsection.

Throughout the thesis, we work with the Epanechnikov kernel $K(v) = (3/4)(1-v^2)$ for $v \in [-1,1]$, and 0 otherwise, but any other usual smoothing kernel would be appropriate. Let $B_\mu > 0$ be the bandwidth parameter. We define the estimator of $\mu_X(x)$ as $\hat{\mu}_X(x) = \hat{c}_0^{(1)}$ by minimizing the weighted sum of squares:

$$\left(\hat{c}_0^{(1)}, \hat{c}_1^{(1)}\right) = \underset{c_0^{(1)}, c_1^{(1)}}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{j=1}^{N_i} K\left(\frac{x - x_{ij}}{B_\mu}\right) \left\{U_{tj} - c_0^{(1)} - c_1^{(1)}(x - x_{tj})\right\}^2. \tag{2.2}$$

For the estimation of lagged autocovariance kernel define the "raw" covariances

$$G_{h,t}^X(x_{t+h,j}, x_{tk}) = (U_{t+h,j} - \hat{\mu}_X(x_{t+h,j}))(U_{tk} - \hat{\mu}_X(x_{tk})) \tag{2.3}$$

where $|h| < T$, $t = \max(1, 1-h), \ldots, \max(T, T-h)$, $j = 1, \ldots, N_{t+h}$ and $k = 1, \ldots, N_t$.

Since the measurement errors $\{\epsilon_{tj}\}$ contribute only to the diagonal of the lag-0 autocovariance kernel, $\operatorname{cov}(U_{t+h,j}, U_{tk}) = R_h^X(x_{t+h,j}, x_{tk}) + \sigma^2 1_{[h=0, j=k]}$ where $1_{[h=0,j=k]} = 1$ if the condition in the subscript is satisfied and zero otherwise, the "raw" covariances satisfy also

$$\mathbb{E}\left[G_{h,t}^X(x_{t+h,j}, x_{tk})\right] \approx R_h^X(x_{t+h,j}, x_{tk}) + \sigma^2 1_{[h=0,j=k]}$$

where the approximation "$\approx$" stems only from the fact that the estimated $\hat{\mu}$ is supplied in (2.3). Hence, the diagonal must be removed only when estimating the lag-0 covariance kernel.

The lag-0 covariance kernel is estimated by letting $\hat{R}_0^X(x,y) = \hat{b}_0$ where $\hat{b}_0$ is obtained by minimizing the following weighted sum of squares:

$$\begin{aligned}\left(\hat{c}_0^{(2)}, \hat{c}_1^{(2)}, \hat{c}_2^{(2)}\right) = \underset{c_0^{(2)}, c_1^{(2)}, c_2^{(2)}}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{\substack{j,k=1 \\ j \neq k}}^{N_i} & K\left(\frac{x - x_{tj}}{B_R}\right) K\left(\frac{x - x_{tk}}{B_R}\right) \\ & \times \left\{G_{0,t}^X(x_{tj}, x_{tk}) - c_0^{(2)} - c_1^{(2)}(x - x_{tj}) - c_2^{(2)}(x - x_{tk})\right\}^2 \end{aligned} \tag{2.4}$$

The measurement error variance σ_X^2 is estimated in the same way as described in Subsection 1.2.3. We estimate the diagonal of $R_0^X(\cdot,\cdot)$ by the surface-smoother that is local-quadratic in the direction perpendicular to the diagonal. For $x \in [0,1]$ we set $\bar{R}_0^X(x) = \hat{c}_0^{(3)}$ where

$$\begin{aligned}\left(\hat{c}_0^{(3)}, \hat{c}_1^{(3)}, \hat{c}_2^{(3)}\right) = \underset{c_0^{(3)}, c_1^{(3)}, c_2^{(3)}}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{\substack{j=1 \\ j \neq k}}^{N_t} \sum_{k=1}^{N_i} & K\left(\frac{x_{tj} - x}{B_R}\right) K\left(\frac{x_{tk} - x}{B_R}\right) \\ & \times \left\{G_{h,t}^X(x_{tj}, x_{tk}) - c_0^{(3)} - c_1^{(3)} \Delta(x_{tj}, x_{tk}) - c_2^{(3)} \Delta(x_{tj}, x_{tk})^2\right\}^2 \end{aligned} \tag{2.5}$$

with $\Delta(\cdot,\cdot)$ defined in (1.13). The function $x \mapsto R^X(x,x) + \sigma^2$, $x \in [0,1]$ is estimated by the

local-linear line smoother by setting $\hat{V}(x) = \hat{c}_0^{(4)}$ where

$$\left(\hat{c}_0^{(4)}, \hat{c}_1^{(4)}\right) = \underset{c_0^{(4)}, c_1^{(4)}}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{j=1}^{N_t} K\left(\frac{x_{tj} - x}{B_V}\right) \left\{ G_{h,t}^X(x_{tj}, x_{tj}) - c_0^{(4)} - c_1^{(4)}(x - x_{tj}) \right\}^2. \tag{2.6}$$

for $x \in [0,1]$ and a bandwidth parameter $B_V > 0$.

The measurement error variance σ_X^2 is estimated by integrating the difference

$$\hat{\sigma}_X^2 = \int_0^1 \left(\hat{V}(x) - \bar{R}_0^X(x)\right) \mathrm{d}x. \tag{2.7}$$

Next, we proceed with the estimation of the lag-h autocovariance kernels for $h > 0$. We define the estimator $\hat{R}_h^X(x,y) = \hat{c}_0^{(5)}$ for $h = 1, \ldots, T-1$ by minimizing

$$\begin{aligned}\left(\hat{c}_0^{(5)}, \hat{c}_1^{(5)}, \hat{c}_2^{(5)}\right) = \underset{c_0^{(5)}, c_1^{(5)}, c_2^{(5)}}{\operatorname{argmin}} \sum_{t=1}^{T-h} \sum_{j=1}^{N_{t+h}} \sum_{k=1}^{N_t} K\left(\frac{x_{t+h,j} - x}{B_R}\right) K\left(\frac{x_{tk} - y}{B_R}\right) \\ \times \left\{ G_{h,t}^X(x_{t+h,j}, x_{tk}) - c_0^{(5)} - c_1^{(5)}(x_{t+h,j} - x) - c_2^{(5)}(x_{tk} - y) \right\}^2 \end{aligned} \tag{2.8}$$

For $h < 0$ we set $\hat{R}_h^X(x,y) = \hat{R}_{-h}^X(y,x)$, $x, y \in [0,1]$. Observe that we did not need to remove the diagonal as in (2.4). Denote the corresponding empirical lagged autocovariance operators, defined as the integral operators, by $\hat{\mathscr{R}}_h^X$.

2.1.3 Spectral density kernel estimation

To estimate the spectral density kernels f_ω one has to resort to smoothing or a different sort of regularization at some point. Panaretos and Tavakoli (2013b) performed kernel smoothing of the periodogram in the spectral domain whereas Hörmann et al. (2015a) made use of Barlett's estimate. Bartlett's estimate involves a weighted average of the lagged autocovariances, with a choice of weights that downweighs higher order lags. From the theoretical perspective, these two approaches are equivalent (Priestley, 1981a, §6.2.3). In fact, the weights supplied in the weighted average correspond to the Fourier coefficients of the smoothing kernel, assumed compactly supported.

We opt for the *Barlett's weights* (or the *triangular window*) defined as $W_h = (1 - |h|/L)$ for $|h| < L$ and 0 otherwise for Barlett's span parameter $L \in \mathbb{N}$ as it seems to be a popular choice (Hörmann and Kokoszka, 2010; Hörmann et al., 2015a). It should be noted that other choices of weights are possible (Rice and Shang, 2017) and the so-called local quadratic windows (Parzen, Bartlett-Pristley, etc.) improve the asymptotic bias. See Priestley (1981a, §7.5) for the detailed discussion in one-dimensional case. The statement seems to also be true for functional time series (van Delft, 2019).

We could use the formula (1.78) and plug-in the smoothed autocovariance kernels obtained in

Section 2.1.2 but instead, we opt to show how to directly construct a smoother-based estimator of the spectral density kernels because this approach is, in a sense canonical. Specifically, we estimate the spectral density kernel at frequency $\omega \in [-\pi, \pi]$ by the local-linear surface-smoother applied to the "raw" covariances multiplied by complex exponentials. The weights for the smoother are based both on the spatial distance from the raw covariances as well as the time lag. Specifically, for each $\omega \in [-\pi, \pi]$ we estimate the spectral density kernel as

$$\hat{f}_\omega(x, y) = \frac{L}{2\pi} \hat{c}_0^{(6)} \in \mathbb{C} \tag{2.9}$$

where $\hat{c}_0^{(6)}$ is obtained by minimizing the following weighted sum of squares

$$\begin{aligned}(\hat{c}_0^{(6)}, \hat{c}_1^{(6)}, \hat{c}_2^{(6)}) = \underset{(c_0^{(6)}, c_1^{(6)}, c_2^{(6)}) \in \mathbb{C}^3}{\arg\min} \sum_{h=-L}^{L} \frac{1}{\mathcal{N}_h} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j \neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \Big| G_{h,t}^X(x_{t+h,j}, x_{tk}) e^{-\mathrm{i}h\omega} - \\ - c_0^{(6)} - c_1^{(6)}(x_{t+h,j} - x) - c_2^{(6)}(x_{tk} - y) \Big|^2 W_h \frac{1}{B_R^2} K\left(\frac{x_{t+h,j} - x}{B_R}\right) K\left(\frac{x_{tk} - y}{B_R}\right)\end{aligned} \tag{2.10}$$

where $\mathcal{N}_h = (T - |h|)(\bar{N})^2$ for $h \neq 0$, $\mathcal{N}_0 = T(\overline{N^2} - \bar{N})$, and where $\bar{N} = (1/n)\sum_{t=1}^T N_t$ and $\overline{N^2} = (1/n)\sum_{t=1}^T N_t^2$.

It turns out that the minimizer of this complex minimization problem can be expressed explicitly. Moreover, the minimizer depends only on a few quantities that are independent of ω, and can be pre-calculated. The estimator can be thus constructed for a given ω by multiplying these quantities by complex exponentials and performing a handful of inexpensive arithmetic operations. Consequently, it is computationally feasible to evaluate the estimator (2.9) on a dense grid of frequencies. The explicit form is stated in (2.72).

Denote the integral operator corresponding to $\hat{f}_\omega(\cdot,\cdot)$ as $\hat{\mathscr{F}}_\omega$. We can go back to the temporal domain by integrating the spectral density and reproduce the estimators of the autocovariance kernels and operators by the inverse formulae (1.29) and (1.30)

$$\tilde{R}_h(\cdot,\cdot) = \int_{-\pi}^{\pi} \hat{f}_\omega(\cdot,\cdot) e^{\mathrm{i}h\omega} \,\mathrm{d}\omega, \qquad \tilde{\mathscr{R}}_h = \int_{-\pi}^{\pi} \hat{\mathscr{F}}_\omega e^{\mathrm{i}h\omega} \,\mathrm{d}\omega. \tag{2.11}$$

The estimators of spectral density kernels $\hat{f}_\omega(\cdot,\cdot), \omega \in [-\pi, \pi]$, are achieved by kernel smoothing. Therefore, especially for smaller sample sizes, the operators $\hat{\mathscr{F}}_\omega$, $\omega \in [-\pi, \pi]$, might not be strictly non-negative, and may feature some tail negative eigenvalues of small modulus. To ensure numerical stability of the method in the following section, it is recommended to truncate these negative eigenvalues of $\hat{\mathscr{F}}_\omega$ at each frequency $\omega \in [-\pi, \pi]$.

2.1.4 Periodic behaviour identification

As discussed at the beginning of Section 2.1.3, the choice of Bartlett's span parameter L is related to the bandwidth for smoothing in the frequency domain. To achieve consistent spectral density estimation, the parameter L needs to be kept quite small (cf. condition (B10) and Theorem 2.3.2). However, for the purpose of exploratory data analysis, it is useful to explore the data for periodic behaviour in a similar way as a periodogram is used in the case of scalar time series.

When the periodicity examination is indeed of interest, we propose to evaluate the estimator (2.9) for a fairly large value of L. The selection of adequate value of L is a question of computational power available because the computational time to evaluate (2.9) grows linearly in L. In the data analysis Section 4.3 we work with $L = 1000$, which is roughly half of the considered time series length.

Once the estimator (2.9) is evaluated for a given value of L we propose to calculate the trace of the spectral density operator at frequency $\omega \in (0, \pi)$. Peaks in this plot indicate periodic behaviour of the functional time series. The existence of periodicity is not only a useful insight into the nature of the data but may us prompt into approaching the periodic behaviour in a different way, for example by modelling the periodicity in a deterministic way as we do it in the data analysis carried out in Section 4.3.

2.2 Functional data recovery framework and confidence bands

We now consider the problem of recovering the latent functional data $\{X_t(x) : x \in [0,1]\}$ given the sparse noisy samples $\{U_{tj}\}$, and provide with confidence bands to quantify uncertainty of the prediction.

2.2.1 Introduction: best linear unbiased predictors

In this short subsection we outline the construction of predictors of random elements given an finite dimensional observation. The construction requires the knowledge of the second order structure and guarantees optimality in the class of unbiased and linear functions of the observation.

Let X be a random element in the Hilbert space $\mathscr{H}$ and $Y \in \mathbb{R}^d$ be a random vector with finite second moments, and denote their means, the covariance operator and matrix, and the cross-covariance operator

$$\mu_X = \mathbb{E}X \in \mathscr{H}, \qquad \mathscr{R}^X = \operatorname{var}(X) = \mathbb{E}\left[(X-\mu_X)\otimes(X-\mu_X)\right] \in \mathscr{L}_1(\mathscr{H}), \tag{2.12}$$

$$\mu_Y = \mathbb{E}Y \in \mathbb{R}^d, \qquad \mathbf{R}^Y = \operatorname{var}(Y) = \mathbb{E}\left[(Y-\mu_Y)(Y-\mu_Y)^\top\right] \in \mathbb{R}^{d\times d}, \tag{2.13}$$

$$\mathscr{R}^{XY} = \operatorname{cov}(X,Y) = \mathbb{E}\left[(X-\mu_X)\otimes(Y-\mu_Y)\right] \in \mathscr{L}_2(\mathbb{R}^d,\mathscr{H}). \tag{2.14}$$

We are interested in constructing a predictor of X as a function of the observed value of Y. The predictor $\Pi(X,Y)$ is called the *best linear unbiased predictor* of X, if it is linear function of Y and

- is *unbiased*: $\mathbb{E}[\Pi(X,Y)-X]=0$,
- the covariance operator of $\operatorname{var}(\Pi(X,Y)-X)=\mathbb{E}[(\Pi(X,Y)-X)\otimes(\Pi(X,Y)-X)]$ is minimized, i.e. for any other linear unbiased predictor $\tilde{X}$, the self-adjoint operator $\operatorname{var}(\tilde{X}-X)-\operatorname{var}(\Pi(X,Y)-X)$ is non-negative definite.

The following proposition provides with a recipe how to construct the best linear unbiased predictors given the second order structure.

Proposition 2.2.1. *Let X and Y admit the moments* (2.12), (2.13), (2.14), *and assume that the matrix $\mathbf{R}^Y$ is invertible. Then the best linear unbiased predictor of X given Y, denoted as $\Pi(X,Y)$, is determined by the formula*

$$\Pi(X,Y) = \mu_X + \mathscr{R}^{XY}\left(\mathbf{R}^Y\right)^{-1}\left(Y-\mu_Y\right). \tag{2.15}$$

Proof. The predictor given the formula (2.15) is clearly linear in Y and unbiased by

$$\mathbb{E}[\Pi(X,Y)] = \mu_X + \mathscr{R}^{XY}\left(\mathbf{R}^Y\right)^{-1}\left(\mu_Y-\mu_Y\right) = \mu_X.$$

It remains to show that the covariance operator $\operatorname{var}(\Pi(X,Y)-X)$ is minimised among unbiased linear predictors. We show that $\Pi(X,Y)$ is in fact the projection of X onto the subspace of random elements on $\mathscr{H}$ of the form $a+\mathscr{B}Y$ where $a\in\mathscr{H}$ and $\mathscr{B}\in\mathscr{L}(\mathbb{R}^d,\mathscr{H})$. It is sufficient to verify that $\Pi(X,Y)-X$ is orthogonal to any such $a+\mathscr{B}Y$. Indeed,

$$\begin{aligned}\operatorname{cov}\left(\Pi(X,Y)-X, a+\mathscr{B}Y\right) &= \operatorname{cov}\Big(\mathscr{R}^{XY}\left(\mathbf{R}^Y\right)^{-1}\left(Y-\mu_Y\right)-X, Y\Big)\mathscr{B}^* = \\ &= \mathscr{R}^{XY}\left(\mathbf{R}^Y\right)^{-1}\mathbf{R}^Y\mathscr{B}^* - \mathscr{R}^{XY}\mathscr{B}^* = 0.\end{aligned}$$

Now, since $\Pi(X,Y)$ is the projection of X onto the subspace of random elements on $\mathscr{H}$ of the form $a+\mathscr{B}Y$, the distance $\operatorname{var}(a+\mathscr{B}Y-X)$ is minimised for $a+\mathscr{B}Y=\Pi(X,Y)$. □

2.2.2 Prediction of latent functional data

Coming back to the problem of sparsely observed functional time series defined in Subsection 2.1.1, consider the random element $\mathbb{X}_T=[X_1,\ldots,X_T]\in\mathscr{H}^{\otimes T}$ composed of "stacked" functional data (formally, it is an element of the product Hilbert space $\mathscr{H}^{\otimes T}$). The first and

second order structure of this random element is given by

$$\mathbb{E}[\mathbb{X}_T] = \mathbb{M}_T = [\mu, \dots, \mu] \in \mathscr{H}^{\otimes T}, \tag{2.16}$$

$$\operatorname{var}(\mathbb{X}_T) = \mathbb{S}_T = \begin{bmatrix} \mathscr{R}_0 & \mathscr{R}_1^* & \mathscr{R}_2^* & \dots & \mathscr{R}_{T-1}^* \\ \mathscr{R}_1 & \mathscr{R}_0 & \mathscr{R}_1^* & \dots & \mathscr{R}_{T-2}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathscr{R}_{T-1} & \mathscr{R}_{T-2} & \mathscr{R}_{T-3} & \dots & \mathscr{R}_0 \end{bmatrix} \in \mathscr{L}(\mathscr{H}^{\otimes T}). \tag{2.17}$$

Define the stacked observables as $\mathbb{U}_T = (U_{11}, \dots, U_{1N_1}, \dots, U_{t1}, \dots, U_{tN_t}, \dots, U_{TN_1}, \dots, U_{TN_T}) \in \mathbb{R}^{\mathscr{N}_1^T}$ where $\mathscr{N}_1^{(T)} = \sum_{t=1}^T N_t$ is the total number of observations up to time T. By analogy to $\mathbb{U}_T$, stack the measurement errors $\{\epsilon_{tj}\}$ and denote this vector $\mathscr{E}_T \in \mathbb{R}^{\mathscr{N}_1^{(T)}}$. Note that $\operatorname{var}(\mathscr{E}_T) = \sigma^2 I_{\mathscr{N}_1^{(T)}}$ where $I_{\mathscr{N}_1^{(T)}}$ is the identity matrix of size $\mathscr{N}_1^{(T)}$. Further define the evaluation operators H_t, $t = 1, \dots, T$, and the joint evaluation operator $\mathbb{H}_T$ by

$$\begin{aligned} H_t : \quad \mathscr{H} &\to \mathbb{R}^{N_t}, \\ g &\mapsto (g(x_{t1}), \dots, g(x_{tN_t})), \\ \mathbb{H}_T : \quad \mathscr{H}^{\otimes T} &\to \mathbb{R}^{\mathscr{N}_1^{(T)}}, \\ [g_1, \dots, g_T] &\mapsto [H_1 g_1, \dots, H_T g_T]. \end{aligned}$$

Finally, for $t = 1, \dots, T$, define the projection operator

$$\begin{aligned} P_t : \quad \mathscr{H}^{\otimes T} &\to \mathscr{H}, \\ [g_1, \dots, g_T] &\mapsto g_t. \end{aligned}$$

In this notation we can rewrite the observation scheme (2.1) as

$$\mathbb{U}_T = \mathbb{H}_T \mathbb{X}_T + \mathscr{E}_T.$$

By Proposition 2.2.1, the best linear unbiased predictor of $\mathbb{X}_T$ given $\mathbb{U}_T$, which we denote by $\Pi(\mathbb{X}_T, \mathbb{U}_T)$, is given by the formula

$$\Pi(\mathbb{X}_T, \mathbb{U}_T) = \mathbb{M}_T + \mathbb{S}_T \mathbb{H}_T^* (\mathbb{H}_T \mathbb{S}_T \mathbb{H}_T^* + \sigma^2 I_{\mathscr{N}_1^{(T)}})^{-1} (\mathbb{U}_T - \mathbb{H}_T \mathbb{M}_T) \quad \in \mathscr{H}^{\otimes T} \tag{2.18}$$

where $*$ denotes the adjoint operator. The term $\mathbb{H}_T \mathbb{S}_T \mathbb{H}_T^*$ is in fact a positive semi-definite matrix. Owing to the fact that $\sigma^2 > 0$, the matrix $\mathbb{H}_T \mathbb{S}_T \mathbb{H}_T^* + \sigma^2 I_{\mathscr{N}_1^{(T)}}$ is always invertible.

Now fix $s \in \{1, \dots, T\}$. The best linear unbiased predictor of the functional datum X_s, which we denote by $\Pi(X_s, \mathbb{U}_T)$, is given by

$$\Pi(X_s, \mathbb{U}_T) = P_s \Pi(\mathbb{X}_T, \mathbb{U}_T) \quad \in \mathscr{H}. \tag{2.19}$$

Hence the recovery of X_s by the formula (2.19) uses the observed data across all $t = 1, \dots, T$, borrowing strength across all the observations.

In practice, however, we need to replace the unknown parameters involved in the construction of the predictor by their estimates. Define $\hat{\mathbb{M}}_T$ and $\hat{\mathbb{S}}_T$ by substituting $\hat{\mu}$ and $\tilde{\mathscr{R}}_h$ for their theoretical counterparts in formulae (2.16) and (2.17) respectively. We stress here the fact that we use the estimator $\tilde{\mathscr{R}}_h$ obtained by the inverse Fourier transformation (2.11) of the estimated spectral density operators $\{\hat{\mathscr{F}}_\omega\}_{\omega\in[-\pi,\pi]}$. We do so, because the full covariance structure induced by $\tilde{\mathscr{R}}_h$ for all $h \in \mathbb{Z}$ is guaranteed to be positive-definite and its estimation error is asymptotically controlled, see Corollary 2.3.3 in the following section. Both these facts are essential for proper prediction.

Now replace $\mathbb{M}_T$, $\mathbb{S}_T$, σ^2 by $\hat{\mathbb{M}}_T$, $\hat{\mathbb{S}}_T$ and $\hat{\sigma}^2$, respectively, in formulae (2.18) and (2.19). The resulting predictors are denoted by

$$\hat{\Pi}(\mathbb{X}_T, \mathbb{U}_T) = \hat{\mathbb{M}}_T + \hat{\mathbb{S}}_T \mathbb{H}_T^* (\mathbb{H}_T \hat{\mathbb{S}}_T \mathbb{H}_T^* + \sigma^2 I_{\mathscr{N}_1^{(T)}})^{-1} (\mathbb{U}_T - \mathbb{H}_T \hat{\mathbb{M}}_T) \tag{2.20}$$

and

$$\hat{\Pi}(X_s, \mathbb{U}_T) = P_s \hat{\Pi}(\mathbb{X}_T, \mathbb{U}_T). \tag{2.21}$$

In order to construct confidence bands later in this subsection we need to impose distributional assumptions, specifically we work under the Gaussian assumption:

(A2) The functional time series $\{X_t\}_t$ as well as the measurement errors $\{\epsilon_{tj}\}_{tj}$ are Gaussian processes.

Thanks to the Gaussian assumption (A2), the predictors of $\mathbb{X}_T$ and X_s given by formulae (2.18) and (2.19) are in fact given by conditional expectations. Furthermore, we can calculate the exact conditional distribution of $\mathbb{X}_T$ given $\mathbb{U}_T$ by the formula

$$\mathbb{X}_T | \mathbb{U}_T \sim N_{\mathscr{H}^{\otimes T}}(\mathbb{M}_{\mathbb{X}_T|\mathbb{U}_T}, \mathbb{S}_{\mathbb{X}_T|\mathbb{U}_T}) \tag{2.22}$$

where

$$\mathbb{M}_{\mathbb{X}_T|\mathbb{U}_T} = \mathbb{M}_T + \mathbb{S}_T \mathbb{H}_T^* (\mathbb{H}_T \mathbb{S}_T \mathbb{H}_T^* + \sigma^2 I_{\mathscr{N}_1^{(T)}})^{-1} (\mathbb{U}_T - \mathbb{H}_T \mathbb{M}_T), \tag{2.23}$$

$$\mathbb{S}_{\mathbb{X}_T|\mathbb{U}_T} = \mathbb{S}_T - \mathbb{S}_T \mathbb{H}_T^* (\mathbb{H}_T \mathbb{S}_T \mathbb{H}_T^* + \sigma^2 I_{\mathscr{N}_1^{(T)}})^{-1} \mathbb{H}_T \mathbb{S}_T. \tag{2.24}$$

From (2.22) we can access the conditional distribution of X_s for fixed $s = 1, \dots, T$, by writing

$$X_s | \mathbb{U}_T \sim N_{\mathscr{H}}(\mathbb{M}_{X_s|\mathbb{U}_T}, \mathbb{S}_{X_s|\mathbb{U}_T}) \tag{2.25}$$

where

$$\mathbb{M}_{X_s|\mathbb{U}_T} = P_s \mathbb{M}_{\mathbb{X}_T|\mathbb{U}_T}, \qquad \mathbb{S}_{X_s|\mathbb{U}_T} = P_s \mathbb{S}_{\mathbb{X}_T|\mathbb{U}_T} P_s^*. \tag{2.26}$$

2.2.3 Confidence bands for functional data

To construct a band for X_s with pointwise coverge, we construct a confidence interval for $X_s(x)$ at each $x \in [0,1]$ — as we will see, the endpoints of these intervals are continuous functions of x, and so automatically define a confidence band. In practice, one constructs bands for a dense collection of locations in $[0,1]$ and interpolates. Given the conditional distribution $X_s(x)|\mathbb{U}_T \sim N(\mathbb{M}_{X_s|\mathbb{U}_T}(x), \mathbb{S}_{X_s|\mathbb{U}_T}(x,x))$, the $(1-\alpha)$-confidence interval for fixed $x \in [0,1]$ is constructed as

$$\mathbb{M}_{X_s|\mathbb{U}_T}(x) \pm \Phi^{-1}(1-\alpha/2)\sqrt{\mathbb{S}_{X_s|\mathbb{U}_T}(x,x)} \tag{2.27}$$

where $\Phi^{-1}(1-\alpha/2)$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

In practice, when we do not know the true dynamics of the functional time series, we have to use the estimates of $\mu_X(\cdot)$ and $R_h(\cdot,\cdot)$. We define $\hat{\mathbb{M}}_{\mathbb{X}_T|\mathbb{U}_T}, \hat{\mathbb{S}}_{\mathbb{X}_T|\mathbb{U}_T}, \hat{\mathbb{M}}_{X_s|\mathbb{U}_T}$ and $\hat{\mathbb{S}}_{X_s|\mathbb{U}_T}$ by replacing $\mathbb{M}_T$ and $\mathbb{S}_T$ with $\hat{\mathbb{M}}_T$ and $\hat{\mathbb{S}}_T$ in the formulae (2.23), (2.24), (2.26) respectively. Therefore the asymptotic confidence interval for $X_s(x)$ is obtain by rewriting (2.27) using the empirical counterparts

$$\hat{\mathbb{M}}_{X_s|\mathbb{U}_T}(x) \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\mathbb{S}}_{X_s|\mathbb{U}_T}(x,x)}. \tag{2.28}$$

For the construction of the simultaneous band we use the method introduced by Degras (2011). Fix $s = 1,\ldots,T$. In the previous section we derived the conditional distribution of X_s given $\mathbb{U}_T$ in formula (2.25). Define the conditional correlation kernel

$$\rho_{X_s|\mathbb{U}_T}(x,y) = \begin{cases} \frac{\mathbb{S}_{X_s|\mathbb{U}_T}(x,y)}{\sqrt{\mathbb{S}_{X_s|\mathbb{U}_T}(x,x)\mathbb{S}_{X_s|\mathbb{U}_T}(y,y)}}, & \mathbb{S}_{X_s|\mathbb{U}_T}(x,x) > 0, \quad \mathbb{S}_{X_s|\mathbb{U}_T}(y,y) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.29}$$

Then, the collection of intervals

$$\left\{\mathbb{M}_{X_s|\mathbb{U}_T}(x) \pm z_{\alpha,\rho}\sqrt{\mathbb{S}_{\mathbb{X}_T|\mathbb{U}_T}(x,x)} : x \in [0,1]\right\}, \tag{2.30}$$

forms a (continuous) confidence band with simultaneous coverage probability $(1-\alpha)$ over $x \in [0,1]$. Here $z_{\alpha,\rho}$ is the $(1-\alpha)$-quantile of the law of $\sup_{x\in[0,1]}|Z(x)|$ where $\{Z(x), x \in [0,1]\}$ is a zero mean Gaussian process with covariance kernel $\rho_{\mathbb{X}_T|\mathbb{U}_T}$. The definition of a quantile specifically requires that $\mathbb{P}(\sup_{x\in[0,1]}|Z(x)| \le z_{\alpha,\rho}) = 1-\alpha$. Degras (2011) explains how to calculate this quantile numerically.

In practice, we replace the population level quantities in (2.30) by their estimated counterparts and define the asymptotic simultaneous confidence band as

$$\left\{\hat{\mathbb{M}}_{X_s|\mathbb{U}_T}(x) \pm z_{\alpha,\hat{\rho}}\sqrt{\hat{\mathbb{S}}_{\mathbb{X}_T|\mathbb{U}_T}(x,x)} : x \in [0,1]\right\}, \tag{2.31}$$

where $\hat{\mathbb{M}}_{X_s|\mathbb{U}_T}(x)$ and $\hat{\mathbb{S}}_{\mathbb{X}_T|\mathbb{U}_T}(x,x)$ are as above and the quantile $z_{\alpha,\hat{\rho}}$ is calculated for the correlation structure $\hat{\rho}_{X_s|\mathbb{U}_T}$ defined as the empirical counterpart to (2.29).

Note that $\Phi^{-1}(1-\alpha/2) < z_{\alpha,\rho}$ for any correlation kernel ρ (Degras, 2011). Therefore, as expected, the pointwise confidence bands are enveloped by the simultaneous band. Once again, in practice, one evaluates the band limits defining (2.31) on a dense grid of $[0,1]$ and interpolates.

2.2.4 Forecasting

A natural next step to consider, and indeed one of the main reasons why one may be interested in recovering the functional time-series dynamics, is that of forecasting. In this section, we comment on how the forecasting problem naturally fits into the functional data recovery framework introduced above.

Assume that we are given sparse data $\{U_{tj} : 1 \le j \le N_t, 1 \le t \le T\}$ and we wish to forecast the functional datum X_{T+r} for $r \in \mathbb{N}$ as well as to quantify the uncertainty of the forecast. We define the random element $\mathbb{X}_{T+r} = [X_1, \ldots, X_T, X_{T+1}, \ldots, X_{T+r}] \in \mathscr{H}^{T+r}$. If the forecasts for the intermediate data $X_{T+1}, \ldots, X_{T+r-1}$ are not of interest, we may delete these elements and naturally alter the explained method below. Nevertheless, we opt to explain the approach for forecasting up to the time $T+r$ simultaneously.

By the same notation as above, we extend formulae (2.16) and (2.17) for $t = 1, \ldots, T+r$ and obtain the law of $\mathbb{X}_{T+r}$, i.e. the joint law of $X_1, \ldots, X_{T+r}$ and we can calculate their conditional distribution given the observed data $\mathbb{U}_T$. In particular, by taking $s = T+r$ in the equations (2.19), (2.27), and (2.30) we obtain the forecast, the pointwise confidence band, and the simultaneous confidence band respectively for the functional datum X_{T+r}. In practice, we substitute the unknown population level quantities by their empirical estimators. Therefore, by taking $s = T+r$ in the equations (2.21), (2.28), and (2.31) we obtain the forecast, the (asymptotic) pointwise confidence band, and the (asymptotic) simultaneous confidence band for X_{T+r}.

2.3 Asymptotic results

2.3.1 On the choice of mixing conditions

In Sections 2.3.2 and 2.3.3 we develop asymptotic theory for our methodology under two different sets of assumptions.

Firstly, in Section 2.3.2 we prove the asymptotic behaviour of the estimators under Brillinger-type cumulant mixing conditions. The corresponding Theorems 2.3.1 and 2.3.2 are in a sense canonical, in that their proofs rely on generalisations of the techniques by Yao et al. (2005a). Nevertheless, the yielded convergence rates for one dimensional smoothing and surface smoothing are $O_{\mathbb{P}}(1/(\sqrt{T}B_\mu))$ and $O_{\mathbb{P}}(1/(\sqrt{T}B_R^2))$, respectively, which are not optimal.

The optimal rates for one dimensional smoothing and surface smoothing are known to be

$O_{\mathbb{P}}(\sqrt{\log T/(TB_\mu)})$ and $O_{\mathbb{P}}(\sqrt{\log T/(TB_R^2)})$ respectively. Recovering such rates using local-regression methods for time-series data relies heavily on the employed measure of weak dependence, namely strong mixing conditions, (Hansen, 2008; Liebscher, 1996; Masry, 1996), (Fan and Yao, 2008, Thm 6.5), geometric strong mixing conditions (Bosq, 2012, Thm. 2.2 and Cor. 2.2), and ρ-mixing conditions (Peligrad, 1992). In Section 2.3.3 and Theorems 2.3.4, 2.3.5 we benefit from the techniques developed by Hansen (2008) to obtain the optimal rates under strong mixing.

Since these two sets of rates rest on qualitatively different conditions, we have chosen to include both results in this section.

2.3.2 Asymptotic results under cumulant mixing conditions

In order to establish the consistency and the convergence rate of the estimators introduced in Section 2.1, we will make use of the following further assumptions on the model (2.1).

(B1) The number of measurements N_t^X in time t are independent random variables with law $N_t^X \sim N$ where $N^X \geq 0$, $\mathbb{E}\left[N^X\right] < \infty$ and $\mathbb{P}(N^X > 1) > 0$.

(B2) The measurement locations x_{tj}, $j = 1, \ldots, N_t^X$, $t = 1, \ldots, T$ are independent random variables generated from the density $g^X(\cdot)$ and are independent of the number of measurements $(N_t^X)_{t=1,\ldots,T}$. The density $g^X(\cdot)$ is assumed to be twice continuously differentiable and strictly positive on $[0,1]$.

We allow the event $\{N_t^X = 0\}$ to potentially have positive probability. This corresponds to the situation where no measurements are available at time t, for example when we additionally have missing data at random. We also need to impose smoothness conditions on the unknown functional parameters

(B3) The common mean function, $\mu_X(\cdot)$, is twice continuously differentiable on $[0,1]$.

(B4) The lag-h autocovariance kernels, $R_h^X(\cdot,\cdot)$, are twice continuously differentiable on $[0,1]^2$ for each $h \in \mathbb{Z}$. Moreover,

$$\sup_{x,y\in[0,1]} \left| \frac{\partial^2}{\partial x^{\alpha_1} \partial y^{\alpha_2}} R_h^X(x,y) \right|$$

is uniformly bounded in h for all combinations of $\alpha_1, \alpha_2 \in \mathbb{N}_0$ where $\alpha_1 + \alpha_2 = 2$.

To prove the consistency of the autocovariance kernels and the cross-covariance kernels estimators we need to further assume some mixing conditions in the time domain. The smoothing estimators are essentially moment-based, therefore it is natural to consider cumulant-type summability conditions. The definition of the cumulants is presented in Subsection 1.3.3

(B5) On top of the generally assumed second order stationarity, assume that $\{X_t\}$ is moreover fourth order stationary and denote the 4-th order cumulant kernel of $\{X_t\}$ as $\mathrm{cum}(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})(\cdot,\cdot,\cdot,\cdot)$. Assume the summability in the supremum norm

$$\sum_{h_1,h_2,h_3=-\infty}^{\infty} \sup_{x_1,x_2,x_3,x_4\in[0,1]} \left|\mathrm{cum}(X_{h_1}, X_{h_2}, X_{h_3}, X_0)(x_1,x_2,x_3,x_4)\right| < \infty.$$

The following condition is required for the proof of the convergence rate of the spectral density kernels estimator.

(B6) Assume

$$\sum_{h=-\infty}^{\infty} |h| \sup_{x,y\in[0,1]} \left|R_h^X(x,y)\right| < \infty.$$

The last two conditions correspond to the conditions **C′(0,4)** and **C′(1,2)** in Definition 1.3.2. Finally, we impose the following assumptions on the decay rate of the bandwidth parameters and the growing rate of the Bartlett's span parameter L

(B7) $B_\mu \to 0$, $TB_\mu^4 \to \infty$,

(B8) $B_R \to 0$, $TB_R^6 \to \infty$,

(B9) $B_V \to 0$, $TB_V^4 \to \infty$,

(B10) $L \to \infty$, $L = o(\sqrt{T}B_R^2)$, $L = o(B_R^{-2})$.

We may now state our asymptotic results on uniform consistency and convergence rates:

Theorem 2.3.1. *Under the assumptions (A1), (B1) — (B3) and (B7):*

$$\sup_{x\in[0,1]} |\hat{\mu}_X(x) - \mu_X(x)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_\mu} + B_\mu^2\right). \tag{2.32}$$

Under the assumptions (B1) — (B5) and (B7) — (B9), for fixed lag $h \in \mathbb{Z}$:

$$\sup_{x,y\in[0,1]} |\hat{R}_h^X(x,y) - R_h^X(x,y)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_R^2} + B_R^2\right), \tag{2.33}$$

$$\hat{\sigma}_X^2 = \sigma_X^2 + O_{\mathbb{P}}\left\{\frac{1}{\sqrt{T}}\left(\frac{1}{B_V} + \frac{1}{B_R^2}\right) + B_\mu^2 + B_R^2\right\}. \tag{2.34}$$

Theorem 2.3.2. *Under the assumptions (A1), (B1) — (B5) and (B7) — (B10), the spectral density is estimated consistently:*

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \left|\hat{f}_\omega^X(x,y) - f_\omega^X(x,y)\right| = o_{\mathbb{P}}(1). \tag{2.35}$$

If we further assume condition (B6), we can additionally obtain the convergence rate:

$$\sup_{\omega\in[-\pi,\pi]}\sup_{x,y\in[0,1]}\left|\hat{f}_\omega^X(x,y)-f_\omega^X(x,y)\right| = O_{\mathbb{P}}\left(L\frac{1}{\sqrt{T}}\frac{1}{B_R^2}+LB_R^2+\frac{1}{L}\right).$$

As a consequence of Theorem 2.3.2 we obtain the consistency and the convergence rate of the entire space-time covariance structure (2.11), i.e. rates uniform in both time index and spatial argument:

Corollary 2.3.3. *Under the assumptions (A1), (B1) — (B5) and (B7) — (B10):*

$$\sup_{h\in\mathbb{Z}}\sup_{x,y\in[0,1]}|\tilde{R}_h^X(x,y)-R_h^X(x,y)| = o_{\mathbb{P}}(1) \tag{2.36}$$

and assuming further (B6):

$$\sup_{h\in\mathbb{Z}}\sup_{x,y\in[0,1]}|\tilde{R}_h^X(x,y)-R_h^X(x,y)| = O_{\mathbb{P}}\left(L\frac{1}{\sqrt{T}}\frac{1}{B_R^2}+LB_R^2+\frac{1}{L}\right). \tag{2.37}$$

2.3.3 Asymptotic results under strong mixing conditions

We begin by listing the assumptions leading to the optimal convergence rates. Besides imposing the key assumption of the strong mixing we need to strengthen some of the other assumptions as well. We require some additional regularity conditions on the smoothing kernel $K(\cdot)$ which until now was assumed only to be a bounded probability density function. The condition is formulated for a generic multivariate function $k(\cdot)$. In the proofs of the theorems in this subsection, we shall plug-in the functions related to $K(\cdot)$, see the assumption (D5), into $k(\cdot)$.

(C1) The function $k:\mathbb{R}^d\to\mathbb{R}$ is bounded and integrable

$$|k(u)|\le\bar{k}<\infty,\qquad\int_{\mathbb{R}^d}|k(u)|\,\mathrm{d}u<\infty,$$

and for some $\Lambda_1<\infty$ and $L<\infty$, either $k(u)=0$ for $|u|>\tilde{L}$ and

$$\left|k(u)-k(u')\right|\le\Lambda_1\|u-u'\|,\qquad u,u'\in\mathbb{R},$$

or $k(\cdot)$ is differentiable, $|(\partial/\partial u)k(u)|\le\Lambda_1$, and for some $\nu>1$, $|(\partial/\partial u)k(u)|\le\Lambda_1\|u\|^{-\nu}$ for $\|u\|>\tilde{L}$.

The following conditions impose more conditions on the functional time series model.

(D1) The functional time series $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary and strong mixing with mixing

coefficients α_m that satisfy

$$\alpha(m) \leq Am^{-\beta},$$

for $A < \infty$ and for some $s > 2$,

$$\mathbb{E}|X_t(x)|^s \leq B_1 < \infty, \qquad x \in [0,1]$$

and moreover

$$\beta > \frac{2s-1}{s-2}.$$

(D2) The number of measurement locations N_t in time t are independent identically-distributed random variables with law $N_t \sim N$ where $N \in \{0,1,\ldots,N^{max}\}$ for some $N^{max} \in \mathbb{N}$ and such that $\mathbb{P}(N > 1) > 0$.

(D3) The measurement errors $\{\epsilon_{tj}\}$ are independent identically-distributed mean-zero random variables satisfying

$$\mathbb{E}|\epsilon_{tj}|^s < \infty.$$

Moreover, $\{\epsilon_{tj}\}$ are independent of the functional time series $\{X_t(\cdot)\}$.

(D4) The marginal density of the observation location $g^X(\cdot)$ satisfies

$$0 < B_2 \leq \inf_{x\in[0,1]} g^X(x) \leq \sup_{x\in[0,1]} g^X(x) \leq B_3 < \infty.$$

For the estimation of the mean function $\mu^X(\cdot)$ the following assumptions are required:

(D5) The smoothing kernel $K(\cdot)$ satisfies $\int |u|^4 K(u)\,\mathrm{d}u < \infty$ and the functions $u \mapsto K(u)$, $u \mapsto uK(u)$, $u \mapsto u^2K(u)$ satisfy the assumption (C1).

(D6) The bandwidth parameter B_μ satisfies

$$\frac{\log T}{T^{\theta_\mu} B_\mu} = o(1), \qquad T \to \infty,$$

with

$$\theta_\mu = \frac{\beta - 2 - (1+\beta)/(s-1)}{\beta + 1 - (1+\beta)/(s-1)}.$$

(D7) The functions $g(\cdot)$ and $g(\cdot)\mu_X(\cdot)$ are twice continuously differentiable on $[0,1]$.

The rates for the lag-h autocovariance kernel estomator(s) will require the following set of assumptions:

(D8) The functional time series $\{X_t(\cdot)\}$ is 4-th order stationary satisfies

$$\sup_{x\in[0,1]} \mathbb{E}|X_0(x)|^4 < \infty.$$

(D9) The smoothing kernel $K(\cdot)$ satisfies $\iint |uv|^4 K(u)K(v)\,\mathrm{d}u\,\mathrm{d}v < \infty$ and the functions $(u,v) \mapsto u^p v^q K(u)K(v)$ satisfy the assumption (C1) for $p, q \in \mathbb{N}_0, 0 \le p+q \le 2$.

(D10) The bandwidth parameter B_R^2 satisfies

$$\frac{\log T}{T^{\theta_R} B_R} = o(1), \qquad T \to \infty,$$

with

$$\theta_R = \frac{\beta - 3 - (1+\beta)/(s-1)}{\beta + 1 - (1+\beta)/(s-1)}.$$

(D11) The functions $(x,y) \mapsto g(x)g(y)$ and $(x,y) \mapsto g(x)g(y)R_h(x,y)$ are twice continuously differentiable and

$$\sup_{x,y\in[0,1]} \left| \frac{\partial^2}{\partial x^{\alpha_1} \partial y^{\alpha_2}} R_h(x,y) \right|$$

is uniformly bounded in h for all combinations of $\alpha_1, \alpha_2 \in \mathbb{N}_0$ where $\alpha_1 + \alpha_2 = 2$.

The following conditions will be required for the rates concerning spectral density estimation.

(D12) Assume the summability in the supremum norm of the 4-th order cumulant kernel of $\{X_t\}$,

$$\sum_{h_1,h_2,h_3=-\infty}^{\infty} \sup_{x_1,x_2,x_3,x_4\in[0,1]} \left|\mathrm{cum}(X_{h_1}, X_{h_2}, X_{h_3}, X_0)(x_1,x_2,x_3,x_4)\right| < \infty.$$

(D13) The Bartlett span parameter L satisfies

$$L = o\left(\left(\sqrt{\frac{\log T}{T B_R^2}} \right)^{-\frac{s-2}{s-1}} \right)$$

(D14) The bandwidth parameter B_R^2 satisfies

$$\frac{\log T}{T^{\theta_F} B_R} = o(1), \qquad T \to \infty,$$

with

$$\theta_F = \frac{\beta(s-2) - 4s + 4}{\beta(s-2)}$$

and

$$LB_R^2 = o(1).$$

We can now state the main consistency and convergence results under the strong mixing conditions.

Theorem 2.3.4. *Under the assumptions (A1), (D1) — (D7),*

$$\sup_{x\in[0,1]} \left|\hat{\mu}_X(x) - \mu_X(x)\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{TB_\mu}} + B_\mu^2\right).$$

For fixed $h \in \mathbb{Z}$, under the assumptions (A1), (D1) — (D11),

$$\sup_{x,y\in[0,1]} \left|\hat{R}_h(x,y) - R_h(x,y)\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{TB_R^2}} + B_R^2\right).$$

Theorem 2.3.5. *Under the assumption (A1), (D1) — (D9) and (D11) — (D14),*

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \left|\hat{f}_\omega(x,y) - f_\omega(x,y)\right| = o_{\mathbb{P}}(1) \tag{2.38}$$

and assuming further (B6),

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \left|\hat{f}_\omega(x,y) - f_\omega(x,y)\right| = O_{\mathbb{P}}\left(L\sqrt{\frac{\log T}{TB_R^2}} + LB_R^2 + \frac{1}{L}\right). \tag{2.39}$$

2.3.4 Functional data recovery and confidence bands

In this section we turn our attention to developing asymptotic theory for the recovered functional data and the associated confidence bands, in particular, the asymptotic behaviour of the plug-in estimator (2.21) vis-à-vis its theoretical counterpart (2.19).

First of all, we need to clarify what asymptotic result we can hope to accomplish. Before venturing into functional time series, let us comment on the asymptotic results for independent identically distributed functional data (Yao et al., 2005a). As the number of sparsely observed functional data grows to infinity, one can consistently estimate the second-order structure of the stochastic process (which in this case consists of the zero-lag autocovariance, due to independence). This is then used in the plug-in prediction of a given functional datum, say $X_s(\cdot)$, given the sparse measurements on this datum. In the limit, this prediction is as good as if we knew the true lag zero covariance of the stochastic process (Yao et al., 2005a, Thm 3). Because the predictor uses the estimate of the lag zero covariance based on all the observed data, Yao et al. (2005a) call this trait as *borrowing strength* from the entire sample.

In the time series setting of this thesis, one can expand the concept of borrowing strength from the entire sample. As the number of sparsely observed functional data (i.e. the time horizon T) grows to infinity, one can not only estimate the dynamics of the functional time series consistently (Theorem 2.3.2 and Corollary 2.3.3), but also further exploit the fact that neighbouring data are correlated to further improve the recovery. Because of the weak dependence, the influence of the observations decreases as we part away from the time s. Therefore we fix a span of times $1,\dots,S$ where $s < S \in \mathbb{N}$ and we will be interested in the prediction of X_s given the data in this span. To be precise, we are going to prove that the prediction of X_s from

the data in the local span and based on the estimated dynamics from complete data is, in the limit, as good as the prediction based on the true (unknown) dynamics. Therefore, in our case, we are *borrowing strength* across the sample in a twofold sense – firstly for the estimation of the functional time series dynamics, and then for prediction of the functional datum X_s.

The span S can in principle be chosen to be as large as one wishes, but is held fixed with respect to T. This is justified by the weak dependence assumption. In practice, one must also entertain numerical considerations and not choose S to be exceedingly large, since the evaluation of the predictors (2.19) and (2.21) based on longer spans requires the inversion of a big matrix.

We formulate Theorems 2.3.6 and 2.3.7 under the cumulant mixing conditions required for Theorems 2.3.1 and 2.3.2. Nevertheless, the conclusions also hold also under the strong mixing condition regime of Theorems 2.3.4 and 2.3.5 since, as is apparent from the proofs, the only requirement coming into play is the consistency of the spectral density operator estimators in the sense of (2.35) or (2.38).

Theorem 2.3.6. *Under the assumptions (A1), (B1) — (B5) and (B7) — (B10), for fixed $s \in \mathbb{N}$, $s < S$,*

$$\sup_{x\in[0,1]} \left|\hat{\Pi}(X_s(x), \mathbb{U}_S) - \Pi(X_s(x), \mathbb{U}_S)\right| = o_{\mathbb{P}}(1), \qquad as \quad T \to \infty.$$

In the following theorem we verify the asymptotic coverage probability of the pointwise and simultaneous confidence bands (2.28) and (2.31) under the Gaussian assumption (A2).

Theorem 2.3.7. *Under the assumptions (A1), (A2), (B1) — (B5) and (B7) — (B10), for fixed $s \in \mathbb{N}$, $s \leq S$:*

- *Asymptotic coverage of the pointwise confidence band for fixed $x \in [0,1]$:*

$$\lim_{T\to\infty} \mathbb{P}\left\{\left|\hat{\Pi}(X_s(x), \mathbb{U}_S) - X_s(x)\right| \leq \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\mathbb{S}}_{\mathbb{X}_T|\mathbb{U}_T}(x,x)}\right\} = 1-\alpha.$$

- *Asymptotic coverage of the simultaneous confidence band:*

$$\lim_{T\to\infty} \mathbb{P}\left\{\forall x \in [0,1] : \left|\hat{\Pi}(X_s(x), \mathbb{U}_S) - X_s(x)\right| \leq z_{\alpha,\hat{\rho}}\sqrt{\hat{\mathbb{S}}_{\mathbb{X}_T|\mathbb{U}_T}(x,x)}\right\} = 1-\alpha.$$

2.4 Estimation of cross-dependence between two time series

2.4.1 Two sparsely observed functional time series

In this subsection we consider two functional time series: the functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ in the function space $\mathscr{H}_1 = L^2([0,1],\mathbb{R})$ and the functional time series $\{Y_t(y) : y \in [0,1]\}_{t\in\mathbb{Z}}$ in the same the function space $\mathscr{H}_2 = L^2([0,1],\mathbb{R})$. An extension to an L^2 space

in a different domain is straightforward but to keep the presentation simple we present the methodology with $\mathscr{H}_1 = \mathscr{H}_2 = L^2([0,1],\mathbb{R})$. Both of the time series X and Y are considered to be observed only sparsely through the observation scheme

$$\begin{aligned} U_{tj} &= X_t(x_{tj}) + \epsilon^X_{tj}, \qquad j = 1,\dots,N^X_t, \quad t = 1,\dots,T, \\ V_{tj} &= Y_t(U_{tj}) + \epsilon^U_{tj}, \qquad j = 1,\dots,N^Y_t, \quad t = 1,\dots,T, \end{aligned} \tag{2.40}$$

where $\{x_{tj}\},\{U_{tj}\} \subset [0,1]$ are the measurement locations, $\{N^X_t\},\{N^Y_t\}$ are the measurement counts, and $\{\epsilon^X_{tj}\},\{\epsilon^Y_{tj}\}$ are independent identically distributed noise processes with mean zero and variances σ^2_X,σ^2_Y. We assume that both sampling regimes are independent. Specifically, the measurement locations $\{x_{tj}\},\{U_{tj}\}$, their counts $\{N^X_t\},\{N^Y_t\}$, and the noise processes $\{\epsilon^X_{tj}\},\{\epsilon^Y_{tj}\}$ are independent, and are moreover independent on the latent functional time series X and Y.

To start with, we estimate the mean functions μ_X and μ_Y of $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ and $\{Y_t(y) : y \in [0,1]\}_{t\in\mathbb{Z}}$, respectively, by the estimator (2.2). Then, define the "raw" lag-h covariances between X and Y by formula

$$G^{XY}_{h,t}(x_{t+h,j}, y_{tk}) = \big(U_{t+h,j} - \hat{\mu}_X(x_{t+h,j})\big)\big(V_{tj} - \hat{\mu}_Y(y_{tk})\big) \tag{2.41}$$

for $|h| < T$, $t = \max(1,1-h),\dots,\min(T,T-h)$, $j = 1,\dots,N^X_{t+h}$, $k = 1,\dots,N^Y_t$. The "raw" covariances (2.41) serve as a basis for non-parametric estimation of the lagged covariance kernels and the cross-spectral density kernels.

Estimation of lagged covariance kernel

Taking the conditional expectation over the "raw" covariances (2.41) reveals that

$$\mathbb{E}\Big[G^{XY}_{h,t}(x_{t+h,j}, y_{tk}) | x_{t+h,j}, y_{tk}\Big] \approx R^{ZX}_h(x_{t+h,j}, y_{tk}) = \operatorname{cov}\big(X_h(x_{t+h,j}), Y_0(y_{tk})\big)$$

where the approximation "$\approx$" come only from the fact that the "raw" covariances (2.41) are defined with the estimated value of $\hat{\mu}_X$ and $\hat{\mu}_Y$. Therefore we may estimate the lag-h covariance kernel between X and Y by applying the local-linear surface smoother over the three dimensional scatter-plot $\{(x_{t+h,j}, y_{tk}, G^{XY}_{h,t}(x_{t+h,j}, y_{tk})\}$. Concretely, for the lag $|h| < T$ we set $\hat{R}^{XY}_h(x,y) = \hat{c}^{(7)}_0$ where

$$\begin{aligned} \Big(\hat{c}^{(7)}_0, \hat{c}^{(7)}_1, \hat{c}^{(7)}_2\Big) = \underset{c^{(7)}_0, c^{(7)}_1, c^{(7)}_2}{\operatorname{argmin}} \sum_{t=\min(1,1-h)}^{\max(T,T-h)} \sum_{j=1}^{N_{t+h}} \sum_{k=1}^{N_t} K\left(\frac{x_{t+h,j}-x}{\tilde{B}_R}\right) K\left(\frac{y_{tk}-y}{\tilde{B}_R}\right) \\ \times \Big\{G^X_{h,t}(x_{t+h,j}, y_{tk}) - c^{(7)}_0 - c^{(7)}_1(x_{t+h,j}-x) - c^{(7)}_2(y_{tk}-y)\Big\}^2 \end{aligned}$$

Unlike for the estimation of the lag-0 autocovariance kernel R^X_0 by the smoother (2.4), we do not need to discard the diagonal "raw" covariances because of the independence of the noise

processes $\{\epsilon^X_{tj}\}, \{\epsilon^Y_{tj}\}$.

Estimation of cross-spectral density kernel

The "raw" covariances (2.41) can also serve for the estimation of the cross-spectral density kernels between the functional time series X and Y. We estimate the spectral density kernel at frequency $\omega \in [-\pi, \pi]$ by the local-linear surface-smoother applied to the "raw" covariances multiplied by complex exponentials, setting

$$\hat{f}_\omega(x,y) = \frac{\tilde{L}}{2\pi}\hat{c}_0^{(8)} \in \mathbb{C} \tag{2.42}$$

where $\hat{c}_0^{(8)}$ is obtained by minimizing the following weighted sum of squares

$$\begin{aligned}(\hat{c}_0^{(8)}, \hat{c}_1^{(8)}, \hat{c}_2^{(6)}) = \underset{(c_0^{(8)}, c_1^{(8)}, c_2^{(8)}) \in \mathbb{C}^3}{\arg\min} \sum_{h=-\tilde{L}}^{\tilde{L}} \frac{1}{T-|h|} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{j=1}^{N_{t+h}} \sum_{k=1}^{N_t} \Big| G^{XY}_{h,t}(x_{t+h,j}, y_{tk}) e^{-\mathrm{i}h\omega} - \\ - c_0^{(8)} - c_1^{(8)}(x_{t_{t+h,j}} - x) - c_2^{(8)}(y_{tk} - y)\Big|^2 W_h \frac{1}{\tilde{B}_R^2} K\left(\frac{x_{t_{t+h,j}} - x}{\tilde{B}_R}\right) K\left(\frac{y_{tk} - y}{\tilde{B}_R}\right). \end{aligned} \tag{2.43}$$

Again due to the independence of the noise processes $\{\epsilon^X_{tj}\}, \{\epsilon^Y_{tj}\}$, we do not need to remove the diagonal in the smoother (2.10). Consequently, the scaling constant $1/(T-|h|)$ in the formula (2.43) is also simplified as opposed to (2.10).

Proposition 2.4.1. *Assume that the conditions (B1) — (B3) hold for both X and Y, the condition (B4) with R^{XY}_h instead of R^X_h,*

$$\sum_{h_1,h_2,h_3=-\infty}^{\infty} \sup_{x_1,x_2,y_1,y_2\in[0,1]} \big|\mathrm{Cum}(X_{h_1}, X_{h_2}, Y_{h_3}, Y_0)(x_1,x_2,y_1,y_2)\big| < \infty, \tag{B5:XY}$$

and that the bandwidths satisfy (B7), (B8), and (B10). Then

$$\sup_{x,y\in[0,1]} |\hat{R}^{XY}_h(x,y) - R^{XY}_h(x,y)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_R^2} + B_R^2\right),$$

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \big|\hat{f}^X_\omega(x,y) - f^X_\omega(x,y)\big| = o_{\mathbb{P}}(1),$$

and assuming further (B6) with R^{XY}_h,

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \big|\hat{f}^X_\omega(x,y) - f^X_\omega(x,y)\big| = O_{\mathbb{P}}\left(\tilde{L}\frac{1}{\sqrt{T}}\frac{1}{\tilde{B}_R^2} + \tilde{L}B_R^2 + \frac{1}{\tilde{L}}\right).$$

Proof. The proof of this proposition is analogous to the proof of Theorems 2.3.1 and 2.3.2 and is therefore omitted. □

2.4.2 One sparsely observed functional time series and a univariate time series

Consider the functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ observed sparsely through the observation regime (2.40) and the univariate time series $Z = \{Z_t\}_{t\in\mathbb{Z}} \in \mathbb{R}$.

We define the "raw" covariances between $\{X_t\}$ and $\{Z_t\}$ as

$$G_{h,t}^{ZX}(x_{tj}) = \left(Z_{t+h} - \overline{Z}\right)\left(U_{tj} - \hat{\mu}_X(x_{tj})\right) \tag{2.44}$$

for $|h| < T$, $t = \min(1, 1-h), \ldots, \max(T, T-h)$, $j = 1, \ldots, N_{t+h}^X$, and where $\overline{Z} = (1/T)\sum_{t=1}^T Z_t$ is the classic univariate empirical mean. The "raw" covariances (2.44) serve again as a basis for non-parametric estimation, this time however we use local-liner line smoothers.

Estimation of lagged covariance kernels

Firstly, we start with the estimation the lag-h covariance kernel $R_h^{ZX}(x) = \mathrm{cov}(Z_t, X_0(x))$ for $x \in [0,1]$. This quantity is then estimated by local-linear line smoother defined by $\hat{R}_h^{Z^{(i)}X}(x) = \hat{c}_0^{(9)}$ where

$$\left(\hat{c}_0^{(9)}, \hat{c}_1^{(9)}\right) = \underset{c_0^{(9)}, c_1^{(9)}}{\arg\min} \sum_{t=\min(1,1-h)}^{\max(T,T-h)} \sum_{j=1}^{N_{t+h}} K\left(\frac{x_{tj}-x}{\tilde{B}_R}\right)\left\{G_{h,t}^{ZX}(x_{tj}) - c_0^{(9)} - c_1^{(9)}(x_{tj}-x)\right\}^2 \tag{2.45}$$

where $B_C > 0$ is a bandwidth parameter. When dealing with a multivariate time series Z as opposed to a univariate one, the smoother (2.45) is repeatedly applied for the all coordinates of the multivariate series to estimate the lag-h covariance kernel.

Estimation of cross-spectral density

The cross-spectral density between the univariate time series $\{Z_t\}$ and the sparsely observed function time series $\{X_t\}$ is estimated by setting

$$\hat{f}_\omega^{ZX}(x) = \frac{\tilde{L}}{2\pi}\hat{c}_0^{(10)} \qquad (\in \mathbb{C}) \tag{2.46}$$

where $\hat{c}_0^{(10)}$ is realised as the minimiser of the following weighted sum of squares

$$\left(\hat{c}_0^{(10)}, \hat{c}_1^{(10)}\right) = \underset{(c_0^{(10)}, c_1^{(10)})\in\mathbb{C}^2}{\arg\min} \sum_{h=-\tilde{L}}^{\tilde{L}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{j=1}^{N_t} W_h K\left(\frac{x_{tj}-x}{\tilde{B}_R}\right) \times \left|G_{h,t}^{ZX}(x_{tj})e^{-\mathrm{i}h\omega} - c_0^{(10)} - c_1^{(10)}(x_{tj}-x)\right|^2.$$

Again, when dealing with a multivariate time series $\{Z_t\}$, we apply (2.46) to each coordinate of $\{Z_t\}$ to estimate the cross-spectral density between $\{Z_t\}$ and $\{X_t\}$.

Proposition 2.4.2. *Assume that the conditions (B1) — (B3) hold for X, the condition (B4) with*

R_h^{ZX} instead of R_h^X,

$$\sum_{h_1,h_2,h_3=-\infty}^{\infty} \sup_{x_1,x_2\in[0,1]} \left|\mathrm{cum}(X_{h_1}(x_1), X_{h_2}(x_2), Z_{h_3}, Z_0)\right| < \infty, \tag{B5:ZX}$$

and that the bandwidths satisfy (B7) and

$$\tilde{B}_R \to 0, \quad T\tilde{B}_R^4 \to \infty, \tag{B8:1d}$$
$$\tilde{L} \to \infty, \quad \tilde{L} = o(\sqrt{T}\tilde{B}_R), \quad \tilde{L} = o(\tilde{B}_R^{-2}). \tag{B10:1d}$$

Then

$$\sup_{x\in[0,1]} |\hat{R}_h^{ZX}(x) - R_h^{ZX}(x)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}\tilde{B}_R} + \tilde{B}_R^2\right), \tag{2.47}$$
$$\sup_{\omega\in[-\pi,\pi]} \sup_{x\in[0,1]} \left|\hat{f}_\omega^{ZX}(x) - f_\omega^{ZX}(x)\right| = o_{\mathbb{P}}(1),$$

and assuming further (B6) with R_h^{ZX},

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \left|\hat{f}_\omega^{ZX}(x) - f_\omega^{ZX}(x)\right| = O_{\mathbb{P}}\left(\tilde{L}\frac{1}{\sqrt{T}}\frac{1}{\tilde{B}_R} + \tilde{L}\tilde{B}_R^2 + \frac{1}{\tilde{L}}\right). \tag{2.48}$$

Proof. The proof of this proposition uses the classical kernel smoothing techniques already established in the proofs of Theorems 2.3.1 and 2.3.2 or Proposition 2.4.1 except the line smoother is used in the time domain and the spectral domain, therefore modifying slightly the convergence rates (2.47) and (2.48). □

2.4.3 One sparsely and one fully observed functional time series

Finally, in this subsection we comment on the estimation method of the lagged covariance kernels and the cross-spectral density kernels between two functional time series, one sparsely observed and the other fully. Such situation is no exception in praxis, for example of the meteorological data at Wank mountain presented in Section 4.4 we analysed the interplay of the sparsely observed functional time series of the atmospheric electricity and the fully observed functional time series of daily temperature profiles.

The estimation of the cross dependence between the functional time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ observed sparsely (2.40) and the fully observed functional time series $\{Y_t(y) : y \in [0,1]\}_{t\in\mathbb{Z}}$ is no different from the analysis of the cross dependence between $\{X_t\}$ and the univariate time series $\{Y_t(y)\}$ for each value of $y \in [0,1]$.

We define the "raw" covariances

$$G_{h,t}^{XY}(x_{t+h,j}, y) = \big(U_{t+h,j} - \mu_X(x_{t+h,j})\big)\big(Y_t(y) - \hat{\mu}_Y(y)\big)$$

for $|h| < T$, $t = \min(1, 1-h), \ldots, \max(T, T-h)$, $j = 1, \ldots, N^X_{t+h}$, and where $\mu_Y = (1/T)\sum_{t=1}^{T} Y_t$ is the standard empirical functional mean.

Estimation of lagged covariance kernels

For fixed $y \in [0,1]$, the lag-h covariance kernel $R^{XY}_h(x,y) = \mathrm{cov}(X_h(x), Y_0(y))$ is estimated by the local-linear line smoother by $\hat{R}^{XY}_h(x,y) = \hat{c}^{(11)}_0$ where

$$\left(\hat{c}^{(11)}_0, \hat{c}^{(11)}_1\right) = \underset{c^{(11)}_0, c^{(11)}_1}{\arg\min} \sum_{t=\min(1,1-h)}^{\max(T,T-h)} \sum_{j=1}^{N_{t+h}} K\left(\frac{x_{t+h,j} - x}{\tilde{B}_R}\right) \left\{G^{XY}_{h,t}(x_{t+h,j}, y) - c^{(11)}_0 - c^{(11)}_1 (x_{t+h,j} - x)\right\}^2$$

with the bandwidth parameter $\tilde{B}_R > 0$.

Estimation of cross-spectral density kernel

The cross-spectral density between the sparsely observed functional time series $\{X_t\}$ and fully observed functional time series $\{Y_t\}$ is estimated by setting

$$\hat{f}^{XY}_\omega(x,y) = \frac{\tilde{L}}{2\pi} \hat{c}^{(12)}_0 \qquad (\in \mathbb{C})$$

where $\hat{c}^{(12)}_0$ is the minimiser of the following weighted sum of squares

$$\left(\hat{c}^{(12)}_0, \hat{c}^{(12)}_1\right) = \underset{(c^{(12)}_0, c^{(12)}_1) \in \mathbb{C}^2}{\arg\min} \sum_{h=-\tilde{L}}^{\tilde{L}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{j=1}^{N_t} W_h K\left(\frac{x_{t+h,j} - x}{\tilde{B}_R}\right) \times \left| G^{XY}_{h,t}(x_{t+h,j}, y) e^{-\mathrm{i}h\omega} - c^{(12)}_0 - c^{(12)}_1 (x_{t+h,j} - x) \right|^2.$$

Proposition 2.4.3. *Assume that the conditions (B1) — (B3) hold for X, the condition (B4) with R^{XY}_h instead of R^X_h, the condition* (B5:XY)*, and that the bandwidths satisfy (B7),* (B10:1d)*, and* (B10:1d)*. Then*

$$\sup_{x,y\in[0,1]} |\hat{R}^{XY}_h(x,y) - R^{XY}_h(x,y)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}\tilde{B}_R} + \tilde{B}^2_R\right),$$
$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \left|\hat{f}^{XY}_\omega(x,y) - f^{XY}_\omega(x,y)\right| = o_{\mathbb{P}}(1),$$

and assuming further (B6) with R^{XY}_h,

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \left|\hat{f}^{XY}_\omega(x,y) - f^{XY}_\omega(x,y)\right| = O_{\mathbb{P}}\left(\tilde{L}\frac{1}{\sqrt{T}}\frac{1}{\tilde{B}_R} + \tilde{L}\tilde{B}^2_R + \frac{1}{\tilde{L}}\right).$$

Proof. The proof of this proposition is once again based on the classical kernel smoothing proof techniques already established in the proofs of Theorems 2.3.1 and 2.3.2 or Proposi-

tion 2.4.1 and 2.4.2, and is therefore omitted. □

2.5 Implementation details

2.5.1 Selection of bandwidths B_μ, B_R, and B_V

Our estimation methodology involves three bandwidth parameters B_μ, B_R, B_V that need to be selected based on some data-driven criterion. To reduce the computational cost we choose to perform the selection of the parameters in successive fashion.

The selection of a bandwidth parameter in kernel smoothing has been extensively studied in literature for the case of locally polynomial regression. The classical selector by Ruppert et al. (1995) calculates the asymptotic mean square error and plugs-in some estimated quantities. However, their methodology applies to the independent case which is distinctly different from the setting of this thesis and hence we opt for a cross-validation selection procedure. The selection of the smoothing parameters by cross-validation has already been implemented by Yao et al. (2005a). Here we use a similar approach.

To further reduce the computational requirements we opt for a K-fold cross-validation strategy instead of the leave-one-curve-out cross-validation originally suggested by Rice and Silverman (1991). For the K-fold cross-validation, we work with $K = 10$ partitions, as follows. We randomly split the functional curves into K partitions and denote the time indices sets as $\mathcal{T}_1, \ldots, \mathcal{T}_K$. For each $k \in \{1, \ldots, K\}$, denote $\hat{\mu}^{(-k),B_\mu^0}$ the estimate of the common mean function μ calculated by the smoother (2.2) from data without the partition k and using the candidate smoothing parameter B_μ^0. We select the smoothing parameter B_μ by minimizing the following loss:

$$B_\mu = \underset{B_\mu^0}{\arg\min} \frac{1}{K} \sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \sum_{j=1}^{N_t} \left\{ U_{tj} - \hat{\mu}^{(-k),B_\mu^0}(x_{tj}) \right\}^2. \tag{2.49}$$

Once the smoothing parameter B_μ is chosen we estimate the function $\hat{\mu}$ from all data and use it in the second step to select B_R and B_V for smoothing the covariance kernels. We choose these smoothing parameters only while smoothing the lag-zero covariance. The reason behind this is that we expect the same smoothness for higher order lags and the selection of the parameters on only one covariance kernel reduces the computational cost, which would otherwise become substantial. We again employ K-fold cross-validation. Denote $\hat{R}_0^{(-k),B_R^0}$ the estimate of R_0 obtained by the smoother (2.4) calculated from the data without the partition k and using the candidate smoothing parameter B_R^0. The smoothing parameters B_R is selected by minimizing the following loss:

$$B_R = \underset{B_R^0}{\arg\min} \frac{1}{K} \sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \sum_{i,j=1}^{N_t} \left\{ \left(U_{ti} - \hat{\mu}(x_{ti})\right)\left(U_{tj} - \hat{\mu}(x_{tj})\right) - \hat{R}_0^{(-k),B_R^0}(x_{ti}, x_{tj}) \right\}^2. \tag{2.50}$$

To select the smoothing parameter B_V, we denote $\hat{V}^{(-k),B_V^0}$ the estimate of the diagonal of $R_0(\cdot,\cdot)$ including the ridge contamination, from the data except the partition k and using the candidate smoothing parameter B_V^0. The parameter B_V is selected by minimizing the following loss:

$$B_V = \underset{B_V^0}{\arg\min} \frac{1}{K} \sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \sum_{i=1}^{N_t} \left\{ \left(U_{ti} - \hat{\mu}(x_{ti})\right)^2 - \hat{V}(x_{ti})^{(-k),B_V^0} \right\}^2 \tag{2.51}$$

Once the minimizers B_R and B_V have been found, we construct the estimate of the lag-zero covariance kernel $\hat{R}_0$ and the measurement error $\widehat{\sigma^2}$ from the full data. The bandwidth parameter B_R will be used for estimation of the spectral density because we expect the same degree of spatial smoothness for spectral density kernels over all frequencies.

To numerically solve the optimization problems (2.49), (2.50), and (2.51) we use Matlab's implementation of the Bayesian optimisation algorithm (BayesOpt). A review of this algorithm can be found for example in Mockus (2012).

2.5.2 Selection of Bartlett's span parameter L

The selection of the parameter L, i.e. the number of lags taken into account when estimating the dynamics, is a challenging problem in general. Selection rules for the bandwidth parameter for smoothing in the frequency domain, which is equivalent to Bartlett's estimate as explained in Subsection 2.1.3, is reviewed in Fan and Yao (2008) for the case of one-dimensional time-series. The selection of the parameter L, or equivalently the bandwidth parameter for frequency domain smoothing, has nevertheless not been explored for the case of functional time-series. Neither Panaretos and Tavakoli (2013b) nor Hörmann et al. (2015a) provide data-dependent criteria, but instead rely on a prior choices based on asymptotic considerations.

The selection of the tuning parameter L is better studied in a related problem — the estimation of the long-run covariance, which is in fact the value of the spectral density at frequency $\omega = 0$. The long-run covariance can be estimated by the Bartlett's formula (1.78) for frequency $\omega = 0$. Data adaptive selection procedures for the tuning parameter L have been suggested in this context by Rice and Shang (2017) and Horváth et al. (2016).

However, it is unclear how to incorporate the sparse sampling scheme to the above-cited rules. To address this issue, we run a number of numerical experiments, simulating datasets from a couple of smooth functional time-series, and estimating the spectral density with a varying value of the parameter L. By investigating the estimation error, we propose guidelines on selecting L in the form of a rule of thumb:

$$L = \lfloor T^{1/3} \left(\bar{N}\right)^{1/4} \rfloor \tag{2.52}$$

where $\bar{N}$ is the average number of measurements per curve and $\lfloor\cdot\rfloor$ is the integer part of a given real number. The selection rule (2.52) was hand-picked for the considered range of variables T and N^{max} and should not be used for extrapolation, especially not for dense observation schemes.

2.6 Possible extensions and future directions

2.6.1 Functional data in other Hilbert spaces

The underlying assumption under the functional data considered in this thesis is the probabilistic model of $L^2([0,1],\mathbb{R})$-valued random elements, i.e. the sparse observation are realised on a latent one-dimensional curve. In the following paragraphs we discuss the extensions to $\mathscr{H}$-valued random elements.

The first extension worth exploring is to the multivariate functional data. Specifically, surface valued data are probabilistically modelled as random elements in $L^2([0,1]^2)$, i.e. we consider the time series of two dimensional surfaces $\{X_t(x,y) : x,y \in [0,1]\}_{t\in\mathbb{Z}}$. It is important to say that the statistical methods for surfaced-valued time series are not even well developed in the fully observed case, let alone for sparsely valued data. We make here a few comments on how the statistical methods could work on sparsely observed surface-valued data sampled by the protocol

$$U_{tij} = X_t(x_{ti}, y_{tj}) + \epsilon_{tij}, \qquad i = 1,\dots,N_t, \quad j = 1,\dots,N_t, \quad t = 1,\dots,T,$$

where $(x_{ti}, y_{tj}) \subset [0,1]^2$ are random locations, N_t their count at time t (itself a random variable), and $\{\epsilon_{tij}\}$ is an ensemble of independent identically distributed random noise contamination. The estimators of the first and second order dynamics follows the general recipe: pool all the data together and smooth the scatter-plots. For example, to estimate the mean surface $\mu(x,y) = \mathbb{E}X_t(x,y)$ one would want to consider the three dimensional scatter-plot consisting of triples $(x_{ti}, y_{tj}, U_{tij})$ and performing a local linear surface smoother. The estimation of the second order dynamics would rely on considering the scatter plots of the vectors

$$\left(x_{t+h,i}, y_{t+h,j}, x_{ti'}, y_{tj'}, \left(U_{t+h,i,j} - \hat{\mu}(x_{t+h,i}, y_{t+h,j})\right)\left(U_{ti'j'} - \hat{\mu}(x_{ti'}, y_{tj'})\right)\right)$$

and performing a four-dimensional local regression smoother. However, the non-parametric regression methods are known to be very data-hungry in higher dimensions and this approach, although working without limitations in theory, is likely to turn out infeasible in practice. Some assumptions on the special structure might turn out to be useful, such as separability of the two spacial dimensions, or stationarity and isotropy. The generalisation of the smoothing methods to use these special structures of covariance kernel retain the same general approach: pool all the data together and construct an appropriate scatter plot (according to the assumed special structure or dimensionality of the data) and perform a local polynomial smoother.

The presented theory cannot, in general, be extended to any abstract separable Hilbert space $\mathscr{H}$. Indeed, the sparse sampling protocol (2.1) requires the notion of point-wise evaluation. This mapping could be seen as a projection into a given fixed direction for an abstract Hilbert space, i.e the data would be sampled as $U_{tj} = \langle X_t, g_{tj} \rangle + \epsilon_{tj}$ where $g_{tj} \in \mathscr{H}$ for $t = 1, \ldots, T$ and $j = 1, \ldots, N_t$, and with $\{\epsilon_{tj}\}$ being an additive noise. A non-parametric statistical estimator would then necessitate a notion of "closeness" of the projection directions $\{g_{tj}\}$ in the same way as the point-wise evaluation are close to each other if the evaluation points are close. We conclude that this sort of generalisation is likely to demand a special structure of the Hilbert space $\mathscr{H}$ in the given application.

2.6.2 Weighted observations

The sparse sampling protocol (2.1) provides a data set composed of the data points $\{U_{tj}\}$ each treated essentially as carrying the same information. It may very well happen, though, that we have some prior information on how each observation is reliable, say, each observation U_{tj} is accompanied by the weight $w_{tj} > 0$. This weight can be assign by the domain expertise and knowledge of the sampling mechanism.

The smoothing methods derived in this chapter can be easily adapted to this set-up. For example, the mean function estimator $\hat{\mu}_X$ defined in (2.2) can be generalised to:

$$\underset{c_0^{(1)}, c_1^{(1)}}{\arg\min} \sum_{t=1}^{T} \sum_{j=1}^{N_i} w_{tj} K\left(\frac{x - x_{ij}}{B_\mu}\right) \left\{U_{tj} - c_0^{(1)} - c_1^{(1)}(x - x_{tj})\right\}^2.$$

The surface smoothers using the "raw" covariances for the estimation of lagged autocovariance kernels or spectral density kernels require cross-products between a pair observations (after centring), say U_{tj} and $U_{t'k}$. In this case, we conjecture that the product of weights $w_{tj} w_{t'k}$ should be used in the local polynomial surface smoothers.

A special case of this weighting scheme is to set up the weights $w_{tj} = 1/N_t$ where N_t is the total number of observation locations at time $t = 1, \ldots, T$. This weighting scheme stems from the fact that the more points are sampled at a given curve the less information is carried per point due within curve correlation. This phenomenon was already observed by Li and Hsing (2010) who proposed essentially the aforementioned weighting scheme for the mean function and covariance kernel estimation in the independent identically distributed regime. As a consequence, the estimators with this weights are robust to hybrid data sets composed of some curves sampled sparsely and some densely.

2.7 Proofs of formal statements

The proofs of the formal statements presented in this chapter that require longer proofs are written down in this section.

2.7.1 Proof of Theorem 2.3.1

We start with the smoother for the common mean function $\mu_X(\cdot)$. Its estimator $\hat{\mu}_X(x)$, the minimizer of (2.2), explicitly:

$$\hat{\mu}_X(x) = \frac{Q_0 S_2 - Q_1 S_1}{S_0 S_2 - S_1^2}, \tag{2.53}$$

where

$$S_r = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N_t}\left(\frac{x_{tj}-x}{B_\mu}\right)^r \frac{1}{B_\mu}K\left(\frac{x_{tj}-x}{B_\mu}\right), \qquad r = 0,1,2,$$

$$Q_r = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N_t}\left(\frac{x_{tj}-x}{B_\mu}\right)^r U_{tj}\frac{1}{B_\mu}K\left(\frac{x_{tj}-x}{B_\mu}\right), \qquad r = 0,1.$$

All of the above quantities are functions of $x \in [0,1]$ and all of the operations are to be understood in the pointwise sense, and this includes the division operation. In Lemma 2.7.1 and Lemma 2.7.2 we determine the asymptotic behaviour of S_r and Q_r, respectively.

Lemma 2.7.1. *Under (A1), (B1), (B2) and (B7), for* $r = 0,1,2$

$$\sup_{x\in[0,1]}\left|S_r - M_{[S_r]}\right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_\mu} + B_\mu^2\right)$$

where $M_{[S_0]} = \mathbb{E}[N]\, g(x), M_{[S_1]} = 0, M_{[S_2]} = \mathbb{E}[N]\,\sigma_K^2 g(x)$ *and* $\sigma_K^2 = \int v^2 K(v)\,\mathrm{d}v$.

Proof. We have the following decomposition

$$\mathbb{E}\left[\sup_{x\in[0,1]}\left|S_r - M_{[S_r]}\right|\right] \le \sup_{x\in[0,1]}\left|\mathbb{E}[S_r] - M_{[S_r]}\right| + \mathbb{E}\left[\sup_{x\in[0,1]}\left|\mathbb{E}[S_r] - S_r\right|\right].$$

For the bias term, by using the Taylor expansion to order 2 it is straightforward to show the formulae for $M_{[S_r]}, r = 0,1,2$ as well as that $\mathbb{E}[S_r] = M_{[S_r]} + O(B_\mu^2)$ where the remainder of the Taylor expansion is uniform in $x \in [0,1]$. Hence

$$\sup_{x\in[0,1]}\left|\mathbb{E}[S_r] - M_{[S_r]}\right| = O(B_\mu^2). \tag{2.54}$$

For the stochastic term, it will be useful to employ the Fourier transform. The inverse Furrier transform of the function $u \mapsto K(u)u^r$ is defined as $\zeta_r(t) = \int e^{-\mathrm{i}ut}K(u)u^r\,\mathrm{d}u$. Therefore we may write

$$w_{tj}\left(\frac{x_{tj}-x}{B_\mu}\right)^r = \frac{1}{2\pi B_\mu}\int e^{\mathrm{i}u(x_{tj}-x)/hB_\mu}\left(\frac{x_{tj}-x}{B_\mu}\right)^r \zeta_r(u)\,\mathrm{d}u =$$

$$= \frac{1}{2\pi}\int e^{\mathrm{i}v(x_{tj}-x)}(x_{tj}-x)^r\zeta_r(vB_\mu)\,\mathrm{d}v.$$

Define

$$\phi_r(v) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N_t} e^{\mathrm{i}\,v x_{tj}}(x_{tj}-x)^r, \tag{2.55}$$

and thus we can write

$$S_r(x) = \frac{1}{2\pi}\int \phi_r(v)e^{-\mathrm{i}\,xv}\zeta_r(vB_\mu)\,\mathrm{d}v.$$

Thanks to the independence of $\{N_t\}$ and $\{x_{tj}\}$ and stationarity we can bound the variance of $\phi_{S_r}(x)$

$$\begin{aligned}\operatorname{var}(\phi_{S_r}(x)) &\le \frac{1}{T}\operatorname{var}\left\{\sum_{j=1}^{N_1} e^{\mathrm{i}\,v x_{1j}(x_{1j}-x)^r}\right\} \le \frac{1}{T}\mathbb{E}\left[\mathbb{E}\left[\left\{\sum_{j=1}^{N_1} e^{\mathrm{i}\,v x_{1j}}(x_{1j}-x)^r\right\}^2 \mid N_1\right]\right] \le \\ &\le \frac{1}{T}\mathbb{E}\left[\mathbb{E}\left[\left\{\sum_{j=1}^{N_1}\left|e^{\mathrm{i}\,v x_{1j}}\right|^2\right\}\left\{\sum_{j=1}^{N_1}(x_{1j}-x)^{2r}\right\} \mid N_1\right]\right] \le \frac{1}{T}E\left[\mathbb{E}[N]\,\mathbb{E}\left\{\sum_{j=1}^{N_1}(x_{1j}-x)^{2r}\right\} \mid N_1\right] \le \\ &\le \frac{\mathbb{E}[N]}{T}\mathbb{E}\left[(x_{11}-x)^{2r}\right] \le \frac{\mathbb{E}[N]}{T}.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}\left\{\sup_x |S_r(x)-\mathbb{E}[S_r(x)]|\right\} &\le \frac{1}{2\pi}\int \mathbb{E}\left[\left|\phi_r(v)-\mathbb{E}\phi_{S_r}(v)\right|\right]|\zeta_r(vB_\mu)|\,\mathrm{d}v \le \\ &\le \frac{1}{2\pi}\int \sqrt{\operatorname{var}(\phi_r(x))}|\zeta_r(vB_\mu)|\,\mathrm{d}v \le \frac{\int|\zeta_r(u)|\,\mathrm{d}u}{2\pi}\frac{\mathbb{E}[N]}{\sqrt{T}B_\mu} = O\left(\frac{1}{\sqrt{T}B_\mu}\right).\end{aligned} \tag{2.56}$$

The proof is concluded by combining (2.54) and (2.56), and by the fact that $\mathbb{E}[|Z_n|] = O(a_n)$ implies $Z_n = O_{\mathbb{P}}(a_n)$ for a sequences of random variables $\{Z_n\}$ and of constants $\{a_n\}$. □

Lemma 2.7.2. *Under (A1), (B1) — (B3) and (B7), for* $r = 0, 1$

$$\sup_{x\in[0,1]}\left|Q_r - M_{[Q_r]}\right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_\mu} + B_\mu^2\right)$$

where $M_{[Q_0]} = \mathbb{E}[N]\,\mu_X(x)g(x)$ *and* $M_{[Q_1]} = 0$.

Proof. The proof of Lemma 2.7.2 follows the same ideas as that of Lemma 2.7.1. We use the bias variance decomposition and a Taylor expansion to order 2 to derive the analogous results as in (2.54) as well as the formulae for $M_{[Q_0]}(x)$ and $M_{[Q_1]}(x)$. We then define

$$\varphi_r(v) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N_t} e^{\mathrm{i}\,v x_{tj}}(x_{tj}-x)^r U_{tj} \tag{2.57}$$

in analogy to (2.55). Thus we can write

$$Q_r(x) = \frac{1}{2\pi}\int \varphi_r(v)e^{-\mathrm{i}\,xv}\zeta_r(vB_\mu)\,\mathrm{d}v.$$

It remains to bound the variance of (2.57). However, the temporal dependence among U_{tj} must be now taken into account. First of all remark that for an arbitrary stationary time-series $\{Z_t\}$ with a summable autocovariance function $\rho_Z(\cdot)$, one has:

$$\operatorname{var}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\right) = \frac{1}{T}\sum_{h=-T+1}^{T-1} \rho_Z(h)\left(1-\frac{|h|}{T}\right) \leq \frac{1}{T}\sum_{h=-\infty}^{\infty} |\rho_Z(h)|. \tag{2.58}$$

Define $Z_t = \sum_{j=1}^{N_t} e^{\mathrm{i}\nu x_{tj}}(x_{tj}-x)^r U_{tj}$ which constitutes a real-valued stationary time series. By conditioning on N_t and x_{tj}, and applying the law of total covariance, we can bound the autocovariance of $\{Z_t\}$ by $|\rho_Z(h)| \leq \max_{x,y}|R_h(x,y)|$ for $h \neq 0$. For $h = 0$, the bound is augmented by σ^2 due to the measurement error but this does not change the summability. The autocovariance function is summable by the assumption (A1) and we conclude that $\operatorname{var}\varphi_r(\nu) = O(1/T)$. The proof is completed by repeating the same steps as in (2.56):

$$E\left\{\sup_{x\in[0,1]} |S_r(x) - \mathbb{E}[S_r(x)]|\right\} = O\left(\frac{1}{\sqrt{T}B_\mu}\right).$$

□

Proof of the first part of Theorem 2.3.1. By combining Lemma 2.7.1, Lemma 2.7.2, the formula (2.53), and the uniform version of Slutsky's theorem, we obtain the rate (2.32). □

Now we turn our attention to the estimation of the lag-0 covariance and lag-h autocovariance kernels. We include the proof only for $h \neq 0$. For $h = 0$ one has to exclude the diagonal to evade the measurement errors but the proof is essentially the same. It is possible to explicitly express the minimizer to (2.8) (cf. Li and Hsing (2010)). The general principles of the explicit formula deviation are also commented on in Section 2.7.2, which uses similar deviation steps as the estimator of lagged autocovariance kernels. The explicit formula yields

$$\hat{R}_h(x,y) = \left(\mathscr{A}_1^{(h)}Q_{00}^{(h)} - \mathscr{A}_2^{(h)}Q_{10}^{(h)} - \mathscr{A}_3^{(h)}Q_{01}^{(h)}\right)\left(\mathscr{B}^{(h)}\right)^{-1}, \tag{2.59}$$

where $|h| < T$ and

$$\mathscr{A}_1^{(h)} = S_{20}^{(h)}S_{02}^{(h)} - \left(S_{11}^{(h)}\right)^2, \qquad \mathscr{A}_2^{(h)} = S_{10}^{(h)}S_{02}^{(h)} - S_{01}^{(h)}S_{11}^{(h)},$$
$$\mathscr{A}_3^{(h)} = S_{01}^{(h)}S_{20}^{(h)} - S_{10}^{(h)}S_{11}^{(h)}, \qquad \mathscr{B}^{(h)} = \mathscr{A}_1^{(h)}S_{00}^{(h)} - \mathscr{A}_2^{(h)}S_{10}^{(h)} - \mathscr{A}_3^{(h)}S_{01}^{(h)},$$

$$S_{pq}^{(h)} = \frac{1}{T-|h|}\sum_{t=\max(1,1-h)}^{\max(T,T-h)} \sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t} \left(\frac{x_{t+h,j}-x}{B_R}\right)^p\left(\frac{x_{tk}-y}{B_R}\right)^q \frac{1}{B_R^2}K\left(\frac{x_{t+h,j}-x}{B_R}\right)K\left(\frac{x_{tk}-y}{B_R}\right),$$

$$Q_{pq}^{(h)} = \frac{1}{T-|h|}\sum_{t=\max(1,1-h)}^{\max(T,T-h)} \sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t} G_{h,t}^X(x_{t+h,j}, x_{tk})$$

$$\times \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q \frac{1}{B_R^2} K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right).$$

All of the above terms are functions of $(x,y) \in [0,1]^2$ and all operations are understood the pointwise sense, including the pointwise inversion of $\left(\mathscr{B}^{(h)}\right)^{-1} = \left(\mathscr{B}^{(h)}(x,y)\right)^{-1}$.

We assess the uniform asymptotic behaviour of $S_{pq}^{(h)}$ and $Q_{pq}^{(h)}$ in Lemma 2.7.3 and Lemma 2.7.4.

Lemma 2.7.3. *Under (A1), (B1), (B2), (B7) and (B8),*

$$\mathbb{E}\left[\sup_{x,y\in[0,1]} \left|S_{pq}^{(h)} - \mathbb{E}S_{pq}^{(h)}\right|\right] \leq U \frac{1}{\sqrt{T-|h|}} \frac{1}{B_R^2} \tag{2.60}$$

$$\sup_{x,y\in[0,1]} \left|\mathbb{E}S_{pq}^{(h)} - M_{[S_{pq}]}\right| = O\left(B_R^2\right) \tag{2.61}$$

where the constant U is independent of $0 \leq p+q \leq 2$, $T \in \mathbb{N}$, $|h| < T$, and

$$\begin{aligned} M_{[S_{00}^{(h)}]} &= c_h g(x) g(y), & M_{[S_{01}^{(h)}]} &= M_{[S_{10}^{(h)}]} = M_{[S_{11}^{(h)}]} = 0, \\ M_{[S_{20}^{(h)}]} &= M_{[S_{02}^{(h)}]} = c_h g(x) g(y) \sigma_K^2, & \sigma_K^2 &= \int v^2 K(v)\,\mathrm{d}v, \end{aligned} \tag{2.62}$$

where $c_h = (\mathbb{E}N)^2$ for $h \neq 0$ and $c_0 = \mathbb{E}\{N(N-1)\}$. Furthermore, the convergence (2.61) *is uniform in h.*

Proof. We decompose the estimation error:

$$\mathbb{E}\left[\sup_{x,y\in[0,1]} \left|S_{pq}^{(h)} - M_{[S_{pq}^{(h)}]}\right|\right] \leq \mathbb{E}\left[\sup_{x,y\in[0,1]} \left|S_{pq}^{(h)} - \mathbb{E}\left[S_{pq}^{(h)}\right]\right|\right] + \sup_{x,y\in[0,1]} \left|\mathbb{E}\left[S_{pq}^{(h)}\right] - M_{[S_{pq}^{(h)}]}\right|. \tag{2.63}$$

Considering a Taylor expansion of order 2, it is easy to show that the formulae (2.62) and that the second term of (2.60) is of order $O(B_R^2)$ uniformly in h and T.

Taking the analogous steps as in the proof of Lemma 2.7.1 while using the Fourier transform of the function $(u,v) \mapsto K(u)K(v)u^p v^q$, one can prove that the first term on the right-hand side of (2.63) are bounded by $1/(T-|h|)$. □

Now assume that the common mean function $\mu_X(\cdot)$ is known for the moment. Thus formally define

$$\tilde{Q}_{pq}^{(h)} = \frac{1}{T-|h|} \sum_{t=\max(1,1-h)}^{\max(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \tilde{G}_{h,t}^X(x_{t+h,j}, x_{tk}) \times \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q \frac{1}{B_R^2} K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right)$$

where

$$\tilde{G}^X_{h,t}(x_{t+h,j}, x_{tk}) = (U_{t+h,j} - \mu_X(x_{t+h,j}))(U_{tk} - \mu_X(x_{tk})). \tag{2.64}$$

We analyse the asymptotics of $\tilde{Q}^{(h)}_{pq}$ in Lemma 2.7.4.

Lemma 2.7.4. *Under (A1), (B1) — (B5) and (B8)*

$$\mathbb{E}\left[\sup_{x,y\in[0,1]}\left|\tilde{Q}^{(h)}_{pq} - \mathbb{E}\tilde{Q}^{(h)}_{pq}\right|\right] \leq U\frac{1}{\sqrt{T-|h|}}\frac{1}{B_R^2}, \tag{2.65}$$

$$\sup_{x,y\in[0,1]}\left|\mathbb{E}\tilde{Q}^{(h)}_{pq} - M_{[Q^{(h)}_{pq}]}\right| = O\left(B_R^2\right) \tag{2.66}$$

where the constant U is uniform for $0 \leq p+q \leq 2$, $T \in \mathbb{N}$, $|h| < T$, and

$$M_{[Q^{(h)}_{00}]} = c_h R_h(x,y)g(x)g(y), \qquad M_{[Q^{(h)}_{01}]} = M_{[Q^{(h)}_{10}]} = 0, \tag{2.67}$$

where $c_h = (\mathbb{E}N)^2$ for $h \neq 0$ and $c_0 = \mathbb{E}\{N(N-1)\}$. Moreover, the convergence (2.66) *is uniform in h.*

Proof. Again, the bias-variance decomposition yields

$$\mathbb{E}\left[\sup_{x,y\in[0,1]}\left|\tilde{Q}^{(h)}_{pq} - M_{[Q^{(h)}_{pq}]}\right|\right] \leq \mathbb{E}\left[\sup_{x,y\in[0,1]}\left|\tilde{Q}^{(h)}_{pq} - \mathbb{E}\left[\tilde{Q}^{(h)}_{pq}\right]\right|\right] + \sup_{x,y\in[0,1]}\left|\mathbb{E}\left[\tilde{Q}^{(h)}_{pq}\right] - M_{[Q^{(h)}_{pq}]}\right|$$

By taking a Taylor expansion of order 2, it is again straightforward to show that the formulae (2.67) and that the second term of (2.65) is of order $O(B_R^2)$ uniformly in h and T.

To treat the first term on the right-hand side of (2.65), we define the Fourier transform of the function $(\alpha,\beta) \mapsto K(\alpha)\alpha K(\beta)\beta$ as

$$\zeta_{pq}(u,v) = \iint e^{-\mathrm{i}(u\alpha+v\beta)}K(\alpha)\alpha^p K(\beta)\beta^q d\alpha d\beta.$$

Thus we may write

$$\begin{aligned}\left(\frac{x_{t+h,j}-x}{B_R}\right)^p\left(\frac{x_{tk}-y}{B_R}\right)^q\frac{1}{B_R^2}K\left(\frac{x_{t+h,j}-x}{B_R}\right)K\left(\frac{x_{tk}-y}{B_R}\right) &= \\ = \frac{1}{(2\pi)^2 B_R^2}\iint \exp\left\{\mathrm{i}\left(\frac{x_{t+h,j}-x}{B_R}\right)u\right\}\exp\left\{\mathrm{i}\left(\frac{x_{tk}-y}{B_R}\right)v\right\}\zeta_{pq}(u,v)\,\mathrm{d}u\,\mathrm{d}v &= \\ = \frac{1}{(2\pi)^2}\iint e^{\mathrm{i}(x_{t+h,j}-y)\tilde{u}}e^{\mathrm{i}(x_{tk}-y)\tilde{v}}\zeta_{pq}(B_R\tilde{u},B_R\tilde{v})\,d\tilde{u}d\tilde{v}&\end{aligned}$$

Define

$$\varphi^{(h)}_{pq} = \varphi^{(h)}_{pq}(u,v,x,y) = \frac{1}{T-|h|}\sum_{t=\max(1,1-h)}^{\max(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t} e^{\mathrm{i}(x_{t+h,j}-x)u}e^{\mathrm{i}(x_{tk}-y)v}\tilde{G}^X_{h,t}(x_{t+h,j},x_{tk})$$

and write

$$\tilde{Q}_{pq}^{(h)} = \frac{1}{(2\pi)^2} \iint \varphi_{pq}^{(h)} \zeta_{pq}(B_R u, B_R v)\, \mathrm{d}u\, \mathrm{d}v.$$

Analogously to (2.56), it now remains to analyse the variance of $\varphi_{pq}^{(h)}$. Define the following stationary time-series

$$Z_t^{(h)} = \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} e^{\mathrm{i}(x_{t+h,j}-x)u} e^{\mathrm{i}(x_{tk}-y)v} \tilde{G}_{h,t}^X(x_{t+h,j}, x_{tk}).$$

As in the proof of Lemma 2.7.2 we want to bound the sum of the autocovariance function $\sum_{\xi\in\mathbb{Z}} |\rho_{Z^{(h)}}(\xi)|$ but the bound must be uniform in h. By conditioning on N_t and x_{tj}, and applying the law of total covariance, the ξ-lag autocovariance $\rho_{Z^{(h)}}(\xi)$ can be bounded by

$$\begin{aligned}
\left|\rho_{Z^{(h)}}(\xi)\right| &= \left|\operatorname{cov}(Z_{t+\xi}, Z_t)\right| \leq \\
&\leq (\mathbb{E}[N])^2 \sup_{x_1,x_2,x_3,x_4\in[0,1]} \Big| \operatorname{cov}\Big\{ (X_{t+\xi+h}(x_1) - \mu_X(x_1))(X_{t+\xi}(x_2) - \mu_X(x_2)), \\
&\qquad\qquad (X_{t+h}(x_3) - \mu_X(x_3))(X_t(x_4) - \mu_X(x_4)) \Big\} \Big| = \\
&= (\mathbb{E}[N])^2 \sup_{x_1,x_2,x_3,x_4\in[0,1]} \Big| \operatorname{cov}\Big\{ (X_{\xi+h}(x_1) - \mu_X(x_1))(X_{\xi}(x_2) - \mu_X(x_2)), \\
&\qquad\qquad (X_{h}(x_3) - \mu_X(x_3))(X_0(x_4) - \mu_X(x_4)) \Big\} \Big| \qquad (2.68)
\end{aligned}$$

for $\xi \notin \{-h, 0, h\}$. For $\xi \in \{-h, 0, h\}$, the bound is augmented by σ^2 but this changes nothing as to the summability with respect to $\xi \in \mathbb{Z}$.

Using the formula for the 4-th order cumulant of centred random variables (Rosenblatt, 2012, p. 36), we express the covariance on the right-hand side of (2.68) as

$$\begin{aligned}
&\operatorname{cov}\Big((X_{\xi+h}(x_1) - \mu_X(x_1))(X_\xi(x_2) - \mu_X(x_2)), (X_h(x_3) - \mu_X(x_1))(X_0(x_4) - \mu_X(x_1)) \Big) = \\
&\quad = \operatorname{cum}\big(X_{\xi+h}(x_1) - \mu_X(x_1), X_\xi(x_2) - \mu_X(x_2), X_h(x_3) - \mu_X(x_3), X_0(x_4) - \mu_X(x_4), \big) + \\
&\qquad + R_\xi(x_1, x_3) R_\xi(x_2, x_4) + R_{\xi+h}(x_1, x_4) R_{\xi-h}(x_2, x_3). \qquad (2.69)
\end{aligned}$$

Taking the absolute value and the supremum, the sum of (2.68) over $\xi \in \mathbb{Z}$ is bounded thanks to the fact that the cumulant on the right-hand side of (2.69) is summable by (B5) and the autocovariances are summable by (1.25). Moreover the sum is bounded uniformly in $h \in \mathbb{Z}$. Therefore

$$\operatorname{var}\Big(Q_{pq}^{(h)}\Big) \leq \frac{1}{T-h} \sum_{\xi\in\mathbb{Z}} |\rho_{Z^{(h)}}(\xi)| \leq U \frac{1}{T-h},$$

where the constant U is independent of h. Observing that $\iint \zeta_{pq}(B_R u, B_R v)\, \mathrm{d}u\, \mathrm{d}v = O(B_R^2)$ concludes the proof of the bound (2.65). □

In the following lemma we modify the previous result for the raw covariances $G_{h,t}^X$ instead of

$\tilde{G}^X_{h,t}$.

Lemma 2.7.5. *Under (A1), (B1) — (B5), (B7) and (B8), for $h \in \mathbb{Z}$ and $0 \le p+q \le 2$; $p,q \in \mathbb{N}_0$*

$$Q^{(h)}_{pq} = M_{[Q_{pq}]} + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right)$$

uniformly in $x, y \in [0,1]$.

Proof. We follow the lines of the discussion at the end of the proof of Yao et al. (2005a, Thm 1). Consider a generic raw covariance and its counterpart

$$\begin{aligned} G^X_{h,t}(x,y) &= \big(X_{t+h}(x) - \hat{\mu}_X(x)\big)\big(X_t(y) - \hat{\mu}_X(y)\big), \\ \tilde{G}^X_{h,t}(x,y) &= \big(X_{t+h}(x) - \mu_X(x)\big)\big(X_t(y) - \mu_X(y)\big). \end{aligned}$$

They can be related to each other by the expansion:

$$\begin{aligned} G^X_{h,t}(x,y) = \tilde{G}^X_{h,t}(x,y) &+ \big(X_{t+h}(x) - \mu_X(x)\big)\big(\mu_X(y) - \hat{\mu}_X(y)\big) + \\ &+ \big(\mu_X(x) - \hat{\mu}_X(x)\big)\big(X_t(y) - \mu_X(y)\big) + \big(\mu_X(x) - \hat{\mu}_X(x)\big)\big(\mu_X(y) - \hat{\mu}_X(y)\big). \end{aligned}$$

By (2.32), the difference of $G^X_{h,t}(x,y)$ and $\tilde{G}^X_{h,t}(x,y)$ is of order $O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_\mu}\right)$ which is negligible with respect to the rate $O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2}\right)$ from Lemma 2.7.4. □

Proof of the second part of Theorem 2.3.1. Combining the results of Lemma 2.7.2 and Lemma 2.7.5, we obtain the following uniform convergence rates:

$$\begin{aligned} \mathscr{A}_1^{(h)} &= \left[c_h g(x) g(y) \sigma_K^2\right]^2 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right), \\ \mathscr{A}_2^{(h)} &= O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right), \\ \mathscr{A}_3^{(h)} &= O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right), \\ \mathscr{B}^{(h)} &= \left[c_h g(x) g(y)\right]^3 \left(\sigma_K^2\right)^2 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right). \end{aligned}$$

The numerator of the ratio (2.59) exhibits the following uniform convergence

$$\mathscr{A}_1^{(h)} Q_{00}^{(h)} - \mathscr{A}_2^{(h)} Q_{10}^{(h)} - \mathscr{A}_3^{(h)} Q_{01}^{(h)} = \left[c_h g(x) g(y)\right]^3 \left(\sigma_K^2\right)^2 R_h(x,y) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right) \quad (2.70)$$

and therefore we have proven the convergence rate for the autocovariance kernel estimator

$$\hat{R}_h(x,y) = R_h(x,y) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right)$$

uniformly in $x, y \in [0,1]$. □

Finally we turn to the estimation of the measurement error variance σ^2. The minimizer of the local quadratic smoother (2.5) can be expressed explicitly as

(2.71)

$$\bar{R}_0(x) = \left(\bar{\mathscr{A}}_1\bar{Q}_0 - \bar{\mathscr{A}}_2\bar{Q}_1 - \bar{\mathscr{A}}_3\bar{Q}_2\right)\bar{\mathscr{B}}^{-1}$$

where

$$\begin{aligned}
&\bar{\mathscr{A}}_1 = \bar{S}_2\bar{S}_4 - \left(\bar{S}_3\right)^2, \qquad \bar{\mathscr{A}}_2 = \bar{S}_1\bar{S}_4 - \bar{S}_2\bar{S}_3, \qquad \bar{\mathscr{A}}_3 = \bar{S}_2\bar{S}_2 - \bar{S}_1\bar{S}_3,\\
&\bar{\mathscr{B}} = \bar{\mathscr{A}}_1\bar{S}_0 - \bar{\mathscr{A}}_2\bar{S}_1 - \bar{\mathscr{A}}_3\bar{S}_2,\\
&\bar{S}_r = \frac{1}{T}\sum_{t=1}^{T}\sum_{j\neq k}\left(\frac{\Delta(x_{tj},x_{tk})}{B_R}\right)^r \frac{1}{B_R^2}K\left(\frac{x_{tj}-x}{B_R}\right)K\left(\frac{x_{tk}-x}{B_R}\right),\\
&\bar{Q}_r = \frac{1}{T}\sum_{t=1}^{T}\sum_{j\neq k}G^X_{0,t}(x_{tj},x_{tk})\left(\frac{\Delta(x_{tj},x_{tk})}{B_R}\right)^r \frac{1}{B_R^2}K\left(\frac{x_{tj}-x}{B_R}\right)K\left(\frac{x_{tk}-x}{B_R}\right).
\end{aligned}$$

All of the above quantities are understood as functions of $x \in [0,1]$ and all operations are considered pointwise, including the pointwise inversion $\bar{\mathscr{B}}^{-1} = (\bar{\mathscr{B}}(x))^{-1}$.

Lemma 2.7.6. *Under (A1), (B1), (B2) and (B7), for* $r \in 0,1,2,3,4$

$$\bar{S}_r(x) = M_{[\bar{S}_r]}(x) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_V^2} + B_V^2\right),$$

uniformly in $x \in [0,1]$ *where*

$$\begin{aligned}
&M_{[\bar{S}_0]} = c_0 g(x)^2, && M_{[\bar{S}_1]} = M_{[\bar{S}_3]} = 0,\\
&M_{[\bar{S}_2]} = \frac{1}{2}c_0 g(x)^2\sigma_K^2, && \sigma_K^2 = \int v^2K(v)\,\mathrm{d}v,\\
&M_{[\bar{S}_4]} = \frac{1}{8}c_0 g(x)^2\left(\mu_4^{(K)} + 3\sigma_K^2\right), && \mu_4^{(K)} = \int v^4K(v)\,\mathrm{d}v
\end{aligned}$$

and $c_0 = \mathbb{E}[N(N-1)]$.

Proof. The proof of Lemma 2.7.6 follows in the footsteps of that of Lemma 2.7.3, and the details are omitted. □

Lemma 2.7.7. *Under (A1), (B1) — (B5) and (B7) — (B9), for* $r \in 0,1,2$

$$\bar{Q}_r(x) = M_{[\bar{Q}_r]}(x) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_V^2} + B_V^2\right),$$

uniformly in $x \in [0,1]$ *where*

$$M_{[\bar{Q}_0]} = c_0 R_0(x,x) g(x)^2, \qquad M_{[\bar{Q}_1]} = 0,$$
$$M_{[\bar{Q}_2]} = \frac{1}{2} c_0 R_0(x,x) g(x)^2 \sigma_K^2, \qquad \sigma_K^2 = \int v^2 K(v)\,\mathrm{d}v.$$

Proof. The proof of Lemma 2.7.7 is analogous to the proofs of Lemma 2.7.4 and Lemma 2.7.5. □

The following corollary is a direct consequence of Lemma 2.7.6 and Lemma 2.7.7, and the formula (2.71).

Corollary 2.7.8. *Under (B1) — (B5) and (B7) — (B9),*

$$\bar{R}_0(x) = R_0(x,x) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_V^2} + B_V^2\right)$$

uniformly in $x \in [0,1]$.

Now we turn our attention to the linear smoother on the diagonal (2.6).

Lemma 2.7.9. *Under (A1), (B1) — (B5) and (B7) — (B9),*

$$\hat{V}(x) = R_0(x,x) + \sigma^2 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}B_V} + B_V^2\right),$$

uniformly in $x \in [0,1]$.

Proof. The proof is similar to the proofs of the above lemmas. An explicit formula for the minimizer of (2.6) can be found analogously. □

Proof of the last part of the Theorem 2.3.1. Combining the results of Lemma 2.7.6, Lemma 2.7.7, and Lemma 2.7.9 yields the rate (2.34). See also the proof of Li and Hsing (2010, Thm 3.4) where the proof with the local-linear smoothing of the diagonal is written out in detail. □

2.7.2 Proof of Theorem 2.3.2

Firstly we comment that the minimizer to (2.10) and hence the estimator can be expressed explicitly (2.9) as

$$\hat{f}_\omega(x,y) = \frac{1}{2\pi}\left(\mathscr{A}_1 Q_{00}^\omega - \mathscr{A}_2 Q_{10}^\omega - \mathscr{A}_3 Q_{01}^\omega\right)\mathscr{B}^{-1}, \tag{2.72}$$

where

$$\begin{aligned}
&\mathscr{A}_1 = S_{20}S_{02} - S_{11}^2, \qquad \mathscr{A}_2 = S_{10}S_{02} - S_{01}S_{11}, \qquad \mathscr{A}_3 = S_{01}S_{20} - S_{10}S_{11},\\
&\mathscr{B} = \mathscr{A}_1 S_{00} - \mathscr{A}_2 S_{10} - \mathscr{A}_3 S_{01},\\
&S_{pq} = \frac{1}{L}\sum_{h=-L}^{L} \frac{W_h}{\widehat{\mathscr{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q \times\\
&\qquad \times \frac{1}{B_R^2} K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right),\\
&Q_{pq}^{\omega} = \sum_{h=-L}^{L} \frac{W_h e^{-\mathrm{i}h\omega}}{\widehat{\mathscr{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} G_{h,t}^{X}(x_{t+h,j}, x_{tk}) \times\\
&\qquad \times \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q \frac{1}{B_R^2} K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right).
\end{aligned}$$

All of the above quantities are understood as functions of $(x,y) \in [0,1]^2$ and all operations are considered in a pointwise sense, including the pointwise inversion $\mathscr{B}^{-1} = (\mathscr{B}(x,y))^{-1}$.

To see why the minimizer has the form (2.72) we simplify the notation of the complex minimisation problem (2.10) to the following:

$$\hat{\mathbf{d}} = \min_{d_0,d_1,d_2} \sum_{j=1}^{J} \left| A_j - d_0 - d_1(x_j - x) - d_2(y_j - y) \right|^2 v_j \tag{2.73}$$

where $A_j \in \mathbb{C}$ represents the raw covariances multiplied by the complex exponential, and $v_j \geq 0$ are the spatial and Barlett's weights. The sum of squares can be rewritten in the as

$$\min_{d_0,d_1,d_2} (\mathbf{A} - \mathbb{X}\mathbf{d})^{\dagger} \mathbb{V} (\mathbf{A} - \mathbb{X}\mathbf{d})$$

where $\dagger$ denotes the complex conjugate, $\mathbf{A} = (A_1, \ldots, A_J)^{\top} \in \mathbb{C}^J$, $\mathbf{d} = (d_0, d_1, d_3) \in \mathbb{C}^3, \mathbb{V} = \operatorname{diag}(v_1, \ldots, v_J) \in \mathbb{R}^{J\times J}$ and

$$\mathbb{X} = \begin{pmatrix} 1 & x_1 - x & y_1 - y \\ \vdots & \vdots & \vdots \\ 1 & x_J - x & y_J - y \end{pmatrix} \in \mathbb{R}^{J\times 3}.$$

Since $\mathbb{X}$ and $\mathbb{V}$ are real, the real and imaginary parts of the minimisation (2.73) are separated:

$$\hat{\mathbf{d}} == \underbrace{\left(\operatorname*{arg\,min}_{\Re\mathbf{d}} (\Re\mathbf{A} - \mathbb{X}\Re\mathbf{d})^{\top} \mathbb{V} (\Re\mathbf{A} - \mathbb{X}\Re\mathbf{d}) \right)}_{\Re\hat{\mathbf{d}}} + \mathrm{i} \underbrace{\left(\operatorname*{arg\,min}_{\Im\mathbf{d}} (\Im\mathbf{A} - \mathbb{X}\Im\mathbf{d})^{\top} \mathbb{V} (\Im\mathbf{A} - \mathbb{X}\Im\mathbf{d}) \right)}_{\Im\hat{\mathbf{d}}}.$$

The above minimisation problems are solved by classical (weighted) normal equations:

$$\hat{\mathbf{d}} = \Re\hat{\mathbf{d}} + \mathrm{i}\,\Im\hat{\mathbf{d}} = \left(\mathbb{X}^{\top}\mathbb{V}\mathbb{X}\right)^{-1}\mathbb{X}^{\top}\mathbb{V}\Re\mathbf{A} + \mathrm{i}\left(\mathbb{X}^{\top}\mathbb{V}\mathbb{X}\right)^{-1}\mathbb{X}^{\top}\mathbb{V}\Im\mathbf{A} = \left(\mathbb{X}^{\top}\mathbb{V}\mathbb{X}\right)^{-1}\mathbb{X}^{\top}\mathbb{V}\mathbf{A}.$$

We can calculate the first element of $\left(\mathbb{X}^\top\mathbb{V}\mathbb{X}\right)^{-1}\mathbb{X}^\top\mathbb{V}\mathbf{A}$ by Cramér's rule. After switching back to the quadruple summation (2.10) we arrive at the formula (2.72).

To investigate the asymptotic behaviour of the estimator (2.9), we need to analyse the asymptotics of the terms in the formula (2.72). We now assess the asymptotics of S_{pq} and Q^{ω}_{pq}.

Lemma 2.7.10. *Under the assumptions (A1), (B1), (B2), and (B8), for any $p, q \in \mathbb{N}_0$, such that $0 \leq p+q \leq 2$:*

$$S_{pq} = M_{[S_{pq}]} + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2\right)$$

uniformly in $x, y \in [0,1]$ and where

$$M_{[S_{00}]} = g(x)g(y), \qquad M_{[S_{01}]} = M_{[S_{10}]} = M_{[S_{11}]} = 0,$$
$$M_{[S_{20}]} = M_{[S_{02}]} = g(x)g(y)\sigma_K^2, \qquad \sigma_K^2 = \int v^2 K(v)\,\mathrm{d}v.$$

Proof. Denote

$$\mathscr{S}^{(pq)}_{htjk} = \left(\frac{x_{t+h,j}-x}{B_R}\right)^p\left(\frac{x_{tk}-y}{B_R}\right)^q\frac{1}{B_R^2}K\left(\frac{x_{t+h,j}-x}{B_R}\right)K\left(\frac{x_{tk}-y}{B_R}\right),$$

for $h = -L,\ldots,L$, $t = 1,\ldots,T-h$, $j = 1,\ldots,N_{t+h}$, $k = 1,\ldots,N_t$ for $j \neq k$ if $h = 0$. Because $L = o(T)$ we may assume (and we do) in the entire proof that $L \leq T/2$. Noting that $L^{-1}\sum_{h=-L}^{L} W_h = 1$ we start with the decomposition

$$\begin{aligned}
&\left|\frac{1}{L}\sum_{h=-L}^{L}\left(\frac{W_h}{\widehat{\mathscr{N}_h}}\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\mathscr{S}^{(pq)}_{htjk}\right) - M_{[S_{pq}]}\right| \leq \\
&\quad\leq \left|\frac{1}{L}\sum_{h=-L}^{L}\frac{W_h}{\widehat{\mathscr{N}_h}}\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\left(\mathscr{S}^{(pq)}_{htjk} - M_{[S_{pq}]}\right)\right| + \left|\frac{1}{L}\sum_{h=-L}^{L}W_h M_{[S_{pq}]}\left(1-\frac{\mathscr{N}_h}{\widehat{\mathscr{N}_h}}\right)\right| \leq \\
&\quad\leq \left|\frac{1}{L}\sum_{h=-L}^{L}\frac{W_h}{\widehat{\mathscr{N}_h}}\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\left(\mathscr{S}^{(pq)}_{htjk} - M_{[S_{pq}]}\right)\right| + \left|\frac{1}{L}M_{[S_{pq}]}\left(1-\frac{\mathscr{N}_t}{\widehat{\mathscr{N}_0}}\right)\right| + \\
&+\left|\frac{1}{L}\sum_{h=-L,h\neq 0}^{L}W_h M_{[S_{pq}]}\left(1-\frac{\mathscr{N}_h}{(T-|h|)(\mathbb{E}N)^2}\right)\right| + \left|\frac{1}{L}\sum_{h=-L,h\neq 0}^{L}W_h M_{[S_{pq}]}\frac{\mathscr{N}_h}{T-|h|}\left(\frac{1}{(\bar{N})^2}-\frac{1}{(\mathbb{E}N)^2}\right)\right|
\end{aligned} \tag{2.74}$$

where $\mathscr{N}_h = \sum_{t=\min(1,1-h)}^{\max(T,T-h)} N_{t+h}N_t$ for $h \neq 0$ and $\mathscr{N}_0 = \sum_{t=1}^{T} N_t(N_t-1)$. The second term on the right-hand side of (2.74) is of order $O_{\mathbb{P}}(L)$. The third term is bounded by bounding the variance $\mathscr{N}_t \leq UT$ for $|h| \leq T/2$ where the constant U is independent of T and h but may

depend on the distribution of N. Thus the third term is of order $O_{\mathbb{P}}(T^{-1/2})$ thanks to

$$\left|\frac{1}{L}\sum_{h=-L,h\neq 0}^{L}\mathbb{E}\left[1-\frac{\mathscr{N}_h}{(T-|h|)(\mathbb{E}N)^2}\right]\right| \leq \frac{1}{L}\sum_{h=-L,h\neq 0}^{L}\left\{\mathrm{var}\left(\frac{\mathscr{N}_h}{(T-|h|)(\mathbb{E}N)^2}\right)\right\}^{1/2} = O(T^{-1/2}).$$

The fourth term on the right hand side of order $O_{\mathbb{P}}(T^{-1/2})$ because $\bar{N} = \mathbb{E}N + O_{\mathbb{P}}(T^{-1/2})$. The first term on the right-hand side of (2.74) is decomposed as

$$\begin{aligned}\left|\frac{1}{L}\sum_{h=-L}^{L}\frac{W_h}{\widehat{\mathscr{N}_h}}\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\left(\mathscr{S}_{htjk}^{(pq)} - M_{[S_{pq}]}\right)\right| \leq \\ \leq \frac{1}{L}\sum_{h=-L}^{L}\frac{W_h}{\widehat{\mathscr{N}_h}}\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\left|\mathscr{S}_{htjk}^{(pq)} - \mathbb{E}\mathscr{S}_{htjk}^{(pq)}\right| + \\ + \frac{1}{L}\sum_{h=-L}^{L}\frac{W_h}{\widehat{\mathscr{N}_h}}\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\left|\mathbb{E}\mathscr{S}_{htjk}^{(pq)} - M_{[S_{pq}]}\right|. \quad (2.75)\end{aligned}$$

The second term on the right hand side of (2.75) is of order $O_{\mathbb{P}}(B_R^2)$ because $\left|\mathbb{E}\mathscr{S}_{htjk}^{(pq)} - M_{[S_{pq}]}\right| = O_{\mathbb{P}}(B_R^2)$ uniformly. The first term on the right hand side of (2.75) is treated using similar steps as in the proof of Lemma 2.7.3, therefore for a constant U independent of B_R, T and $|h| < T/2$,

$$\mathbb{E}\left[\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\sup_{x,y\in[0,1]}\left|\mathscr{S}_{htjk}^{(pq)} - \mathbb{E}\mathscr{S}_{htjk}^{(pq)}\right| | N_1,\ldots,N_T\right] \leq U\frac{\sqrt{\mathscr{N}_h}}{B_R^2}.$$

We conclude that the right hand side of (2.75) is of order $O_{\mathbb{P}}(\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + B_R^2)$ by observing

$$\frac{1}{L}\sum_{h=-L}^{L}\frac{\sqrt{\mathscr{N}_h}}{\widehat{\mathscr{N}_h}} = O_{\mathbb{P}}(T^{-1/2}).$$

□

Lemma 2.7.11. *Let $p, q \in \mathbb{N}_0$, be such that $0 \leq p+q \leq 2$. We have the following results.*

1. *Assuming (A1), (B1) — (B5), (B7), (B8) and (B10), and letting $T \to \infty$:*

$$\tilde{Q}_{pq}^{\omega} = M_{[Q_{pq}^{\omega}]} + o_{\mathbb{P}}(1), \qquad Q_{pq}^{\omega} = M_{[Q_{pq}^{\omega}]} + o_{\mathbb{P}}(1).$$

2. *Assuming (B1) — (B8) and (B10), and letting $T \to \infty$:*

$$\tilde{Q}_{pq}^{\omega} = M_{[Q_{pq}^{\omega}]} + O_{\mathbb{P}}\left(L\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + LB_R^2 + \frac{1}{L}\right), \qquad Q_{pq}^{\omega} = M_{[Q_{pq}^{\omega}]} + O_{\mathbb{P}}\left(L\frac{1}{\sqrt{T}}\frac{1}{B_R^2} + LB_R^2 + \frac{1}{L}\right).$$

All convergences are uniformly in $\omega \in [-\pi, \pi]$ *and* $x, y \in [0,1]$ *and*

$$M_{[Q^{\omega}_{00}]} = 2\pi g(x) g(y) f_{\omega}(x,y), \qquad M_{[Q^{\omega}_{10}]} = M_{[Q^{\omega}_{01}]} = 0.$$

Proof. Analogously to Lemma 2.7.4, we first assume that $\mu_X(\cdot)$ is known. Hence we define

$$\tilde{Q}^{\omega}_{pq} = \sum_{h=-L}^{L} \frac{W_h e^{-\mathrm{i}h\omega}}{\widehat{\mathcal{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \tilde{G}^X_{h,t}(x_{t+h,j}, x_{tk}) \times \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q \frac{1}{B_R^2} K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right).$$

Denote $M_{[Q_{00,h}]}(x,y) = g(x)g(y)R_h(x,y)$ and $M_{[Q_{10,h}]}(x,y) = M_{[Q^{\omega}_{01,h}]}(x,y) = 0$. Further denote

$$\tilde{\mathcal{Q}}^{(pq)}_{htjk} = \tilde{G}^X_{h,t}(x_{t+h,j}, x_{tk}) \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q \frac{1}{B_R^2} K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right)$$

$h = -L, \dots, L$, $t = 1, \dots, T-h$, $j = 1, \dots, N_{t+h}$, $k = 1, \dots, N_t$ for $j \neq k$ if $h = 0$, we can write

$$\left|\tilde{Q}^{\omega}_{pq} - M_{[Q^{\omega}_{pq}]}\right| \leq \left| \sum_{h=-L}^{L} \frac{W_h e^{-\mathrm{i}h\omega}}{\widehat{\mathcal{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \tilde{\mathcal{Q}}^{(pq)}_{htjk} - \sum_{h=-\infty}^{\infty} M_{[Q_{pq,h}]} e^{-\mathrm{i}h\omega} \right| \leq$$

$$\leq \sum_{h=-L}^{L} \frac{W_h}{\widehat{\mathcal{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \left|\tilde{\mathcal{Q}}^{(pq)}_{htjk} - M_{[Q_{pq,h}]}\right| + \frac{1}{L} \sum_{h=-L}^{L} |h| \left|M_{[Q^{\omega}_{pq,h}]}\right| + \sum_{|h|>L} \left|M_{[Q_{pq,h}]}\right| \tag{2.76}$$

Under the assumption (B5), the second and the third term on the right-hand side of (2.76) converge to zero uniformly in $x, y \in [0,1]$ by Kronecker's lemma. Assuming further the assumption (B6), these terms are in fact of order $O(L^{-1})$ uniformly in $x, y \in [0,1]$.

The first term on the right-hand side of (2.76) is treated similarly as in the proof of Lemma 2.7.5.

$$\sum_{h=-L}^{L} \frac{W_h}{\widehat{\mathcal{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \left|\tilde{\mathcal{Q}}^{(pq)}_{htjk} - M_{[Q_{pq,h}]}\right| \leq$$

$$\leq \sum_{h=-L}^{L} \frac{W_h}{\widehat{\mathcal{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \left|\tilde{\mathcal{Q}}^{(pq)}_{htjk} - \mathbb{E}\tilde{\mathcal{Q}}^{(pq)}_{htjk}\right| +$$

$$+ \sum_{h=-L}^{L} \frac{W_h}{\widehat{\mathcal{N}_h}} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \left|\mathbb{E}\tilde{\mathcal{Q}}^{(pq)}_{htjk} - M_{[Q_{pq,h}]}\right| \tag{2.77}$$

The second term on the right-hand side of (2.77) is of order $O(LB_R^2)$ uniformly in $x, y \in [0,1]$. The first term on the right-hand side of (2.77) is treated analogously as in the proof of Lemma 2.7.5, thus there exists a constant U independent of B_R, T and $|h| < T/2$ such that

$$\mathbb{E}\left[\sum_{h=-L}^{L}\frac{W_h}{\widehat{\mathscr{N}_h}}\sum_{t=\max(1,1-h)}^{\min(T,T-h)}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}}\sum_{k=1}^{N_t}\sup_{x,y\in[0,1]}\left|\tilde{\mathscr{Q}}_{htjk}^{(pq)}-\mathbb{E}\tilde{\mathscr{Q}}_{htjk}^{(pq)}\right| | N_1,\ldots,N_t\right] \leq U\frac{\sqrt{\mathscr{N}_h}}{B_R^2}.$$

The rates $o_{\mathbb{P}}(1)$, and $O_{\mathbb{P}}(LT^{-1/2}B_R^{-2})$ under the assumption (B6), are concluded by the fact

$$\sum_{h=-L}^{L}\frac{\sqrt{\mathscr{N}_h}}{\widehat{\mathscr{N}_h}} = O_{\mathbb{P}}(LT^{-1/2}).$$

The proof is completed by the same steps as in the proof of Lemma 2.7.5, switching to the $O_{\mathbb{P}}$ notation and noting that the difference between $\tilde{Q}_{pq}^{\omega}$ and Q_{pq}^{ω} is asymptotically negligible. □

Proof of Theorem 2.3.2. Combining the above derived results in lemmas 2.7.10 and 2.7.11 we are ready to establish the asymptotic behaviour of the terms that enter the formula (2.72).

$$\begin{aligned}
&\mathscr{A}_1 = \left[g(x)g(y)\sigma_K^2\right]^2 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2}+B_R^2\right), && \mathscr{A}_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2}+B_R^2\right),\\
&\mathscr{B} = \left[g(x)g(y)\right]^3\left(\sigma_K^2\right)^2, && \mathscr{A}_3 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2}+B_R^2\right), + O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\frac{1}{B_R^2}+B_R^2\right),\\
&Q_{00} = 2\pi g(x)g(y)f_{\omega}(x,y) + o_{\mathbb{P}}(1), && Q_{10} = o_{\mathbb{P}}(1),\\
& && Q_{01} = o_{\mathbb{P}}(1)
\end{aligned}$$

uniformly in $\omega \in [-\pi,\pi]$ and $x, y \in [0,1]$. Finally, the numerator of (2.72) is

$$\mathscr{A}_1 Q_{00}^{\omega} - \mathscr{A}_2 Q_{10}^{\omega} - \mathscr{A}_3 Q_{01}^{\omega} = 2\pi f_{\omega}(x,y)\left[g(x)g(y)\right]^3\left(\sigma_K^2\right)^2 + o_{\mathbb{P}}(1)$$

uniformly in $\omega \in [-\pi,\pi]$ and $x, y \in [0,1]$ which completes the proof of consistency. Under the assumption (B6) we replace $o_{\mathbb{P}}(1)$ by $O_{\mathbb{P}}(L/(\sqrt{T}B_R^2) + LB_R^2 + 1/L)$. □

2.7.3 Proof of Corollary 2.3.3

Proof of Corollary 2.3.3. By Theorem 2.3.2, we have for $h \in \mathbb{Z}$ and $x, y \in [0,1]$:

$$\begin{aligned}
\tilde{R}_h(x,y) - R_h(x,y) &= \int_{-\pi}^{\pi}\left\{\tilde{f}_{\omega}(x,y) - f_{\omega}(x,y)\right\}e^{\mathrm{i}h\omega}\,\mathrm{d}\omega,\\
\sup_{h\in\mathbb{Z}}\sup_{x,y\in[0,1]}\left|\tilde{R}_h(x,y) - R_h(x,y)\right| &\leq 2\pi\sup_{\omega\in[-\pi,\pi]}\sup_{x,y\in[0,1]}\left|\tilde{f}_{\omega}(x,y) - f_{\omega}(x,y)\right| = o_{\mathbb{P}}(1).
\end{aligned}$$

Assuming further (B6), proving the statement (2.37) is analogous to the previous line. □

2.7.4 Proof of Theorem 2.3.4

Proof of Theorem 2.3.4 . We begin with the estimation of the mean function $\mu_X(\cdot)$. We are going to make use of Hansen (2008, Thm 10). Define the two-dimensional time-series $\{\tilde{U}_i, \tilde{X}_i\}_i$ composed of the sparse observations and their observation locations according to the observation scheme (2.1)

$$\begin{aligned}\left(\tilde{U}_1, \tilde{U}_2, \ldots\right) &= \left(U_{1,1}, \ldots, U_{1,N_1}, U_{2,1}, \ldots, U_{2,N_2}, U_{3,1}, \ldots\right),\\ \left(\tilde{X}_1, \tilde{X}_2, \ldots\right) &= \left(x_{1,1}, \ldots, x_{1,N_1}, x_{2,1}, \ldots, x_{2,N_2}, x_{3,1}, \ldots\right).\end{aligned}$$

Under the assumptions (D1) — (D7), the time-series $\{\tilde{U}_i, \tilde{X}_i\}_i$ is strictly stationary and strongly mixing, with mixing coefficients $\tilde{\alpha}(m) \leq \tilde{A}m^{-\beta}$, and satisfies the conditions (2) — (7) of Hansen (2008). In particular:

- the condition of (3) Hansen (2008)

$$\mathbb{E}|\tilde{U}_1|^s \leq 2^s \left(\mathbb{E}|X_1(x_{11})|^s + \mathbb{E}|\epsilon_{1,1}|^s\right) < \infty,$$

- the condition (6) of Hansen (2008)

$$\sup_{x\in[0,1]} \mathbb{E}\left[|\tilde{U}_1|^s | \tilde{X} = x\right] g(x) \leq B_1 B_3 < \infty,$$

- the condition (7) of Hansen (2008)

$$\begin{aligned}&\sup_{x,x'\in[0,1]} \mathbb{E}\left[|\tilde{U}_1\tilde{U}_j| | \tilde{X}_1 = x, \tilde{X}_j = x'\right] g(x)g(x') \leq\\ &\quad\leq \left\{\sup_{h\in\mathbb{Z}} \sup_{x,x'\in[0,1]} \left[\mathbb{E}\left|X_h(x)X_0(x')\right|\right] + 2\sup_{x\in[0,1]} \left[\mathbb{E}|X_0(x)|\mathbb{E}|\epsilon_{1,1}|\right] + \sigma^2\right\} g(x)g(x') < \infty.\end{aligned}$$

The conditions (10) — (13) of Hansen (2008) are also satisfied taking $q = \infty$ and $c_n = 1$. Therefore all conditions of Hansen (2008, Thm 10) are satisfied. Noting that the length $n = n(T)$ of the time-series $\{\tilde{U}_i, \tilde{X}_i\}_i$ is asymptotically of the same order as T of the functional time series $\{X_t(\cdot)\}$, formally $n = n(T) \asymp T, \quad T \to \infty$, yields

$$\sup_{x\in[0,1]} \left|\hat{\mu}_X(x) - \mu_X(x)\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{TB_\mu}} + B_\mu^2\right).$$

Next we turn to the estimation of the lag-h autocovariance kernels $R_h(\cdot,\cdot)$. Fix $h \in \mathbb{Z}$. For simplicity consider $h \neq 0$. The proof for $h = 0$ is essentially the same, only the diagonal "raw" covariances must be removed. For the moment assume that the mean function $\mu_X(\cdot)$ is known and we shall work with the "raw" covariances $\tilde{G}^X_{h,t}(x_{t+h,j}, x_{tk})$ as defined in (2.64). Similarly as in the first part of this proof, define now the three dimensional time-series $\{\tilde{U}_i, \tilde{X}_i\}_i$ composed

of the "raw" covariances and their locations

$$\begin{aligned}\left(\tilde{U}_1, \tilde{U}_2, \ldots\right) &= \Big(\tilde{G}_{h,1}(x_{1+h,1}, x_{1,1}), \ldots, \tilde{G}_{h,1}(x_{1+h,N_{1+h}}, x_{1,N_t}),\\ &\qquad \tilde{G}_{2,1}(x_{2+h,1}, x_{2,1}), \ldots\Big),\\ \left(\tilde{X}_1, \tilde{X}_2, \ldots\right) &= \left(\begin{bmatrix} x_{1+h,1} \\ x_{1,1} \end{bmatrix}, \ldots, \begin{bmatrix} x_{1+h,N_{t+h}} \\ x_{1,N_t} \end{bmatrix}, \begin{bmatrix} x_{2+h,1} \\ x_{2,1} \end{bmatrix}, \ldots\right).\end{aligned}$$

We are again going to make us of Hansen (2008, Thm 10). Under the assumptions (D1) — (D11) it is easy to verify (analogously as in the first part of this proof) that the time series $\{\tilde{U}_i, \tilde{X}_i\}_i$ satisfies the conditions (2) — (7) of Hansen (2008). The conditions (10) — (13) also follow directly from our assumptions. It remains to repeat the discussion as in the proof of Lemma 2.7.5 to conclude that the difference between $\tilde{G}^X_{h,t}(x_{t+h,j}, x_{tk})$ and $G^X_{h,t}(x_{t+h,j}, x_{tk})$ is asymptotically negligible with respect to the rate bellow.

Therefore by Hansen (2008, Thm 10), for fixed $h \in \mathbb{Z}$,

$$\sup_{x,y\in[0,1]} \left|\hat{R}_h(x,y) - R_h(x,y)\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{TB_R^2}} + B_R^2\right).$$

□

2.7.5 Proof of Theorem 2.3.5

The proof of Theorem 2.3.5 is more involved. Rather than make direct use of, we shall need to modify the proof techniques of Hansen (2008) in order to construct our proof. We express the spectral density kernel estimator (2.9) in a similar way as in the proof of Theorem 2.3.2.

$$\hat{f}_\omega(x,y) = \frac{1}{2\pi}\left(\mathscr{A}_1 Q^\omega_{00} - \mathscr{A}_2 Q^\omega_{10} - \mathscr{A}_3 Q^\omega_{01}\right)\mathscr{B}^{-1}, \tag{2.78}$$

where

$$\begin{aligned}\mathscr{A}_1 &= S_{20}S_{02} - S_{11}^2, \qquad \mathscr{A}_2 = S_{10}S_{02} - S_{01}S_{11}, \qquad \mathscr{A}_3 = S_{01}S_{20} - S_{10}S_{11},\\ \mathscr{B} &= \mathscr{A}_1 S_{00} - \mathscr{A}_2 S_{10} - \mathscr{A}_3 S_{01},\\ S_{pq} &= \frac{1}{L}\sum_{h=-L}^{L} \frac{W_h \mathscr{N}_h}{\widehat{\mathscr{N}_h}} S^{(h)}_{pq},\\ S^{(h)}_{pq} &= \frac{1}{\mathscr{N}_h B_R^2} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q \times\\ &\qquad \times K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right),\end{aligned}$$

$$Q_{pq}^{\omega} = \sum_{h=-L}^{L} \frac{W_h e^{-\mathrm{i}h\omega}\mathscr{N}_h}{\widehat{\mathscr{N}_h}} Q_{pq}^{(h)},$$
$$Q_{pq}^{(h)} = \frac{1}{\mathscr{N}_h B_R^2} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \tilde{G}_{h,t}^{X}(x_{t+h,j}, x_{tk}) \times$$
$$\times \left(\frac{x_{t+h,j}-x}{B_R}\right)^p \left(\frac{x_{tk}-y}{B_R}\right)^q K\left(\frac{x_{t+h,j}-x}{B_R}\right) K\left(\frac{x_{tk}-y}{B_R}\right).$$

All of the above quantities are understood as functions of $(x,y) \in [0,1]^2$ and all operations are considered in a pointwise sense, including the pointwise inversion $\mathscr{B}^{-1} = (\mathscr{B}(x,y))^{-1}$.

Similarly as in the proof of Theorem 2.3.4 define for $h \in \mathbb{Z}$,

$$\left(\tilde{U}_1^{h,r}, \tilde{U}_2^{h,r}, \ldots\right) = \left(\left\{\tilde{G}_{h,1}(x_{1+h,1}, x_{1,1})\right\}^r, \ldots, \left\{\tilde{G}_{h,1}(x_{1+h,N_{1+h}}, x_{1,N_t})\right\}^r, \left\{\tilde{G}_{2,1}(x_{2+h,1}, x_{2,1})\right\}^r, \ldots\right),$$
$$\left(\tilde{X}_1^h, \tilde{X}_2^h, \ldots\right) = \left(\begin{bmatrix} x_{1+h,1} \\ x_{1,1} \end{bmatrix}, \ldots, \begin{bmatrix} x_{1+h,N_{t+h}} \\ x_{1,N_t} \end{bmatrix}, \begin{bmatrix} x_{2+h,1} \\ x_{2,1} \end{bmatrix}, \ldots\right).$$

Let $k(\cdot): \mathbb{R}^2 \to \mathbb{R}$ be a function satisfying the assumption (C1) and denote for $r = 0,1$ and $h \in \mathbb{Z}$ define

$$\hat{\Psi}^{h,r}(x,y) = \frac{1}{\mathscr{N}_h B_R^2} \sum_{t=\max(1,1-h)}^{\min(T,T-h)} \sum_{\substack{j=1 \\ j\neq k \text{ if } h=0}}^{N_{t+h}} \sum_{k=1}^{N_t} \left\{\tilde{G}_{h,t}^{X}(x_{t+h,j}, x_{tk})\right\}^r k\left(\frac{x_{t+h,j}-x}{B_R}, \frac{x_{tk}-y}{B_R}\right) \tag{2.79}$$
$$= \frac{1}{\mathscr{N}_h B_R^2} \sum_{i=1}^{\mathscr{N}_h} \tilde{U}_i^{h,r} k\left(\frac{\tilde{X}_i^h - (x,y)}{B_R}\right) \tag{2.80}$$
$$= \frac{1}{\mathscr{N}_h B_R^2} \sum_{i=1}^{\mathscr{N}_h} Z_i^{h,r}(x,y) \tag{2.81}$$

where we are denoting

$$Z_i^{h,r}(x,y) = \tilde{U}_i k\left(\left(\tilde{X}_i^{h,r} - (x,y)\right)/B_R\right). \tag{2.82}$$

Lemma 2.7.12. *Under the assumptions (A1), (D1) — (D11),*

$$\operatorname{var}(\hat{\Psi}^{h,r}(x,y)|\mathscr{N}_h) \leq \frac{\Theta}{\mathscr{N}_h B_R^2}$$

for $\mathscr{N}_h > 0$ and where the constant Θ is uniform in $h \in \mathbb{Z}$, $x,y \in [0,1]$, $r = 0,1$.

Proof. Note that the sequence $\{Z_i^{h,r}(x,y)\}_i$ is a stationary scalar time-series and denote its autocovariance function as $\rho_{Z_i^{h,r}(x,y)}(\xi)$ for lag ξ. Therefore we have the bound (2.58). Conditioning on $\mathscr{N}_h$ yields

$$\operatorname{var}\left(\frac{1}{\mathscr{N}_h}\sum_{i=1}^{\mathscr{N}_h} Z_i^{h,r}(x,y)|\mathscr{N}_h\right) \leq \frac{1}{\mathscr{N}_h}\sum_{\xi=-\infty}^{\infty}\left|\rho_{Z_i^{h,r}(x,y)}(\xi)\right|. \tag{2.83}$$

The sum on the right hand side of (2.83) can be bounded by

$$\sum_{\xi=-\infty}^{\infty}\left|\rho_{Z_i^{h,r}(x,y)}(\xi)\right| \le \left(N^{max}\right)^2 \sum_{\xi=-\infty}^{\infty} \sup_{x_1,x_2,x_3,x_4\in[0,1]} \Big|\operatorname{cum}\left(X_{\xi+h}(x_1), X_\xi(x_2), X_h(x_3), X_0(x_4)\right) + \\ + R_\xi(x_1,x_2)R_\xi(x_3,x_4) + R_{\xi+h}(x_1,x_4)R_{\xi-h}(x_2,x_3)\Big|. \quad (2.84)$$

The bound (2.84) is uniform in h and constitutes the constant Θ in Lemma 2.7.12. □

The key tool for our proof is an exponential-type inequality for strongly mixing random sequences. This result was given by Liebscher (1996, Thm 2.1) and Rio (1995, Thm 5).

Lemma 2.7.13 (Liebscher/Rio)**.** *Let Z_i be a stationary mean-zero real-valued process such that $|Z_i| \le b$, with strong mixing coefficients α_m. Denote $\sigma_m^2 = \mathbb{E}\left(\sum_{i=1}^m Z_i\right)^2$. Then for each positive integer $m \le n$ and $\epsilon > 0$ such that $m < \epsilon b/4$*

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| > \epsilon\right) \le 4\exp\left(-\frac{\epsilon^2}{64\frac{n\sigma_m^2}{m} + \frac{8}{3}\epsilon m b}\right) + 4\frac{n}{m}\alpha_m.$$

Lemma 2.7.14. *Under the assumptions (A1), (D1) — (D9) and (D11) — (D14)*

$$\sup_{\omega\in[-\pi,\pi]}\sup_{x,y\in[0,1]}\left|Q_{pq}^\omega - M_{[Q_{pq}^\omega]}\right| = o_{\mathbb{P}}(1), \quad (2.85)$$

and assuming further assumption (B6),

$$\sup_{\omega\in[-\pi,\pi]}\sup_{x,y\in[0,1]}\left|Q_{pq}^\omega - M_{[Q_{pq}^\omega]}\right| = O_{\mathbb{P}}\left(L\sqrt{\frac{\log T}{TB_R^2}} + B_R^2 + \frac{1}{L}\right) \quad (2.86)$$

where

$$M_{[Q_{00}^\omega]} = 2\pi g(x)g(y)f_\omega(x,y), \qquad M_{[Q_{10}^\omega]} = M_{[Q_{01}^\omega]} = 0.$$

Proof. Denote

$$M_{[Q_{00,h}]}(x,y) = g(x)g(y)R_h(x,y),$$
$$M_{[Q_{10,h}]}(x,y) = M_{[Q_{01,h}^\omega]}(x,y) = 0.$$

Similarly as in the proof of Lemma 2.7.11, decompose

$$\left|Q_{pq}^\omega - M_{[Q_{pq}^\omega]}\right| \le \left|\sum_{h=-L}^{L} W_h e^{-\mathrm{i}h\omega}\frac{\mathcal{N}_h}{\hat{\mathcal{N}}_h}Q_{pq}^{(h)} - \sum_{h=-\infty}^{\infty} M_{[Q_{pq,h}]}e^{-\mathrm{i}h\omega}\right| \le \sum_{h=-L}^{L} W_h\left|Q_{pq}^{(h)}\right|\left|\frac{\mathcal{N}_h}{\hat{\mathcal{N}}_h} - 1\right| + \\ + \sum_{h=-L}^{L} W_h\left|Q_{pq}^{(h)} - \mathbb{E}Q_{pq}^{(h)}\right| + \sum_{h=-L}^{L} W_h\left|\mathbb{E}Q_{pq}^{(h)} - M_{[Q_{pq,h}]}\right| + \frac{1}{L}\sum_{h=-L}^{L}|h|\left|M_{[Q_{pq,h}]}\right| + \sum_{|h|\ge L}\left|M_{[Q_{pq,h}]}\right|. \quad (2.87)$$

Under the assumption (D12), the last two terms on the right-hand side of (2.87) converge to zero uniformly in $x, y \in [0,1]$ by Kronecker's lemma. Assuming further the assumption (B6), these terms are in fact of order $O(L^{-1})$ uniformly in $x, y \in [0,1]$. The first term on the right-hand side of (2.87) is of order $O_{\mathbb{P}}(LT^{-1/2})$ uniformly in $x, y \in [0,1]$. The bias term, third term on the right-hand side of (2.87), is of order $O_{\mathbb{P}}(LB_R^2)$ which is shown exactly as in the proof of Lemma 2.7.4.

It remains to treat the second term on the right-hand side of (2.87), for which we start with the observation

$$\sup_{\omega\in[-\pi,\pi]} \sup_{x,y\in[0,1]} \sum_{h=-L}^{L} W_h \left| Q_{pq}^{(h)} - \mathbb{E}Q_{pq}^{(h)} \right| \leq \sum_{h=-L}^{L} \sup_{x,y\in[0,1]} \left| Q_{pq}^{(h)} - \mathbb{E}Q_{pq}^{(h)} \right|. \tag{2.88}$$

Denote $a_T = \left(\log T/(TB_R^2)\right)^{-1/2}$. To show the order $O_{\mathbb{P}}(La_T)$ of the right-hand side of (2.88) we investigate the probabilities for some $M > 0$

$$\begin{aligned}\mathbb{P}\left(\sum_{h=-L}^{L} \sup_{x,y\in[0,1]} \left| Q_{pq}^{(h)} - \mathbb{E}Q_{pq}^{(h)} \right| > MLa_T \right) &\leq \sum_{h=-L}^{L} \mathbb{P}\left(\sup_{x,y\in[0,1]} \left| Q_{pq}^{(h)} - \mathbb{E}Q_{pq}^{(h)} \right| > \frac{MLa_T}{2L+1} \right) \leq \\ &\leq \sum_{h=-L}^{L} \mathbb{P}\left(\sup_{x,y\in[0,1]} \left| Q_{pq}^{(h)} - \mathbb{E}Q_{pq}^{(h)} \right| > \frac{1}{3}Ma_T \right) \end{aligned} \tag{2.89}$$

We bound the probabilities on the right-hand side of (2.89) using the proof techniques presented in Hansen (2008, Thm 2). For the simplification of the notation and the proof we shall assume that the numbers of observation locations are deterministic and constant,

$$N_1 = \cdots = N_T = N^{max} \equiv N \geq 2. \tag{2.90}$$

Without this assumption, all bounds must be conditioned on these counts and the unconditional statements follow from the fact that $(1/T)\mathcal{N}_h = (\mathbb{E}N)^2 + O_{\mathbb{P}}(T^{-1/2})$ for $h \neq 0$ and $(1/T)\mathcal{N}_0 = (\mathbb{E}\{N(N-1)\}) + O_{\mathbb{P}}(T^{-1/2})$ where the convergences are uniform in $|h| < T/3$. Under the technical assumption (2.90), $\mathcal{N}_h = (T-|h|)N^2$ for $h \neq 0$ and $\mathcal{N}_0 = TN(N-1)$.

From the assumption (D13) we may take T to be sufficiently large so that

$$L \leq \frac{1}{2}\sqrt{\frac{\log T}{TB_R^2}}^{-\frac{s-2}{s-1}}. \tag{2.91}$$

Our proof follows essentially the same steps Hansen (2008, Thm 2), the only difference is that we need to keep track of the uniformity in h and adjust the convergence rate for the growing L.

Using the notation (2.79), (2.80), and (2.81) rewrite $Q_{pq}^{(h)}$ as

$$Q_{pq}^{(h)}(\tilde{x}) = \frac{1}{\mathcal{N}_h B_R^2} \sum_{i=1}^{\mathcal{N}_h} \tilde{U}_i^{h,1} k\left(\frac{\tilde{X}_i^h - \tilde{x}}{B_R} \right) = \frac{1}{\mathcal{N}_h B_R^2} \sum_{i=1}^{\mathcal{N}_h} \tilde{Z}_i^{h,1}(\tilde{x})$$

where $k(u,v) = u^p v^q K(u)K(v)$.

The proof consists of three steps. Firstly we replace the values of $\tilde{U}_i^{h,r}$ with the truncated process $\tilde{U}_i^{h,r}\mathbb{1}_{[|\tilde{U}_i^{h,r}|\le\tau_T]}$ where $\tau_T = a_T^{-1/(s-1)}$. Secondly, we replace the supremum over $\tilde{x} \equiv (x,y) \in [0,1]$ with a maximisation over a finite N_g-point grid. And finally, with the help of the exponential inequality from Lemma 2.7.13 we bound the remainder.

Define

$$R^{h,r}(\tilde{x}) = \hat{\Psi}^{h,r}(\tilde{x}) - \frac{1}{\mathcal{N}_h B_R^2}\sum_{i=1}^{\mathcal{N}_h} Z_i^{h,r}(\tilde{x})\mathbb{1}_{[\tilde{U}_i\le\tau_T]}.$$

Following the same steps as in the proof of Hansen (2008, Thm 2), we bound

$$\left|\mathbb{E}\left[R^{h,r}(\tilde{x})\right]\right| = O_{\mathbb{P}}\left(\tau_T^{-(s-1)}\right) = O_{\mathbb{P}}(a_T)$$

uniformly in $|h| < T/3$.

Thus replacing $\tilde{U}_i$ with $\tilde{U}_i\mathbb{1}_{[|\tilde{U}_i|\le\tau_T]}$ yields only an error of order $O_{\mathbb{P}}(a_T)$ and we therefore assume for the rest of the proof that $\tilde{U}_i \le \tau_T$.

The second step of the proof introduces a discretization of the square $[0,1]^2$ which can be covered by a regular grid of $N_g = 2B_R^{-2}a_T^{-2}$ points such that for each $(x,y) \in [0,1]^2$, the closest grid point $\tilde{x}_j \equiv (x_j, y_j)$ is at a distance of at most $B_R a_T$ distance. Denote this discretization as $A_j \subset [0,1]^2$, $j = 1,\dots,N_g$.

Thanks to the assumption (D9), for all $\tilde{x}_1, \tilde{x}_2 \in [0,1]^2$ satisfying $\|\tilde{x}_1 - \tilde{x}_2\| \le \delta \le \tilde{L}$, we have the bound

$$|k(\tilde{x}_1) - k(\tilde{x}_2)| \le \delta k^*(\tilde{x}_1) \tag{2.92}$$

where $k^* : \mathbb{R}^2 \to \mathbb{R}$ is a bounded integrable function. Indeed, if $k(\cdot)$ satisfies the compact support condition of (C1) and is Lipschitz then $k^*(u) = \Lambda_1 1_{[\|u\|\le 2\tilde{L}]}$. If on the other hand $k(u)$ satisfies the differentiability condition of (C1), then we may put $k^*(u) = \Lambda_1 1_{[\|u\|\le 2\tilde{L}]} + \|u-\tilde{L}\|^{-\eta}$.

The inequality (2.92) implies that if $a_T \le \tilde{L}$ then for $\tilde{x} \in A_j$ we have $\|\tilde{x} - \tilde{x}_j\|/B_R \le a_T$ and, for T large enough such that $a_T \le \tilde{L}$,

$$\left|k\left(\frac{\tilde{x} - \tilde{X}_i^h}{B_R}\right) - k\left(\frac{\tilde{x}_j - \tilde{X}_i^h}{B_R}\right)\right| \le a_T k^*\left(\frac{\tilde{x}_j - X_i}{B_R}\right).$$

Define

$$\tilde{\Psi}^{h,r}(\tilde{x}) = \frac{1}{\mathcal{N}_h B_R^2}\sum_{i=1}^{\mathcal{N}_h} \tilde{U}_i^{h,r} k^*\left(\frac{\tilde{x} - \tilde{X}_i^h}{B_R}\right),$$

that is, a modification of $\hat{\Psi}^{h,r}$ where $k(\cdot)$ is replaced by $k^*(\cdot)$. Note that by the assumptions (D4) and (D8), $\mathbb{E}\left|\tilde{\Psi}^{h,r}(\tilde{x})\right|$ is bounded uniformly in $h \in \mathbb{Z}$ and $r = 0,1$. Following the steps in

the proof of Hansen (2008, Thm 2), we conclude that

$$\sup_{\tilde{x}\in A_j}\left|\hat{\Psi}^{h,r}(\tilde{x})-\mathbb{E}\hat{\Psi}^{h,r}(\tilde{x})\right| \le \left|\hat{\Psi}^{h,r}(\tilde{x}_j)-\mathbb{E}\hat{\Psi}^{h,r}(\tilde{x}_j)\right| + \left|\tilde{\Psi}^{h,r}(\tilde{x}_j)-\mathbb{E}\tilde{\Psi}^{h,r}(\tilde{x}_j)\right| + 2a_T M,$$

for $M > \mathbb{E}|\tilde{\Psi}^{h,r}(\tilde{x})|$, and

$$\begin{aligned}
\mathbb{P}\left(\sup_{\tilde{x}\in[0,1]^2}\left|\hat{\Psi}^{h,r}(x)-\mathbb{E}\hat{\Psi}^{h,r}(x)\right| > 3Ma_T\right) \le \\
\le N_g \max_{j=1,\dots,N_g}\mathbb{P}\left(\left||\hat{\Psi}^{h,r}(\tilde{x}_j)-\mathbb{E}\hat{\Psi}^{h,r}(\tilde{x}_j)|\right| > Ma_T\right) + \qquad (2.93)\\
+ N_g \max_{j=1,\dots,N_g}\mathbb{P}\left(\left||\tilde{\Psi}^{h,r}(\tilde{x}_j)-\mathbb{E}\tilde{\Psi}^{h,r}(\tilde{x}_j)|\right| > Ma_T\right) \qquad (2.94)
\end{aligned}$$

The terms (2.93) and (2.94) are bounded likewise because the only difference between them is the presence of $k(\cdot)$ and $k^*(\cdot)$. Next we show how to bound (2.93).

By the definition (2.82) of $Z_i^{h,r}(\tilde{x})$ we notice that $|Z_i^{h,r}(\tilde{x})| \le \tau_T \bar{K} \equiv b_T$ because $|\tilde{U}_i^{h,r}| \le \tau_T$ and $\left|k((\tilde{x}-\tilde{X}_i^h)/B_R)\right| \le \bar{k}$ where $\bar{k}$ is the upper bound of the bounded function $k(\cdot)$. Therefore, by Lemma 2.7.12, for m sufficiently large we have, uniformly in $|h| < m/3$,

$$\sup_{\tilde{x}\in[0,1]^2}\mathbb{E}\left(\sum_{i=1}^{m} Z_i^{h,r}(\tilde{x})\right)^2 \le \Theta m B_R^2.$$

Put $m = (a_T\tau_T)^{-1}$ and we conclude that $m < T$ and $m < \epsilon b_T/4$ for $\epsilon = Ma_T T B_R^2$ for T sufficiently large. Therefore by Lemma 2.7.13 for any $\tilde{x} \in [0,1]$

$$\begin{aligned}
\mathbb{P}\left(\left|\hat{\Psi}^{h,r}(\tilde{x})-\mathbb{E}\hat{\Psi}^{h,r}(\tilde{x})\right| > Ma_T\right) = \mathbb{P}\left(\left|\sum_{i=1}^{\mathcal{N}_h} Z_i^{h,r}(\tilde{x})\right| > Ma_T\mathcal{N}_h B_R^2\right) \le \\
\le 4\exp\left(-\frac{M^2a_T^2T^2B_R^2}{64\Theta\mathcal{N}_hB_R^2 + 6\bar{k}MTB_R^2}\right) + 4\frac{\mathcal{N}_h}{m}\alpha_m \le \\
\le 4\exp\left(-\frac{M^2\log T}{64\,(N^{max})^2\Theta + 6\bar{k}M}\right) + 4\left(N^{max}\right)^2 T\tilde{A}(m-|h|)^{-\beta}m^{-1} \le \\
\le 4T^{-M/\left(64(N^{max})^2+\bar{k}\right)} + 4\left(N^{max}\right)^2\tilde{A}T\left(\frac{1}{2}m\right)^{-\beta}m^{-1} \le \\
\le 4T^{-M/\left(64(N^{max})^2+\bar{k}\right)} + 4(2^\beta)\left(N^{max}\right)^2\tilde{A}Ta_T^{1+\beta}\tau_T^{1+\beta}
\end{aligned}$$

where the second inequality comes from the fact that the time-series $\{Z_i^{h,r}(\tilde{x})\}$ is strong mixing with coefficients $\alpha_m \le \tilde{A}(m-|h|)^{-\beta}$ for $m \ge |h|$, the third inequality is due to (2.91), and the final one by taking $M > \Theta$. Since $N_g \le 2B_R^{-2}a_T^{-2}$ we have from the above inequality and (2.93) and (2.94) that

$$\mathbb{P}\left(\sup_{\tilde{x}\in[0,1]^2}\left|\hat{\Psi}^{h,r}(\tilde{x})-\mathbb{E}\hat{\Psi}^{h,r}(\tilde{x})\right| > 3Ma_T\right) \le O\left(C_{1,T}\right) + O\left(C_{2,T}\right) \qquad (2.95)$$

where

$$C_{1,T} = B_R^{-2} a_T^{-2} T^{-M/(64+6\bar{k})}$$
$$C_{2,T} = B_R^{-2} T a_T^{-1+\beta} \tau_T^{1+\beta}.$$

Returning to the inequalities (2.88) and (2.89), we conclude that

$$\mathbb{P}\left(\sum_{h=-L}^{L} \sup_{x,y\in[0,1]} \left| Q_{pq}^{(h)} - \mathbb{E}Q_{pq}^{(h)} \right| > MLa_T \right) \leq L\left[O(C_{1,T}) + O(C_{2,T}) \right] \tag{2.96}$$

Assumption (D10) implies that $(\log T)B_R^{-2} = o(T^\theta)$ and therefore also $B_R^{-2} = o\left(T^\theta\right)$ and $a_T = ((\log T)B_R^{-2}T^{-1})^{1/2} = o(T^{-(1-\theta)/2})$. For M sufficiently large and by the assumptions (D13) and (D14)

$$LC_{1,n} = o\left(T^{\theta_F + (1-\theta_F) - M/\left(64(N^{max})^2 + 6\bar{k}\right) + (1-\theta_F)(s-2)/(s-1)/2} \right) = o(1),$$
$$LC_{2,n} = o\left(T^{\theta_F + 1 - (1-\theta_F)[1+\beta-2-(1+\beta)/(s-1)-(s-2)/(s-1)]/2} \right) = o(1).$$

Thus (2.96) is of order $o(1)$ and we conclude, together with the rates of the other terms of (2.87), the rates (2.85) and (2.86). □

Lemma 2.7.15. *Under the assumptions (A1), (D1) — (D9) and (D11) — (D14),*

$$\sup_{x,y\in[0,1]} \left| S_{pq} - M_{[S_{pq}]} \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{TB_R^2}} + B_R^2 \right). \tag{2.97}$$

Proof. We decompose the estimation error as follows:

$$\begin{aligned} \left| S_{pq} - M_{[S_{pq}]} \right| &= \left| \frac{1}{L} \sum_{|h|<L} W_h \left(\frac{\mathcal{N}_h}{\hat{\mathcal{N}}_h} S_{pq}^{(h)} - M_{[S_{pq}]} \right) \right| \leq \\ &\leq \frac{1}{L} \sum_{|h|<L} W_h \left| \left(\frac{\mathcal{N}_h}{\hat{\mathcal{N}}_h} - 1 \right) S_{pq}^{(h)} \right| + \frac{1}{L} \sum_{|h|<L} W_h \left| S_{pq}^{(h)} - \mathbb{E}S_{pq}^{(h)} \right| + \frac{1}{L} \sum_{|h|<L} W_h \left| \mathbb{E}S_{pq}^{(h)} - M_{[S_{pq}]} \right| \end{aligned} \tag{2.98}$$

The first term on the right hand side of (2.98) is of order $O(T^{-1/2})$, uniformly in $x, y \in [0,1]$, because $(1/T)\mathcal{N}_h = c_h + O_{\mathbb{P}}(T^{-1/2})$ and $(1/T)\hat{\mathcal{N}}_h = c_h + O_{\mathbb{P}}(T^{-1/2})$ uniformly in $|h| \leq L$.

The third term on the right hand side of (2.98) is of order $O(B_R^2)$, uniformly in $x, y \in [0,1]$. This is shown identically as in the proof of Lemma 2.7.3.

The second term on the right hand side of order

$$S_{pq} = \mathbb{E}S_{pq} + O_{\mathbb{P}}\left(\sqrt{\frac{\log T}{TB_R^2}} \right)$$

uniformly in $x, y \in [0,1]$ and $|h| < L$. This is shown analogously as the proof of Lemma 2.7.14. The difference is that the normalising factor $1/L$ improves the rate to $(\log T/(TB_R^2))^{1/2}$ as opposed to $L(\log T/(TB_R^2))^{1/2}$ as in Lemma 2.7.14. □

Proof of Theorem 2.3.5. We start with assuming that the mean function $\mu_X(\cdot)$ is known. Combining the results of Lemmas 2.7.14 and 2.7.15, and the formula (2.78) provides the rate (2.38), and the rate (2.39) if (B6) is assumed.

The proof is completed by the discussion that the difference between the "raw" covariances with known $\mu_X(\cdot)$ versus estimated
$hatmu_X(\cdot)$ is negligible. □

2.7.6 Proof of Theorem 2.3.6

The following lemma ensures the convergence of $\hat{\mathbb{M}}_{X_s|\mathbb{U}_S}$ and $\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}$ to their population level counterparts (2.26). We investigate the convergence without the Gaussianity assumption.

Lemma 2.7.16. *Under the assumptions (A1), (B1) — (B5) and (B7) — (B10),*

$$\sup_{x\in[0,1]} \left|\hat{\mathbb{M}}_{X_s|\mathbb{U}_S}(x) - \mathbb{M}_{X_s|\mathbb{U}_S}(x)\right| = o_{\mathbb{P}}(1) \qquad \text{as} \quad T\to\infty,$$
$$\sup_{x,y\in[0,1]} \left|\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x,y) - \mathbb{S}_{X_s|\mathbb{U}_S}(x,y)\right| = o_{\mathbb{P}}(1) \qquad \text{as} \quad T\to\infty.$$

Proof. We start with $\hat{\mathbb{M}}_{X_s|\mathbb{U}_S}$. Decompose the difference as

$$\begin{aligned}\left|\hat{\mathbb{M}}_{X_s|\mathbb{U}_S} - \mathbb{M}_{X_s|\mathbb{U}_S}\right| \le \underbrace{\left|\hat{\mu}_X(x) - \mu_X(x)\right|}_{J_1} + \\ + \underbrace{\left|\left[P_s\hat{\mathbb{S}}_S\mathbb{H}_S^*\left\{\left(\mathbb{H}_S\hat{\mathbb{S}}_S\mathbb{H}_S^* + \hat{\sigma}^2 I_{\mathcal{N}_1^{(T)}}\right)^{-1} - \left(\mathbb{H}_S\mathbb{S}_S\mathbb{H}_S^* + \sigma^2 I_{\mathcal{N}_1^{(T)}}\right)^{-1}\right\}\left(\mathbb{U}_S - \mathbb{H}_S\hat{\mathbb{M}}_S\right)\right](x)\right|}_{J_2} + \\ + \underbrace{\left|\left\{\left(P_s\hat{\mathbb{S}}_S\mathbb{H}_S^* - P_s\mathbb{S}_S\mathbb{H}_S^*\right)\left(\mathbb{H}_S\mathbb{S}_S\mathbb{H}_S^* + \sigma^2 I_{\mathcal{N}_1^{(T)}}\right)^{-1}\left(\mathbb{U}_S - \mathbb{H}_S\hat{\mathbb{M}}_S\right)\right\}(x)\right|}_{J_3}. \qquad (2.99)\end{aligned}$$

The first term J_1 on the right-hand side of (2.99) tends to zero, uniformly in x, as $T\to\infty$ by Theorem 2.3.1. The second term J_2 and the third term J_3 can be rewritten as

$$\begin{aligned}J_2 = \left|\left[P_s\hat{\mathbb{S}}_S\mathbb{H}_S^*\left\{\left(\mathbb{H}_S\hat{\mathbb{S}}_S\mathbb{H}_S^* + \hat{\sigma}^2 I_{\mathcal{N}_1^{(T)}}\right)^{-1} - \left(\mathbb{H}_S\mathbb{S}_S\mathbb{H}_S^* + \sigma^2 I_{\mathcal{N}_1^{(T)}}\right)^{-1}\right\}\left(\mathbb{U}_S - \mathbb{H}_S\hat{\mathbb{M}}_S\right)\right](x)\right| = \\ = \left|\widehat{\mathrm{cov}}(X_s(x),\mathbb{U}_S)^*\left(\mathrm{var}(\mathbb{U}_S)^{-1} - \widehat{\mathrm{var}}(\mathbb{U}_S)^{-1}\right)\left(\mathbb{U}_S - \mathbb{H}_S\hat{\mathbb{M}}_S\right)\right|\end{aligned}$$

$$J_3 = \left| \left\{ \left(P_s \hat{\mathbb{S}}_S \mathbb{H}_S^* - P_s \mathbb{S}_S \mathbb{H}_S^* \right) \left(\mathbb{H}_S \mathbb{S}_S \mathbb{H}_S^* + \sigma^2 I_{\mathcal{N}_1^{(T)}} \right)^{-1} \left(\mathbb{U}_S - \mathbb{H}_S \hat{\mathbb{M}}_S \right) \right\} (x) \right| =$$
$$= \left| \left\{ \widehat{\mathrm{cov}}(X_s(x), \mathbb{U}_S) - \mathrm{cov}(X_s(x), \mathbb{U}_S) \right\}^* \left(\mathrm{var}(\mathbb{U}_S)^{-1} \right) \left(\mathbb{U}_S - \mathbb{H}_S \hat{\mathbb{M}}_S \right) \right|$$

where $\mathrm{cov}(X_s(x), \mathbb{U}_S)$ is a random vector in $\mathbb{R}^{\mathcal{N}_1^{(S)}}$ whose elements are $\{R_{h_k}(x, x_{t_k,j_k})\}_{k=1}^{N_S}$ for some lags h_k and locations x_{t_k,j_k} and $\mathrm{var}(\mathbb{U}_S)$ is a random matrix in $\mathbb{R}^{\mathcal{N}_1^{(S)} \times \mathcal{N}_1^{(S)}}$ whose elements are of the form $\{R_{t_{k'}-t_k}(x_{t_k,j_k}, x_{t_{k'},j_{k'}})\}_{k,k'=1}^{\mathcal{N}_1^{(S)}}$. The terms $\widehat{\mathrm{cov}}(X_s(x), \mathbb{U}_S)$ and $\widehat{\mathrm{var}}(\mathbb{U}_S)^{-1}$ are defined using the estimated autocovariance kernels.

To treat the term J_2 note that $\widehat{\mathrm{var}}(\mathbb{U}_S)^{-1} - \mathrm{var}(\mathbb{U}_S)^{-1} \to 0$ as $T \to \infty$ by Corollary 2.3.3. The term $\left(\mathbb{U}_S - \mathbb{H}_S \hat{\mathbb{M}}_S \right)$ is bounded as $T \to \infty$ thanks to the convergence $\hat{\mu} \to \mu$. The term $\widehat{\mathrm{cov}}(X_s(x), \mathbb{U}_S)$ is bounded uniformly in x due to its convergence to $\mathrm{cov}(X_s(x), \mathbb{U}_S)$, uniformly in x, by Corollary 2.3.3. The term J_3 is treated similarly. $\widehat{\mathrm{cov}}(X_s(x), \mathbb{U}_S) - \mathrm{cov}(X_s(x), \mathbb{U}_S) \to 0$, uniformly in x, by Corollary 2.3.3. The formula for the variance $\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x, y)$ can be written as

$$\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x, y) = \hat{R}_0(x, y) - \widehat{\mathrm{cov}}(X_s(x), \mathbb{U}_S)^* \widehat{\mathrm{var}}(\mathbb{U}_S)^{-1} \widehat{\mathrm{cov}}(X_s(y), \mathbb{U}_S).$$

Its convergence, uniform in $(x, y) \in [0,1]^2$, is treated similarly as above by Corollary 2.3.3. □

Proof of Theorem 2.3.6. The first statement of Lemma 2.7.16 is the statement of Theorem 2.3.6. □

2.7.7 Proof of Theorem 2.3.7

Proof of Theorem 2.3.7 . We start with the pointwise confidence band. Fix $x \in [0,1]$. From (A2) and the conditional distribution

$$\frac{X_s(x) - \mathbb{M}_{X_s|\mathbb{U}_S}(x)}{\sqrt{\mathbb{S}_{X_s|\mathbb{U}_S}(x, x)}} \sim N(0,1).$$

Therefore

$$\mathbb{P}\left\{ \left| X_s(x) - \mathbb{M}_{X_s|\mathbb{U}_S}(x) \right| \leq \Phi^{-1}(1 - \alpha/2) \sqrt{\mathbb{S}_{X_s|\mathbb{U}_S}(x, x)} \right\} = 1 - \alpha.$$

By Lemma 2.7.16 and Slutsky theorem,

$$\frac{X_s(x) - \hat{\mathbb{M}}_{X_s|\mathbb{U}_S}(x)}{\sqrt{\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x, x)}} \xrightarrow{d} N(0,1)$$

and thus

$$\mathbb{P}\left\{ \left| X_s(x) - \hat{\mathbb{M}}_{X_s|\mathbb{U}_S}(x) \right| \leq \Phi^{-1}(1 - \alpha/2) \sqrt{\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x, x)} \right\} \to 1 - \alpha.$$

Now we turn our attention to the simultaneous confidence band. By the definition of the

conditional distribution,

$$X_s - \mathbb{M}_{X_s|\mathbb{U}_S} \sim N(0, \mathbb{S}_{X_s|\mathbb{U}_S}).$$

By the definition of the simultaneous confidence bands (Degras, 2011), which was reviewed in Section 2.2,

$$\mathbb{P}\left\{\forall x \in [0,1] : \left|X_s(x) - \mathbb{M}_{X_s|\mathbb{U}_S}(x)\right| \leq z_{\alpha,\rho}\sqrt{\mathbb{S}_{X_s|\mathbb{U}_S}(x,x)}\right\} = 1-\alpha.$$

Define the correlation kernel $\rho_{X_s|\mathbb{U}_T}(x,y)$ as in (2.29). Assume for simplicity of the proof that $\rho_{X_s|\mathbb{U}_T}(x,x) > 0$ for all $x \in [0,1]$. Then

$$\frac{X_s(\cdot) - \mathbb{M}_{X_s|\mathbb{U}_S}(\cdot)}{\sqrt{\mathbb{S}_{X_s|\mathbb{U}_S}(\cdot,\cdot)}} \sim N\left(0, \rho_{X_s|\mathbb{U}_T}\right)$$

where the square root and the division is understood pointwise. Denote W_ρ the law of $\sup_{x\in[0,1]}|Z_\rho|$ where $Z_\rho \sim N(0,\rho)$. Then

$$\sup_{x\in[0,1]}\left|\frac{X_s(x) - \mathbb{M}_{X_s|\mathbb{U}_S}(x)}{\sqrt{\mathbb{S}_{X_s|\mathbb{U}_S}(x,x)}}\right| \sim W_{\rho_{X_s|\mathbb{U}_T}}$$

By Lemma 2.7.16,

$$\sup_{x\in[0,1]}\left|\frac{X_s(x) - \hat{\mathbb{M}}_{X_s|\mathbb{U}_S}(x)}{\sqrt{\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x,x)}}\right| \xrightarrow{d} W_{\rho_{X_s|\mathbb{U}_T}}.$$

Note also that if $\rho_n \to \rho$ uniformly then $N(0,\rho_n) \to N(0,\rho)$ weakly, $W_{\rho_n} \to W_\rho$ weakly and therefore $z_{\alpha,\rho_n} \to z_{\alpha,\rho}$. Therefore:

$$\mathbb{P}\left\{\sup_{x\in[0,1]}\left|\frac{X_s(x) - \hat{\mathbb{M}}_{X_s|\mathbb{U}_S}(x)}{\sqrt{\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x,x)}}\right| \leq z_{\alpha,\hat{\rho}}\right\} = \mathbb{P}\left\{\sup_{x\in[0,1]}\left|\frac{X_s(x) - \hat{\mathbb{M}}_{X_s|\mathbb{U}_S}(x)}{\sqrt{\hat{\mathbb{S}}_{X_s|\mathbb{U}_S}(x,x)}}\right| \frac{z_{\alpha,\rho}}{z_{\alpha,\hat{\rho}}} \leq z_{\alpha,\rho}\right\} \to 1-\alpha,$$

as $T \to \infty$.

□

3 Lagged functional regression with sparse noisy observations

The analysis of a regression link between two time series has been widely studied having its origins in the work by Kolmogoroff (1941) and Wiener (1950). Generally speaking, the lagged regression model of two discrete-time stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Z_t\}_{t\in\mathbb{Z}}$ with values in some vector spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ is given by the equation

$$Z_t = a + \sum_{k\in\mathbb{Z}} B_k X_{t-k} + e_t, \quad t \in \mathbb{Z}, \tag{3.1}$$

where $a \in \mathscr{H}_2$ is a constant, called the *intercept*, $\{e_t\}_{t\in\mathbb{Z}}$ is an $\mathscr{H}_2$-valued sequence of disturbances, and $B_k : \mathscr{H}_1 \to \mathscr{H}_2$, $k \in \mathbb{Z}$ is a sequence of linear mappings, called the *filter coefficients*. This linear coupling would be the typical dependence model, for instance, if $\{(X_t, Z_t)\}_{t\in\mathbb{Z}}$ were a jointly Gaussian stationary process in $\mathscr{H}_1 \times \mathscr{H}_2$, and is also known as a time-invariant linearly filtered time series model. The estimation problem is then to estimate the unknown transformations $\{B_k\}_{k\in\mathbb{Z}}$ given the realisation of a finite stretch of the joint series $\{(X_t, Z_t)\}_{t\in\mathbb{Z}}$ (a problem also known as system identification, particularly in signal processing).

The problem is well understood in the case where $\mathscr{H}_1$ and $\mathscr{H}_2$ are Euclidean spaces of possibly different dimensions (the dimension 1 relates to the scalar time series scenario) (Brillinger, 1983; Priestley, 1981a,b; Shumway and Stoffer, 2000). The generalisation to the setting with infinite dimensional spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ is not straightforward, in particular the infinite dimensionality of the input time series $\{X_t\}$ is problematic and leads to an ill-posed inverse problem, more in Subsection 1.1.4. The lagged regression problem when $\mathscr{H}_1$ and $\mathscr{H}_2$ are infinite dimensional Hilbert spaces, corresponding to the situation of functional data, has been presented only recently (Hörmann et al., 2015b; Pham and Panaretos, 2018) and we review their approaches in Sections 3.1.

In section 3.2 we present the main result of this chapter, the set-up with the regressors time series with values in $\mathscr{H}_1 = L^2([0,1],\mathbb{R})$ but which is sparsely observed. The response time series can be either a sparsely observed functional time series with values in $\mathscr{H}_2 = L^2([0,1],\mathbb{R})$, or a univariate time series with values in $\mathscr{H}_2 = \mathbb{R}$. We show how to estimate the filter coefficients from sparse data and how to forecast the response process from the sparse observations.

Section 3.3 extends the methodology to the models with multiple input regressors time series that can be either fully observed or sparsely observed functional time series.

The main results of this chapter are based on Rubín and Panaretos (2020a) while extending some results further.

3.1 Spectral analysis of the functional lagged regression model

3.1.1 Frequency response function and estimation equations

We consider the functional lagged regression equation (3.1) where $\{X_t\}$ is a stationary functional time series with finite second moments and with values in $\mathscr{H}_1$. Moreover it is assumed to satisfy the weak dependence condition (A1) introduced in Subsection 2.1.1 and therefore, by Proposition 1.3.5, it admits the spectral density operator given by

$$\mathscr{F}_\omega^X = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \mathscr{R}_h^X e^{-\mathrm{i}h\omega}, \qquad \omega\in[-\pi,\pi].$$

The mappings $B_k : \mathscr{H}_1 \to \mathscr{H}_2$, $k\in\mathbb{Z}$ are considered to be Hilbert-Schmidt operators satisfying the summability condition

$$\sum_{k\in\mathbb{Z}} \|B_k\|_{\mathscr{L}_2(\mathscr{H}_1,\mathscr{H}_2)} < \infty. \tag{3.2}$$

Thanks to the inequality $\|B_k\|_{\mathscr{L}(\mathscr{H}_1,\mathscr{H}_2)} \le \|B_k\|_{\mathscr{L}_2(\mathscr{H}_1,\mathscr{H}_2)}$, the filter $\{B_k\}_k$ with its frequency response function given by $\mathscr{B}_\omega = \sum_{k\in\mathbb{Z}} B_k e^{-\mathrm{i}h\omega}$ satisfies the assumptions of Proposition 1.3.15, and the $\mathscr{H}_2$-valued functional time series $\{Z_t\}_t$ is stationary.

Moreover, the spectral density of $\{Z_t\}$ and the cross-spectral density between $\{Z_t\}$ and $\{X_t\}$ are given by

$$\begin{aligned}\mathscr{F}_\omega^Z &= \mathscr{B}_\omega \mathscr{F}_\omega^X \mathscr{B}_\omega^*, \qquad \omega\in[-\pi,\pi],\\ \mathscr{F}_\omega^{ZX} &= \mathscr{B}_\omega \mathscr{F}_\omega^X, \qquad \omega\in[-\pi,\pi].\end{aligned} \tag{3.3}$$

It is the equation (3.3) that constitutes the foundation of the estimation of the filter coefficients. Heuristically, one would want to solve (3.3) and invert $\mathscr{F}_\omega^X$ for each $\omega\in[-\pi,\pi]$. However, this inversion is not well defined in the case of the infinite dimensional $\mathscr{H}_1$. Indeed, the spectral density operator $\mathscr{F}_\omega^X$ is trace-class, hence its spectral decomposition is of the form

$$\mathscr{F}_\omega^X = \sum_{j=1}^{\infty} \lambda_j(\omega)\varphi_j(\omega)\otimes\varphi_j(\omega)$$

where the harmonic eigenvalues $\lambda_j(\omega)\to 0$ as $j\to\infty$. Therefore it is obvious that $(\mathscr{F}_\omega^X)^{-1}$ cannot be well defined as a bounded linear operator.

Nevertheless, one can give a rigorous solution to (3.3) by writing it as

$$\mathscr{B}_\omega(f) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j(\omega)} \left\langle f, \varphi_j(\omega) \right\rangle_{\mathscr{H}_1} \mathscr{F}_\omega^{ZX}\left(\varphi_j(\omega)\right), \qquad f \in \mathscr{H}_1, \quad \omega \in [-\pi,\pi]. \tag{3.4}$$

This fact, however, does not change the intrinsic ill-posedness of the inverse problem (3.3) and regularisation is required when plugging in the estimated values of $\mathscr{F}_\omega^X$ and $\mathscr{F}_\omega^{ZX}$.

3.1.2 Regularised estimation of frequency response function

To start with the presentation of the regularised inversion of (3.3), assume that we posses estimates of the spectral density operators $\{\hat{\mathscr{F}}_\omega^X\}_{\omega\in[-\pi,\pi]}$ and the cross-spectral density operators $\{\hat{\mathscr{F}}_\omega^{ZX}\}_{\omega\in[-\pi,\pi]}$. In the case of fully observed functional data, such estimators can be obtained by the means reviewed in Subsection 1.3.9. The case of sparsely observed functional data is developed in detail in Section 3.2.

Being a self-adjoint trace class operator, $\hat{\mathscr{F}}_\omega^X$ admits the spectral representation

$$\hat{\mathscr{F}}_\omega^X = \sum_{j=1}^{\infty} \hat{\lambda}_j^\omega \hat{\varphi}_j^\omega \otimes \hat{\varphi}_j^\omega, \qquad \omega \in [-\pi,\pi],$$

which can be viewed as the empirical version of the decomposition (1.36). The difficulty in inverting $\hat{\mathscr{F}}_\omega^X$ can be seen from the fact that $\sum_j \hat{\lambda}_j^\omega = \mathrm{Tr}\{\hat{\mathscr{F}}_\omega^X\} < \infty$, implying that $\hat{\lambda}_j^\omega$ decays at least as fast as $j^{-(1+\delta)}$, $\delta > 0$. It is the small values of $\hat{\lambda}_j^\omega$ that cause problems and there are two classical strategies to overcome the issue: *Tikhonov regularisation* and the *spectral truncation*.

Spectral truncation regularisation

Hörmann et al. (2015b) suggested to replace the inverse of $\hat{\mathscr{F}}_\omega^X$ is replaced by

$$\sum_{j=1}^{K_\omega} \frac{1}{\hat{\lambda}_j^\omega} \hat{\varphi}_j^\omega \otimes \hat{\varphi}_j^\omega, \qquad \omega \in [-\pi,\pi],$$

where $K_\omega \in \mathbb{N}$, $\omega \in [-\pi,\pi]$, is the spectral truncation parameter that needs to grow to infinity sufficiently slowly to allow for the consistency. It may or may not depend on the frequency $\omega \in [-\pi,\pi]$. We opt to implement the spectral truncation by relying on the *eigenvalue thresholding* approach (Hörmann et al., 2015b) where we implement the eigenvalue threshold selection by cross-validation (more in Section 4.2).

The estimator of the spectral transfer function becomes

$$\hat{\mathscr{B}}_\omega^{trunc} = \sum_{j=1}^{K_\omega} \frac{1}{\hat{\lambda}_j^\omega} \left\langle \hat{\varphi}_j^\omega, \cdot \right\rangle \hat{\mathscr{F}}_\omega^{ZX} \hat{\varphi}_j^\omega, \qquad \omega \in [-\pi,\pi]. \tag{3.5}$$

Tikhonov regularisation

Pham and Panaretos (2018) proposed to replace the inverse of $\hat{\mathscr{F}}_\omega^X$ by

$$\left(\hat{\mathscr{F}}_\omega^X + \rho \mathrm{I}\right)^{-1} = \sum_{j=1}^{\infty} \frac{1}{\hat{\lambda}_j^\omega + \rho} \hat{\varphi}_j^\omega \otimes \hat{\varphi}_j^\omega, \qquad \omega \in [-\pi, \pi],$$

where I is the identity operator on $\mathscr{H}$ and the Tikhonov regularisation parameter $\rho > 0$ tends to zero as $T \to \infty$ slowly enough to allow for consistency. Even though the parameter ρ may, in general, depend on ω we carry out further analysis with the frequency independent parameter. We do so because in the implementation (more in Section 4.2) we select the tuning parameter ρ using the cross-validation where it is feasible to optimize over a single (frequency independent) tuning parameter.

The estimator of the spectral transfer function becomes

$$\hat{\mathscr{B}}_\omega^{Tikh} = \hat{\mathscr{F}}_\omega^{ZX} \left(\hat{\mathscr{F}}_\omega^X + \rho \mathrm{I}\right)^{-1} = \sum_{j=1}^{\infty} \frac{1}{\hat{\lambda}_j^\omega + \rho} \left\langle \hat{\varphi}_j^\omega, \cdot \right\rangle \hat{\mathscr{F}}_\omega^{ZX} \hat{\varphi}_j^\omega, \qquad \omega \in [-\pi, \pi]. \tag{3.6}$$

Estimation of filter coefficients

Once the estimators of the spectral transfer operator $\hat{\mathscr{B}}_\omega$ have been constructed by either of the above regularisation techniques, the filter coefficients are estimated by

$$\hat{B}_k^{trunc} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\mathscr{B}}_\omega^{trunc} e^{\mathrm{i}\omega k} \, \mathrm{d}\omega, \qquad k \in \mathbb{Z}, \tag{3.7}$$

$$\hat{B}_k^{Tikh} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\mathscr{B}}_\omega^{Tikh} e^{\mathrm{i}\omega k} \, \mathrm{d}\omega, \qquad k \in \mathbb{Z}. \tag{3.8}$$

3.2 Lagged functional regression with sparsely observed regressors

3.2.1 Estimation in sparsely observed regime

Consider the functional time series of regressors $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ as a time series of continuous curves in the Hilbert space $\mathscr{H}_1 = L^2([0,1], \mathbb{R})$ which is observed only sparsely by the observation scheme introduced in Subsection 2.1.1 for times $t = 1, \ldots, T$ where $T \in \mathbb{N}$. The response time series $\{Z_t\}_{t\in\mathbb{Z}}$ is considered to be a functional time series, observed sparsely or fully functionally, or univariate time series. In any of the set-ups, the cross-spectral density operator $\{\mathscr{F}_\omega^{ZX}\}_{\omega\in[-\pi,\pi]}$ can be estimated by the one of the estimators established in Section 2.4, while the spectral density operators $\{\mathscr{F}_\omega^X\}_{\omega\in[-\pi,\pi]}$ are estimated by the tools developed in Section 2.1.

Once the estimators $\{\hat{\mathscr{F}}_\omega^X\}_{\omega\in[-\pi,\pi]}$ and $\{\hat{\mathscr{F}}_\omega^{ZX}\}_{\omega\in[-\pi,\pi]}$ are constructed, we may produce the estimate of the frequency response function $\mathscr{B}_\omega$ by either the Tikhonov regularisation or the

spectral truncation regularisation (Subsection 3.1.2). For either of the regularisation techniques, denote the estimated filter coefficients $\hat{B}_k$, be it in the set-up of functional response ($\mathscr{H}_2 = L^2([0,1],\mathbb{R})$) in which case the filter coefficients are operators in $\mathscr{L}(L^2([0,1],\mathbb{R}))$, or in the set-up of univariate response ($\mathscr{H}_2 = \mathbb{R}^d$) in which case the filter coefficients are operators in $\mathscr{L}(L^2([0,1],\mathbb{R}),\mathbb{R})$ or equivalently the functionals on $L^2([0,1],\mathbb{R})$.

3.2.2 Forecasting the response process

Having the estimators of the filter coefficients $\{\hat{B}_k\}_{k\in\mathbb{Z}}$ we may wish to predict the values of the response process $\{Z_t\}$ from the sparsely observed functional time series of regressors $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$. The objectives of the prediction may include various set-ups:

- Both the sparsely observed regressor functional time series $\{X_t\}$ and the response time series $\{Z_t\}$ are available at times $t = 1,\dots,T$. Once we estimate the spectral and the cross-spectral density operators $\{\hat{\mathscr{F}}_\omega^X\}$ and $\{\hat{\mathscr{F}}_\omega^{ZX}\}$ and construct the estimates of the filter coefficients $\{\hat{B}_k\}_{k\in\mathbb{Z}}$, we may wish to test the goodness of fit and construct the predictions of $\hat{Z}_1,\dots,\hat{Z}_T$ produced only from the sparse measurements on $\{X_t\}$ and compare them with the true $Z_1,\dots,Z_T$.
- The response time series $\{Z_t\}$ is observed up to time $S \in \mathbb{N}$ while the measurements realised on the sparsely observed regressors time series $\{X_t\}$ continue up to time $T > S$. In this case, we may want to forecast the time series $\{Z_t\}$ beyond the observation horizon S and predict the values $Z_{S+1},\dots,Z_S$ from the estimated filter coefficients $\{\hat{B}_k\}$ and the regressors $\{X_t\}$ observed sparsely beyond the horizon S.

Assume that we have constructed the estimates of the spectral density operators $\{\mathscr{F}_\omega^X\}$ and $\{\mathscr{F}_\omega^{ZX}\}$, and the filter coefficients $\{\hat{B}_k\}$. We have access to the sparse measurements

$$U_{tj} = X_t(x_{tj}) + \epsilon_{tj}^X, \qquad j = 1,\dots,N_t^X, \quad t = 1,\dots,T, \tag{3.9}$$

and we wish to construct the prediction of $Z_1,\dots,Z_T$. In the following we show how to actually construct the best linear unbiased predictors of Z_t, $t = 1,\dots$ given all the observed data $\{U_{tj}\}$. A key observation that simplifies the forecasting is the fact that we may predict the latent functional data first, and then plug them into the filter coefficients. The following proposition summarises this assertion formally.

Proposition 3.2.1. *The best linear unbiased predictor of Z_s given $\mathbb{U}$, denoted as $\Pi(Z_s|\mathbb{U})$, is equivalent to constructing the best linear unbiased predictors of X_t given $\mathbb{U}$, denoted as $\Pi(X_t|\mathbb{U})$, for all $t \in \mathbb{Z}$ and then applying the filter coefficients $\{B_k\}_{k\in\mathbb{Z}}$ to these predictions:*

$$\Pi(Z_s|\mathbb{U}) = \sum_{k\in\mathbb{Z}} B_k \Pi(X_{s-k}|\mathbb{U}), \qquad s \in \mathbb{Z}.$$

Proof. The proof follows directly from the formula for the best linear unbiased predictors and

the theory reviewed in Subsection 2.2.1.

$$\begin{aligned}\Pi(Z_s|\mathbb{U}_T) &= \operatorname{cov}(Z_s,\mathbb{U}_T)\,(\operatorname{var}(\mathbb{U}_T))^{-1}\,\mathbb{U}_T\\ &= \operatorname{cov}\left(\sum_{k\in\mathbb{Z}} B_k X_{s-k} + e_s, \mathbb{U}_T\right)(\operatorname{var}(\mathbb{U}_T))^{-1}\,\mathbb{U}_T\\ &= \sum_{k\in\mathbb{Z}} B_k \operatorname{cov}(X_{s-k},\mathbb{U}_T)\,(\operatorname{var}(\mathbb{U}_T))^{-1}\,\mathbb{U}_T\\ &= \sum_{k\in\mathbb{Z}} B_k \Pi(X_{s-k}|\mathbb{U}_T).\end{aligned}$$

□

The predictors $\Pi(X_s|\mathbb{U})$ for $t=-M+1,\dots,T+M$ can be constructed by the methods given in Section 2.2 with the estimated dynamics of the process X manifested by the spectral density kernels estimator $\{\hat{f}^X_\omega(\cdot,\cdot)\}_{\omega\in[-\pi,\pi]}$ and the estimated measurement error term $\hat{\sigma}^2_X$.

In summary, the forecasting algorithm consists of the following steps:

1. From the measurements $\mathbb{U}$ realised on the regressor time series $\{X_t\}$ estimate the spectral density $\{\hat{\mathscr{F}}^X_\omega\}_{\omega\in[-\pi,\pi]}$ and the measurement error variance $\hat{\sigma}^2$. Using the formula (2.11), integrate the estimated spectral density to obtain the complete space time covariance $\{\hat{\mathscr{R}}^X_h\}_{h\in\mathbb{Z}}$ of the regressor time series $\{X_t\}$.
2. From the measurements $\mathbb{U}$ and the observed response times series (be it $Z_1,\dots,Z_S$ or $Z_1,\dots,Z_T$), estimate the cross-spectral density $\{\hat{\mathscr{F}}^{ZX}_\omega\}_{\omega\in[-\pi,\pi]}$. Using either truncation regularisation (3.5) or Tikhonov regularisation (3.6), estimate the spectral transfer function $\{\mathscr{B}_\omega\}_{\omega\in[-\pi,\pi]}$. By means of formula (3.7) or (3.8), depending on the regularisation scheme, integrate the spectral transfer function to obtain the filter coefficients $\{\hat{B}_k\}_{k\in\mathbb{Z}}$.
3. Choose M such that the estimated filter coefficients $\hat{B}_k$ are negligible for $|k|>M$ and construct the prediction of the latent functional data $\hat{\Pi}(X_{-M+1}|\mathbb{U}),\dots,\hat{\Pi}(X_{T+M}|\mathbb{U})$.
4. For each $t=1,\dots,T$, construct the forecasts

$$\hat{\Pi}(Z_t|\mathbb{U}) = \sum_{k=-M}^{M} \hat{B}_k \hat{\Pi}(X_{t-k}|\mathbb{U}).$$

3.2.3 Asymptotic results

In this subsection we list the conditions for the consistent filter coefficients estimation obtained via Tikhonov regularisation (3.8) and via truncation regularisation (3.7) and prove the said consistency in the two set-ups:

(i) The regressors functional time series $\{X_t(x) : x\in[0,1]\}_{t\in\mathbb{Z}}$ is sparsely observed and the response time series $Z\equiv\{Z_t\}_{t\in\mathbb{Z}}$ is univariate.

(ii) Both the regressors time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ and the response time series $\{Z_t(z) : z \in [0,1]\}_{t\in\mathbb{Z}}$ are sparsely observed functional time series.

The conditions for the consistent estimation of the spectral and cross-spectral density operators are given in Chapter 2, therefore we proceed now with the conditions specific for the filter coefficient estimation by the two considered regularisation techniques.

First, the following condition is required for the regression model (3.1) to be identifiable, regardless of the regularisation method used.

(E) For all $\omega \in [-\pi,\pi]$ the operators $\mathscr{F}^X_\omega : \mathscr{H}_1 \to \mathscr{H}_1$ satisfy $\ker\left(\mathscr{F}^X_\omega\right) = 0$.

To ensure the consistency of the filter coefficients estimator by the Tikhonov method we only need to guarantee that the regularisation parameter vanishes slowly relative to the smoothing bandwidths. Denote B_R the bandwidth for the smoother estimator $\hat{\mathscr{F}}^X_\omega$ and $\tilde{B}_R$ the bandwidth for the smoother estimator $\hat{\mathscr{F}}^{ZX}_\omega$.

(F:i) The Tikhonov regularisation parameter satisfies

$$\frac{1}{\rho}L\frac{1}{\sqrt{T}}\frac{1}{\tilde{B}_R} \to 0, \qquad \text{as} \quad T\to\infty,$$
$$\frac{1}{\rho^2}L\frac{1}{\sqrt{T}}\frac{1}{B_R^2} \to 0, \qquad \text{as} \quad T\to\infty.$$

(F:ii) The Tikhonov regularisation parameter satisfies

$$\frac{1}{\rho}L\frac{1}{\sqrt{T}}\frac{1}{\tilde{B}_R^2} \to 0, \qquad \text{as} \quad T\to\infty,$$
$$\frac{1}{\rho^2}L\frac{1}{\sqrt{T}}\frac{1}{B_R^2} \to 0, \qquad \text{as} \quad T\to\infty.$$

The following theorem establishes the consistency of the Tikhonov filter coefficient estimators.

Theorem 3.2.2. *Let the regressor time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ be a functional time series in $\mathscr{H}_1 = L^2([0,1],\mathbb{R})$ satisfying (A1), (B1) — (B10) (listed in Chapter 2) and (E). Moreover:*

*(i) The **univariate response** time series $\{Z_t\}$, i.e. $\mathscr{H}_2 = \mathbb{R}$, satisfies the conditions (B1) — (B4), (B6) with R_h^{ZX},* (B5:ZX)*, and the parameters $\tilde{B}_R$, $\tilde{L}$ and ρ satisfy the rates* (B8:1d), (B10:1d), *(F:i).*

*(ii) The **sparsely observed functional response** time series $\{Z_t(\cdot)\}$, i.e. $\mathscr{H}_2 = L^2([0,1],\mathbb{R})$, satisfies the conditions (B1) — (B4), B5:XY, (B6) with R_h^{ZX}, and the parameters $\tilde{B}_R$, $\tilde{L}$ and ρ satisfy the rates (B8), (B10), (F:ii).*

Then the filter coefficient estimators (3.8) *constructed by means of Tikhonov regularisation are consistent*

$$\sup_{k\in\mathbb{Z}}\left\|\hat{B}_k^{Tikh}-B_k\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}=o_{\mathbb{P}}(1)\qquad as\quad T\to\infty.$$

Note that for the case of univariate response $\mathcal{H}_2=\mathbb{R}$, the norm $\|\cdot\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}$ reduces to the operator norm $\|\cdot\|_{\mathcal{L}(\mathcal{H}_1,\mathbb{R})}$.

Proof. Thanks to the fact that

$$\sup_{k\in\mathbb{Z}}\left\|\hat{B}_k^{Tikh}-B_k\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}\le\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\|\hat{\mathscr{B}}_\omega-\mathscr{B}_\omega\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}\mathrm{d}\omega$$

the proof reduces to establishing the convergence rate of the frequency response operator $\hat{\mathscr{B}}_\omega$.

For the Tikhonov regularisation parameter $\rho=\rho(T)>0$ define

$$\tilde{\mathscr{B}}_\omega=\mathscr{F}_\omega^{ZX}\left(\mathscr{F}_\omega^{X}+\rho\,\mathrm{I}\right)^{-1},\qquad\omega\in[-\pi,\pi].$$

Further, split the desired difference into three terms:

$$\begin{aligned}\left\|\hat{\mathscr{B}}_\omega-\mathscr{B}_\omega\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}&\le\left\|\hat{\mathscr{B}}_\omega-\tilde{\mathscr{B}}_\omega\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}+\left\|\tilde{\mathscr{B}}_\omega-\mathscr{B}_\omega\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}\le\\ &\le\underbrace{\left\|\left(\mathscr{F}_\omega^{ZX}-\hat{\mathscr{F}}_\omega^{ZX}\right)\left(\mathscr{F}_\omega^{X}+\rho\,\mathrm{I}\right)^{-1}\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}}_{\mathscr{S}_1}+\underbrace{\left\|\hat{\mathscr{F}}_\omega^{ZX}\left[\left(\mathscr{F}_\omega^{X}+\rho\,\mathrm{I}\right)^{-1}-\left(\hat{\mathscr{F}}_\omega^{X}+\rho\,\mathrm{I}\right)^{-1}\right]\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}}_{\mathscr{S}_2}+\\ &\qquad+\underbrace{\left\|\tilde{\mathscr{B}}_\omega-\mathscr{B}_\omega\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}}_{\mathscr{S}_3}.\end{aligned}$$

We bound each of the terms $\mathscr{S}_1$, $\mathscr{S}_2$, and $\mathscr{S}_3$ and show the convergence of the bound to zero uniformly in $\omega\in[-\pi,\pi]$. We start with bounding $\mathscr{S}_1$.

$$\mathscr{S}_1\le\left\|\mathscr{F}_\omega^{ZX}-\hat{\mathscr{F}}_\omega^{ZX}\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}\left\|\left(\mathscr{F}_\omega^{X}+\rho\,\mathrm{I}\right)^{-1}\right\|_{\mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)}\le\left\|\mathscr{F}_\omega^{ZX}-\hat{\mathscr{F}}_\omega^{ZX}\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}\frac{1}{\rho}\tag{3.10}$$

Now, bounding $\mathscr{S}_2$:

$$\begin{aligned}\mathscr{S}_2=\left\|\hat{\mathscr{F}}_\omega^{ZX}\left[\left(\mathscr{F}_\omega^{X}+\rho\,\mathrm{I}\right)^{-1}\left(\hat{\mathscr{F}}_\omega^{X}-\mathscr{F}_\omega^{X}\right)\left(\hat{\mathscr{F}}_\omega^{X}+\rho\,\mathrm{I}\right)^{-1}\right]\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}&\le\\ \le\left\|\hat{\mathscr{F}}_\omega^{ZX}\right\|_{\mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)}\frac{1}{\rho}\left\|\hat{\mathscr{F}}_\omega^{X}-\mathscr{F}_\omega^{X}\right\|_{\mathcal{L}_2(\mathcal{H}_1,\mathcal{H}_2)}\frac{1}{\rho}&\end{aligned}\tag{3.11}$$

The right-hand sides of (3.10) tend to zero uniformly in $\omega\in[-\pi,\pi]$ as $T\to\infty$ thanks to assumption (F:i) and Proposition 2.4.2, or (F:ii) and Proposition 2.4.1, for the univariate or functional response respectively. The right-hand side of (3.11) tend to zero uniformly in $\omega\in[-\pi,\pi]$ as $T\to\infty$ thanks to assumption (F:i) (or (F:ii)) and Theorem 2.3.2. Remember also that the results of Proposition 2.4.1 and 2.4.2, and Theorem 2.3.2 are stated in the supremum

norm over the (cross)-spectral density kernels which implies the convergence in the Hilbert-Schmidt norm.

It remains to handle the deterministic term $\mathscr{S}_3$. Since the spectral density operator $\mathscr{F}_\omega^X$ is self-adjoint and trace-class, it admits the series decomposition

$$\mathscr{F}_\omega^X = \sum_{j=1}^{\infty} \lambda_j^\omega \varphi_j^\omega \otimes \varphi_j^\omega = \sum_{j=1}^{\infty} \lambda_j^\omega \left\langle \varphi_j^\omega, \cdot \right\rangle \varphi_j^\omega$$

where $\{\lambda_j^\omega\}_{j\in\mathbb{N}}$ and $\{\varphi_j^\omega\}_{j\in\mathbb{N}}$ are the harmonic (dynamic) eigenvalues and the harmonic (dynamic) eigenfunctions at frequency $\omega \in [-\pi,\pi]$ (Panaretos and Tavakoli, 2013a; Hörmann et al., 2015a). From the relations (3.3) and (3.4) we obtain the following series expansions

$$\begin{aligned}
\mathscr{F}_\omega^{ZX} &= \sum_{j=1}^{\infty} \lambda_j^\omega \left\langle \varphi_j^\omega, \cdot \right\rangle \mathscr{B}_\omega \varphi_j^\omega, \\
\mathscr{F}_\omega^{ZX} \left(\mathscr{F}_\omega^X\right)^{-1} &= \sum_{j=1}^{\infty} \left\langle \varphi_j^\omega, \cdot \right\rangle \mathscr{B}_\omega \varphi_j^\omega, \\
\mathscr{F}_\omega^{ZX} \left(\mathscr{F}_\omega^X + \rho \mathrm{I}\right)^{-1} &= \sum_{j=1}^{\infty} \frac{\lambda_j^\omega}{\lambda_j^\omega + \rho} \left\langle \varphi_j^\omega, \cdot \right\rangle \mathscr{B}_\omega \varphi_j^\omega,
\end{aligned}$$

Combining the above expansions yields

$$\begin{aligned}
\mathscr{S}_3^2 = \left\| \tilde{\mathscr{B}}_\omega - \mathscr{B}_\omega \right\|_{\mathscr{L}_2(\mathscr{H}_1,\mathscr{H}_2)}^2 &\leq \sum_{j=1}^{\infty} \left\| \left(\frac{\lambda_j^\omega}{\lambda_j^\omega + \rho} - 1 \right) \left\langle \varphi_j^\omega, \cdot \right\rangle \mathscr{B}_\omega \varphi_j^\omega \right\|^2 \leq \\
&\leq \sum_{j=1}^{\infty} \left(\frac{\lambda_j^\omega}{\lambda_j^\omega + \rho} - 1 \right)^2 \left\| \mathscr{B}_\omega \varphi_j^\omega \right\|^2 = \sum_{j=1}^{\infty} \left(\frac{\rho}{\lambda_j^\omega + \rho} \right)^2 \left\| \mathscr{B}_\omega \varphi_j^\omega \right\|^2. \quad (3.12)
\end{aligned}$$

Since $\sum_{j=1}^{\infty} \left\| \mathscr{B}_\omega \varphi_j^\omega \right\|^2 = \|\mathscr{B}_\omega\|_{\mathscr{L}_2(\mathscr{H}_1,\mathscr{H}_2)}^2 < \infty$ and $\rho/(\lambda_j^\omega + \rho) \to 0$ as $\rho \searrow 0$ for each $j \in \mathbb{N}$, the right-hand side of (3.12) tends to zero, completing the proof. □

We now turn to the truncation estimator (3.7) of the filter coefficients, whose consistency requires more technical assumptions. We use the result by Hörmann et al. (2015b, Thm 1) relies on having consistent estimators of the spectral density and cross-spectral density operators with a known rate of convergence, on a condition on the eigenvalue spacing, and on an assumption that the spectral truncation parameter K_ω grows sufficiently slowly. In what follows, we review their conditions and adapt them to the setting when the spectral density kernels and the cross-spectral density are estimated by the kernel smoothing methods from sparse noisy observations.

Recall the eigendecomposition of the spectral frequency operator

$$\mathscr{F}_\omega^X = \sum_{j=1}^{\infty} \lambda_j^\omega \varphi_j^\omega \otimes \varphi_j^\omega = \sum_{j=1}^{\infty} \lambda_j^\omega \left\langle \varphi_j^\omega, \cdot \right\rangle \varphi_j^\omega$$

with its harmonic eigenvalues and harmonic eigenfunction are denoted $\{\lambda_k^\omega\}_{k=1}^\infty$ and $\{\varphi_k^\omega\}_{k=1}^\infty$ respectively. Define

$$\begin{aligned}\Lambda_1^\omega &= \lambda_1^\omega - \lambda_2^\omega,\\ \Lambda_k^\omega &= \min\{\lambda_k^\omega - \lambda_{k+1}^\omega, \lambda_{k-1}^\omega - \lambda_k^\omega\}, \qquad k \geq 2.\end{aligned}$$

The following condition guarantees that the eigenspaces belonging to each of the eigenvalues $\{\lambda_m^\omega\}_{m=1}^\infty$ are one-dimensional, hence the eigenfunctions $\{\varphi_m^\omega\}_{m=1}^\infty$ can be identified (up to multiplication by a complex number with modulus 1).

(G1) For all $k \geq 1$ we assume $\inf_{\omega\in[-\pi,\pi]} \Lambda_k^\omega > 0$.

Furthermore we need to assume that the truncation parameter K_ω grows to infinity as $T \to \infty$ but does so sufficiently slowly.

(G2) Set $r = 1$ if $\{Z_t\}$ is a univariate time series and $r = 2$ if it is a sparsely observed functional and

$$\begin{aligned}K_\omega &= \min\{K^{(i)}, 1 \leq i \leq 4\},\\ K^{(1)} &= \max\left\{k \geq 1 : \inf_{\omega\in[-\pi,\pi]} \hat{\lambda}_k \geq 2LT^{-1/2}B_R^{-2}\right\},\\ K^{(2)} &= \max\left\{k \geq 1 : LT^{-1/2}\tilde{B}_R^{-r}\int_{-\pi}^{\pi} W_\lambda^K(\omega)\,\mathrm{d}\omega \leq 1\right\},\\ K^{(3)} &= \max\left\{k \geq 1 : \int_{-\pi}^{\pi}\left(W_\lambda^k(\omega)\right)^2\mathrm{d}\omega \leq L^{-1/2}T^{1/4}B_R\right\},\\ K^{(4)} &= \max\left\{k \geq 1 : \int_{-\pi}^{\pi}\left(W_\Lambda^k(\omega)\right)^2\mathrm{d}\omega \leq L^{-1/2}T^{1/4}B_R\right\}\end{aligned}$$

where

$$W_\lambda^k(\omega) = \left(\sum_{m=1}^k \frac{1}{\left[\hat{\lambda}_m^\omega\right]^2}\right)^{1/2}, \qquad W_\Lambda^k(\omega) = \left(\sum_{m=1}^k \frac{1}{\left[\hat{\Lambda}_m^\omega\right]^2}\right)^{1/2}$$

and $\{\hat{\Lambda}_m^\omega\}$ are the empirical counterparts of $\{\Lambda_m^\omega\}$ with the estimates $\{\hat{\lambda}_j^\omega\}$ plugged-in.

Under the above stated assumptions, the filter coefficient estimator (3.7) obtained by means of truncation regularisation is consistent.

Theorem 3.2.3. *Let the regressor time series $\{X_t(x) : x \in [0,1]\}_{t\in\mathbb{Z}}$ be a functional time series in $\mathscr{H}_1 = L^2([0,1],\mathbb{R})$ satisfying (A1), (B1) — (B10) (listed in Chapter 2), (E) and (G1). Moreover:*

*(i) The **univariate response** time series $\{Z_t\}$, i.e. $\mathscr{H}_2 = \mathbb{R}$, satisfies the conditions (B1) — (B4), (B6) with R_h^{ZX},* (B5:ZX)*, and the parameters $\tilde{B}_R$, $\tilde{L}$ and K_ω satisfy the rates* (B8:1d), (B10:1d)*, (G2).*

*(ii) The **sparsely observed functional response** time series $\{Z_t(\cdot)\}$, i.e. $\mathscr{H}_2 = L^2([0,1],\mathbb{R})$, satisfies the conditions (B1) — (B4), B5:XY, (B6) with R_h^{ZX}, and the parameters $\tilde{B}_R$, $\tilde{L}$ and K_ω satisfy the rates (B8), (B10), (G2).*

Then the filter coefficient estimators (3.8) *constructed by means of Tikhonov regularisation are consistent*

$$\sup_{k\in\mathbb{Z}} \left\| \hat{B}_k^{trunc} - B_k \right\|_{\mathscr{L}_2(\mathscr{H}_1,\mathscr{H}_2)} = o_p(1) \qquad as \quad T\to\infty.$$

Proof. The proof of this theorem is an application of the result by Hörmann et al. (2015b, Thm 1). Their theorem requires that the spectral density operators $\{\mathscr{F}_\omega^X\}_{\omega\in[-\pi,\pi]}$ are estimated with a certain rate, say $(\psi_T^X)_{T=1}^\infty$, and the cross-density operators $\{\mathscr{F}_\omega^{ZX}\}_{\omega\in[-\pi,\pi]}$ with another rate, say $(\psi_T^{ZX})_{T=1}^\infty$. For $T\in\mathbb{N}$ we put $\psi_T^X = LT^{-1/2}B_R^{-2}$ and $\psi_T^{ZX} = LT^{-1/2}\tilde{B}_R^{-r}$ by Theorem 2.3.2 and Proposition 2.4.1 (or Proposition 2.4.2) respectively. Note that the convergence rates for the spectral density kernels $\{f_\omega^X\}_{\omega\in[-\pi,\pi]}$ and the cross-spectral density kernels $\{f_\omega^{ZX}\}_{\omega\in[-\pi,\pi]}$ are in the supremum norm which is stronger than the operator norm of the assumptions of Hörmann et al. (2015b, Thm 1).

We can now replicate all the steps of the proof of Hörmann et al. (2015b, Thm 1) and see that they only require the above stated rates and the assumptions (E), (G1), and (G2).

It is worth noting that the results by Hörmann et al. (2015b, Thm 1) are derived under the assumption of L^p-m-approximability as opposed to cumulant mixing conditions considered in this chapter. Nevertheless, they use the assumption of L^p-m-approximability only to prove the rate of convergence of the spectral density and cross-spectral density estimators in (Hörmann et al., 2015b, Lemma 1). Once these rates are established, the proof of (Hörmann et al., 2015b, Thm 1) does not use L^p-m-approximability. Therefore our proof of Theorem 3.2.3 is indeed its simple adaptation which takes as inputs the convergence rates $\psi_T^X = LT^{-1/2}B_R^{-2}$ and $\psi_T^{ZX} = LT^{-1/2}\tilde{B}_R^{-r}$ proved under the cumulant mixing conditions. □

3.3 Lagged regression with multiple inputs

3.3.1 Joint regression model and its spectral analysis

It may very well happen that we want to analyse the relationship between multiple time series, including functional time series observed either sparsely or fully, multivariate time series or univariate time series. To keep the notation simple we focus our presentation on the setting of a scalar response and two functional time series regressors, one of which is observed sparsely and the other fully. This setting corresponds to the model **(E+T)** of the data analysis in Section 4.4. The generalisation to a higher number of functional time series, observed sparsely or fully, is a straightforward extension of the bellow presented equations. the further incorporation of multivariate or scalar time series is uncomplicated because the regularisation of the spectral density inversion is no longer required.

Consider the model with two functional time series regressors $\{X_t^{(1)}(\cdot)\}_{t\in\mathbb{Z}}$ and $\{X_t^{(2)}(\cdot)\}_{t\in\mathbb{Z}}$. To simplify the notation we assume the both functional time series to take values in the same Hilbert space $\mathscr{H} = L^2([0,1],\mathbb{R})$. The lagged regression with a scalar response $\{Z_t\}_{t\in\mathbb{Z}}$ becomes:

$$Z_t = a + \sum_{k\in\mathbb{Z}} B_k^{(1)} X_{t-k}^{(1)} + \sum_{k\in\mathbb{Z}} B_k^{(2)} X_{t-k}^{(2)} + e_t$$

where $a \in \mathbb{R}$ is the intercept, $\{B_k^{(1)}\}_{k\in\mathbb{Z}}$ and $\{B_k^{(1)}\}_{k\in\mathbb{Z}}$ are two sequences of linear mappings from $\mathscr{H}$ to $\mathbb{R}$, and $\{e_t\}$ is a sequence of zero mean independent identically distributed real random variables. Denote $\{b_k^{(1)}\}_{k\in\mathbb{Z}}$ and $\{b_k^{(1)}\}_{k\in\mathbb{Z}}$ the filter functions corresponding to the functionals $\{B_k^{(1)}\}$ and $\{B_k^{(1)}\}$ by the Riesz representation theorem. We assume that the first functional time series, $\{X_t^{(1)}\}$, is sparsely observed, that is we have access to only the observations generated by

$$U_{tj}^{(1)} = X_t^{(1)}(x_{tj}) + \epsilon_{tj}, \qquad j = 1,\dots,N_t, \quad t = 1,\dots,T,$$

where N_t is a number of observation locations $\{x_{tj}\}$ at time $t = 1,\dots,T$. The second functional time series is fully observed, therefore we have access to $X_1^{(2)},\dots,X_T^{(2)}$. The mean functions of $\{X_t^{(1)}\}$ and $\{X_t^{(2)}\}$ are denoted as $\mu^{(1)}(\cdot)$ and $\mu^{(2)}(\cdot)$ respectively.

Denote $R_h^{\boxdot}$ and $\mathscr{R}_h^{\boxdot}$ the (cross-)covariance kernel and (cross-)covariance operator where $\boxdot$ is substituted by a single functional time series or a pair theoreof, as the case may be. Likewise denote the (cross-)spectral density kernel and operator as $f_\omega^{\boxdot}$ and $\mathscr{F}_\omega^{\boxdot}$ respectively.

Assuming

$$\sum_{h\in\mathbb{Z}} \left|\!\left|\!\left| \begin{bmatrix} R_h^{X^{(1)}} & R_h^{X^{(1)}X^{(2)}} \\ R_h^{X^{(2)}X^{(1)}} & R_h^{X^{(2)}} \end{bmatrix} \right|\!\right|\!\right|_\infty < \infty, \qquad \sum_{h\in\mathbb{Z}} \left\| \begin{bmatrix} \mathscr{R}_h^{X^{(1)}} & \mathscr{R}_h^{X^{(1)}X^{(2)}} \\ \mathscr{R}_h^{X^{(2)}X^{(1)}} & \mathscr{R}_h^{X^{(2)}} \end{bmatrix} \right\|_1 < \infty$$

and $\sum_{k\in\mathbb{Z}} \left|\!\left|\!\left| b_k^{(1)} \right|\!\right|\!\right|_\infty < \infty$, $\sum_{k\in\mathbb{Z}} \left|\!\left|\!\left| b_k^{(2)} \right|\!\right|\!\right|_\infty < \infty$, $\sum_{k\in\mathbb{Z}} \|B_k^{(1)}\|_{\mathscr{H}} < \infty$, $\sum_{k\in\mathbb{Z}} \|B_k^{(2)}\|_{\mathscr{H}} < \infty$ implies that the (cross)-spectral density kernels and operators $\{f_\omega^{X^{(1)}}\}_\omega$, $\{f_\omega^{X^{(2)}}\}_\omega$, $\{f_\omega^{X^{(1)}X^{(2)}}\}_\omega$, $\{f_\omega^{ZX^{(1)}}\}_\omega$, $\{f_\omega^{ZX^{(2)}}\}_\omega$, $\{\mathscr{F}_\omega^{X^{(1)}}\}_\omega$, $\{\mathscr{F}_\omega^{X^{(2)}}\}_\omega$, $\{\mathscr{F}_\omega^{X^{(1)}X^{(2)}}\}_\omega$, $\{\mathscr{F}_\omega^{ZX^{(1)}}\}_\omega$, $\{\mathscr{F}_\omega^{ZX^{(2)}}\}_\omega$ are well defined.

The joint frequency response operator

$$\mathscr{B}_\omega = \begin{bmatrix} \mathscr{B}_\omega^{(1)} & \mathscr{B}_\omega^{(2)} \end{bmatrix} = \sum_{k\in\mathbb{Z}} \begin{bmatrix} B_k^{(1)} & B_k^{(2)} \end{bmatrix} e^{-\mathrm{i}k\omega}, \qquad \omega \in [-\pi,\pi],$$

satisfies the relation

$$\begin{bmatrix} \mathscr{F}_\omega^{ZX^{(1)}} & \mathscr{F}_\omega^{ZX^{(2)}} \end{bmatrix} = \begin{bmatrix} \mathscr{B}_\omega^{(1)} & \mathscr{B}_\omega^{(2)} \end{bmatrix} \begin{bmatrix} \mathscr{F}_\omega^{X^{(1)}} & \mathscr{F}_\omega^{X^{(1)}X^{(2)}} \\ \mathscr{F}_\omega^{X^{(2)}X^{(1)}} & \mathscr{F}_\omega^{X^{(2)}} \end{bmatrix}, \qquad \omega \in [-\pi,\pi]. \tag{3.13}$$

The filter coefficients can be recovered by the formulae

$$B_k^{(1)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathscr{B}_\omega^{(1)} e^{\mathrm{i}k\omega}\,\mathrm{d}\omega, \qquad B_k^{(2)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathscr{B}_\omega^{(2)} e^{\mathrm{i}k\omega}\,\mathrm{d}\omega, \qquad k \in \mathbb{Z}.$$

3.3.2 Nonparametric estimation

The mean function of the sparsely observed $\{X_t^{(1)}\}$ can be estimated by the local linear smoother (Rubín and Panaretos, 2020b) while the mean function of the fully observed $\{X_t^{(2)}\}$ is estimated by the classical empirical pointwise mean. We denote these estimates $\hat{\mu}^{(1)}(\cdot)$ and $\hat{\mu}^{(2)}(\cdot)$ respectively.

The estimation of the spectral density operator $\{\mathscr{F}_\omega^{X^{(1)}}\}_{\omega\in[-\pi,\pi]}$ and the cross-spectral density $\{\mathscr{F}_\omega^{ZX^{(1)}}\}_{\omega\in[-\pi,\pi]}$ from the sparse observations is explained in Chapter 2, and the estimation of the spectral density operator $\{\mathscr{F}_\omega^{X^{(2)}}\}_{\omega\in[-\pi,\pi]}$ and the cross-spectral density $\{\mathscr{F}_\omega^{ZX^{(2)}}\}_{\omega\in[-\pi,\pi]}$ in the case of fully functional observations are presented in Chapter 1.

Once we have estimated all the above (cross-)spectral densities, we wish to recover the filter coefficients from the equation (3.13). The inversion of the joint spectral density of $\{X_t^{(1)}\}$ and $\{X_t^{(2)}\}$ is ill-conditioned due to the reasons explained in Section 3.1, where we overcame this issue by two regularisation techniques: spectral truncation and Tikhonov regularisation. Here we do the same while allowing each of the regressor time series to have a different degree of regularisation. This is important because, generally speaking, the estimation from sparse data will require more regularisation.

Denote the spectral decompositions of the spectral density operators by

$$\hat{\mathscr{F}}_\omega^{X^{(1)}} = \sum_{i=1}^{\infty} \hat{\lambda}_i^\omega \hat{\varphi}_i^\omega \otimes \hat{\varphi}_i^\omega, \qquad \omega\in[-\pi,\pi],$$
$$\hat{\mathscr{F}}_\omega^{X^{(2)}} = \sum_{i=1}^{\infty} \hat{\eta}_i^\omega \hat{\psi}_i^\omega \otimes \hat{\psi}_i^\omega, \qquad \omega\in[-\pi,\pi],$$

and express the cross-spectral basis with respect to the bases induced in the spectral decompositions above

$$\hat{\mathscr{F}}_\omega^{X^{(1)}X^{(2)}} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \hat{\gamma}_{ij}^\omega \hat{\psi}_j^\omega \otimes \hat{\varphi}_i^\omega, \qquad \omega\in[-\pi,\pi],$$

where $\{\hat{\lambda}_i^\omega\}_{i=1}^\infty$ and $\{\hat{\eta}_i^\omega\}_i^\infty$ are the harmonic eigenvalues of $\mathscr{F}_\omega^{X^{(1)}}$ and $\mathscr{F}_\omega^{X^{(2)}}$ respectively, for given $\omega\in[-\pi,\pi]$. The sequences of functions $\{\hat{\varphi}^\omega\}_i^\infty$ and $\{\hat{\psi}^\omega\}_i^\infty$ are orthonormal bases of $\mathscr{H}$ and are called the harmonic eigenfunctions of $\mathscr{F}_\omega^{X^{(1)}}$ and $\mathscr{F}_\omega^{X^{(2)}}$ respectively. The complex numbers $\{\hat{\gamma}_{ij}^\omega\}_{i,j=1}^\infty$ are the basis coefficients of $\hat{\mathscr{F}}_\omega^{X^{(1)}X^{(2)}}$ with respect to $\{\hat{\varphi}_i^\omega\otimes\hat{\psi}_j^\omega\}_{i,j=1}^\infty$, a basis of the space of Hilbert-Schmidt operators on $\mathscr{H}$.

Truncation regularisation

Select two (possibly different) truncation parameters $K_\omega^{(1)}\in\mathbb{N}$ and $K_\omega^{(2)}\in\mathbb{N}$ that may depend on $\omega\in[-\pi,\pi]$. The idea of the truncation regularisation is to replace the empirical joint spectral

density operator on the right hand side of the empirical version of (3.13) by

$$\begin{bmatrix} \sum_{i=1}^{K_\omega^{(1)}} \hat{\lambda}_i^\omega \hat{\varphi}_i^\omega \otimes \hat{\varphi}_i^\omega & \sum_{i=1}^{K_\omega^{(1)}} \sum_{j=1}^{K_\omega^{(2)}} \hat{\gamma}_{ij}^\omega \hat{\psi}_j^\omega \otimes \hat{\varphi}_i^\omega \\ \sum_{i=1}^{K_\omega^{(2)}} \sum_{j=1}^{K_\omega^{(1)}} \hat{\gamma}_{ji}^\omega \hat{\varphi}_j^\omega \otimes \hat{\psi}_i^\omega & \sum_{i=1}^{K_\omega^{(2)}} \hat{\eta}_i^\omega \hat{\psi}_i^\omega \otimes \hat{\psi}_i^\omega \end{bmatrix}. \tag{3.14}$$

Alternatively, the operator in the form (3.14) can be expressed with respect to the reduced basis of $\mathscr{H} \times \mathscr{H}$ composed of $\{[\hat{\varphi}_1^\omega, \mathbf{0}], \ldots, [\hat{\varphi}_{K_\omega^{(1)}}^\omega, \mathbf{0}], [\mathbf{0}, \hat{\psi}_1^\omega], \ldots, [\mathbf{0}, \hat{\psi}_{K_\omega^{(2)}}^\omega]\}$ where $\mathbf{0}$ is the zero element of $\mathscr{H}$. Denote $M \in \mathbb{C}^{(K_\omega^{(1)}+K_\omega^{(2)}) \times (K_\omega^{(1)}+K_\omega^{(2)})}$ the complex matrix given by the inverse of the following matrix, assuming it is invertible,

$$M = \left(\begin{array}{ccc|ccc} \hat{\lambda}_1^\omega & 0 & 0 & \hat{\gamma}_{11} & \cdots & \hat{\gamma}_{1,K_\omega^{(2)}} \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \hat{\lambda}_{K_\omega^{(1)}}^\omega & \hat{\gamma}_{K_\omega^{(1)},1} & \cdots & \hat{\gamma}_{K_\omega^{(1)},K_\omega^{(2)}} \\ \hline \hat{\gamma}_{11} & \cdots & \hat{\gamma}_{K_\omega^{(1)},1} & \hat{\eta}_1^\omega & 0 & 0 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 \\ \hat{\gamma}_{1,K_\omega^{(2)}} & \cdots & \hat{\gamma}_{K_\omega^{(1)},K_\omega^{(2)}} & 0 & 0 & \hat{\eta}_{K_\omega^{(2)}}^\omega \end{array} \right)^{-1}.$$

Denote the elements of M according to the blocks using the following scheme

$$M = \left(\begin{array}{ccc|ccc} m_{11}^{(1)} & \cdots & m_{1,K_\omega^{(1)}}^{(1)} & m_{11}^{(12)} & \cdots & m_{1,K_\omega^{(2)}}^{(12)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ m_{K_\omega^{(1)},1}^{(1)} & \cdots & m_{K_\omega^{(1)},K_\omega^{(1)}}^{(1)} & m_{K_\omega^{(1)},1}^{(12)} & \cdots & m_{K_\omega^{(1)},K_\omega^{(2)}}^{(12)} \\ \hline m_{11}^{(21)} & \cdots & m_{1,K_\omega^{(1)}}^{(21)} & m_{11}^{(2)} & \cdots & m_{1,K_\omega^{(2)}}^{(2)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ m_{K_\omega^{(2)},1}^{(21)} & \cdots & m_{K_\omega^{(2)},K_\omega^{(1)}}^{(21)} & m_{K_\omega^{(2)},1}^{(2)} & \cdots & m_{K_\omega^{(2)},K_\omega^{(2)}}^{(2)} \end{array} \right).$$

Then, the frequency response operators can be recovered by

$$\hat{\mathscr{B}}_\omega^{(1)} = \sum_{i=1}^{K_\omega^{(1)}} \sum_{j=1}^{K_\omega^{(1)}} m_{ij}^{(1)} \left\langle \hat{\varphi}_j^\omega, \cdot \right\rangle \mathscr{F}_\omega^{ZX^{(1)}}\left(\hat{\varphi}_i^\omega\right) + \sum_{i=1}^{K_\omega^{(2)}} \sum_{j=1}^{K_\omega^{(1)}} m_{ij}^{(21)} \left\langle \hat{\varphi}_j^\omega, \cdot \right\rangle \mathscr{F}_\omega^{ZX^{(2)}}\left(\hat{\psi}_i^\omega\right),$$

$$\hat{\mathscr{B}}_\omega^{(2)} = \sum_{i=1}^{K_\omega^{(1)}} \sum_{j=1}^{K_\omega^{(2)}} m_{ij}^{(12)} \left\langle \hat{\psi}_j^\omega, \cdot \right\rangle \mathscr{F}_\omega^{ZX^{(1)}}\left(\hat{\varphi}_i^\omega\right) + \sum_{i=1}^{K_\omega^{(2)}} \sum_{j=1}^{K_\omega^{(2)}} m_{ij}^{(2)} \left\langle \hat{\psi}_j^\omega, \cdot \right\rangle \mathscr{F}_\omega^{ZX^{(2)}}\left(\hat{\psi}_i^\omega\right).$$

Tikhonov regularisation

We consider again two possibly different regularisation parameters $\rho^{(1)} > 0$ and $\rho^{(2)} > 0$.

The empirical joint spectral density operator on the right hand side of the empirical version of

(3.13) is replaced by

$$\begin{bmatrix} \hat{\mathscr{F}}_\omega^{X^{(1)}} + \rho^{(1)}\mathrm{I} & \hat{\mathscr{F}}_\omega^{X^{(1)}X^{(2)}} \\ \hat{\mathscr{F}}_\omega^{X^{(2)}X^{(1)}} & \hat{\mathscr{F}}_\omega^{X^{(2)}} + \rho^{(2)}\mathrm{I} \end{bmatrix}$$

and the joint frequency response operator is estimated as

$$\begin{bmatrix} \hat{\mathscr{B}}_\omega^{(1)} & \hat{\mathscr{B}}_\omega^{(2)} \end{bmatrix} = \begin{bmatrix} \hat{\mathscr{F}}_\omega^{ZX^{(1)}} & \hat{\mathscr{F}}_\omega^{ZX^{(2)}} \end{bmatrix} \begin{bmatrix} \hat{\mathscr{F}}_\omega^{X^{(1)}} + \rho^{(1)}\mathrm{I} & \hat{\mathscr{F}}_\omega^{X^{(1)}X^{(2)}} \\ \hat{\mathscr{F}}_\omega^{X^{(2)}X^{(1)}} & \hat{\mathscr{F}}_\omega^{X^{(2)}} + \rho^{(2)}\mathrm{I} \end{bmatrix}^{-1}.$$

3.3.3 Forecasting the response process through multiple regressors

Once the estimates of the frequency response functions have been constructed, the filter coefficients are estimated by integrating back into the temporal domain

$$\hat{B}_k^{(1)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{\mathscr{B}}_\omega^{(1)} e^{\mathrm{i}\omega k}, \qquad k \in \mathbb{Z},$$
$$\hat{B}_k^{(2)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{\mathscr{B}}_\omega^{(2)} e^{\mathrm{i}\omega k}, \qquad k \in \mathbb{Z}.$$

In this section we extend the forecasting algorithm introduced in Subsection 3.2.2. We assume here again that the response time series is observed only at times $1,\dots,S$ where $S < T$ and we wish to predict $Z_{S+1},\dots,Z_T$.

1. Denote the ensemble of the measurements $U_{tj}, t = 1,\dots,T, j = 1,\dots,N_t$ as $\mathbb{U}^{(1)}$ realised on the regressor time series $\{X_t^{(1)}\}$. From the measurements $\mathbb{U}^{(1)}$, the fully observed functional observations $X_1^{(2)},\dots,X_T^{(2)}$, and the response time series $Z_1,\dots,Z_T$ (or only $Z_1,\dots,Z_S$ for $1 < S < T$), construct the estimates of the model: the mean functions $\hat{\mu}^{(1)}, \hat{\mu}^{(2)}$ the spectral density operators $\{\hat{\mathscr{F}}_\omega^{X^{(1)}}\}_{\omega\in[-\pi,\pi]}, \{\hat{\mathscr{F}}_\omega^{X^{(2)}}\}_{\omega\in[-\pi,\pi]}$, and the cross-spectral density operators $\{\hat{\mathscr{F}}_\omega^{X^{(1)}X^{(2)}}\}_{\omega\in[-\pi,\pi]}, \{\hat{\mathscr{F}}_\omega^{ZX^{(1)}}\}_{\omega\in[-\pi,\pi]}, \{\hat{\mathscr{F}}_\omega^{ZX^{(2)}}\}_{\omega\in[-\pi,\pi]}$. Using either the truncation regularisation or the Tikhonov regularisation, estimate the spectral transfer functions $\{\hat{\mathscr{B}}_\omega^{(1)}\}_{\omega\in[-\pi,\pi]}$ and $\{\hat{\mathscr{B}}_\omega^{(2)}\}_{\omega\in[-\pi,\pi]}$ and the filter coefficients $\{\hat{B}_k^{(1)}\}_{k\in\mathbb{Z}}$ and $\{\hat{B}_k^{(2)}\}_{k\in\mathbb{Z}}$.

2. Denoting the sparse measurements $\mathbb{U}^{(1)}$, predict the latent values of $X_{-M+1}^{(1)},\dots,X_{T+M}^{(1)}$ using the method recalled in Section 2.2. Denote these predictions as

   $$\hat{\Pi}(X_{-M+1}^{(1)}|\mathbb{U}^{(1)}),\dots,\hat{\Pi}(X_{T+M}^{(1)}|\mathbb{U}^{(1)}).$$

   The constant M is determined in such a way that the filter coefficients $\hat{B}_k^{(1)}$ and $\hat{B}_k^{(2)}$ are negligible for $|k| > M$.

3. Pad the fully observed functional time series $\{X_t^{(2)}\}$ by the mean function $\hat{\mu}^{(2)}$, i.e. set $X_t = \hat{\mu}^{(2)}$ for $t < 1$ or $t > T$. This approach was suggested in functional lagged prediction for fully functional regressor (Hörmann et al., 2015b) where the mean-zero functional

regressor time series is padded by zeros.

4. For each $s = S+1, \ldots, T$, construct the forecast

$$\hat{Z}_s = \bar{Z} + \sum_{k=-M}^{M} \hat{B}_k^{(1)} \left(\hat{\Pi}(X_{s-k} | \mathbb{U}^{(1)}) - \hat{\mu}^{(1)} \right) + \sum_{k=-M}^{M} \hat{B}_k^{(2)} \left(X_{s-k} - \hat{\mu}^{(2)} \right)$$

where $\bar{Z}$ is the sample mean of the response process.

3.3.4 Final comments on models combining multiple inputs

The framework of functional lagged regression introduced in this chapter is firstly explained on the set-up consisting of a sparsely observed functional time series viewed as the regressor and a univariate time series or sparsely observed functional time series response. Later, we discussed an extension of the lagged regression model to the set-up with multiple regressors inputs in Section 3.3. Concretely, we considered as a case study the model incorporating two functional time series as regressors: one of which is observed sparsely with measurement error, the other as fully functional observations. We discuss how to adapt the Tikhonov regularisation and the spectral truncation techniques to allow for a different degree of regularisation for each regressor, and how to forecast the response process having the estimates of the filter coefficient.

The further extension to an arbitrary number and type of regression inputs is straightforward. Indeed, one can consider a regression model with univariate/multivariate or functional response whose dependence is modelled via the lagged regression model blending together univariate/multivariate time series regressors and/or a number of functional time series regressors, some may be observed sparsely with measurement noise contamination, some may be observed fully functionally.

The estimation and the prediction routine consist then of the following steps.

1. Estimate all the spectral densities, be it operators or matrices, and all the cross-spectral densities among the regressors time series. Chapter 2 and Subsection 1.3.9 list the estimators for all the required (cross-)spectral densities.

2. Estimate all the cross-spectral densities between the response time series and each of the regressors time series.

3. Estimate the frequency response function by inverting the generalisation of the equation (3.13). The components corresponding to the functional time series need to be regularised by the Tikhonov regularisation or the spectral truncation. The regularisation by a possibly different magnitude is explained in Subsection 3.3.2. The univariate and the multivariate regressors time series do not need to be regularised as long as the sample size is large enough. Integrate the estimated frequency response function to obtain the estimates of the filter coefficients.

4. Recover the latent functional data for those functional time series regressors that are observed sparsely using the tools developed in Section 2.2. Pad to the left and the right all the fully observed functional time series and the univariate/multivariate time series by their respective empirical means.

5. Plug-in the fully observed functional time series regressors, univariate/multivariate regressors, and the predicted latent functional time series regressors into the estimated filter coefficients to construct the forecasts of the response process.

3.4 Code availability

The estimation, prediction, and regression methods presented in Chapter 2 and 3 for sparsely observed functional time series are implemented in MATLAB and the code is available on https://github.com/tomasrubin/sparse-functional-lagged-regression.

4 Empirical results and applications

This chapter provides with empirical demonstrations for the sparsely observed functional time series toolbox. It includes two simulation studies: the simulation study in Section 4.1 assesses the spectral density estimation and the functional data recovery predictor (Chapter 2), and the dependence of their performance on the sparse observation scheme design parameters, and the simulation study in Section 4.2 compares the spectral truncation and the Tikhonov regularisation techniques for lagged regression model estimation (Chapter 3).

The non-parametric inference for sparsely observed functional time series is further illustrated on three data applications: the fair-weather atmospheric electricity data measured at Tashkent, Uzbekistan, presented in Section 4.3 depicts the spectral density estimation and the functional data recovery; the visibility data recorded at Wank mountain, Germany, reported in Section 4.4 compares the spectral truncation and the Tikhonov regularisation techniques for lagged regression model estimation on a real data example; and Section 4.5 represents a functional lagged regression model with a sparsely observed functional time series response.

4.1 Simulation study: estimation of spectral density operators

4.1.1 Simulation set-up

In this subsection, we present a simulation study in order to prove the finite-sample performance of the spectral density operator estimator presented in Chapter 2. To this aim, we simulate realisations of functional linear processes (Subsection 1.3.8), namely the functional autoregressive process and the functional moving average process. Specifically, we consider:

- **Functional autoregressive process**

  The (Gaussian) functional autoregressive process $\{X_t\}$ of order 1 with values in the Hilbert space $\mathscr{H} = L^2([0,1])$ is defined by the iteration

$$(X_{t+1} - \mu) = \mathscr{A}(X_t - \mu) + E_t \tag{4.1}$$

where the mean function $\mu(x) = 4\sin(1.5\pi x)$ and the autoregression operator $\mathscr{A}$ associated with the kernel $A_c(x,y) = \kappa_c \exp\left(-(x+2y)^2\right)$ where the scaling constant κ_c is chosen so that $\|\mathscr{A}_c\| = c$. We vary c to control the degree of temporal dependence and let $c \in \{0.7, 0.9\}$. The stochastic innovation noise $\{E_t\}_t$ is a sequence consisting of independent identically distributed random elements with the covariance operator $\mathscr{S}$ defined as the integral operator with kernel $S(x,y) = 1.4\sin(2\pi x)\sin(2\pi y) + 0.6\cos(2\pi x)\cos(2\pi y)$. In the simulation results we denote the resulting two processes as $\mathbf{FAR(1)_{0.7}}$ and $\mathbf{FAR(1)_{0.9}}$ for $c = 0.7$ and $c = 0.9$ respectively.

By Proposition 1.3.17 and Theorem 1.3.18, there exists a unique Gaussian stationary solution to the equation (4.1) and its spectral density operator is given by the formula (1.63).

- **Functional moving average process** The (Gaussian) functional moving average process of order q is given by the formula

$$X_t = \mu + E_t + \mathscr{B}_1 E_{t-1} + \mathscr{B}_2 E_{t-2} + \cdots + \mathscr{B}_q E_{t-q}$$

where μ and E_{t-q} are as above, $\mathscr{B}_j, j = 1,\ldots,q$ are bounded linear operators on $\mathscr{H}$. Specifically we consider the orders $q \in \{4,8\}$ and define the moving average operators $\mathscr{B}_1,\ldots,\mathscr{B}_8$ as integral operators with kernels $B_1(x,y) = B_5(x,y) = 5\exp(-(x^2+y^2))$, $B_2(x,y) = B_6(x,y) = 5\exp(-((1-x)^2+y^2))$, $B_3(x,y) = B_7(x,y) = 5\exp(-(x^2+(1-y)^2))$, and $B_4(x,y) = B_8(x,y) = 5\exp(-((1-x)^2+(1-y)^2))$. We denote these functional moving average processes as $\mathbf{FMA(4)}$ and $\mathbf{FMA(8)}$ for $q = 4, 8$ respectively.

We simulate the considered processes over the temporal periods of varying length, specifically $T \in \{150, 300, 450, 600, 900, 1200\}$. The sparse observations are then obtained by the following process. We set a maximum number of locations to be sampled $N^{max} \in \{5, 10, 20, 30, 40\}$. For each $t = 1,\ldots,T$, a random integer N_t is independently drawn from the uniform distribution on $0, 1, \ldots, N^{max}$. Next, for each $t = 1,\ldots,T$, we independently draw N_t random locations $x_{tj}, j = 1,\ldots,N_t$ from the uniform distribution on $[0,1]$. At each location, an independent identically distributed Gaussian measurement error $\epsilon_{tj} \sim N(0,\sigma^2)$ is added and the ensemble $Y_{tj} = X_t(x_{tj}) + \epsilon_{tj}, j = 1,\ldots,N_t, t = 1,\ldots,T$ is used as the dataset for the estimation procedure. Therefore the observation protocol satisfies the assumptions (B1) and (B2).

The measurement error variance is chosen in the way that the ratio $\mathrm{Tr}(\mathscr{R}_0)/\sigma^2$, which we interpret as a basic signal-to-noise ratio metric, is 20. The same signal-to-noise ratio was used in the simulation study by Yao et al. (2005a). Further simulation results of ours not reported here indicate that moderate variations of the signal-to-noise ratio do not change the conclusions of this simulation study.

4.1.2 Spectral density operator estimation error dependence on sample size

In this subsection we quantify the estimation error of the spectral density estimator (2.9) in our simulation setting. In particular, we want to explore the dependence of the estimation error on the length T of the time series and the number N^{max} impacting the average number of measurements per curve.

For each of the considered process and for each pair of the sample size parameters $T \in \{150, 300, 450, 600, 900, 1200\}$ and $N^{max} \in \{5, 10, 20, 30, 40\}$ we simulated 100 independent realisations. We have run the estimation procedure introduced in Subsections 2.1.2 and 2.1.3. The tuning parameters were selected in accordance with the discussion in Section 2.5.

We measure the quality of the spectral density estimation by the relative mean square error defined as

$$RMSE = \frac{\int_{-\pi}^{\pi}\int_0^1\int_0^1 |\hat{f}_\omega(x,y) - f_\omega(x,y)|^2 \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\omega}{\int_{-\pi}^{\pi}\int_0^1\int_0^1 |f_\omega(x,y)|^2 \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\omega} \tag{4.2}$$

where $\hat{f}_\omega(\cdot,\cdot)$ and $f_\omega(\cdot,\cdot)$ are respectively the estimated and the true spectral density kernels at the frequency $\omega \in (-\pi, \pi)$. Table 4.1 presents the results for the considered functional autoregressive and functional moving average processes.

Concerning the results of Table 4.1, one can raise an interesting design question:

> Provided one has a fixed budget for the total number of measurements to be made, should opt to record fewer spatial measurements over a longer time interval (lengthy but sparsely observed time series), or rather record dense spatial measurements over a shorter time period (short but densely observed time series)?

In order to answer this question we define a simple linear model to assess the dependence of the relative mean square error on the considered sample size parameters T and N^{max}. For each of the considered processes we fit the linear model

$$\log(RMSE(N^{max}, T)) = \beta_0 + \beta_1 \log(N^{max}) + \beta_2 \log(T) + e \tag{4.3}$$

where $RMSE(N^{max}, T)$ is the average relative mean square error for the considered parameters T and N^{max}, $(\beta_0, \beta_1, \beta_2)$ are the regression parameters, and e is a homoskedastic model error. We consider the logarithmic model (4.3) because we expect multiplicative effect of sample size parameters on the estimation error. This hypothesis was confirmed by fitting various models with different transformations and the logarithmic model (4.3) provided with the best fit.

The least square estimates of (4.3) are presented in Table 4.2. The coefficient $\hat{\beta}_2$ is larger than $\hat{\beta}_1$ in absolute value for all four considered processes, therefore the relative increase of the time-length T has a stronger effect in reducing the relative mean square error of the estimated spectral density than the same relative increase in the number of points per curve. The apparent conclusion is that, in order to estimate the spectral density of a smooth functional time

Table 4.1 – Average relative mean square errors (defined in (4.2)) of the spectral density estimators for the above defined functional autoregressive and functional moving average processes ($\mathbf{FAR(1)_{0.7}}, \mathbf{FAR(1)_{0.9}}, \mathbf{FMA(4)}, \mathbf{FMA(8)}$) and varying sample sizes. The numbers in parentheses are the standard deviations of the relative mean square error. Each cell of the table (each error and its standard deviation) is the result of 100 independent simulations. The Bartlett's span parameter L was selected by the rule (2.52)

$T \backslash N^{max}$	5	10	20	30	40
			FAR(1)$_{0.7}$		
150	0.359 (0.082)	0.289 (0.070)	0.232 (0.069)	0.211 (0.064)	0.213 (0.066)
300	0.257 (0.067)	0.195 (0.048)	0.154 (0.044)	0.142 (0.042)	0.138 (0.042)
450	0.212 (0.041)	0.155 (0.037)	0.123 (0.032)	0.114 (0.031)	0.111 (0.030)
600	0.187 (0.047)	0.129 (0.029)	0.108 (0.024)	0.100 (0.026)	0.090 (0.025)
900	0.147 (0.031)	0.107 (0.022)	0.084 (0.019)	0.075 (0.018)	0.069 (0.020)
1200	0.125 (0.022)	0.094 (0.019)	0.073 (0.018)	0.063 (0.015)	0.060 (0.015)
			FAR(1)$_{0.9}$		
150	0.564 (0.097)	0.466 (0.112)	0.460 (0.117)	0.454 (0.149)	0.399 (0.135)
300	0.433 (0.075)	0.372 (0.102)	0.334 (0.101)	0.272 (0.098)	0.291 (0.113)
450	0.374 (0.074)	0.324 (0.078)	0.283 (0.092)	0.239 (0.081)	0.216 (0.077)
600	0.305 (0.068)	0.272 (0.062)	0.216 (0.074)	0.216 (0.083)	0.192 (0.073)
900	0.282 (0.054)	0.227 (0.068)	0.179 (0.061)	0.165 (0.072)	0.146 (0.061)
1200	0.241 (0.061)	0.194 (0.058)	0.152 (0.059)	0.137 (0.059)	0.125 (0.059)
			FMA(4)		
150	0.312 (0.060)	0.225 (0.063)	0.184 (0.060)	0.170 (0.049)	0.165 (0.050)
300	0.206 (0.040)	0.157 (0.042)	0.124 (0.028)	0.115 (0.030)	0.110 (0.033)
450	0.167 (0.033)	0.126 (0.034)	0.097 (0.022)	0.092 (0.027)	0.081 (0.021)
600	0.137 (0.027)	0.107 (0.027)	0.083 (0.017)	0.077 (0.023)	0.071 (0.017)
900	0.115 (0.020)	0.082 (0.015)	0.067 (0.016)	0.061 (0.015)	0.056 (0.016)
1200	0.096 (0.019)	0.072 (0.015)	0.056 (0.013)	0.050 (0.012)	0.047 (0.012)
			FMA(8)		
150	0.352 (0.071)	0.263 (0.064)	0.213 (0.064)	0.188 (0.074)	0.178 (0.069)
300	0.253 (0.055)	0.170 (0.043)	0.143 (0.050)	0.129 (0.053)	0.127 (0.053)
450	0.176 (0.048)	0.148 (0.049)	0.114 (0.044)	0.091 (0.031)	0.086 (0.043)
600	0.159 (0.041)	0.123 (0.039)	0.093 (0.036)	0.080 (0.031)	0.081 (0.036)
900	0.128 (0.030)	0.098 (0.030)	0.074 (0.029)	0.062 (0.023)	0.060 (0.026)
1200	0.101 (0.026)	0.071 (0.023)	0.055 (0.020)	0.049 (0.017)	0.051 (0.018)

Table 4.2 – The estimated regression parameters $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ of the model (4.3) for all considered functional autoregressive and functional moving average processes.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
FAR(1)$_{0.7}$	2.12	-0.32	-0.56
FAR(1)$_{0.9}$	2.25	-0.24	-0.49
FMA(4)	1.98	-0.32	-0.57
FMA(8)	2.37	-0.34	-0.61

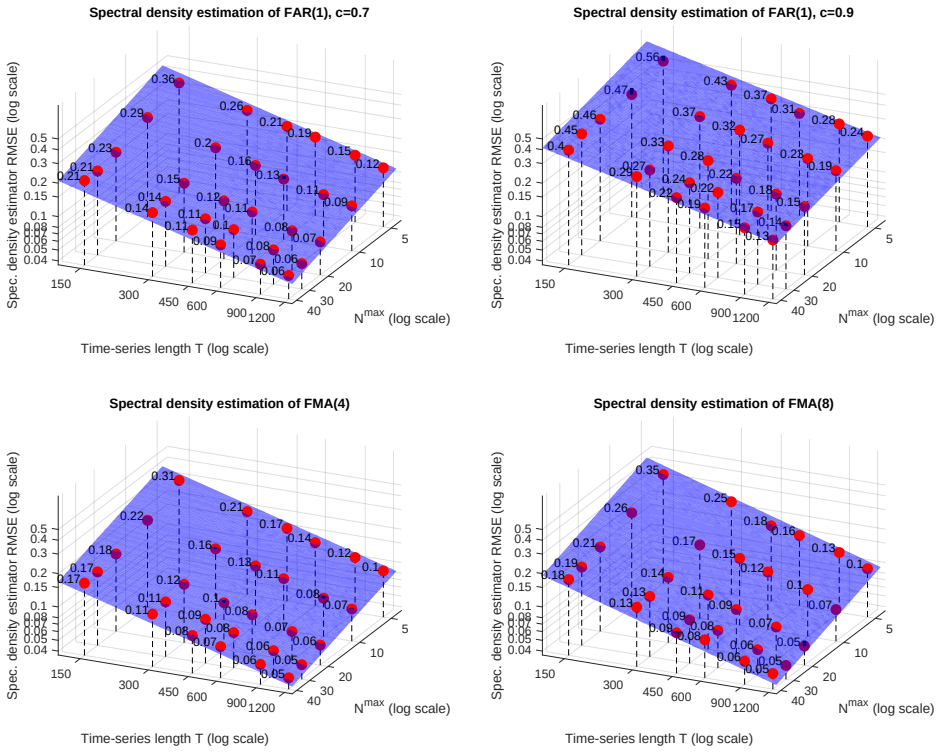

Figure 4.1 – The dependence of spectral density estimation relative mean square error (red points with labels of the magnitude of this error) on the sample size parameters T and N^{max}. The blue plane is the estimated regression surface in model (4.3).

series, the better strategy is to invest in longer time-horizon T rather than denser sampling regime. Figure 4.1 displays the relative mean square errors (4.2) dependence on the sample size parameters together with the regression surface fitted by the model (4.3).

4.1.3 Recovery of functional data from sparse observations

In this section, we examine the performance of the functional recovery procedure proposed in Section 2.2. We compare the recovery performance of our dynamic predictor (2.21), in the following denoted as the *dynamic recovery*, with its static version that relies only on the lag-zero covariance and hence does not exploit the temporal dependence, essentially following the method by Yao et al. (2005a). In the following, we call this predictor the *static recovery*. This static recovery is in fact the predictor (2.21) with the Bartlett's span parameter L set to 1.

We simulate 100 independent realisations for each of the considered functional moving average processes **FMA**(**4**), **FMA**(**8**), and the considered functional autoregressive processes $\mathbf{FAR}(\mathbf{1})_{\mathbf{0.7}}$, $\mathbf{FAR}(\mathbf{1})_{\mathbf{0.9}}$, see their definitions in Section 4.1.2, and each combination of the sample size parameters $T \in \{150, 300, 450, 600, 900, 1200\}$ and $N^{max} \in \{5, 10, 20, 30, 40\}$.

For each dataset we run the estimation procedure from Sections 2.1.2 and 2.1.3. The tuning parameters B_μ, B_R, and B_V are selected by K-fold cross-validation as explained in Subsection 2.5. The parameter L is selected again by the rule (2.52).

We define the functional recovery (either dynamic or static) relative mean square error as

$$RMSE = \frac{1}{T}\sum_{t=1}^{T} \frac{\int_0^1 \left(\hat{\Pi}(X_t|\mathbb{U}_T)(x) - X_t(x)\right)^2 \mathrm{d}x}{\mathrm{Tr}\{\mathscr{R}_0\}} \tag{4.4}$$

where $\hat{\Pi}(X_t|\mathbb{U}_T)$ is the recovered functional curve at $t = 1,\ldots,T$, either dynamically or statically, and X_t is the true (unobserved) functional datum.

The key factor contributing to the quality of the functional recovery is the estimate $\hat{\sigma}^2$ of the additive measurement error variance parameter σ^2. A very small value of the estimated $\hat{\sigma}^2$ can lead to an ill-conditioned matrix needed to be inverted in (2.20), thus resulting in a defective recovery of the functional data. Because this circumstance affects the relative mean square error metric, we opt to calculate the median of the relative mean square errors as a better indicator of the typical recovery error instead.

We calculate the *relative gain* as

$$Relative\ gain = \left(\frac{RMSE(static)}{RMSE(dynamic)} - 1\right) * 100\% \tag{4.5}$$

where $RMSE(static)$ is the median relative mean square error of the static recovery and $RMSE(dynamic)$ is the median mean square error of the dynamic recovery.

Table 4.3 summarizes the relative gains of dynamic recovery over the static recovery. Unsurprisingly, the relative gain is strikingly large for sparser designs. This can be explained by the fact that in sparse designs there is not sufficient information to interpolate the functional curves themselves, and the observed data in neighbouring curves are crucial for the recovery of the curves. That being said, it is observed that even when the number of points sampled per curve are as many as 40, the improvement remains substantial, demonstrating that the new methodology should be preferred over methods designed for the independent identically distributed case when dependence is present.

4.2 Simulation study: regularisation techniques comparison for lagged regression

4.2.1 Simulation set-up

In this simulation study we assess the performance of the methods of Chapter 3 on the basis of two criteria: the estimation error of the filter coefficients estimator (3.8), and the prediction error of the forecasts of the response process (Section 3.2.2). We also compare the performance

Table 4.3 – Relative gain (4.5) between median relative mean square error of the dynamic recovery and median relative mean square error of the static recovery. Positive percentage signifies that dynamic recovery has smaller error. Simulations from the functional autoregressive and functional moving average processes, $\mathbf{FAR(1)_{0.7}}$, $\mathbf{FAR(1)_{0.9}}$, $\mathbf{FMA(4)}$, $\mathbf{FMA(8)}$. Each cell of the table is the result of 100 independent simulations

	$T \backslash N^{max}$	5	10	20	30	40
$\mathbf{FAR(1)_{0.7}}$	150	51 %	39 %	14 %	16 %	4 %
	300	43 %	29 %	21 %	6 %	4 %
	450	39 %	28 %	18 %	7 %	4 %
	600	36 %	30 %	12 %	7 %	5 %
	900	37 %	27 %	15 %	8 %	5 %
	1200	35 %	28 %	17 %	8 %	3 %
$\mathbf{FAR(1)_{0.9}}$	150	73 %	47 %	46 %	34 %	23 %
	300	70 %	43 %	37 %	30 %	15 %
	450	66 %	45 %	32 %	16 %	18 %
	600	59 %	45 %	25 %	22 %	18 %
	900	60 %	43 %	29 %	18 %	15 %
	1200	59 %	44 %	29 %	24 %	16 %
$\mathbf{FMA(4)}$	150	67 %	38 %	38 %	23 %	30 %
	300	53 %	39 %	33 %	31 %	26 %
	450	52 %	45 %	38 %	30 %	24 %
	600	45 %	41 %	32 %	26 %	24 %
	900	54 %	41 %	37 %	30 %	22 %
	1200	54 %	45 %	34 %	26 %	21 %
$\mathbf{FMA(8)}$	150	55 %	36 %	34 %	34 %	26 %
	300	49 %	45 %	30 %	28 %	18 %
	450	51 %	37 %	31 %	22 %	20 %
	600	55 %	45 %	35 %	19 %	19 %
	900	56 %	47 %	34 %	26 %	22 %
	1200	54 %	44 %	31 %	24 %	24 %

of the two regularisation techniques, and demonstrate that neither dominates the other. To illustrate this, we introduce two different filter coefficient functions, see (4.9) and (4.10), for which one of the techniques is expected to perform better than the other, and vice versa.

We simulate the functional regressor series $\{X_t\}_{t\in\mathbb{Z}}$ as either the functional autoregressive process of order 1 and the functional moving average process of order 4:

We simulate realisations of the functional autoregressive process of order 1 and the functional moving average process of order 4 defined in the following two settings:

($\widetilde{\mathbf{FAR(1)}}$) The process $\{X_t\}_{t\in\mathbb{Z}}$ is a functional autoregressive process of order 1 (Bosq, 2000) defined by the iteration

$$X_{t+1} = \mathscr{A} X_t + E_t, \qquad t \in \mathbb{Z}. \tag{4.6}$$

The operator $\mathscr{A}$ is assumed to be a Hilbert-Schmidt operator and we define its kernel as $A(x,y) = \kappa \sin(x-y), x, y \in [0,1]$, where $\kappa > 0$ is chosen such that $\|\mathscr{A}\|_{L(\mathscr{H})} = 0.7$ and $\|\cdot\|_{L(\mathscr{H})}$ is the operator norm in the space of linear operators on $\mathscr{H}$.

Moreover, the sequence $\{E_t\}_{t\in\mathbb{Z}}$ is defined to be a collection of independent identically distributed zero-mean Gaussian random variables in $\mathscr{H} = L^2([0,1])$ with the covariance kernel $K(x,y)$ given by

$$\begin{aligned} K(x,y) = \sin(2\pi x)\sin(2\pi y) + 0.6\cos(2\pi x)\cos(2\pi y) + 0.3\sin(4\pi x)\sin(4\pi y) + \\ + 0.1\cos(4\pi x)\cos(4\pi y) + 0.1\sin(6\pi x)\sin(6\pi y) + 0.1\cos(6\pi x)\cos(6\pi y) + \\ + 0.05\sin(8\pi x)\sin(8\pi y) + 0.05\cos(8\pi x)\cos(8\pi y) + 0.05\sin(10\pi x)\sin(10\pi y) + \\ + 0.05\cos(10\pi x)\cos(10\pi y), \quad x, y \in [0,1]. \end{aligned} \tag{4.7}$$

($\widetilde{\textbf{FMA}(\textbf{4})}$) The process $\{X_t\}_{t\in\mathbb{Z}}$ is considered to be the functional moving average process of order 4 defined by

$$X_t = E_t + \mathscr{M}_1 E_{t-1} + \mathscr{M}_2 E_{t-2} + \mathscr{M}_3 E_{t-3} + \mathscr{M}_4 E_{t-4}, \qquad t \in \mathbb{Z}. \tag{4.8}$$

The operators $\mathscr{M}_1, \ldots, \mathscr{M}_4$ are assumed to be Hilbert-Schmidt and given by their kernels $M_1(x,y) = \kappa_1 \sin(x+y), M_2(x,y) = \kappa_2 \sin(1-x+y), M_3(x,y) = \kappa_3 \sin(1+x-y), M_4(x,y) = \kappa_4 \sin(2-x-y)$, for $x, y \in [0,1]$, respectively. The constants $\kappa_1 > 0, \ldots, \kappa_4 > 0$ are chosen so that $\|\mathscr{M}_1\|_{L(\mathscr{H})} = 0.8, \|\mathscr{M}_2\|_{L(\mathscr{H})} = 0.6, \|\mathscr{M}_3\|_{L(\mathscr{H})} = 0.4, \|\mathscr{M}_4\|_{L(\mathscr{H})} = 0.2$ respectively.

The functional autoregressive process ($\widetilde{\textbf{FAR}(\textbf{1})}$), defined uniquely by the equation (4.6), and the functional moving average process ($\widetilde{\textbf{FMA}(\textbf{4})}$) are stationary and Gaussian (Bosq, 2000).

Each of the above defined functional processes is simulated with a varying time length $T \in \{300, 600, 900, 1200\}$. The sparse observations (2.1) are generated by fixing the maximal number of observations per curve $N^{max} \in \{10, 20, 40, 60\}$. For each curve, an integer valued random variable is drawn with uniform distribution on $\{0, \ldots, N^{max}\}$ corresponding to the number of spatial locations where the X_t is observed (with measurement error, to be defined). The measurement locations x_{tj} are sampled as uniform random variables on $[0,1]$. At each x_{tj} location, the measurement error is added as a realisation of a centred Gaussian random variable with variance $\sigma^2 > 0$. The variance $\sigma^2 > 0$ is chosen so the signal-to-noise ratio is $\mathrm{Tr}(\mathscr{R}_0^X)/\sigma^2 = 20$.

For the lagged regression model (3.1) we assess the setting with a sparsely observed functional time series predictor ($\mathscr{H}_1 = L^2([0,1],\mathbb{R})$) and a scalar response ($\mathscr{H}_2 = \mathbb{R}$). We consider regression models where only certain filter coefficients B_k are nonzero functions. In particular, the nonzero filter coefficients are considered to be either of the two following options

(A) $$\beta_A(x) = \cos(4\pi x), \qquad x \in [0,1] \tag{4.9}$$

(B) $$\beta_B(x) = \sin(2\pi x), \qquad x \in [0,1] \tag{4.10}$$

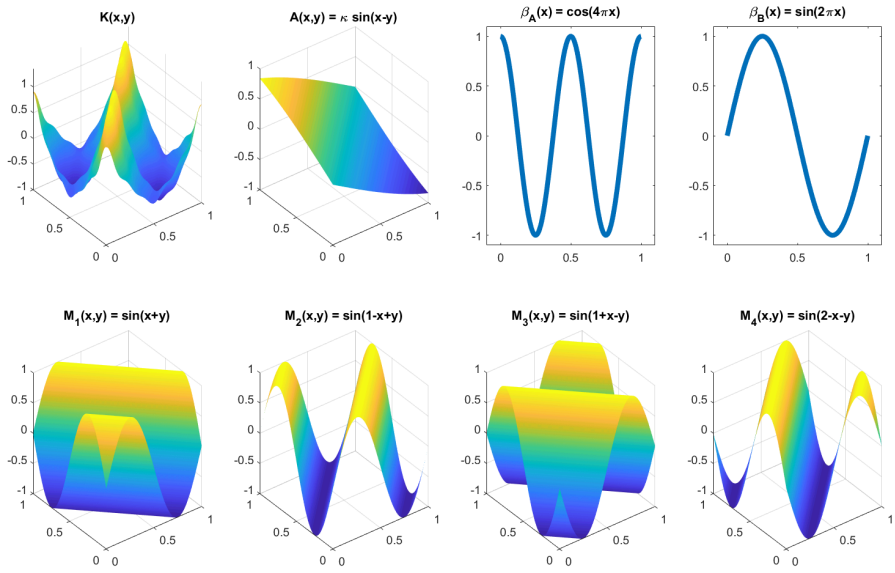

Figure 4.2 – **Top row:** The covariance kernel of the stochastic innovation K (4.7), the autoregressive kernel A (4.6), the filter coefficients β_A (4.9) and β_B (4.10) respectively. **Bottom row:** The kernels of the moving average process $M_1, \ldots, M_4$ (4.8).

The considered kernels and filter functions are visualised on Figure 4.2.

We consider 3 lagged regression schemata with a varying set of nonzero filter coefficients.

(reg1) The filter coefficients B_0, B_1 are set to either β_A or β_B.

(reg2) The filter coefficients B_0, B_3 are set to either β_A or β_B.

(reg3) The filter coefficients $B_0, B_1, B_2, B_3, B_4, B_5$ are nonzero but with decaying magnitude. They are set to either

$$(B_0, B_1, B_2, B_3, B_4, B_5) = \left(\beta_A, 0.9\beta_A, 0.7\beta_A, 0.5\beta_A, 0.3\beta_A, 0.1\beta_A\right)$$

or

$$(B_0, B_1, B_2, B_3, B_4, B_5) = \left(\beta_B, 0.9\beta_B, 0.7\beta_B, 0.5\beta_B, 0.3\beta_B, 0.1\beta_B\right).$$

The variance of the measurement error $\{e_t\}_{t\in\mathbb{Z}}$ is set to be $\tau^2 = 0.001$.

For each combination of the settings, i.e. each of the 2 linear processes of $\{X_t\}_{t\in\mathbb{Z}}$, each of 4 length parameters $T \in \{300, 600, 900, 1200\}$, each of 4 sampling density parameters $N^{max} \in \{10, 20, 40, 60\}$, each of 3 regression schemata, and 2 shapes of the filter coefficients, we run 90 independent runs. Moreover, we consider also the regime of complete functional observations

in the setting of Hörmann et al. (2015b) in order to compare how much information is lost due to sparse sampling.

4.2.2 Estimation procedure

The proposed methodology requires the selection of the tuning parameters. We implemented the choice of the bandwidths B_R and B_C for the estimation of $\{\mathscr{F}_\omega^X\}_{\omega\in[-\pi,\pi]}$ and σ^2 by means of K-fold cross validation, as explained in detail in Rubín and Panaretos (2020b). The Bartlett span parameter for the estimation of the spectral density is set to $L = \lfloor 2T^{1/3} \rfloor$.

In order to select the regularisation parameters for either of the two proposed regularsation methods we resort to holdout cross-validation. We split the response time series $Z_1,\dots,Z_T$ into the training set $Z_1,\dots,Z_S$ and the test set $Z_{S+1},\dots,Z_T$. The split is set to be 80:20 in favour of the training set, i.e. $S = 0.8T$. The cross-spectral density $\{\mathscr{F}_\omega^{ZX}\}_{\omega\in[-\pi,\pi]}$ is estimated from the data in the training set $Z_1,\dots,Z_S$ and the bandwidth parameter B_C is selected by K-fold cross-validation within the training set.

The truncation estimator (3.5) is constructed by means of *eigenvalue thresholding* (Hörmann et al., 2015b). Specifically, we set $K_\omega(\upsilon) = \arg\max_{m\geq 1}\{\hat{\lambda}_m^\omega > \upsilon\}$ where $\upsilon > 0$ is a parameter to be chosen by holdout cross-validation in the following way. Having estimated $\{\mathscr{F}_\omega^X\}_{\omega\in[-\pi,\pi]}$ and σ^2 from the sparsely observed regressor time series $\{X_t\}_{t=1}^T$, and the cross-spectral density $\{\mathscr{F}_\omega^{ZX}\}_{\omega\in[-\pi,\pi]}$ from $\{X_t\}_{t=1}^T$ and the training partition of the response $Z_1,\dots,Z_S$, the frequency response function is estimated by the formula (3.5) using a candidate value of $K_\omega(\upsilon)$. The forecasts $\hat{Z}_{S+1},\dots,\hat{Z}_T$ are produced by the methodology outlined in Section 3.2.2. Comparing the forecasts with the true values of $Z_{S+1},\dots,Z_T$ yields a mean square forecast error on the holdout partition which we minimse with respect to υ. The Tikhonov estimator (3.6) involves the selection of the parameter ρ. In the same way as for the truncation regularisation, we chose ρ by holdout cross-validation based on the mean square forecast error on $Z_{S+1},\dots,Z_T$.

In the case of complete functional observations we again use holdout cross-validation for the selection of the eigenvalue thresholding parameter as well as the Tikhonov parameter.

4.2.3 Evaluation criteria

We assess the estimation error of the filter coefficients by the mean square error criterion:

$$\delta^{\mathscr{B}} = \sum_{k\in\mathbb{Z}} \left\| \hat{\mathscr{B}}_k - \mathscr{B}_k \right\|^2 \tag{4.11}$$

Next we want to assess the forecasting performance of the proposed methodology. Because the entire sample was used for fitting the model dynamics, we simulate an independent copy of the regressor time series, denoted as $\{X_t^{copy}\}_{t=1}^T$, and the response process $\{Z_t^{copy}\}_{t=1}^T$. Using the estimates of the model dynamics from the original data, we produce the predictions

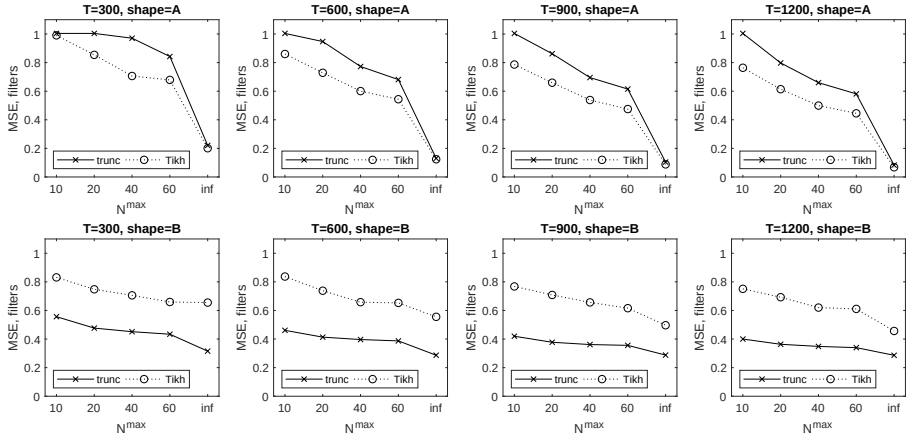

Figure 4.3 – The median mean square error $\delta^{\mathscr{B}}$ of the filter coefficient estimates (4.11) for the truncation regularsation ("trunc") and Tikhonov regularsation method ("Tikh"), displayed as a function of the time series length $T \in \{300, 600, 900, 1200\}$ and the the maximum number of the observation locations $N^{max} \in \{10, 20, 40, 60, \text{inf}\}$ where "inf" stands for the fully observed functional data. The top and the bottom row show the results for the filter coefficients of the shape A, defined in (4.9), and the shape B, defined in (4.10), respectively.

$\{\hat{Z}_t^{copy}\}_{t=1}^T$ from the sparsely observed $\{X_t^{copy}\}_{t=1}^T$ and compare with the true values $\{Z_t^{copy}\}_{t=1}^T$. The prediction relative mean square error is then defined

$$\delta^{pred} = \frac{1}{T}\sum_{t=S}^{T} \frac{\left(\hat{Z}_t^{copy} - Z_t^{copy}\right)^2}{\text{var}(Z_0)}. \tag{4.12}$$

Moreover, we include the prediction error of the oracle estimator that assumes that both the dynamics of the regressor time series $\{X_t\}_{t\in\mathbb{Z}}$ and the filter coefficients $\{\mathscr{B}_k\}_{k\in\mathbb{Z}}$ are known. The oracle estimator completes the steps 3 and 4 of the algorithm of Section 3.2.2 where the estimates $\{\hat{R}_h^X(\cdot,\cdot)\}_{h\in\mathbb{Z}}$ and $\{\hat{\mathscr{B}}_k^{trunc}\}_{k\in\mathbb{Z}}$ (or $\{\hat{\mathscr{B}}_k^{Tikh}\}_{k\in\mathbb{Z}}$) are replaced by the true values of $\{R_h^X(\cdot,\cdot)\}_{h\in\mathbb{Z}}$ and $\{\mathscr{B}_k\}_{k\in\mathbb{Z}}$.

4.2.4 Results of numerical experiments

Due to the large number of simulation settings considered, we display the results in an aggregated form. Figures 4.3 and 4.5 present the results for the filter coefficient estimation and the prediction performance as a function of the sample size parameters T and N^{max}. The results are aggregated over both types of simulated dynamics of $\{X_t\}_t$, i.e. the functional autoregressive process ($\widetilde{\textbf{FAR}(\textbf{1})}$) and the functional moving average process ($\widetilde{\textbf{FMA}(\textbf{4})}$), and over all three regression schemes **(reg1)**, **(reg2)**, and **(reg3)**. Figures 4.4 and 4.6, on the other hand, present the results for different the simulated dynamics of the process and the different considered regression schemes separately, aggregated over all time series length parameters $T \in \{300, 600, 900, 1200\}$ and all sparse observation regimes $N^{max} \in \{10, 20, 40, 60\}$.

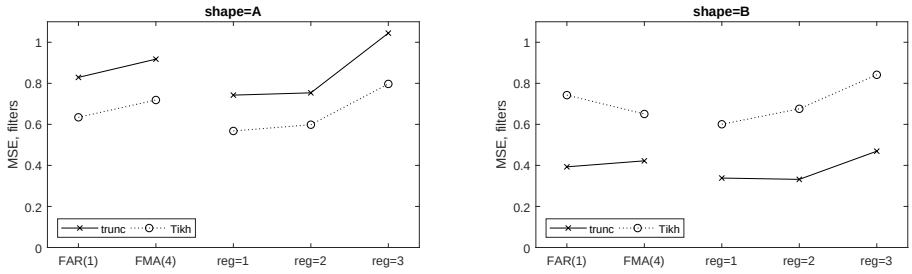

Figure 4.4 – The median mean square error $\delta^{\mathscr{B}}$ of the filter coefficient estimates (4.11) for the truncation regularsation ("trunc") and Tikhonov regularsation method ("Tikh") with respect to the simulated dynamics of $\{X_t\}_t$ and the regression scheme. The results are aggregated over all sparse observation setups $T \in \{300, 600, 900, 1200\}$ and $N^{max} \in \{10, 20, 40, 60\}$. The left and the right figures show the results for the filter coefficients of the shapes (4.9) and (4.10) respectively.

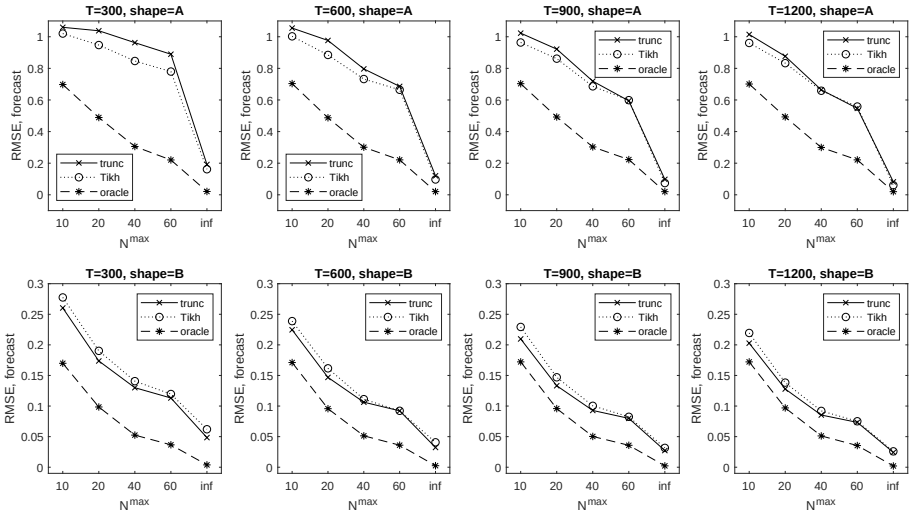

Figure 4.5 – The median mean square prediction error δ^{pred} (4.12) for the truncation regularisation ("trunc"), Tikhonov regularisation method ("Tikh"), and the oracle estimator, displayed as a function of the time series length $T \in \{300, 600, 900, 1200\}$ and the the maximum number of the observation locations $N^{max} \in \{10, 20, 40, 60, \text{inf}\}$ where "inf" stands for the fully observed functional data. The top and the bottom row show the results for the filter coefficients of the shapes (4.9) and (4.10) respectively.

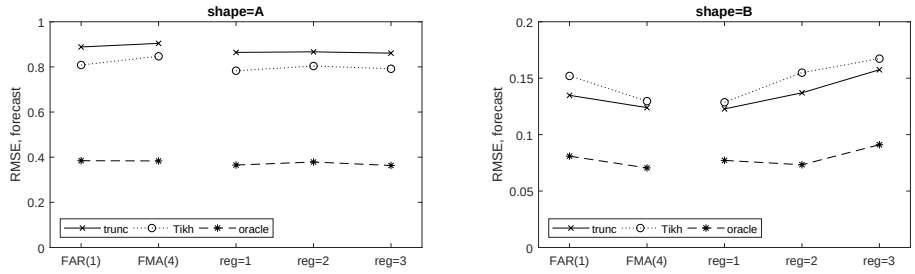

Figure 4.6 – The median mean square prediction error δ^{pred} (4.12) for the truncation regularisation ("trunc"), Tikhonov regularisation method ("Tikh"), and the oracle estimator with respect to the simulated dynamics of $\{X_t\}_t$ and the regression scheme. The results are aggregated over all sparse observation setups $T \in \{300, 600, 900, 1200\}$ and $N^{max} \in \{10, 20, 40, 60\}$. The left and the right figures show the results for the filter coefficients of the shapes (4.9) and (4.10) respectively.

An inspection of Figures 4.3 and 4.5 reveals that there is no clear winner between the truncation and the Tikhonov methods. The numerical experiments with the shape **(A)** defined by (4.9) show that the Tikhonov method dominates the truncation method in all considered settings for the estimation of the filter coefficients. The simulations with the shape **(B)** defined by (4.10) yield the opposite behaviour: the truncation regularisation prevails. We attribute this dichotomy to the following reasons:

- The shape **(B)** corresponds to the leading eigenfunction of the covariance kernel (4.7). Even though the functional regression is performed in the spectral domain, the frequency response function $\mathscr{B}_\omega$ is still well aligned with the first eigenfunction of $\mathscr{F}^X_\omega$ and therefore it is enough to cut off after the first eigenvalue, thus favouring truncation regularisation. See Figure 4.7 to visualise the alignment in the spectral domain.
- The shape **(A)** corresponds to the fourth eigenfunction of the covariance kernel (4.7). Moreover the fourth eigenvalue is tied with the fifth and the sixth one. This structure is preserved also in the spectral domain, c.f. Figure 4.7. Since the Tikhonov regularisation does not discard the eigenspace corresponding to any of these eigenvalues, it achieves lower estimation error in this non-aligned case. Moreover, the Tikhonov regularisation enjoys the advantage of being stable to spectral eigenvalue ties (Hall and Horowitz, 2007; Pham and Panaretos, 2018).
- Note that the shape **(A)** is generally more difficult to estimate than shape **(B)**. This is not surprising because the signal in the fourth (or higher) eigenfunction is much weaker than in the first one.

There is no notable difference in the difficulty of estimation between the two simulated dynamics of the regressor time series $\{X_t\}_t$, i.e. the functional autoregressive process ($\widetilde{\mathbf{FAR}(\mathbf{1})}$) and the functional moving average process ($\widetilde{\mathbf{FMA}(\mathbf{4})}$). The considered regression schemes do

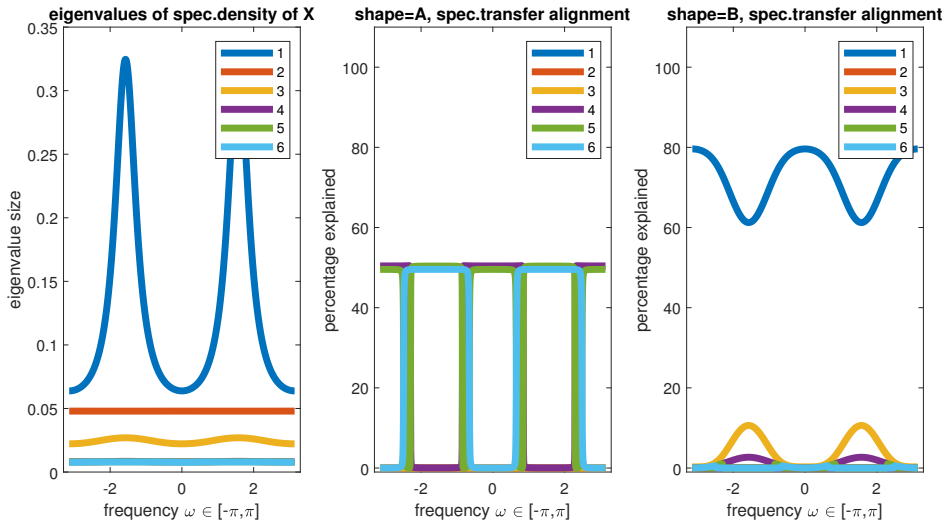

Figure 4.7 – **Left:** The six leading eigenvalues $\lambda_1^\omega, \ldots, \lambda_6^\omega$ of the spectral density operator $\mathscr{F}_\omega^X$ as a function of the frequency $\omega \in [-\pi, \pi]$. **Center:** The norm percentage of the frequency response function $\mathscr{B}_\omega$ for the shape **(A)**, the simulated process **(FAR(1))**, and the regression scheme **(reg1)** explained by the eigenspaces corresponding to the six leading eigenvalues as a function of the frequency $\omega \in [-\pi, \pi]$. **Right:** The same as the central plot but for the shape **(B)**.

not reveal any surprises: the longer the lagged dependence is, the more difficult estimation becomes. Therefore the **(reg3)** scheme produces the largest errors while **(reg1)** produces the lowest.

The prediction error of the response process δ^{pred}, which is presented in Figures 4.5 and 4.6, follows the same conclusions as the estimation of the filter coefficients. The shape **(A)** is more challenging to predict and the Tikhonov regularisation is seen to be preferable not only for estimation, but for prediction too. The shape **(B)** is easier to predict using the truncation regularisation. The predictions by either of the two techniques, feature twice to thrice greater prediction error δ^{pred} than the oracle estimator, i.e. the prediction assuming the model of the data to be known and the uncertainty coming only from sparse noisy sampling regime.

4.3 Data analysis: fair weather electricity in Tashkent, Uzbekistan

The atmosphere is weakly conductive due to the ionization of molecules and this conductivity can be continuously measured by a variable called *atmospheric electricity* (Tammet, 2009). The ionization is the outcome of complex physical-chemical processes that are subject to the current weather conditions. Since unfair weather conditions affect and alter these processes (Israelsson and Tammet, 2001), climatologists are interested in analysing the atmospheric electricity variable only under fair weather conditions (the definition of fair weather is given in this section). The analyses under fair weather conditions are of particular interest because the

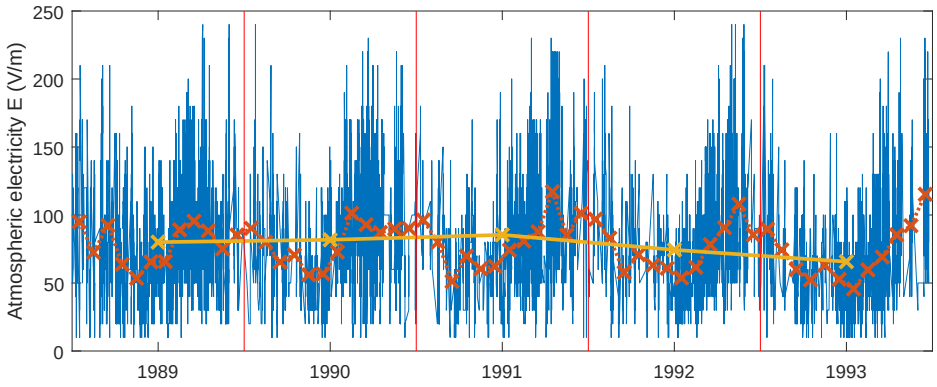

Figure 4.8 – Overview of the fair-weather atmospheric electricity time series measured in Tashkent, Uzbekistan. All fair-weather hourly measurements (blue line) accompanied by monthly means (brown crosses, brown dotted line) and yearly means (yellow crosses, yellow solid line).

fair-weather electricity variable is a valuable source of information in global climate research (Tammet, 2009) as well as with regards to air pollution (Israelsson and Tammet, 2001).

Tammet (2009) published an open-access database of atmospheric electricity time series accompanied by some meteorological variables. Most of the data come from weather stations across the former Soviet Union states and their data quality is assessed as high (Tammet, 2009). In this example, we analyse the time series of one weather station, namely that measured at the station near Tashkent, Uzbekistan. The atmospheric electricity was recorded between the years 1989 and 1993 in the form of hourly averages. Besides the atmospheric electricity, a number of other meteorological variables were measured, of which we use two: the wind speed and the cloudiness.

The definition of the fair-weather criteria is not simple and can often be relatively subjective (Xu et al., 2013). Inspired by criteria in climatology research (Xu et al., 2013; Israelsson and Tammet, 2001), we define the weather conditions as fair if the particular hourly measurement satisfies all of the following conditions:

- the wind speed is less than $20\,km/h$,
- the sky is clear (the total cloudiness variable is equal to 0),
- the atmospheric electricity E satisfies $0 < E < 250\,V/m$.

Because of the above stated fair-weather criteria (and some genuinely missing data in the database), the resulting fair-weather electricity time series is, in fact, unevenly sampled time series. Nevertheless, we assume there exists an underlying continuous truth, corresponding to the atmospheric electricity if the weather was fair. The latent process of fair-weather

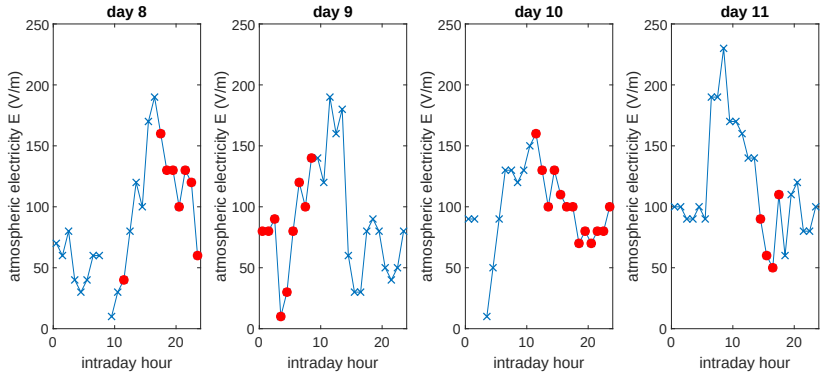

Figure 4.9 – Example of atmospheric electricity profiles over 4 consecutive days. The fair-weather atmospheric electricity measurements are highlighted as red points. The unfair-weather measurements (blue crosses) are not used for the analysis. Some hourly measurements are genuinely missing in the data (e.g. 9 AM on day 8) and we treat them as missing completely at random.

atmospheric electricity is considered smooth and its values are observed only under the fair-weather conditions, possibly with a deviation from the truth (noise). Based on the above discussed natural mechanisms, we justify the assumption that the censoring protocol is independent of the underlying fair-weather atmospheric electricity process.

The underlying fair-weather atmospheric electricity process is a scalar continuous time series. Previous research (Hörmann and Kokoszka, 2010; Hörmann et al., 2015a, 2018; Aue et al., 2015) has demonstrated the usefulness of segmenting a continuous scalar time series into segments of an obvious periodicity, usually days, and thus constructing a functional time series. A key benefit of this practice is the separation of intra-day variability and the temporal dependence across the days while preserving a fully non-parametric model.

We use the same approach in our analysis as well. We segment the (latent) continuous time series into days and consider each day as an unobserved (latent) functional datum defined on $[0, 24]$. We place the hourly observations in the middle of the hour interval, i.e. $0.5, 1.5, 2.5, \ldots, 23.5$. Because of the above fair-weather criteria, the constructed fair-weather atmospheric electricity time series falls into the sparsely observed functional time series framework defined in Subsection 2.1.1.

Figure 4.8 presents an overview of the considered fair-weather atmospheric electricity time series accompanied by monthly and yearly means. Figure 4.9 provides a zoomed-in perspective into a stretch of data in 4 consecutive days.

In summary, the fair-weather atmospheric electricity functional time series has the following features:

- the data are recorded over 5 years, therefore the time horizon of the functional time series is $T = 1826$ (days),

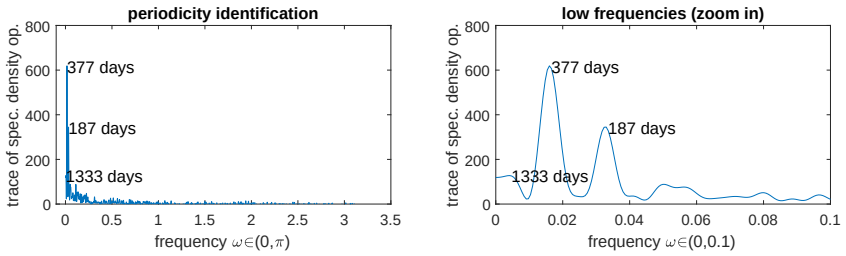

Figure 4.10 – **Left:** The periodicity identification plot with $L = 1000$. The labels at first 4 peaks convert the frequency into the corresponding periodicity. **Right:** Zoom-in into low frequencies.

- there are 1118 days (61 %) have at least 1 fair-weather measurement,
- there are 251 gaps in time series (we define a gap as a stretch of days where there is no measurement within these days) with the average length of 2.8 days,
- there are 12997 fair-weather measurements in total, i.e. 7.1 on average per day, or 11.6 on average per day among the days with at least one measurement.

The statistical question raised is the following. Benefiting from the separation of intra-day variability and temporal dependence across the days, can we fit an interpretable model of the process dynamics? Additionally, we aim to recover the latent functional data, fill in the gaps in the data, remove the noise, and construct confidence bands.

We analyse the fair-weather atmospheric electricity data by the means of Chapter 2. After removing the intra-day dependence by subtracting the estimate $\hat{\mu}(\cdot)$ we inspect the periodicity identification chart introduced in Section 2.1.4. Specifically, we construct the said chart with $L = 1000$ and plot the trace of the estimated spectral density operator against frequencies $\omega \in (0, \pi)$. We identify the peaks of this plot as suggesting the presence of periodicities in the corresponding frequencies.

The largest peak in Fig. 4.10 clearly corresponds to yearly periodicity together with a half-year harmonic. The peak is not exactly at 365 days because of the combination of the following factors: discretisation of the frequency grid, numerical rounding, and most likely the slight smoothing by $L = 1000$.

A yearly periodicity is discovered, we opt to model it deterministically, as is usual in (scalar) time series. Thus we propose the model

$$Y_{tj} = \mu(x_{tj}) + s_t + X_t(x_{tj}) + \epsilon_{tj} \tag{4.13}$$

where Y_{tj} are the observed measurements at locations x_{tj}, $\mu(\cdot)$ is the intra-day mean, s_t is yearly seasonality adjustment, and the "residual" process $X_t(\cdot)$ is a zero-mean stationary weakly-dependent functional time series. The assumptions of an additive relation of $\mu(\cdot)$ and s_t as well as the stationarity of $X_t(\cdot)$ were justified by exploratory analysis.

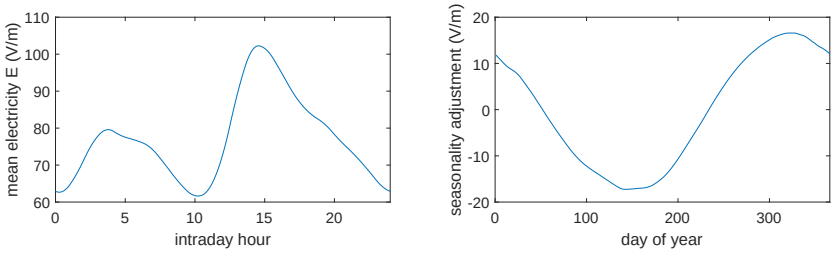

Figure 4.11 – **Left:** The estimated intra-day mean $\hat{\mu}(\cdot)$. **Right:** the estimated yearly seasonality adjustment $\hat{s}_t$

We fit the model (4.13) in the following order. First, we estimate $\mu(\cdot)$ by a local-linear smoother. We expect the mean function to be periodic and assume $\mu(0) = \mu(24)$. Thus we modify the estimator (2.2) to measure the distance between x and x_{tj} as if the endpoints of the interval $[0,24]$ were connected. Having estimated $\hat{\mu}(\cdot)$, we estimate the yearly periodic seasonality adjustment s_t again by a local-linear smoother, again by assuming continuity between first day and last day of the year. The smoothing parameter was chosen by leave-one-year-out cross-validation. Figure 4.11 presents the estimates $\hat{\mu}(\cdot)$ and $\hat{s}_t$. We observe that the intraday mean exhibits two peaks at around 4 a.m. and 3 p.m. The yearly seasonality is almost sinusoidal with low values in the spring and summer and high values in the autumn and winter.

Once the first-order structure given by $\mu(\cdot)$ and s_t is estimated, we calculate the raw covariance (2.3) by subtracting both $\hat{\mu}(x)$ and $\hat{s}_t$. The lag-0 covariance kernel $R_0(\cdot,\cdot)$ is estimated by (2.4). For the estimation of the components of (2.7), namely $\hat{V}(\cdot)$ and $\bar{R}_0(\cdot)$, we use the same periodicity adjustment as for $\hat{\mu}(\cdot)$ because we expect the marginal variance (with and without the ridge contamination) to be continuous across midnight. For illustration and interpretation purposes we estimate also the lag-1 autocovariance $R_1(\cdot,\cdot)$ by (2.8). Figure 4.12 shows the surface plots of these estimates. An interesting element of the estimated lag-0 covariance kernel is the peak at afternoon hours signifying higher marginal variance of the fair-weather atmospheric electricity in the afternoon hours. The estimated lag-0 correlation kernel demonstrates that the observations measured close to each other are highly correlated and the correlation diminishes as the distance grows. The estimated lag-1 autocovariance and autocorrelation kernels show that the correlation between two consecutive days is positive. The lag-1 autocorrelation kernel features a lifted-up surface up to correlation 1 in the eastern corner of the surface plot. This is expected since the late hours of one day are strongly correlated with early morning hours of the following day.

In order to estimate the spectral density consistently, we need to select a moderate value of Bartlett's span parameter L. Plugging in the size of the dataset into the formula (2.52) we set $L = 19$. Figure 4.13 presents a few views on the estimated spectral density kernels.

Once the spectral density is estimated, we apply the functional recovery method of Section 2.2 and estimate the unobserved functional data. The method produces estimates of intraday profiles of fair-weather atmospheric electricity that can be interpreted as the predicted

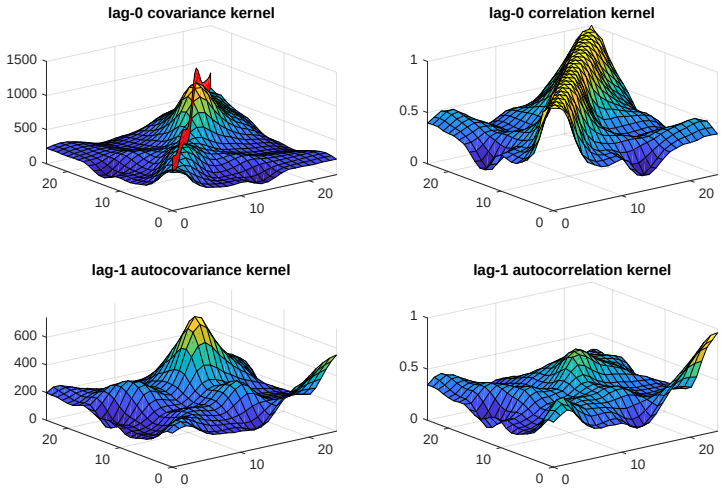

Figure 4.12 – **Top-left:** $\hat{R}_0(\cdot,\cdot)$, the estimated lag-0 covariance kernel $R_0(\cdot,\cdot)$ (surface), the ridge contamination by the measurement error (red) **Top-right:** the correlation kernel corresponding to $\hat{R}_0(\cdot,\cdot)$. **Bottom-left:** $\hat{R}_1(\cdot,\cdot)$, the estimated lag-1 covariance kernel $R_1(\cdot,\cdot)$. **Bottom-right:** the correlation kernel corresponding to $\hat{R}_1(\cdot,\cdot)$.

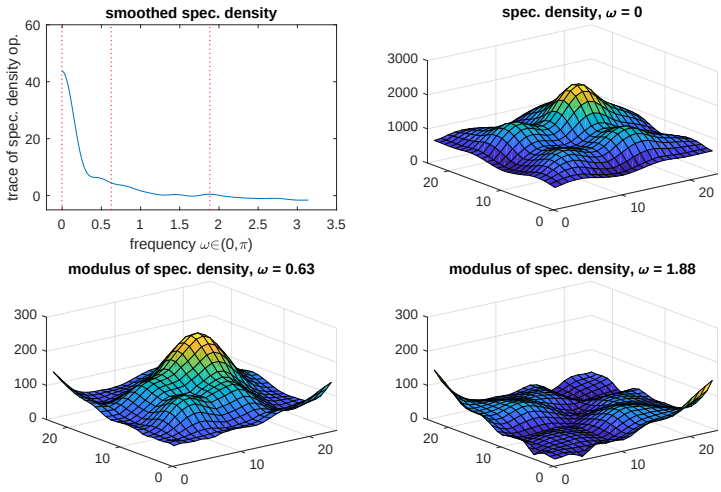

Figure 4.13 – **Top-left:** the traces of estimated spectral density operators with the highlighted frequencies considered in the next plots. **Top-right:** the estimated spectral density at frequency $\omega = 0$ (it is a real-valued kernel for $\omega = 0$). **Bottom-left and bottom-right:** the modulus of the complex-valued estimated spectral density at frequencies $\omega = 0.63$ and $\omega = 1.88$ respectively.

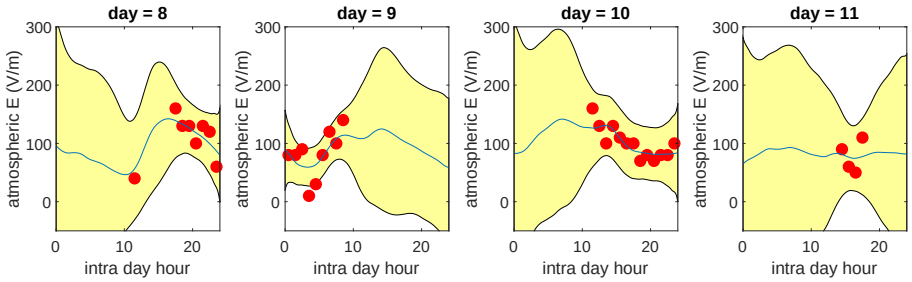

Figure 4.14 – Fair-weather atmospheric electricity hourly measurements (red points) over 4 consecutive days; functional recovery of the latent smooth fair-weather atmospheric electricity process (blue); 95%-simultaneous confidence bands for the functional data of the said latent process (yellow).

atmospheric electricity if the weather was fair at given time, without the modelled noise. As a by-product, the method fills in the gaps in the data (the stretches of days without any measurement). Another output is the construction of confidence bands (under the Gaussianity assumption). Figure 4.14 presents 4 consecutive days with estimated (noiseless) fair-weather atmospheric electricity together with 95%-simultaneous confidence bands. It is important to note that these bands are supposed to cover the assumed smooth underlying functional data, not the observed data produced by adding measurement errors to the smooth underlying process.

The approach used in this case study follows from the aforementioned meteorological methodology suggesting to discard the atmospheric electricity measurements if the fair-weather conditions are not met. One could ask if there is a way to incorporate the unfair-weather atmospheric electricity measurements, perhaps by defining the weights to each observation related to the magnitude in which the fair-weather conditions are broken. Such set-up would then rely on weighted smoothers introduced in Subsection 2.6.2. However, such analysis is wrong. The atmospheric electricity measurements do not become less reliable as the fair-weather conditions are not met, it is a bias due to different ionisation and radiation processes that is introduced. The incorporation of unfair-weather measurements is a question on meteorological research and is out of scope of this thesis.

4.4 Data analysis: visibility data in Wank, Germany

In this section we illustrate the proposed methodology on measurements recorded (Tammet, 2009) at the scientific observatory located at mount Wank located in southern Germany. We remark that this analysis should be seen primarily as an example of the type of data that fall in our framework, rather than a complete data analysis, since there are presumably further important covariates that otherwise should be included. The considered period is January 1, 1977 – December 31, 1979 consisting of $T = 1095$ days. In particular, we analyse the interdependence of three time series:

- *Atmospheric electricity.* The ionisation processes in the atmosphere causes the air to be conductive and the conductivity can be measured in terms of electric potential difference per distance, expressed in volts per meter (V/m). The atmospheric electricity is an important indicator for climate research (Tammet, 2009) and air pollution (Israelsson and Tammet, 2001). However, the atmospheric electricity can be reliably measured only under fair-weather conditions, otherwise the atmospheric ionisation processes are changed and a different quantity is recorded. The standard meteorological methodologies (Xu et al., 2013; Israelsson and Tammet, 2001) suggest to discard the observations under unfair conditions and analyse only those observations recorded under fair weather. Given these guidelines, we take into account only those hourly observations of atmospheric electricity where the wind speed was below $20\,km/h$ and the atmospheric electricity E itself satisfies $0 < E < 250\,V/m$. The meteorological community also advocate discarding the data based on cloud coverage (Xu et al., 2013; Israelsson and Tammet, 2001), but, unfortunately, the mount Wank dataset (Tammet, 2009) does not contain cloud coverage information.

  Based on the above criteria, we eventually retain an unevenly sampled scalar time series which we consequently decatenate into individual days. This technique is useful (Aue et al., 2015; Hörmann et al., 2015a; Hörmann and Kokoszka, 2010; Hörmann et al., 2018) in separating the intra-day variability and the temporal dependence across days. Thus we arrive at a sparsely observed functional time series, where the latent functional data are interpreted as the "atmospheric electricity had the weather been fair".

  In what follows, we denote the fair weather electricity time series as $\{X_t^{(E)}\}_{t=1}^T$. The time series features total of 18326 measurements or 16.7 measurement per day on average.

- *Temperature.* The temperature was recorded hourly at Wank over the considered period. First, we remove the yearly periodicity in the data, then divide the time domain into individual days, and finally convert the hourly observations into functional data using the cubic B-splines with knots at the hourly measurements. The produced fully observed functional time series is denoted as $\{X_t^{(\tau)}\}_{t=1}^T$. The time series includes 21 missing days which we treat as missing completely at random.

- *Recorded visibility.* The reported visibility (in kilometres) was recorded hourly at a range of locations. We define the scalar response time series $\{Z_t\}_{t=1}^T$ as the average visibility on the given day. The time series includes 42 missing values which we treat as missing completely at random.

Since the goal of our analysis is to illustrate the lagged regression methodology and compare the Tikhonov and the truncation regularisation, we split the response time series into two parts, the training component $Z_1,\dots,Z_{822}$ and the test component $Z_{823},\dots,Z_{1095}$ consisting of roughly 75% and 25% of the observations respectively. We fit three models on the entire time span of $\{X_t^{(E)}\}_{t=1}^T$ and/or $\{X_t^{(\tau)}\}_{t=1}^T$ and the training set $\{Z_t\}_{t=1}^{822}$:

- *Atmospheric electricity model* **(E)**. In this model we use the sparsely observed functional

time series $\{X_t^{(E)}\}_{t=1}^T$ as the regressor time series for the response $\{Z_t\}_{t=1}^T$ exploiting the methodology outlined in Chapter 3.

- *Temperature model* **(T)**. This model handles the fully observed functional time series $\{X_t^{(\tau)}\}_{t=1}^T$ as the regressor for the response $\{Z_t\}_{t=1}^T$ using the methodology developed by Hörmann et al. (2015b).

- *Joint model* **(E+T)**. Finally this model includes the information from both the sparsely observed functional time series $\{X_t^{(E)}\}_{t=1}^T$ and the fully observed functional time series $\{X_t^{(\tau)}\}_{t=1}^T$ to predict the response process $\{Z_t\}_{t=1}^T$ outlined by the equation

$$Z_t = a + \sum_{k\in\mathbb{Z}} \mathscr{B}_k^{(E)} X_{t-k}^{(E)} + \sum_{k\in\mathbb{Z}} \mathscr{B}_k^{(\tau)} X_{t-k}^{(\tau)} + e_t \tag{4.14}$$

  which extends the model (3.1) and corresponds to the model discussed in Section 3.3.

  Figure 4.15 displays a schematic visualisation of the prediction for a specific day.

We estimated the filter coefficients in all three models. Figure 4.16 displays the estimated filter coefficients in the joint model **(E+T)**. The filter coefficients estimates in the marginal models **(E)** and **(T)** (not presented here) are very similar to the corresponding filters estimated in **(E+T)** thanks to the fact that the time series $\{X_t^{(E)}\}_{t=1}^T$ and $\{X_t^{(\tau)}\}_{t=1}^T$ are essentially uncorrelated.

A first look at Figure 4.16 reveals that truncation-based estimates (depicted via their Riesz-representers) feature more spikes. This, combined with a worse prediction performance commented upon later, suggests that the Tikhonov regularisation provides a better fit. Therefore we comment only on the interpretation of the filter coefficients estimated by the Tikhonov regularisation. The filter coefficient $\mathscr{B}_0^{(E)}$ is negative, especially in the morning hours, suggesting an obvious interpretation: high atmospheric electricity (which is linked to pollution (Israelsson and Tammet, 2001)) implies a reduction of visibility. The filter coefficients $\mathscr{B}_k^{(\tau)}$ corresponding to the temperature time series reveal an opposite effect. The filters $\mathscr{B}_0^{(\tau)}$ and $\mathscr{B}_{-1}^{(\tau)}$ are positive, therefore high temperatures on the same day and the next day predict a higher visibility today. We recall that our model is not causal, therefore the filter coefficient $\mathscr{B}_{-1}^{(\tau)}$ indeed predicts an effect backwards in time.

Table 4.4 – The mean square prediction error (MSE) and R^2 coefficients of determination of each of the model with either truncation or Tikhonov regularisation based estimation of the filter coefficients. Both the MSE's and the R^2 coefficients were determined on the test partition of the response $\{Z_t\}_{t=823}^{1095}$

Model	Truncation		Tikhonov	
	MSE	R^2	MSE	R^2
(E)	541	0.14	481	0.23
(T)	407	0.35	379	0.39
(E+T)	392	0.37	335	0.46

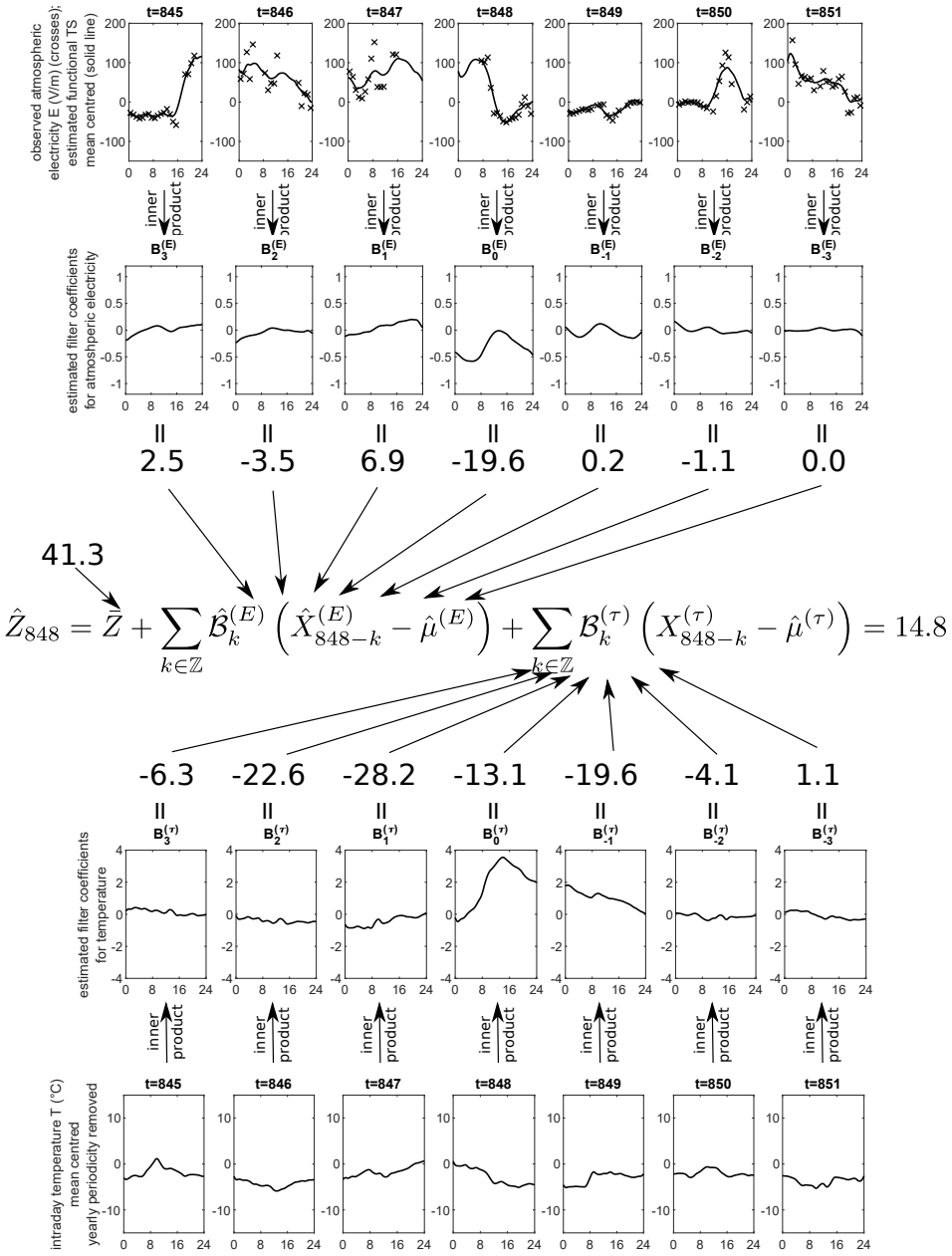

Figure 4.15 – A schema demonstrating prediction of the response time series at time $t = 848$ in the joint model **(E+T)**. Recall that because the response $\{Z_t\}$ is scalar, the filter coefficients $\mathscr{B}_k^{(E)}$ and $\mathscr{B}_k^{(\tau)}$ are functionals and thus can be viewed as inner products with fixed functions which are visualised here. **Top part of the figure:** the contribution of the atmospheric electricity, **bottom part:** the contribution of the temperature.

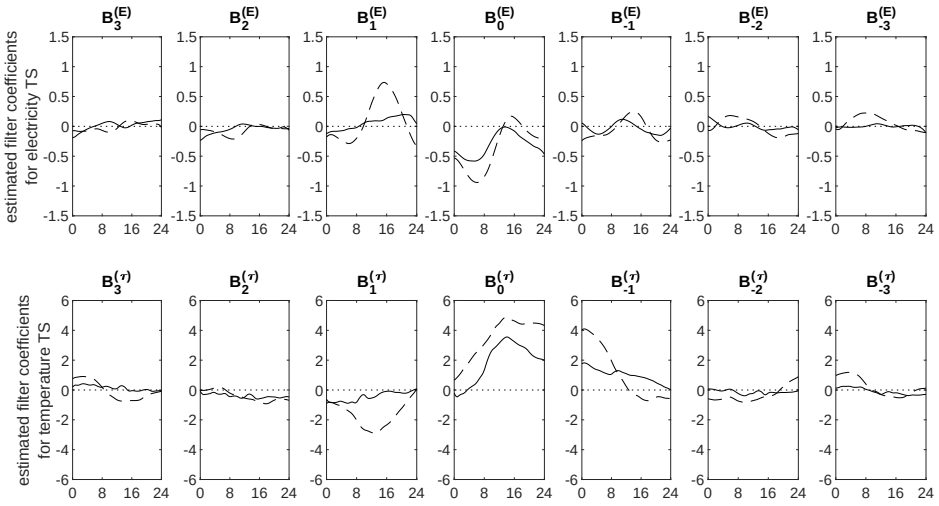

Figure 4.16 – The estimated filter coefficients for lags $k \in \{-3,-2,-1,0,1,2,3\}$ for the joint model **(E+T)**. Solid line: the estimates by Tikhonov regularisation, dashed line: the estimates by truncation regularisation, dotted line: the reference line for zero. **Top row:** the filter coefficients $\mathscr{B}_k^{(E)}$ for the atmospheric electricity, **bottom row:** the filter coefficients $\mathscr{B}_k^{(\tau)}$ for the temperature time series

The results in Table 4.4 represent the prediction performance of the considered models. We calculate the mean square error (MSE) on the test partition $\{Z_t\}_{t=823}^{1095}$ and calculate the R^2 coefficient of determination. The table reveals that the Tikhonov regularisation delivers a better prediction performance for all considered models.

4.4.1 Discussion of the regularisation methods

Chapter 3 presented the methodology for the functional lagged regression problem where the regressor time series is observed sparsely and with noise contamination. We have shown how to estimate the (cross)-spectral density using surface smoothers. The estimation of the spectral transfer function and consequently the filter coefficients using the estimated (cross)-spectral density are ill-posed problems and therefore require regularisation. We considered two regularisation strategies, namely spectral truncation and Tikhonov regularisation, and compared them on a simulation study and the analysis of a data set. In the following we summarise some observations on the differences, strengths, and weaknesses of the two approaches.

The simulation study presented in Section 4.2 illustrates that neither of the two regularisation method can dominate the other. In one of the considered scenarios the spectral transfer function $\mathscr{B}_\omega$ is well-aligned with the leading eigenfunction of the spectral density operator $\mathscr{F}_\omega^X$ thus being estimated better by means of truncation. In the other considered setting, $\mathscr{B}_\omega$ is explained by the fourth, the fifth and the sixth leading eigenvalue of $\mathscr{F}_\omega^X$, which are moreover

nearly tied, resulting in a more challenging estimation task where the stability of Tikhonov regularisation to ties leads to better results than truncation.

The data analysis illustration given in Section 4.4 analysed the dependence of visibility on atmospheric electricity and temperature. The comparative analysis of the the two regularisation methods revealed that the estimates obtained by the Tikhonov regularisation feature better predictive performance. Moreover we found that the filter coefficients estimated by the Tikhonov regularisation were easier to interpret. At least in this application setting, the spectral transfer function $\mathscr{B}_\omega$ does not seem to be well-aligned with the spectral density operator $\mathscr{F}_\omega^X$.

We conclude that both of the regularisation techniques should belong to the statistician's repertoire as neither can dominate the other. However, if we were to choose only one to broadly recommend, this would be the Tikhonov approach, as it seems more robust to "spectral misalignment" and eigenvalue ties, and its theoretical treatment requires fewer assumptions.

4.5 Data analysis: US Treasury yield curve and macroeconomics

The example in this section constitutes a case study analysing the dependence of the US Treasury yield curve on macroeconomic variables. Aim at statistical description of the dependence, we consider the lagged regression model introduced in Chapter 3. The US Treasury yield curve is treated as a sparsely observed functional time series response while the considered macroeconomic variables constitute a multivariate time series predictor. Our non-parametric analysis confirms previous findings established under parametric assumptions, namely a strong impact of the federal funds rate on the short end of the yield curve and a moderate effect of the annual inflation on the longer end of the yield curve.

The considered data, the US Treasury yield curve between the years 1985 and 2000 and the macroeconomic variables, are visualised in Figures 4.17 and 4.18.

4.5.1 Introduction: on yield curve modelling

The *yield curve* is a collection of yields corresponding to traded debt contracts indexed by varying maturity length, ranging from 1 month to 30 years, whose construction is well explained by Filipovic (2009). As an important indicator of the financial sector health, the yield curve is watched closely by traders and investors alike in order to gain understanding about the conditions in financial markets and to discover investment opportunities, and by economists whose analyses provide conclusions about the economic conditions of the national and global economy. Therefore the statistical understanding of the yield curve dynamics is important. The statistical analysis of yield curves is traditionally split into two perspectives: the no-arbitrage approach and the econometric descriptive modelling.

The *no-arbitrage approach* aims to perfectly describe the market data by precisely fitting the term structure in such way that no arbitrage can exist. This fact is quintessential for derivatives

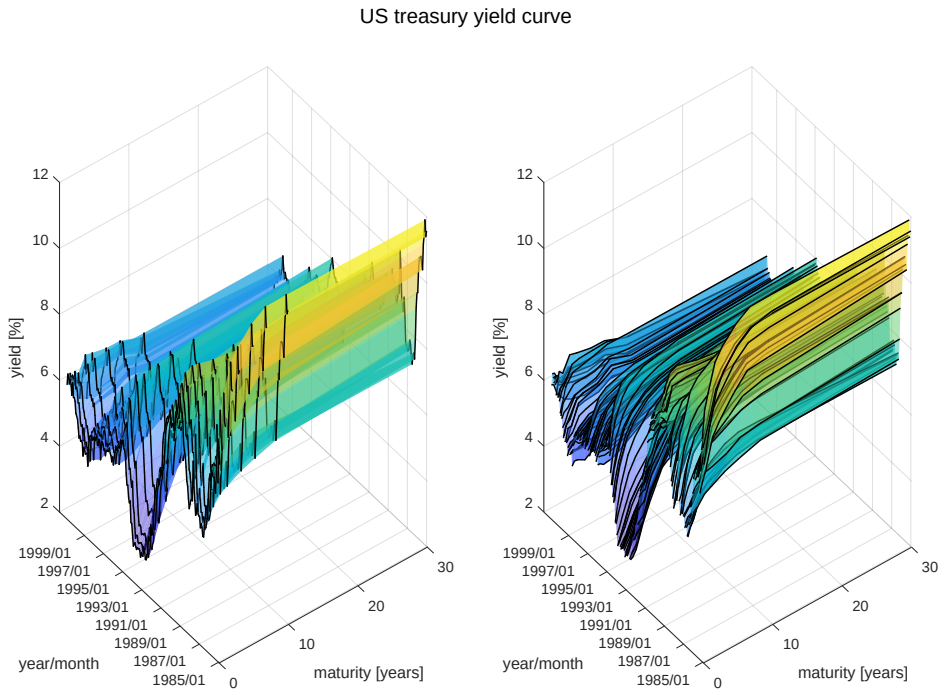

Figure 4.17 – The surface on both the left and the right plots displays the evolution of the US Treasury yield curve between the years 1985 and 2000, interpolating the observed yields linearly. **Left:** the black curves highlight the temporal evolution of the yields at the observed maturities $\tau \in \{1/12, 6/12, 1, 2, 3, 5, 7, 10, 30\}$ (years). **Right:** the black curves constitute the linearly interpolated yields at any given time point, thus plotting the individual yield curves.

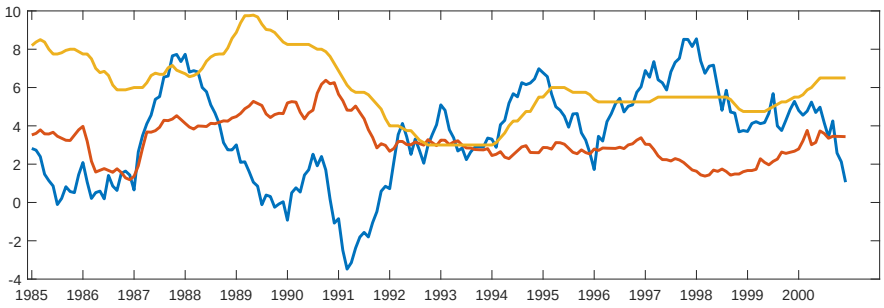

Figure 4.18 – The three macroeconomic time series considered in our analysis. **Blue:** Annual change in industrial production [%]. **Red:** Annual inflation rate [%]. **Yellow:** US federal funds target rate [%].

pricing formulae that are ultimately based on the same no-arbitrage assumption. Notable contributions in no-arbitrage yield curve modelling include Vašíček (1977); Hull and White (1990); Heath et al. (1992); Cox et al. (2005).

The *econometric perspective* of yield curve analysis, which we are going to adopt in this case study, aims at a statistical description of the temporal yield curve evolution. Duffee (2002) argued that while the no-arbitrage models admit a good intra-curve fit, they do not depict well the temporal development of the yield curve and such description is necessary for the modelling of the link with the macroeconomy. In their seminal work, Diebold and Li (2006) extended the *Nelson-Siegel factor model* (Nelson and Siegel, 1987) to model the yield curve dynamics. Their state-space framework admits three latent factors, interpreted as level, slope, and curvature, and turned out to be useful for the yield curve forecasting in a single market as well as in the interaction analysis among numerous markets (Diebold et al., 2008). Moreover, the framework showed the interaction between the US macroeconomic variables and the US Treasury yield curve (Diebold et al., 2006), notably that a positive increase of the federal funds rate, the target rate set by the US Federal Reserve, almost immediately pushes up the slope factor of the yield curve, and a positive increase of the annual inflation influences largely the long-run level of the yield curve. Similar findings have been also demonstrated by Rudebusch and Svensson (1999); Kozicki and Tinsley (2001).

Following the parametric model of Diebold and Li (2006) and its aforementioned variants, the yield curves dynamics have been further explored by the non-parametric functional time series apparatus in the few following articles. Hays et al. (2012) estimated the yield curves dynamics by functional dynamic factor framework where the factor loading curves are estimated non-parametrically using smoothness penalisation, with emphasis oon the factor loading curves interpretation and yield curve forecasting. Kowal et al. (2017) approached the yield curve modelling. fitting functional autoregressive process by the means of Bayesian hierarchical Gaussian models. They derived a Gibbs sampler for inference and forecasting, and conducted an extensive comparative study of yield curve forecasting methods. Finally, Sen and Klüppelberg (2019) estimated the lag-0 covariance operator of the yield curve by the local-polynomial smoothers, estimated the functional principal components scores by the *PACE* methodology (Yao et al., 2005a) and fitted a vector autoregression to the principal component scores.

In our analysis, we consider the novel spectral domain tools for the functional time series modelling established in this thesis. We consider the US Treasury yield curve as sparsely observed functional time series and estimate the cross-dependence between this data set and the US macroeconomic variables using the local-polynomial smoother techniques developed in Chapter 2. We model the dependence between the yield curve and the macroeconomic variables by the means of the functional lagged regression (Chapter 3) and estimated the filter-based regression coefficients. The results of our analysis confirm the findings of Diebold and Li (2006) obtained under parametric assumptions. This fact provides with additional supporting argument in favour of the Nelson-Siegel parametric family and might allow for the

use of the aforementioned parametric models with greater confidence.

4.5.2 Nonparametric estimation and economic implications

We assume the existence of the yields running across the considered interval of maturities $[0,30]$ and denote this curve as Z_t at time t. The collection of this latent curves, $\{Z_t\}_t$ is treated as a sparsely observed functional time series. Concretely, $Z_t(\tau), \tau \in [0,30]$ is considered as a random element in $L^2([0,30];\mathbb{R})$. Moreover we assume that the sample paths of Z_t are continuous and smooth, and that the functional time series $\{Z_t\}$ is stationary, weak-dependent, and with finite second moments. The yield curves are modelled by the observation equation (2.1), that is we assume

$$y_t(\tau_j) = Z_t(\tau_j) + \epsilon_{tj}, \qquad j = 1,\dots,N, \quad t = 1,\dots,T,$$

where $y_t(\tau_j)$, $j = 1,\dots,N$, are the quoted marked-data yields at month t, the independent identically distributed zero-mean additive perturbations ϵ_{tj} constitute the deviation of observed yield from the smooth latent curves Z_t, and $T = 192$ is the number of months in our considered interval ranging from 1985 until 2000.

Denote the vector of the macroeconomic variables at time t by $X_t \in \mathbb{R}^3$ and assume that it is a stationary weak-dependent vector time series with finite second moments, and that it admits the spectral density matrix $\mathbf{F}_\omega^X \in \mathbb{C}^{3\times 3}$ for $\omega \in [-\pi,\pi]$. In order to study the impact of the macroeconomic variables $\{X_t\}$ on the yield curve evolution, we consider a lagged regression model

$$Z_t = a + \sum_{k\in\mathbb{Z}} B_k X_{t-k} + e_t, \qquad t \in \mathbb{Z},$$

where $a \in L^2([0,30];\mathbb{R})$ is the intercept, B_k are linear operators from $\mathbb{R}^3$ to $L^2([0,30];\mathbb{R})$, and $\{e_t\}$ is the model error consisting of independent identically distributed random elements in $L^2([0,30];\mathbb{R})$ that are independent of the vector time series $\{X_t\}$. The primary object of interest of our analysis are the filter coefficients $\mathscr{B}_k$ as they model the link between the economy and the yield curves. The rest of this case study is devoted to their estimation.

Starting with the spectral analysis of the macroeconomic variables $\{X_t\}$ time series, we estimate the spectral density matrices by the Bartlett's formula

$$\hat{\mathbf{F}}_\omega^X = \frac{1}{2\pi} \sum_{h=-L}^{L} W_h \hat{\mathbf{R}}_h^X e^{-\mathrm{i}h\omega}, \qquad \omega \in [-\pi,\pi]$$

where $W_h = (1-|h|)/L$ for $|h| < L$, and zero otherwise, are the Bartlett's weights and $\hat{\mathbf{R}}_h^X$ is the standard empirical lag-h autocovariance matrix for $|h| < L$. We set the Bartlett's span parameter $L = \lceil\sqrt{192}\rceil = 14$. Figure 4.19 visualises the estimated spectral density matrices. The estimated shape of the spectral densities is typical for autoregressive processes with the autoregressive parameter being close to one. This observation might be the first indication

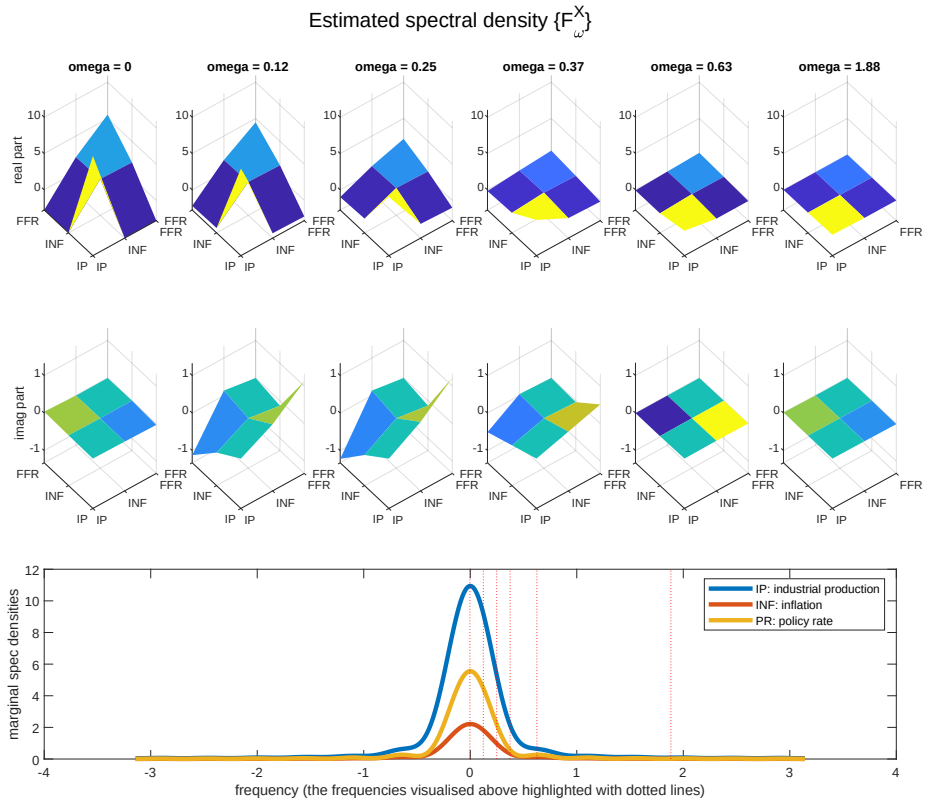

Figure 4.19 – A visualisation of the estimated spectral density matrices $\{\hat{\mathbf{F}}_\omega^X\}$ corresponding to the macroeconomic variable time series $\{X_t\}$. **Top and center rows:** The real and the imaginary parts of the empirical spectral density matrix $\hat{\mathbf{F}}_\omega^X$ evaluated at six different frequencies. **Bottom:** The marginal spectral densities of the individual macroeconomic variables which coincide with the diagonal elements of the spectral density matrices $\{\hat{\mathbf{F}}_\omega^X\}$. The dotted vertical lines denote the frequencies visualised above.

that the parametric model of Diebold et al. (2006), in particular the autoregressive model for the temporal evolution, seems to be provide an appropriate statistical fit.

Secondly, we estimate the mean yield curve defined as $\mu_Z(\tau) = \mathbb{E}[Z_t(\tau)]$, $\tau \in [0,30]$ by the local linear smoother (2.2) with the following modification. Because of the fact that the observed maturities $\tau_1,\dots,\tau_J$ are concentrated rather on the left-hand side of the interval $[0,30]$, we redefine the maturities $\tilde{\tau}_j = (j-1)/(J-1) * 30$ for $j = 1,\dots,J$. The numbers $\tilde{\tau}_1,\dots,\tilde{\tau}_J$ constitute an equidistant partition of $[0,30]$, therefore a local linear line smoother on the scatter-plot $(\tilde{\tau}_j, y_t(\tau_j))$, $j = 1,\dots,J$, $t = 1,\dots,T$ is expected to perform better. Figure 4.20 shows the estimated mean yield curve μ_Z.

Thirdly, we estimate the cross-spectral density $\{\mathscr{F}_\omega^{ZX}\}$ of the functional time series $\{Z_t\}$ and the vector time series $\{X_t\}$ by the local liner line smoother by the methods of Subsection (2.4.2)

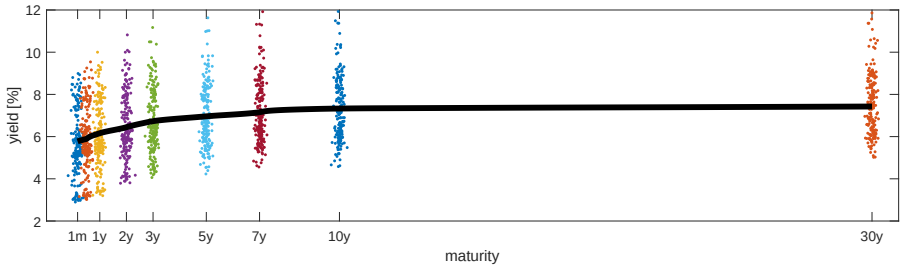

Figure 4.20 – The estimated mean yield curve $\{Z_t\}$ (the black line) obtained by the local-linear line smoother. The data clouds of individual colours constitute the jittered yields at given maturity.

where we use the same transformation for the maturities in order to have an equidistant partition of $[0,30]$. The Bartlett's span parameter is again set by $L = \lceil\sqrt{192}\rceil = 14$. Figure 4.21 visualises the estimated cross-spectral density operators $\{\hat{\mathscr{F}}_\omega^{ZX}\}$.

Finally, the inversion of the relation

$$\mathscr{F}_\omega^{ZX} = \mathscr{B}(\omega)\mathbf{F}_\omega^X, \qquad \omega \in [-\pi,\pi],$$

where $\{\mathscr{B}(\omega)\}_{\omega\in[-\pi,\pi]}$ is the frequency response function between $\{Z_t\}$ and $\{X_t\}$, yields the estimator

$$\hat{\mathscr{B}}(\omega) = \hat{\mathscr{F}}_\omega^{ZX}\left(\hat{\mathbf{F}}_\omega^X\right)^{-1}, \qquad \omega \in [-\pi,\pi].$$

Because the regressor space $\mathbb{R}^3$ is finite dimensional and $\mathbf{F}_\omega^X$ is positive definite, the inversion of $\hat{\mathbf{F}}_\omega^X$ is well-posed and the regularisation is not required. Integration of $\hat{\mathscr{B}}(\omega)$ back into the temporal domain reveals the estimates of the filter coefficients

$$\hat{B}_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{\mathscr{B}}(\omega)e^{\mathrm{i}k\omega}\,\mathrm{d}\omega, \qquad k \in \mathbb{Z}.$$

Figure 4.22 shows the estimates of the filter coefficients $\hat{B}_k$ for $k = -3,-2,\ldots,3$. The interpretation of the estimated filter coefficients is the following:

- The changes of the macroeconomic variables have only imminent impact on the yield curve, the impact is not delayed.
- The industrial production index (IP) has minimal impact on the yield curves, nevertheless a positive increase of the industrial production seems to increase the yield curve at the shortest of maturities by a tiny bit.
- The annual inflation (INF) seems to affect the yield curve more. Concretely, an increase of the inflation results into the increase of the yield curve at higher maturities but hardly affects the short maturities.
- The federal funds rate (FFR) is linked with the yield curve the strongest among the

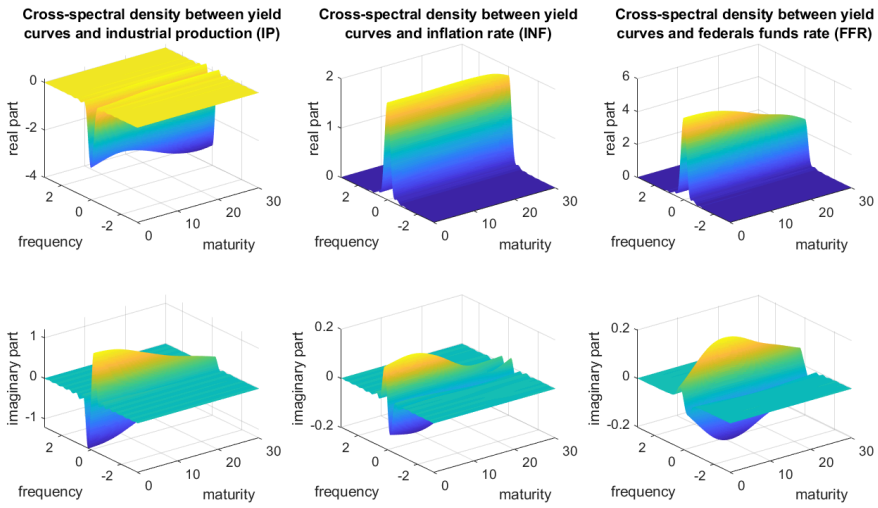

Figure 4.21 – The estimated cross-spectral density $\{\hat{\mathscr{F}}_\omega^{ZX}\}$ between the yield curves time series $\{X_t\}$ and the economic variables time series $\{Z_t\}$. For fixed $\omega \in [-\pi,\pi]$, the operator $\hat{\mathscr{F}}_\omega^{ZX} : \mathbb{R}^3 \to L^2([0,30];\mathbb{C})$ can be visualised component wise for each macroeconomic variable separately; the figure displays the surfaces $(\omega, x) \mapsto \hat{\mathscr{F}}_\omega^{ZX} v \in \mathbb{C}$ where we plug in $v = (1,0,0)$, $v = (0,1,0)$, and $v = (0,0,1)$ in order to visualise the cross-spectral density between the yield curves and the industrial production (first column), inflation rate (second column), and the federal funds rate (third column) respectively.

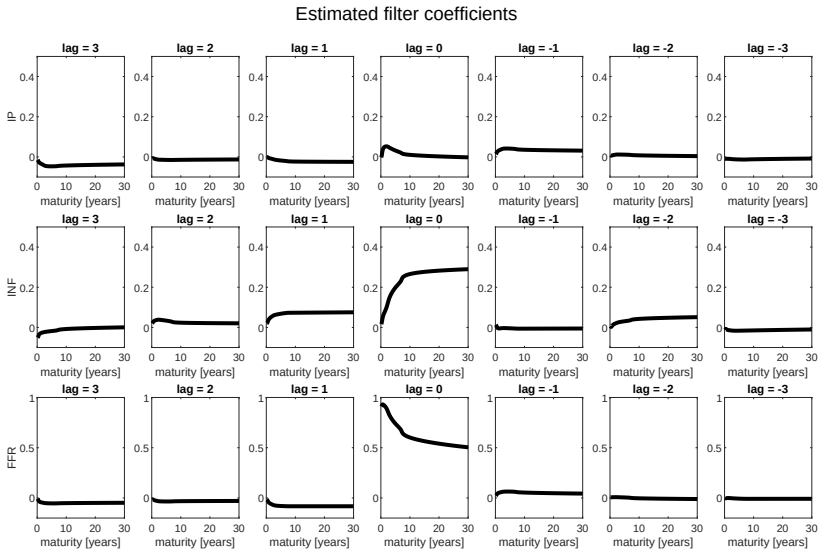

Figure 4.22 – The estimated filter coefficients $\hat{B}_k$ for $k = -3,-2,\ldots,3$. For fixed lag $k \in \mathbb{Z}$, the operator $\hat{B}_k : \mathbb{R}^3 \to L^2([0,30];\mathbb{R})$ can be visualised as a collection of three functions by plotting $\hat{B}_k v$ for $v = (1,0,0)$, $v = (0,1,0)$, and $v = (0,0,1)$ in order to visualise the impact of the industrial production (IP), inflation rate (INF), and the federal funds rate (FFR) respectively on the yield curves.

considered macroeconomic variables. The linkage is the most profound at the short maturities, in particular the filter coefficient function reaches a value close to one at short maturities, signifying that the short end of the market driven yield curve follows closely the federal funds rate. The impact of the federal funds rate is nevertheless significant also for the longer maturities.

Our conclusions reflect the same findings as Diebold et al. (2006) who studied the macroeconomic interactions with the yield curve using a parametric model, modelling the yield curve using the Nelson-Siegel parametric family (Nelson and Siegel, 1987; Diebold and Li, 2006) and assuming a vector autoregressive model of the temporal evolution.

Having estimated the filter coefficients $\{\hat{B}_k\}_{k\in\mathbb{Z}}$ allows us to make predictions of the yield curve based on the macroeconomic variables by the formula

$$\hat{Z}_t(\tau) = \hat{\mu}_Z(\tau) + \sum_{k\in\mathbb{Z}} \hat{B}_k \left(X_{t-k} - \hat{\mu}_X\right), \qquad t = 1,\dots,T, \tag{4.15}$$

where $\hat{\mu}_X = (1/T)\sum_{t=1}^{T} X_t \in \mathbb{R}^3$ is the classical empirical mean of the vector time series $\{X_t\}$. Moreover, we impute the data outside of our observation window by the mean value $X_t := \hat{\mu}_X$ for $t < 1$ or $t > T$. This imputation has however a minimal effect as the estimated coefficients $\hat{B}_k$ are close to zero for $k \neq 0$. By the formula (4.15) we predict the yield curves given the macroeconomic variables $\{X_t\}$ and compare the prediction with the observed yield for observable maturities. The R^2 coefficient of determination of such model evaluation yields a quite high value of 0.78.

4.5.3 Code availability

The MATLAB code supporting the results of this case study is available as a repository on https://github.com/tomasrubin/us-yield-curve-macroeconomics.

5 Spectral simulation of functional time series

Any methodological development in functional time series will be accompanied by a finite sample performance assessment of the novel method, given the complexity of the data involved. Such simulations require the generation of functional time series with prescribed model dynamics. Despite many new methods being generally applicable to time series (whether linear or not), their assessments is carried out predominately on simulated data coming from functional autoregressive moving average (FARMA) processes, typically only functional autoregressive (FAR) processes, because their simulation is straightforward in the time-domain by applying the autoregressive equation sequentially on white noise (or a moving average of white noise). In order to assess the applicability of a method beyond linear processes, however, one should aim to cover as broad as possible a range of possible functional time series dynamics (including non-linear dynamics). This is especially true for methods that are not specific to linear processes but whose assumptions, theory, and implementation are more generally valid. Indeed, many functional time series methods (Hörmann et al., 2015a,b; Zhang, 2016; Tavakoli and Panaretos, 2016) rely on the eigendecomposition of spectral density operators (the harmonic/dynamic principal components) and present performance trade-offs that are best captured by their spectral structure. It is thus beneficial to be able to simulate functional time series specified by means of their spectral density structure.

The objective of this chapter is to develop a general-purpose simulation method that is able to efficiently simulate stationary functional time series not restricted to the linear class. The approach is to use the spectral specification of such a time series, by means of its *spectral density operator*. The general method, presented in Section 5.1, hinges on a discretisation and dimension reduction of the functional Cramér representation (Panaretos and Tavakoli, 2013a). It simulates an ensemble of independent complex random elements whose covariance operators match the designated spectral density operators, and transposes this ensemble into the time-domain by the means of the (inverse) fast Fourier transform (Cooley and Tukey, 1965). We show that this strategy is particularly effective when the series is defined by means of the eigendecomposition of its spectral density operator or by filtering a white noise, but consider various other specification scenarios, too.

Our functional time series simulation method in the spectral domain is inspired in part by the methods for scalar and multivariate time series simulation. The original idea of simulating a signal in the spectral domain and converting it to the time-domain by the inverse fast Fourier transform seems to be due to Thompson (1973). This approach was further explored by Percival (1993) who reviewed some variants of the algorithm and addressed some practical implementation questions, and Davies and Harte (1987) used the method for simulation of fractionally integrated noise processes. Furthermore, the simulation of multivariate time series with given spectral density matrices is due to Chambers (1995). However, pushing the general ideas forward to functional time series is not a matter of simple generalisation of the multivariate time series simulation methods. The intrinsic infinite dimensionality of functional data calls for the approximate generation of infinite dimensional objects approximated in finite dimension, which requires optimally reducing dimension (which we implement either via the Karhunen-Loéve or the Cramér-Karhunen-Loève representation (Panaretos and Tavakoli, 2013a)) and/or judicious discretisation (pixelisation) of the spatial domain (the argument of each function). An additional side effect of this, in contrast to the multivariate case, is that one must pay particular attention that the simulation algorithms scale well as the discretisation resolution refines and the dimension parameter grows, and these need to be incorporated in the time complexity assessments.

Our spectral domain simulation method constitutes a general approach, is able to simulate arbitrary functional time series that are specified in the frequency domain, with additional computational speed-ups that can be realised when assuming a special structure of the spectral density operators. In particular, simulation of the important FARFIMA(p,d,q) processes can be much faster in the spectral domain than in the time-domain, while the spectral domain simulation of FARMA(p,q) processes is competitive with time-domain methods.

The results in this chapter are based on Rubín and Panaretos (2020c).

5.1 Simulation of functional time series with given spectrum

We will present a functional time series simulation method in the spectral domain. We focus our presentation on functional time series with values in $\mathscr{H} = L^2([0,1],\mathbb{R})$ whose trajectories are continuous and whose spectral density operators are integral operators with continuous kernels, but note that our discussion equally applies to other function spaces constituting separable Hilbert spaces.

The objective of the simulation is to generate a Gaussian sample $X_1,\dots,X_T$ for some $T \in \mathbb{N}$ given the spectral density operator $\{\mathscr{F}_\omega^X\}_{\omega\in[-\pi,\pi]}$. Without loss of generality, we assume that T is even and we furthermore define the canonical frequencies

$$\omega_k = \begin{cases} (2\pi k)/T, & k = 1,\dots,T/2, \\ (2\pi(k-T))/T, & k = T/2+1,\dots,T. \end{cases}$$

At a high level, our spectral domain simulation method mimics the discrete approximation of the Cramér representation (1.35), which boils down to performing the following two steps.

1. Generate an ensemble of independent complex mean-zero Gaussian random elements Z'_k, $k = 1, \dots, T/2, T$ such that

$$\mathbb{E}\left[Z'_k \otimes Z'_k\right] = \mathscr{F}^X_{\omega_k}, \qquad k = 1, \dots, T/2, T, \tag{5.1}$$

   and, for $k = 1, \dots, T/2 - 1$, generate independent copies Z''_k thereof. Define

$$Z_k = \begin{cases} \sqrt{2} Z'_k & k = T/2, T, \\ Z'_k + \mathrm{i} Z''_k & k = 1, \dots, T/2 - 1, \\ Z'_{T-k} - \mathrm{i} Z''_{T-k} & k = T/2 + 1, \dots, T/2 - 1. \end{cases} \tag{5.2}$$

2. By the inverse fast Fourier transform algorithm calculate

$$X_t = \left(\frac{\pi}{T}\right)^{1/2} \sum_{k=1}^{T} Z_k e^{\mathrm{i} t \omega_k}, \qquad t = 1, \dots, T. \tag{5.3}$$

   The formula (5.2) ensures that the sample of $\{Z_k\}$ is symmetric and thus inverse Fourier transform constitutes a real-valued functional time series, as will be proved later in Theorem 5.1.1.

While the application of the inverse fast Fourier transform in Step 2 of the algorithm is computationally fast, the generation of the complex random elements $\{Z'_k\}$ in Step 1, whose covariance operators may in general have no structure in common, is not a trivial matter, and is discussed in the next three subsections, for three different specifications of the operator $\mathscr{F}^X_{\omega_k}$. In Subsection 5.1.1, these random elements are generated by their Karhunen-Loève expansions, therefore essentially enacting the Cramér-Karhunen-Loève representation (1.37). On the other hand, the filtering specification discussed in Subsection 5.1.2 leverages the special structure of the filtered white noise spectral density operators to generate the random elements $\{Z_k\}$ efficiently. This approach is further tailored to simulation of FARFIMA processes in Subsection 5.1.3.

Before moving on to the specifics, though, we establish that the sample generated by formula (5.3) will indeed follow the correct dependence structure:

Theorem 5.1.1. *Assume either of the two following conditions:*

(i) The condition (1.22) *holds and thus the spectral density operator* $\{\mathscr{F}^X_\omega\}_{\omega \in [-\pi, \pi]}$ *exists in the sense of Proposition 1.3.5.*

(ii) The weak spectral density operator $\mathscr{F}^X_\omega \in L^1([-\pi, \pi], \mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$ *is continuous with respect to the norm* $\|\cdot\|_1$ *on* $(-\pi, 0) \cap (0, \pi)$*, and we additionally set* $\mathscr{F}^X_0 = 0$.

Then, the functional time series sample $X = \{X_t\}_{t=1}^T$ *generated by* (5.3) *is a real-valued stationary Gaussian time series of zero mean, and asymptotically admits* $\{\mathscr{F}_\omega\}$ *as its weak spectral density operator when* $T \to \infty$.

Proof. The Gaussianity, stationarity, and mean-zero properties of $X_1, \dots, X_T$ are clear thanks to linearity.

First we show that the functional time series defined by (5.3) is real-valued. For $k = 1, \dots, T/2-1$ we have that

$$Z_k e^{\mathrm{i}\, t\omega_k} + Z_{T-k} e^{\mathrm{i}\, t\omega_{T-k}} = Z_k e^{\mathrm{i}\, t\omega_k} + \overline{Z_k} e^{-\mathrm{i}\, t\omega_k} = 2\Re\{Z_k e^{\mathrm{i}\, t\omega_k}\} \quad \in \mathbb{R}.$$

For $k = T/2$ or $k = T$, the spectral density operator $\mathscr{F}_{\omega_k}^X$ is real, thus Z_k is real-valued, and $e^{\mathrm{i}\, t\omega} \in \{-1, 1\}$ for $\omega \in \{\pi, 2\pi\}$. Therefore (5.3) defines a real-valued functional time series.

Let us calculate the lag-h autocovariance operators of (5.3) for $h \in \mathbb{N}$.

$$\begin{aligned} \mathbb{E}[X_{t+h} \otimes X_t] &= \frac{\pi}{T} \mathbb{E}\left[\left(\sum_{k=1}^T Z_k e^{\mathrm{i}(t+h)\omega_k} \right) \otimes \left(\sum_{l=1}^T Z_l e^{\mathrm{i}\, t\omega_l} \right) \right] \\ &= \frac{\pi}{T} \sum_{k=1}^T \sum_{l=1}^T \mathbb{E}[Z_k \otimes Z_l]\, e^{\mathrm{i}(t+h)\omega_k} e^{-\mathrm{i}\, t\omega_l} \end{aligned} \tag{5.4}$$

We shall calculate the term $\mathbb{E}[Z_k \otimes Z_l]$ on the right-hand side of (5.4). Firstly, $\mathbb{E}[Z_k \otimes Z_k] = 2\mathscr{F}_{\omega_k}^X$ for $k \in \{T/2, T\}$, and $\mathbb{E}[Z_k \otimes Z_l] = 0$ for $k \in \{T/2, T\}$ and $l \neq k$.

Secondly, fix $k \in \{1, \dots, T/2\}$. Then

$$\begin{aligned} \mathbb{E}[Z_k \otimes Z_k] &= \mathbb{E}\left[Z_k' \otimes Z_k' + \mathrm{i}\, Z_k'' \otimes Z_k' - \mathrm{i}\, Z_k' \otimes Z_k'' + Z_k'' \otimes Z_k'' \right] = 2\mathscr{F}_{\omega_k}^X, \\ \mathbb{E}[Z_k \otimes Z_{T-k}] &= \mathbb{E}\left[Z_k' \otimes Z_k' + \mathrm{i}\, Z_k'' \otimes Z_k' + \mathrm{i}\, Z_k' \otimes Z_k'' - Z_k'' \otimes Z_k'' \right] = \\ &= \mathbb{E}\left[Z_k' \otimes Z_k' - Z_k'' \otimes Z_k'' \right] = 0. \end{aligned}$$

Furthermore, for $l \notin \{k, T-k\}$, we have $\mathbb{E}[Z_k \otimes Z_l] = 0$ from the independence of Z_k's.

We continue with the calculations on (5.4) as

$$\mathbb{E}[X_{t+h} \otimes X_t] = \frac{2\pi}{T} \sum_{k=1}^T \mathscr{F}_{\omega_k}^X e^{\mathrm{i}\, h\omega_k}. \tag{5.5}$$

The right-hand side of (5.5) constitutes the Riemann sum of the integral (1.31). The convergence of the Riemann sums (5.5), as $T \to \infty$, towards (1.31) is justified by the assumptions (i) or (ii). □

Due to the periodicity of Fourier transform, the values X_1 and X_T will tend to be similar

which might be an undesirable trait, depending on the application. To overcome this artefact, Mitchell and McPherson (1981); Percival (1993) propose to simulate a sample of length $\tilde{T} = kT$ for some integer $k \geq 2$ and sub-sample a functional time series of length T.

5.1.1 Simulation under spectral eigendecomposition specification

Perhaps the most direct means to generate (approximate versions of) the random elements $\{Z'_k\}$ considered in Step 1 of the algorithm introduced at the beginning of Section 5.1 is by means of a finite rank approximation to the spectral density operator at the corresponding frequencies, appearing in the definition (see equation (5.1)). For a given rank, the optimal such approximation is obtained by truncating the eigenexpansion (1.36) at that value, thus using a finite number of the harmonic eigenfunctions and corresponding eigenvalues to approximately generate $\{Z'_k\}$.

Concretely, denoting $\{\lambda_n(\omega)\}_{n=1}^{\infty}$ and $\{\varphi_n(\omega)\}_{n=1}^{\infty}$ the harmonic eigenvalues and the harmonic eigenfunctions of the spectral density operator $\mathscr{F}_{\omega}^{X}$ at the frequency $\omega \in [-\pi, \pi]$, we may generate exact versions of Z'_k by setting

$$Z'_k = \sum_{n=1}^{\infty} \sqrt{\lambda_n(\omega_k)}\varphi_n(\omega_k)\xi_n^{(k)} \tag{5.6}$$

where $\{\xi_n^{(k)}\}$ is an ensemble of i.i.d. standard Gaussian real-valued random variables. The random elements defined by (5.6) clearly satisfy the requirement (5.1). In practice one has to truncate the series in (5.6) at a finite level, say N. This truncation is optimal in terms of preserving the second order structure of the functional time series (Proposition 1.3.11) and requires only a low number of inexpensive operations. If we are to evaluate the functional time series X on a spatial grid of $[0,1]$ at resolution $M \in \mathbb{N}$, the simulation requires $O(NMT + MT\log T)$ operations, provided we have direct access to the decomposition (1.36). The $O(MT\log T)$ comes from the inverse discrete Fourier transform (5.3).

When the decomposition (1.36) is not directly available, as for example is the case for the FARMA(p,q) process with non-trivial autoregressive part, the evaluation of the spectral density operator (1.63) requires inversion of a bounded linear operator different at each frequency ω. Unless a special structure of the autoregressive operator is assumed (e.g. as in Example 5.2.2), the evaluation of this inversion is expensive. One could discretise the operator on a grid of $[0,1]^2$ and invert the resulting matrix, but this will become slow for dense grids, especially considering to do it for each frequency ω_k, $k = 1, \ldots, T/2, T$. Moreover, to obtain the harmonic eigenvalues and eigenfunctions (1.36) one would need to perform the eigendecomposition at each frequency ω_k which is also slow for large matrices. These operations, if performed on a spatial grid of resolution $M \times M$, require $O(M^3)$ operations, bringing the overall cost to $O(M^3 T + MT\log T)$. This can be reduced by calling a truncated eigendecomposition algorithm instead, e.g. the truncated singular value decomposition (SVD) algorithm, and evaluating only $N < M$ eigenfunctions. This yields computational gains when $N \ll M$, namely reducing the

complexity of the said operations from $O(M^3)$ to $O(NM^2)$, and the overall cost to $O(NM^2T + MT\log T)$.

Though the simulation cost is high when the decomposition (1.36) is not directly available, the approach still constitutes a general method to simulate a functional time series with arbitrary spectrum. Example 5.2.1 illustrates a functional time series whose dynamics are defined through its Cramér-Karhunen-Loève expansion where we show that simulation is possible even when we do not leverage our knowledge of this expansion, but rather calculate it numerically.

Finally, it is worth remarking that even though the functions $\{\varphi_n(\omega)\}_{n=1}^{\infty}$ appearing in (5.6) are orthonormal for each $\omega \in [-\pi,\pi]$, orthonormality is not *required* for the correct simulation of $\{Z_k'\}$ by (5.6). In other words, a practitioner can specify a spectral density operator by a sum similar to (1.36) without insisting on using orthonormal functions, and still achieve rapid simulation in the spectral domain.

5.1.2 Simulation under filtering specification

The second implementation of Step 1 of the abstract algorithm introduced at the beginning of Section 5.1 leverages the set-up where a white noise with covariance operator $\mathscr{S}$ is plugged into a filter with given frequency response function $\Theta(\omega)$ in which case the spectral density operator is given directly by the formula

$$\mathscr{F}_{\omega}^{X} = \frac{1}{2\pi}\Theta(\omega)\mathscr{S}\Theta(\omega)^*, \qquad \omega \in [-\pi,\pi], \tag{5.7}$$

where $\mathscr{S}$ is a non-negative definite self-adjoint trace class operator and $\Theta : [-\pi,\pi] \to \mathscr{L}(\mathscr{H}^{\mathbb{C}})$, i.e. $\Theta(\omega)$ is a bounded linear operator on $\mathscr{H}^{\mathbb{C}}$ for each $\omega \in [-\pi,\pi]$. We only require that

$$\int_{-\pi}^{\pi} \|\Theta(\omega)\|^2_{\mathscr{L}(\mathscr{H}^{\mathbb{C}})}\,\mathrm{d}\omega < \infty$$

and $\Theta(\omega)g = \overline{\Theta(2\pi-\omega)(g)}$ for $\omega \in [0,\pi]$ and $g \in \mathscr{H}^{\mathbb{C}}$, which implies that $\{X\}$ is a stationary mean-zero functional time series admitting the weak spectral density operator $\mathscr{F}^X \in L^1([-\pi,\pi],\mathscr{L}_1(\mathscr{H}^{\mathbb{C}}))$.

The operator $\mathscr{S}$, being a non-negative definite self-adjoint trace class operator, admits the decomposition

$$\mathscr{S} = \sum_{n=1}^{\infty} \eta_n e_n \otimes e_n \tag{5.8}$$

where $\{\eta_n\}$ are the eigenvalues and $\{e_n\}$ are the eigenfunctions of $\mathscr{S}$.

We may simulate real random elements $\{Y_k\}$, corresponding the sequence of independent

identically distributed random variables with the covariance operator $\mathscr{S}$, by setting

$$Y_k = \sum_{n=1}^{\infty} \sqrt{\eta_n} e_n \tilde{\xi}_n^{(k)} \tag{5.9}$$

with an ensemble $\{\tilde{\xi}_n^{(k)}\}$ of i.i.d. standard Gaussian random variables. In practice, the sum (5.9) is truncated at some $N \in \mathbb{N}$.

If the decomposition (5.8) is unknown, it can be numerically calculated by discretisation of the kernel corresponding to the operator $\mathscr{S}$ on the grid of $[0,1]^2$, say constituting an $M \times M$ matrix, and numerically calculating its eigendecomposition, in which case we may select $N = M$ eigenvalues and eigenvectors. The advantage of this approach over numerical evaluation of the spectral density operators at each ω, performing the numerical eigendecomposition of each spectral density operator, and applying the Cramér–Karhunen–Loève-based simulation algorithm presented in Subsection 5.1.1 is that the filtered white noise approach requires only one runtime of this expensive step.

Having defined the random elements $\{Y_k\}$ by (5.9), we define the elements $\{Z'_k\}$ in the notation of the algorithm presented at the beginning of Section 5.1 by putting

$$Z'_k = \frac{1}{\sqrt{2\pi}} \Theta(\omega_k) Y_k, \qquad k = 1, \dots, T/2, T. \tag{5.10}$$

Such $\{Z'_k\}$ obviously satisfy (5.1).

If the decomposition (5.8) is unknown and we opt to numerically evaluate it on a grid of size M, the total computational complexity turns out to be $O(M^3 + M^2 T + MT \log T)$ where $O(M^2 T)$ comes from the matrix application (5.10) and $O(MT \log T)$ from the inverse fast Fourier transform (5.3).

5.1.3 Simulation under linear time domain specification

One of the typical functional time series dynamics specifications is a linear process in the time domain. In this subsection we consider the flexible class of the FARFIMA(p,d,q) processes, one of the most general classes of such linear processes, and show how to generate their trajectories by spectral domain simulation methods.

The FARFIMA(p,d,q) processes, thanks to being defined as a linear filter of white noise, admit the spectral density operators of the form (5.7). However, the application of the simulation algorithm presented in Subsection 5.1.2 requires the frequency response function $\Theta(\omega)$ to be readily available, which is not always the case: the FARFIMA(p,d,q) (or FARMA(p,q)) process with a non-degenerate autoregressive part admit the frequency response function given by the formula prompting operator inversion:

$$\Theta(\omega) = \mathcal{A}(e^{-\mathrm{i}\omega})^{-1} \mathcal{B}(e^{-\mathrm{i}\omega}), \qquad \omega \in [-\pi, \pi]. \tag{5.11}$$

Therefore a naive implementation would require inversion of the linear bounded operator $\mathcal{A}(e^{-\mathrm{i}\omega})$ for each frequency ω. It may very well happen that $\mathcal{A}(e^{-\mathrm{i}\omega})$ has special structure, e.g. as is the case for the FARFIMA(1,d,0) process considered in Example 5.2.2, in which case the inversion evaluation is rapid. In the general case, however, the inversion on a spatial domain discretisation would require $O(M^3)$ operations where M is the discretisation resolution. Fortunately, there are two ways to avoid this computational cost:

- A *fully spectral approach* which consists in the efficient evaluation of (5.10). The discretization of this formula for the FARFIMA(p,d,q) process involves evaluation of

$$Z'_k = \frac{[2\sin(\omega/2)]^{-d}}{\sqrt{2\pi}}\mathbf{A}(e^{-\mathrm{i}\omega_k})^{-1}\mathbf{B}(e^{-\mathrm{i}\omega_k})\,Y_k \tag{5.12}$$

  where the matrices $\mathbf{A}(e^{-\mathrm{i}\omega_k})$ and $\mathbf{B}(e^{-\mathrm{i}\omega_k})$ are the discretized operators $\mathcal{A}(e^{-\mathrm{i}\omega_k})$ and $\mathcal{B}(e^{-\mathrm{i}\omega_k})$ respectively. The numerical evaluation of (5.12) requires solving the matrix equation

$$\mathbf{A}(e^{-\mathrm{i}\omega_k})\,Z'_k = \frac{[2\sin(\omega/2)]^{-d}}{\sqrt{2\pi}}\mathbf{B}(e^{-\mathrm{i}\omega_k})\,Y_k,$$

  thus resulting in $O(M^2)$ complexity, as opposed to the $O(M^3)$ complexity of matrix inversion.

- A *hybrid simulation approach*, where we simulate the FARFIMA(p,d,q) processes by simulating the corresponding FARFIMA$(0,d,q)$ process in the spectral domain and then applying the autoregressive recursion in the time-domain. Concretely, we:

  1. Choose a burn-in length $\tilde{T} \gg p$, and simulate a FARFIMA$(0,d,q)$ process with degenerate autoregressive part, denoted as $X'_1,\ldots,X'_{T+\tilde{T}}$, by the means of the tools in Subsection 5.1.2. Such a functional time series admits the spectral density operator

$$\mathscr{F}^{X'}_{\omega} = \frac{[2\sin(\omega/2)]^{-2d}}{2\pi}\mathcal{B}(e^{-\mathrm{i}\omega})\mathscr{S}\mathcal{B}(e^{\mathrm{i}\omega})^*$$

     whose corresponding frequency response function $\Theta(\omega) = [2\sin(\omega/2)]^{-d}\mathcal{B}(e^{-\mathrm{i}\omega})$ can be evaluated fast.

  2. Set $X_1,\ldots,X_p = 0$ and run the recursion

$$X_t = \mathscr{A}_1 X_{t-1} + \cdots + \mathscr{A}_p X_{t-p} + X'_t, \qquad t = p+1,\ldots,T+\tilde{T}.$$

  3. Discard the first $\tilde{T}$ values of $X_1,\ldots,X_{T+\tilde{T}}$ and keep only the last T elements.

Both the fully spectral and the hybrid implementations involve the numerical eigendecomposition of the noise covariance operator $\mathscr{S}$, incurring an $O(M^3)$ computation cost, the applications of matrices on vectors or solving linear equations, yielding $O(M^2T)$ operations, and the inverse fast Fourier transform at each point of the discretisation with the $O(MT\log T)$

complexity. Thus the total computational complexity is $O(M^3 + M^2 T + MT \log T)$. Nevertheless, even though the application of a matrix on a vector has the same complexity as solving a linear system of equations, the constant hidden in the "O" is different and the hybrid simulation method is faster than the fully spectral approach, which requires the solution of linear systems at each frequency, as the simulation study in Example 5.2.3 demonstrates.

5.2 Examples and numerical experiments

This section presents three examples of functional time series specified according in various ways where the spectral density operator may be directly or indirectly defined, depending on the scenario. The examples are accompanied by a small simulation study assessing the simulation speed and the simulation accuracy by comparing the lagged autocovariance operators of the simulated processes with the ground truth. The purpose of the simulation study is to illustrate the performance of the method in terms of speed and accuracy, and draw some qualitative conclusions about the choice of methods and parameters, rather than to provide with an extensive quantitative comparison.

A parallel objective is to provide code that is accessible (Section 5.4), simple to run, and easy to tailor for custom-defined spectral density operators used in functional time series research.

5.2.1 Specification by spectral eigendecomposition

Consider the spectral density operator defined by its eigendecomposition

$$\begin{aligned} \mathscr{F}_\omega^X &= \sum_{n=1}^{\infty} \lambda_n(\omega)\varphi_n(\omega) \otimes \varphi_n(\omega), \qquad \omega \in [-\pi, \pi], \\ \lambda_n(\omega) &= \frac{1}{(1 - 0.9\cos(\omega))\pi^2 n^2}, \qquad \omega \in [-\pi, \pi], \\ \big(\varphi_n(\omega)\big)(x) &= \begin{cases} \sqrt{2}\sin(n(\pi\delta_{\omega/\pi}(x)), & x \in [0,1], \quad \omega \in [0, \pi], \\ \sqrt{2}\sin(n(\pi\delta_{-\omega/\pi}(x))), & x \in [0,1], \quad \omega \in (\pi, 2\pi], \end{cases} \end{aligned} \tag{5.13}$$

where

$$\delta_a(\cdot) = x - a \mod 1$$

is the periodic shift by $a \in \mathbb{R}$ with "mod" denoting the modulo operation, the remainder after the division. Under such definition, which guarantees that $\delta_a(x) \in [0,1]$, the harmonic eigenfunctions at distinct frequencies are phase-shifted versions of each other. It turns out that the spectral density operator given by the sum (5.13) can be expressed in closed analytical form, as an integral operator with kernel

$$f_\omega^X(x,y) = \begin{cases} \frac{1}{(1-0.9\cos(\omega))} K_{BB}(\delta_{\omega/\pi}(x), \delta_{\omega/\pi}(x)), & \omega \in [0, \pi], \\ \frac{1}{(1-0.9\cos(\omega))} K_{BB}(\delta_{-\omega/\pi}(x), \delta_{-\omega/\pi}(x)), & \omega \in (\pi, 2\pi]. \end{cases} \tag{5.14}$$

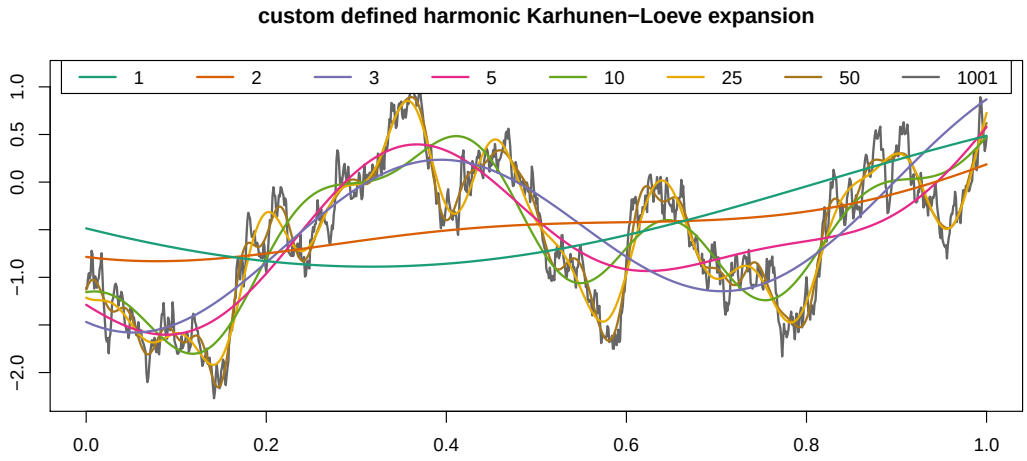

Figure 5.1 – Sample trajectories $X_1(\cdot)$ of the process defined in Example 5.2.1 with varying number of harmonic principal components N chosen in the truncation of (5.6). Simulated with $T = 100$ and the grid resolution $M = 1001$.

where $K_{BB}(\cdot,\cdot)$ is the covariance kernel of Brownian bridge (Deheuvels and Martynov, 2003) defined as

$$K_{BB}(x,y) = \min(x,y) - xy, \qquad x,y \in [0,1].$$

Figure 5.1 illustrates the simulated trajectories of X_1 with varying number of the harmonic principal components N used in the truncation of the sum (5.6) when simulating by the means presented in Subsection 5.1.1.

In order to assess the simulation accuracy we opt to: simulate $I = 1000$ independent realisations $\{X_t^{(1)}\}_{t=1}^T, \ldots, \{X_t^{(I)}\}_{t=1}^T$ of the process $\{X_t\}_{t\in\mathbb{Z}}$; evaluate its empirical autocovariance operators $\hat{\mathscr{R}}_{h,[i]}^X$ for each $i = 1,\ldots,I$ and some lags h; and define the average empirical autocovariance operator $\overline{\mathscr{R}_h^X} = \frac{1}{I}\sum_{i=1}^I \hat{\mathscr{R}}_{h,[i]}^X$. We then compare this with the true covariance operator $\mathscr{R}_h^X$ by calculating

$$rel.error(h) = \frac{\left\| \overline{\mathscr{R}_h^X} - \mathscr{R}_h^X \right\|_1}{\left\| \mathscr{R}_0^X \right\|_1}, \qquad \text{for some lags } h. \tag{5.15}$$

The true autocovariance operators $\mathscr{R}_h^X$ were calculated by numerically integrating (1.31).

Figure 5.2 the manner of error decay as $N \to \infty$ and the number of harmonic components $N = 100$ seems to be satisfactory. The relative simulation errors for $N > 100$ seem to be dominated by the random component of (5.15) rather than the simulation error itself. We note that the kernel of the spectral density operator (5.14) is non-differentiable near the spatial diagonal, and consequently features a relatively slow (quadratic) decay of its eigenvalues. It thus repre-

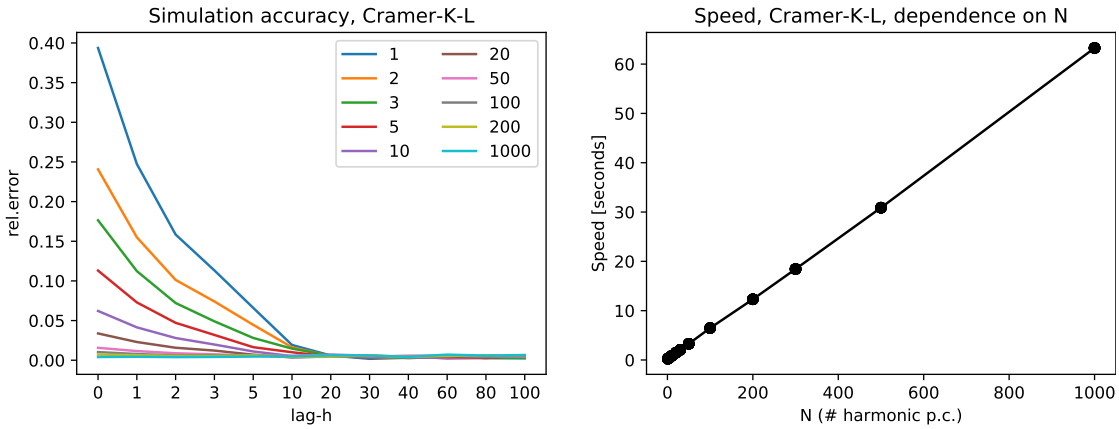

Figure 5.2 – The simulation accuracy (5.15) and computation speed of the process defined in Example 5.2.1. **Left:** The simulation accuracy for lag-h autocovariance operator with varying number of harmonic principal components used $N \in \{1,2,3,5,10,20,50,100,200,1000\}$ visualised as a function of the lag $h \in \{0,1,2,3,5,10,20,30,40,60,80,100\}$. The sample size parameters are set $T = 1000$ and $M = 1001$. **Right:** The simulation speed as a function of N with fixed $T = 1000$ and $M = 1001$. The dots represent the results of our experiments for $N \in \{\}$ while the solid line linearly interpolates those.

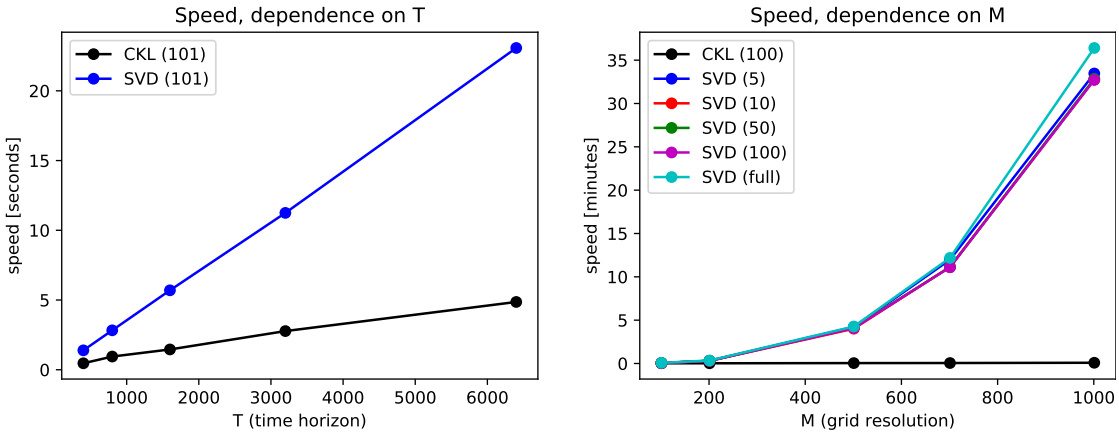

Figure 5.3 – The simulation speed of the process defined in Example 5.2.1. **Left:** The dependence on varying the time horizon $T \in \{400,800,1600,3200,6400\}$ while setting the spatial resolution $M = 101$. Both the simulation using the known Cramér-Karhunen-Loève expansion (CKL) and the method calculating this decomposition by the SVD algorithm use $N = 101$ eigenfunctions. **Right:** The dependence on varying $M \in \{101,201,501,701,1001\}$ while setting $T = 1000$. The simulation using the known Cramér-Karhunen-Loève expansion (CKL) uses 100 eigenfunctions while the numerical SVD (N) decomposition finds $N \in \{5,10,50,100\}$ leading eigenfunctions (the lines mostly overlap each other) or all of them $N = M$ for SVD (FULL). The CKL method has the running time below 0.1 minutes (6 seconds) even for $M = 1001$.

sents one of the more challenging cases one might wish to simulate from in an FDA context: functional data analyses typically feature smooth curves and differentiable corresponding operators, including spectral density operators, admitting a faster quicker eigenvalue requiring $N \ll 100$ eigenfunctions to capture a substantial amount of their variation.

Figure 5.3 presents the simulation speed results with varying sample size parameters: the time horizon T and the spatial resolution M. We compared the simulation using the known Cramér-Karhunen-Loève decomposition (5.13) with the method finding this decomposition numerically starting from the kernel (5.14). Such method finds the harmonic eigendecomposition using the (truncated) SVD algorithm applied to discretization of (5.14). Figure 5.3 shows that such routine can become very costly for higher spatial resolutions M, but if no other method is available, the method still constitutes an general approach how to simulate process with any dynamics structure defined through weak spectral density operators.

5.2.2 Long-range dependent FARFIMA(p,d,q) process

The next example is sourced from the work of Li et al. (2019) and Shang (2020) on long-range dependent functional time series. They consider the FARFIMA(1,0.2,0) process defined by (1.71) with the autoregressive operator $\mathscr{A}_1$ and the innovation covariance operator $\mathscr{S}$ defined as integral operators with respective kernels

$$A_1(x,y) = 0.34 \exp\left\{(x^2+y^2)/2\right\}, \qquad x,y \in [0,1], \tag{5.16}$$

$$S(x,y) = \min(x,y), \qquad x,y \in [0,1], \tag{5.17}$$

depicted in Figure 5.7. Recall that $S(x,y) = \min(x,y)$ is the covariance kernel of the standard Brownian motion on $[0,1]$. Because $d = 0.2 > 0$, the process exhibits long-range dependence (Li et al., 2019).

The constant 0.34 in (5.16) ensures that condition (1.62) is satisfied, and thus the process is stationary and admits a weak spectral density operator (Theorem 1.3.20) given by

$$\mathscr{F}_\omega^X = \frac{[2\sin(\omega/2)]^{-2d}}{2\pi} \left(\mathrm{I} - \mathscr{A}_1 e^{-\mathrm{i}\omega}\right)^{-1} \mathscr{S} \left(\mathrm{I} - \mathscr{A}_1^* e^{\mathrm{i}\omega}\right)^{-1}, \qquad \omega \in [-\pi,\pi]. \tag{5.18}$$

In fact, the operator $\mathscr{A}_1$ is of rank 1 and can be written as $\mathscr{A}_1 = 0.34 g \otimes g$ with $g(x) = \exp(x^2/2)$ for $x \in [0,1]$. This fact hugely simplifies the evaluation of (5.18) because the inversion of the autoregressive part can be written by the Sherman–Morrison formula as

$$\left(\mathrm{I} - \mathscr{A}_1 e^{-\mathrm{i}\omega}\right)^{-1} = \mathrm{I} + \frac{0.34 e^{-\mathrm{i}\omega}}{1 - 0.34 e^{-\mathrm{i}\omega} \|g\|^2_{L^2([0,1],\mathbb{R})}} g \otimes g, \qquad \omega \in [-\pi,\pi], \tag{5.19}$$

thus allowing for fast evaluation. Further computation gains, though less considerable, are made by using the Mercer decomposition of the Brownian motion covariance kernel (De-

heuvels and Martynov, 2003)

$$S(x,y) = \sum_{n=1}^{\infty} \frac{1}{[(n-0.5)\pi]^2} \sqrt{2}\sin\{(n-0.5)\pi x\}\sqrt{2}\sin\{(n-0.5)\pi y\}, \qquad x,y \in [0,1], \tag{5.20}$$

instead of numerical evaluation on a grid followed by an SVD decomposition.

In what follows, we consider the following implementations the spectral and time-domain, and hybrid simulation methods:

- SPECTRAL (BM): This method uses the known Mercer decomposition of the Brownian motion (BM) kernel (5.20) and simulates the process in the spectral domain using the method of Subsection 5.1.2 with the help of the Sherman-Morrison formula (5.19).
- HYBRID (BM): This method again uses the known Mercer decomposition of the Brownian motion (BM) kernel (5.20) and simulates the FARFIMA$(0,d,0)$ process and then applies the autoregressive recustion in the time-domain as explained in Subsection 5.1.3, thus constituting a HYBRID simulation method combining spectral and time-domain.
- SPECTRAL (SVD), HYBRID (SVD): These method correspond to SPECTRAL (BM) and HYBRID (BM) but the Mercer decomposition of the Brownian motion kernel is calculated numerically using the SVD algorithm.
- TEMPORAL: We use the original code by Li et al. (2019) available in the on-line supplement of their article and treat is as the benchmark for comparison with our spectral simulation methods. They simulate the realisations of the process by discretising the space domain $[0,1]$ and evaluating the integral operator $\mathscr{A}_1$ as a sum on this grid. Moreover, they perform the fractional integration (1.71) by analytically calculating the filter coefficients in the time-domain and thus expressing the process as FMA(∞), the functional moving average process of infinite order. Details on the FMA(∞) representation can be found in Li et al. (2019); Hosking (1981). The computational complexity of this method is $O(M^2T^2)$.

In order to assess the simulation accuracy we opt to simulate $I = 100$ independent realisations, and compare the mean empirical autocovariance operators (5.15) with the true autocovariance operator for varying $T \in \{400, 800, 1600, 3200, 6400\}$ and $M \in \{101, 201, 501, 1001\}$. We simulate the process with varying parameter T, the time horizon of the simulation, as well as varying spatial resolution M, based on a regular grid $\{x_m = (m-1)/(M-1)\}_{m=1}^{M} \subset [0,1]$. The simulation accuracy error, reported in Figure 5.9 (in Appendix 5.5), is negligible for all the simulation methods and (5.15) is dominated rather by the random component.

Figures 5.4 summarise how fast the different simulation methods were. We then see that the simulation by the TEMPORAL method used by Li et al. (2019) scales badly in T, while the other methods are linear in T, performing significantly better. On the other hand, the

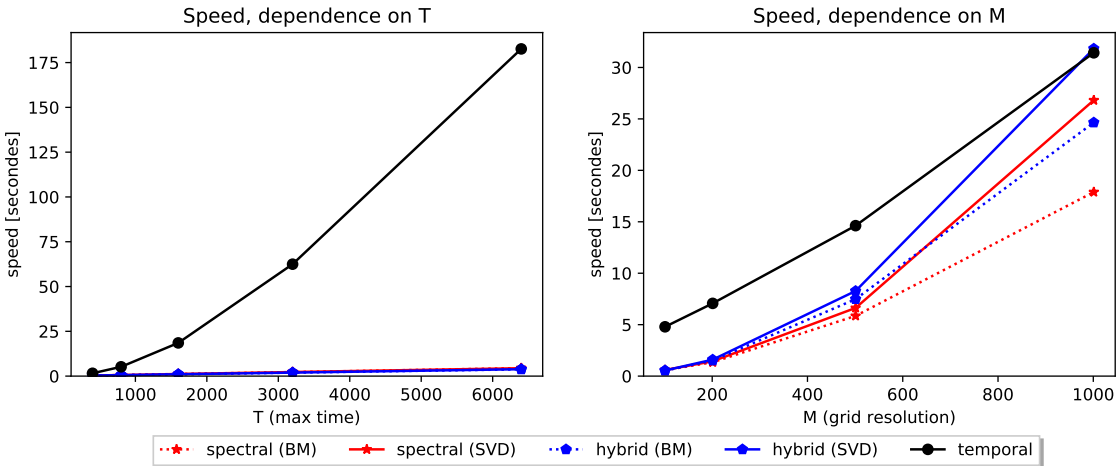

Figure 5.4 – The dependence of the **simulation speed** for the long-range dependent FARFIMA(1,0.2,0) process defined in Example 5.2.2 on the simulation parameters. **Left:** The simulation speed for varying time horizon $T \in \{400, 800, 1600, 3200, 6400\}$ with the spatial resolution is set $M = 101$. **Right:** The dependence of the simulation speed on the grid size $M \in \{101, 201, 501, 1001,\}$ with $T = 800$.

SPECTRAL (BM), HYBRID (BM), and TEMPORAL methods taking advantage of the innovation error covariance eigendecomposition have complexity dominated by $O(M^2 T)$ and scale similarly. The SPECTRAL (SVD) and HYBRID (SVD) methods require a further $O(M^3)$ operations for the SVD algorithm and this contribution becomes visible for $M \in \{501, 1001\}$.

5.2.3 FARMA(p, q) process with smooth parameters

In this example we consider the FARMA(4,3) process (1.61) with the autoregressive operators $\mathscr{A}_1, \ldots, \mathscr{A}_4$, the moving average operators $\mathscr{B}_1, \ldots, \mathscr{B}_3$, and the innovation covariance operator $\mathscr{S}$ defined as integral operators with kernels

$$\begin{aligned} A_1(x,y) &= 0.3\sin(x-y), & B_1(x,y) &= x+y, \\ A_2(x,y) &= 0.3\cos(x-y), & B_2(x,y) &= x, \\ A_3(x,y) &= 0.3\sin(2x), & B_3(x,y) &= y, \\ A_4(x,y) &= 0.3\cos(y), \end{aligned}$$

and

$$\begin{aligned} S(x,y) = {} & \sin(2\pi x)\sin(2\pi y) + 0.6\cos(2\pi x)\cos(2\pi y) + \\ & + 0.3\sin(4\pi x)\sin(4\pi y) + 0.1\cos(4\pi x)\cos(4\pi y) + 0.1\sin(6\pi x)\sin(6\pi y) + \\ & + 0.1\cos(6\pi x)\cos(6\pi y) + 0.05\sin(8\pi x)\sin(8\pi y) + 0.05\cos(8\pi x)\cos(8\pi y) + \\ & + 0.05\sin(10\pi x)\sin(10\pi y) + 0.05\cos(10\pi x)\cos(10\pi y), \qquad x, y \in [0,1]. \end{aligned} \tag{5.21}$$

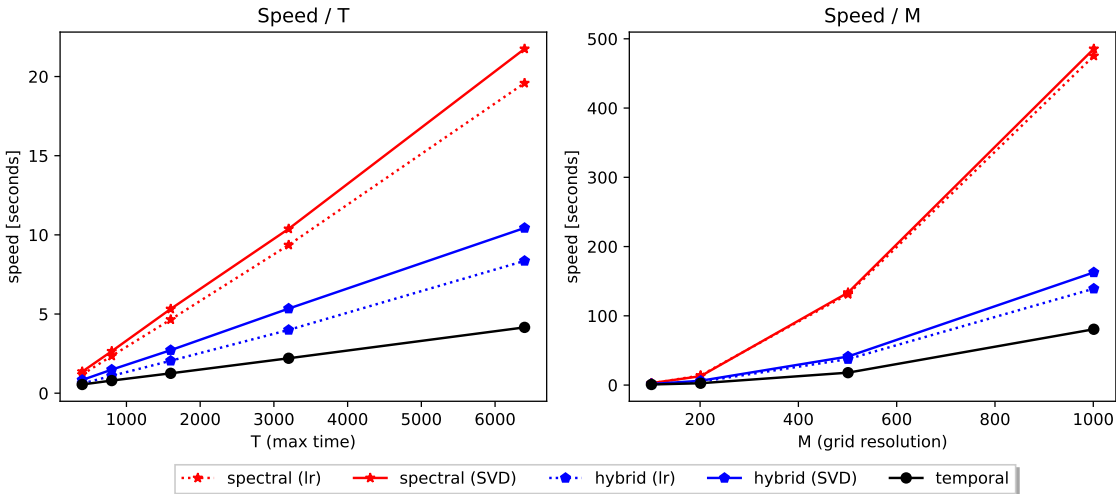

Figure 5.5 – The dependence of the **simulation speed** for the FARMA(4,3) process defined in Example 5.2.3 on the simulation parameters. **Left:** The simulation speed for varying time horizon $T \in \{400, 800, 1600, 3200, 6400\}$ with the spatial resolution is set $M = 101$. **Right:** The dependence of the simulation speed on the grid size $M \in \{101, 201, 501, 1001\}$ with $T = 800$.

These are depicted in Appendix 5.5, Figure 5.8. The constant 0.3 guarantees stationarity of the process, hence it admits the spectral density (1.63). Figure 5.10, included in Appendix 5.5, confirms that all the simulation methods approximate well the simulated process as the relative simulation error metric is affected more by the stochastic component. Figure 5.5 presents the simulation speed comparison between the spectral domain methods and the time-domain autoregressive recursion approach (TEMPORAL). The four considered spectral domain methods are:

- SPECTRAL (LR): This method uses the eigendecomposition (5.21) of the innovation noise covariance kernel. The simulation is conducted fully in the spectral domain as explained in Subsection 5.1.3.
- HYBRID (LR): This method uses the eigendecomposition (5.21) of the innovation noise covariance kernel, simulates the corresponding moving average process in the spectral domain and applies the autoregressive part in the time-domain as explained in Subsection 5.1.3.
- SPECTRAL (SVD), HYBRID (SVD): As above, but the eigendecomposition of $S(x, y)$ is calculated numerically by the SVD algorithm.

Even though the time complexity, which is dominated by the term $O(M^2 T)$, of the spectral domain simulation method matches the time complexity of the TEMPORAL domain approach with $O(M^2 T)$ complexity, the results presented in Figure 5.5 show that the simulation of the FARMA(p, q) process in the spectral domain, requiring solving matrix equation at each frequency, as well as the hybrid simulation are slower than the TEMPORAL approach.

The low-rank definition of (5.21) does not yield any computational speed-up compared to infinite rank covariance kernels (such as the Brownian motion kernel in Example 5.2.2). The purpose of such a definition is to allow for easy modification of the code if one wishes to specify the process via its harmonic eigenfunctions.

5.3 General recommendations for simulations

Our methodology provides a general purpose toolbox for simulating stationary (Gaussian) functional time series, leveraging their spectral representation. The high-level skeleton outlined at the beginning of Section 5.1 essentially reduces the problem to simulating a finite ensemble of independent random elements, and then applying the inverse fast Fourier transform. The generation of this i.i.d. ensemble depends on how one chooses to carry out discretisation and/or dimension reduction. We have demonstrated how knowledge of additional structure can significantly speed up the computations.

Some take-away messages and recommendations are as follows.

- **Simulation of functional time series specified through their spectral density operator.** To date, this problem had not been addressed, presumably because the assessment of the functional time series methods has traditionally been done based on simulation of functional *linear* processes. Key methods pertaining to regression and prediction, however, present performance tradeoffs that depend on the frequency domain properties, rather than the time domain properties of the time series (Hörmann et al., 2015a,b, 2018; Zhang, 2016; Tavakoli and Panaretos, 2016; Pham and Panaretos, 2018) or the results of this thesis (Chapters 2 and 3). One then wishes to simulate from a spectrally specified functional time series. More generally, our method can in principle be applied to any stationary model, linear or nonlinear, going well beyond the classical families of functional FARMA(p,q) or FARFIMA(p,d,q) processes, provided the process admits a weak spectral density operator.

  The method is fast and produces accurate results when the process is spectrally specified, courtesy of the Cramér-Karhunen-Loève expansion (Subsection 5.1.1) which is provably the optimal way to carry out dimension reduction. Excellent performance can also be expected when the dynamics of a functional time series are specified by means of white noise filtering (Subsection 5.1.2). For a general specification, the spectral domain simulation method of Subsection 5.1.1 still provides means how to simulate arbitrary functional time series. If the Cramér-Karhunen-Loève expansion is unknown, or a filtering representation is not available, the spectral density evaluation and the numeric eigendecomposition might require more time-consuming operations. Still, the approach constitutes the only general purpose recipe, where no previous method was available.

- **Simulation of FARFIMA(p,d,q) processes.** The advantages of the spectral approach

compared to time domain methods become quote considerable when dealing with processes that have an infinite order moving average representation, while having a simple formulation in the spectral domain. An important example being the FARFIMA(p,d,q) processes with $d>0$ (long memory process) or $d<0$ (anti-persistent) as the fractional integration is straightforward in the spectral domain while it produces an infinite order dependence in the time-domain. Example 5.2.2 showed how to efficiently and effortlessly simulate a long-range dependent FARFIMA process. Therefore we conclude that the simulation of FARFIMA(p,d,q) processes with $d\neq 0$ is more accessible and easy to implement in the spectral domain.

- **Simulation of FARMA(p,q) processes.** If one does specifically want to simulate a FARMA(p,q) processes, simulation in the time-domain is straightforward and fast. Still, our spectral domain simulation method matches the time complexity of the time domain methods in these cases. The constant hidden in "O", however, seems to be higher for the spectral domain methods, as Example 5.2.3 confirms. One advantage that the simulation in the spectral domain attains over the time-domain, though, is that we do not need to worry about the burn-in to reach the stationary distribution. We tentatively conclude that if a practitioner wishes to simulate a FARMA(p,q) process, then both the time-domain and the spectral simulation domain methods are equally applicable, though the time-domain simulation seems to be more straightforward to implement.

Overall the presented methods provide a useful toolbox of simulation methods in the spectral domain which are fast and accurate, and allow for simulation of standard as well as unusual or "custom defined" stationary time series defined through their weak spectral density operators. We hope that the accompanying code can be helpful for carrying out numerical experiments in future functional time series methodological research.

5.4 Code availability and R package specsimfts

To facilitate the implementation of the spectral domain simulation methods, we have created an R package specsimfts available on GitHub at https://github.com/tomasrubin/specsimfts. The package includes the implementations of all the presented methods as well as the examples considered in Section 5.2 as demos that are easy to use and modify.

5.5 Supplementary figures for examples 5.2.2 and 5.2.3

Figure 5.6 displays the trajectories of the FARFIMA(1, 0.2, 0) process simualted in Example 5.2.2 while Figures 5.7 and 5.8 depict the kernels of the integral operators used in Examples 5.2.2 and 5.2.3. Figures 5.9 and 5.10 illustrate the results on simulation accuracy discussed in Examples 5.2.2 and 5.2.3.

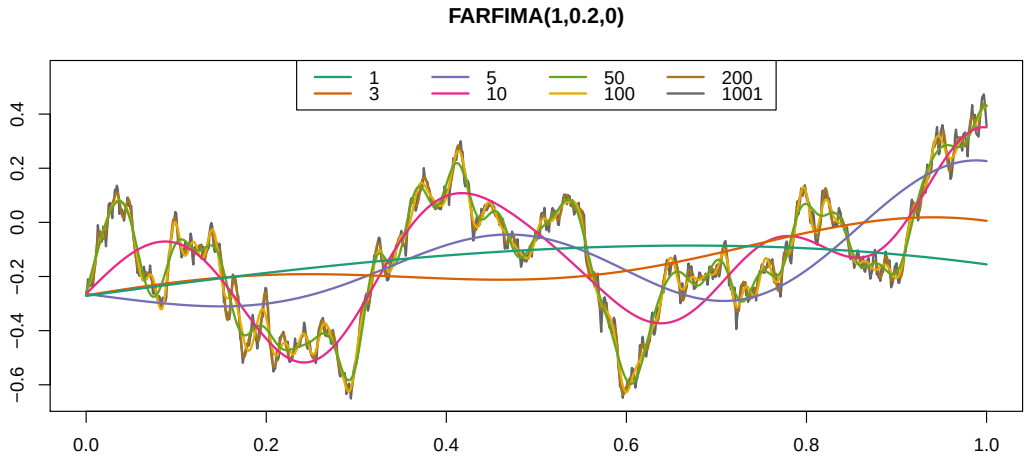

Figure 5.6 – Sample trajectories $X_1(\cdot)$ of the **long-range dependent FARFIMA(1,0.2,0)** process defined in Example 5.2.2 with varying number of N chosen in the truncation of (5.9). Simulated with $T = 100$ and the grid resolution $M = 1001$.

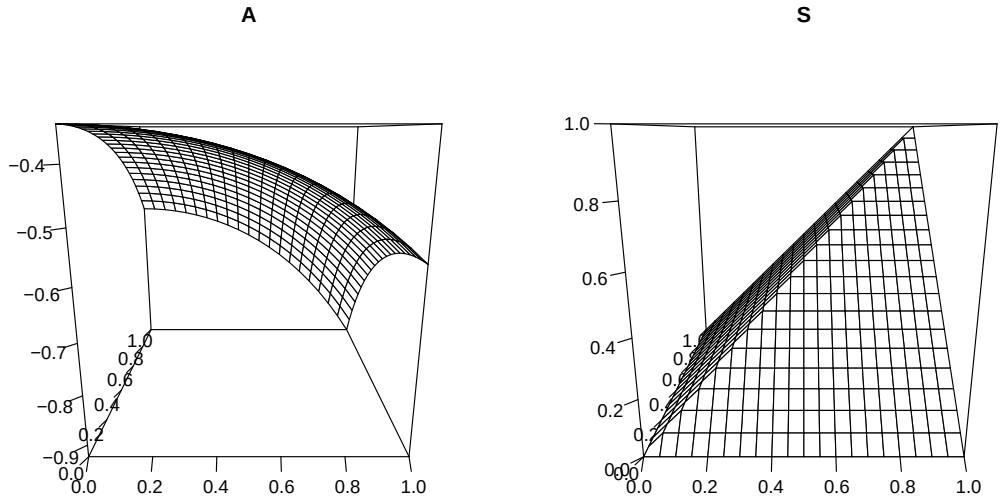

Figure 5.7 – The kernels of the autoregressive operator and the innovation covariance operator for the FARFIMA(1,0.2,0) process scrutinized in Example 5.2.2.

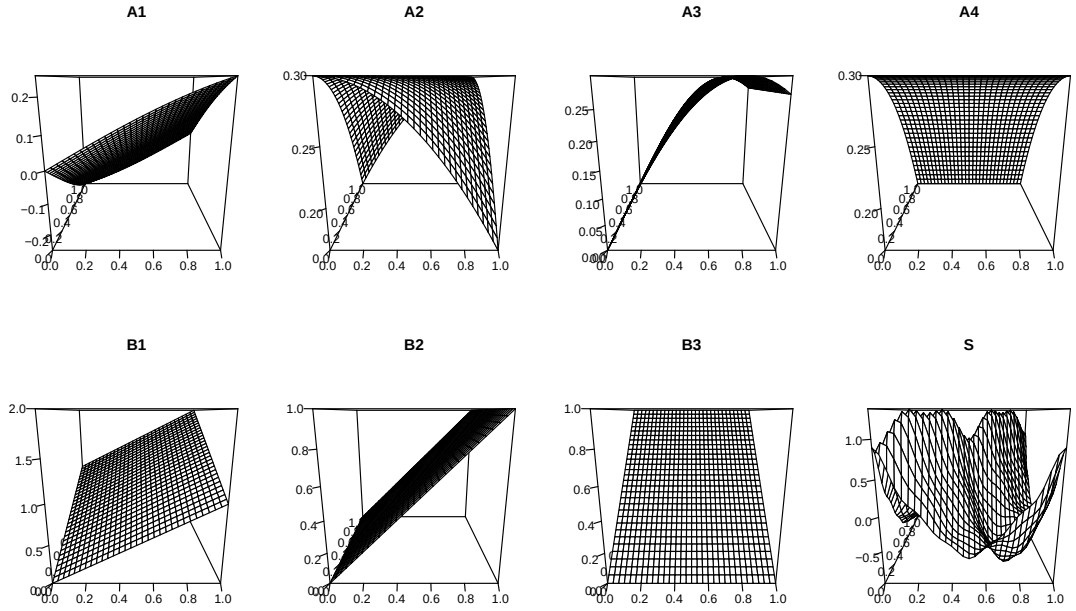

Figure 5.8 – The kernels of the autoregressive operators, moving average operators, and the innovation covariance operator for the FARMA(4,3) process scrutinized in Example 5.2.3.

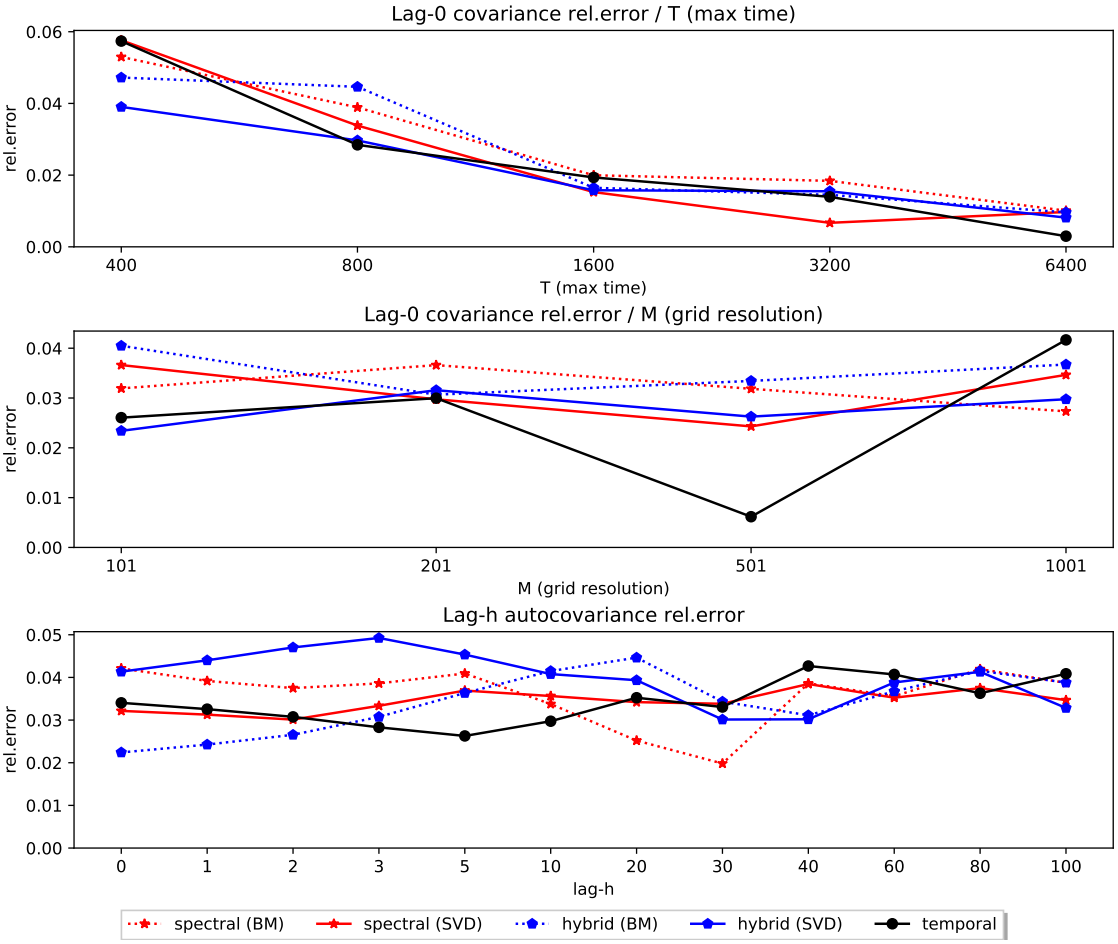

Figure 5.9 – The dependence of the **simulation accuracy** (relative error defined in (5.15)) for the long-range dependent FARFIMA(1,0.2,0) process defined in Example 5.2.2 on simulation parameters **Top:** The dependence of the lag-0 covariance operator simulation accuracy on time horizon $T \in \{400, 800, 1600, 3200, 6400\}$, while the spatial resolution is set $M = 101$. **Center:** the dependence of the lag-0 covariance operator simulation accuracy on the grid size $M \in \{101, 201, 501, 1001\}$, while the the time horizon is set $T = 800$. **Bottom:** the dependence of the lag-h covariance operator simulation accuracy on $h \in \{0, 1, 2, 3, 5, 10, 20, 30, 40, 60, 80, 100\}$, with $T = 800$ and $M = 101$.

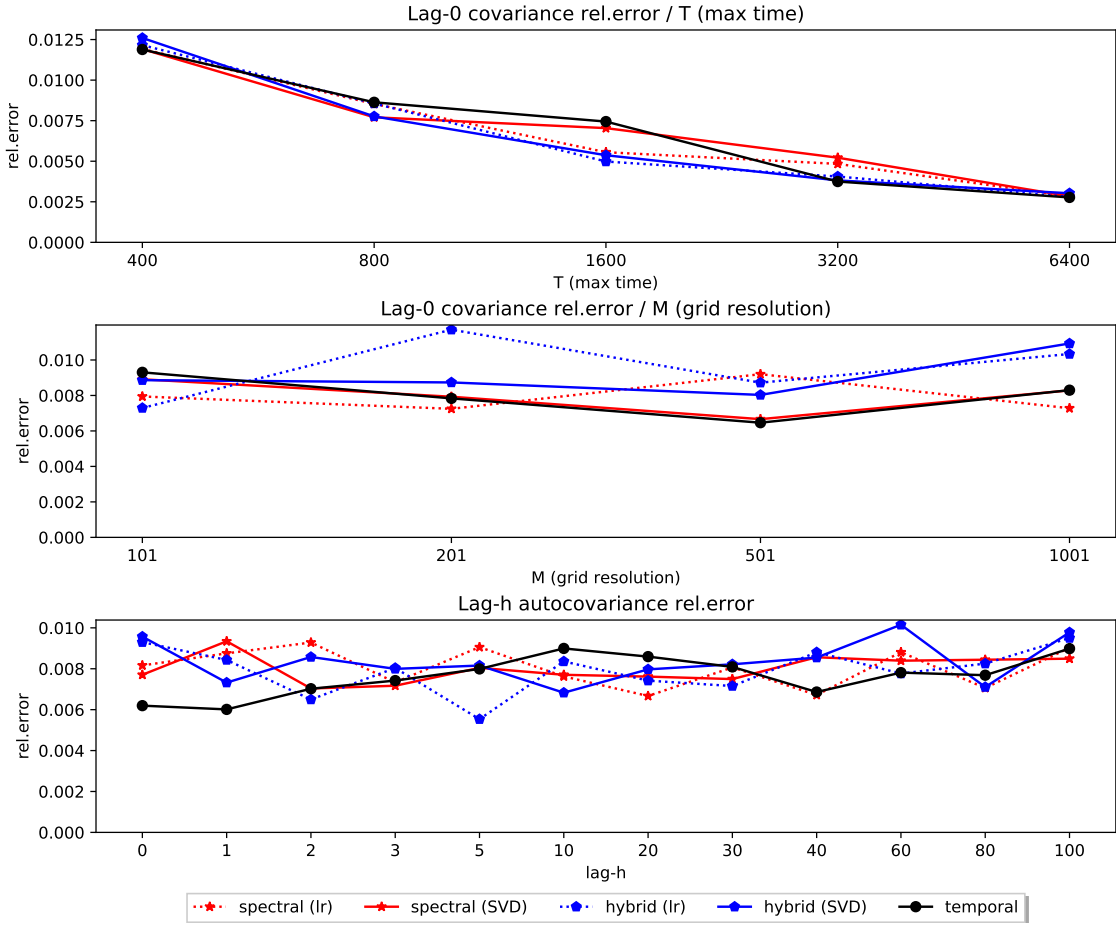

Figure 5.10 – The dependence of the **simulation accuracy** (relative error defined in (5.15)) for the FARMA(4,3) process defined in Example 5.2.3 on simulation parameters **Top:** The dependence of the lag-0 covariance operator simulation accuracy on time horizon $T \in \{400, 800, 1600, 3200, 6400\}$, while the spatial resolution is set $M = 101$. **Center:** the dependence of the lag-0 covariance operator simulation accuracy on the grid size $M \in \{101, 201, 501, 1001\}$, while the the time horizon is set $T = 800$. **Bottom:** the dependence of the lag-h covariance operator simulation accuracy on $h \in \{0, 1, 2, 3, 5, 10, 20, 30, 40, 60, 80, 100\}$, with $T = 800$ and $M = 101$.

6 Conclusions and future work

The presented thesis introduces a toolbox of methods for the estimation, prediction, and regression tasks for the sparsely observed functional time series. The key component turned out to be the smoother-based spectral density estimator allowing for consistent estimation of the entire second order dependence structure of the data, providing with helpful insights into the dependence or periodicity, and allowing prediction of the latent functional time series. The cross-spectral density estimator together with either the spectral truncation or the Tikhonov regularisation led to the estimation of the filter coefficients in the lagged regression dependence model among multiple functional time series. In this concluding chapter we outline a several future research questions that came from the results derived in this thesis.

The leading assumption of our functional time series set-up is the stationarity in the time variable t. The assessment of this assumption's validity is done either by visually checking the data evolution or by conducting rigorous statistical tests. The later approach has been well developed in the domain of fully observed functional time series in the works by Horváth et al. (2014); van Delft et al. (2017); Aue and van Delft (2020) who introduced various stationarity test. The key assumption for the construction of confidence bands within our functional recovery approach presented in Section 2.2 relies heavily on the Gaussianity assumption. This assumption can be tested for fully observed functional time series by the method of Górecki et al. (2018). It would be interesting to see how these test can be adjusted to sparsely observed functional time series data but this extension is beyond the scope of this thesis.

The principal design assumption overhanging the considered sparsely observed functional time series data comes down to the sampling protocol (2.1). The assumption of this scheme, although quintessential in sparsely observed literature (Yao et al., 2005a; Rice and Wu, 2001; James et al., 2000; Kowal et al., 2017), might turn out to be too strong and unrealistic in some applications. The example of the *Argo profiling float data* (Kuusela and Stein, 2018; Yarger et al., 2020), where the seawater temperature and salinity are measured by the buoy-like devices, comes to mind. The buoys provide sparsely scattered measurements in space and time but their locations are certainly not independent: the buoy at certain time cannot be far away from where it was yesterday. Moreover, their movement is subject ocean currents that may in fact

depend on the underlying temperature and salinity fields. It would be an interesting question to see how to incorporate the dependent sampling regime into the non-parametric estimation framework and how to non-parametrically predict the entire salinity and temperature fields across space and time, as the interpolation task is the primary objective of Kuusela and Stein (2018).

Moreover, the Argo float data mentioned in the previous paragraph hints to other possible generalisation directions. The Argo buoys do not float on the ocean surface, they dive in depth to probe the subsurface ocean properties, constructing a truly four dimensional data set (three dimension in space, and time). Even though the functional data analysis perspective has been pioneered for the Argo data (Yarger et al., 2020), their approach considers the ocean as a field of one dimensional strings (functions) in depth and interpolates using the spatial and spatio-temporal statistics techniques. On the other hand, one could wonder if considering the Argo data as a truly surface-valued (or higher dimensional) time series could be useful. Of course, a truly non-parametric approach might not turn to be feasible, and some assumptions such as separability of the spatial variable and the temporal one could help. Furthermore, the data come actually from the globe, i.e. a sphere and not a flat surface. Nevertheless, before venturing into the domain of sparsely observed *sphere or manifold valued functional data*, one should consider starting from the probabilistic and statistical analysis of fully observed design as spectral analysis of such random objects is still not well developed. The ideas in this paragraph are developed in more details in Section 2.6.

Another possible extension of the methods presented in this thesis is to abandon the fully non-parametric realm and consider some structure for the sparsely observed functional time series. In particular, the *functional autoregressive process* of order 1 defined in (1.60) is a popular modelling tool (Besse et al., 2000). The non-parametric estimation of the autoregressive operator, the key parameter for the modelling and forecasting, relies on the inversion of the functional version of the Yule-Walker equation (Bosq, 2000)

$$\hat{\mathscr{R}}_1^X = \mathscr{A}\hat{\mathscr{R}}_0^X$$

where $\hat{\mathscr{R}}_1^X$ and $\hat{\mathscr{R}}_0^X$ are the empiric lag-1 and lag-0 autocovariance operators, and $\mathscr{A}$ is the unknown autoregressive operator. In the sparse setting, one could directly employ the smoother-based estimators $\hat{\mathscr{R}}_1^X$ and $\hat{\mathscr{R}}_0^X$ constructed in Chapter 2 instead of the classical empiric autocovariance operators in the fully functional case and immediately obtain, after regularisation of the ill-posed inversion (Bosq, 2000), the estimator of $\mathscr{A}$. Having estimated the parameters of the sparsely observed autoregressive process, the latent functional data can be recovered by the Kalman filter and Rauch-Tung-Striebel smoother (Durbin and Koopman, 2012).

Finally we note that the novel derivation of the spectral density operators for the FARMA(p,q) and FARFIMA(p,d,q) processes (Subsection 1.3.8) could serve as a basis for model estimation in the spectral domain. The concept of *Whittle likelihood* is known in the univariate (Whittle, 1954) and the multivariate (Whittle, 1963) time series analysis to provide with computationally

quick approximation of the Gaussian likelihood function for time series parametric models. The objective for the the functional time series model with operator-valued parameters is clear: can one generalise Whittle likelihood to infinite dimensional function spaces? Such model fitting would require regularisation, most likely the one demanding smoothness and penalizing roughness of the operator-valued parameters as these objects are naturally infinite dimensional, and the penalized Whittle likelihood approach (Pawitan and O'sullivan, 1994) is likely to be a necessity.

Bibliography

Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2009). Bandwidth selection for functional time series prediction. *Statistics & Probability Letters*, 79(6):733–740.

Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using hilbert-valued autoregressive processes. *Journal of Multivariate Analysis*, 87(1):133–158.

Ash, R. B. and Gardner, M. F. (2014). *Topics in Stochastic Processes: Probability and Mathematical Statistics: A Series of Monographs and Textbooks*, volume 27. Academic press.

Aston, J. A. D. and Kirch, C. (2012a). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220.

Aston, J. A. D. and Kirch, C. (2012b). Evaluating stationarity via change-point alternatives with applications to fMRI data. *The Annals of Applied Statistics*, 6(4):1906–1948.

Aue, A., Hörmann, S., Horváth, L., and Hušková, M. (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica*, pages 1043–1073.

Aue, A., Horváth, L., and Pellatt, D. F. (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis*, 38(1):3–21.

Aue, A. and Klepsch, J. (2017). Estimating functional time series by moving average model fitting. *arXiv preprint arXiv:1701.00770*.

Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509):378–392.

Aue, A. and van Delft, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Annals of Statistics*, 48(5):2505–2547.

Barigozzi, M., Hallin, M., Soccorsi, S., and von Sachs, R. (2019). Time-varying general dynamic factor models and the measurement of financial connectedness. *Available at SSRN 3329445*.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, 37(1/2):1–16.

Beran, J. (1994). *Statistics for long-memory processes*, volume 61. CRC press.

Berlinet, A. and Thomas-Agnan, C. (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media.

Besse, P. C. and Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics*, 24(4):467–487.

Besse, P. C., Cardot, H., and Stephenson, D. B. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 27(4):673–687.

Blanke, D. and Bosq, D. (2007). Inference and prediction in large dimensions. In *Wiley Series in Probability and Statistics*. Dunod.

Bosq, D. (1983). Sur la prédiction non paramétrique de variables aléatoires et de mesures aléatoires. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 64(4):541–553.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In *Nonparametric functional estimation and related topics*, pages 509–529. Springer.

Bosq, D. (1996). Limit theorems for banach-valued autoregressive processes: applications to real continuous time processes. *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 3(5):537–555.

Bosq, D. (1999). Autoregressive Hilbertian processes. In *Annales de l'ISUP*, volume 43, pages 25–55.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Lecture Notes in Statistics. Springer London, Limited.

Bosq, D. (2012). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, volume 110. Springer Science & Business Media.

Bosq, D. and Blanke, D. (2008). *Inference and prediction in large dimensions*, volume 754. John Wiley & Sons.

Brillinger, D. (1983). *Time Series in the Frequency Domain*. Handbook of statistics. North-Holland.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, volume 36. SIAM.

Brockwell, P. and Davis, R. (2009). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer.

Cai, T. and Yuan, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. *University of Pennsylvania and Georgia institute of technology*.

Casas, I. and Gao, J. (2008). Econometric estimation in long-range dependent volatility models: Theory and practice. *Journal of Econometrics*, 147(1):72–83.

Chambers, M. (1995). The simulation of random vector time series with given spectrum. *Mathematical and Computer Modelling*, 22(2):1 – 6.

Chen, S. X., Lei, L., and Tu, Y. (2016). Functional coefficient moving average model with applications to forecasting chinese CPI. *Statistica Sinica*, pages 1649–1672.

Chiou, J.-M. and Li, P.-L. (2008). Correlation-based functional clustering via subspace projection. *Journal of the American Statistical Association*, 103(484):1684–1692.

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (2005). A theory of the term structure of interest rates. In *Theory of valuation*, pages 129–164. World Scientific.

Damon, J. and Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, 13(7):759–774.

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

Davies, R. B. and Harte, D. S. (1987). Tests for Hurst effect. *Biometrika*, 74(1):95–101.

de Boor, C. (1978). A practical guide to splines. volume 27. springer-verlag New York.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, pages 1735–1765.

Deheuvels, P. and Martynov, G. (2003). Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. In Hoffmann-Jørgensen, J., Wellner, J. A., and Marcus, M. B., editors, *High Dimensional Probability III*, pages 57–93, Basel. Birkhäuser Basel.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of econometrics*, 130(2):337–364.

Diebold, F. X., Li, C., and Yue, V. Z. (2008). Global yield curve dynamics and interactions: a dynamic nelson–siegel approach. *Journal of Econometrics*, 146(2):351–363.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, 131(1-2):309–338.

Doukhan, P. (2012). *Mixing: Properties and Examples*. Lecture Notes in Statistics. Springer New York.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57(1):405–443.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

Efromovich, S. (2008). *Nonparametric Curve Estimation: Methods, Theory, and Applications*. Springer Series in Statistics. Springer New York.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Fan, J. and Yao, Q. (2008). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Science & Business Media.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Science & Business Media.

Filipovic, D. (2009). *Term-Structure Models. A Graduate Course*. Springer.

Gao, Y., Shang, H. L., and Yang, Y. (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis*, 170:232–243.

Górecki, T., Hörmann, S., Horváth, L., and Kokoszka, P. (2018). Testing normality of functional time series. *Journal of time series analysis*, 39(4):471–487.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.

Green, P. and Silverman, B. (1993). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Grenander, U. (1950). Stochastic processes and statistical inference. *Arkiv för matematik*, 1(3):195–277.

Grenander, U. (1981). *Abstract Inference*. Probability and Statistics Series. John Wiley & Sons.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3):1493–1517.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Hays, S., Shen, H., and Huang, J. Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, 6(3):870–894.

Heath, D., Jarrow, R., and Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, pages 77–105.

Heckman, N. (2012). The theory and application of penalized methods or reproducing kernel hilbert spaces made easy. *Statistics Surveys*, 6:113–141.

Heil, C. (2018). *Metrics, Norms, Inner Products, and Operator Theory*. Springer.

Hewitt, E. and Ross, K. (2012). *Abstract Harmonic Analysis: Volume I Structure of Topological Groups Integration Theory Group Representations*. Grundlehren der mathematischen Wissenschaften. Springer New York.

Hörmann, S., Horváth, L., and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory*, 29(2):267–288.

Hörmann, S., Kidziński, Ł., and Hallin, M. (2015a). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):319–348.

Hörmann, S., Kidziński, Ł., and Kokoszka, P. (2015b). Estimation in functional lagged regression. *Journal of Time Series Analysis*, 36(4):541–561.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884.

Hörmann, S., Kokoszka, P., and Nisol, G. (2018). Testing for periodicity in functional time series. *The Annals of Statistics*, 46(6A):2960–2984.

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

Horváth, L., Kokoszka, P., and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179(1):66–82.

Horváth, L., Rice, G., and Whipple, S. (2016). Adaptive bandwidth selection in the long run covariance estimator of functional time series. *Computational Statistics & Data Analysis*, 100:676 – 693.

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

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Wiley.

Hull, J. and White, A. (1990). Pricing interest-rate-derivative securities. *The review of financial studies*, 3(4):573–592.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–799.

Israelsson, S. and Tammet, H. (2001). Variation of fair weather atmospheric electricity at Marsta observatory, Sweden, 1993–1998. *Journal of atmospheric and solar-terrestrial physics*, 63(16):1693–1703.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99(10):2508–2526.

Karhunen, K. (1946). Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Fennicae, AI*, 34.

Kidziński, Ł., Kokoszka, P., and Jouzdani, N. M. (2018). Principal components analysis of periodically correlated functional time series. *Journal of Time Series Analysis*, 39(4):502–522.

Klepsch, J. and Klüppelberg, C. (2017). An innovations algorithm for the prediction of functional linear processes. *Journal of Multivariate Analysis*, 155:252–271.

Klepsch, J., Klüppelberg, C., and Wei, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econometrics and Statistics*, 1:128–149.

Kokoszka, P. and Jouzdani, N. M. (2020). Frequency domain theory for functional time series: Variance decomposition and an invariance principle. *Bernoulli*, 26(3):2383–2399.

Kokoszka, P., Miao, H., and Zhang, X. (2015). Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics*, 13(2):456–477.

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

Kolmogoroff, A. (1941). Interpolation und extrapolation von stationaren zufalligen folgen. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 5(1):3–14.

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017). Functional autoregression for sparsely sampled data. *Journal of Business & Economic Statistics*, pages 1–13.

Kozicki, S. and Tinsley, P. A. (2001). Shifting endpoints in the term structure of interest rates. *Journal of monetary Economics*, 47(3):613–652.

Kuusela, M. and Stein, M. L. (2018). Locally stationary spatio-temporal interpolation of argo profiling float data. *Proceedings of the Royal Society A*, 474(2220):20180400.

Laurini, M. P. (2014). Dynamic functional data analysis with non-parametric state space models. *Journal of Applied Statistics*, 41(1):142–163.

Leng, X. and Müller, H.-G. (2006). Time ordering of gene coexpression. *Biostatistics*, 7(4):569–584.

Leucht, A., Paparoditis, E., and Sapatinas, T. (2018). Testing equality of spectral density operators for functional linear processes. *arXiv preprint arXiv:1804.03366*.

Li, D., Robinson, P. M., and Shang, H. L. (2019). Long-range dependent curve time series. *Journal of the American Statistical Association*, pages 1–30.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications*, 65(1):69–80.

Loève, M. (1946). Fonctions aléatoires à décomposition orthogonale exponentielle. *La Revue Scientifique*, 84:159–162.

Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

Mandelbrot, B. B. and van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437.

Mas, A. (2007). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, 98(6):1231–1261.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599.

Mitchell, R. and McPherson, D. (1981). Generating nonstationary random sequences. *IEEE Transactions on aerospace and electronic systems*, (4):553–560.

Mockus, J. (2012). *Bayesian Approach to Global Optimization: Theory and Applications*, volume 37. Springer Science & Business Media.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, 2:321–359.

Müller, H.-g. (2005). Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32(2):223–240.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, pages 473–489.

Palma, W. (2007). *Long-memory time series: theory and methods*, volume 662. John Wiley & Sons.

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

Panaretos, V. M. and Tavakoli, S. (2013a). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, 123(7):2779–2807.

Panaretos, V. M. and Tavakoli, S. (2013b). Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 41(2):568–603.

Paul, D. and Peng, J. (2011). Principal components analysis for sparsely observed correlated functional data using a kernel smoothing approach. *Electronic Journal of Statistics*, 5:1960–2003.

Pawitan, Y. and O'sullivan, F. (1994). Nonparametric spectral density estimation using penalized whittle likelihood. *Journal of the American Statistical Association*, 89(426):600–610.

Peligrad, M. (1992). Properties of uniform consistency of the kernel estimators of density and regression functions under dependence assumptions. *Stochastics and Stochastic Reports*, 40(3-4):147–168.

Peng, J. and Müller, H.-G. (2008). Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics*, 2(3):1056–1077.

Percival, D. B. (1993). Simulating Gaussian random processes with specified spectra. *Computing Science and Statistics*, pages 534–534.

Pham, T. and Panaretos, V. M. (2018). Methodology and convergence rates for functional time series regression. *Statistica Sinica*, 28(4):2521–2539.

Priestley, M. (1981a). *Spectral Analysis and Time Series*, volume 1 of *Probability and mathematical statistics : A series of monographs and textbooks*. Academic Press.

Priestley, M. (1981b). *Spectral Analysis and Time Series*, volume 2 of *Probability and mathematical statistics : A series of monographs and textbooks*. Academic Press.

Ramsay, J. and Silverman, B. (2013). *Functional Data Analysis*. Springer Series in Statistics. Springer New York.

Ramsay, J. O. and Silverman, B. W. (2007). *Applied functional data analysis: methods and case studies*. Springer.

Rice, G. and Shang, H. L. (2017). A plug-in bandwidth selection procedure for long-run covariance estimation with stationary functional time series. *Journal of Time Series Analysis*, 38(4):591–609.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 233–243.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259.

Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *The Annals of Probability*, 23(3):1188–1203.

Rosenblatt, M. (2012). *Stationary Sequences and Random Fields*. Birkhäuser Boston.

Rubín, T. (2020). Yield curve and macroeconomy interaction: evidence from the nonparametric functional lagged regression approach. *arXiv preprint arXiv:2007.02763*.

Rubín, T. and Panaretos, V. M. (2020a). Functional lagged regression with sparse noisy observations. *Journal of Time Series Analysis*, 41(6):858–882.

Rubín, T. and Panaretos, V. M. (2020b). Sparsely observed functional time series: Estimation and prediction. *Electronic Journal of Statistics*, 14(1):1137–1210.

Rubín, T. and Panaretos, V. M. (2020c). Spectral simulation of functional time series. *arXiv preprint arXiv:2007.08458*.

Rudebusch, G. and Svensson, L. E. (1999). Policy rules for inflation targeting. In *Monetary policy rules*, pages 203–262. University of Chicago Press.

Ruiz-Medina, M. D. (2019). Spectral analysis and parameter estimation of Gaussian functional time series. *arXiv preprint arXiv:1912.07086*.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.

Sen, R. and Klüppelberg, C. (2019). Time series of functional data with application to yield curves. *Applied Stochastic Models in Business and Industry*, 35(4):1028–1043.

Shang, H. L. (2020). A comparison of hurst exponent estimators in long-range dependent curve time series. *arXiv preprint arXiv:2003.08787*.

Shi, M., Weiss, R. E., and Taylor, J. M. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(2):151–163.

Shumway, R. H. and Stoffer, D. S. (2000). Time series analysis and its applications. *Studies In Informatics And Control*, 9(4):375–376.

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 93(444):1403–1418.

Stoica, P. and Moses, R. (1997). *Introduction to Spectral Analysis*. Prentice Hall.

Tammet, H. (2009). A joint dataset of fair-weather atmospheric electricity. *Atmospheric Research*, 91(2-4):194–200.

Tavakoli, S. (2014). *Fourier Analysis of Functional Time Series, with Applications to DNA Dynamics*. PhD thesis, EPFL.

Tavakoli, S., Nisol, G., and Hallin, M. (2020). High-dimensional functional factor models. *arXiv preprint arXiv:1905.10325*.

Tavakoli, S. and Panaretos, V. M. (2016). Detecting and localizing differences in functional time series dynamics: a case study in molecular biophysics. *Journal of the American Statistical Association*, 111(515):1020–1035.

Thompson, R. (1973). Generation of stochastic processes with given spectrum. *Util. Math*, 3:127–137.

van Delft, A. (2019). A note on quadratic forms of stationary functional time series under mild conditions. *arXiv e-prints*, page arXiv:1905.13186.

van Delft, A., Characiejus, V., and Dette, H. (2017). A nonparametric test for stationarity in functional time series. *arXiv preprint arXiv:1708.05248*.

van Delft, A. and Dette, H. (2018). A similarity measure for second order properties of non-stationary functional time series with applications to clustering and testing. *arXiv preprint arXiv:1810.08292*.

van Delft, A. and Eichler, M. (2018). Locally stationary functional time series. *Electronic Journal of Statistics*, 12(1):107–170.

van Delft, A. and Eichler, M. (2020). A note on Herglotz's theorem for time series on function spaces. *Stochastic processes and their applications*, 130(6):3687–3710.

Vašíček, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188.

Wand, M. and Jones, M. (1994). *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449.

Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50(1-2):129–134.

Wiener, N. (1950). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: with Engineering Applications*. Technology Press.

Wong, R. K. and Zhang, X. (2019). Nonparametric operator-regularized covariance function estimation for functional data. *Computational statistics & data analysis*, 131:131–144.

Xu, B., Zou, D., Chen, B. Y., Zhang, J. Y., and Xu, G. W. (2013). Periodic variations of atmospheric electric field on fair weather conditions at YBJ, Tibet. *Journal of Atmospheric and Solar-Terrestrial Physics*, 97:85–90.

Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A., and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics*, 59(3):676–685.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903.

Yarger, D., Stoev, S., and Hsing, T. (2020). A functional-data approach to the Argo data. *arXiv preprint arXiv:2006.05020*.

Zhang, X. (2016). White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics*, 194(1):76–95.

TOMAS RUBIN

@ tomas.rubin@gmail.com (+41) 78 690 42 82 live:tomas.rubin_1
Friedaustrasse 11, 8003 Zürich, Switzerland www.tomasrubin.com
in linkedin.com/in/tomas-rubin github.com/tomasrubin

STRENGTHS

- excellent quantitative background in machine learning and data science, including time series and stochastic analysis,
- strong coding skills in Matlab, Python, R, SQL,
- supervision of interdisciplinary projects experience,
- strong science and business communication skills.

CORE EXPERIENCE

Swiss Federal Institute of Technology, Lausanne (EPFL)

"Sparsely observed functional time series: theory and applications"

09/2016 – 09/2020 Lausanne, Switzerland

- My PhD thesis introduces a novel statistical method to analyse either time series data or spatial-temporal data. The novelty lies in removal of parametric assumptions of the existing methods, adaptability to a large number of missing values, and reduction of the prediction error up to 60 %.
- I was responsible for the design of the methodology, its implementation in Matlab, and its theoretical and simulation based justifications.
- I have successfully applied my methods for meteorological data and in yield curve modelling.

Advanced Risk Management (consulting company)

Business Analyst

09/2015 – 06/2016 Prague, Czech Republic

- Worked directly with clients in Central and Easter Europe.
- Developed a predictive model of electricity generation in client's wind farm used for trading of electricity contracts using R, SQL.
- Developed a predictive model of a client's customers' retention probability used for targeted marketing using R, SQL.
- Prepared materials for in-house seminars on banking regulation.
- Audited client's credit risk model, and worked on other projects.
- www.arm.cz/?lang=en_US

RozpisSluzeb.cz (tech start-up)

Software Developer

02/2013 – 03/2014 Prague, Czech Republic

- On-line service delivering easy and fast shift scheduling of personnel in hospitals and other medical institutions.
- I developed and maintained the algorithm for shifts scheduling using non-convex optimisation techniques, integer programming. Implemented in PHP and SQL.
- www.rozpissluzeb.cz (Czech only)

EDUCATION

Swiss Federal Institute of Technology, Lausanne (EPFL)

PhD in mathematics/statistics

09/2016 – 09/2020 Lausanne, Switzerland

Charles University in Prague

Master's degree in probability theory

09/2013 – 06/2016 Prague, Czech Republic

- Academic exchange at Humboldt University of Berlin, Germany

Charles University in Prague

Bachelor's degree in mathematics

09/2010 – 06/2013 Prague, Czech Republic

SKILL SET

Matlab | SQL | Python | numpy | R | Matplotlib | Mathematica | LaTeX

machine learning | time series analysis | data visualisation | stochastic calculus | stochastic differential equations models

interest rates & yield curve modelling | econometrics | financial forecasting | volatility modelling | stochastic volatility | options pricing | credit risk modelling

cloud computing | cluster computing

project management | problem solving | teaching | organisation of seminars | science & business communication | presentation skills | project supervision | intercultural & interdisciplinary skills

ADDITIONAL EXPERIENCE

Yield curve modelling

- I applied the novel non-parametric toolbox I developed in my core PhD research to model the dependence of the US Treasury yield curve on the US macroeconomy. My analysis confirms the previous findings under parametric assumptions and thus providing with additional arguments in favour of the standard econometric tools. Publication: [4].
- In a semester project I supervised at EPFL, we compared various forecasting methods for the yield curves in MINT economies (Mexico, Indonesia, Nigeria, Turkey). We discovered that the yield curve in these emerging markets is hardly predictable as opposed to the US Treasury yield curve. Publication: [5].

Implied volatility surface modelling

- Currently working on the application of the tools of my core PhD research to implied volatility surface modelling with objective of statistical description and forecasting. *Publication in preparation.*

Stochastic analysis & stochastic models

- Proficient in stochastic analysis and stochastic (partial) differential equations and stochastic models: interest rate models, market models, option pricing.
- Wrote Master's thesis on stochastic partial differential equations driven by singular fractional noise.

LIST OF PUBLICATIONS

[1] **Rubin** and Panaretos (2020). "Sparsely observed functional time series: estimation and prediction." *EJS*. 14 (1), 1137-1210.

[2] **Rubin** and Panaretos (2020). "Functional lagged regression with sparse noisy observations." *J. Time Ser. Anal.*, 41: 858-882.

[3] **Rubin** and Panaretos (2020). "Spectral simulation of functional time series." *arXiv:2007.08458*.
- software: `R` package `specsimfts` (available on my GitHub)

[4] **Rubin** (2020). "Yield curve and macroeconomy interaction: evidence from the non-parametric functional lagged regression approach." *arXiv:2007.02763*.

[5] Ayliffe and **Rubin** (2020). "A quantitative comparison of yield curve models in the MINT economies." *EPFL Infoscience*.

AWARDS

- 2016 — Gave a commencement speech on behalf of the students at the graduation ceremony at Charles University in Prague,
- 2016 — First place in student research competition *SVOC*,
- 2016 — Second best Master's thesis in probability & statistics at Charles University in Prague,
- 2016 — Second place in *Competition for the Best Student Paper in Theoretical Economics*,
- 2013 — First place in student research competition *SVOC*,
- 2013 — Best Bachelor's thesis in mathematics at Charles University in Prague.

LANGUAGES

English, Full professional proficiency, C2

German, Advanced proficiency, C1
- exchange semester at HU Berlin, Germany,
- *Leistungsnachweis* issued by HU Berlin,
- followed C1 level course at EPFL.

French, Limited proficiency, B1
- teaching experience in French at EPFL,
- individually tutored students in French.

Czech, Native tongue

TRAINING, TEACHING EXPERIENCE

Supervised 4 student semester projects in quantitative finance, statistics, and machine learning. In addition to [5]:
- Kalman filter and stochastic volatility,
- Stochastic differential equations and Vašíček model for interest rates modelling,
- Classification methods for life sciences.

Teaching assistant for:
- Autumn 2016, 2017, 2018, 2019, 2020 — *Statistical theory*, Master's level course, taught in English.
- Spring 2017, 2018, 2019 — *Statistique*, Bachelor's level course, taught in French.

Repeatedly given positive feedback from students' evaluation forms.

CONFERENCE TALKS

- 12/2020 (forthcoming) — CMStatistics 2020, London, UK, (invited talk),
- 06/2019 — European Meeting of Statisticians, Palermo, Italy,
- 04/2019 — CRoNoS & MDA 2019, Limassol, Cyprus.

EXTRACURRICULAR

- mountaineering, climbing, hiking, skiing,
- 2017 – 2020, committee member of student association *Club Montagne* at EPFL, responsible for organising rock climbing events,
- 2017 – 2018, committee member of *Society for Industrial and Applied Mathematics* (SIAM) chapter at EPFL.

PERSONAL INFO

29, single, Czech Republic (EU) citizenship, Swiss residency permit B, no military obligation, Swiss driving licence A,B.