The Bootstrap for the Functional Autoregressive Model FAR(1)

Vom Fachbereich Mathematik

der Technischen Universität Kaiserslautern

zur Verleihung des akademischen Grades

Doktor der Naturwissenschaften

(Doctor rerum naturalium, Dr. rer. nat.)

genehmigte

Dissertation

von

Euna Gesare Nyarige

Gutachten:

Prof. Dr. J. Franke, Technische Universität Kaiserslautern

Prof. Dr. J-P. Kreiß, Technische Universität Braunschweig

Datum der Disputation: 22. Juni 2016

D 386

To my Dad and Mom

Acknowledgement

I am deeply and sincerely grateful to my supervisor Prof. Jürgen Franke for his continuous guidance, patience, and support during my entire research period.

I would also like to express my gratitude to Danilo Pezo for his continuous support and suggestions concerning my research, his friendship, and for proof reading my manuscript. I am grateful too to Dr. Habil. Joseph Tadjuidje for his suggestions concerning my work.

I am grateful to Dr. Falk Triebsch, Graduate School of Mathematics for his support with regard to administartive matters concerning my Doctral studies and my schorlaship. Special thanks to the *Deutscher Akademischer Austausch Dienst* (DAAD) for their funding which enabled me to undertake my research and also to the Technical University of Kaiserslautern for the condusive environment and facilities which enabled me to carry out my work.

I thank the entire Statistics working group at TU Kaiserslautern for the amicable atmosphere I had during my studies. I am also deeply indebted to all my friends for all their support.

Finally I am deeply grateful to my family for their continuous support and encouragement through it all.

Abstract

Functional data analysis is a branch of statistics that deals with observations $X_1, \ldots, X_n$ which are curves. We are interested in particular in time series of dependent curves and, specifically, consider the functional autoregressive process of order one (FAR(1)), which is defined as $X_{n+1} = \Psi(X_n) + \epsilon_{n+1}$ with independent innovations ϵ_t. Estimates $\hat{\Psi}$ for the autoregressive operator Ψ have been investigated a lot during the last two decades, and their asymptotic properties are well understood. Particularly difficult and different from scalar- or vector-valued autoregressions are the weak convergence properties which also form the basis of the bootstrap theory.

Although the asymptotics for $\hat{\Psi}(X_n)$ are still tractable, they are only useful for large enough samples. In applications, however, frequently only small samples of data are available such that an alternative method for approximating the distribution of $\hat{\Psi}(X_n)$ is welcome. As a motivation, we discuss a real-data example where we investigate a changepoint detection problem for a stimulus response dataset obtained form the animal physiology group at the Technical University of Kaiserslautern.

To get an alternative for asymptotic approximations, we employ the naive or residual-based bootstrap procedure. In this thesis, we prove theoretically and show via simulations that the bootstrap provides asymptotically valid and practically useful approximations of the distributions of certain functions of the data. Such results may be used to calculate approximate confidence bands or critical bounds for tests.

Zusammenfassung

Die Funktionaldatenanalyse ist ein vergleichsweise neues Gebiet der Statistik, das sich mit Daten $X_1, \ldots, X_n$ befasst, die Funkktionen sind, zum Beispiel Kurven über einem Intervall der reellen Achse. Wir interessieren uns insbesondere für Zeitreihen abhängiger Kurven und betrachten ein einfaches Modell, den funktionalen autoregressiven Prozess der Ordnung 1 oder FAR(1)-Prozess, der folgendermaßen definiert ist: $X_{n+1} = \Psi(X_n) + \epsilon_{n+1}$ mit unabhängigen Innovationen ϵ_t. Schätzer $\hat{\Psi}$ des autoregressiven Operators Ψ sind in den vergangenen zwei Jahrzehnten ausgiebig untersucht worden, und ihre asymptotischen Eigenschaften sind gut bekannt. Besonders schwierig ist dabei die Verteilungskonvergenz, die die Grundlage für Bootstrapverfahren ist. Sie unterscheidet sich deutlich von den skalaren oder vektorwertigen Autoregressionen.

Obwohl die asymptotische Theorie für $\hat{\Psi}(X_n)$ noch handhabbar ist, nützt sie in der Praxis nur bei Stichproben, die groß genug sind. In Anwendungen ist das gerade bei Funktionaldaten oft nicht der Fall, so dass alternative Methoden zur Approximation der Verteilung von $\hat{\Psi}(X_n)$ wünschenswert sind. Als Motivation betrachten wir ein Changepoint- Detektionsproblem für einen funktionalen Realdatensatz aus dem Bereich der kognitiven Wissenschaften, den die Arbeitsgruppe Tierphysiologie der TU Kaiserslautern zur Verfügung gestellt hat.

Um eine Alternative zu asymptotischen Approximationen zu entwickeln, setzen wir das naive oder residuenbasierte Bootstrapverfahren ein. In dieser Arbeit beweisen wir theoretisch, dass das Bootstrap asymptotisch valide und praktisch nützliche Verteilungsapproximationen für gewisse Funktionen der Daten liefert. Die Theorie wird durch Simulationen unterstützt. Derartige Ergebnisse können dazu benutzt werden, um approximative Konfidenzbänder oder kritische Bereiche für Tests zu berechnen.

Contents

List of Tables

List of Figures

Chapter 1

Functional Data

1.1 Introduction

Functional Data Analysis (FDA) is a field of research concerned with observations which are curves. The data are of the form $X_n(t)$, where $t \in [a, b]$ and $X_n, n = 1, ..., N$ denotes the different observations (curves), see [28]. Considering a random function Z, a realisation $z(t)$ can be the height at time t, intensity at location t implying that the set T over which the functions are defined may be but is not necessarily limited to time. A random variable X is called a functional variable if it takes values in a functional space, for instance a Hilbert space or Banach space, and a functional dataset is the observation of N functional variables $X_1, X_2, ..., X_N$ which may be identically distributed as X.

Functional data are usually sampled in a discrete manner from continuous time processes and are represented as $X_n(t_i), n = 1, 2, ..., N, i = 1, 2, ..., m$, where n represents the index of observations or curves and i denotes the sampling points for each curve which may or may not be equidistant. An example is weather data which is recorded once each day for ten years. In this case taking each year as a curve, we have $N = 10$ and $m = 365$, it is easily noted that the sampling points m are equi-distant. On the other hand, if we consider the Berkeley Growth Study data, see Ramsay and Silverman, [49] then the sampling points are not equidistant. In the study the height of 10 girls was recorded in 31 instances, with the aim of studying their growth. According to how the measurements were recorded, it is clear that the data is not equidistant. Another example of irregularly spaced data is that of transactions at the point of sale (POS) which differ for different

days and times. It should be noted that most standard statistical techniques are based on equidistant data. However, aggregation of irregularly spaced data to conform to the standard models leads to a loss of information when the intervals of averaging are too big. On the other hand, for very small intervals, complexity may be introduced unnecessarily since each interval may not necessarily contain useful information. Once the discrete data has been sampled, to convert it to functional form we assume that there exists an underlying function X which generates it. Conversion to functional form may involve some smoothing which essentially means that we assume some relationship exists between two adjacent observations $X_n(t_i)$ and $X_n(t_j)$. Smoothing further implies that the underlying function possesses a given number of regularity conditions which are useful for describing the data or making inferences about it.

When the data is recorded during observation, some noise may be present so that what is recorded is of the form $Y = X + \epsilon$. In this case, as in standard regression, Y represents the response variable, X the explanatory variable (regressor) and ϵ is the noise also called disturbance which adds roughness to the observed data. This noise may be filtered during smoothing of data to functional form or later after the desired computations have been done on the data in functional form. It should be quickly noted that when dealing with functional linear models, three cases can be distinguished. In the first case we have the fully functional model where both the response and regressor are curves, the second and third cases consider a scalar response and functional regressors and vice versa. Details of these models can be found in Horváth and Kokoszka, [28]. In the event that there is no observational noise, then the conversion process from discrete to functional form is called interpolation. As already noted above, for different values of n, the argument values t_i may be different and the interval of observation may also vary. This non-uniformity of data collected therefore implies that a functional approach would be better suited for the representation and later analysis of the data since it is able to deal well with such cases as opposed to existing methods such as multivariate methods which require that observations be equidistant.

The resolution of the data is also of importance when dealing with functional data as it dictates, to some extent, what is achievable or not with the data. Resolution can

be described as the density of the argument values t_i relative to the curvature of the data, i.e it is a question of whether the number of observations recorded are enough to describe the data well including the minute details, see for instance Ramsay and Silverman, [52]. Higher curvature (data that includes a lot of detail) requires more data points for estimation so as to ensure satisfactory description and vice versa. An example of this is the data considered by Laukaitis and Račkauskas [38] which involves cash flow and transactions intensities at POS (Point of Sale) and ATMs (Automatic Teller Machines). In this case the curvature is high since there is a need to capture all the patterns for instance when the POS and ATMs are busiest among other things. They employ the limit of the measure precision in that they consider all the recorded transactions and therefore they are in a position to make reasonable inferences. In the event that all points are not considered, then there needs to be a way of ensuring that the collected data represents the high curvature observed.

Smoothing can be achieved by use of different methods. The most optimal method to be used depends on the underlying characteristics of the sampled data. Smoothing methods available include basis function methods, local weighting and roughness penalty approach. Basis function methods involve representing the observations as a weighted linear combination. Local weighting emphasizes on the local dependence of observations such that those closest to the argument values get the highest weights. Roughness penalty on the other hand defines smoothing at the level of the criterion being optimised. See [52].

Basis functions are most commonly employed for smoothing (estimating functions from discrete observations). A functional observation X_i smoothed by an appropriate basis is expressed as a weighted linear combination,

$$X_i(t) \approx \sum_{k=1}^{K} c_{ik}\phi_k(t)$$

where $\phi_k(t)$ are known functions which are orthogonal of each other and c_{ik} are weighting coefficients. The number K of basis functions used depends on the amount of smoothing that is required. In the case of interpolation, $K = m$, where m is the number of sampling points per curve. Choice of basis functions depends on the underlying characteristics of

the data. Some of the most commonly used basis functions include,

1. Fourier basis: This can be represented for a given function x as $\hat{x}(t) = c_0 + c_1 \sin\ \omega t + c_2 \cos\ \omega t + c_3 \sin\ 2\omega t + ...$ where the basis is defined as $\phi_0(t) = 1$, $\phi_{2k-1}(t) = \sin\ k\omega t$, and $\phi_{2k}(t) = \cos k\ \omega t$. This basis is periodic with period $2\pi/\omega$. Its coefficients can be efficiently computed by use of the Fast Fourier Transform when N is a power of 2. This basis is very useful for data that displays some periodicity, which may be distorted by noise, without local fluctuations and with uniform curvature. However, in the event of discontinuities in the function or its lower derivatives, the basis may perform poorly to some extent, see for instance [48] or [52] for more details.

2. Spline basis: This is employed for non-periodic data. A spline is composed of a combination of polynomials of a given order defined over specified sub-intervals of the function to be estimated with each polynomial being connnected to the next at a point called the breakpoint or knot. To define a spline one requires the order of the polynomial and the sequence of break points or knots. When the curvature of the function does not exhibit a lot of variation, the breakpoints can be set at equally spaced intervals. However flexibility can be enhanced by increasing the number of breakpoints for instance where the variation or curvature is high and vice versa. It should be noted though that increasing the value of K (the number of basis functions), does not necessarily lead to a better fit because the best fit depends on among others the knot sequence and the number of knots, see [**?**] for a detailed exposition of the same. The B-spline basis system is the most popular among spline systems because of its efficient computation. Like other splines, it consists of polynomial pieces that are joined at knots. A B-spline of degree z consists of $z + 1$ polynomial pieces, each of degree z. At the knots, there exists derivatives of order $z - 1$. As mentioned above, the problem of overfitting or underfitting plays a role here and to correct it penalties are employed. This involves penalising a feature of interest for instance the second derivative of the fitted curve. We apply this basis here.

3. Wavelets: These combine properties of the Fourier and Spline basis systems. This

implies that they are well adapted to coping with functions with discontinuities unlike in the case of the Fourier basis. In the case of Wavelet basis a mother wavelet is constructed which has the property of compact support and is built in such a way as to ensure orthogonality of the basis functions. As is the case in the Fourier transform there exists a Discrete Wavelet Transform which ensures speeds of $O(n)$ in computation as long as $n = 2^m$, which is even faster than the Fast Fourier Transform, see for instance [1] and [38].

4. Other basis systems include; the constant basis which can be used to view constant scalar observations as functions, polynomial basis which has good approximations in the center and is very poor at the boundaries, the step function basis among others.

It should be stressed that no particular basis can be uniformly applied but, depending on the underlying characteristics of the data a suitable basis can be chosen.

Another method of converting discrete observations to functional form is by use of least squares. In this case the observations are assumed to follow the model $y_j = x(t) + \epsilon_j$ where y_j are the discrete observations, $x(t_j)$ is a suitable basis function expansion (chosen according to the underlying characteristics of the data) and ϵ_j is the error or exogeneous variable. It should be noted that at times more than one basis function can be used to approximate the function x. In this case the first set of basis functions takes into consideration large features in the data while the other set considers local features and a combination of these two is then used. The method of least squares is applied as in the multivariate or scalar situation in that we choose the fit that minimizes the sum of squared errors, i.e

$$\sum_{j=1}^{m} \left(y_j - \sum_{k=1}^{K} c_{jk}\phi_k \right)^2 \tag{1.1}$$

As in the multivariate and scalar cases, fitting can be done using ordinary, weighted or localised least squares. Localised least squares for instance is preferred as compared to weighted or ordinary least squares as it considers local features near a given point of interest and will therefore provide a fit that better represents the data.

The roughness penalty approach is also useful for smoothing. This method smoothes by penalising a feature of interest in order to get the best fit, for instance the curvature can be the feature that we are interested in controlling. In this case the penalty is simply the squared norm of the second derivative, which is zero for a straight line. It is practical since it allows for control of features of interest.

Once the smoothing method and the proper basis have been selected, the number of basis functions K required to provide a good fit should be considered. K large implies that we have minimal bias but large variance and vice versa. This is the common problem of bias-variance or mean squared error trade-off and therefore K should be chosen such that there is a balance between the two. A thorough exposition of the above methods can be obtained from Ramsay and Silverman [49] for the functional case while several books on regression contain detailed information on the method of least squares.

1.2 Dependent functional data

Functional data can either be dependent or independent in nature. An example of independent functional data is the Berkeley growth study where the height of one girl is independent from the height of another. Dependent functional data arises for instance when an almost continuous time record is separated into consecutive intervals, for instance days, minutes or seconds among others. This implies and is mostly the case that some temporal dependence will exist. An example is Electro Encophalography (EEG) data recorded before during and after a given task which the subject is given. It is obvious that the recorded observations have some dependence as they represent three consecutive states of a subject. Data of such nature can be represented by the equation $X_n(t) = \xi_{nh+t}, 0 \leq t \leq h,\ n \in \mathbb{Z}, n = 1, ..., N$. Here a function X_n is an interval whose length is h. This dependence can be approximated with a known model for instance the Functional Autoregressive Model (FAR) also referred to as Autoregressive Hilbertian Model (ARH) and in cases where the exact model is unknown the effect of the dependence on the procedure under investigation should be checked.

Although consecutive observations may have some level of dependence, in some cases, as

the distance between the observations increases, then we have asymptotic independence of observations. Two observations A and B are asymptotically independent if they satisfy certain mixing conditions for instance strong mixing i.e

$$\alpha(k) = \sup\{|\mathcal{P}(A \cap B) - P(A)P(B)|: -\infty < t < +\infty, A \in X_{-\infty}^{t}, B \in X_{t+k}^{+\infty}\} \to 0$$

as $k \to \infty$. Asymptotic independence of observations implies that procedures and theorems for independent data for instance the Central Limit Theorem can be applied to these data. In some cases however, the approximation quality of asymptotic results is limited to rather large samples. An alternative is the bootstrap which allows us to obtain approximative distributions of an estimate of a parameter with frequently minimal assumptions if any. Once we have these approximative distributions we can then obtain confidence intervals, test hypotheses and so on. We quickly note that although the bootstrap is very useful in the event that we do not know the underlying distribution of the data, care must be taken when applying it so that the underlying characteristics of the original sample are replicated in the bootstrap pseudosamples. The bootstrap technique will be discussed in more detail in the last two chapters.

1.3 Selected applications

Functional data analysis has been employed in different fields of research to solve different problems. Castro *et. al.* [12] employ functional methods to predict sulfur dioxide levels emmited by a power plant. This is required in order to control air quality. The aim of their study was to predict the amount of sulfer dioxide that would be emmited, so that the staff in the plant could be in a position to take contol action in the event that the amount exceeded a certain set limit. In particular they use the functional autoregressive model and manage to show that its predictions are better and more reliable than other semiparametric and neural methods which have been employed before. They also apply bootstrap techniques in the functional context to deal with the confidence of their predictions a technique which we employ later on in this study.

Besse *et. al.* [5] compare different methods for forecasting functional data. They propose a hybrid method which is a local adaptation of the FAR(1) by introducing a weighted

kernel estimator of the covariance operator. The FAR forecasts were seen to outperform the scalor methods like SARIMA and ARIMA. This could be attributed to the fact that for real data and longer forecast horizons, the assumptions of linearity and stationarity are more likely to be violated thus making the scalar methods less powerful.

Antoniadis and Sapatinas [1] propose three linear wavelet methods to address the problem of continuous time prediction. They employ the wavelet basis for smoothing from discrete to continuous functional data, and to solve the ill-posed inverse problem they turn to regularization methods, in particular the Tikhonov-Phillips regularization method. They also prove the consistency of their estimators under certain assumptions, for instance, second order differentiability is not required for the smoothed curves inorder to obtain asymptotic rates, a condition that is required in the classical case. They apply their methods for prediction of the El Niño Southern Oscillation (ENSO) together with other methods that have been applied in literature for comparison. They compared their method with the smoothing spline interpolation estimator from Besse *et. al.* [5] and a SARIMA model with 12 months seasonality. One of their proposed methods was the same as that of Besse *et. al.* [5] save for the fact that they use wavelet basis as opposed to splines. In terms of perfomance the two methods were the same which implies that the proposed wavelet method should be chosen considering that its computation time is 20 times faster. In general the functional methods were better and the SARIMA model had the worst perfomance both graphically and in terms of prediction error.

Damon and Guillas [14] use functional methods and in particular the Autoregressive Hilbertian model (ARH) to estimate the concentration of ozone in the atmosphere. Ozone is an atmospheric pollutant and therefore the study aims at predicting the amount that will be in the air at a particular point in time in order to take control or corrective measures when a certain limit is exceeded. The main enhancement in their work is that apart from employing the ARH model, they also consider the influence of exogeneous variables temperature, wind speed and wind direction, which they incorporate into this model. They compare the predictions from this model with those from the ARH model without exogeneous variables, a functional kernel model and generalised additive models. In this case the ARH with exogeneous variables gave the best results followed by the

ARH without exogeneous variables and Generalised additive models. This implies that although the functional methods seem optimal because they consider the continuity of the process which inturn implies a bulk of information as compared to multivariate methods, their predictions in some instances can be improved by taking into consideration external influences as in this case.

Laukaitis and Račkauskas [38] employ functional methods in particular the Hilbert-valued autoregresssive process to study cash flow and transactions intensity in ATM (Automatic Teller Machine) and POS (Point of Sale) networks. Their work extends to econometrics the powerful functional data analysis methods which allow irregularly spaced data to be analysed, a feature not supported by standard econometric methods. They use wavelet-vaguelette estimators proposed in Antoniadis and Sapatinas, [1] and the projection type method introduced and extensively studied in Bosq [7]. Both methods were found to perform well (in terms of prediction). To be noted is that linear wavelet methods have a much faster implementation than B spline bases and are therefore suitable for large scale problems.

1.4 Motivation

We see that the functional data approach can be used in several different contexts with very good results. Among the reasons this approach is powerful is the fact that it considers the data as curves (functions), which implies that we assume there exists infinite dimensional data points which translates to a bulk of information. In this case no information is lost due to averaging as is the case when we consider scalar or vector methods. Another point in favour of this approach is the fact that it can be applied even in the case where the data is irregularly spaced. As the data are functions, we are not limited to the functions themselves only but we could also consider their derivatives for instance the acceleration of growth which is the second derivative, a feature not available for data that is not functional. The fact that the curvature of the data can also be captured in detail goes a long way in recommending the use of this approach as we are able to obtain better predictions if we have a good representation of the data. Also with increased

computation capacity this means that having the data as functions although bulky is still feasible in terms of computation time. The above are only some of the reasons for applying the functional approach and they all suggest that whenever possible it should be considered as its performance is superior to that of existing scalar and vector valued methods.

Although we see that when we consider functional data we have availability of a bulk of information per observation in the sample, the sample size in general may be small. This poses a problem as it introduces a bias in the estimation of parameter estimates. To overcome this problem we consider the naive bootstrap and seek to show that the asymptotics obtained from it are close to those obtained empirically. In our work we focus on the Functional Autoregressive model of order one (FAR(1)). We study existing results on weak convergence properties of the predicton obtained from the model. We seek to show theoretically and numerically that results obtained by employing the naive bootstrap method mimic the weak convergence behaviour of the empirical estimates obtained from the model, thereby solving the bias problem created by having a small sample size and obtaining accurate estimates.

Apart from parameter estimation, the bootstrap can be used to estimate the asymptotic distribution of the parameter estimates, which is useful when we want to investigate properties of these estimates, for instance their confidence bands, or when we would like to carry out tests of hypotheses as the critical values could be obtained using the bootstrap.

1.5 Structure of the thesis

This thesis is organised as follows:

In Chapter 2 we give some introductory notions and theory of the functional autoregressive model of order 1 (FAR(1)) and its properties. These are required to understand the rest of the chapters and are therefore given here as preliminaries.

In Chapter 3 we explore the weak convergence of the estimates obtained from the model.

We report some existing results and extend them slightly.

Chapter 4 looks at a practical change point problem involving functional data. We apply existing theoretical results to functional stimulus response data from the animal physiology group at the University of Kaiserslautern. Here, we use critical values of changepoint tests derived from asymptotics, but the example provides also a motivation for the bootstrap which promises better approximate critical values for small samples.

In Chapter 5 we have a first look at the bootstrap for FAR(1) models. We introduce the naive or residual-based bootstrap which is the focus of this thesis, and for comparison the block bootstrap. To illustrate the basic ideas and results we first study a simple toy model where we can circumvent some of the major technical difficulties of the general case.

In the central Chapter 6, we first prove that the basis for residual-based bootstrap methods holds, i.e. the distribution of the innovations may be approximated by the distribution of the bootstrap residuals which is the empirical distribution of the centered sample residuals. Then we prove that the bootstrap principle asymptotically holds for the crucial covariance operator estimate. Moreover, we also prove that the bootstrap may be applied to the estimate of the autoregressive operator under an artificial condition which is likely to hold. To show the latter conjecture, perturbation theory for linear operators like in the deep proof of weak convergence given by Mas [44] has to be extended to the world of the bootstrap data which would go beyond the scope of this thesis.

Finally, in Chapter 7 we present some simulations.

Chapter 2

The Functional Autoregressive Model (FAR(1)) and its Properties

This chapter concentrates on the theoretical properties of the FAR(1) (Functional Autoregressive Model of order one) also denoted as ARH (Autoregressive Hilbertian Model). The two terms will be used interchangeably thoughout this work. In the first section we define some terms that appear in this chapter. In Section 2 the model is introduced and its composition, together with conditions required for it to exist and be stationary are described. The next section deals with dimension reduction which is a prerequisite for the ill posed inverse problem that arises when estimating the models operator.

2.1 Definitions

This subsection contains definitions which are useful in understanding the material in the following chapters. These definitions are from Brockwell and Davis, [8] and Gohberg *et.al.*, [23] and Bosq, [7] which are useful references in case the reader would like a deeper understanding of the Hilbert space or to check for the proofs of the theorems stated below.

Definition 2.1.1. Strict stationarity: A real process $\xi = (\xi_t, t \in T)$ is said to be strictly stationary if and only if $P_{(\xi_{t_1+h},...,\xi_{t_k+h})} = P_{(\xi_{t_1},...,\xi_{t_k})}$, $k \geq 1$ and $t_1, ..., t_k, h \in T$. See [7].

Definition 2.1.2. Weak stationarity: A second order process is said to be weakly stationary if its mean $m(t) = m$ is constant and independent of t and its covariance is given by $c(s + h, t + h) = c(s, t)$, for $s, t, h \in T$. See [7].

Definition 2.1.3. *$L^p - m -$ approximability*: A sequence $\{X_n\} \in L^p_{\mathcal{H}}$ is called $L^p - m -$ *approximable* if each X_n admits the representation

$$X_n = f(\epsilon_n, \epsilon_{n-1,...})$$

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

$$X_n^{(m)} = f(\epsilon_n, \epsilon_{n-1}, ..., \epsilon_{n-m+1}, \epsilon_{n-m}^{'}, \epsilon_{n-m-1}^{'}, ...)$$

we have

$$\sum_{m=1}^{\infty} \nu_p \left(X_n - X_n^{(m)}\right) < \infty$$

In this case ν_p is a distance measure.

Definition 2.1.4. A Hilbert space $\mathcal{H}$ is an inner product space which is complete, i.e an inner product space in which every Cauchy sequence X_n converges in norm to some element $x \in \mathcal{H}$.

Definition 2.1.5. Closure: Given a set $S \subset \mathcal{H}$, the closure of S, written as $\bar{S}$ is the set of those vectors in $\mathcal{H}$ which are limits of sequences of vectors in S, i.e $x \in \bar{S}$ if $x_n \to x$ for some sequence $x_n \in S$. If $\bar{S} = S$, then S is a closed set.

Definition 2.1.6. Separability: The Hilbert space $\mathcal{H}$ is separable if $\mathcal{H} = \bar{sp}\ \{\nu_t, t \in T\}$, i.e. the closure of the span, with $\nu_t,\ t \in T$, a finite or countably infinite orthonormal set. We include Theorem 2.4.2 of [8] without proof to further explain the concept of separability.

Theorem 2.1.1. *If $\mathcal{H}$ is the separable Hilbert space defined above with $\nu_i, i = 1, 2, ...$ an orthonormal set, then*

1. *The set of all finite linear combinations of $\nu_1, \nu_2, ...$ is dense in $\mathcal{H}$, i.e for each $x \in \mathcal{H}$ and $\epsilon > 0$, there exists a positive integer k and constants $c_1, c_2, ..., c_k$ such that $\|x - \sum_{i=1}^k c_i \nu_i\| < \epsilon$*
2. *$x = \sum_{i=1}^{\infty} \langle x, \nu_i \rangle \nu_i$, for each $x \in \mathcal{H}$ i.e $\|x - \sum_{i=1}^n \langle x, \nu_i \rangle \nu_i\| \to 0$ as $n \to \infty$.*
3. *$\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, \nu_i \rangle|^2$ for each $x \in \mathcal{H}$.*
4. *Parsevals identity: $\langle x, y \rangle = \sum_{i=1}^{\infty} \langle x, \nu_i \rangle \langle \nu_i, y \rangle$ for each $x, y \in \mathcal{H}$.*
5. *$x = 0$ if and only if $\langle x, \nu_i \rangle = 0$, for all $i = 1, 2, ...$*

For later reference, we state without proof Theorems 2.3.1 and 2.4.1 of [8] in order to illustrate the projection from a given space (for instance the Hilbert space) to a closed subspace of this space.

Theorem 2.1.2. *If $\mathcal{M}$ is a closed subspace of the Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$, then*

1. *there is a unique element* $x \in \mathcal{M}$ *such that* $\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|$
2. $\hat{x} \in \mathcal{M}$ *and* $\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|$ *if and only if* $(x - \hat{x}) \in \mathcal{M}^{\perp}$, *i.e* $\hat{x}$ *is the (orthogonal) projection* $P_{\mathcal{M}}x$ *of* x *onto* $\mathcal{M}$.

Theorem 2.1.3. *If* $\nu_1, ..., \nu_k$ *is an orthonormal subset of the Hilbert space* $\mathcal{H}$ *and* $\mathcal{M} = \bar{sp}\{\nu_1, ..., \nu_k\}$, *then* $P_{\mathcal{M}}x = \sum_{i=1}^{k} \langle x, \nu_i \rangle \nu_i$ *for all* $x \in \mathcal{H}$.

2.2 Functional Autoregressive Model

We consider the Functional Autoregressive model of order one, FAR(1). This is an example of a model based on dependent observations. The model is defined in a separable Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ and generating norm $\|\cdot\|$. To have a specific situation, we consider the Hilbert space $L^2[0,1]$ but our considerations can be easily transfered to other L^2-spaces. The inner product for the L^2 space is defined as

$$\langle x, y \rangle = \int x(t)y(t)dt, \quad x, y \in \mathcal{H}$$

A sequence X_n, $n \in \mathbb{Z}$, of $\mathcal{H}$-random variables is called an autoregressive Hilbertian process of order 1 (ARH(1)) if it is strictly stationary and such that

$$X_{n+1} - \mu = \Psi(X_n - \mu) + \epsilon_{n+1}, \ n \in \mathbb{Z} \tag{2.1}$$

where X_n are functions in $L^2[0,1]$, $E\epsilon_n = 0$, Ψ is a bounded linear operator and $(\Omega, A) = (L^2[0,1], \mathcal{B})$, $\mathcal{B}$ Borel σ-algebra, as in Bosq [7] or Horváth et al., [28]. When the ϵ_n are such that $\mathbb{E}(\epsilon_n | X_{n-1}) = 0$, then X_n is called a Markov process. μ is assumed without loss of generality to be zero, although in applications this is generally not the case. The main attraction of this model lies in its ability to predict X_{n+1} using $\Psi(X_n)$, which is a very useful property in various fields where the occurence of an event depends to some extent on its past and an innovation or shock, ϵ. These fields include medicine, finance, meteorology among others.

The autoregressive operator $\Psi : L^2[0,1] \to L^2[0,1]$ is assumed to be a bounded linear operator. In the next section, we collect some definitions related to such operators on the Hilbert space $\mathcal{H}$.

2.2.1 Operators in the Hilbert Space ($L^2[0,1]$)

The norm generated by the inner product of $\mathcal{H}$ for the space $\mathcal{L}$ of bounded linear operators is given by

$$\|\Psi\|_{\mathcal{L}} = \sup_{\|x\|\leq 1} \|\Psi(x)\|$$

An operator Ψ is said to be compact if for orthonormal bases ν_j and f_j of $\mathcal{H}$ and a sequence λ_j of real numbers tending to zero the following relation holds

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle f_j, \ x \in \mathcal{H}$$

which can also be represented as

$$\Psi = \sum_{j=1}^{\infty} \lambda_j \nu_j \otimes f_j$$

Here, the Kronecker product of $x, y \in \mathcal{H}$ is a linear operator given by $x \otimes y(z) = \langle x, z \rangle y, \ z \in \mathcal{H}$. It is also possible to replace f_j with $-f_j$ therefore the λ_j can be assumed positive. For further reference, we state a few properties of the Kronecker product here, where $x, y \in \mathcal{H}$, A is a linear operator and A^T denotes the corresponding adjoint operator. They follow immediately from the definition.

$$\begin{aligned} x \otimes y &= (y \otimes x)^T \\ x \otimes (Ay) &= Ax \otimes y \\ (Ax) \otimes y &= x \otimes yA^T \end{aligned} \tag{2.2}$$

A compact operator is said to be Hilbert-Schmidt if $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$. The space of Hilbert-Schmidt operators $\mathcal{S}$ is separable and admits the inner product

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} = \sum_{1 \leq i,j \leq \infty} \langle \Psi_1(g_i), h_j \rangle \langle \Psi_2(g_i), h_j \rangle$$

where $(g_i), (h_j)$ are orthonormal bases in $\mathcal{H}$. The associated norm is

$$\|\Psi\|_{\mathcal{S}} = \left(\sum_j \lambda_j^2 \right)^{1/2} = \left(\sum_j |\Psi(g_j)|^2 \right)^{1/2}$$

An operator $\Psi \in \mathcal{L}$ is symmetric if $\langle \Psi(x), y\rangle = \langle x, \Psi(y)\rangle$, $x, y \in \mathcal{H}$ and positive if $\langle \Psi(x), x\rangle \geq 0$. A symmetric positive Hilbert-Schmidt operator admits the decomposition

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle \nu_j, \ x \in \mathcal{H} \tag{2.3}$$

thus $\Psi(\nu_j) = \lambda_j \nu_j$, $j \geq 1$, and λ_j, ν_j are the eigenvalues and eigenvectors of Ψ. A compact operator is called nuclear if $\sum_j |\lambda_j| < \infty$ with norm

$$\|\Psi\|_{\mathcal{N}} = \sum_{j=1}^{\infty} |\lambda_j|$$

The relationship between the different norms can be summarised as
$\|.\|_{\mathcal{N}} \geq \|.\|_{\mathcal{S}} \geq \|.\|_{\mathcal{L}}$. For a detailed exposition the references Bosq, [7], Horváth and Kokoszka, [28] and Gohberg *et. al.*, [23] are among the vast literature on the topic.

Assume, e.g., that the operator Ψ in our model 2.1 belongs to the class of integral operators in L^2 defined by

$$\Psi(x)(t) = \int \psi(t, s) x(s) ds, \quad x \in L^2$$

where $\psi(.,.)$ is a real kernel, see Horváth et al [28]. Then it is a Hilbert-Schmidt operator if and only if

$$\int \int \psi^2(t, s) dt ds < \infty$$

which is evident from Section 2.2.1. The kernel in this case represents the covariance of the functions which is normally symmetric and therefore the operator is Hilbert-Schmidt symmetric positive definite and admits the decomposition 2.3, where ν_j is an orthonormal basis. The model is nonparametric since Ψ is an infinite dimensional parameter.

2.3 Estimation of the Operator Ψ

The operator Ψ which is normally unknown, can be estimated by say $\hat{\Psi}_n$ from the sample. As in the scalar case, we require a condition on the operator Ψ in 2.1 in order to obtain a stationary solution to the process. To achieve this we consider one of the following assumptions

Assumption 2.1. *There exists an integer* $j_0 \geq 1$ *such that* $\|\Psi^{j_0}\|_{\mathcal{L}} < 1$

Assumption 2.2. *There exist* $a > 0$ *and* $0 < b < 1$ *such that* $\|\Psi^j\|_{\mathcal{L}} \leq ab^j$, $j \geq 0$

If the assumption 3.1 above holds, then we have a unique strictly stationary solution, see for instance Theorem 3.1 of Bosq, [7] or Theorem 13.1 of Horváth and Kokoszka, [28]. Moreover, Lemma 3.1 of [7] states that the two assumptions above are equivalent. The proof of the Lemma can be found in the above mentioned references for the interested reader.

It should be noted that an estimate of Ψ cannot be based on likelihood since the Lebesgue measure does not exist on non-locally compact spaces and the notion of density is not yet available for functional data. Having this in mind and from the classical method of moments, the operator Ψ, can be represented as $\Psi = C\Gamma^{-1}$, where,

$$\Gamma = \mathrm{E}(X_n \otimes X_n) \quad \text{and} \quad C = \mathrm{E}(X_n \otimes X_{n+1}) \tag{2.4}$$

are the covariance and cross covariance operators of the process and $\otimes$ is the Kronecker product. We denote the sample versions by $\hat{\Gamma}_n, \hat{C}_n$. To simplify an already involved notation, we follow Horváth and Kokoszka, [28] and assume that $\mathbb{E}X_t = 0$ is known throughout the thesis, i.e. we consider

$$\hat{\Gamma}_n = \frac{1}{n}\sum_{t=0}^{n-1} X_t \otimes X_t \quad \text{and} \quad \hat{C}_n = \frac{1}{n}\sum_{t=0}^{n-1} X_t \otimes X_{t+1} \tag{2.5}$$

In the general case, we would have to subtract the sample mean $\bar{X}_n$, e.g.

$$\hat{\Gamma}_n = \frac{1}{n}\sum_{t=0}^{n-1} \left(X_t - \bar{X}_n\right) \otimes \left(X_t - \bar{X}_n\right) = \frac{1}{n}\sum_{t=0}^{n-1} X_t \otimes X_t - \bar{X}_n \otimes \bar{X}_n$$

From Bosq, [7], however, we know that the first term on the right-hand side is of order $\mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)$, whereas the second term is of order $\mathcal{O}_p\left(\frac{1}{n}\right)$ such that our simplifying assumption has no influence on asymptotic results.

Γ is a symmetric, positive definite and compact operator which admits the spectral decomposition

$$\Gamma(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle \nu_j \tag{2.6}$$

However, Γ^{-1} is not a bounded operator. Indeed, Γ^{-1} admits the representation

$$\Gamma^{-1}(x) = \sum_{j=1}^{\infty} \lambda_j^{-1} \langle x, \nu_j \rangle \nu_j$$

which does not have a bounded inverse, considering $\|\Gamma^{-1}(\nu_j)\| = \lambda_j^{-1} \to \infty$ as $j \to \infty$.

A practical solution to this problem is to consider the first p most important Functional Principal Components, and since in practice the population principal components are unknown, these are replaced by the EFPC's (Empirical Functional Principal Components), $\hat{\nu_j}$. $\hat{\Gamma}_n^{-1}$(also referred to as $\hat{\Gamma}_n^{\dagger}$) is then given by

$$\hat{\Gamma}_n^{-1}(x) = \sum_{j=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{\nu}_j \rangle \hat{\nu}_j = \hat{\Gamma}_n^{\dagger}(x) \tag{2.7}$$

which is defined in the whole of L^2 and is bounded if $\hat{\lambda}_j > 0,\ j \leq p$ and the estimate thus obtained is unique. $\hat{\lambda}_j$ are the empirical eigenvalues.

From Equation 2.1, and as in the scalar autoregressive case multiplying through by X_n we have

$$X_n \otimes X_{n+1} = X_n \otimes (\Psi X_n) + X_n \otimes \epsilon_{n+1} = \Psi X_n \otimes X_n + X_n \otimes \epsilon_{n+1}$$

Considering the definitions of the covariance and cross covariance operators 2.4, and the fact that the ϵ term vanishes, if we consider expectation, we have

$$C = \Psi\Gamma \tag{2.8}$$

and $\Psi = C\Gamma^{-1}$. We therefore have a representation of the form

$$\hat{\Psi}_n(x) = \hat{C}_n\hat{\Gamma}_n^{\dagger}(x) = \frac{1}{n-1} \sum_{k=1}^{n-1} \left\langle X_k, \sum_{j=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{\nu}_j \rangle \hat{\nu}_j \right\rangle X_{k+1}$$

The estimate of Ψ is therefore given by

$$\hat{\Psi}_n(x) = \frac{1}{n-1} \sum_{k=1}^{n-1} \sum_{j=1}^{p} \sum_{i=1}^{p} \hat{\lambda}_j^{-1} \langle x, \hat{\nu}_j \rangle \langle X_k, \hat{\nu}_j \rangle \langle X_{k+1}, \hat{\nu}_i \rangle \hat{\nu}_i \tag{2.9}$$

where the last term is obtained by an additional smoothing step on X_{n+1} and $\hat{\nu}_j$. The empirical eigenfunctions, are known to converge to the population eigenfunctions asymptotically. See Bosq [7], or Kokoszka et. al. [28] for more details.

Having obtained the estimator of the population parameter Ψ, it is important that we gauge its optimality, i.e how good the estimator can estimate the true parameter. In the case of the FAR parameter Ψ, Didericksen *et. al.* [16] show that the above predictor is the most optimal in the Mean Squared Error and Mean Absolute error sense, since its prediction error is comparable to that of the infeasible predictor $\Psi(x)$ for appropriately chosen p.

2.4 Predictor Estimation and Dimension reduction

As noted above, the model exists in an infinite dimensional space but we need to work with it in a finite dimensional space. To do this we need to reduce the dimensionality of the data from infinite to finite while at the same time retaining as much information as possible. Two different methods have been proposed in the functional case for this purpose as discussed in the following subsections.

2.4.1 Functional Principal Component Analysis

Functional Principal Component Analysis is a dimension reduction technique that transforms the data from infinite dimensional space to finite dimensional space. This is done with the aim of representing the data in such a way that as much information as possible of the original data is retained. The data are represented using the first say p principal components which are normally arranged in descending order of magnitude. The principal components are chosen in such a way that they represent the highest variability in the data and that they are orthogonal to each other. Orthogonality is required to ensure that the variability represented by each component is unique. The number of functional principal components to be computed can be chosen using several criteria. These include cross-validation, cumulative percentage of variance (CPV), Akaikes Information Criterion (AIC), graphical methods (Scree plot), among others.

Given the covariance operator Γ defined above, see Equation 2.4 and additionally assuming that Γ is symmetric, positive definite and compact, it admits representation 2.6, where λ_j and ν_j are the eigenvalues and eigenfunctions respectively. Now these are pop-

ulation Functional Principal Components which are not normally available in practice, therefore they have to be estimated by their empirical counterparts $\hat{\lambda}_j$ and $\hat{\nu}_j$. It should be noted that the estimated eigenvalues and eigenfunctions are not necessarily unique. This is due to the fact that the estimate of the operator obtained with $\hat{\nu}_j$ is the same as that obtained with $-\hat{\nu}_j$. For the eigenvalues, their multiplicity may be greater than one. To solve these problems, the eigenvalues identifiability can be obtained by assuming $\lambda_1 > \lambda_2 > .... > \lambda_p > 0$. As for the eigenfunctions we introduce $\hat{c}_j = sign\left(\langle \hat{\nu}_j, \nu_j \rangle\right)$ such that $\hat{c}_j \hat{\nu}_j$ is close to ν_j, see Bosq, [7].

In order that the empirical counterparts are employed satisfactorily, their consistency should be proven. The empirical estimator of Γ, $\hat{\Gamma}_n$ is given by

$$\hat{\Gamma}_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} \langle X_i, x \rangle X_i \tag{2.10}$$

The expectation of $\hat{\Gamma}_n$ is therefore given by

$$\mathbb{E}\hat{\Gamma}_n = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}\langle X_i, x \rangle X_i \underset{n\to\infty}{\to} \Gamma$$

Thus $\hat{\Gamma}_n$ is an asymptotically unbiased estimator of Γ. This result can be found for instance in Bosq, [7]. Looking at the empirical eigenvalues and eigenfunctions, we have a collection of results that are used to prove consistency. The notation used is that of subsection 2.2.1. Lemma 4.2 of [7] proves the following result,

Lemma 1. *Suppose $A, B \in \mathcal{L}$ are two compact operators with singular value decomposition $A(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle f_j$ and $B(x) = \sum_{j=1}^{\infty} \gamma_j \langle x, u_j \rangle g_j$. Then, for $j \geq 1$,*

$$|\lambda_j - \gamma_j| \leq \|A - B\|_{\mathcal{L}}$$

On the other hand, from Lemma 2.3 of Horváth and Kokoszka, [28], (compare also Lemma 4.3 of Bosq, [7]) we have

Lemma 2. *$\lambda_1 > \lambda_2 > .... > \lambda_p > 0$, $\nu_j' = c_j \nu_j$ with $c_j = sign(\langle u_j, \nu_j \rangle)$, A symmetric, i.e. $\nu_j = f_j$, then*

$$\|u_j - \nu_j'\| \leq a_j \|A - B\|_{\mathcal{L}},\ 1 \leq j \leq p$$

where $a_j = 2\sqrt{2} \max[(\lambda_{j-1} - \lambda_j)^{-1}, (\lambda_j - \lambda_{j+1})^{-1}]$ if $j \geq 2$ and $a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}$

Having the above results in mind and from the fact that it has been proven that asymptotic results of operators can be transferred successfully to the eigenelements, see Mas and Menneteau, [43], it is then possible to use the empirical estimators of the population eigenelements when carrying out functional principal component analysis. This works when we have strict stationarity as this will ensure that the covariance operator is bounded. Further, Hörmann and Kokoszka [25] prove that for $\{X_n\}$ that is L^4-m-approximable, and for some constant $U_X < \infty$ then $n\mathbb{E}\|\hat{\Gamma}_n - \Gamma\|^2_{\mathcal{S}} \leq U_X$. From this result they further prove in their Theorem 3.2 that principal components are consistent in the case of weak dependence in particular that consistency holds in the case of L^4-m-approximability.

This method has been used extensively and successfully for dimension reduction, and we employ it in this study. It should be noted though that due to the fact that only the first few FPC's are incorporated, the estimated curves will be smoother than the original curves.

2.4.2 Predictive Factor Decomposition

This method was introduced by Kargin and Onatski [31], in which they proposed a reduced rank approximation of the autoregression operator that minimizes the prediction error, that is

$$\min_{\Psi_k \in R_k} \mathrm{E}\|X_{n+1} - \Psi_k X_n\|^2 = \min_{\Psi_k \in R_k} \mathrm{E}\|(\Psi - \Psi_k)\, X_n\|^2 \tag{2.11}$$

where R_k represents the set of all finite operators acting on $\mathcal{H}$. Unlike the case of Functional Principal Component analysis (FPCA) they focus on obtaining a reduced rank approximation, which is a linear combination of factors that contribute most to prediction as opposed to the variability in the data, arguing that directions based on FPCA are not necessarily justified by efficiency in the problem of prediction. Their estimator is given as

$$\hat{\Psi}_{\alpha,k} = \sum_{i=1}^{k} \left| \left\langle X_n, \hat{b}_{\alpha,i} \right\rangle \right| C\hat{b}_{\alpha,i} \tag{2.12}$$

where $\hat{b}_{\alpha,i} = \hat{\Gamma}_{\alpha}^{-1/2}\hat{x}_{\alpha,i}$, $\hat{x}_{\alpha,i}$ are the eigenfunctions of the operator Φ defined by the polar decomposition $\Psi\Gamma^{1/2} = U\Phi^{1/2}$ and U is a unitary operator. They argue that to get a consistent estimator of Ψ, $\hat{\Gamma}$ should be defined as $\hat{\Gamma}_{\alpha} = \hat{\Gamma} + \alpha I$, α a positive real number since the eigenfunctions of $\hat{\Phi}$ do not converge to those of Φ. In their work they showed that the Predictive Factor method perfomed better than the Functional Principal Component analysis method when it came to predicting Euro dollar future contracts. On the contrary, Didericksen *et. al.* [16] argue that this method never dominates the Functional Principal Component (FPC) method although with good tuning of the parameters α and k good results may be obtained.

2.4.3 Principal Component Analysis through Conditional Expectation (PACE)

This method extends functional principal component analysis (FPCA) to situations involving functional longitudinal data with few repetitions which are irregularly spaced but with sufficiently dense pooled time points. FPCA as described above is widely applied for dimension reduction in the functional case but in the case of very few data per subject it encounters difficulties. Few data in this case may imply one or two observations/measurements for some subjects. This method was suggested by Yao *et. al.*, [58] and it was found to give the best prediction under Gaussian assumptions and the best linear prediction for the non-Gaussian case of the function principal scores.

In their work, [58] consider sparse and irregular longitudinal data, where presence of measurement error is also accounted for. Because of the sparseness of the data, the classical FPCA would give biased approximations, due to sparseness of time points and presence of measurement error. The PACE method suggested on the other hand gives the functional principal score of a given subject as the conditional expectation of the score on the observation of that subject at a given time, i.e for an observation $Y_{ij} = X_i(T_{ij}) + \epsilon_j$ where Y_{ij} is the jth observation of the random function X with measurement error ϵ_j, and the PACE score is given by $\xi_{ik} = \mathbb{E}\left[\xi_{ik}|\tilde{Y}_i\right]$ where $\tilde{Y}_i$ denotes measurement on the ith subject. Their method was applied to simulated data and in the real case to longitudinal CD4 counts. In both cases it was found to give substantial gain over the classical FPCA

for sparse data. It was also noted from simulations that the method performed well with regular dense data. Asymptotics for the parameters involved were also established.

2.5 Weak Dependence

Owing to the fact that functional data arise from almost continuous data partitioned into intervals, for instance daily weather records, existance of some temporal dependence is inevitable. Hörmann and Kokoszka [25] quantify this notion of weak dependence using a moment based measure. They use the idea of m-dependence which is to approximate $X_n,\ n \in \mathbb{Z}$ by m-dependent processes $X_n^{(m)}, m \geq 1$. Now for a sequence X_n taking values in a given measurable space, the σ-algebras generated by the observations up to and from time l, that is $\mathcal{F}_l^- = \sigma\{..., X_{l-2}, X_{l-1}, X_l\}$ and $\mathcal{F}_l^+ = \sigma\{X_l, X_{l+1}, X_{l+2}, ...\}$ are said to be m-dependent if for any l the σ-algebras $\mathcal{F}_l^-$ and $\mathcal{F}_{l+m}^+$ are independent. Their theory is based on $L^p - m - approximability$.

Their definition is an extension of the scalar and vector $L^p - approximability$ with a few differences, for instance independence of the innovations in their case. It can also be noted that since m dependence implies the CLT, so does $L^p - m - approximability$. Although not directly comparable with classical mixing coefficients such as α-mixing or β-mixing, it can be shown that this form of dependence holds for time series models, for example the FAR model as illustrated in [25].

Chapter 3

Weak Convergence in the Functional Autoregressive Model (FAR(1))

In this chapter we explore the results on the weak convergence of the predictor in the FAR(1). In the first section some definitions required within the chapter are given. In section 2 we give existing results on the weak convergence of the autoregressive operator and predictor together with assumptions required. In the next section we extend the existing results on the prediction from 1-step ahead to n-steps ahead and in the last section we give conditions under which the random projector (from section 1) can be replaced by a non-random one. The weak convergence results in this chapter motivate the bootstrap chapter which follows as we are interested in carrying out predictions and obtaining confidence bands in practice even in the event of small sample size.

3.1 Definition

Definition 3.1.1. (Martingale differences): For a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset ... \subset \mathcal{F}_n \subset ....$ of sub-σ-algebras of $\mathcal{F}$ and a sequence $(X_i, i \geq 1)$ of integrable Banach space random variables, (X_i) is a martingale difference with respect to $(\mathcal{F}_i)$ if it is adapted to $(\mathcal{F}_i)$ and $\mathbb{E}\left(X_i \| \mathcal{F}_{i-1}\right) = 0$. See Bosq, [7].

3.2 Known Results

Here we introduce the results by Mas, [44] which we use extensively in this chapter and the next. Mas studies weak convergence of the functional autoregressive model which

elaborates on the rate of convergence and gives an exact asymptotic distribution. The fact that the exact asymptotic distribution is known ensures that confidence set bounds for $\Psi(X_n)$ can be obtained. The model is studied in a separable Hilbert space $\mathcal{H} = W^{m,2}$, with smoothness index $m = 0$ belonging to $\mathbb{N}$ ($W^{0,2} = \mathcal{L}^2$). The linear operator Ψ is supposed to be a compact operator, which is advantageous since the class of compact operators is a larger class than the normally used class of Hilbert-Schmidt operators.

Although the operator Ψ_n and the predictor $\Psi_n(X_n)$ have been shown to converge almost surely by Bosq [7] and hence, in probability, by Mas [41], from Theorem 3.2.1 below (Theorem 3.2 of Mas, [44]) it is shown that it is impossible for the operator to converge in distribution as it is. This convergence is impossible as from the proof of Theorem 3.2.1, in specific equation 3.5, it can be deduced that weak convergence i.e. $\Psi_n - \Psi$ depends only on the second term of 3.5 i.e $S_n\Gamma^\dagger$ where S_n is defined as,

$$S_n = \sum_{k=1}^{n} X_{k-1} \otimes \epsilon_k \tag{3.1}$$

and $\Gamma^\dagger$ is given by Equation 2.7. Due to the ill posed inverse problem involving Γ^{-1}, in the event that convergence is considered for $x \in \mathcal{H}$, if x belongs to the dense subset where Γ^{-1} is defined, then convergence will be achieved, otherwise the sum will diverge implying that there won't be a uniform limiting distribution. Because of the above mentioned reasons therefore, convergence in distribution cannot be achieved. A thorough exposition of the same fact can be found in Mas, [44] (proof of Theorem 3.2.2). It should also be noted that the problem of non-convergence weakly is not limited to the dependent case, as it is also observed in the i.i.d case, see Cardot *et. al.*[10].

Since convergence in distribution is very desirable in the sense that it allows for inference concerning point estimators and their asymptotic confidence bands, asymptotic mean and variances e.t.c, Theorem 3.2.1 (Theorem 3.1 of Mas, [44]) seeks to correct the above mentioned convergence problem. They prove a modified version of the result, i.e. the unknown operator Ψ is replaced by $\Psi\hat{\Pi}_{k_n}$ and with this random operator as the center convergence in distribution is proved. They employ a sample consisting of $n + 1$ observations where the first n are used to estimate the parameter after which the estimated parameter and X_{n+1} are used to predict X_{n+2}. For completeness, we include their

assumptions $A_0 - A_3$ and state the Theorems below.

Assumption 3.1. *ker* $\Gamma = 0$, $\mathbb{E}\|\epsilon\|^2 < \infty$, $\|\Psi\|_{\mathcal{L}} < 1$

Assumption 1 focuses on the existence of an inverse of the symmetric operator Γ, without which boundedness of Ψ cannot be achieved. The norm of $\Psi < 1$ is required so as to ensure that a unique, strictly stationary causal solution to the FAR(1) process exists i.e $X_n = \sum_{j=0}^{\infty} \Psi^j (\epsilon_{n-j})$, and $\mathbb{E}\|\epsilon\|^2 < \infty$ ensures finite variance. The proof for stationarity and causality can be found in Bosq, [7].

Assumption 3.2. $\|\Gamma^{-1/2}\Psi\|_{\mathcal{L}} < +\infty$

Assumption 2 can be considered as a smoothness constraint of the unknown operator Ψ. If Ψ is a diagonal operator in a complete orthonormal system such that $\Psi = diag\left[(\mu_i)_{i\geq 1}\right]$ with $\mu_i \geq \mu_{i+1}$, then the degree of smoothness of Ψ will be strictly determined by the rate of decrease to zero of $(|\mu_i|)_{i\geq 1}$. If we consider that Ψ is symmetric and has the same basis of eigenvectors as Γ, the assumption reduces to $\mu_i/\sqrt{\lambda_i} < \infty$, $i \in \mathbb{N}$. We are looking at the rate at which the eigenvalues of Ψ decay to zero. When this rate is high, then Ψ can be estimated well by a finite dimensional approximation based on the first p eigenvalues and vice versa.

The Karhunen-Loéve expansion of X is given by

$$X \underset{d}{=} \sum_{l=1}^{+\infty} \sqrt{\lambda_l}\xi_l \nu_l \tag{3.2}$$

where the ξ_l's are centered real random variables such that $E\xi_l\xi_{l'} = 1$ if $l = l'$ and 0 otherwise.

Assumption 3.3. $\sup_k \mathbb{E}\xi_k^4 < M$

This assumption is required since the definition of Ψ already contains second order moments and therefore in case of estimating for instance its variance the fourth moments will be required.

Assumption 3.4. $\lambda_j = \Lambda(j)$ *for all* $j \geq j_0$, *where the function* Λ *is convex.*

Convexity of the eigenvalues is required since in projecting Ψ to a finite dimensional space, it is normally highly desirable that only a few eigenvalues explain a high cumulative

percentage of variance. Additionally, as seen in Assumption three above, the smoothness of Ψ is linked to its eigenvalues, which convexity would guarantee to some extent. Further details and explanations on the assumptions can be obtained from Mas, [44].

Theorem 3.2.1. *When assumptions 3.1-3.4 hold and* $k_n = o\left(\frac{n^{1/4}}{\log\ n}\right)$,

$$\sqrt{\frac{n}{k_n}}(\hat{\Psi}_n(X_{n+1}) - \Psi\hat{\Pi}_{k_n}(X_{n+1})) \xrightarrow{w} \mathcal{G} \tag{3.3}$$

where $\mathcal{G}$ *is a H-valued Gaussian centered random variable with covariance operator* Γ_ϵ. $\hat{\Pi}_{k_n} = \sum_{j=1}^{k_n} \hat{\pi}_j$ *is the random projector onto the space spanned by the first* k_n *eigenvectors of the covariance operator* Γ_n, k_n *an increasing sequence tending to infinity.*

The Theorem states that the projection of the operator onto a finite dimensional subspace allows for the convergence in distribution of the operator. Now

$$\hat{\Psi}_n\left(X_{n+1}\right) - \Psi\hat{\Pi}_{k_n}\left(X_{n+1}\right) = \hat{\Psi}_n\left(X_{n+1}\right) - \Psi\left(X_{n+1}\right) + B_n \tag{3.4}$$

where $B_n = \Psi\left(I - \hat{\Pi}_{k_n}\right)(X_{n+1})$ can be considered as a bias term which will vanish when the sample size increases. For the sake of completeness, we now state Theorem 3.2 of Mas, [44].

Theorem 3.2.2. *It is impossible for* $\hat{\Psi}_n - \Psi$ *to converge in distribution for the norm topology on* $\mathcal{K}$, *where* $\mathcal{K}$ *is the space of compact operators.*

3.3 n-Step ahead Prediction of X

We show that Theorem 3.2.1 holds for X_{n+2} and include some remarks for the case of any given X. From the proof of Theorem 3.1 of Mas, [44] and from Equation 2.1 we have $\Gamma = \Psi\Gamma\Psi^* + \Gamma_\epsilon$. Let $S_n = \sum_{k=1}^{n} X_{k-1} \otimes \epsilon_k$. From Lemma 10, we have

$$\hat{\Psi}_n - \Psi\hat{\Pi}_{k_n} = \frac{1}{n}S_n(\hat{\Gamma}_n^\dagger - \Gamma^\dagger) + \frac{1}{n}S_n\Gamma^\dagger \tag{3.5}$$

It is shown in Mas, [44] for X_{n+1} that $\frac{1}{n}S_n(\hat{\Gamma}_n^\dagger - \Gamma^\dagger)$ is a vanishing term and that the asymptotic distribution depends only on $\frac{1}{n}S_n\Gamma^\dagger$. Note that Mas includes the factor $\frac{1}{n}$ into S_n, but in the explicit definition, there, the factor is missing due to a misprint. We show that the same holds for X_{n+2} and in extension for n steps ahead.

Using a combination of pertubation theory and functional calculus for bounded operators, Mas [44] defines $\Gamma^\dagger$ and the projector onto the space spanned by the first k_n eigen values of Γ as

$$\Gamma^\dagger = \int_{\mathcal{C}_n} z^{-1} (zI - \Gamma)^{-1} dz = \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} z^{-1} (zI - \Gamma)^{-1} dz$$

and

$$\Pi_{k_n} = \frac{1}{2\pi i} \int_{\mathcal{C}_n} (zI - \Gamma)^{-1} dz$$

where $\mathcal{C}_n = \bigcup_{i=1}^{k_n} \mathcal{B}_i$, and $\mathcal{B}_i$ is the oriented circle on the complex plane with center λ_i and radius $\delta_i/3$ and $\delta_i = \min(|\lambda_i - \lambda_{i-1}|, |\lambda_i - \lambda_{i+1}|)$ Their random counterparts are then given by

$$\hat{\Gamma}_n^\dagger = \int_{\hat{\mathcal{C}}_n} z^{-1} \left(zI - \hat{\Gamma}_n\right)^{-1} dz = \sum_{j=1}^{k_n} \int_{\hat{\mathcal{B}}_j} z^{-1} \left(zI - \hat{\Gamma}_n\right)^{-1} dz$$

and

$$\hat{\Pi}_{k_n} = \frac{1}{2\pi i} \int_{\hat{\mathcal{C}}_n} \left(zI - \hat{\Gamma}_n\right)^{-1} dz$$

respectively. From these we have

$$\begin{aligned} & S_n \left(\hat{\Gamma}_n^\dagger - \Gamma^\dagger\right)(X_{n+2}) \\ &= \int_{\hat{\mathcal{C}}_n} z^{-1} S_n \left(zI - \hat{\Gamma}_n\right)^{-1} (X_{n+2}) dz - \int_{\mathcal{C}_n} z^{-1} S_n (zI - \Gamma)^{-1} (X_{n+2}) dz \end{aligned} \tag{3.6}$$

For the residual term (3.6); Lemma 5.6 and Proposition 5.1 of [44] are unaffected by the change from X_{n+1} to X_{n+2}. This is because in the case of Lemma 5.6 what is proved is that the random contour $\hat{\mathcal{C}}_n$ can be replaced by the non random one $\mathcal{C}_n$. In this case an event A_n is introduced which allows the consideration of ordered eigenvalues of $\hat{\Gamma}_n$ which are very close to those of Γ. This effectively splits the set into two; sample eigenvalues (the first k_n) very close to their true counterparts such that they can be considered to be in the non random contour $\mathcal{C}_n$ (A_n) and those that do not fulfill the requirements of the event (A_n^c). It is then shown that the probability of the complement set of the event goes to zero in probability when the conditions of the lemma hold thus the random contour can be replaced by the non random one. In the proof of the lemma, we are majorly concerned

with the covariance operator which is constructed from the observations without taking into account the observation to be predicted. This therefore implies that whether we consider X_{n+1} or X_{n+k} where $k \geq 2$ the outcome of this lemma will not be affected.

When we consider Proposition 5.1 on the other hand this involves splitting X_{n+2} into ϵ_{n+2} and $\Psi(X_{n+1})$ and bounding the respective terms separately. In this case too considering $X_{n+k},\ k \geq 2$ does not alter the results since among the properties considered in order to bound both terms is identical distribution, which we have as a result of the assumption of strict stationarity.

Thus in our case we have, under the assumptions of Theorem 3.2.1

$$\sqrt{\frac{n}{k_n}} S_n(\hat{\Gamma}_n^\dagger - \Gamma^\dagger)(X_{n+2}) \xrightarrow{p} 0$$

The weakly convergent term, $S_n\Gamma^\dagger(X_{n+2})$, should fully determine the asymptotics of the predictor. Now

$$\begin{aligned} S_n\Gamma^\dagger(X_{n+2}) &= \sum_{k=1}^{n} \left\langle X_{k-1}, \Gamma^\dagger(X_{n+2}) \right\rangle \epsilon_k \\ &= \sum_{k=1}^{n} Z_{k,n} \end{aligned}$$

Note that $X_{n+2} = \epsilon_{n+2} + \Psi(\epsilon_{n+1}) + ... + \Psi^{n+2-k}(\epsilon_k) + \Psi^{n+2-(k-1)}(X_{k-1})$. Decomposing $Z_{k,n}$ into three terms we have,

$$\begin{aligned} Z_{k,n}^+ &= \left\langle \Gamma^\dagger X_{k-1}, \epsilon_{n+2} + \Psi(\epsilon_{n+1}) + ... + \Psi^{n+2-(k+1)}(\epsilon_{k+1}) \right\rangle \epsilon_k \\ Z_{k,n}^0 &= \left\langle \Gamma^\dagger X_{k-1}, \Psi^{n+2-k}(\epsilon_k) \right\rangle \epsilon_k \\ Z_{k,n}^- &= \left\langle \Gamma^\dagger X_{k-1}, \Psi^{n+2-(k-1)}(X_{k-1}) \right\rangle \epsilon_k \end{aligned}$$

The random sequences $Z_{k,n}^+$ and $Z_{k,n}^-$ remain Hilbert-valued martingale difference arrays according to lemma 5.7 of [44] even when we replace X_{n+1} by X_{n+2}, noting that in this case $X_{k,n}^\# = \epsilon_{n+2} + \Psi(\epsilon_{n+1}) + ... + \Psi^{n+2-(k+1)}(\epsilon_{k+1})$. Indeed, $\mathbb{E}(Z_{k,n}^+|\mathcal{F}_{k-1}) = \mathbb{E}\left(\left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle \epsilon_k | \mathcal{F}_{k-1}\right)$ and since ϵ_k is independent from $X_{k,n}^\#$, we have;
$\mathbb{E}\left(Z_{k,n}^+|\mathcal{F}_{k-1}\right) = \left(\mathbb{E}\left\langle \Gamma^\dagger X_{k-1}, X_{k,n}^\# \right\rangle \mathbb{E}\left(\epsilon_k|\mathcal{F}_{k-1}\right)\right)$, where the $\mathcal{F}_{k-1}$ is a σ-algebra generated by $(\epsilon_l)_{l\leq j}$.

Considering that both sequences are centered we then obtain $\mathbb{E}(Z_{k,n}^+|\mathcal{F}_{k-1}) = 0$.

Next we show that weak convergence to the Gaussian limit depends only on the term

$Z^+_{k,n}$. Define

$$S^+_n = \frac{1}{\sqrt{nk_n}} \sum_{k=1}^{n} Z^+_{k,n} \xrightarrow{w} \mathcal{G}(0, \Gamma_\epsilon)$$

We prove that S^+_n is a uniformly tight sequence and that finite distributions, when computed on a sufficiently large set of functionals converge to Gaussian limits as in Proposition 5.2 of [44].

We first compute the covariance and cross covariance of $Z^+_{k,n}$, where

$$Z^+_{k,n} = \left\langle \Gamma^\dagger X_{k-1}, \epsilon_{n+2} + \Psi(\epsilon_{n+1}) + ... + \Psi^{n+2-(k+1)}(\epsilon_{k+1}) \right\rangle \epsilon_k \tag{3.7}$$

Lemma 3. *If $k < j$, $\mathbb{E}\left(Z^+_{k,n} \otimes Z^+_{j,n}\right) = 0$ and*

$$\mathbb{E}\left(Z^+_{k,n} \otimes Z^+_{k,n}\right) = \Gamma_\epsilon \left(k_n - tr\left(\Gamma^\dagger \Psi^{n-k+1} \Gamma (\Psi^*)^{n-k+1}\right)\right)$$

The lemma is analogous to Lemma 5.8 of [44], and the proof is as follows

$$\begin{aligned} Z^+_{k,n} \otimes Z^+_{j,n} &= \left\langle \Gamma^\dagger X_{k-1}, X^\#_{k,n} \right\rangle \left\langle \Gamma^\dagger X_{j-1}, X^\#_{j,n} \right\rangle (\epsilon_k \otimes \epsilon_j) \\ X_{j-1} &= \Psi^{j-1-(k-1)}(X_{k-1}) + \epsilon_{j-1} + \Psi(\epsilon_{j-2}) + ... + \Psi^{j-1-k}(\epsilon_k) \end{aligned}$$

Splitting $Z^+_{k,n} \otimes Z^+_{j,n}$ into two terms, for $j < k$:

$$\mathbb{E}\left[\left\langle \Gamma^\dagger X_{k-1}, X^\#_{k,n} \right\rangle \left\langle \Gamma^\dagger \left(\epsilon_{j-1} + ... + \Psi^{j-1-k}(\epsilon_k)\right), X^\#_{j,n} \right\rangle (\epsilon_k \otimes \epsilon_j)\right] = 0$$

since X_{k-1} is centered and independent from the other terms. The second term

$$\mathbb{E}\left[\left\langle \Gamma^\dagger X_{k-1}, X^\#_{k,n} \right\rangle \left\langle \Gamma^\dagger \Psi^{j-k}(X_{k-1}), X^\#_{j,n} \right\rangle (\epsilon_k \otimes \epsilon_j)\right] = 0$$

since ϵ_k is independent from all other terms and its expectation is zero.

Further,

$$\begin{aligned} \mathbb{E}\left(Z^+_{k,n} \otimes Z^+_{k,n}\right) &= \left(\mathbb{E}\left\langle \Gamma^\dagger X_{k-1}, X^\#_{k,n} \right\rangle^2\right) \mathbb{E}\left(\epsilon_k \otimes \epsilon_k\right) \\ &= \left(\mathbb{E}\left\langle \Gamma^\dagger X_{k-1}, X^\#_{k,n} \right\rangle^2\right) \Gamma_\epsilon \end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left\langle \Gamma^{\dagger}X_{k-1}, X_{k,n}^{\#}\right\rangle^{2} &= \mathbb{E}\left(\mathbb{E}\left\langle X_{k-1}, \Gamma^{\dagger}X_{k,n}^{\#}\right\rangle^{2} | X_{k,n}^{\#}\right) \\ &= \mathbb{E}\left\|\Gamma^{1/2}\Gamma^{\dagger}X_{k,n}^{\#}\right\|^{2} \\ &= \mathbb{E}\left\|\Gamma^{\dagger 1/2}X_{k,n}^{\#}\right\|^{2} \\ &= tr\left(\Gamma^{\dagger}\Gamma_{k,n}^{\#}\right)\end{aligned}$$

where the second last step follows from the underlying idea that $\Gamma^{\dagger} \underset{n\to\infty}{\to} \Gamma^{-1}$ thus $\Gamma^{1/2}\Gamma^{\dagger} \approx \Gamma^{1/2}\Gamma^{-1} \approx \Gamma^{-1/2} \approx \Gamma^{\dagger 1/2}$,

$$\begin{aligned}\Gamma_{k,n}^{\#} &= \mathbb{E}\left(X_{k,n}^{\#} \otimes X_{k,n}^{\#}\right) \\ &= \Gamma_{\epsilon} + \Psi\Gamma_{\epsilon}\Psi^{*} + ... + \Psi^{n+1-k}\Gamma_{\epsilon}(\Psi^{*})^{n+2-(k+1)} \\ &= \Gamma - \Psi^{n-k+2}\Gamma(\Psi^{*})^{n-k+2}\end{aligned}$$

and

$$\begin{aligned}tr\left(\Gamma^{\dagger}\Gamma_{k,n}^{\#}\right) &= tr\left(\Gamma^{\dagger}\Gamma\right) - tr\left(\Gamma^{\dagger}\Psi^{n-k+2}\Gamma(\Psi^{*})^{n-k+2}\right) \\ &= k_n - tr\left(\Gamma^{\dagger}\Psi^{n-k+2}\Gamma(\Psi^{*})^{n-k+2}\right)\end{aligned}$$

which completes the proof of the lemma.

Lemma 4. *With* $\sigma_{\epsilon,x}^{2} = \mathbb{E}\left\langle \epsilon_k, x\right\rangle^{2}$,

$$\frac{1}{\sqrt{nk_n}}\sum_{k=1}^{n}\left\langle Z_{k,n}^{+}, x\right\rangle \xrightarrow{w} \mathcal{N}\left(0, \sigma_{\epsilon,x}^{2}\right)$$

This lemma aims at proving convergence of the finite dimensional distributions to a Gaussian limit. To prove the lemma, we consider Theorem 2.3 of McLeish, [39] which deals with convergence to Gaussian limits for martingale difference arrays and apply it to the real valued martingale difference array $\left\langle Z_{k,n}^{+}, x\right\rangle$. It is enough to show that $\sum_{k=1}^{n} tr\left(\Gamma^{\dagger}\Gamma_{k,n}^{\#}\right) \sim nk_n$ i.e

$$\frac{\sum_{k=1}^{n} tr\left(\Gamma^{\dagger}\Gamma_{k,n}^{\#}\right) - nk_n}{nk_n} = \frac{-\sum_{k=1}^{n} tr\left(\Gamma^{\dagger}\Psi^{n-k+2}\Gamma(\Psi^{*})^{n-k+2}\right)}{nk_n} \to 0$$

When this holds we have that the limiting covariance will be $\sigma_{\epsilon,x}^{2}$ as required in condition (c) of the Theorem. From the cyclic property of the trace which implies that the trace

is invariant under cyclic permutations, i.e. $tr(ABCD) = tr(BCDA) = tr(DABC)$ for A, B, C, D (square) matrices, we have

$$\begin{aligned} |tr\left(\Gamma^{\dagger}\Psi^{n-k+2}\Gamma(\Psi^{*})^{n-k+2}\right)| &= |tr\left((\Psi^{*})^{n-k+2}\Gamma^{\dagger}\Psi^{n-k+2}\Gamma\right)| \\ &\leq \|(\Psi^{*})^{n-k+2}\Gamma^{\dagger}\Psi^{n-k+2}\|_{\mathcal{L}}|tr\Gamma| \\ &= \|(\Psi^{*})^{n-k+1}\tilde{\Psi}^{*}\Gamma^{1/2}\Gamma^{\dagger}\Gamma^{1/2}\tilde{\Psi}\Psi^{n-k+1}\|_{\mathcal{L}}|tr\Gamma| \\ &\leq \|\Psi^{n-k+1}\|^{2}\|\tilde{\Psi}^{*}\|_{\mathcal{L}}\|\tilde{\Psi}\|_{\mathcal{L}}|tr\Gamma| \end{aligned}$$

Now $\|\Psi^{n-k+1}\| < 1$ which implies that its square is bounded as the next two terms. $|tr\Gamma|$ is also finite dimensional since from the assumptions we have that Γ is a symmetric compact operator which allows the decomposition given in Equation 2.6 so that $tr\Gamma = \sum_{j} \lambda_{j} < \infty$. Therefore as $nk_{n} \to \infty$; $\sum_{k=1}^{n} \left(\mathbb{E}\left\langle \Gamma^{\dagger}X_{k-1}, X_{k,n}^{\#}\right\rangle^{2}\right) \sim nk_{n}$ which ends the proof of the Lemma. In order to complete the proof of convergence to Gaussian limits of S_{n}^{+}, we need to show that S_{n}^{+} is tight. Looking at Lemma 5.9 of [44], we have

$$\limsup_{m\to+\infty} \sup_{n} \mathbb{P}\left(\left\|\left(I - \mathcal{P}_{m}\right) S_{n}^{+}\right\| > \epsilon\right) = 0$$

where $\mathcal{P}_{m}$ denotes the projector associated to the m first eigenvectors of the covariance operator Γ_{ϵ} of ϵ_{1}. The lemma proves that with a prescribed probability the sequence S_{n}^{+} is concentrated in the ϵ-neighbourhood of a finite dimensional space, (for some $\epsilon > 0$). The probability above is bounded using Chebyshevs inequality, so that we have

$$\mathbb{P}\left(\left\|\left(I - \mathcal{P}_{m}\right) S_{n}^{+}\right\| > \epsilon\right) \leq \frac{\mathbb{E}\left(\left\|\left(I - \mathcal{P}_{m}\right) S_{n}^{+}\right\|^{2}\right)}{\epsilon^{2}}$$

Expanding the right hand side of the equation and noting that X_{k-1} and ϵ_{k} are not correlated, it is possible to write the squared expectation as a product of two terms which are non correlated and which can be shown to converge to zero. Now the results of this lemma are not altered when $X_{n+k},\ k \geq 2$ is considered since it is still possible to separate the terms as in the case of X_{n+1}. Thus the lemma holds in our case too and since we have tightness and convergence in distribution to Gaussian limits we conclude that the Theorem 3.2.1 holds for the case $X_{n+k},\ k \geq 2$.

For the extension to all other x in $\mathcal{H}$, we quickly note that, as before the inverse of the covariance operator is not bounded and therefore, as in the i.i.d case, see Cardot *et. al.*,

[10], this will only hold for those x on the dense vector space given by $\mathcal{D}(\Gamma^{-1}) = Im\ \Gamma = \left\{ x \in \mathcal{H}, \sum_{i=1}^{n} \frac{\langle x, \nu_p \rangle^2}{\lambda_p^2} < +\infty \right\}$, see Mas, [44].

We finally show that the sequences $Z_{k,n}^{-}$ and $Z_{k,n}^{0}$ converge to zero in probability and thus have no influence on the weak convergence in our case too. This follows from Lemma 5.10 of Mas. In their proof, they show that the squared expectations of the sequences which serve as an upper bound converge to zero. The proof consists in bounding the terms which involve ϵ and X_k separately, taking into account that due to strict stationarity the X_k are identically distributed as are the ϵ. Finally the Karhunen-Loève extension of X is pulled in to complete the bounding of the terms. Although in the case X_{n+2} we have some slight changes in the representation of the terms, this does not affect the results of the lemma and thus we can also conclude that the lemma holds for the case X_{n+2}. The proof of the Theorem is therefore complete. 1

3.4 Random Projector

We check whether Theorem 3.2.1 holds when the random projector $\hat{\Pi}_{k_n}$ is replaced by the non-random projector Π_{k_n}. Although in the case of the i.i.d data the same holds, see Cardot *et.al.*, [10], in the case of dependence the same holds under an additional condition. This is because unlike in the i.i.d case, the dependence of the data introduces a cross covariance term which does not vanish asymptotically unless an additional condition is introduced.

Proposition 3.4.1. *If* $\frac{1}{\sqrt{n}} k_n^{5/2} (log\ k_n)^2 \to 0$ *as* n *goes to infinity, then*

$$\sqrt{\frac{n}{k_n}} \Psi \left(\hat{\Pi}_{k_n} - \Pi_{k_n} \right) (X_{n+1}) \underset{p}{\to} 0 \tag{3.8}$$

where Π_{k_n} is the projector onto the eigenspace associated to the k_n first eigenvalues and $\hat{\Pi}_{k_n}$ its empirical counterpart, i.e projector on the eigenspace associated to $\hat{\lambda_1}, \hat{\lambda_2}, \cdots, \hat{\lambda_{k_n}}$ Before we state the proof, we state without proof some lemmas from Cardot *et. al.*[10] which will be required together with those at the beginning of this Chapter.

Below is Lemma 1 of [10]

Lemma 5. *Consider two large enough positive integers j and k such that $k > j$. Then*

$$j\lambda_j \geq k\lambda_k \text{ and } \quad \lambda_j - \lambda_k \geq \left(1 - \frac{j}{k}\right)\lambda_j$$

Besides

$$\sum_{j \geq k} \lambda_j \leq (k+1)\lambda_k$$

Next we state Lemma 4 of the same paper
Lemma 6. *Denoting*

$$\mathcal{E}_j(z) = \left\{ \left\| (zI - \Gamma)^{-1/2} \left(\hat{\Gamma}_n - \Gamma\right) (zI - \Gamma)^{-1/2} \right\|_{\mathcal{L}} < \frac{1}{2},\ z \in \mathcal{B}_j \right\},$$

the following holds

$$\left\| (zI - \Gamma)^{1/2} \left(zI - \hat{\Gamma}_n\right)^{-1} (zI - \Gamma)^{1/2} \right\|_{\mathcal{L}} \mathbb{1}_{\mathcal{E}_j(z)} \leq C,\ a.s.$$

where C is some positive constant. Besides

$$\mathbb{P}\left(\mathcal{E}_j(z)\right) \leq \frac{j\ \log\ j}{\sqrt{n}}$$

Proof. This proof follows closely that of Cardot *et. al.* and seeks to extend it to the dependent case. We consider another stationary process $\{X_n^{'}\}$ which is independent of $\{X_n\}$ and satisfies the same equation as $\{X_n\}$ i.e.

$$X_n^{'} = \Psi X_{n-1}^{'} + \epsilon_n \qquad \epsilon_n\ i.i.d\ \mathcal{N}\left(0, \sigma_\epsilon^2\right) \tag{3.9}$$

We use the process $\{X_n^{'}\}$ to estimate $\hat{\Pi}_{k_n}\Psi$, after which we apply the operator to our original process in which case the operator is independent from X_{n+1}, see Shibata, [51]. Now

$$\begin{aligned}
\left(\hat{\Pi}_{k_n} - \Pi_{k_n}\right) &= \frac{1}{2\pi i} \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} \left[\left(zI - \hat{\Gamma}_n\right)^{-1} - (zI - \Gamma)^{-1} \right] dz \\
&= \frac{1}{2\pi i} \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} \left[\left(zI - \hat{\Gamma}_n\right)^{-1} \left(\hat{\Gamma}_n - \Gamma\right) (zI - \Gamma)^{-1} \right] dz \\
&= \mathcal{S}_n + \mathcal{R}_n
\end{aligned} \tag{3.10}$$

where

$$\mathcal{S}_n = \frac{1}{2\pi i} \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} \left[(zI - \Gamma)^{-1} \left(\hat{\Gamma}_n - \Gamma\right) (zI - \Gamma)^{-1} \right] dz \tag{3.11}$$

and

$$\mathcal{R}_n = \frac{1}{2\pi i}\sum_{j=1}^{k_n}\int_{\mathcal{B}_j}\left[(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)\left(zI-\hat{\Gamma}_n\right)^{-1}\right]dz \quad (3.12)$$

The second step (3.10) can be verified as

$$\begin{aligned}\left[\left(zI-\hat{\Gamma}_n\right)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)(zI-\Gamma)^{-1}\right] &= \left(zI-\hat{\Gamma}_n\right)^{-1}\left(\hat{\Gamma}_n-zI+zI-\Gamma\right)(zI-\Gamma)^{-1}\\ &= \left(zI-\hat{\Gamma}_n\right)^{-1}\left((zI-\Gamma)-\left(zI-\hat{\Gamma}_n\right)\right)(zI-\Gamma)^{-1}\\ &= \left(zI-\hat{\Gamma}_n\right)^{-1}(zI-\Gamma)(zI-\Gamma)^{-1}\\ &\quad -\left(zI-\hat{\Gamma}_n\right)^{-1}\left(zI-\hat{\Gamma}_n\right)(zI-\Gamma)^{-1}\\ &= \left(zI-\hat{\Gamma}_n\right)^{-1}-(zI-\Gamma)^{-1}\end{aligned}$$

From the relation $\hat{\Pi}_{k_n}-\Pi_{k_n}=\mathcal{S}_n+\mathcal{R}_n$, we have that $-\mathcal{S}_n+\hat{\Pi}_{k_n}-\Pi_{k_n}=\mathcal{R}_n$. Looking at the term in $\mathcal{R}_n$ we have

$$\begin{aligned}&\left[(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)\left(zI-\hat{\Gamma}_n\right)^{-1}\right]\\ &=-(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)\left[(zI-\Gamma)^{-1}-\left(zI-\hat{\Gamma}_n\right)^{-1}\right]\\ &=-(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)(zI-\Gamma)^{-1}-(zI-\Gamma)^{-1}\left(\Gamma-\hat{\Gamma}_n\right)\left(zI-\hat{\Gamma}_n\right)^{-1}\\ &=-(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n-\Gamma\right)(zI-\Gamma)^{-1}+\left(zI-\hat{\Gamma}_n\right)^{-1}-(zI-\Gamma)^{-1}\end{aligned}$$

Thus $\hat{\Pi}_{k_n}-\Pi_{k_n}=\mathcal{S}_n+\mathcal{R}_n$. Since the operator is estimated from a different process, we have that $\mathcal{S}_n$ and $\mathcal{R}_n$ are both independent from X_{n+1}, therefore

$$\begin{aligned}\mathbb{E}\left\|\Psi\mathcal{S}_n\left(X_{n+1}\right)\right\|^2 &= \mathbb{E}\sum_{l,l'=1}^{+\infty}\left\langle\Psi\mathcal{S}_n(e_l),\Psi\mathcal{S}_n(e_l)'\right\rangle\left\langle X_{n+1},e_{l'}\right\rangle\left\langle X_{n+1},e_l\right\rangle &(3.13)\\ &= \mathbb{E}\sum_{l=1}^{+\infty}\left\|\Psi\mathcal{S}_n(e_l)\right\|^2\left\langle X_{n+1},e_l\right\rangle^2 &(3.14)\\ &= \sum_{l=1}^{+\infty}\lambda_l\mathbb{E}\left\|\Psi\mathcal{S}_n(e_l)\right\|^2 &(3.15)\end{aligned}$$

since $\mathbb{E}\left(\left\langle X_{n+1},e_l\right\rangle\left\langle X_{n+1},e_{l'}\right\rangle\right)=0$ if $l\neq l'$ by the definition of the e_l i.e orthonormality. $\mathbb{E}\left\langle X_{n+1},e_l\right\rangle^2=\lambda_l$ follows from the Karhunen-Loève expansion of X, i.e.

$$X=\sum_{l=1}^{\infty}\sqrt{\lambda_l}\xi_l e_l \quad (3.16)$$

so that

$$\mathbb{E}\left\langle X_{n+1}, e_l\right\rangle \left\langle X_{n+1}, e_{l'}\right\rangle = \begin{cases} \lambda_l & \text{if } l = l' \\ 0 & \text{if } l \neq l' \end{cases}$$

Next

$$\mathbb{E}\left\|\Psi\mathcal{S}_n(e_l)\right\|^2 = \mathbb{E}\left\|\sum_{l'=1}^{+\infty} \Psi(e_{l'})\left\langle \mathcal{S}_n(e_l), e_{l'}\right\rangle\right\|^2 \tag{3.17}$$

This follows from the fact that $\mathcal{S}_n e_l$ can be expressed as

$$\mathcal{S}_n e_l = \sum_{j=1}^{\infty} \left\langle \mathcal{S}_n e_l, e_j\right\rangle e_j$$

From Cardot *et. al.* [10] the explicit computation of the operator $\mathcal{S}_n$ can be given by

$$\frac{1}{2\pi i}\int_{\mathcal{B}_j}\left[(zI-\Gamma)^{-1}\left(\hat{\Gamma}_n - \Gamma\right)(zI-\Gamma)^{-1}\right]dz = v_j\left(\hat{\Gamma}_n - \Gamma\right)\pi_j + \pi_j\left(\hat{\Gamma}_n - \Gamma\right)v_j$$

with $v_j = \sum_{j'\neq j}\frac{1}{\lambda_{j'}-\lambda_j}\pi_{j'}$ where π_j is the projector on the eigenspace associated to the jth eigenfunction of Γ. We note that $\Gamma, \hat{\Gamma}_n, v_j, \pi_j$ are self-adjoint operators and that a linear combination of self adjoint transformations will result in a self adjoint transformation. We can therefore write

$$\left\langle \mathcal{S}_n e_l, e_{l'}\right\rangle = \sum_{j=1}^{k_n}\left[\left\langle\left(\hat{\Gamma}_n - \Gamma\right)\pi_j e_l, v_j e_{l'}\right\rangle + \left(\hat{\Gamma}_n - \Gamma\right)v_j e_l, \pi_j e_{l'}\right] \tag{3.18}$$

This gives rise to four different possibilities which we list below

1. $l' > k_n$ and $l > k_n$
   In this case $\pi_j e_l = 0$ and $\pi_j e_{l'} = 0$ since $1 \leq j \leq k_n$ and because of orthonormality the terms will be zero.

2. $l' \leq k_n$ and $l \leq k_n$
   Here, for some $1 \leq l, l' \leq k_n$ and keeping in mind that e_l are orthonormal, the sum 3.18 reduces to

$$\begin{aligned} &= \left\langle\left(\hat{\Gamma}_n - \Gamma\right)\pi_l e_l, v_l e_{l'}\right\rangle + \left(\hat{\Gamma}_n - \Gamma\right)v_{l'} e_l, \pi_{l'} e_{l'} \\ &= \left\langle\left(\hat{\Gamma}_n - \Gamma\right) e_l, v_l e_{l'}\right\rangle + \left(\hat{\Gamma}_n - \Gamma\right)v_{l'} e_l, e_{l'} \end{aligned}$$

   If $l = l'$ then $v_l e_{l'} = v_{l'} e_l = 0$ from the definition of v_j. If $l \neq l'$ then $\left\langle\left(\hat{\Gamma}_n - \Gamma\right) e_l, v_l e_{l'}\right\rangle + \left(\hat{\Gamma}_n - \Gamma\right)v_{l'} e_l, e_{l'}$ reduces to

$$\begin{aligned} &= \left\langle\left(\hat{\Gamma}_n - \Gamma\right) e_l, \frac{1}{\lambda_{l'}-\lambda_l}\pi_{l'} e_{l'}\right\rangle + \left\langle\left(\hat{\Gamma}_n - \Gamma\right)\frac{1}{\lambda_l - \lambda_{l'}}\pi_l e_l, e_{l'}\right\rangle \\ &= \frac{1}{\lambda_{l'}-\lambda_l}\left\langle\left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'}\right\rangle + \frac{1}{\lambda_l - \lambda_{l'}}\left\langle\left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'}\right\rangle \\ &= \frac{1}{\lambda_{l'}-\lambda_l}\left\langle\left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'}\right\rangle - \frac{1}{\lambda_{l'} - \lambda_l}\left\langle\left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'}\right\rangle \\ &= 0 \end{aligned}$$

3. $l^{'} > k_n$ and $l \leq k_n$
   In this case 3.18 reduces to

$$\begin{aligned}\left\langle \left(\hat{\Gamma}_n - \Gamma\right)\pi_l e_l, v_l e_{l'} \right\rangle &= \left\langle \left(\hat{\Gamma}_n - \Gamma\right) e_l, v_l e_{l'} \right\rangle \\ &= \left\langle \left(\hat{\Gamma}_n - \Gamma\right) e_l, \frac{1}{\lambda_{l'} - \lambda_l}\pi_{l'} e_{l'} \right\rangle \\ &= \frac{\left\langle \left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_{l'} - \lambda_l}\end{aligned}$$

4. $l^{'} \leq k_n$ and $l > k_n$ 3.18 reduces to

$$\begin{aligned}\left\langle \left(\hat{\Gamma}_n - \Gamma\right) v_{l'} e_l, \pi_{l'} e_{l'} \right\rangle &= \left\langle \left(\hat{\Gamma}_n - \Gamma\right) \frac{1}{\lambda_l - \lambda_{l'}}\pi_l e_l, e_{l'} \right\rangle \\ &= \frac{\left\langle \left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_l - \lambda_{l'}}\end{aligned}$$

Collecting all the cases together we then have that

$$\langle \mathcal{S}_n e_l, e_{l'} \rangle = \begin{cases} 0 & \text{if } \left(l^{'} \leq k_n \text{ and } l \leq k_n\right) \text{ or if } \left(l^{'} > k_n \text{ and } l > k_n\right) \\ \dfrac{\left\langle \left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_{l'} - \lambda_l} & \text{if } \left(l^{'} > k_n \text{ and } l \leq k_n\right) \\ \dfrac{\left\langle \left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_l - \lambda_{l'}} & \text{if } \left(l^{'} \leq k_n \text{ and } l > k_n\right) \end{cases}$$

Considering the instance when $l > k_n$ and $l^{'} \leq k_n$,

$$\begin{aligned}\mathbb{E}\left\|\Psi \mathcal{S}_n(e_l)\right\|^2 &= \mathbb{E}\left\| \sum_{l'=1}^{k_n} \Psi(e_{l'}) \frac{\left\langle \left(\hat{\Gamma}_n - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_l - \lambda_{l'}} \right\|^2 \\ &= \mathbb{E}\left\| \frac{1}{n} \sum_{j=1}^{n} \sum_{l'=1}^{k_n} \Psi(e_{l'}) \frac{\left\langle \left(X_j^{'} \otimes X_j^{'} - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_l - \lambda_{l'}} \right\|^2 \\ &= \mathbb{E}\left\| \frac{1}{n} \sum_{j=1}^{n} R_{j,l,n}^{*} \right\|^2\end{aligned}$$

where

$$R_{j,l,n}^{*} = \sum_{l'=1}^{k_n} \Psi(e_{l'}) \frac{\left\langle \left(X_j^{'} \otimes X_j^{'} - \Gamma\right) e_l, e_{l'} \right\rangle}{\lambda_l - \lambda_{l'}}$$

and $R^*_{j,l,n}$ are centered and correlated random vectors, i.e.,

$$\begin{aligned}\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n}R^*_{j,l,n}\right\|^2 &= \frac{1}{n}\mathbb{E}\left\|\sum_{l'=1}^{k_n}\Psi(e_{l'})\frac{\left\langle X_1^{'},e_l\right\rangle\left\langle X_1^{'},e_{l'}\right\rangle}{\lambda_l-\lambda_{l'}}\right\|^2 \\ &\quad+\frac{2}{n^2}\mathbb{E}\left\|\sum_{k=2}^{n}\sum_{j<k}\sum_{l',m=1}^{k_n}\left\langle\Psi(e_{l'}),\Psi(e_m)\right\rangle\frac{\left\langle X_j^{'},e_l\right\rangle\left\langle X_j^{'},e_{l'}\right\rangle}{\lambda_l-\lambda_{l'}}\frac{\left\langle X_k^{'},e_l\right\rangle\left\langle X_k^{'},e_{l'}\right\rangle}{\lambda_l-\lambda_m}\right\|\end{aligned}$$

Looking at the first term on the right hand side, and considering that our process is strictly stationary, we have

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\left\|\sum_{l'=1}^{k_n}\Psi(e_{l'})\frac{\left\langle\left(X_1^{'}\otimes X_1^{'}\right)e_l,e_{l'}\right\rangle}{\lambda_{l'}-\lambda_l}\right\|^2 \\ &=\frac{1}{n}\mathbb{E}\left\|\sum_{l'=1}^{k_n}\Psi(e_{l'})\frac{\left\langle X_1^{'},e_l\right\rangle\left\langle X_1^{'},e_{l'}\right\rangle}{\lambda_{l'}-\lambda_l}\right\|^2 \\ &=\frac{1}{n}\mathbb{E}\sum_{m,l'=1}^{k_n}\left\langle\Psi(e_{l'}),\Psi(e_m)\right\rangle\frac{\left\langle X_1^{'},e_l\right\rangle\left\langle X_1^{'},e_{l'}\right\rangle\left\langle X_1^{'},e_l\right\rangle\left\langle X_1^{'},e_m\right\rangle}{\left(\lambda_{l'}-\lambda_l\right)\left(\lambda_m-\lambda_l\right)}\end{aligned}$$

By the Karhunen-Loève expansion of X, this reduces to

$$\frac{1}{n}\sum_{m,l'=1}^{k_n}\left\langle\Psi(e_{l'}),\Psi(e_m)\right\rangle\frac{\lambda_l\sqrt{\lambda_{l'}\lambda_m}\mathbb{E}\left(\xi_l^2\xi_{l'}\xi_m\right)}{\left(\lambda_{l'}-\lambda_l\right)\left(\lambda_m-\lambda_l\right)}$$

Applying Cauchy-Schwarz twice to ξ_k's yields

$$\begin{aligned}\mathbb{E}\left(\xi_l^2\xi_{l'}\xi_m\right) &\leq \sqrt{\mathbb{E}\left(\xi_l^4\right)}\sqrt{\mathbb{E}\left(\xi_{l'}^2\xi_m^2\right)} \\ &\leq \sqrt{M}\sqrt{M}\end{aligned}$$

Thus the first term on the right hand side is bounded by

$$\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n}R^*_{1,l,n}\right\|^2 \leq \frac{M}{n}\left\|\sum_{l'=1}^{k_n}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2$$

by using the Karhunen-Loève expansion of X, assumption 3.3 and Lemma 5.1 of [44].

Considering the second term on the right hand side, we have

$$\frac{2}{n^2}\mathbb{E}\sum_{k=2}^{n}\sum_{j<k}\sum_{l',m=1}^{k_n}\langle\Psi(e_{l'}),\Psi(e_m)\rangle\,\frac{\left\langle X_j^{'},e_l\right\rangle\left\langle X_j^{'},e_{l'}\right\rangle}{\lambda_l-\lambda_{l'}}\,\frac{\left\langle X_k^{'},e_l\right\rangle\left\langle X_k^{'},e_m\right\rangle}{\lambda_l-\lambda_m}$$

using the same arguments employed above of stationarity and Karhunen-Loève expansion,

$$=\frac{2}{n^2}\sum_{k=2}^{n}\sum_{j<k}\sum_{l',m=1}^{k_n}\langle\Psi(e_{l'}),\Psi(e_m)\rangle\,\frac{\lambda_l\sqrt{\lambda_{l'}\lambda_m}\mathbb{E}\left(\xi_l^2\xi_{l'}\xi_m\right)}{\left(\lambda_l-\lambda_{l'}\right)\left(\lambda_l-\lambda_m\right)}$$

The expectation term can be bounded as above so that

$$=\frac{2}{n^2}\sum_{k=2}^{n}\sum_{j<k}M\sum_{l',m=1}^{k_n}\langle\Psi(e_{l'}),\Psi(e_m)\rangle\,\frac{\lambda_l\sqrt{\lambda_{l'}\lambda_m}\mathbb{E}\left(\xi_l^2\xi_{l'}\xi_m\right)}{\left(\lambda_l-\lambda_{l'}\right)\left(\lambda_l-\lambda_m\right)}$$

$$\leq\frac{2}{n^2}\sum_{k=2}^{n}\sum_{j<k}M\sum_{l'=1}^{k_n}\left\|\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2$$

where the summand is independent of j and k.

Now

$$\sum_{k=2}^{n}\sum_{j=1}^{k-1}1=\sum_{k=2}^{n}(k-1)=\sum_{k=1}^{n-1}k=\frac{n(n-1)}{2}$$

Finally we have

$$\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n}R_{j,l,n}^{*}\right\|^2 \leq \frac{M}{n}\left\|\sum_{l'=1}^{k_n}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 \tag{3.19}$$

$$+\frac{2M}{n^2}\left(1-\frac{1}{n}\right)\left\|\sum_{l'=1}^{k_n}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 \tag{3.20}$$

$$\leq M\left\|\sum_{l'=1}^{k_n}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 \tag{3.21}$$

Similar calculations for the case $l'>k_n$ and $l\leq k_n$ lead to

$$\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n}R_{j,l,n}^{*}\right\|^2\leq M\left\|\sum_{l'\geq k_n+1}^{+\infty}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 \tag{3.22}$$

We see that Cardot *et. al.*'s, [10] method does not follow through for the case of depen-

dence since the covariance term leads to

$$\frac{1}{k_n}\mathbb{E}\left\|\Psi\mathcal{S}_n(X_{n+1})\right\|^2 \leq \frac{M}{k_n}\sum_{l=1}^{k_n}\lambda_l\left\|\sum_{l'\geq k_n+1}^{+\infty}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 \tag{3.23}$$

$$+\frac{M}{k_n}\sum_{l>k_n}\lambda_l\left\|\sum_{l'\geq 1}^{k_n}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 \tag{3.24}$$

whereas we are interested in $\frac{n}{k_n}\mathbb{E}\left\|\Psi\mathcal{S}_n(X_{n+1})\right\|^2$. To overcome this problem, we introduce a slightly stronger assumption on the λ_i's. Cardot *et. al.* [10] have the condition that for a sufficiently large i, $\lambda_i i \; log \; i < C$. We on the other hand consider the conditions

$$n(1+k_n\log k_n)\sum_{j>k_n+\sqrt{k_n/\log k_n}}\lambda_j\to 0,\ n\to\infty \quad \text{and} \quad nk_n^2\sqrt{\frac{k_n}{\log k_n}}\lambda_{k_n}\to 0,\ n\to\infty.$$

If, e.g., $\lambda_j = ba^j$ for some $b>0$, $0<a<1$, then both conditions are satisfied if $\limsup\frac{\log n}{k_n} < \log\frac{1}{a}$. Having this in mind we can then write the above equation as

$$\frac{n}{k_n}\mathbb{E}\left\|\Psi\mathcal{S}_n(X_{n+1})\right\|^2 \leq \underbrace{\frac{n}{k_n}M\sum_{l=1}^{k_n}\lambda_l\left\|\sum_{l'\geq k_n+1}^{+\infty}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2}_{I} \tag{3.25}$$

$$+\underbrace{\frac{n}{k_n}M\sum_{l>k_n}\lambda_l\left\|\sum_{l'\geq 1}^{k_n}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2}_{II} \tag{3.26}$$

We need a minimal condition to bound I and II. Looking at I (3.25), we have by Lemma 5 and from $\|\Psi\|_{\mathcal{L}}<1$,

$$\begin{aligned}\left\|\sum_{l'\geq k_n+1}^{+\infty}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 &= \left\|\Psi\left(\sum_{l'\geq k_n+1}^{+\infty}\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}e_{l'}\right)\right\|^2\\ &\leq \left\|\sum_{l'\geq k_n+1}^{+\infty}\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}e_{l'}\right\|^2 = \sum_{l'\geq k_n+1}^{+\infty}\frac{\lambda_l\lambda_{l'}}{(\lambda_l-\lambda_{l'})^2}\\ &\leq \sum_{l'\geq k_n+1}^{+\infty}\frac{\lambda_{l'}}{\lambda_l\left(1-(l/l')\right)^2}\end{aligned}$$

as $e_{l'}$ are an orthonormal basis. We therefore have for I:

$$\frac{n}{k_n}M\sum_{l=1}^{k_n}\lambda_l\left\|\sum_{l'\geq k_n+1}^{+\infty}\Psi(e_{l'})\frac{\sqrt{\lambda_l\lambda_{l'}}}{\lambda_l-\lambda_{l'}}\right\|^2 \leq nM\sum_{l'\geq k_n+1}^{+\infty}\frac{\lambda_{l'}}{\left(1-\frac{l}{l'}\right)^2} \tag{3.27}$$

Setting $h_n = \left[\sqrt{\frac{k_n}{log\ k_n}}\right]$ where $[x]$ denotes the largest integer smaller than $[x]$. We can split the right-hand side into summation from k_n+h_n+1 to ∞ and from k_n+1 to k_n+h_n, and we use that for $l^{'} \geq k_n + h_n$ and $l < k_n$, $l'k_n > l\,(k_n + h_n)$, thus

$$1 - \frac{l}{l'} \geq 1 - \frac{k_n}{k_n + h_n} = \frac{h_n}{k_n + h_n}$$

hence,

$$\begin{aligned} nM \sum_{l'>k_n+h_n}^{\infty} \frac{\lambda_l^{'}}{\left(1 - \frac{l}{l'}\right)^2} &\leq nM \sum_{l'>k_n+h_n}^{\infty} \lambda_{l'} \left(1 + \frac{k_n}{h_n}\right)^2 \\ &\leq nM \sum_{l'>k_n+h_n} \lambda_{l'}(1 + k_n \log k_n) \to 0 \end{aligned}$$

by our assumption on $n, k_n, \lambda_j, j > k_n + h_n$.

Analogously, we get for $1 \leq l \leq k_n$

$$\begin{aligned} nM \sum_{l'=k_n+1}^{k_n+h_n} \frac{\lambda_{l'}}{\left(1 - \frac{l}{l'}\right)^2} &\leq nM(k_n + 1)^2 \lambda_{k_n+1} h_n \\ &\leq nM(k_n + 1)^2 \sqrt{\frac{k_n}{\log k_n}} \lambda_{k_n} \to 0 \end{aligned}$$

Hence, the term II (3.26) also goes to zero using similar arguments as $n \to \infty$, and we get

$$\frac{n}{k_n} \mathbb{E} \left\| \Psi \mathcal{S}_n(X_{n+1}) \right\|^2 \to 0 \tag{3.28}$$

The second part of the proof, i.e showing that

$$\sqrt{\frac{n}{k_n}} \left\| \Psi \mathcal{R}_n(X_{n+1}^{'}) \right\|^2 = \mathcal{O}_p\left(\frac{1}{\sqrt{n}} k_n^{5/2}\ (\log\ k_n)^2\right) \quad \text{when} \quad \frac{k_n^2\ \log\ k_n}{\sqrt{n}} \to 0$$

is quite similar to the first one and uses again the same kind of arguments as in Cardot *et. al.*, [10], so we do not give the details here. □

Chapter 4

Change Point Detection for Functional Observations

In this chapter we consider the change point problem for functional observations and an application to cognitive science data. We look at the possibility of a change point in the mean first studied by Berkes *et. al.* [4] for independent identically distributed (i.i.d) observations. In [4], they applied the test to temperature data and made the observation that apart from a change in level, change of shape of the curve can be informative, for instance the summers may be warmer and winters colder. This serves to show that functional methods incorporate a lot of data from which finer and more accurate observations and inferences can be made. Horváth *et. al.* [26] consider a change in mean for dependent observations, noting that the test for i.i.d. data fails to account for the long run covariance and therefore may give false results. Aston and Kirch [2] evaluate stationarity via change point with an application to resting state fMRI (functional Magnetic Resonance Imaging) data. In this case too we have change point detection for data with dependence. Horváth *et. al.* [26] develop a test for the stability of the FAR(1) model. In all the situations considered, we see that failure to account for a change when one exists will lead to erroneous results and inferences. In [4], [26] and [25], the tests developed are based on the CUSUM method, and as rightly noted in Aue *et. al.*, [3], a lot of open questions still remain, for instance applicability of MOSUM test in the functional case, sequential change points for functional data among others.

The chapter is divided into two sections considering a change in mean and stability of

the functional autoregressive process. In each case we give a brief introduction, overview of the test procedure and application of the procedure on a real data set consisting of stimulus response data of cells from the animal physiology group at the University of Kaiserslautern and the results obtained under asymptotic considerations. This chapter is an application of the methods developed in Berkes *et. al.* [4], Horváth *et. al.* [26] and Hörmann and Kokoszka, [25]. For completeness we include the technical assumptions used for developing the test statistics.

4.1 Change Point in the Mean

When considering observations, although for theoretical purposes it is easier to consider observations with mean zero, in reality, this is not the case. A suitable assumption would be $X_i = \mu + Y_i$ where $\mathbf{E}Y_i = 0$. When working with data, various procedures consider mean adjusted variables for instance the dimension reduction Functional Principal Components ν_k are those of $X - \mu$, with the following L^2 expansion,

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ki}\nu_k(t),\ 1 \leq i \leq N \tag{4.1}$$

This simply means that if there is a change in the mean function at some point, then the results of the (Functional Principal Component Analysis) FPCA will no longer be valid.

As noted in Berkes *et. al.* [4], detecting a change point of the mean does not necessarily mean an abrupt change from one curve to the next but that the assumption of a constant mean for the whole series is not acceptable. We consider possibility of a change in the mean while assuming that the data are independent and identically distributed (i.i.d.) in one case and dependent in the other. This is to ensure that the possibility of spurious dependence is ruled out.

4.1.1 Assumptions and Detection Procedure for i.i.d. Observations

In the case of independent observations, we test the null hypothesis

$$H_0 : \mathbb{E}X_1 = \mathbb{E}X_2 = \cdots \mathbb{E}X_N \tag{4.2}$$

against the alternative of a single or more change points.

Under the null hypothesis each functional observation is represented as

$$X_i(t) = \mu(t) + Y_i(t),\ \mathbb{E}Y_i(t) = 0 \tag{4.3}$$

where $X_i \in L^2$.

Assumptions

Assumption 4.1. *The mean $\mu(.)$ is in L^2. The errors $Y_i(.)$ are i.i.d. mean zero random elements of L^2 which satisfy*

$$\mathbb{E}\|Y_i\|^2 = \int \mathbb{E}Y_i^2(t)dt < \infty \tag{4.4}$$

The assumption above implies that the covariance function

$$c(t,s) = \mathbb{E}\left[Y_i(t)Y_i(s)\right]\ \ t,s \in [0,1] \tag{4.5}$$

is square integrable, and consequently the expansions

$$c(t,s) = \sum_{1 \le k < \infty} \lambda_k \nu_k(t)\nu_k(s) \tag{4.6}$$

and

$$Y_i(t) = \sum_{1 \le l < \infty} \xi_{l,i}\nu_l(t) \tag{4.7}$$

where $\xi_{l,i} = \langle Y_i, \nu_l \rangle$ are implied.

The estimated eigenelements are defined by

$$\int \hat{c}(t,s)\hat{\nu}_l(s)ds = \hat{\lambda}_l\hat{\nu}_l(t),\ l = 1,2,\cdots \tag{4.8}$$

where
$\hat{c}(t,s) = \frac{1}{N}\sum_{1 \le i \le N}\left(X_i(t) - \bar{X}_N(t)\right)\left(X_i(s) - \bar{X}_N(s)\right)$ and $\bar{X}_N(t) = \frac{1}{N}\sum_{1 \le i \le N} X_i(t)$

The next two assumptions together with Lemmas 1 and 2 in chapter 2 control the distance between the population and estimated (empirical) eigenelements.

Assumption 4.2. *The eigenvalues λ_l satisfy, for some $d > 0$*

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1}$$

Assumption 4.3. *The Y_i in Assumption 4.1 satisfy*

$$\mathbb{E}\|Y_i\|^4 = \int \mathbb{E}Y_i^4(t)dt < \infty. \tag{4.9}$$

Considering a single change point alternative:

Assumption 4.4. *The observations follow the model*

$$X_i(t) = \begin{cases} \mu_1(t) + Y_i(t), & 1 \le i \le k* \\ \mu_2(t) + Y_i(t), & k* < i \le N \end{cases}$$

where Y_i satisfy 4.1, μ_1, μ_2 are in $L^2(\mathcal{T})$ and $k^* = [n\theta]$ for some $0 < \theta < 1$.

Detection Procedure

Under the specified technical assumptions, we denote

$$\hat{\mu}_k(t) = \frac{1}{k}\sum_{1\le i\le k} X_i(t), \quad \tilde{\mu}_k(t) = \frac{1}{N-k}\sum_{k<i\le N} X_i(t) \tag{4.10}$$

If the mean is constant, the difference $\Delta_k(t) = \hat{\mu}_k(t) - \tilde{\mu}_k(t)$ is small for all $1 \le k < N$ and all $t \in [0, 1]$. To account for effects of chance variability when k is close to 1 or N, we consider the weighted differences

$$P_k(t) = \sum_{1\le i\le k} X_i(t) - \frac{k}{N}\sum_{1\le i\le N} X_i(t) = \frac{k(N-k)}{N}[\hat{\mu}_k(t) - \tilde{\mu}_k(t)] \tag{4.11}$$

in which the variability at the end points is attenuated by a parabolic weight function.

Since the functions are infinite dimensional, we project them onto the first say p principal components which explain most of the variability. The population eigenelements are also replaced by those from the sample so that we have

$$\int \left\{ \sum_{1\le i\le Nx} X_i(t) - \frac{[Nx]}{N}\sum_{1\le i\le N} X_i(t) \right\} \hat{\nu}_l(t)dt = \sum_{1\le i\le Nx} \hat{\xi}_{l,i} - \frac{[Nx]}{N}\sum_{1\le i\le N} \hat{\xi}_{l,i} \tag{4.12}$$

To derive the test, consider the statistic

$$T_N(x) = \frac{1}{N}\sum_{l=1}^{p} \hat{\lambda}_l^{-1} \left(\sum_{1\le i\le Nx} \hat{\xi}_{l,i} - x\sum_{1\le i\le N} \hat{\xi}_{l,i} \right)^2 \tag{4.13}$$

where $\hat{\lambda}_i$ estimates the eigen values of the covariance operator of the observations (see for instance (4.6)) and $B_1(.), \cdots, B_d(.)$ denote independent standard Brownian bridges.

Theorem 4.1.1. *Suppose Assumptions 4.1, 4.2, 4.3 hold. Then, under* H_0

$$T_N(x) \xrightarrow{d} \sum_{1 \le l \le p} B_l^2(x), \quad 0 \le x \le 1 \tag{4.14}$$

in the Skorokhod topology of $D[0,1]$.

The proof of Theorem 4.1.1 can be found in [28].

By Theorem 4.1.1, $U(T_N) \xrightarrow{d} U\left(\sum_{1 \le l \le p} B_l^2(.)\right)$ for any continuous functional $U : D[0,1] \to \mathbb{R}$. Considering the Cramér -von-Mises functional, we look at the convergence

$$\int_0^1 T_N(x)dx \xrightarrow{d} \int_0^1 \sum_{1 \le l \le p} B_l^2(x)dx$$

which can be rewritten as

$$S_{N,p} := \frac{1}{N^2} \sum_{l=1}^{p} \hat{\lambda}_l^{-1} \sum_{k=1}^{N} \left(\sum_{1 \le i \le k} \hat{\xi}_{l,i} - \frac{k}{N} \sum_{1 \le i \le N} \hat{\xi}_{l,i} \right)^2 \xrightarrow{d} \int_0^1 \sum_{1 \le l \le p} B_l^2(x)dx. \tag{4.15}$$

The distribution of the random variable

$$K_p = \int_0^1 \sum_{1 \le l \le p} B_l^2(x)dx \tag{4.16}$$

was derived by Kiefer [27]. The test rejects the null hypothesis if $S_{N,p} > c_p(\alpha)$, where $c_p(\alpha)$ is the $(1-\alpha)th$ quantile of K_p.

Results on the consistency of the test and the behaviour of the test statistic under the alternative can be found in [28].

To estimate the change point, provided the test detects one, the estimator

$$\hat{\theta}_N = \inf \left\{ x : T_N(x) = \sup_{0 \le y \le 1} T_N(y) \right\} \tag{4.17}$$

is employed, the weak consistency of which can be found in Proposition 6.1 of [28].

4.1.2 Assumptions and Detection Procedure for Dependent Observations

In this subsection we consider the test to be employed when the observations are dependent. The null hypothesis is the same as that in (4.2). Under the null hypothesis, each functional observation can be represented as (4.3).

To establish convergence of the test statistic under H_0, some assumptions are required, which we include here for completeness.

Assumption 4.5. *The mean μ is in $\mathcal{H}$. The error functions $Y_i \in \mathcal{L}_H^4$ are $L^4 - m -$ approximable mean zero random elements such that the eigenvalues of their covariance operator satisfy Assumption 4.2 above.*

$L^4 - m -$ *approximability* is a technical condition which implies that the Y_i are weakly dependent with $\nu_4(Y_i) < \infty$ (finite fourth moment). A detailed definition of the condition can be obtained from Hörmann and Kokoszka, [25]. In this case, the covariance function (4.5) is square integrable, i.e is in $L^2([0,1] \times [0,1])$.

In the case of a single change point;

Assumption 4.6. *The observations follow the model*

$$X_i(t) = \begin{cases} \mu_1(t) + Y_i(t), & 1 \leq i \leq k* \\ \mu_2(t) + Y_i(t), & k* < i \leq N \end{cases}$$

where Y_i satisfy 4.5, μ_1, μ_2 are in L^2 and $k^* = [N\theta]$ for some $0 < \theta < 1$.

The testing procedure follows from that of the i.i.d. observations with the challenge of how to accommodate the dependence. The main difference lies in the fact that when the data has dependence we consider the long run variance which accommodates the dependence between observations as opposed to the sample variance.

To obtain the new test statistic, we first define the partial sums

$$\mathbf{S}_N(x, \boldsymbol{\xi}) = \sum_{n=1}^{\lfloor Nx \rfloor} \boldsymbol{\xi}_n, \ x \in [0,1]$$

and the bridge process

$$\mathbf{L}_N(x, \boldsymbol{\xi}) = \mathbf{S}_N(x, \boldsymbol{\xi}) - x\mathbf{S}_N(1, \boldsymbol{\xi}) \tag{4.18}$$

where $\boldsymbol{\xi} = (\xi_n, n \geq 1)$ is a generic R^d-valued stochastic process. We denote by $\boldsymbol{\Sigma}(\boldsymbol{\xi})$ the long run variance of the sequence $\boldsymbol{\xi}$, and by $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\xi})$ its kernel estimator, where

$$\boldsymbol{\Sigma} = \sum_{h=-\infty}^{\infty} \mathbb{E}\left[(\xi_0 - \mathbb{E}\xi) (\xi_h - \mathbb{E}\xi)^T \right]$$

and

$$\hat{\boldsymbol{\Sigma}} = \sum_{h=-\infty}^{\infty} \hat{\gamma}_h(N), \quad \hat{\gamma}_h(N) = \frac{1}{N} \sum_{1 \leq i \leq N-h} \left(\hat{\xi}_i - \frac{1}{N} \sum \hat{\xi}_i \right) \left(\hat{\xi}_{i+h} - \frac{1}{N} \sum \hat{\xi}_i \right)^T$$

The test statistic is then given by

$$T_N(d) = \frac{1}{N}\int_0^1 \mathbf{L}_N(x,\hat{\boldsymbol{\eta}})^T \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1} \mathbf{L}_N(x,\hat{\boldsymbol{\eta}})dx \tag{4.19}$$

with the scores $\hat{\eta}_{li}$ given by

$$\hat{\eta}_{li} = \int \left(X_i(t) - \bar{X}_N(t)\right)\hat{\nu}_l(t)dt,\ 1 \le l \le d \tag{4.20}$$

Theorem 4.1.2. *Suppose H_0 and Assumption 4.5 hold. If the estimator $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})$ is consistent, then*

$$T_N(p) \xrightarrow{d} \sum_{l=1}^{p}\int_0^1 B_l^2(x)dx, \tag{4.21}$$

where $B_l(x), x \in [0,1], 1 \le l \le p$, are independent Brownian bridges.

We quickly note that the asymptotic distribution of the test statistic under the null hypothesis is the same as that for the independent case since the long-run variance soaks up the dependence.

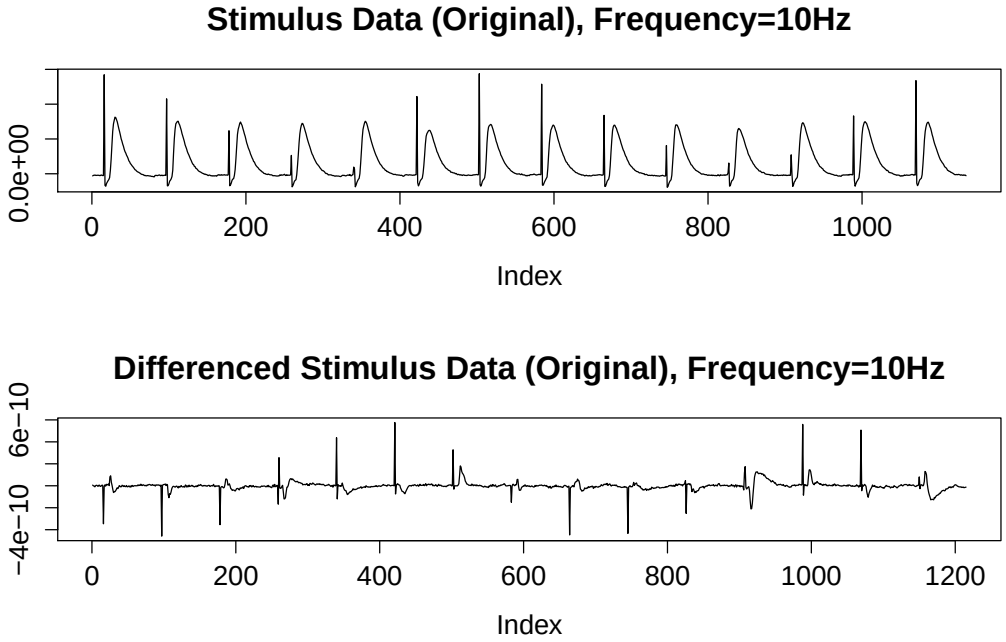
Results to prove the consistency of the test statistic under a single change point alternative and consistency of the kernel estimator for the long run covariance can be found in Horváth and Kokoszka [28]. Hörmann and Kokoszka [25] give more details on the notions of $L_p - approximability$ and results for dealing with dependence in functional observations.

4.1.3 Application to Stimulus Response Data

In this section we report the results of analysing stimulus response data of cells from the animal physiology group at the University of Kaiserslautern. The data is obtained from experiments in which a stimulus is repeatedly given to a cell at a certain speed (1,2,5,10,50Hz) and the reaction of the cell is recorded.

In order to attain stationarity and constant mean, the data is differenced as a first step to some of the analyses. Figure 4.1 shows a subsection of the data (original and its differenced counterpart) series for the frequency 10Hz. The subsection is taken from the first part of the data, curves (responses) 10-24.

Figure 4.1: Original stimulus data



It is clear that the differenced series in Figure 4.1 (lower panel) looks more like a stationary time series as compared to the original series in the top panel of the same figure. At the beginning of each response there is a noticeable sharp spike (circled in red) in Figure 4.2. This is an artifact which represents the direct effect of the stimulus onto the measuring device, but not the response of the cell. The cell reacts to the stimulus after a short delay and therefore it is safe to remove a few data points after the stimulus is given as these are not part of the cells reaction. We therefore cut the data points in the circle and form a vector consisting of the first and last part of the observations. Once the adjustment has been made, the local random noise in the differenced data can be seen more clearly as in Figure 4.4 bottom panel. We also include the adjusted and differenced plots of subsections of the 1,2,5 and 50Hz frequencies in Figures 4.3, 4.4 and 4.5 respectively.

Once the truncation has been done, we have 68, 73, 78, 73, 73 points per curve in the case of 1, 2, 5, 10 and 50Hz frequencies respectively, which are then smoothed to form curves.

Since all the tests include projection of the curves from an infinite dimensional space to the finite dimensional space of the first p principal components, we include a figure of the scree plot which is used in selecting the number of principal components p to be

Figure 4.2: Artifact

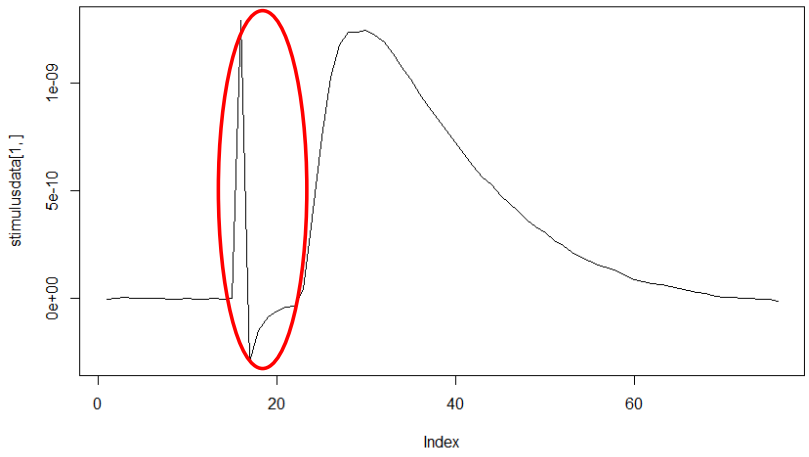


Table 4.1: Test for change in the mean function (i.i.d. Test)
$\alpha = 0.05$, p=4, Asymptotic crit. value=1.239675

	1Hz	2Hz	5Hz	10Hz	50Hz
N	60	120	300	600	3000
Test statistic	2.0872	3.8249	8.5994	54.7244	212.0775

employed. The idea of the scree plot is that we visually select the number of principal components at which the curve dies off. Another method that can be used for this purpose is the cumulative percentage of variance in which we consider the number of principal components for which the cumulative percentage of variance explained is greater than say 85%. In this case the number of principal components selected was $p = 4$ which has a cumulative percentage of variance of approximately 96%. Figure 4.6 gives the scree plot.

Assuming the data is independent and identically distributed, Table 4.1 reports the results obtained. The data used is adjusted to remove the artificial artifact but not differenced.

Against the asymptotic critical value, in all cases we reject the null hypothesis of constant mean. In Table 4.2 we list the change points in order of significance for the 1,2,5 and 10Hz frequencies. These will be used for comparison with the change points in the dependent case. A test for change in mean was then carried out but in this case the data was first differenced. Once the data was differenced, under all frequencies no change point was observed which implies that the data did not have a trend. Note that the change points

Figure 4.3: Adjusted Responses (left) and their Differenced Counterparts 1,2 Hz

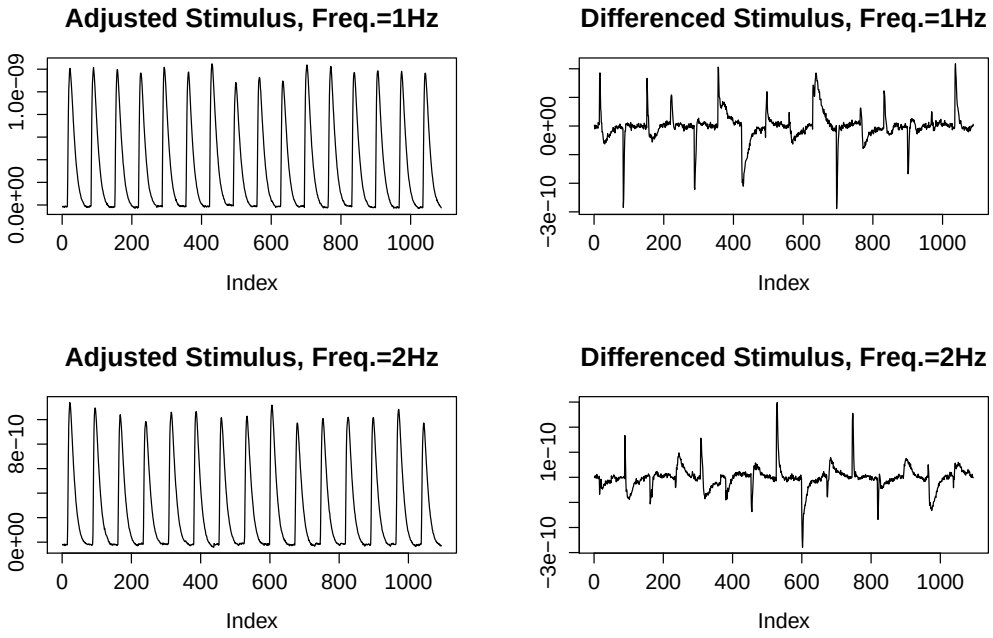


Table 4.2: Change points in order of significance (i.i.d. Test)

Frequency	Change Points						
1Hz	20						
2Hz	70	100					
5Hz	155	85					
10Hz	361	164	62	10	472	396	547

refer to the number of stimulus-response and not to time, i.e. the units on the horizontal axes of the fugures. The figures only show a small part of the total sample of curves.

Since the iid test for the change in mean is known to give wrong results when the data has some dependency, we carried out a Portmanteau test of independence, see Gabrys *et. al.* [21] for details of the test. The main assumptions required are that the fourth moment of the observations exists and that the eigenvalues are in decreasing order. When these assumptions hold then the test statistic converges to a chi-square distribution under the null hypothesis.

The results of the test are given in Table 4.3. The main idea of the Portmanteau test is to test the null hypothesis of independent identical distributions against an open ended alternative of lack of independence or identicality. We test the null hypothesis

H_0: the $X_n(.)$ are independent and identically distributed (i.i.d)

Figure 4.4: Adjusted responses (left) and their differenced counterparts 5, 10 Hz

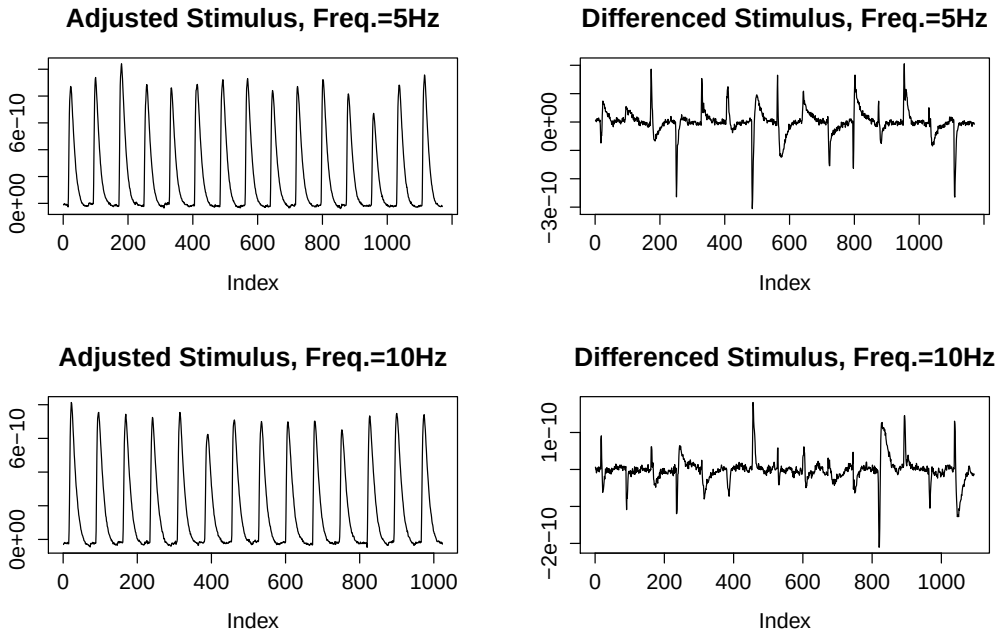


Table 4.3: Portmanteau Test
$\alpha = 0.05$, p=4, Asymptotic crit. value=67.5050

	1Hz	2Hz	5Hz	10Hz	50Hz
N	59	119	299	599	
Test statistic	176.3522	313.8736	334.5219	552.3081	2574.5181

versus

H_A: H_0 does not hold.

The data was differenced as a first step to the analysis. In all cases the assumption of independence is rejected.

Next we investigated whether there was a change in mean while dropping the assumption of independence of the data. In this case we consider that the data are dependent and follow the work of Hörmann and Kokoszka. The results of the test are reported in Table 4.4, where we have a change point in the mean in all cases. As in the iid case, when the

Table 4.4: Test for change in the mean function (Dependent Test)
$\alpha = 0.05$, p=4, Asymptotic crit. value=1.239675

	1Hz	2Hz	5Hz	10Hz	50Hz	
N	60	120	300	600	3000	6000
Test statistic	1.5847	2.0715	3.6859	8.6769	32.6208	

Figure 4.5: Adjusted responses (top) and their differenced counterparts 50 Hz

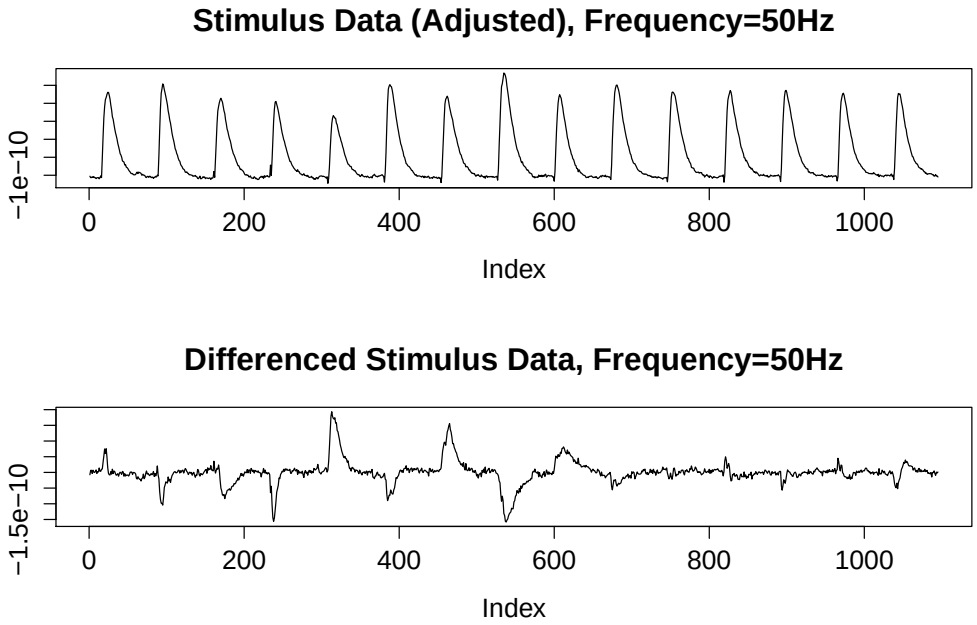


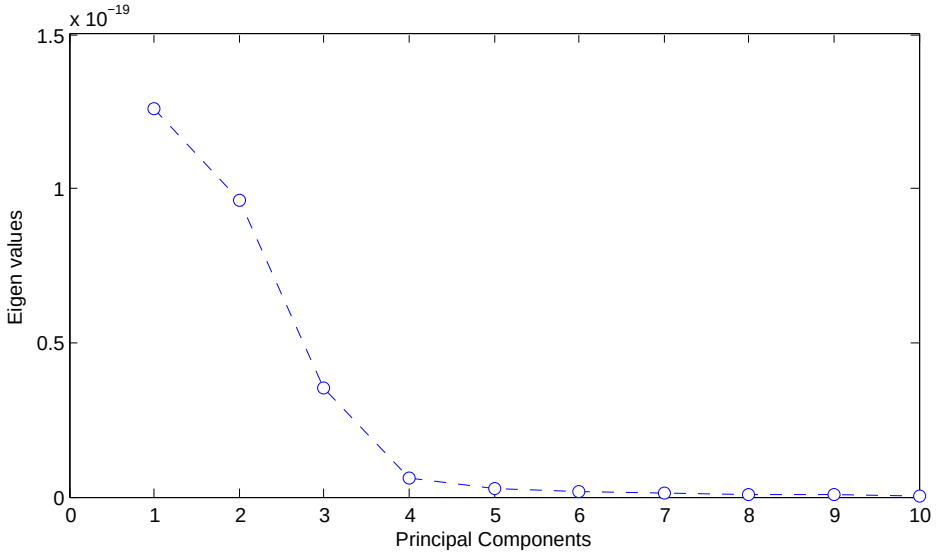
Table 4.5: Change points in order of significance (Dependent Test)

Frequency	Change Points								
1hz	20								
2Hz	74								
5Hz	155								
10Hz	359	163	62	472	389				
50Hz	2067	1213	679	288	182	542	358	987	1081
	1787	1632	1924	2459	2330	2748	2591	2830	

data is differenced then there is no change in mean observed.

In Table 4.5, we give the change points in their order of significance for the dependent test. Comparing the two test we see that the iid test gives many false change points as a result of failure to account for the long run variance. Also, it is noticeable as expected, that with increasing frequency of the stimulus there are more change points. This can be attributed to the fact that at high frequency the cell does not have enough time to go back to its resting state before the next stimulus is given and this is what is observed through the many changes in mean.

Figure 4.6: Scree Plot



4.2 Change Point Detection in the Functional Autoregressive Model (FAR(1))

Our aim is to test (through the linear operator Ψ) whether the stochastic structure of the model changes, so that we need a different model (operator) to represent the data. The first subsection gives a brief introduction of the change point problem for FAR(1) model which is based on the paper by Horváth *et. al.* [26]. Subsection 2 gives an overview of the test procedure while the third subsection deals with the application of the test to the Stimulus response data.

Given the observations $X_n(t),\ t \in [0,1], n = 1, ..., N$, we are concerned with testing the hypothesis

$$H_0 : \Psi_1 = \Psi_2 = \cdots = \Psi_N = \Psi$$

versus

$$H_A : \text{there is } 1 \leq k* < N : \Psi_1 = \cdots = \Psi_{k^*} \neq \Psi_{k^*+1} = \cdots = \Psi_N$$

The structure of the observations under the null hypothesis can be formalised by Assumption 2.1 of Horváth *et. al.* [26], which we include here for completion.

Assumption 4.7. *The functional observations $X_n \in L^2$ satisfy*

$$X_{n+1} = \Psi X_n + \epsilon_{n+1}, \qquad n = 0, 1, \cdots, N-1 \tag{4.22}$$

where Ψ is an integral operator with the kernel $\psi(t,s)$ satisfying

$$\int\int \psi^2(t,s)dsdt < 1 \tag{4.23}$$

and the iid mean zero innovations $\epsilon_n \in L^2$ satisfy

$$\mathbb{E}\left\|\epsilon_n\right\|^4 = \mathbb{E}\left[\int \epsilon_n^2(t)dt\right]^2 < \infty \tag{4.24}$$

4.2.1 Test Procedure

Here we include an overview of the test but complete details can be obtained from Horváth *et al.* [26].

Since the observations are infinite dimensional, they are first projected onto a finite dimensional space before the test is carried out. The test is applied to check whether the action of Ψ on the span of the p most important principal components of the observations changes at some unknown time point i.

Under H_0, $\langle \Psi\nu_j, \nu_l\rangle = \lambda_j^{-1}\langle C\nu_j, \nu_l\rangle$ where $Cx = \mathbb{E}\left[\langle X_n, x\rangle X_{n+1}\right]$ is the lag-1 autocovariance operator. From this representation it is obvious that a change in the operator Ψ can be reflected by a change in $\langle C\nu_j, \nu_l\rangle$, $1 \le j, l \le p$. Looking at the representation $\langle C\nu_j, \nu_l\rangle = \lambda_j \langle \Psi\nu_j, \nu_l\rangle$, we note that a change in $\langle \Psi\nu_j, \nu_l\rangle$ may be obscured by a change in λ_j, thus potentially reducing power, but the test developed is nevertheless effective in practice.

A test against the alternative of a change point involves estimating the above scalar products from observations $X_1, X_2, \cdots, X_k$ and $X_{k+1}, X_{k+2}, \cdots, X_N$ and comparing the estimates. Moreover, defining the $p \times p$ lag-1 autocovariance matrices:

$$\mathbf{R}_k = \frac{1}{k}\sum_{2\le i\le k} \mathbf{X}_{i-1}\mathbf{X}_i^T, \quad \mathbf{R}^*_{N-k} = \frac{1}{N-k}\sum_{k\le i\le N} \mathbf{X}_{i-1}\mathbf{X}_i^T$$

where $\mathbf{X}_i = \left(\langle X_i, \nu_1\rangle, \cdots, \langle X_i, \nu_p\rangle\right)^T$. By the ergodic theorem, as $k \to \infty$,

$$\mathbf{R}_k(j,l) = \frac{1}{k}\sum_{2\le i\le k} \langle X_{i-1}, \nu_j\rangle \langle X_i, \nu_l\rangle \xrightarrow{a.s} \mathbb{E}\langle X_{n-1}, \nu_j\rangle \langle X_n, \nu_l\rangle = \langle C\nu_j, \nu_l\rangle$$

Since the matrices $\mathbf{R}_k$ and $\mathbf{R}^*_{N-k}$ approximate the matrix $[\langle C\nu_j, \nu_l\rangle \;\; j,l = 1,2,\cdots,p]$ based on observations before and after time k, basing the test on their difference is a good choice.

Test Statistic

Define $Y_i(j,l) = \langle X_{i-1}, \nu_j\rangle \langle X_i, \nu_l\rangle$ and $\hat{Y}_i(j,l) = \langle X_{i-1}, \hat{\nu}_j\rangle \langle X_i, \hat{\nu}_l\rangle$ and consider the column vectors of length p^2:

$$\begin{aligned}\mathbf{Y}_i &= [Y_i(1,1),\cdots,Y_i(1,p),Y_i(2,1),\cdots,Y_i(2,p),\cdots,Y_i(p,1),\cdots,Y_i(p,p)]^T;\\ \hat{\mathbf{Y}}_i &= \left[\hat{Y}_i(1,1),\cdots,\hat{Y}_i(1,p),\hat{Y}_i(2,1),\cdots,\hat{Y}_i(2,p),\cdots,\hat{Y}_(p,1),\cdots,\hat{Y}_i(p,p)\right]^T\end{aligned}$$

Since X_i follow a functional AR(1) model, vectors $\mathbf{Y}_i$ form a weakly dependent stationary sequence, and as $k \to \infty$,

$$\sqrt{k}\left[\frac{1}{k}\sum_{2\le i\le k}\mathbf{Y}_i - \mathbb{E}\mathbf{Y}_k\right] \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbf{D}) \tag{4.25}$$

where $\mathbf{D}$ is the $p^2 \times p^2$ long run covariance matrix defined by

$$\mathbf{D} = \mathbb{E}\left[(\mathbf{Y}_0 - \mathbb{E}\mathbf{Y}_0)(\mathbf{Y}_0 - \mathbb{E}\mathbf{Y}_0)^T\right] + 2\sum_{1\le h<\infty}\mathbb{E}\left[(\mathbf{Y}_0 - \mathbb{E}\mathbf{Y}_0)(\mathbf{Y}_h - \mathbb{E}\mathbf{Y}_h)^T\right] \tag{4.26}$$

A quick introduction for the central limit theorem for weakly dependent stationary sequences can be found for instance in Yao *et. al.*, [58], Ibragimov and Linnik, [29]. Define further $\mathbf{Z}_k = \sum_{2\le i\le k}\mathbf{Y}_i$ and $\mathbf{Z}^*_{N-k} = \sum_{k<i\le N}\mathbf{Y}_i$ and their sample counterparts with $\hat{\mathbf{Z}}_k = \sum_{2\le i\le k}\hat{\mathbf{Y}}_i$ and $\hat{\mathbf{Z}}^*_{N-k} = \sum_{k<i\le N}\hat{\mathbf{Y}}_i$.

Now if the autoregressive operator Ψ is constant, then the difference

$$\frac{1}{k}\mathbf{Z}_k - \frac{1}{N-k}\mathbf{Z}^*_{N-k}$$

should be small for all $2 \le k < N$. However this difference may be large due to chance variability if k is close to 2 or N. To circumvent this, a parabolic weight function is employed instead so that the difference is large only when there is a change point. The new representation is as in equation 4.27.

Denoting by $\{\mathbf{W}_D(t), t \geq 0\}$ a p^2-dimensional Brownian motion with covariance matrix $\mathbf{D}$ and from equation 4.25 we have that

$$\mathbf{Z}_k - k\mathbb{E}\mathbf{Y}_N \approx \mathcal{N}(0, k\mathbf{D}) \approx \mathbf{W}_D(k),$$
$$\mathbf{Z}^*_{N-k} - (N-k)\mathbb{E}\mathbf{Y}_N \approx \mathcal{N}(0, (N-k)\mathbf{D}) \approx \mathbf{W}_D(N) - \mathbf{W}_D(k)$$

Under H_0 we have,

$$\frac{1}{k}\mathbf{Z}_k - \frac{1}{N-k}\mathbf{Z}^*_{N-k} \approx \frac{N}{k(N-k)}\left[\mathbf{W}_D(k) - \frac{k}{N}\mathbf{W}_D(N)\right]$$

Denoting

$$\mathbf{U}_N(k) = \frac{k(N-k)}{N}\left(\frac{1}{k}\mathbf{Z}_k - \frac{1}{N-k}\mathbf{Z}^*_{N-k}\right) \tag{4.27}$$

and comparing covariances, we have

$$\frac{1}{N}\left[\mathbf{W}_D(k) - \frac{k}{N}\mathbf{W}_D(N)\right]^T \mathbf{D}^{-1}\frac{1}{N}\left[\mathbf{W}_D(k) - \frac{k}{N}\mathbf{W}_D(N)\right], \quad 1 \leq k \leq N$$

has the same distribution as

$$\sum_{1 \leq m \leq p^2} B^2_m(k/N), \quad 1 \leq k \leq N \tag{4.28}$$

where the $B_m(.)$ are independent standard Brownian bridges on $[0, 1]$. Thus any functional of

$$G_N(k) = \frac{1}{N}\mathbf{U}_N(k)^T\mathbf{D}^{-1}\mathbf{U}_N(k), \quad 1 \leq k \leq N \tag{4.29}$$

can be approximated in distribution by 4.28.

To implement the test, the long run covariance matrix $\mathbf{D}$ is estimated by the Bartlett estimator. The lag h $p^2 \times p^2$ autocovariance matrix for the first k observations is given by

$$\hat{\gamma}_h(k) = \frac{1}{k}\sum_{1 \leq i \leq k-h}\left(\hat{\mathbf{Y}}_i - \frac{1}{k}\sum_{1 \leq i \leq k}\hat{\mathbf{Y}}_i\right)\left(\hat{\mathbf{Y}}_{i+h} - \frac{1}{k}\sum_{1 \leq i \leq k}\hat{\mathbf{Y}}_i\right)^T$$

and its corresponding Bartlett estimator is

$$\hat{\mathbf{D}}_k = \sum_{|h| \leq q}\left(1 - \frac{h}{q+1}\right)\hat{\gamma}_h(k) \tag{4.30}$$

The same steps are employed to obtain the Bartlett estimator for the last $N-k$ observations. The sequence $G_N(k)$ is approximated by its sample counterpart

$$\frac{1}{N}\hat{\mathbf{U}}_N(k)^T\left[\frac{k}{N}\hat{\mathbf{D}}_k+\left(1-\frac{k}{N}\right)\hat{\mathbf{D}}^*_{N-k}\right]\hat{\mathbf{U}}_N(k),\quad 1\leq k\leq N \tag{4.31}$$

For the test statistic, a Cramér-von-Mises type functional $K_{p^2}:=\int_0^1\sum_{1\leq m\leq p^2}B_m^2(u)du, u\in[0,1]$ is employed, although asymptotic theory exists for maximal selected statistics and weighted sums. Defining the critical value $c(\alpha,p^2)$ by $P(K_{p^2}>c(\alpha,p^2))=\alpha$, and

$$\hat{I}_N=\frac{1}{N}\sum_{k=1}^{N}\hat{G}_N(k), \tag{4.32}$$

the test rejects if $\hat{I}_N>c(\alpha,p^2)$.

We quickly note that the constancy or lack thereof of $\langle\Psi\nu_j,\nu_l\rangle$ is investigated for the first p principal components after which any change that may occur is assumed to be negligible since the amount of variability explained by the first p principal components is assumed to be sufficiently large.

4.2.2 Application to Stimulus Response Data

In order to attain stationarity and constant mean, the data is differenced as a first step to the analysis. Once we have evidence that the data are not i.i.d (Portmanteau test above), we test for the nullity of the operator Ψ, (goodness of fit test). In this case we acknowledge that although the data are dependent, the FAR(1) model (or a linear representation) may not be a good way to represent the data. This is an extension of the work of Kokoszka *et. al.*, [32] and a brief overview together with the test statistic can be obtained from Section 4.2 of this thesis. We quickly note that for change point detection, although the FAR(1) model may not be the best to represent the functional time series, this does not matter since a change in a FAR(1) model fitted to the data still signifies a change in the dependence structure of the functional time series even if it is not well approximated by a FAR(1) scheme.

We test the hypothesis

$$H_0:\Psi=0\quad versus\quad H_A:\Psi\neq 0.$$

Table 4.6: Goodness of Fit
$\alpha = 0.05$, p=4, Asymptotic crit. value=26.296

	1Hz	2Hz	5Hz	10Hz	50Hz
N	59	119	299	599	2999
Test statistic	76.0413	147.4256	255.1255	509.0306	2520.8195

Table 4.7: Change Point
$\alpha = 0.05$, p=4, d=pˆ2=16, Asymptotic crit. value=3.740248

	1Hz	2Hz	5Hz	10Hz	50Hz
N	59	119	299	599	2999
Test statistic	-12.5624	-9.6	-30.6264	2.8461	11.7638

In this case we check whether the relationship between the curves is linear, and once we have a confirmation that $\Psi \neq 0$ we fit a functional autoregressive model of order one (FAR(1)) to the data and proceed to investigate whether there is a change point. For each stimulus frequency we have a different data set. The results of the goodness of fit test are given in Table 4.6. Note that the value of the test statistic grows with N as in the Portmanteau test of independence.

Once we establish that a linear relationship exists between the observations, we test for a change in the linear operator Ψ. The results are recorded in Table 4.7. From Table 4.7, we see that for the 50Hz frequency a change point exists. This implies that modelling the data with the same linear operator Ψ, will give erroneous results. In this case the change point can be estimated and the data modelled in segments with a uniform parameter within and different parameters between.

In all cases here we used the asymptotic critical values which are justified since the samples are quite large. However, in practice, frequently, there are small samples, in which case asymptotic values are not quite representative. In such cases then the bootstrap would be useful as it would provide better approximations of the distribution of the test statistics. We use this as a motivation for studying bootstrap techniques for functional data in the rest of the thesis.

Chapter 5

Bootstrap for FAR - Preliminaries

In this chapter the bootstrap methodology is introduced. A general description is given in the first two sections. Then, we consider a simple toy example of a FAR(1) process which allows to prove that the residual-based or naive bootstrap holds for the estimate of the autoregressive operator.

5.1 The bootstrap procedure

The bootstrap is a computer intensive resampling method which can be employed as a cheaper alternative to obtain the asymptotics of a given estimator especially when there exists no theoretical formula or when the theoretical formula exists but is highly complex or in the event that the sample available is too small and therefore it is impossible to obtain reasonable approximations of asymptotics. Among its advantages is that it can be employed even in the case of small or moderate sample size as is frequently the case in practice, with reasonable results. Highly desirable also is its ability to give satisfactory results under minimum assumptions, if any, as opposed to normal asymptotic approximations which require that certain assumptions be specified and fulfilled, failure of which results in wrong or suboptimal conclusions.

The idea of the bootstrap revolves around generating a pseudosample from the existing sample in such a way that the pseudosample shares the characteristics or features of the original sample. That is necessary in order to lead to a consistent bootstrap estimator

of some distributional characteristics of some functions of the data, for example of an estimate $\hat{\theta}$ of a parameter θ. Given a sample $X_1, ..., X_n$ with joint distribution F_n, we construct an estimate F_n^* which should be as close as possible to F_n and from it generate a bootstrap sample $X_1^*, ..., X_n^*$. Once we have the bootstrap sample, we compute a bootstrap estimate $\hat{\theta}^*$ in the same way we compute the original estimator $\hat{\theta}$, only this time we use the bootstrap sample. We resample from the original sample a large number of times, each time computing the bootstrap estimator. From the resulting Monte Carlo sample of realizations of $\hat{\theta}^*$, we may approximate distributional characteristics of $\hat{\theta}^*$ by averaging. If the bootstrap is valid, they provide approximations of the corresponding characteristics of the distribution of $\hat{\theta}$.

It should quickly be noted that not only the data generation process is at stake here but also the statistic or parameter of interest. This means that if in order to estimate the statistic of interest we need only the first two moments which completely define the asymptotic distribution, then it is sufficient to generate a bootstrap pseudosample that is similar to the underlying process up to the second moment. An example of this is if we are interested in proving the Central Limit Theorem for means, in this case, asymptotically we need the mean and covariance of the process to fully describe the asymptotic normal distribution. This is advantageous since in the dependent data case the relationships between the variables of the underlying process can be very complex and therefore almost impossible or very expensive to mimic [33]. The disadvantage obviously is that if the underlying process is not fully mimicked the results may not be as good as in the case when it is fully mimicked, unless the estimator being estimated is robust to this situation.

The bootstrap can be implemented in different ways. The unknown distribution function can be estimated using the parametric bootstrap or the empirical (nonparametric) bootstrap depending on how much information we have about the underlying sample. When the distribution of the underlying sample is known, the parametric bootstrap is appropriate as it is best suited to give the most fitting results. Employing the nonparametric bootstrap in the case of a known parametric model that generates the underlying data can only work in the event that the asymptotics of the (parametric) model are

not sensitive to a change in model which is rare. On the other hand, if the underlying parametric model is not known, then the nonparametric bootstrap is highly advisable. Paparoditis and Sapatinas [46] propose a novel bootstrap-based functional testing procedure which is applicable for different testing problems, different test statistics and more than two populations. The main idea behind their method is that the pseudo sampling is done in such a way as to satisfy the null hypothesis of interest. They applied their method to test statistics proposed in literature for comparison of means and covariance functions. It is evident from their simulations that the bootstrap is robust as compared to asymptotic approximations to departures from Gaussianity. Also since no theoretical approximations are required for the bootstrap, Berkes *et. al.* [4] employ it to assess the empirical size and power of their test for changes in means of functional observations. This is especially advantageous since in practical settings the distribution of the random variables is normally unknown.

Depending on the nature of the underlying data, several bootstrap procedures exist. We have bootstrap procedures for i.i.d data and those for data with dependence. In all cases it is highly important to mimic the stochastic properties of the underlying data set. Since we are dealing with time series data, we focus mainly on those methods that maintain the dependence structure. These include the nonparametric naive or residual-based bootstrap (algorithm below), the block bootstrap which has several variations, see for instance Lahiri, [36], who gives the description of several bootstrap resampling methods including the different forms of the block bootstrap for scalar observations, Künsch, [35], who extended the bootstrap method of estimating standard errors to the case of stationary observations in the scalar case, the autoregressive (AR)-sieve bootstrap, the Markovian bootstrap, the wild bootstrap among others. These majorly are found in the time domain but it should be noted that bootstrap methods for the frequency domain such as the local and hybrid bootstrap procedures exist. We focus on bootstrap methods in the time domain.

Considering the functional case, Politis and Romano, [47] were the first to develop weak convergence results for approximate sums of weakly dependent variables in the Hilbert space and in a triangular array setting. They applied their results to estimators obtained

under the stationary bootstrap resampling method and in particular prove a bootstrap central limit theorem for the stationary bootstrap. We discuss their approach a bit more in detail in section 2. Dehling *et. al.*, [15], develop weak convergence results for Hilbert space valued random variables. The random variables are assumed to be weakly dependent in the sense of near epoch dependence. They also show that the nonoverlapping block bootstrap is consistent. In terms of application of the bootstrap in the functional case, we have for instance De Castro *et. al.* [12] who employ the bootstrap for predicting Sulfur Dioxide levels, Aston and Kirch [2] in checking for stationarity in resting state fMRI data, Besse *et. al.* among others.

Since its introduction by Efron [17], the bootstrap methodology has been documented extensively. [54], [56], [40] give introductory notions on the bootstrap methodology and some applications. Several papers have also been devoted to the topic with some giving reviews of the topic for instance the introductory paper by [17], reviews by [9] and [33], [34] among others.

As summarised in Cuevas *et. al.*, [13], the goal is to show that the distribution of $a_n(T(F_n) - T(F))$ is close to that of its corresponding bootstrap version $a_n(T(F_n^*) - T(F_n))$, where a_n is an increasing sequence tending to infinity, F is the unknown population distribution function, F_n is the sample distribution function computed from the sample, F_n^* is based on the bootstrap pseudosample and therefore the distribution of the centered and scaled bootstrap estimator consistently estimates the corresponding distribution of the true estimator. The increasing sequence a_n of nonnegative real numbers is chosen so that the sequence of distributions converges to a nondegenerate limit. This can be done by showing that a suitable distance between the sampling distributions of both sequences tends almost surely to zero or that both sequences converge weakly to the same limit distribution asymptotically. The bootstrap technique, as used in the latter context provides an approximate distribution of the unknown underlying distribution. This approximate distribution is required when checking the accuracy of the empirical estimator as compared to the true parameter, through for instance, confidence intervals and also for constructing tests of hypotheses.

5.1.1 Nonparametric residual (Naive) bootstrap

As mentioned in the previous section, the bootstrap was originally developed for i.i.d observations. However, Singh, [53] in his seminal paper showed that in the event that the data had some dependence then the i.i.d. bootstrap would fail. He proved this for the case where the parameter of interest was the sample mean and showed that even though both the bootstrap and empirical distributions converge asymptotically to a normal distibution, they do so with different asymptotic variances. This led to a need for developing methods that could take into account the dependence structure in the data.

This method of resampling tries to retain the original bootstrap idea of resampling from i.i.d. random variables. Since the data has some dependence, in this method resampling is done of the residuals obtained from an optimal predictor of the observations. For a given sample $(X_0, X_1, ..., X_N)$, the standard residual bootstrap algorithm is given by Algorithm 1 where resampling is done for the centered residuals. We quickly note that lack of centering of residuals may introduce a bias which is random, in which case the bootstrap would fail. An example of the same problem for the regression scenario can be found for instance in Freedman, [19].

For the algorithm, we need an estimate $\hat{\Psi}$ of the operator Ψ, which, of course has to satisfy some assumptions for the bootstrap to work. Here, as usual, pr^* denotes the conditional probability given the original data $X_0, X_1, ..., X_N$. The naive bootstrap works quite well if the model assumptions hold true for the data. If the model is only an approximation, or if no good model is known in advance at all, model-free resampling is advisable as discussed in the following section.

Algorithm 1 Naive Bootstrap Algorithm

1. $\hat{\epsilon}_{n+1} = X_{n+1} - \hat{\Psi}(X_n), n = 0, ..., N-1$, $\tilde{\epsilon}_n = \hat{\epsilon}_n - \frac{1}{N}\sum_{s=1}^{N} \hat{\epsilon}_s$.
2. Draw randomly from $\{\tilde{\epsilon}_1, ..., \tilde{\epsilon}_N\}$ new bootstrap residuals $\epsilon_1^*, ..., \epsilon_N^*$ with probability $pr^*(\epsilon_n^* = \tilde{\epsilon}_k) = \frac{1}{N}$,$k = 1, ..., N$.
3. Generate bootstrap data $X_{n+1}^* = \hat{\Psi}(X_n^*) + \epsilon_{n+1}^*$, $X_0^* = X_0$.
4. Calculate $\hat{\Psi}^*$ from $X_0^*, ..., X_N^*$ in the same manner as $\hat{\Psi}$ from $X_0, ..., X_N$.

5.1.2 Block bootstrap

When Efron [17] introduced the bootstrap it was mainly focused on independent and identically distributed observations. The bootstrap provides under relatively few conditions approximations to distributions of statistics and confidence intervals which for sample sizes available in practice are more accurate than those from first-order asymptotic theory, see [24]. In this case resampling is done randomly with replacement from the empirical distribution or from a parametric model which includes the distribution of the data.

In practice however, many data have some form of dependence which should be taken into consideration when doing any form of estimation. The dependence structure of the observations cannot be ignored without adverse effects in most cases. In order to overcome this challenge, block bootstrap methods were introduced and they consider the dependence structure between the observations when resampling, see for instance Künsch, [35], Carlstein, [11] who were among the first to introduce block bootstrap methods and the monograph by Lahiri, [36] who gives an insightful description of the different bootstrapping methods, their advantages and disadvantages when applied to different situations. Several other monographs exist together with review papers some of which were already mentioned in the previous section. In the functional data situation, Politis and Romano, [47] prove the asymptotic validity of the stationary resampling algorithm while Dehling *et. al.* show the consistency of the non overlapping block bootstrap.

There are several types of block bootstrap methods which can mainly be differentiated on whether the blocks are overlapping or not and also on whether the block length is fixed or increasing. The overall idea however involves partitioning the sample into blocks and then resampling independently from these blocks. The idea is that after a certain lag, the observations are almost independent and therefore the observations in a block retain the dependency while the different blocks are almost independent of each other. A slightly different but almost similar idea is the block of blocks bootstrap. Here, the sample is split into blocks first, which may overlap, and then subblocks of those blocks are drawn randomly. From [24] we have that the finite sample performance of the block of blocks

bootstrap is better (more accurate) as compared to that of the normal block bootstrap in certain situations. Once the blocks have been resampled they are ordered from end to end to form a new sample which should retain some if not all of the characteristics of the underlying distribution.

It should also be noted from [24] that although there exist several methods to tackle the problem of dependence of observations, the relative accuracy of the block bootstrap as well as first-order approximations is poorer as compared to i.i.d data. Estimation errors are also larger for the block bootstrap since partitioning the sample into blocks before resampling the blocks independently distorts the dependence structure of the data. Another reason is that partitioning into blocks has the effect of reducing the effective sample size and therefore increasing sample variation, see [24]. As noted in [24], choice of block length affects the accuracy of the bootstrap, therefore care should be taken in order to ensure accurate results.

The algorithm for the non overlapping block bootstrap is given below. It should be noted

Algorithm 2 Block Bootstrap Algorithm

1. Given the sample $X_1, \cdots, X_n$, using a suitable rule choose the fixed block length(l) and partition the sample into B blocks each of the given length l.
2. Draw randomly with replacement from the resulting blocks and order the drawn blocks from end to end to form a new bootstrap sample $X_1^*, \cdots, X_n^*$.
3. Compute the statistic of interest in the same way as in the true case.

that although overlapping block bootstrap is preferred in practice, the numerical results from overlapping and nonoverlapping block bootstrap are very close, provided the overlap is short compared to the block length, and theoretical arguments are simpler with the nonoverlapping block bootstrap [33]. Since we assume that our data has the dependence structure of a functional autoregressive model, we carry out further theoretical investigation and simulation considering the naive bootstrap.

5.2 Triangular arrays of Hilbert-space valued random variables

Here we summarise results from the technical report by Politis and Romano [47] which we use in the next chapter for studying the asymptotics of the error of estimating Ψ. The authors develop convergence results for approximate sums of weakly dependent stationary Hilbert space valued random variables in a triangular array setting. They consider random variables in a separable Hilbert space with inner product $\langle .,. \rangle$ and norm $\|.\|$. We consider Theorems 2.1 and 4.2 of this technical report which are given below.

Theorem 5.2.1. *Let $X_{n,1}, ..., X_{n,n}$ be H-valued, stationary, mean zero random variables such that $\mathbf{E}\left(\|X_{n,i}\|^2\right) < \infty$. Assume, for any integer $k \geq 1$, $(X_{n,1}, ..., X_{n,k})$, regarded as a random element of H^k, converges in distribution to say $(X_1, ..., X_k)$. Moreover, assume, $\mathbf{E}\left[\langle X_{n,1}, X_{n,k} \rangle\right] \to \mathbf{E}\left[\langle X_1, X_k \rangle\right]$ as $n \to \infty$ and*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{E}\left(\langle X_{n,1}, X_{n,k} \rangle\right) \to \sum_{k=1}^{\infty} \mathbf{E}\left(\langle X_1, X_k \rangle\right) < \infty. \tag{5.1}$$

Let $Z_n = n^{-1/2} \sum_{i=1}^{n} X_{n,i}$. Then, Z_n is weakly compact.

Theorem 4.2 of the same technical report shows that this kind of asymptotics may be used to show that the bootstrap works for independent data.

Theorem 5.2.2. *Suppose $X_1, X_2, ...$ are independent and identically distributed H˙valued random variables with common distribution μ such that $\mathbf{E}\|X_1\|^2 < \infty$. Conditional on $X_1, ..., X_n$, let $X_1^*, ..., X_n^*$ be independent and identically distributed according to $\hat{\mu}_n$, where $\hat{\mu}_n$ is the empirical measure: $\hat{\mu}_n(E) = n^{-1} \sum_{i=1}^{n} 1\left(X_i \in E\right)$. Let*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i/n \text{ and } \bar{X}_n^* = \frac{1}{n} \sum_{i=1}^{n} X_i^* \tag{5.2}$$

Then, along almost all sample sequences $X_1, X_2, ...$, given $(X_1, ..., X_n)$, the conditional distribution of $n^{1/2}\left(\bar{X}_n^* - \bar{X}_n\right)$ converges weakly to the normal distribution on $\mathcal{H}$ having mean 0 and covariance operator $\mathcal{S}$, where $\mathcal{S}$ is the covariance operator of $X - m$ when X has distribution μ.

Unfortunately, the technique applied for proving this result may not be extended to dependent data. We may show that the quantities of interest converge in distribution in the real and in the bootstrap world, but it is not possible to prove that the asymptotic

distributions coincide. The main problem is that mixing properties of $\{X_t\}$ which guarantee asymptotic normality in real world do not necessarily hold for $\{X_t^*\}$.

5.3 A first look at the bootstrap theory

Our aim, as already stated, is to prove that the distribution of the bootstrap estimation error provides a consistent estimator of the estimation error for the real data. The key to that result is that the distribution of the innovations ϵ_t is well approximated by the distribution of the bootstrap innovations. In the next section 5.4 we prove such a result for a special case. In this section, however, we discuss a toy example where we unrealistically assume that the eigenfunctions of Γ (and C) are known. This reduces the functional setting essentially to a scalar one (for the individual eigenvalues) and circumvents a fundamental difficulty of bootstrapping FAR data. On the other hand, some of the basic features of the FAR bootstrap problem can already be studied at that simple example.

We start with describing the special case to be considered in the next section 5.4 as it is the basis for the toy example, too. We study a FAR(1) process with covariance and lag 1 autocovariance operators Γ and C respectively. Let $\lambda_j,\ \nu_j$ denote the eigenvalues and eigenvectors of the symmetric, positive definite operator Γ. We assume that C is also symmetric, compact with the singular value decomposition

$$\textbf{S:} \qquad C(x) = \sum_{j=1}^{\infty} \gamma_j \langle x, \nu_j \rangle \nu_j \tag{5.3}$$

i.e. it has the same eigenbasis as Γ!

In general, $\Gamma,\ C$ may be estimated by the sample versions

$$\hat{\Gamma}_n(x) = \frac{1}{n} \sum_{t=0}^{n-1} \left\langle X_t - \bar{X}_n, x \right\rangle \left(X_t - \bar{X}_n \right)$$

and

$$\hat{C}_n(x) = \frac{1}{n} \sum_{t=1}^{n} \left\langle X_{t-1} - \bar{X}_n, x \right\rangle \left(X_t - \bar{X}_n \right) \tag{5.4}$$

where $\bar{X}_n = \frac{1}{n}\sum_{t=0}^{n-1} X_t$. We assume $X_0, \cdots, X_n$ to be given, but we use X_n only in $\hat{C}_n(x)$ to simplify notation.

Assuming $\mathbb{E}X_t = 0$, we have $\Gamma = \mathbb{E}X_t \otimes X_t,\ C = \mathbb{E}X_t \otimes X_{t+1}$,

$$\begin{aligned}\lambda_j &= \lambda_j||\nu_j||^2 = <\lambda_j\nu_j, \nu_j> = <\Gamma(\nu_j), \nu_j> \\ &= <\mathbb{E}X_t \otimes X_t(\nu_j), \nu_j> = \mathbb{E}\langle <X_t, \nu_j> X_t, \nu_j\rangle = \mathbb{E} <X_t, \nu_j>^2\end{aligned}$$

and analogously

$$\gamma_j = \mathbb{E}(<X_t, \nu_j><X_{t+1}, \nu_j>).$$

Under suitable assumptions, $\sqrt{n}\left(\hat{C}_n - C\right)$ and $\sqrt{n}\left(\hat{\Gamma}_n - \Gamma\right)$ are asymptotically Gaussian, see Theorem 3 of Mas and Pumo, [42]. Let $\hat{\lambda}_j,\ \hat{\gamma}_j$ denote the eigenvalues of $\hat{\Gamma}_n$ respectively $\hat{C}_n$. From the asymptotic normality of $\hat{\Gamma}_n,\ \hat{C}_n$, we get a corresponding asymptotic normality of the eigenvalue estimates from Theorem 2.2 of Mas and Menneteau, [43], provided that the multiplicity of the eigenvalues is 1. By their Remark 2.1, the same rate of convergence also holds for eigenvalues with larger multiplicity.

5.3.1 A toy example

For a simpler presentation of the arguments, we make an additional restrictive assumption like, e.g., in section 2 of Guillas, [22]:

T.1: The common eigenvectors ν_j of Γ respectively C are known.

Moreover, just to simplify notation, we also assume

T.2: $\mathbb{E}X_t = 0$ is known.

Then, we can estimate λ_j, γ_j by

$$\hat{\lambda}_j = \frac{1}{n}\sum_{t=1}^{n} <X_{t-1}, \nu_j>^2, \quad \hat{\gamma}_j = \frac{1}{n}\sum_{t=1}^{n} <X_{t-1}, \nu_j><X_t, \nu_j>,$$

and, similar as in [22], we set in this subsection

$$\hat{C}_n = \sum_{j=1}^{k_n} \hat{\gamma}_j \nu_j \otimes \nu_j, \quad \hat{\Gamma}_n = \sum_{j=1}^{k_n} \hat{\lambda}_j \nu_j \otimes \nu_j.$$

We denote as usual by $\hat{\Gamma}_n^\dagger$ the inverse of $\hat{\Gamma}_n$ on the linear hull V_{k_n} of $\nu_1, \ldots, \nu_{k_n}$, i.e.

$$\hat{\Gamma}_n^\dagger = \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j} \nu_j \otimes \nu_j,$$

and, finally, we estimate Ψ by

$$\hat{\Psi}_n = \hat{C}_n \hat{\Gamma}_n^\dagger.$$

Let Π_{k_n} denote the projection onto V_{k_n}, and

$$\Gamma_n = \sum_{j=1}^{k_n} \lambda_j \nu_j \otimes \nu_j, \quad \Gamma_n^\dagger = \sum_{j=1}^{k_n} \frac{1}{\lambda_j} \nu_j \otimes \nu_j.$$

Lemma 7. $\Pi_{k_n} = \hat{\Gamma}_n \hat{\Gamma}_n^\dagger = \hat{\Gamma}_n^\dagger \hat{\Gamma}_n, \quad \hat{\Psi}_n \Pi_{k_n} = \hat{\Psi}_n$

Proof.

$$\hat{\Gamma}_n \hat{\Gamma}_n^\dagger(x) = \sum_{j=1}^{k_n} \hat{\lambda}_j < \hat{\Gamma}_n^\dagger(x), \nu_j > \nu_j = \sum_{j=1}^{k_n} \hat{\lambda}_j < x, \hat{\Gamma}_n^\dagger(\nu_j) > \nu_j = \sum_{j=1}^{k_n} \frac{\hat{\lambda}_j}{\hat{\lambda}_j} < x, \nu_j > \nu_j = \Pi_{k_n}(x),$$

using the selfadjointness of $\hat{\Gamma}_n^\dagger$. $\Pi_{k_n} = \hat{\Gamma}_n^\dagger \hat{\Gamma}_n$ follows analogously. Finally,

$$\begin{aligned} \hat{\Psi}_n &= \hat{C}_n \hat{\Gamma}_n^\dagger(x) = \hat{C}_n \left(\sum_{j=1}^{k_n} < x, \nu_j > \frac{1}{\hat{\lambda}_j} \nu_j \right) = \sum_{j=1}^{k_n} \frac{\hat{\gamma}_j}{\hat{\lambda}_j} < x, \nu_j > \nu_j \\ \hat{\Psi}_n \Pi_{k_n} &= \hat{\Psi}_n \left(\sum_{j=1}^{k_n} < x, \nu_j > \nu_j \right) = \sum_{j=1}^{k_n} < x, \nu_j > \hat{\Psi}_n(\nu_j) = \sum_{j=1}^{k_n} < x, \nu_j > \frac{\hat{\gamma}_j}{\hat{\lambda}_j} \nu_j. \end{aligned}$$

□

We consider the coefficients of the projections of the innovations ϵ_t and of their sample analogues $\hat{\epsilon}_t = X_t - \hat{\Psi}_n(X_{t-1})$ onto ν_m for some fixed m,

$$\eta_t = \langle \epsilon_t, \nu_m \rangle, \ \hat{\eta}_t = \langle \hat{\epsilon}_t, \nu_m \rangle$$

Then,

$$\begin{aligned} Z_t &= \langle X_t, \nu_m \rangle = \langle \Psi(X_{t-1}) + \epsilon_t, \nu_m \rangle = \langle \Psi(X_{t-1}), \nu_m \rangle + \eta_t \\ \hat{Z}_t &= \left\langle \hat{\Psi}_n(X_{t-1}) + \hat{\epsilon}_t, \nu_m \right\rangle = \left\langle \hat{\Psi}_n(X_{t-1}), \nu_m \right\rangle + \hat{\eta}_t \end{aligned}$$

As $C = \Psi\Gamma$, we have

$$\gamma_m \nu_m = C(\nu_m) = \Psi(\Gamma(\nu_m)) = \lambda_m \Psi(\nu_m) \tag{5.5}$$

Therefore, as $\lambda_j > 0$ for all j, setting $\alpha_m = \gamma_m/\lambda_m$,

$$\begin{aligned} \langle \Psi(X_{t-1}), \nu_m \rangle &= \left\langle \Psi\left(\sum_{j=1}^{\infty} \langle X_{t-1}, \nu_j \rangle \nu_j \right), \nu_m \right\rangle \\ &= \sum_{j=1}^{\infty} \langle X_{t-1}, \nu_j \rangle \langle \Psi(\nu_j), \nu_m \rangle \\ &= \alpha_m \langle X_{t-1}, \nu_m \rangle \end{aligned} \tag{5.6}$$

Analogously, assuming $k_n \geq m$,

$$\begin{aligned} \hat{\Psi}_n(\nu_m) &= \hat{C}_n\left(\hat{\Gamma}_n^{\dagger}(\nu_m)\right) = \hat{C}_n\left(\frac{1}{\hat{\lambda}_m}\nu_m\right) \\ &= \frac{1}{\hat{\lambda}_m}\hat{\gamma}_m \nu_m = \hat{\alpha}_m \nu_m \end{aligned} \tag{5.7}$$

and

$$\left\langle \hat{\Psi}_n(X_{t-1}), \nu_m \right\rangle = \hat{\alpha}_m \langle X_{t-1}, \nu_m \rangle$$

Hence, we have

$$\eta_t = Z_t - \alpha_m Z_{t-1}, \quad \hat{\eta}_t = Z_t - \hat{\alpha}_m Z_{t-1}$$

where η_t are i.i.d. zero mean real random variables, and Z_t corresponds to a stationary AR(1) process as we have $|\alpha_m| < 1$, as otherwise, $\|\Psi(\nu_m)\| \geq 1$ in contradiction to our general assumption $\|\Psi\|_{\mathcal{L}} < 1$.

Let $\tilde{\eta}_t = \hat{\eta}_t - \frac{1}{n}\sum_{k=1}^{n} \hat{\eta}_k$ be the centered sample residuals. Then we can apply the same arguments as in Theorem 3.1 of Kreiss and Franke [20] to get that the distribution F of η_t and the empirical distribution $\hat{F}_n$ of $\hat{\eta}_1, \cdots, \hat{\eta}_n$ converge to each other in the Mallows metric d_2 (compare, e.g., Bickel and Freedman [6]).

Proposition 5.3.1. *Let $X_0, \cdots, X_n$ be a sample from a stationary FAR (1)-process satisfying **T.1** and **T.2** and*

i) $\mathbb{E}\|\epsilon_t\|^4 < \infty$, $\mathbb{E}\epsilon_t = 0$, ϵ_t *i.i.d.*

ii) $\|\Psi\|_{\mathcal{L}} < 1$

iii) C has the singular value decomposition $\boldsymbol{S}$ with the eigenfunctions ν_j of Γ.

Then, $d_2\left(\hat{F}_n, F\right) \xrightarrow{p} 0$ for $n \to \infty$.

Proof. First, we note that from i), ii) and Theorem 3 of Mas and Pumo, [42], the asymptotic normality of $\sqrt{n}\left(\hat{\Gamma}_n - \Gamma\right)$ and $\sqrt{n}\left(\hat{C}_n - C\right)$ follows, as exhibited in Example 16.1 of Horváth and Kokoszka, [28], the FAR(1)-process satisfies the assumptions of said theorem. From Mas and Menneteau, [43], Theorem 2.2, we then get $\sqrt{n}\left(\tilde{\gamma}_m - \gamma_m\right)$ and $\sqrt{n}\left(\tilde{\lambda}_m - \lambda_m\right)$ are asymptotically normal, where here $\tilde{\lambda}_m, \tilde{\gamma}_m$ denote the eigenvalues of the usual nontruncated sample covariance and autocovariance operators $\hat{\Gamma}_n, \hat{C}_n$. They coincide with $\hat{\lambda}_m, \hat{\gamma}_m$ from above as

$$\tilde{\lambda}_m =< \hat{\Gamma}_n(\nu_m), \nu_m >=< \frac{1}{n}\sum_{t=1}^{n} X_{t-1} \otimes X_{t-1}(\nu_m), \nu_m >= \frac{1}{n}\sum_{t=1}^{n} < X_{t-1}, \nu_m >^2 = \hat{\lambda}_m$$

and analogously for $\tilde{\gamma}_m = \hat{\gamma}_m$.

By Slutsky's Lemma we also have,

$$\sqrt{n}\left(\hat{\alpha} - \alpha\right) = \frac{\sqrt{n}\left(\hat{\gamma}_m - \gamma_m\right)}{\hat{\lambda}_m} - \frac{\gamma_m \sqrt{n}\left(\hat{\lambda}_m - \lambda_m\right)}{\hat{\lambda}_m \lambda_m}$$

is asymptotically normal too, and in particular

$$\sqrt{n}\left(\hat{\alpha}_n - \alpha\right) = \mathcal{O}_p(1) \tag{5.8}$$

Let F_n denote the empirical distribution function of $\eta_1, \cdots, \eta_n$. From Bickel and Freedman, [6], Lemma 8.4, we have

$$d_2\left(F_n, F\right) \xrightarrow{a.s.} 0 \tag{5.9}$$

Next, let J be Laplace distributed on $1, \cdots, n$, i.e. $pr\left(J = t\right) = \frac{1}{n},\ 1 \le t \le n$. We define random variables

$$U_0 = \eta_J,\ V_0 = \hat{\eta}_J - \frac{1}{n}\sum_{k=1}^{n} \hat{\eta}_j$$

with distribution F_n respectively $\hat{F}_n$. By definition of the Mallows metric

$$\begin{aligned} d_2^2\left(F_n, \hat{F}_n\right) &= \inf_{U,V} \mathbb{E}\left(U - V\right)^2 \le \mathbb{E}\left(U_0 - V_0\right)^2 && (5.10)\\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\hat{\eta}_j - \eta_j - \frac{1}{n}\sum_{k=1}^{n}\hat{\eta}_k\right)^2 && (5.11)\\ &\le \frac{6}{n}\sum_{j=1}^{n}\left(\hat{\eta}_j - \eta_j\right)^2 + \frac{3}{n^2}\left(\sum_{k=1}^{n}\eta_k\right)^2 && (5.12) \end{aligned}$$

As $\eta_t = \langle \epsilon_t, \nu_m \rangle$ are i.i.d. with finite variance and $\mathbb{E}\eta_t = 0$, the second term on the right hand side is of order $\frac{1}{n}$ by the central limit theorem. Therefore, plugging in $\eta_j = Z_j - \alpha Z_{j-1}$ and $\hat{\eta}_j = Z_j - \hat{\alpha}_n Z_{j-1}$, we have

$$d_2^2\left(F_n, \hat{F}_n\right) \leq \frac{6}{n}\sum_{j=1}^{n}\left(\hat{\alpha}_n - \alpha\right)^2 Z_{j-1}^2 + \mathcal{O}_p\left(\frac{1}{n}\right) = \mathcal{O}_p\left(\frac{1}{n}\right)$$

as, by our assumption above, $n\left(\hat{\alpha}_n - \alpha\right)^2 = \mathcal{O}_p(1)$ and $Q = \frac{1}{n}\sum_{j=1}^{n} Z_{j-1}^2 = \mathcal{O}_p(1)$ too, from $\mathbb{E}Q = \mathbb{E}Z_1^2 < \infty$ and Markov's inequality. □

5.3.2 Bootstrapping the eigenvalue estimates

We are mainly interested in the bootstrap for the estimation error of the autoregressive operator Ψ, again projected to a one-dimensional subspace. First, let us remark, that we do not consider $\hat{\Psi}_n - \Psi$ but, analogously to Mas [44], we compare $\hat{\Psi}_n$ to $\Psi\Pi_{k_n}$ where Π_{k_n} is the orthogonal projection onto the linear hull of $\nu_1, \cdots, \nu_{k_n}$. Note that Π_{k_n} replaces $\hat{\Pi}_{k_n}$ from Mas [44], as in this subsection, we assume the eigenvectors ν_j to be known. As a first step towards the bootstrap for $\hat{\Psi}_n - \Psi\Pi_{k_n}$ we first show that the bootstrap works for an auxiliary operator S_n which dominates the asymptotics of $\hat{\Psi}_n - \Psi\Pi_{k_n}$ and, then, that the bootstrap works for the error of eigenvalue estimates $\hat{\lambda}_j - \lambda_j$. We start with the introduction of S_n.

Lemma 8. $\hat{\Psi}_n - \Psi\Pi_{k_n} = \frac{1}{n} S_n \hat{\Gamma}_n^{\dagger}$ *with*

$$S_n = \sum_{t=1}^{n}\sum_{j=1}^{k_n}\langle X_{t-1}, \nu_j\rangle \langle \epsilon_t, \nu_j\rangle (\nu_j \otimes \nu_j) = n\left(\hat{C}_n - \Psi\hat{\Gamma}_n\right)$$

Proof. Using Lemma 7, we have

$$\hat{\Psi}_n - \Psi\Pi_{k_n} = \left(\hat{C}_n - \Psi\hat{\Gamma}_n\right)\hat{\Gamma}_n^{\dagger}$$

Note that

$$\Psi\left(\nu_j \otimes \nu_j(x)\right) = \Psi\left(\langle \nu_j, x\rangle \nu_j\right) = \frac{\gamma_j}{\lambda_j}\langle \nu_j, x\rangle \nu_j$$

as $\Psi(\nu_j) = \frac{\gamma_j}{\lambda_j}\nu_j$, i.e. we have $\Psi\nu_j \otimes \nu_j = \frac{\gamma_j}{\lambda_j}\nu_j \otimes \nu_j$. Therefore,

$$\begin{aligned}\hat{C}_n - \Psi\hat{\Gamma}_n &= \sum_{j=1}^{k_n}\left(\hat{\gamma}_j - \hat{\lambda}_j\Psi\right)\nu_j \otimes \nu_j \\ &= \sum_{j=1}^{k_n}\left(\hat{\gamma}_j - \gamma_j\frac{\hat{\lambda}_j}{\lambda_j}\right)\nu_j \otimes \nu_j\end{aligned}$$

Now, we plug in the definition of $\hat{\lambda}_j,\ \hat{\gamma}_j$ and get

$$\begin{aligned}\hat{\gamma}_j - \gamma_j\frac{\hat{\lambda}_j}{\lambda_j} &= \frac{1}{n}\sum_{t=1}^{n}\left(\langle X_{t-1},\nu_j\rangle\langle X_t,\nu_j\rangle - \frac{\gamma_j}{\lambda_j}\langle X_{t-1},\nu_j\rangle\langle X_{t-1},\nu_j\rangle\right) \\ &= \frac{1}{n}\sum_{t=1}^{n}\langle X_{t-1},\nu_j\rangle\left\langle X_t - \frac{\gamma_j}{\lambda_j}X_{t-1},\nu_j\right\rangle \\ &= \frac{1}{n}\sum_{t=1}^{n}\langle X_{t-1},\nu_j\rangle\langle\epsilon_t,\nu_j\rangle\end{aligned}$$

Hence, $\hat{C}_n - \Psi\hat{\Gamma}_n = \frac{1}{n}S_n$. □

Corollary 5.3.1. *With $Z_t = \langle X_t,\nu_m\rangle,\ \eta_t = \langle\epsilon_t,\nu_m\rangle$, we have for $m \le k_n$*

$$a)\qquad \left\langle\hat{C}_n - \Psi\hat{\Gamma}_n(y),\nu_m\right\rangle = \frac{1}{n}\sum_{t=1}^{n}Z_{t-1}\eta_t\langle y,\nu_m\rangle$$

$$b)\qquad \left\langle\hat{\Psi}_n - \Psi\Pi_{k_n}(x),\nu_m\right\rangle = \frac{1}{n}\sum_{t=1}^{n}Z_{t-1}\eta_t\frac{1}{\hat{\lambda}_m}\langle x,\nu_m\rangle$$

Proof. a) Follows immediately from $\langle\nu_j\otimes\nu_j(y),\nu_m\rangle = \langle\nu_j,y\rangle\langle\nu_j,\nu_m\rangle = \delta_{jm}\langle y,\nu_m\rangle$
b) Follows from a) with $y = \hat{\Gamma}_n^{\dagger}(x)$ and $\left\langle\hat{\Gamma}_n^{\dagger}(x),\nu_m\right\rangle = \left\langle x,\hat{\Gamma}_n^{\dagger}(\nu_m)\right\rangle = \frac{1}{\hat{\lambda}_m}\langle x,\nu_m\rangle$. □

Proposition 5.3.2. *Under the assumptions of Proposition 5.3.1, we have for $n\to\infty$*

$$\begin{aligned}& n\ d_2^2\left(\left\langle\hat{C}_n - \Psi\hat{\Gamma}_n(W),\nu_m\right\rangle,\left\langle\hat{C}_n^* - \hat{\Psi}_n\hat{\Gamma}_n^*(W),\nu_m\right\rangle\right) \\ =\ & \frac{1}{n}d_2^2\left(\langle S_n(W),\nu_m\rangle,\langle S_n^*(W),\nu_m\rangle\right) \underset{p}{\to} 0\end{aligned}$$

where W is an arbitrary L^2-valued random variable independent of $\{X_t\},\{X_t^*\}$.

Note that, due to the fact that the Mallows metric is a distance between distributions and that the distribution of $\{X_t^*\}$ is conditional on $X_0,\ldots,X_n$, the independence assumption on W has to be interpreted in the right manner: W is independent of the data in the real world used to calculate S_n which may be $X_0,\ldots,X_n$ or an independent realization of it. Moreover, W has to be conditionally independent of S_n^* given $X_0,\ldots,X_n$.

Proof. We follow essentially the line of arguments of the proof of Theorem 4.1 of Franke and Kreiss, [20]. Therefore, we consider a particular realisation of the two time series $\{X_t\}$ and $\{X_t^*\}$ with a specific kind of dependence. Let the pairs of innovations $(\epsilon_t', \epsilon_t^*)$ be i.i.d. and that for $\eta_t' = \langle \epsilon_t', \nu_m \rangle$, $\eta_t^* = \langle \epsilon_t^*, \nu_m \rangle$ we have $L(\eta_t') = F$, $L(\eta_t^*) = \hat{F}_n$ and $\mathbb{E}(\eta_t' - \eta_t^*)^2 = d_2^2\left(F, \hat{F}_n\right) \underset{p}{\to} 0$ for $n \to \infty$ by Proposition 5.3.1. Of course, we do all the calculations conditional on $X_0, \cdots, X_n$ such that $\hat{F}_n$ is treated as given, but $\underset{p}{\to}$ relates to its randomness.

Let X_0' be independent of $(\epsilon_t', \epsilon_t^*)$, $t \geq 1$, and $\mathcal{L}(X_0') = \mathcal{L}(X_0)$, and set recursively with $X_0^* = X_0'$

$$X_t' = \Psi\left(X_{t-1}'\right) + \epsilon_t', \; X_t^* = \Psi\left(X_{t-1}^*\right) + \epsilon_t^*, \; t \geq 1.$$

Writing again $Z_t' = \langle X_t', \nu_m \rangle$, we have as previously $Z_t' = \alpha_m Z_{t-1}' + \eta_t$ and, with $Z_t^* = \langle X_t^*, \nu_m \rangle$, we analogously have $Z_t^* = \hat{\alpha}_m Z_{t-1}^* + \eta_t^*$ as

$$\begin{aligned}
\left\langle \hat{\Psi}_n\left(X_{t-1}^*\right), \nu_m \right\rangle &= \left\langle \hat{C}_n \hat{\Gamma}_n^{\dagger} \left(\sum_{k=1}^{\infty} \left\langle X_{t-1}^*, \nu_k \right\rangle \nu_k \right), \nu_m \right\rangle \\
&= \sum_{k=1}^{\infty} \left\langle X_{t-1}^*, \nu_k \right\rangle \frac{\hat{\gamma}_k}{\hat{\lambda}_k} \langle \nu_k, \nu_m \rangle \\
&= \hat{\alpha}_m \left\langle X_{t-1}^*, \nu_m \right\rangle = \hat{\alpha}_m Z_{t-1}^*
\end{aligned}$$

a) First, we show the following analogue relationship to (4.18) of Franke and Kreiss [20]. We have with probability converging to 1

$$\begin{aligned}
\mathbb{E}\left\{ (Z_t' - Z_t^*)^2 \mid X_0, \cdots, X_n \right\} &\leq D\left(\mathbb{E}Z_0^2 + \mathbb{E}^*(\eta_1^*)^2\right)(\alpha_m - \hat{\alpha}_m)^2 \\
&\quad + D d_2^2\left(F, \hat{F}_n\right)
\end{aligned}$$

for some generic constant D. Using the autoregressive form of Z_t', Z_t^* respectively, we get recursively

$$\begin{aligned}
Z_t' &= \alpha_m Z_{t-1}' + \eta_t' = \alpha_m^2 Z_{t-2}' + \alpha_m \eta_{t-1}' + \eta_t' = \cdots \\
&= \alpha_m^t Z_0' + \sum_{k=0}^{t-1} \alpha_m^k \eta_{t-k}'
\end{aligned}$$

and, as $X_0^* = X_0'$,

$$Z_t^* = \hat{\alpha}_m^t Z_0' + \sum_{k=0}^{t-1} \hat{\alpha}_m^k \eta_{t-k}^*$$

Note that $\left|\alpha_m^k - \hat{\alpha}_m^k\right| = |\alpha_m - \hat{\alpha}_m| \left|\sum_{j=0}^{k-1} \alpha_m^j \hat{\alpha}_m^{k-1-j}\right|$.

From Lemma 7 and the preceding remarks, we know that $|\alpha_m| < 1$. Moreover, from the remarks in the first paragraph of the proof of Proposition 5.3.1, we have $\hat{\lambda}_m =$

$\lambda_m + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)$ and $\hat{\gamma}_m = \gamma_m + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)$, which implies $\hat{\alpha}_m \underset{p}{\to} \alpha_m$. Therefore, we can choose $1 > \hat{\delta} > |\alpha_m|$ such that $pr\left(|\hat{\alpha}_m| \leq \hat{\delta}\right) \to 1$. If $|\hat{\alpha}_m| \leq \hat{\delta}$, we have

$$\left|\sum_{j=0}^{k-1} \alpha_m^j \hat{\alpha}_m^{k-1-j}\right| \leq \hat{\delta}^{k-1} \sum_{j=0}^{k-1} \left|\frac{\alpha_m}{\hat{\delta}}\right|^j \leq \frac{\hat{\delta}^{k-1}}{1 - \left|\frac{\alpha_m}{\hat{\delta}}\right|} = \frac{\hat{\delta}^k}{\hat{\delta} - |\alpha_m|}$$

and with probability converging to 1:

$$\left|\alpha_m^k - \hat{\alpha}_m^k\right| \leq |\alpha_m - \hat{\alpha}_m| \frac{\hat{\delta}^k}{\hat{\delta} - |\alpha_m|}$$

Hence, with probability converging to 1, writing short hand $\mathbb{E}$ for $\mathbb{E}\left\{\cdot \mid X_0, \cdots, X_n\right\}$

$$\begin{aligned}
\mathbb{E}\left(Z_t' - Z_t^*\right)^2 &= \mathbb{E}\left(\left(\alpha_m^t - \hat{\alpha}_m^t\right) Z_0' + \sum_{k=0}^{t-1} \alpha_m^k \left(\eta_{t-k}' - \eta_{t-k}^*\right) + \sum_{k=0}^{t-1} \left(\alpha_m^k - \hat{\alpha}_m^k\right) \eta_{t-k}^*\right)^2 \\
&\leq 3\mathbb{E}\left(Z_0'\right)^2 \left(\alpha_m^t - \hat{\alpha}_m^t\right)^2 + 3\sum_{k=0}^{t-1} \alpha_m^{2k} \mathbb{E}\left(\eta_{t-k}' - \eta_{t-k}^*\right)^2 + 3\sum_{k=0}^{t-1} \left(\alpha_m^k - \hat{\alpha}_m^k\right)^2 \mathbb{E}\left(\eta_{t-k}^*\right)^2
\end{aligned}$$

using that (η_k', η_k^*) are i.i.d. with mean 0. Hence,

$$\begin{aligned}
\mathbb{E}\left(Z_t' - Z_t^*\right)^2 &\leq 3\left(\mathbb{E}\left(Z_0'\right)^2 \frac{\hat{\delta}^{2t}}{\left(\hat{\delta} - |\alpha_m|\right)^2} + \mathbb{E}^*\left(\eta_t^*\right)^2 \sum_{k=0}^{t-1} \frac{\hat{\delta}^{2k}}{\left(\hat{\delta} - |\alpha_m|\right)^2}\right) (\alpha_m - \hat{\alpha}_m)^2 \\
&\quad + 3\frac{1}{1 - \alpha_m^2} d_2^2\left(F, \hat{F}_n\right) \\
&\leq \frac{3}{\left(\hat{\delta} - |\alpha_m|\right)^2} \left(\hat{\delta}^{2t}\mathbb{E}\left(Z_0'\right)^2 + \mathbb{E}^*\left(\eta_t^*\right)^2\right) (\alpha_m - \hat{\alpha}_m)^2 + \frac{3}{1 - \alpha_m^2} d_2^2\left(F, \hat{F}_n\right)
\end{aligned}$$

b) From Corollary 5.3.1 a), we have

$$\begin{aligned}
&nd_2^2\left(\left\langle \hat{C}_n - \Psi\hat{\Gamma}_n\left(W\right), \nu_m\right\rangle, \left\langle \hat{C}_n^* - \hat{\Psi}_n\hat{\Gamma}_n^*\left(W\right), \nu_m\right\rangle\right) \\
&\leq \frac{1}{n}\mathbb{E}\left\langle W, \nu_m\right\rangle^2 \mathbb{E}\left(\sum_{t=1}^n Z_{t-1}'\eta_t' - \sum_{t=1}^n Z_{t-1}^*\eta_t^*\right)^2 \\
&\leq \frac{2}{n}\mathbb{E}\left\langle W, \nu_m\right\rangle^2 \left\{\mathbb{E}\left(\sum_{t=1}^n \left(Z_{t-1}' - Z_{t-1}^*\right)\eta_t'\right)^2 + \mathbb{E}\left(\sum_{t=1}^n Z_{t-1}^*\left(\eta_t' - \eta_t^*\right)\right)^2\right\} \\
&= \frac{2}{n}\mathbb{E}\left\langle W, \nu_m\right\rangle^2 \left\{\sum_{t=1}^n \mathbb{E}\left(Z_{t-1}' - Z_{t-1}^*\right)^2 \mathbb{E}\eta_1^2 + \sum_{t=1}^n \mathbb{E}^*\left(Z_{t-1}^*\right)^2 \mathbb{E}\left(\eta_t' - \eta_t^*\right)^2\right\} \\
&\leq D\mathbb{E}\left\langle W, \nu_m\right\rangle^2 \left\{(\hat{\alpha}_m - \alpha_m)^2 + d_2^2\left(F, \hat{F}_n\right)\right\} \underset{p}{\to} 0
\end{aligned}$$

where all the calculations are conditional w.r.t. $X_0, \cdots, X_n$, and D is some generic constant. We use, e.g., $\mathbb{E}^* \left(\eta_t^*\right)^2 = \frac{1}{n}\sum_{t=1}^n \hat{\eta}_t^2 \leq 2\mathbb{E}\eta_1^2$ and

$$\mathbb{E}^* \left(Z_t^*\right)^2 = \mathbb{E}\left(Z_t^* - Z_t' + Z_t'\right)^2 \leq 2\mathbb{E}\left(Z_t^* - Z_t'\right)^2 + 2\mathbb{E}Z_1'^2 \leq 6\mathbb{E}Z_1^2$$

with probability converging to 1. □

We mainly are interested in the autoregressive operator and its approximation by the bootstrap, in particular, for predictions, in $\left\langle \hat{\Psi}_n - \Psi\Pi_{k_n}(W), \nu_m \right\rangle$ for some random variable W. By Corollary 5.3.1, this is a ratio of the quantity studied in Proposition 5.3.2 and $\hat{\lambda}_m$. Therefore, we next show that the bootstrap works for $\hat{\lambda}_m$.

Proposition 5.3.3. *Under the assumptions of Proposition 5.3.1, we have for* $n \to \infty$

$$nd_2^2\left(\hat{\lambda}_m - \lambda_m, \hat{\lambda}_m^* - \hat{\lambda}_m\right) \underset{p}{\to} 0$$

Proof. Recall that $\mathbb{E}Z_t^2 = \lambda_m$ (compare with the remarks preceding Lemma 7) such that

$$\hat{\lambda}_m - \lambda_m = \frac{1}{n}\sum_{t=1}^n \left(Z_t^2 - \mathbb{E}Z_t^2\right)$$

a) First we show that $\hat{\lambda}_m$ coincides with $\mathbb{E}^*\hat{\lambda}_m^*$ up to asymptotically negligible terms. We have from the representation of Z_t^* in the proof of Proposition 5.3.2

$$\mathbb{E}^* \left(Z_t^*\right)^2 = \mathbb{E}^*\left(\hat{\alpha}_m^t Z_0' + \sum_{k=0}^{t-1}\hat{\alpha}_m^k \eta_{t-k}^*\right)^2 = \hat{\alpha}_m^{2t}\mathbb{E}Z_0'^2 + \sum_{k=0}^{t-1}\hat{\alpha}_m^{2k}\mathbb{E}^*\left(\eta_{t-k}^*\right)^2$$

as $Z_0', \eta_1^*, \cdots, \eta_t^*$ are i.i.d. in the bootstrap world. Moreover,

$$\begin{aligned}
\mathbb{E}^*\left(\eta_{t-k}^*\right)^2 &= \mathbb{E}^*\left(\eta_1^*\right)^2 = \frac{1}{n}\sum_{t=1}^n \tilde{\eta}_t^2 = \frac{1}{n}\sum_{t=1}^n\left(\hat{\eta}_t - \frac{1}{n}\sum_{k=1}^n \hat{\eta}_k\right)^2 \\
&= \frac{1}{n}\sum_{t=1}^n\left(Z_t - \hat{\alpha}_m Z_{t-1} - \frac{1}{n}\sum_{k=1}^n\left(Z_k - \hat{\alpha}_m Z_{k-1}\right)\right)^2 \\
&= \frac{1}{n}\sum_{t=1}^n\left(Z_t - \hat{\alpha}_m Z_{t-1}\right)^2 - \left(\frac{1}{n}\sum_{k=1}^n\left(Z_k - \hat{\alpha}_m Z_{k-1}\right)\right)^2
\end{aligned}$$

where

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^n\left(Z_t - \hat{\alpha}_m Z_{t-1}\right)^2 &= \frac{1}{n}\sum_{t=1}^n Z_t^2 - 2\hat{\alpha}_m\frac{1}{n}\sum_{t=1}^n Z_{t-1}Z_t + \hat{\alpha}_m^2\frac{1}{n}\sum_{t=1}^n Z_{t-1}^2 \\
&= \hat{\lambda}_m + \frac{1}{n}\left(Z_n^2 - Z_0^2\right) - 2\hat{\alpha}_m\hat{\gamma}_m + \hat{\alpha}_m^2\hat{\lambda}_m \\
&= \hat{\lambda}_m\left(1 - \hat{\alpha}_m^2\right) + \frac{1}{n}\left(Z_n^2 - Z_0^2\right)
\end{aligned}$$

as $\hat{\alpha}_m\hat{\gamma}_m = \hat{\alpha}_m^2\hat{\lambda}_m$. As $\{Z_t\}$ is a stationary AR(1) process, we have (compare, e.g., Brockwell and Davis, [8], p.219),

$$\frac{1}{n}\sum_{t=1}^{n}\left(Z_t - \hat{\alpha}_m Z_{t-1}\right) = \left(1-\hat{\alpha}_m\right)\mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)$$

as $\mathbb{E}Z_t = 0$ and the sample mean is asymptotically normal. More precisely, the asymptotic rate also holds in the L^2-sense, not only in probability, as

$$\begin{aligned}
\mathbb{E}^*\hat{\lambda}_m^* &= \frac{1}{n}\sum_{t=0}^{n-1}\mathbb{E}^*\left(Z_{t-1}^*\right)^2 = \frac{\mathbb{E}\left(Z_0'\right)^2}{n}\frac{1-\hat{\alpha}_m^{2n}}{1-\hat{\alpha}_m^2} + \frac{\mathbb{E}^*\left(\eta_1^*\right)^2}{n}\sum_{t=0}^{n-1}\sum_{k=0}^{t-1}\hat{\alpha}_m^{2k} \\
&= \frac{\mathbb{E}\left(Z_0'\right)^2}{n}\frac{1-\hat{\alpha}_m^{2n}}{1-\hat{\alpha}_m^2} + \frac{\mathbb{E}^*\left(\eta_1^*\right)^2}{n}\sum_{t=0}^{n-1}\frac{1-\hat{\alpha}_m^{2t}}{1-\hat{\alpha}_m^2} \\
&= \left(\hat{\lambda}_m + \frac{1}{n}\frac{Z_n^2 - Z_0^2}{1-\hat{\alpha}_m^2}\right)\frac{1}{n}\left(n - \frac{1-\hat{\alpha}_m^{2n}}{1-\hat{\alpha}_m^2}\right) + \frac{\mathbb{E}\left(Z_0'\right)^2}{n}\frac{1-\hat{\alpha}_m^{2n}}{1-\hat{\alpha}_m^2}
\end{aligned}$$

As with probability converging to 1, $|\alpha_m| \leq \hat{\delta} < 1$ for $\hat{\delta} > |\hat{\alpha}_m|$, we get

$$\mathbb{E}^*\lambda_m^* = \hat{\lambda}_m + \frac{1}{n}R_n$$

with $\mathbb{E}R_n^2$ uniformly bounded for all large enough n. Note that here and in the following, expectations are conditional on $X_0, \cdots, X_n$.

b) We now choose Z_0', (η_t', η_t^*) as in the proof of Proposition 5.3.2. Then, from a),

$$\begin{aligned}
d_2^2\left(\hat{\lambda}_m - \lambda_m, \hat{\lambda}_m^* - \hat{\lambda}_m\right) &\leq \mathbb{E}\left\{\frac{1}{n}\sum_{t=1}^{n}\left[\left(Z_t'\right)^2 - \mathbb{E}\left(Z_t'\right)^2\right] - \frac{1}{n}\sum_{t=1}^{n}\left[\left(Z_t^*\right)^2 - \mathbb{E}\left(Z_t^*\right)^2\right] + \frac{1}{n}R_n\right\}^2 \\
&\leq 2\mathbb{E}\left\{\frac{1}{n}\sum_{t=1}^{n}\left[\left(Z_t'\right)^2 - \mathbb{E}\left(Z_t'\right)^2\right] - \frac{1}{n}\sum_{t=1}^{n}\left[\left(Z_t^*\right)^2 - \mathbb{E}\left(Z_t^*\right)^2\right]\right\}^2 \\
&\quad + \frac{2}{n^2}\mathbb{E}R_n^2
\end{aligned}$$

The remainder term is of order $\dfrac{1}{n^2}$ such that it remains to show that

$$\begin{aligned}
&\frac{1}{n^2}\mathbb{E}\left(\sum_{t=1}^{n}\left[\left(Z_t'\right)^2 - \mathbb{E}\left(Z_t'\right)^2 - \left\{\left(Z_t^*\right)^2 - \mathbb{E}\left(Z_t^*\right)^2\right\}\right]\right)^2 \\
&= \frac{1}{n^2}\sum_{s,t=1}^{n}\mathbb{E}\left[\left(Z_t'\right)^2 - \mathbb{E}\left(Z_t'\right)^2 - \left(Z_t^*\right)^2 + \mathbb{E}\left(Z_t^*\right)^2\right]\left[\left(Z_s'\right)^2 - \mathbb{E}\left(Z_s'\right)^2 - \left(Z_s^*\right)^2 + \mathbb{E}\left(Z_s^*\right)^2\right] \\
&= \frac{1}{n^2}\sum_{s,t=1}^{n}\mathbb{E}A_tA_s = o_p\left(\frac{1}{n}\right)
\end{aligned}$$

Again, we use the representation of Z_t', Z_t^* in terms of $Z_0',\ \eta_k', \eta_k^*$ from the proof of Proposition 5.3.2 such that

$$\begin{aligned}
Z_t'^2 - \mathbb{E}\left(Z_t'\right)^2 &= \left(\alpha_m^t Z_0' + \sum_{k=1}^{t-1} \alpha_m^k \eta_{t-k}'\right)^2 - \mathbb{E}\left(\alpha_m^t Z_0' + \sum_{k=1}^{t-1} \alpha_m^k \eta_{t-k}'\right)^2 \\
&= \alpha_m^{2t}\left[\left(Z_0'\right)^2 - \mathbb{E}\left(Z_0'\right)^2\right] + 2\alpha_m^t Z_0' \sum_{k=1}^{t-1} \alpha_m^k \eta_{t-k}' \\
&\quad + \sum_{k,l=1}^{t-1} \alpha_m^{k+l}\left[\eta_{t-k}'\eta_{t-l}' - \mathbb{E}\left(\eta_{t-k}'\eta_{t-l}'\right)\right] \\
&= \alpha_m^{2t}\left[\left(Z_0'\right)^2 - \mathbb{E}\left(Z_0'\right)^2\right] + 2\alpha_m^t Z_0' \sum_{k=1}^{t-1} \alpha_m^{t-k} \eta_k' + \sum_{k\neq l=1}^{t-1} \alpha_m^{2t-k-l} \eta_k'\eta_l' \\
&\quad + \sum_{k=1}^{t-1} \alpha_m^{2(t-k)}\left[\left(\eta_k'\right)^2 - \mathbb{E}\left(\eta_k'\right)^2\right]
\end{aligned}$$

using that $Z_0', \eta_1', \cdots \eta_k'$ i.i.d. with mean 0. With the same decomposition for $\left(Z_t^*\right)^2 - \mathbb{E}^*\left(Z_t^*\right)^2$, we have

$$A_t = a_t + 2b_t + c_t + d_t$$

with, using $Z_0^* = Z_0'$

$$\begin{aligned}
a_t &= \left(\alpha_m^{2t} - \hat{\alpha}_m^{2t}\right)\left[\left(Z_0'\right)^2 - \mathbb{E}\left(Z_0'\right)^2\right] \\
b_t &= \alpha_m^t Z_0' \sum_{k=1}^{t-1} \alpha_m^{t-k}\eta_k' - \hat{\alpha}_m^t Z_0' \sum_{k=1}^{t-1} \hat{\alpha}_m^{t-k}\eta_k^* \\
c_t &= \sum_{k\neq l=1}^{t-1} \alpha_m^{2t-k-l}\eta_k'\eta_l' - \sum_{k\neq l=1}^{t-1} \hat{\alpha}_m^{2t-k-l}\eta_k^*\eta_l^* \\
d_t &= \sum_{k=1}^{t-1} \alpha_m^{2(t-k)}\left[\left(\eta_k'\right)^2 - \mathbb{E}\left(\eta_k'\right)^2\right] - \sum_{k=1}^{t-1} \hat{\alpha}_m^{2(t-k)}\left[\left(\eta_k^*\right)^2 - \mathbb{E}^*\left(\eta_k^*\right)^2\right]
\end{aligned}$$

So, we have to show

$$\sum_{s,t=1}^{n} \mathbb{E}\left(a_t + 2b_t + c_t + d_t\right)\left(a_s + 2b_s + c_s + d_s\right) = o_p(n)$$

with probability converging to 1 (w.r.t. the randomness of $X_0, \cdots, X_n$). We only consider some of the 16 products where the others can be dealt with in exactly the same manner. We repeatedly use

$$\left|\alpha_m^k - \hat{\alpha}_m^k\right| \leq c\left|\alpha_m - \hat{\alpha}_m\right| \hat{\delta}^k \tag{5.13}$$

with probability converging to 1 with some $|\alpha_m| < \hat{\delta} < 1$ and $c > 0$, (compare with the proof of Proposition 5.3.2).

$$\begin{aligned}
\sum_{s,t=1}^{n} \mathbb{E}a_t a_s &= \sum_{s,t=1}^{n} \left(\alpha_m^{2t} - \hat{\alpha}_m^{2t}\right)\left(\alpha_m^{2s} - \hat{\alpha}_m^{2s}\right) \mathbb{E}\left[\left(Z_0'\right)^2 - \mathbb{E}\left(Z_0'\right)^2\right]^2 \\
&\leq c\left(\alpha_m - \hat{\alpha}_m\right)^2 \sum_{s,t=1}^{n} \hat{\delta}^{2(t+s)} \mathbb{E}\left[\left(Z_0'\right)^2 - \mathbb{E}\left(Z_0'\right)^2\right]^2 \\
&\leq c\left(\alpha_m - \hat{\alpha}_m\right)^2 \left(\frac{1-\hat{\delta}^{2n}}{1-\hat{\delta}^2}\right)^2 \mathbb{E}\left[\left(Z_0'\right)^2 - \mathbb{E}\left(Z_0'\right)^2\right]^2 \\
&= \mathcal{O}_p\left(\frac{1}{n}\right)
\end{aligned}$$

using 5.8.

$$\sum_{s,t=1}^{n} \mathbb{E}c_t c_s = \sum_{s,t=1}^{n} \sum_{k\neq l=1}^{t-1} \sum_{i\neq j=1}^{s-1} \mathbb{E}B_{k,l,i,j}$$

with $B_{k,l,i,j} = \left(\alpha_m^{2t-k-l}\eta_k'\eta_l' - \hat{\alpha}_m^{2t-k-l}\eta_k^*\eta_l^*\right)\left(\alpha_m^{2s-i-j}\eta_i'\eta_j' - \hat{\alpha}_m^{2s-i-j}\eta_i^*\eta_j^*\right)$.
As $\left(\eta_k', \eta_k^*\right),\ k = 1, \cdots, n$ are i.i.d. with mean 0, and as $k \neq l,\ i \neq j,\ \mathbb{E}B_{k,l,i,j} = 0$ except for $k = i,\ l = j$ and $k = j,\ l = i$. In particular, $k + l = i + j$, and hence, $B_{klkl} = B_{kllk}$ and

$$\begin{aligned}
\mathbb{E}B_{klkl} &= \mathbb{E}B_{kllk} \\
&= \alpha_m^{2(t+s-k-l)}\left(\mathbb{E}\left(\eta_1'\right)^2\right)^2 + \hat{\alpha}_m^{2(t+s-k-l)}\left(\mathbb{E}^*\left(\eta_1^*\right)^2\right)^2 \\
&\quad -\alpha_m^{2t-k-l}\hat{\alpha}_m^{2s-k-l}\mathbb{E}\left(\eta_1'\right)^2\mathbb{E}^*\left(\eta_1^*\right)^2 - \alpha_m^{2s-k-l}\hat{\alpha}_m^{2t-k-l}\mathbb{E}\left(\eta_1'\right)^2\mathbb{E}^*\left(\eta_1^*\right)^2 \\
&= \alpha_m^{2t-k-l}\mathbb{E}\left(\eta_1'\right)^2\left\{\left(\alpha_m^{2s-k-l} - \hat{\alpha}_m^{2s-k-l}\right)\mathbb{E}\left(\eta_1'\right)^2 + \hat{\alpha}_m^{2s-k-l}\left(\mathbb{E}\left(\eta_1'\right)^2 - \mathbb{E}^*\left(\eta_1^*\right)^2\right)\right\} \\
&\quad +\hat{\alpha}_m^{2t-k-l}\mathbb{E}^*\left(\eta_1^*\right)^2\left\{\left(\hat{\alpha}_m^{2s-k-l} - \alpha_m^{2s-k-l}\right)\mathbb{E}^*\left(\eta_1^*\right)^2 + \alpha_m^{2s-k-l}\left(\mathbb{E}^*\left(\eta_1^*\right)^2 - \mathbb{E}\left(\eta_1'\right)^2\right)\right\}
\end{aligned}$$

Using 5.13, $|\alpha_m|,\ |\hat{\alpha}_m| \leq \hat{\delta},\ \mathbb{E}^*\left(\eta_1^*\right)^2 \leq c$ for some constant c with probability converging to 1, we get for some constant D

$$|\mathbb{E}B_{klkl}| \leq D\hat{\delta}^{2(t+s-k-l)}\left\{|\alpha_m - \hat{\alpha}_m| + \left|\mathbb{E}\left(\eta_1'\right)^2 - \mathbb{E}^*\left(\eta_1^*\right)^2\right|\right\}$$

As $k=i,\ l=j$ or $k=j,\ l=i$ is only possible if $k,l \leq \min(s,t)-1$, we get

$$\begin{aligned}
\left|\sum_{s,t=1}^{n} \mathbb{E}c_t c_s\right| &\leq 2 \sum_{1\leq s\leq t\leq n} \sum_{k\neq l=1}^{s-1} \left(|\mathbb{E}B_{klkl}| + |\mathbb{E}B_{kllk}|\right) \\
&\leq 4D \sum_{1\leq s\leq t\leq n} \sum_{k\neq l=1}^{s-1} \hat{\delta}^{2(t+s-k-l)} \left\{|\alpha_m - \hat{\alpha}_m| + \left|\mathbb{E}\left(\eta_1'\right)^2 - \mathbb{E}^*\left(\eta_1^*\right)^2\right|\right\} \\
&= 4D \sum_{1\leq s\leq t\leq n} \sum_{k\neq l=1}^{s-1} \hat{\delta}^{2(t-s+k+l)} \left\{|\alpha_m - \hat{\alpha}_m| + \left|\mathbb{E}\left(\eta_1'\right)^2 - \mathbb{E}^*\left(\eta_1^*\right)^2\right|\right\} \\
&\leq \frac{4D}{\left(1-\hat{\delta}^2\right)^2} \sum_{t=1}^{n} \sum_{s=1}^{t} \hat{\delta}^{2(t-s)} \left\{|\alpha_m - \hat{\alpha}_m| + \left|\mathbb{E}\left(\eta_1'\right)^2 - \mathbb{E}^*\left(\eta_1^*\right)^2\right|\right\} \\
&= \frac{4D}{\left(1-\hat{\delta}^2\right)^2} \sum_{t=1}^{n} \frac{1-\hat{\delta}^{2t}}{1-\hat{\delta}^2} \left\{|\alpha_m - \hat{\alpha}_m| + \left|\mathbb{E}\left(\eta_1'\right)^2 - \mathbb{E}^*\left(\eta_1^*\right)^2\right|\right\} \\
&= o_p(n)
\end{aligned}$$

as $|\alpha_m - \hat{\alpha}_m| = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)$ and $\mathbb{E}^*\left(\eta_1^*\right)^2 \underset{p}{\to} \mathbb{E}\left(\eta_1'\right)^2$ for $n\to\infty$ by Proposition 5.3.1 and Lemma 8.3 of Bickel and Freedman, [6].

$$\sum_{s,t=1}^{n} \mathbb{E}d_t d_s = \sum_{s,t=1}^{n} \sum_{k=1}^{t-1} \sum_{j=1}^{s-1} \mathbb{E}B_{kj}$$

with, abbreviating $S_k = \left(\eta_k'\right)^2 - \mathbb{E}\left(\eta_k'\right)^2,\ S_k^* = \left(\eta_k^*\right)^2 - \mathbb{E}^*\left(\eta_k^*\right)^2$

$$B_{kj} = \left(\alpha_m^{2(t-k)} S_k - \hat{\alpha}_m^{2(t-k)} S_k^*\right)\left(\alpha_m^{2(s-j)} S_j - \hat{\alpha}_m^{2(s-j)} S_j^*\right)$$

Note that $\mathbb{E}S_k = \mathbb{E}^* S_k^* = 0$ such that from the independence of $\left(S_k', S_k^*\right),\ k=1,\cdots,n,\ \mathbb{E}B_{kj} = 0$ for $k\neq j$. Moreover,

$$\begin{aligned}
\mathbb{E}B_{kk} &= \alpha_m^{2(t+s-2k)} \mathbb{E}S_k^2 + \hat{\alpha}_m^{2(t+s-2k)} \mathbb{E}^*\left(S_k^*\right)^2 \\
&\quad -\alpha_m^{2(t-k)} \hat{\alpha}_m^{2(s-k)} \mathbb{E}\left(S_k S_k^*\right) - \alpha_m^{2(s-k)} \hat{\alpha}_m^{2(t-k)} \mathbb{E}\left(S_k S_k^*\right) \\
&= \alpha_m^{2(t-k)} \left\{\left(\alpha_m^{2(s-k)} - \hat{\alpha}_m^{2(s-k)}\right) \mathbb{E}S_k^2 + \hat{\alpha}_m^{2(s-k)} \left(\mathbb{E}S_k^2 - \mathbb{E}\left(S_k S_k^*\right)\right)\right\} \\
&\quad +\hat{\alpha}_m^{2(t-k)} \left\{\left(\hat{\alpha}_m^{2(s-k)} - \alpha_m^{2(s-k)}\right) \mathbb{E}^*\left(S_k^*\right)^2 + \alpha_m^{2(s-k)} \left(\mathbb{E}^*\left(S_k^*\right)^2 - \mathbb{E}\left(S_k S_k^*\right)\right)\right\}
\end{aligned}$$

As $\left(S_k, S_k^*\right)$ are i.i.d., we get again, as $\mathbb{E}^*\left(S_1^*\right)^2 \leq C$ for some constant C from part c) of the proof below, and hence by Cauchy-Schwarz

$$\left|\mathbb{E}^* S_k^*\left(S_k^* - S_k\right)\right| \leq C\left\{\mathbb{E}^*\left(S_k - S_k^*\right)^2\right\}^{1/2}$$

for some constant C,

$$\begin{aligned}
\mathbb{E}B_{kk} &\leq D\hat{\delta}^{2(t+s-2k)} \left\{|\alpha_m - \hat{\alpha}_m| + \left[\mathbb{E}^*\left(S_1 - S_1^*\right)^2\right]^{1/2}\right\} \\
&= \hat{\delta}^{2(t+s-2k)} o_p(1)
\end{aligned}$$

as $\hat{\alpha}_m \underset{p}{\to} \alpha_m$ and from part c) below. Remarking again that $k = j$ is only possible for $k \leq min(t,s) - 1$, we have

$$\begin{aligned}
\left| \sum_{s,t=1}^{n} \mathbb{E} d_t d_s \right| &\leq 2 \sum_{1 \leq s \leq t \leq n} \sum_{k=1}^{s-1} |\mathbb{E} B_{kk}| \\
&= 2 \sum_{1 \leq s \leq t \leq n} \sum_{k=1}^{s-1} \hat{\delta}^{2(t+s-2k)} o_p(1) \\
&= 2 \sum_{1 \leq s \leq t \leq n} \sum_{l=1}^{s-1} \hat{\delta}^{2(t-s+2l)} o_p(1) \\
&\leq \frac{2}{1-\hat{\delta}^4} \sum_{t=1}^{n} \sum_{s=1}^{t} \hat{\delta}^{2(t-s)} o_p(1) \\
&= o_p(n)
\end{aligned}$$

c) It remains to show that

$$\mathbb{E} \left(S_1 - S_1^* \right)^2 \underset{p}{\to} 0 \tag{5.14}$$

First we note that this implies that $\mathbb{E}^* \left(S_1^* \right)^2 \leq C$ with probability converging to 1 for all large enough C, as, writing $\left\| S_1^* \right\|^2 = \mathbb{E}^* \left(S_1^* \right)^2$ for the usual L^2-norm of real random variables,

$$\left| \|S_1\| - \|S_1^*\| \right| \leq \|S_1 - S_1^*\|$$

implying

$$\|S_1^*\| \leq \min \left\{ \|S_1\|, \ \|S_1 - S_1^*\| \right\}$$

which is asymptotically bounded if $\|S_1 - S_1^*\| \underset{p}{\to} 0$ and $\|S_1\| < \infty$. The latter follows from our assumption $\mathbb{E} \left\| \epsilon_1 \right\|^4 < \infty$, which implies
$\mathbb{E} \eta_1^4 = \mathbb{E} \left\langle \epsilon_1, \nu_m \right\rangle^4 \leq \mathbb{E} \left\| \epsilon_1 \right\|^4 \left\| \nu_m \right\|^4 < \infty$. Now

$$\begin{aligned}
\mathbb{E} \left(S_1 - S_1^* \right)^2 &\leq \mathbb{E} \left(\left(\eta_1' \right)^2 - \left(\eta_1^* \right)^2 \right)^2 \\
&= \mathbb{E} \left(\eta_1^* - \eta_1' \right)^2 \left(\eta_1^* + \eta_1' \right)^2
\end{aligned}$$

As $\mathbb{E} \left(\eta_1^* - \eta_1' \right)^2 = d_2^2 \left(\hat{F}_n, F \right) \underset{p}{\to} 0$ by Proposition 5.3.1, we have $\left(\eta_1^* - \eta_1' \right)^2 \underset{p}{\to} 0$, and 5.14 follows from a dominated convergence argument if we can show that there is a random variable U with

$$\left(\left(\eta_1' \right)^2 - \left(\eta_1^* \right)^2 \right)^2 \leq 2 \left[\left(\eta_1' \right)^4 + \left(\eta_1^* \right)^4 \right] \leq 2 \left[\left(\eta_1' \right)^4 + U^4 \right]$$

and, as $\mathbb{E} \left(\eta_1' \right)^4 = \mathbb{E} \left(\eta_1 \right)^4 < \infty$, we have $\mathbb{E} U^4 < \infty$. Let J be a Laplace random variable in $\{1, \cdots, n\}$, i.e. $pr \left(J = k \right) = \dfrac{1}{n}$, $1 \leq k \leq n$, and we set

$$U = |\eta_J| + |\alpha_m - \hat{\alpha}_m| \left| Z_{J-1} \right| + \left| \frac{1}{n} \sum_{j=1}^{n} Z_j \right| + |\hat{\alpha}_m| \left| \frac{1}{n} \sum_{j=1}^{n} Z_j \right|$$

Then, $\mathbb{E}U^4 < \infty$ (recall that all expectations in this proof are conditional on $X_0, \cdots, X_n$), as, by the law of large numbers,

$$\mathbb{E}\left|\eta_J\right|^4 = \frac{1}{n}\sum_{j=1}^{n}\eta_j^4 \leq C$$

for any $C > \mathbb{E}\eta_1^4$ with probability converging to 1, and

$$\mathbb{E}\left|Z_{J-1}\right|^4 = \frac{1}{n}\sum_{j=1}^{n}Z_{j-1}^4 \leq C$$

for any $C > \mathbb{E}Z_1^4$ analogously. □

5.3.3 Bootstrapping the estimate of the autoregressive operator

Finally, we can consider $\hat{\Psi}_n - \Psi\Pi_{k_n}$ which represents the additional forecasting error caused by using $\hat{\Psi}_n$ instead of Ψ. Note that, by Theorems 3.1 and 3.2 of Mas, [44], we have to consider $\Psi\Pi_{k_n}$ instead of Ψ. Let us assume that we use $\hat{\Psi}_n$ to forecast an independent realisation $\{X_t^\circ\}$ of the FAR(1)-process $\{X_t\}$. More precisely, we predict X_{n+2}° by $\hat{\Psi}_n\left(X_{n+1}^\circ\right)$ like in Theorem 3.1 of Mas, [44]. The following result shows that the bootstrap works for approximating the distribution of $\hat{\Psi}_n - \Psi\Pi_{k_n}\left(X_{n+1}^\circ\right)$ projected onto ν_m.

Proposition 5.3.4. *Under the assumptions of Proposition 5.3.1, we have for $n \to \infty$ that*

$$\sqrt{n}\left\langle\hat{\Psi}_n - \Psi\Pi_{k_n}\left(X_{n+1}^\circ\right), \nu_m\right\rangle \text{ and } \sqrt{n}\left\langle\hat{\Psi}_n^* - \hat{\Psi}_n\Pi_{k_n}\left(X_{n+1}^\circ\right), \nu_m\right\rangle$$

converge in distribution to the same limit.

Proof. The result follows from Proposition 5.3.2 and 5.3.3 by the delta method for the bootstrap (Theorem 23.5 of van der Vaart, [55]), as

$$\begin{aligned}\sqrt{n}\left\langle\hat{\Psi}_n - \Psi\Pi_{k_n}\left(X_{n+1}^\circ\right), \nu_m\right\rangle &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{t-1}\eta_t\left\langle X_{n+1}^\circ, \nu_m\right\rangle\frac{1}{\hat{\lambda}_m} \\ &= \sqrt{n}\phi\left(\left\langle\hat{C}_n - \Psi\hat{\Gamma}_n\left(X_{n+1}^\circ\right), \nu_m\right\rangle, \hat{\lambda}_m\right)\end{aligned}$$

by Corollary 5.3.1 where $\phi(u, v)$ is continuously differentiable for all u and for all v in a neighbourhood of $\lambda_m > 0$. □

5.4 Approximation of the innovations by the bootstrap innovations for a special case

We now drop the assumption that the eigenvectors ν_m are known, and we have to estimate them by the eigenvectors $\hat{\nu}_m$ of $\hat{\Gamma}_n$. Under suitable assumptions (compare, e.g., Theorem 16.2 of Horváth and Kokoszka, [28]), we have

$$\limsup_{n\to\infty} n\mathbb{E}\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2 < \infty$$

i.e. by Markov's inequality

$$\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2 = \mathcal{O}_p\left(\frac{1}{n}\right).$$

Here, $\hat{c}_j = sgn\left(\hat{\nu}_j, \nu_j\right)$ has to be added to the relationship due to the inherent nonidentifiability of the eigenbasis (compare the discussion in Section 2.5 of Horváth and Kokoszka, [28]).

Our arguments remain unchanged, except that now we no longer have $\hat{\Psi}_n\left(\nu_m\right) = \hat{\alpha}_m\nu_m$. The corresponding relationship with $\hat{\nu}_m$ instead of ν_m does not hold either, as the coincidence of eigenvectors of Γ and C does not imply that $\hat{\Gamma}_n$ and $\hat{C}_n$ have the same eigenvectors. Nevertheless, we get that the distribution of the centered sample residuals $\hat{\epsilon}_t$ projected onto ν_m converges to the distribution of the corresponding quantity involving the true residuals ϵ_t.

Theorem 5.4.1. *Let $X_0, \cdots, X_n$ be a sample from a stationary FAR(1) process satisfying*

i) $\mathbb{E}\left\|\epsilon_t\right\|^4 < \infty,\ \mathbb{E}\epsilon_t = 0,\ \epsilon_t$ *i.i.d.*

ii) $\left\|\Psi\right\|_{\mathcal{L}} < 1$

iii) C *has the singular value decomposition* $\boldsymbol{S}$ *with the eigenfunctions* ν_j *of* Γ.

iv) The eigenvalues $\lambda_1 > \lambda_2 > \cdots$ *of* Γ *have multiplicity 1 and satisfy* $\lambda_j - \lambda_{j+1} \geq ba^j$ *for some* $0 < a < 1, b > 0$ *and all* $j \geq 1$.

v) $\sup_{j\leq k_n} \frac{\sigma_j\sqrt{2\log\log n}}{\sqrt{n}} \to 0$ *for* $n \to \infty$, *where* $\sigma_j^2 = \mathbb{E}\left\langle\left[\left(I-R\right)^{-1}\Delta\right]\left(\nu_j\right), \nu_j\right\rangle^2$ *with* I *the identity in the space of Hilbert-Schmidt operators on* L^2, R *a linear operator on that space given by* $R\left(S\right) = \Psi S\Psi^*$ *and* Δ *the Hilbert-Schmidt operator given by* $\Delta\left(x\right) = \left\langle\Psi\left(X_0\right), x\right\rangle\epsilon_1 + \left\langle\epsilon_1, x\right\rangle\Psi\left(X_0\right) + \left\langle\epsilon_1, x\right\rangle\epsilon_1 - C_\epsilon\left(x\right)$, C_ϵ *denoting the covariance operator of the innovations* ϵ_t

Then, $d_2\left(\hat{F}_n, F\right) \to 0$ *for* $n \to \infty$ *if* $k_n \to \infty$ *such that*

$$\lim_{n\to\infty} \frac{k_n}{\log\ n} < \frac{1}{4\log\frac{1}{a}}.$$

Assumption iv) is a regularity condition which could be enforced by regularizing the eigenvalue estimators like in Guillas [22]. Assumption v) states that the law of the iterated logarithm which holds for $\hat{\lambda}_j$ by Theorem 4 of Menneteau [45], holds uniformly in $j \leq k_n$. Note that in the definition of R we could use $\Psi^* = \Psi$ which follows from iii), but we prefer to stick to the notation of Menneteau.

Proof. The proof coincides with that of proposition 5.3.1 until we start to evaluate $(\hat{\eta}_j - \eta_j)^2$. Now, we have for $t \geq 1$

$$\begin{aligned}
(\eta_t - \hat{\eta}_t)^2 &= \left(\left\langle \hat{\Psi}_n\left(X_{t-1}\right), \nu_m\right\rangle - \left\langle \Psi\left(X_{t-1}\right), \nu_m\right\rangle\right)^2 \qquad (5.15)\\
&= \left(\left\langle \left(\hat{C}_n - C\right)\hat{\Gamma}_n^{\dagger}\left(X_{t-1}\right), \nu_m\right\rangle + \left\langle C\left(\hat{\Gamma}_n^{\dagger} - \Gamma^{-1}\right)\left(X_{t-1}\right), \nu_m\right\rangle\right)^2\\
&\leq 2\left\|\left(\hat{C}_n - C\right)\hat{\Gamma}_n^{\dagger}\left(X_{t-1}\right)\right\|^2 + 2\left\langle C\left(\hat{\Gamma}_n^{\dagger} - \Gamma^{-1}\right)\left(X_{t-1}\right), \nu_m\right\rangle^2
\end{aligned}$$

using the Cauchy-Schwarz inequality and $\|\nu_m\| = 1$ for the first term.
a) We first consider the first term on the right hand side of (5.15). From the definition of the operator norm $\|\cdot\|_{\mathcal{L}}$ we immediately have the well known inequality

$$\left\|\left(\hat{C}_n - C\right)\hat{\Gamma}_n^{\dagger}\left(X_{t-1}\right)\right\|^2 \leq \left\|\hat{C}_n - C\right\|_{\mathcal{L}}^2 \left\|\hat{\Gamma}_n^{\dagger}\left(X_{t-1}\right)\right\|^2$$

From Theorem 16.1 of Horváth and Kokoszka, [28] we have, as the Hilbert-Schmidt norm always dominates the operator norm, i.e. $\|\cdot\|_{\mathcal{S}} \geq \|\cdot\|_{\mathcal{L}}$, that

$$n\mathbb{E}\left\|\hat{C}_n - C\right\|_{\mathcal{L}}^2 = \mathcal{O}\left(1\right)$$

and hence, by Markov's inequality, $\left\|\hat{C}_n - C\right\|_{\mathcal{L}}^2 = \mathcal{O}_p\left(\frac{1}{n}\right)$. Note that, under our assumptions i), ii), the FAR(1)-process is L^4-m-approximable (see Example 16.1 of Horváth and Kokoszka, [28]).

Now, as $\hat{\nu}_1, \hat{\nu}_2, \cdots$ are orthonormal,

$$\begin{aligned}
\left\|\hat{\Gamma}_n^{\dagger}\left(X_{t-1}\right)\right\|^2 &= \left\|\sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j}\left\langle X_{t-1}, \hat{\nu}_j\right\rangle \hat{\nu}_j\right\|^2 \qquad (5.16)\\
&= \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j^2}\left\langle X_{t-1}, \hat{\nu}_j\right\rangle^2 \qquad (5.17)\\
&\leq \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j^2}\left\|X_{t-1}\right\|^2 \qquad (5.18)
\end{aligned}$$

Moreover, by stationarity,

$$\mathbb{E}\left(\frac{1}{n}\sum_{t=1}^{n}\left\|X_{t-1}\right\|^2\right)=\mathbb{E}\left\|X_0\right\|^2<\infty$$

i.e. $\frac{1}{n}\sum_{t=1}^{n}\left\|X_{t-1}\right\|^2=\mathcal{O}_p\left(1\right)$ by Markov's inequality. Therefore, we finally have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left\|\left(\hat{C}_n-C\right)\hat{\Gamma}_n^{\dagger}\left(X_{t-1}\right)\right\| &\leq \left\|\left(\hat{C}_n-C\right)\right\|_{\mathcal{L}}^2\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}\mathcal{O}_p\left(1\right)\\ &= \mathcal{O}_p\left(\frac{1}{n}\right)\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}\end{aligned}$$

b) Now, we consider the second term on the right hand side of 5.15. From (5.6), we have $\left\langle\Psi\left(X_{t-1}\right),\nu_m\right\rangle=\left\langle C\Gamma^{-1}\left(X_{t-1}\right),\nu_m\right\rangle=\alpha_m\left\langle X_{t-1},\nu_m\right\rangle$. Moreover, from 5.3, C is selfadjoint and

$$\begin{aligned}\left\langle C\Gamma_n^{\dagger}\left(X_{t-1}\right),\nu_m\right\rangle &= \left\langle\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}\left\langle X_{t-1},\hat{\nu}_j\right\rangle\hat{\nu}_j,C\left(\nu_m\right)\right\rangle\\ &= \gamma_m\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}\left\langle X_{t-1},\hat{\nu}_j\right\rangle\left\langle\hat{\nu}_j,\nu_m\right\rangle\end{aligned}$$

From assumptions i), ii) and iv) the assumptions of Theorem 16.2 of Horváth and Kokoszka, [28] are satisfied, and we have in particular from the inequalities preceding said Theorem in [28],

$$n\mathbb{E}\left\|\hat{c}_j\hat{\nu}_j-\nu_j\right\|^2\leq\frac{D}{a^{2j}},\ n\mathbb{E}\left|\hat{\lambda}_j-\lambda_j\right|^2\leq D$$

where D stands for some generic constant independent of $j,\ n$. As $\left|\hat{c}_j\right|=1$, we have, using $\alpha_m=\dfrac{\gamma_m}{\lambda_m}$

$$\begin{aligned}\left\langle C\Gamma_n^{\dagger}\left(X_{t-1}\right)-C\Gamma^{-1}\left(X_{t-1}\right),\nu_m\right\rangle &= \gamma_m\sum_{j=1,\ j\neq m}^{k_n}\frac{1}{\hat{\lambda}_j}\left\langle X_{t-1},\hat{c}_j\hat{\nu}_j\right\rangle\left\langle\hat{c}_j\hat{\nu}_j,\nu_m\right\rangle\\ &\quad+\gamma_m\left(\frac{1}{\hat{\lambda}_m}\left\langle X_{t-1},\hat{c}_m\hat{\nu}_m\right\rangle\left\langle\hat{c}_m\hat{\nu}_m,\nu_m\right\rangle-\frac{1}{\lambda_m}\left\langle X_{t-1},\nu_m\right\rangle\right)\end{aligned}$$

For $j\neq m$, we have, by Cauchy-Schwarz,

$$\mathbb{E}\left\langle\hat{c}_j\hat{\nu}_j,\nu_m\right\rangle^2=\mathbb{E}\left\langle\hat{c}_j\hat{\nu}_j-\nu_j,\nu_m\right\rangle^2\leq\mathbb{E}\left\|\hat{c}_j\hat{\nu}_j-\nu_j\right\|^2\leq\frac{D}{na^{2j}}$$

Analogously, for $j=m$,

$$\mathbb{E}\left(\left\langle\hat{c}_m\hat{\nu}_m,\nu_m\right\rangle-1\right)^2\leq\frac{D}{na^{2m}}$$

Therefore, we have again by Markov's inequality

$$\langle \hat{c}_j \hat{\nu}_j, \nu_m \rangle = \langle \nu_j, \nu_m \rangle + \mathcal{O}_p\left(\frac{1}{a^j \sqrt{n}}\right)$$

uniformly in j. Therefore, we get for the average of the second term of 5.15.

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\left\langle C\left(\hat{\Gamma}_n^{\dagger} - \Gamma^{-1}\right)(X_{t-1})\,\nu_m\right\rangle^2 &= \frac{1}{n}\sum_{t=1}^{n}\left\{\gamma_m \sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}\langle X_{t-1}, \hat{c}_j\hat{\nu}_j\rangle \frac{1}{a^j}\mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)\right. \\
&\quad \left. +\gamma_m\left(\frac{1}{\hat{\lambda}_m}\langle X_{t-1}, \hat{c}_m\hat{\nu}_m\rangle - \frac{1}{\lambda_m}\langle X_{t-1}, \nu_m\rangle\right)\right\}^2 \\
&\leq \frac{2\gamma_m^2}{n}\sum_{t=1}^{n}\left(\sum_{j=1}^{k_n}\frac{1}{a^j\hat{\lambda}_j}\langle X_{t-1}, \hat{c}_j\hat{\nu}_j\rangle\right)^2 \mathcal{O}_p\left(\frac{1}{n}\right) \\
&\quad +\frac{2\gamma_m^2}{n}\sum_{t=1}^{n}\left\{\left(\frac{1}{\hat{\lambda}_m} - \frac{1}{\lambda_m}\right)\langle X_{t-1}, \hat{c}_m\hat{\nu}_m\rangle\right. \\
&\quad \left. +\frac{1}{\lambda_m}\langle X_{t-1}, \hat{c}_m\hat{\nu}_m - \nu_m\rangle\right\}^2 \\
&\leq \frac{2\gamma_m^2 k_n}{n}\sum_{t=1}^{n}\sum_{j=1}^{k_n}\frac{1}{a^{2j}\hat{\lambda}_j^2}\langle X_{t-1}, \hat{c}_j\hat{\nu}_j\rangle^2 \mathcal{O}_p\left(\frac{1}{n}\right) \\
&\quad +4\gamma_m^2\left(\frac{1}{\hat{\lambda}_m} - \frac{1}{\lambda_m}\right)^2 \frac{1}{n}\sum_{t=1}^{n}\langle X_{t-1}, \hat{c}_m\hat{\nu}_m\rangle^2 \\
&\quad +\frac{4\gamma_m^2}{\lambda_m^2}\frac{1}{n}\sum_{t=1}^{n}\langle X_{t-1}, \hat{c}_m\hat{\nu}_m - \nu_m\rangle^2
\end{aligned}$$

where we use Jensen's inequality for the last inequality.

Now, again by Cauchy-Schwarz and stationarity of X_t

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\langle X_{t-1}, \hat{c}_m\hat{\nu}_m - \nu_m\rangle^2 &\leq \frac{1}{n}\sum_{t=1}^{n}\|X_{t-1}\|^2\,\|\hat{c}_m\hat{\nu}_m - \nu_m\|^2 \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\|X_{t-1}\|^2\,\mathcal{O}_p\left(\frac{1}{na^{2m}}\right) = \mathcal{O}_p\left(\frac{1}{na^{2m}}\right)
\end{aligned}$$

as $\frac{1}{n}\sum_{t=1}^{n}\|X_{t-1}\|^2 = \mathcal{O}_p(1)$ as discussed above. As $\|\hat{c}_m\hat{\nu}_m\| = 1$,

$$\frac{1}{n}\sum_{t=1}^{n}\langle X_{t-1}, \hat{c}_m\hat{\nu}_m\rangle^2 \leq \frac{1}{n}\sum_{t=1}^{n}\|X_{t-1}\|^2 = \mathcal{O}_p(1)$$

and

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{k_n}\frac{1}{a^{2j}\hat{\lambda}_j^2}\langle X_{t-1}, \hat{c}_j\hat{\nu}_j\rangle^2 \leq \sum_{j=1}^{k_n}\frac{1}{a^{2j}\hat{\lambda}_j^2}\frac{1}{n}\sum_{t=1}^{n}\|X_{t-1}\|^2 = \sum_{j=1}^{k_n}\frac{1}{a^{2j}\hat{\lambda}_j^2}\mathcal{O}_p(1)$$

such that

$$\frac{1}{n}\sum_{t=1}^{n}\left\langle C\left(\hat{\Gamma}_n^{\dagger}-\Gamma^{-1}\right)(X_{t-1}),\nu_m\right\rangle^2 = k_n\sum_{j=1}^{k_n}\frac{1}{a^{2j}\hat{\lambda}_j^2}\mathcal{O}_p\left(\frac{1}{n}\right) + 4\gamma_m^2\left(\frac{1}{\hat{\lambda}_m}-\frac{1}{\lambda_m}\right)^2\mathcal{O}_p(1)+\mathcal{O}_p\left(\frac{1}{na^{2m}}\right)$$

c) As the next step, we have to investigate

$$\frac{1}{\hat{\lambda}_j}-\frac{1}{\lambda_j}=\frac{\lambda_j-\hat{\lambda}_j}{\hat{\lambda}_j\lambda_j}$$

From assumption iv) we in particular have $\lambda_j \geq a^j$. As all eigenvalues have multiplicity 1, we have from Theorem 4 of Mas and Menneteau, [43], a law of the iterated logarithm stating that $\sqrt{\frac{n}{2\ \log\log\ n}}\left(\hat{\lambda}_j-\lambda_j\right)$ is relatively compact with limit set $[-\sigma_j,\sigma_j]$. In particular, for all large enough n and some $\delta>0$,

$$\left|\hat{\lambda}_j-\lambda_j\right| \leq (\sigma_j+\delta)\sqrt{\frac{2\ \log\log\ n}{n}}\ \text{a.s,}$$

We choose n large enough such that $\left|\hat{\lambda}_j-\lambda_j\right| \leq \frac{1}{2}\lambda_j$ a.s. for all $j \leq k_n$ using assumption v), i.e.

$$\frac{1}{\hat{\lambda}_j}=\frac{1}{\lambda_j+\left(\hat{\lambda}_j-\lambda_j\right)} \leq \frac{1}{\left|\lambda_j-|\hat{\lambda}_j-\lambda_j|\right|} \leq \frac{2}{\lambda_j}\ \text{a.s.}$$

Therefore, using $\lambda_j \geq a^j$,

$$\begin{aligned} k_n\sum_{j=1}^{k_n}\frac{1}{a^{2j}\hat{\lambda}_j^2} &\leq k_n\sum_{j=1}^{k_n}\frac{4}{a^{2j}\lambda_j^2} \leq 4k_n\sum_{j=1}^{k_n}\frac{1}{a^{4j}} = \frac{4k_n}{a^{4k_n}}\sum_{j=0}^{k_n-1}\left(a^4\right)^j \\ &= \frac{4k_n}{a^{4k_n}}\frac{1-a^{4k_n}}{1-a^4} \leq \frac{4k_n}{(1-a^4)\,a^{4k_n}} \end{aligned}$$

Analogously,

$$\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2} \leq \frac{4}{(1-a^2)\,a^{2k_n}}$$

Finally,

$$\left(\frac{1}{\hat{\lambda}_m}-\frac{1}{\lambda_m}\right)^2=\frac{\left(\hat{\lambda}_m-\lambda_m\right)^2}{\hat{\lambda}_m^2\lambda_m^2} \leq \frac{4\left(\hat{\lambda}_m-\lambda_m\right)^2}{\lambda_m^4}=\mathcal{O}_p\left(\frac{1}{n}\right)$$

as $n\mathbb{E}\left(\hat{\lambda}_m-\lambda_m\right)^2 \leq D$ by the remark after Theorem 16.1 of Horváth and Kokoszka, [28].

d) Combining a), b), c), we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left(\eta_t-\hat{\eta}_t\right)^2 &= \mathcal{O}_p\left(\frac{1}{na^{2k_n}}\right)+\mathcal{O}_p\left(\frac{k_n}{na^{4k_n}}\right)+\mathcal{O}_p\left(\frac{1}{n}\right)\\ &= \mathcal{O}_p\left(\frac{k_n}{na^{4k_n}}\right)\xrightarrow{p}0 \text{ for } n\to\infty\end{aligned}$$

Under our assumptions on the rate of $k_n \to \infty$. It follows $d_2^2\left(F_n, \hat{F}_n\right) \xrightarrow{p} 0$ with the same rate and finally $d_2^2\left(\hat{F}_n, F\right) \xrightarrow{p} 0$, which ends the proof. □

We do not consider this special case further. It already shows one of the main difficulties of the general case, which is due to the fact that $\hat{\Gamma}_n$ and $\hat{\Gamma}_n^\dagger$ do not have the same eigenvectors in contrast to the toy example of the previous section. This has nothing to do with the special form of C. Therefore, we immediately study general FAR(1)-processes in the next chapter.

Chapter 6

Bootstrap for FAR - Some Theory

In this chapter we consider the bootstrap for the general FAR(1)-model. We start with some auxiliary results, then prove an analogue of the weak convergence result of Mas [44] for the simpler case of fixed $k_n = p$ using the results of [47] on triangular arrays of Hilbert space-valued random variables. Then, we show that the bootstrap innovations approximate the true innovations in Mallows metric. In the next two sections, we show that the bootstrap principle holds for the sample mean and the sample covariance operator. Finally we discuss the problems of bootstrapping the estimate of the autoregressive operator Ψ when it is projected onto a finite random subspace with fixed dimension $k_n = p$ which is independent of the sample size n.

6.1 Some auxiliary results

We start with introducing some notation and properties which we shall need in the rest of this chapter. We consider data $X_0, \ldots, X_n$ from a stationary FAR(1)-process with autoregressive operator Ψ, covariance operator $\Gamma = \mathbb{E}X_t \otimes X_t$ and lag-1-covariance operator $C = \mathbb{E}X_t \otimes X_{t+1}$. We use $\hat{\Gamma}_n, \hat{C}_n$ for the sample versions of Γ, C. λ_j resp. ν_j are the eigenvalues resp. eigenfunctions of Γ, and $\hat{\lambda}_j, \hat{\nu}_j$ their estimates. Then,

$$\hat{\Psi}_n = \hat{C}_n \hat{\Gamma}_n^\dagger, \quad \text{with} \quad \hat{\Gamma}_n^\dagger = \sum_{j=1}^{p} \frac{1}{\hat{\lambda}_j} \hat{\nu}_j \otimes \hat{\nu}_j$$

is the estimate of Ψ, where p may be fixed or depend on n depending on the circumstances.

$$\Pi_p = \sum_{j=1}^{p} \nu_j \otimes \nu_j, \qquad \hat{\Pi}_p = \sum_{j=1}^{p} \hat{\nu}_j \otimes \hat{\nu}_j$$

denote the projections onto the span of the first p eigenfunctions resp. empirical eigenfunctions.

Lemma 9. $\hat{\Pi}_p = \hat{\Gamma}_n \hat{\Gamma}_n^\dagger = \hat{\Gamma}_n^\dagger \hat{\Gamma}_n,\ \hat{\Psi}_n \hat{\Pi}_p = \hat{\Psi}_n$

Proof.

$$\begin{aligned}
\hat{\Gamma}_n \hat{\Gamma}_n^\dagger (x) &= \hat{\Gamma}_n \left(\sum_{k=1}^{p} \frac{1}{\hat{\lambda}_k} \langle \hat{\nu}_k, x \rangle \hat{\nu}_k \right) \\
&= \sum_{k=1}^{p} \frac{1}{\hat{\lambda}_k} \langle \hat{\nu}_k, x \rangle \hat{\Gamma}_n (\hat{\nu}_k) \\
&= \sum_{k=1}^{p} \hat{\nu}_k \otimes \hat{\nu}_k (x) = \hat{\Pi}_p (x)
\end{aligned}$$

as $\hat{\Gamma}_n (\hat{\nu}_k) = \hat{\lambda}_k \hat{\nu}_k$. Analogously, we get $\hat{\Gamma}_n^\dagger \hat{\Gamma}_n (x) = \hat{\Pi}_p (x)$. Now,

$$\hat{\Psi}_n (x) = \hat{C}_n \hat{\Gamma}_n^\dagger (x) = \sum_{k=1}^{p} \frac{1}{\hat{\lambda}_k} \langle \hat{\nu}_k, x \rangle \hat{C}_n (\hat{\nu}_k)$$

$$\begin{aligned}
\hat{\Psi}_n \hat{\Pi}_p (x) &= \hat{\Psi}_n \left(\sum_{j=1}^{p} \langle \hat{\nu}_j, x \rangle \hat{\nu}_j \right) \\
&= \sum_{j=1}^{p} \langle \hat{\nu}_j, x \rangle \hat{\Psi}_n (\hat{\nu}_j) \\
&= \sum_{j=1}^{p} \langle \hat{\nu}_j, x \rangle \frac{1}{\hat{\lambda}_j} \hat{C}_n (\hat{\nu}_j)
\end{aligned}$$

as $\hat{\Gamma}_n^\dagger (\hat{\nu}_j) = \frac{1}{\hat{\lambda}_j} \hat{\nu}_j$. □

Lemma 10. $\hat{\Psi}_n - \Psi \hat{\Pi}_p = \frac{1}{n} S_n \hat{\Gamma}_n^\dagger$ *with* $S_n = \sum_{t=1}^{n} X_{t-1} \otimes \epsilon_t = n \left(\hat{C}_n - \Psi \hat{\Gamma}_n \right)$

Proof. From Lemma 9, we have $\hat{\Psi}_n - \Psi\hat{\Pi}_p = \left(\hat{C}_n - \Psi\hat{\Gamma}_n\right)\hat{\Gamma}_n^{\dagger}$

$$\begin{aligned} n\left(\hat{C}_n - \Psi\hat{\Gamma}_n\right)(x) &= \sum_{t=1}^{n} X_{t-1} \otimes X_t(x) - \Psi\left(\sum_{t=1}^{n} X_{t-1} \otimes X_{t-1}(x)\right) \\ &= \sum_{t=1}^{n} \langle X_{t-1}, x\rangle X_t - \sum_{t=1}^{n} \Psi\left(\langle X_{t-1}, x\rangle X_{t-1}\right) \\ &= \sum_{t=1}^{n} \langle X_{t-1}, x\rangle \left(X_t - \Psi\left(X_{t-1}\right)\right) = \sum_{t=1}^{n} X_{t-1} \otimes \epsilon_t(x) \end{aligned}$$

□

Both Lemmas are essentially used already by Mas, [44]. We give their proofs here for the sake of completeness.

Lemma 11. *Let $\{X_t\}$ be a L^2-valued time series with $\mathbb{E}\|X_t\|^4 < \infty$, which is $L^4 - m$-approximable. Let the eigenvalues $\lambda_j,\ j \geq 1$ of Γ satisfy $\lambda_1 > \lambda_2 > \ldots > \lambda_{p+1}$ and all have multiplicity 1. Then, for any L^2-valued random variable Y with $\mathbb{E}\|Y\|^2 < \infty$, which is independent of $X_0, \ldots, X_{n-1}$, we have for some constant D*

$$\mathbb{E}\left\|\hat{\Pi}_p(Y) - \Pi_p(Y)\right\|^2 \leq D\mathbb{E}\|Y\|^2 \frac{p}{n}\sum_{j=1}^{p}\frac{1}{a_j^2}$$

where $a_1 = \lambda_1 - \lambda_2,\ a_j = \min\left(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\right),\ j = 2, \ldots, p.$

Proof. Let $\hat{c}_j = sgn\langle\hat{\nu}_j, \nu_j\rangle$, where $\nu_1, \ldots, \nu_p,\ \hat{\nu}_1, \ldots, \hat{\nu}_p$ are the first p eigenvectors of Γ respectively $\hat{\Gamma}_n$. Then,

$$\begin{aligned} \mathbb{E}\|\langle Y, \hat{c}_j\hat{\nu}_j\rangle \hat{c}_j\hat{\nu}_j - \langle Y, \nu_j\rangle \nu_j\|^2 &\leq 2\mathbb{E}\|\langle Y, \hat{c}_j\hat{\nu}_j\rangle (\hat{c}_j\hat{\nu}_j - \nu_j)\|^2 + 2\mathbb{E}\|\langle Y, \hat{c}_j\hat{\nu}_j - \nu_j\rangle \nu_j\|^2 \\ &\leq 4\mathbb{E}\left(\|Y\|^2 \|\hat{c}_j\hat{\nu}_j - \nu_j\|^2\right) \end{aligned}$$

using Cauchy-Schwarz and $\|\hat{\nu}_j\| = \|\nu_j\| = 1$. Hence,

$$\begin{aligned} \mathbb{E}\left\|\hat{\Pi}_p(Y) - \Pi_p(Y)\right\|^2 &= \mathbb{E}\left\|\sum_{j=1}^{p}\langle Y, \hat{c}_j\hat{\nu}_j\rangle \hat{c}_j\hat{\nu}_j - \sum_{j=1}^{p}\langle Y, \nu_j\rangle \nu_j\right\|^2 \\ &\leq 4p\sum_{j=1}^{p}\mathbb{E}\left(\|Y\|^2 \|\hat{c}_j\hat{\nu}_j - \nu_j\|^2\right) \\ &\leq Dp\mathbb{E}\|Y\|^2 \sum_{j=1}^{p}\frac{1}{na_j^2} \end{aligned}$$

by the remarks following Theorem 16.1 of Horvath and Kokoszka, [28]. □

Corollary 6.1.1. *If, under the assumptions of Lemma 11, $\lambda_j - \lambda_{j+1} \geq a^j b$ for some $0 < a < 1$ and all $j = 2, \ldots, p+1$, and some constant $b > 0$, then*

$$\mathbb{E}\left\|\hat{\Pi}_p(Y) - \Pi_p(Y)\right\|^2 \leq D\mathbb{E}\|Y\|^2 \frac{p}{na^{2p}} \frac{1}{1-a^2}$$

Proof. We have $a_j \geq a^j b$, and

$$\sum_{j=1}^{p} \frac{1}{a^{2j}} = \frac{1}{a^{2p}} \frac{1-a^{2p}}{1-a^2} \leq \frac{1}{a^{2p}} \frac{1}{1-a^2}$$

□

6.2 Weak convergence for fixed p

We consider the FAR(1) model given by

$$X_{t+1} = \Psi X_t + \epsilon_{t+1}.$$

Regarding Lemma 10 we have

$$\hat{\Psi}_n - \Psi\hat{\Pi}_p = \frac{1}{n} S_n \hat{\Gamma}_p^\dagger,$$

where again $S_n(x) = \sum_{t=1}^n X_{t-1} \otimes \epsilon_t(x)$, and we are interested in showing that for fixed p

$$\sqrt{n}\left(\hat{\Psi}_n - \Psi\hat{\Pi}_p\right) = \frac{1}{\sqrt{n}} S_n \hat{\Gamma}_p^\dagger \xrightarrow{w} S_\infty \Gamma_p^\dagger, \tag{6.1}$$

where S_∞ is Gaussian and

$$\Gamma_p^\dagger = \sum_{j=1}^{p} \frac{1}{\lambda_j} \nu_j \otimes \nu_j.$$

Note that this is the analogue of Theorem 3.1 of Mas, [44], who considers $p = k_n$ increasing with n. However, the proof is different and simpler for the simpler kind of asymptotics with fixed p. We investigate S_n and $\hat{\Gamma}_p^\dagger$ separately and then apply an argument like Slutsky's Lemma to conclude the proof. We first compute the mean and covariance operator of $\frac{1}{\sqrt{n}} S_n$. Note that

$$S_n(x) = \sum_{j=1}^{n} Y_j(x) \qquad \text{with} \quad Y_t = X_{t-1} \otimes \epsilon_t(x) =< X_{t-1}, x > \epsilon_t$$

As ϵ_t is independent of X_{t-1} and has mean 0, the $Y_t(x)$ are white noise, i.e. pairwise uncorrelated, with mean 0. From this we get

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathbb{E}S_n(z) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}Y_t(z) = 0 \\
\frac{1}{n}\mathbb{E}\{S_t(x)\otimes S_t(x)\}(z) &= \frac{1}{n}\mathbb{E}\left\langle S_t(x), z\right\rangle S_t(x) \\
&= \mathbb{E}\left\langle \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\langle X_{t-1}, x\right\rangle \epsilon_t \ , z \right\rangle \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\langle X_{k-1}, x\right\rangle \epsilon_k \\
&= \frac{1}{n}\sum_{k,t}\mathbb{E}\left\langle X_{t-1}, x\right\rangle \left\langle X_{k-1}, x\right\rangle \epsilon_t \otimes \epsilon_k\,(z) \\
&= \mathbb{E}\left\langle X_1, x\right\rangle^2 \mathbb{E}\epsilon_2\otimes\epsilon_2\,(z) = \mathbb{E}\left\langle X_1, x\right\rangle^2 \Gamma_\epsilon\,(z) \qquad (6.2)
\end{aligned}$$

with Γ_ϵ denoting the covariance operator of the innovations, using stationarity of X_t and independence of the ϵ_t's for $k \neq t$.

We have from our assumptions that $X_0, \ldots, X_n$ are from a strictly stationary process with zero mean and finite second moment. We therefore can conclude that $\mathbb{E}\left\langle X_1, x\right\rangle^2 \leq ||x||^2\mathbb{E}||X_1||^2$ is bounded. We now want to check the assumptions of Theorem 5.2.1. First, we note that $S_n(x)$ is the sum of the pairwise uncorrelated random variables $Y_t(x)$, which are, moreover, strictly stationary by stationarity of X_t. As the Y_t are pairwise uncorrelated, $\mathbb{E} < Y_1(x), Y_t(x) >= 0$ for $t > 1$ and $\mathbb{E} < Y_1(x), Y_1(x) >= \mathbb{E} < X_0, x >^2 \mathbb{E}||\epsilon_1||^2$, i.e. the long-run variance exists and is just $\mathbb{E} < X_{t-1}, x >^2 \mathbb{E}||\epsilon_t||^2$. Therefore, condition (2.1) of [47] is trivially satisfied, and we conclude from their Theorem 2.1 that $\frac{1}{\sqrt{n}}S_n(x)$ is weakly compact, and a subsequence converges in distribution to a limit with mean 0 and covariance given by (6.2).

To get Gaussianity of the limit, we assume a mixing condition like in Theorem 2.3 of [47] or Theorem 2.17 of [7]. We combine these results in the following proposition.

Proposition 6.2.1. *Let $X_0, \ldots, X_n$ be a sample from a stationary FAR(1) process with i.i.d. innovations ϵ_t, having mean 0 and covariance operator Γ_ϵ. Assume $\mathbb{E}||X_t||^2 < \infty$.*

i) $\frac{1}{\sqrt{n}}S_n(x) \xrightarrow{w} S_\infty(x)$
where $S_\infty(x)$ has mean 0 and covariance operator $\mathbb{E}\left\langle X_1, x\right\rangle^2 \Gamma_\epsilon$.

ii) If $\mathbb{E}||X_t||^{2+\delta} < \infty$ for some $\delta > 0$, and if the time series $X_t, -\infty < t < \infty$ is strongly

mixing with mixing coefficients $\alpha_j, j \geq 1$, *satisfying* $\sum_j |\alpha_j|^{\delta/(2+\delta)} < \infty$, *then* $S_\infty(x)$ *is Gaussian.*

Proof. i) We have already calculated the mean and the covariance operator of $\frac{1}{\sqrt{n}} S_n(x)$ above which both do not depend on n. The assertion then follows from Theorem 5.2.1, which is Theorem 2.1 of Politis and Romano, [47] for the conditions of said result are fulfilled as we have discussed above.

ii) The assertion follows from Theorem 2.3 of [47] as, due to independence of the ϵ_t

$$\mathbb{E}||Y_t(x)||^{2+\delta} = \mathbb{E}|< X_{t-1}, x >|^{2+\delta} \mathbb{E}||\epsilon_t||^{2+\delta} \leq \mathbb{E}||X_t||^{2+\delta} ||x||^{2+\delta} \mathbb{E}||\epsilon_t||^{2+\delta}$$

by Cauchy-Schwarz, and

$$\mathbb{E}||\epsilon_t||^{2+\delta} = \mathbb{E}||X_t - \Psi(X_{t-1})||^{2+\delta} \leq \left(\mathbb{E}||X_t - \Psi(X_{t-1})||^2\right)^{(2+\delta)/2} \leq c\mathbb{E}||X_t||^{2+\delta}$$

for some constant c, using Hölder's inequality and $||\Psi(X_{t-1})|| \leq ||\Psi||_{\mathcal{L}} ||X_{t-1}||$. Moreover,

$$Y_t(x) = < X_{t-1}, x > (X_t - \Psi(X_{t-1}))$$

is a measurable function of finitely many X_t and, hence, has the same mixing rate as X_t. □

Let us remark that, in the mixing case, we would get ii) with a somewhat weaker condition on the quantile function of $||X_t||$ from Theorem 2.17 of Bosq, [7]. Alternatively to assuming mixing, we could use that due to the independence of ϵ_t from $X_{t-1}, X_{t-2}, \ldots$

$$\mathbb{E}\{Y_t(x) | X_{t-1}, X_{t-2}, \ldots\} = 0,$$

i.e. the $Y_t(x)$ are martingale differences. Then, under some additional assumptions, $\frac{1}{\sqrt{n}} S_n(x)$ is also asymptotically normal from Theorem 2.16 of [7].

From an argument similar to Slutsky's Lemma, we can conclude our main result of this section, the asymptotic normality of $\frac{1}{\sqrt{n}} S_n \hat{\Gamma}_n^\dagger(x)$, from the following auxiliary result.

Proposition 6.2.2. *Under the assumptions of Theorem 6.3.1 below, we have uniformly in* x

$$\mathbb{E}||\hat{\Gamma}_n^\dagger(x) - \Gamma_p^\dagger(x)||^2 = ||x||^2 \mathcal{O}(\frac{1}{n})$$

Proof. Note first that from the proof of Theorem 6.3.1 below and by our assumption on the rate of the eigenvalues we have, as p is fixed,

$$\hat{\lambda}_j \geq \frac{1}{2}\lambda_j \geq ba^j, \ j = 1, \ldots, p$$

for all large enough n. Then, with $\hat{c}_j = sgn\left(\langle \hat{\nu}_j, \nu_j \rangle\right)$,

$$\begin{aligned}
\mathbb{E}\left\|\hat{\Gamma}^{\dagger}_{n,p}(x) - \Gamma^{\dagger}_{p}(x)\right\|^2 &= \mathbb{E}\left\|\sum_{j=1}^{p} \frac{1}{\hat{\lambda}_j}\langle \hat{\nu}_j, x\rangle \hat{\nu}_j - \sum_{j=1}^{p}\frac{1}{\lambda_j}\langle \nu_j, x\rangle \nu_j\right\|^2 \\
&\leq 3\mathbb{E}\left\|\sum_{j=1}^{p}\left(\frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j}\right)\langle \hat{\nu}_j, x\rangle \hat{\nu}_j\right\|^2 \\
&\quad +3\mathbb{E}\left\|\sum_{j=1}^{p}\frac{1}{\lambda_j}\langle \hat{c}_j\hat{\nu}_j, x\rangle\left(\hat{c}_j\hat{\nu}_j - \nu_j\right)\right\|^2 \\
&\quad +3\mathbb{E}\left\|\sum_{j=1}^{p}\frac{1}{\lambda_j}\langle \hat{c}_j\hat{\nu}_j - \nu_j, x\rangle \nu_j\right\|^2 \\
&= 3\mathbb{E}\sum_{j=1}^{p}\frac{\left(\hat{\lambda}_j - \lambda_j\right)^2}{\hat{\lambda}_j^2\lambda_j^2}\langle \hat{\nu}_j, x\rangle^2 \\
&\quad +3\mathbb{E}\left\|\sum_{j=1}^{p}\frac{1}{\lambda_j}\langle \hat{c}_j\hat{\nu}_j, x\rangle\left(\hat{c}_j\hat{\nu}_j - \nu_j\right)\right\|^2 \\
&\quad +3\mathbb{E}\sum_{j=1}^{p}\frac{1}{\lambda_j^2}\langle \hat{c}_j\hat{\nu}_j - \nu_j, x\rangle^2
\end{aligned}$$

The first term on the right-hand side is bounded by, using $\langle \hat{\nu}_j, x\rangle^2 \leq \|x\|^2$ by Cauchy-Schwarz,

$3\sum_{j=1}^{p}\frac{\mathbb{E}\left(\hat{\lambda}_j - \lambda_j\right)^2}{b^4a^{4j}}\|x\|^2 = \|x\|^2\mathcal{O}\left(\frac{1}{n}\right)$

by the remarks after Theorem 16.1 of Horvath and Kokoszka, [28]. The third term on the right-hand side is bounded by, again using Cauchy-Schwarz

$$3\sum_{j=1}^{p}\frac{1}{b^2a^{2j}}\mathbb{E}\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2\|x\|^2 \leq D\sum_{j=1}^{p}\frac{1}{b^4a^{4j}}\frac{1}{n}\|x\|^2 = \|x\|^2\mathcal{O}\left(\frac{1}{n}\right)$$

from the same remarks of Horvath and Kokoszka, [28] for some constant D. Finally, the second term is bounded by

$$3p\mathbb{E}\sum_{j=1}^{p}\left\|\frac{1}{\lambda_j}\langle \hat{c}_j\hat{\nu}_j, x\rangle\left(\hat{c}_j\hat{\nu}_j - \nu_j\right)\right\|^2 \leq 3p\sum_{j=1}^{p}\frac{1}{b^2a^{2j}}\|x\|^2\mathbb{E}\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2 = \|x\|^2\mathcal{O}\left(\frac{1}{n}\right)$$

by the same argument. □

Combining the two propositions, we get

Theorem 6.2.1. *Under the assumptions of Theorem 6.3.1 we have for fixed p*
i) With S_∞ as in Proposition 6.2.1,

$$\sqrt{n}(\hat{\Psi}_n - \Psi\hat{\Pi}_p)(z) = \frac{1}{\sqrt{n}}S_n\hat{\Gamma}^{\dagger}_n(z) \xrightarrow{w} S_\infty\Gamma^{\dagger}_p(z)$$

ii) If additionally the mixing and moment conditions of Proposition 6.2.1, ii) are satisfied, we have that $S_\infty \Gamma_p^\dagger(z)$ *is Gaussian with mean 0 and covariance operator* $\mathbb{E}\left\langle X_1, \Gamma_p^\dagger(z)\right\rangle^2 \Gamma_\epsilon$.

Proof. With $x = \Gamma_p^\dagger(z)$, we have immediately from Proposition 6.2.1 the results for $S_n\Gamma_p^\dagger(z)$ instead of $S_n\hat{\Gamma}_n^\dagger(z)$. So, it suffices to show that $S_n\hat{\Gamma}_n^\dagger(z) - S_n\Gamma_p^\dagger(z) = o_p(\sqrt{n})$. We have, recalling $||S_n||_{\mathcal{L}} = O_p(\sqrt{n})$ from the proof of Theorem 6.3.1

$$||S_n(\hat{\Gamma}_n^\dagger(z) - \Gamma_p^\dagger(z))|| \leq ||S_n||_{\mathcal{L}}\, ||\hat{\Gamma}_n^\dagger(z) - \Gamma_p^\dagger(z)|| = \mathcal{O}_p(1) = o_p(\sqrt{n})$$

applying Proposition 6.2.2. □

6.3 Approximation of the innovation distribution by the empirical measure of sample residuals

The basis for residual-based bootstrapping in scalar regression and autoregression models is the approximability of the innovations by the bootstrap innovations where the latter are drawn from the centered sample residuals. This is stated in the following theorem in terms of the Mallows metric. Note that now we have to allow k_n to increase as, otherwise, $\hat{\Psi}_n$ would not be a consistent estimate of Ψ and we could not expect the sample residuals to mimic the distribution of the true innovations.

Theorem 6.3.1. *Let* $X_0, \ldots, X_n$ *be a sample from a stationary FAR(1) process satisfying*

i) $\mathbb{E}\,\|\epsilon_t\|^4 < \infty,\ \mathbb{E}\epsilon_t = 0,\ \{\epsilon_t\}$ *i.i.d.*

ii) Ψ *is a Hilbert-Schmidt operator with* $\|\Psi\|_{\mathcal{L}} < 1$

iii) The eigenvalues $\lambda_1 > \lambda_2 > \ldots$ *of* Γ *have multiplicity 1 and satisfy* $\lambda_j - \lambda_{j+1} \geq ba^j$ *for some* $0 < a < 1$ *and some* $b > 0$ *for all* $j \geq 1$.

Let $F, \hat{F}_n$ *be the distribution of* ϵ_j *respectively the empirical distribution of* $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_n$ *with*

$$\tilde{\epsilon}_j = \hat{\epsilon}_j - \frac{1}{n}\sum_{k=1}^{n} \hat{\epsilon}_k, \quad \hat{\epsilon}_j = X_j - \hat{\Psi}_n\left(X_{j-1}\right), \quad j = 1, \ldots, n.$$

Then,

$$d_2\left(\hat{F}_n, F\right) \underset{p}{\to} 0,\ n \to \infty$$

if $k_n < \dfrac{1}{4\log\dfrac{1}{a}} \log\ n.$

Proof. Let F_n denote the empirical distribution of $\epsilon_1, \ldots, \epsilon_n$. Then, again from Lemma 8.4 of Bickel and Freedman, [6], we have $d_2(F_n, F) \underset{a.s}{\to} 0$. Hence it suffices to show that $d_2\left(F_n, \hat{F}_n\right) \underset{p}{\to} 0$. As in the proof of Proposition 5.3.1, let J be Laplace distributed on $\{1, \ldots, n\}$, i.e. $pr(J = t) = \frac{1}{n},\ 1 \leq t \leq n$, and consider the random variables
$U_0 = \epsilon_J,\ V_0 = \tilde{\epsilon}_J = \hat{\epsilon}_J - \frac{1}{n}\sum_{j=1}^n \hat{\epsilon}_j$
with distributions F_n respectively $\hat{F}_n$. As in the proof of Theorem 3.1 of Franke and Kreiss, [20], we have from the definition of the Mallows metric

$$\begin{aligned} d_2^2\left(F_n, \hat{F}_n\right) &\leq \mathbb{E}\left\|U_0 - V_0\right\|^2 = \frac{1}{n}\sum_{k=1}^n \left\|\hat{\epsilon}_k - \epsilon_k - \frac{1}{n}\sum_{j=1}^n \hat{\epsilon}_j\right\|^2 \\ &\leq \frac{6}{n}\sum_{k=1}^n \left\|\hat{\epsilon}_k - \epsilon_k\right\|^2 + \frac{3}{n^2}\left\|\sum_{j=1}^n \epsilon_j\right\|^2 \end{aligned}$$

From the law of large numbers for i.i.d. random variables we have

$$\frac{1}{n}\sum_{j=1}^n \epsilon_j \underset{p}{\to} \mathbb{E}\epsilon_j = 0,\ n \to \infty$$

such that the second term on the right-hand side vanishes for $n \to \infty$. For the first term, we show in the following

$$\left\|\hat{\epsilon}_t - \epsilon_t\right\|^2 \leq \left\|X_{t-1}\right\|^2 R_n + 3\left\|\Pi_{k_n}(X_{t-1}) - X_{t-1}\right\|^2$$

where R_n does not depend on t, and $R_n \underset{p}{\to} 0$. Π_{k_n} denotes the projection on the eigenvectors $\nu_1, \ldots, \nu_{k_n}$ of the covariance operator Γ of $\{X_t\}$ as in Lemma 11. Hence,

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^n \left\|\hat{\epsilon}_t - \epsilon_t\right\|^2 &\leq \frac{1}{n}\sum_{t=1}^n \left\|X_{t-1}\right\|^2 R_n + 3\frac{1}{n}\sum_{t=1}^n \left\|\Pi_{k_n}(X_{t-1}) - X_{t-1}\right\|^2 \\ &\underset{p}{\to} 0,\ n \to \infty, \end{aligned}$$

as $\frac{1}{n}\sum_{t=1}^n \left\|X_{t-1}\right\|^2 \to \mathbb{E}\left\|X_1\right\|^2 < \infty$, and, by stationarity of $\{X_t\}$

$$\mathbb{E}\left(\frac{1}{n}\sum_{t=1}^n \left\|\Pi_{k_n}(X_{t-1}) - X_{t-1}\right\|^2\right) = \mathbb{E}\sum_{j=k_n+1}^{\infty} \langle X_1, \nu_j\rangle^2 \to 0$$

for $k_n \to \infty$, using a monotone convergence argument and $\mathbb{E}\sum_{j=1}^{\infty}\langle X_1, \nu_j\rangle^2 = \mathbb{E}\left\|X_1\right\|^2$.

a) By definition of $\epsilon_t,\ \hat{\epsilon}_t$, we have

$$\begin{aligned}
\left\|\epsilon_t - \hat{\epsilon}_t\right\|^2 &= \left\|X_t - \Psi\left(X_{t-1}\right) - X_t + \hat{\Psi}_n\left(X_{t-1}\right)\right\|^2 \\
&= \left\|\left(\hat{\Psi}_n - \Psi\right)\left(X_{t-1}\right)\right\|^2 \\
&= \left\|\left(\hat{\Psi}_n - \Psi\hat{\Pi}_{k_n}\right)\left(X_{t-1}\right) + \Psi\left(\hat{\Pi}_{k_n} - \Pi_{k_n}\right)\left(X_{t-1}\right) + \Psi\left(\Pi_{k_n}\left(X_{t-1}\right) - X_{t-1}\right)\right\|^2 \\
&\leq 3\left\|\left(\hat{\Psi}_n - \Psi\hat{\Pi}_{k_n}\right)\left(X_{t-1}\right)\right\|^2 + 3\left\|\left(\hat{\Pi}_{k_n} - \Pi_{k_n}\right)\left(X_{t-1}\right)\right\|^2 + 3\left\|\Pi_{k_n}\left(X_{t-1}\right) - X_{t-1}\right\|^2
\end{aligned}$$

using $\|\Psi\|_{\mathcal{L}} < 1$, where $\hat{\Pi}_{k_n}$ is defined as in Lemma 11. We now show that the first and the second terms are bounded in the required manner.

b) We split $\left(\hat{\Pi}_{k_n} - \Pi_{k_n}\right)\left(X_{t-1}\right)$ into two terms

$$\begin{aligned}
\left(\hat{\Pi}_{k_n} - \Pi_{k_n}\right)\left(X_{t-1}\right) &= \sum_{j=1}^{k_n}\left\langle X_{t-1}, \hat{\nu}_j\right\rangle\hat{\nu}_j - \sum_{j=1}^{k_n}\left\langle X_{t-1}, \nu_j\right\rangle\nu_j \\
&= \sum_{j=1}^{k_n}\left\langle X_{t-1}, \hat{c}_j\hat{\nu}_j\right\rangle\left(\hat{c}_j\hat{\nu}_j - \nu_j\right) + \sum_{j=1}^{k_n}\left\langle X_{t-1}, \hat{c}_j\hat{\nu}_j - \nu_j\right\rangle\nu_j
\end{aligned}$$

where, again, $\hat{c}_j = sgn\left\langle\hat{\nu}_j, \nu_j\right\rangle$. As $\nu_1, \nu_2, \ldots$ are orthonormal, we have for the second term

$$\begin{aligned}
\left\|\sum_{j=1}^{k_n}\left\langle X_{t-1}, \hat{c}_j\hat{\nu}_j - \nu_j\right\rangle\nu_j\right\|^2 &= \sum_{j=1}^{k_n}\left\langle X_{t-1}, \hat{c}_j\hat{\nu}_j - \nu_j\right\rangle^2 \\
&\leq \left\|X_{t-1}\right\|^2\sum_{j=1}^{k_n}\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2
\end{aligned}$$

where the right hand side converges to 0 in probability, as, from the remarks after Theorem 16.1 of Horvath and Kokoszka, [28],

$$\mathbb{E}\sum_{j=1}^{k_n}\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2 \leq \frac{1}{n}\sum_{j=1}^{k_n}\frac{D}{a^{2j}} = \frac{D}{na^{2k_n}}\frac{1-a^{2k_n}}{1-a^2} \leq \frac{D'}{na^{2k_n}} \to 0$$

for $n \to \infty$ under the condition on the rate of k_n.

For the first term, we have, as $\|\hat{c}_j\hat{\nu}_j\| = 1$,

$$\begin{aligned}
\left\|\sum_{j=1}^{k_n}\left\langle X_{t-1}, \hat{c}_j\hat{\nu}_j\right\rangle\left(\hat{c}_j\hat{\nu}_j - \nu_j\right)\right\|^2 &\leq k_n\sum_{j=1}^{k_n}\left\langle X_{t-1}, \hat{c}_j\hat{\nu}_j\right\rangle^2\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2 \\
&\leq \left\|X_{t-1}\right\|^2 k_n\sum_{j=1}^{k_n}\left\|\hat{c}_j\hat{\nu}_j - \nu_j\right\|^2
\end{aligned}$$

where again the right hand side converges to 0 in probability as, from above,

$$\mathbb{E}k_n \sum_{j=1}^{n} \|\hat{c}_j\hat{\nu}_j - \nu_j\|^2 \leq \frac{k_n D}{na^{2k_n}} \to 0$$

for $n \to \infty$ under the condition on the rate of k_n.

c) Using Lemma 10, we have

$$\begin{aligned}
\left\|\left(\hat{\Psi}_n - \Psi\hat{\Pi}_{k_n}\right)(X_{t-1})\right\|^2 &= \left\|\frac{1}{n}S_n\hat{\Gamma}_n^{\dagger}(X_{t-1})\right\|^2 \\
&\leq \left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2 \left\|\hat{\Gamma}_n^{\dagger}(X_{t-1})\right\|^2 \\
&\leq \left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2 \|X_{t-1}\|^2 \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j^2}
\end{aligned}$$

as in the proof, part a), of Theorem 5.4.1. Moreover, as $C = \Psi\Gamma$,

$$\begin{aligned}
\left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2 &= \left\|\hat{C}_n - \Psi\hat{\Gamma}_n\right\|^2 \leq 2\left\|\hat{C}_n - C\right\|_{\mathcal{L}}^2 + 2\left\|\Psi\left(\Gamma - \hat{\Gamma}_n\right)\right\|_{\mathcal{L}}^2 \\
&\leq 2\left\|\hat{C}_n - C\right\|_{\mathcal{L}}^2 + 2\left\|\hat{\Gamma}_n - \Gamma\right\|_{\mathcal{L}}^2
\end{aligned}$$

as $\|\Psi\|_{\mathcal{L}} \leq 1$. From the remarks after Theorem 16.1 of Horvath and Kokoszka, [28], we have $\mathbb{E}\left\|\hat{\Gamma}_n - \Gamma\right\|_{\mathcal{L}}^2 = \mathcal{O}\left(\frac{1}{n}\right)$, and from Theorem 3 of Mas and Pumo, [42], analogously $\mathbb{E}\left\|\hat{C}_n - C\right\|_{\mathcal{L}}^2 = \mathcal{O}\left(\frac{1}{n}\right)$. Hence

$$\left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2 = \mathcal{O}_p\left(\frac{1}{n}\right)$$

From assumption iii) and the rate condition on k_n, we have, as $a < 1$,

$$\frac{1}{\lambda_{k_n}} \leq \frac{1}{ba^{k_n}} \leq \frac{1}{b}n^{1/4}$$

From Theorem 4.1 of Bosq, [7], we then have, as $\|\cdot\|_{\mathcal{L}} \leq \|\cdot\|_{\mathcal{S}}$,

$$\frac{1}{\lambda_{k_n}}\left\|\hat{\Gamma}_n - \Gamma\right\|_{\mathcal{L}} \leq \frac{1}{b}n^{1/4}\left\|\hat{\Gamma}_n - \Gamma\right\|_{\mathcal{L}} \underset{a.s.}{\to} 0 \text{ for } n \to \infty$$

Therefore, we have for all large enough n,
$\left\|\hat{\Gamma}_n - \Gamma\right\|_{\mathcal{L}} \leq \frac{1}{2}\lambda_{k_n}$ *a.s.* and, as in the proof of Theorem 8.7 of Bosq, [7],
$\hat{\lambda}_{k_n} \geq \lambda_{k_n} - \left\|\hat{\Gamma}_n - \Gamma\right\|_{\mathcal{L}} \geq \frac{1}{2}\lambda_{k_n}$ *a.s.*

using $\sup_{j\geq 1}\left|\hat{\lambda}_j-\lambda_j\right|\leq\left\|\hat{\Gamma}_n-\Gamma\right\|_{\mathcal{L}}$. Therefore, for large enough n,

$$\begin{aligned}\left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2} &\leq 4\left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2\sum_{j=1}^{k_n}\frac{1}{\lambda_j^2}\\ &\leq 4\left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2\sum_{j=1}^{k_n}\frac{1}{b^2a^{2j}}\\ &\leq \frac{4}{b^2}\left\|\frac{1}{n}S_n\right\|_{\mathcal{L}}^2\frac{1}{a^{2k_n}}\frac{1}{1-a^2}=\mathcal{O}_p\left(\frac{1}{na^{2k_n}}\right)=o_p\left(1\right)\end{aligned}$$

□

Let us remark that the logarithmic rate for k_n in case of exponentially bounded $\lambda_j-\lambda_{j+1}$ like in assumption iii) also appears in the convergence rate results of Guillas, [22].

6.4 Bootstrap for S_n and the sample mean

We have seen in section 6.2 that the asymptotic behaviour of $\hat{\Psi}_n$ is mainly determined by the behaviour of $S_n=n(\hat{C}_n-\Psi\hat{\Gamma}_n)$. In this section we show that the bootstrap works for S_n. The proof of this result can be easily modified to show that the bootstrap works for the sample mean. We start with stating that the well-known strong consistency of $\hat{\Psi}_n$ as an estimate of Ψ in particular holds under our set of assumptions.

Lemma 12. *Under the conditions of Theorem 6.3.1, we have*

$$\left\|\hat{\Psi}_n-\Psi\right\|_{\mathcal{L}}\underset{a.s.}{\to}0,\quad n\to\infty$$

Proof. The result is a slight modification of Theorem 8.7 of Bosq, [7], taking into account that Bosq considers $\tilde{\Psi}_n=\hat{\Pi}_{k_n}\hat{\Psi}_n$ instead of $\hat{\Psi}_n$ as an estimate of Ψ. From the discussion in the proof of Theorem 6.3.1, the conditions of that theorem of Bosq are satisfied. In our notation,

$$\left\|\hat{\Psi}_n-\Psi\right\|_{\mathcal{L}}\leq\left\|\hat{\Psi}_n-\Psi\Pi_{k_n}\right\|_{\mathcal{L}}+\left\|\Psi\left(\Pi_{k_n}-\hat{\Pi}_{k_n}\right)\right\|_{\mathcal{L}}+\left\|\Psi\hat{\Pi}_{k_n}-\Psi\right\|_{\mathcal{L}}$$

From the proof of Theorem 8.7 of Bosq, [7], in particular (8.92), (8.93), the second and third terms converge to 0 a.s.

For the first term, we have in our notation for every x,

$$\begin{aligned}\left(\hat{\Psi}_n - \Psi\Pi_{k_n}\right)(x) &= \hat{C}_n\left(\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}\langle x, \hat{c}_j\hat{\nu}_j\rangle\,\hat{c}_j\hat{\nu}_j\right) - C\left(\sum_{j=1}^{k_n}\frac{1}{\lambda_j}\langle x,\nu_j\rangle\,\nu_j\right)\\ &= \hat{C}_n\left(\sum_{j=1}^{k_n}\left(\frac{1}{\hat{\lambda}_j}-\frac{1}{\lambda_j}\right)\langle x, \hat{c}_j\hat{\nu}_j\rangle\,\hat{c}_j\hat{\nu}_j\right)\\ &\quad +\hat{C}_n\left(\sum_{j=1}^{k_n}\frac{1}{\lambda_j}\left(\langle x, \hat{c}_j\hat{\nu}_j\rangle - \langle x,\nu_j\rangle\right)\hat{c}_j\hat{\nu}_j\right)\\ &\quad +\hat{C}_n\left(\sum_{j=1}^{k_n}\frac{1}{\lambda_j}\langle x,\nu_j\rangle\left(\hat{c}_j\hat{\nu}_j-\nu_j\right)\right)\\ &\quad +\left(\hat{C}_n - C\right)\left(\sum_{j=1}^{k_n}\frac{1}{\lambda_j}\langle x,\nu_j\rangle\,\nu_j\right)\\ &= a_{n1}(x)+a_{n2}(x)+a_{n3}(x)+a_{n4}(x)\end{aligned}$$

with $A_{ni} = \sup_{\|x\|\le 1} a_{ni}(x),\ 1\le i\le 4$, we have $\left\|\hat{\Psi}_n - \Psi\Pi_{k_n}\right\|_{\mathcal{L}} \le \sum_{i=1}^4 A_{ni}$ and, from the proof of Theorem 8.7 of Bosq, [7], (8.84), (8.86), (8.88) and (8.90), we have $A_{ni} \underset{a.s.}{\to} 0$ for $i = 1,\ldots,4$. □

Proposition 6.4.1. *Under the assumptions of Theorem 6.3.1, we have for $n\to\infty$*

$$nd_2^2\left(\hat{C}_n - \Psi\hat{\Gamma}_n(W), \hat{C}_n^* - \hat{\Psi}_n\hat{\Gamma}_n^*(W)\right) = \frac{1}{n}d_2^2\left(S_n(W), S_n^*(W)\right) \underset{p}{\to} 0$$

where W is an arbitrary L^2-valued random variable independent of $X_0', \epsilon_1', \ldots, \epsilon_n', \epsilon_1^*, \ldots, \epsilon_n^*$.

Proof. Again, we follow essentially the proof of Theorem 4.1 of Franke and Kreiss, [20]. As in the proof of Proposition 5.3.2 we choose $(\epsilon_t', \epsilon_t^*)$ i.i.d. such that $L(\epsilon_t') = F,\ L^*(\epsilon_t^*) = \hat{F}_n$ and $\mathbb{E}\|\epsilon_t' - \epsilon_t^*\|^2 = d_2^2\left(F, \hat{F}_n\right)$. Moreover, we choose $X_0^* = X_0'$ in the recursive definition of X_t^*, and we assume $(\epsilon_t', \epsilon_t^*),\ t\ge 1$, independent of X_0'. Then, we have

$$X_t' = \Psi\left(X_{t-1}'\right) + \epsilon_t',\ X_t^* = \hat{\Psi}_n\left(X_{t-1}^*\right) + \epsilon_t^*,\ t\ge 1 \tag{6.3}$$

a) First, we show the following analogue relationship to (4.18) of Franke and Kreiss, [20]:

$$\begin{aligned}\mathbb{E}\left\{\|X_t' - X_t^*\|^2 \mid X_0,\ldots,X_n\right\} &\le D\left(\mathbb{E}\|X_0'\|^2 + \mathbb{E}^*\|\epsilon_1^*\|^2\right)\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}^2\\ &\quad + Dd_2^2\left(F, \hat{F}_n\right)\end{aligned}$$

for some generic constant D, for all n large enough.

We use the recursive definition 6.3 to get

$$X_t' = \Psi^t(X_0') + \sum_{k=0}^{t-1}\Psi^k\left(\epsilon_{t-k}'\right),\ X_t^* = \hat{\Psi}_n^t(X_0') + \sum_{k=0}^{t-1}\hat{\Psi}_n^k\left(\epsilon_{t-k}^*\right)$$

First, we note that

$$\Psi^k - \hat{\Psi}_n^k = \left(\Psi - \hat{\Psi}_n\right) \sum_{j=0}^{k-1} \Psi^j \hat{\Psi}_n^{k-1-j}$$

From Lemma 12, we get for $\|\Psi\|_{\mathcal{L}} < \hat{\delta} < 1$

$$\left\|\hat{\Psi}_n\right\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{L}} + \left\|\hat{\Psi}_n - \Psi\right\|_{\mathcal{L}} \leq \hat{\delta}$$

for large enough n and

$$\begin{aligned}\left\|\sum_{j=0}^{k-1} \Psi^j \hat{\Psi}_n^{k-1-j}\right\|_{\mathcal{L}} &\leq \sum_{j=0}^{k-1} \left\|\Psi^j\right\| \left\|\hat{\Psi}_n^{k-1-j}\right\| \leq \sum_{j=0}^{k-1} \|\Psi\|^j \left\|\hat{\Psi}_n\right\|^{k-1-j} \\ &\leq \hat{\delta}^{k-1} \sum_{j=0}^{k-1} \left(\frac{\|\Psi\|}{\hat{\delta}}\right)^j \leq \frac{\hat{\delta}^{k-1}}{1 - \frac{\|\Psi\|}{\hat{\delta}}} = \frac{\hat{\delta}^k}{\hat{\delta} - \|\Psi\|}\end{aligned}$$

i.e. we have for large enough n

$$\left\|\Psi^k - \hat{\Psi}_n^k\right\|_{\mathcal{L}} \leq \frac{\hat{\delta}^k}{\hat{\delta} - \|\Psi\|} \left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}$$

Now, writing $\mathbb{E}$ as an abbreviation of $\mathbb{E}\{\cdot \mid X_0, \ldots, X_n\}$ and using

$$X_t' - X_t^* = \left(\Psi^t - \hat{\Psi}_n^t\right)(X_0') + \sum_{k=0}^{t-1} \Psi^k \left(\epsilon_{t-k}' - \epsilon_{t-k}^*\right) + \sum_{k=0}^{t-1} \left(\Psi^k - \hat{\Psi}_n^k\right)\left(\epsilon_{t-k}^*\right)$$

we get, as X_0', $(\epsilon_t', \epsilon_t^*)$ are independent and $\left\|\Psi^k\right\| \leq \|\Psi\|^k$

$$\begin{aligned}\mathbb{E}\left\|X_t' - X_t^*\right\|^2 &\leq 3\left\|\Psi^t - \hat{\Psi}_n^t\right\|_{\mathcal{L}}^2 \mathbb{E}\left\|X_0'\right\|^2 + 3\sum_{k=0}^{t-1} \|\Psi\|^{2k} \mathbb{E}\left\|\epsilon_{t-k}' - \epsilon_{t-k}^*\right\|^2 \\ &\quad + 3\sum_{k=0}^{t-1} \left\|\Psi^k - \hat{\Psi}_n^k\right\|_{\mathcal{L}}^2 \mathbb{E}\left\|\epsilon_{t-k}^*\right\|^2 \\ &\leq 3\left(\mathbb{E}\left\|X_0'\right\|^2 \frac{\hat{\delta}^{2t}}{\left(\hat{\delta} - \|\Psi\|\right)^2} + \mathbb{E}^*\left\|\epsilon_1^*\right\|^2 \sum_{k=0}^{t-1} \frac{\hat{\delta}^{2k}}{\left(\hat{\delta} - \|\Psi\|\right)^2}\right) \left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}^2 \\ &\quad + 3\frac{1}{1 - \|\Psi\|^2} d_2^2\left(F, \hat{F}_n\right) \\ &\leq \frac{3}{\left(\hat{\delta} - \|\Psi\|\right)^2}\left(\hat{\delta}^{2t}\mathbb{E}\left\|X_0'\right\|^2 + \mathbb{E}^*\left\|\epsilon_1^*\right\|^2\right) \left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}^2 \\ &\quad + \frac{3}{1 - \|\Psi\|^2} d_2^2\left(F, \hat{F}_n\right)\end{aligned}$$

b) From the definition of S_n in Lemma 10, we have

$$\begin{aligned}
nd_2^2\left(\hat{C}_n - \Psi\hat{\Gamma}_n(W), \hat{C}_n^* - \hat{\Psi}_n\hat{\Gamma}_n^*(W)\right) &= \frac{1}{n}d_2^2\left(S_n(W), S_n^*(W)\right) \\
&\leq \frac{1}{n}\mathbb{E}\left\|\sum_{t=1}^{n}\left\langle X'_{t-1}, W\right\rangle\epsilon'_t - \sum_{t=1}^{n}\left\langle X^*_{t-1}, W\right\rangle\epsilon^*_t\right\|^2 \\
&\leq \frac{2}{n}\mathbb{E}\left\|\sum_{t=1}^{n}\left\langle X'_{t-1} - X^*_{t-1}, W\right\rangle\epsilon'_t\right\|^2 \\
&\quad + \frac{2}{n}\mathbb{E}\left\|\sum_{t=1}^{n}\left\langle X^*_{t-1}, W\right\rangle\left(\epsilon'_t - \epsilon^*_t\right)\right\|^2 \\
&= \frac{2}{n}\sum_{t=1}^{n}\mathbb{E}\left\langle X'_{t-1} - X^*_{t-1}, W\right\rangle^2\mathbb{E}\left\|\epsilon'_t\right\|^2 \\
&\quad + \frac{2}{n}\sum_{t=1}^{n}\mathbb{E}\left\langle X^*_{t-1}, W\right\rangle^2\mathbb{E}\left\|\epsilon'_t - \epsilon^*_t\right\|^2 \\
&\leq \frac{2}{n}\sum_{t=1}^{n}\mathbb{E}\left\|X'_{t-1} - X^*_{t-1}\right\|^2\mathbb{E}\left\|W\right\|^2\mathbb{E}\left\|\epsilon'_1\right\|^2 \\
&\quad + \frac{2}{n}\sum_{t=1}^{n}\mathbb{E}\left\langle X^*_{t-1}, W\right\rangle^2 d_2^2\left(F, \hat{F}_n\right)
\end{aligned}$$

using the independence of $(\epsilon'_t, \epsilon^*_t)$ from $X'_{t-1}, X^*_{t-1},\ W$ and $\mathbb{E}\epsilon'_t = \mathbb{E}^*\epsilon^*_t = 0$, and Cauchy-Schwarz. From **a**), we have for some constant D

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\|X'_t - X^*_t\right\|^2 &\leq D\left(\mathbb{E}\left\|X'_0\right\|^2 + \mathbb{E}^*\left\|\epsilon^*_1\right\|^2\right)\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}^2 + Dd_2^2\left(F, \hat{F}_n\right) \\
&\underset{p}{\to} 0
\end{aligned}$$

from Theorem 6.3.1 and Lemma 12, as

$$\mathbb{E}^*\left\|\epsilon^*_1\right\|^2 = \frac{1}{n}\sum_{t=1}^{n}\left\|\hat{\epsilon}_t\right\|^2 \underset{p}{\to} \mathbb{E}\left\|\epsilon_1\right\|^2$$

from Theorem 6.3.1 and Lemma 8.3 of Bickel and Freedman, [6].

Moreover,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\langle X^*_{t-1}, W\right\rangle^2 &\leq \mathbb{E}\left\|W\right\|^2\frac{2}{n}\sum_{t=0}^{n-1}\mathbb{E}\left\|X'_t\right\|^2 + \mathbb{E}\left\|W\right\|^2\frac{2}{n}\sum_{t=0}^{n-1}\mathbb{E}\left\|X'_t - X^*_t\right\|^2 \\
&= 2\mathbb{E}\left\|W\right\|^2\mathbb{E}\left\|X_1\right\|^2 + o_p(1)
\end{aligned}$$

such that the assertion of the proposition follows from $d_2^2\left(F, \hat{F}_n\right) \underset{p}{\to} 0$ by Theorem 6.3.1. □

Checking the proof carefully and using $\mathbb{E} < Y, W >^2 \leq \mathcal{O}(\mathbb{E}||Y||^2)$ for bounded random variables W instead of $\mathbb{E} < Y, W >^2 = \mathbb{E}||Y||^2\mathbb{E}||W||^2$ for independent Y, W we get also

Corollary 6.4.1. *Proposition 6.4.1 continues to hold if the independence assumption on W is replaced by the assumption that W is a bounded random variable.*

6.4.1 Bootstrapping the sample mean

In this subsection we do still assume without loss of generality that $\mathbb{E}X_t = 0$, but we do no longer assume that it is known, but that we have to estimate it by the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{t=0}^{n-1} X_t.$$

Note that $\mathbb{E}\bar{X}_n = 0$, and that we also have the bootstrap analogue $\mathbb{E}^*\bar{X}_n^* = 0$ as, with X_t^* in the proof of Proposition 6.4.1,

$$\mathbb{E}^* X_t^* = \mathbb{E}^*\left(\hat{\Psi}_n^t(X_0') + \sum_{k=1}^{t}\hat{\Psi}_n^{t-k}(\epsilon_k^*)\right) = 0$$

due to linearity, $\mathbb{E}X_t' = \mathbb{E}X_t = 0$ and, by definition, $\mathbb{E}^*\epsilon_t^* = 0$. Therefore, we have to compare $\bar{X}_n$ and $\bar{X}_n^*$ without additional centering.

Theorem 6.4.1. *Under the assumptions of Theorem 6.3.1, we have for $n \to \infty$*

$$nd_2^2\left(\bar{X}_n, \bar{X}_n^*\right) \underset{p}{\to} 0$$

Proof. **a)** We choose $X_t', X_t^*, 0 \leq t \leq n-1$, as in the proof of Proposition 6.4.1. Then,

$$nd_2^2\left(\bar{X}_n, \bar{X}_n^*\right) \leq n\mathbb{E}||\bar{X}_n' - \bar{X}_n^*||^2 = \frac{1}{n}\sum_{t,s=0}^{n-1}\mathbb{E} < X_t' - X_t^*, X_s' - X_s^* > .$$

As in the proof of Proposition 6.4.1, part a), we have $X_t' - X_t^* = a_t + b_t + c_t$ with

$$X_t' - X_t^* = \left(\Psi^t - \hat{\Psi}_n^t\right)(X_0') + \sum_{k=1}^{t}\Psi^{t-k}(\epsilon_k' - \epsilon_k^*) + \sum_{k=1}^{t}\left(\Psi^{t-k} - \hat{\Psi}_n^{t-k}\right)(\epsilon_k^*) = a_t + b_t + c_t.$$

So, we have to study

$$\frac{1}{n}\sum_{t,s=0}^{n-1}\mathbb{E} < a_t + b_t + c_t, a_s + b_s + c_s > .$$

We show in the following two parts of the proof that the dominant terms

$$\frac{1}{n}\sum_{t,s=0}^{n-1}\mathbb{E}<b_t,b_s> \quad \text{and} \quad \frac{1}{n}\sum_{t,s=0}^{n-1}\mathbb{E}<c_t,c_s>$$

are of order $o_p(1)$. The remaining terms can be handled analogously, and the assertion follows.

b) As $(\epsilon_k', \epsilon_k^*), k = 1, \ldots, n$, are independent with mean 0, we have for $s \leq t$

$$\begin{aligned}
\mathbb{E}<b_t,b_s> &= \sum_{k=1}^{t}\sum_{l=1}^{s}<\Psi^{t-k}\left(\epsilon_k'-\epsilon_k^*\right),\Psi^{s-l}\left(\epsilon_l'-\epsilon_l^*\right)> \\
&= \sum_{k=1}^{s}\mathbb{E}<\Psi^{t-k}\left(\epsilon_k'-\epsilon_k^*\right),\Psi^{s-k}\left(\epsilon_k'-\epsilon_k^*\right)> \\
&\leq \sum_{k=1}^{s}||\Psi^{t-k}||_{\mathcal{L}}||\Psi^{s-k}||_{\mathcal{L}}\mathbb{E}||\epsilon_k'-\epsilon_k^*||^2 \\
&\leq \sum_{k=1}^{s}\hat{\delta}^{t+s-2k}d_2^2(F,\hat{F}_n) \\
&= \hat{\delta}^{t-s}\sum_{k=1}^{s}\hat{\delta}^{2(s-k)}d_2^2(F,\hat{F}_n) = \hat{\delta}^{t-s}d_2^2(F,\hat{F}_n)\mathcal{O}(1)
\end{aligned}$$

where $\hat{\delta} < 1$ is as in the proof of Proposition 6.4.1. We conclude

$$\frac{1}{n}\sum_{t,s=0}^{n-1}\mathbb{E}<b_t,b_s>\leq \frac{2}{n}\sum_{t=0}^{n-1}\sum_{s=0}^{t}\hat{\delta}^{t-s}d_2^2(F,\hat{F}_n)\mathcal{O}(1) = d_2^2(F,\hat{F}_n)\mathcal{O}(1) = o_p(1)$$

by Theorem 6.3.1.

c) As $\epsilon_k^*, k = 1, \ldots, n$, are independent, we have for $s \leq t$ and some generic constant D

$$\begin{aligned}
\mathbb{E}<c_t,c_s> &= \sum_{k=1}^{t}\sum_{l=1}^{s}\mathbb{E}<\left(\Psi^{t-k}-\hat{\Psi}_n^{t-k}\right)\left(\epsilon_k^*\right),\left(\Psi^{s-l}-\hat{\Psi}_n^{s-l}\right)\left(\epsilon_l^*\right)> \\
&= \sum_{k=1}^{s}\mathbb{E}<\left(\Psi^{t-k}-\hat{\Psi}_n^{t-k}\right)\left(\epsilon_k^*\right),\left(\Psi^{s-k}-\hat{\Psi}_n^{s-k}\right)\left(\epsilon_k^*\right)> \\
&\leq D\sum_{k=1}^{s}\hat{\delta}^{t+s-2k}||\Psi-\hat{\Psi}_n||_{\mathcal{L}}^2\mathbb{E}^*||\epsilon_k^*||^2 \\
&= \hat{\delta}^{t-s}\sum_{k=1}^{s}\hat{\delta}^{2(s-k)}||\Psi-\hat{\Psi}_n||_{\mathcal{L}}^2\mathcal{O}_p(1) = \hat{\delta}^{t-s}||\Psi-\hat{\Psi}_n||_{\mathcal{L}}^2\mathcal{O}_p(1)
\end{aligned}$$

using the bounds on $||\Psi^t - \hat{\Psi}_n^t||_{\mathcal{L}}$ and $\mathbb{E}^*||\epsilon_k^*||^2$ from the proof of Proposition 6.4.1. We

conclude

$$\frac{1}{n}\sum_{t,s=0}^{n-1}\mathbb{E}< c_t, c_s >\leq \frac{2}{n}\sum_{t=0}^{n-1}\sum_{s=0}^{t}\hat{\delta}^{t-s}||\Psi-\hat{\Psi}_n||^2_{\mathcal{L}}\mathcal{O}_p(1)=||\Psi-\hat{\Psi}_n||^2_{\mathcal{L}}\mathcal{O}_p(1)=o_p(1)$$

by Lemma 12. □

6.5 Bootstrapping the covariance operator

In this section, we show that the bootstrap works for the covariance operator estimate $\hat{\Gamma}_n$. It is obvious from the proof that a similar result would also hold for the lag 1-autocovariance operator estimate $\hat{C}_n$, at least under the condition that C is a symmetric operator, which is automatically satisfied for Γ. $\hat{\Gamma}_n$ is an unbiased estimate of Γ. We need that this is asymptotically true for the bootstrap analogue too which we formulate as an own result.

Lemma 13. $\mathbb{E}^*\hat{\Gamma}^*_n=\hat{\Gamma}_n+\mathcal{O}_p\left(\frac{1}{n}\right)$, *if* $\left\|\hat{\Psi}_n-\Psi\right\|_{\mathcal{L}}\underset{a.s.}{\to}0$

Proof. From the representation of X^*_t in terms of $X^*_0=X'_0$ and $\epsilon^*_1,\ldots,\epsilon^*_t$ (compare the proof of Theorem 6.3.1), we have

$$X^*_t=\hat{\Psi}^t_n\left(X'_0\right)+\sum_{k=1}^{t-k}\hat{\Psi}^{t-k}_n\left(\epsilon^*_k\right)$$

As $\mathbb{E}^*\epsilon^*_k=0$ and as X'_0 is independent of $X_0,\ldots,X_n$ with mean 0, we know, using linearity of $\hat{\Psi}^l_n,\ l\geq 1$, that we have $\mathbb{E}^*X^*_t=0$ such that we consider

$$\hat{\Gamma}^*_n=\frac{1}{n}\sum_{t=0}^{n-1}X^*_t\otimes X^*_t$$

as the version of the covariance operator estimate in the bootstrap world to simplify notation. Plugging in the recursive definition of X^*_t, we get

$$\begin{aligned}\hat{\Gamma}^*_n=\frac{1}{n}\sum_{t=0}^{n-1}\Bigg(&\hat{\Psi}^t_n\left(X'_0\right)\otimes\hat{\Psi}^t_n\left(X'_0\right)+\sum_{k=1}^{t}\hat{\Psi}^t_n\left(X'_0\right)\otimes\hat{\Psi}^{t-k}_n\left(\epsilon^*_k\right)\\&+\sum_{k=1}^{t}\hat{\Psi}^{t-k}_n\left(\epsilon^*_k\right)\otimes\hat{\Psi}^t_n\left(X'_0\right)+\sum_{k,l=1}^{t}\hat{\Psi}^{t-k}_n\left(\epsilon^*_k\right)\otimes\hat{\Psi}^{t-l}_n\left(\epsilon^*_l\right)\Bigg)\end{aligned}\tag{6.4}$$

As $\mathbb{E}^*\hat{\Psi}^l\left(\epsilon^*_k\right)=0$ due to linearity and as $\epsilon^*_1,\ldots,\epsilon^*_n,X'_0$ are independent, we get

$$\mathbb{E}\hat{\Gamma}^*_n=\frac{1}{n}\sum_{t=0}^{n-1}\left(\mathbb{E}\hat{\Psi}^t_n\left(X'_0\right)\otimes\hat{\Psi}^t_n\left(X'_0\right)+\sum_{k=1}^{t}\mathbb{E}^*\hat{\Psi}^{t-k}_n\left(\epsilon^*_k\right)\otimes\hat{\Psi}^{t-k}_n\left(\epsilon^*_k\right)\right)$$

As in the bootstrap world, $\hat{\Psi}_n^l$ are fixed, and as for any linear operator A and $y \in L^2$, we have $A(y) \otimes A(y) = Ay \otimes yA^T$, with A^T denoting the adjoint of A, we have to investigate mainly $\mathbb{E}^* \epsilon_k^* \otimes \epsilon_k^*$, as we know already $\mathbb{E} X_0' \otimes X_0' = \mathbb{E} X_0 \otimes X_0 = \Gamma$.

$$\mathbb{E}^* \epsilon_k^* \otimes \epsilon_k^* = \frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t \otimes \tilde{\epsilon}_t = \frac{1}{n} \sum_{t=1}^{n} \left(\hat{\epsilon}_t - \bar{\hat{\epsilon}}_n \right) \otimes \left(\hat{\epsilon}_t - \bar{\hat{\epsilon}}_n \right) = \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t \otimes \hat{\epsilon}_t - \bar{\hat{\epsilon}}_n \otimes \bar{\hat{\epsilon}}_n$$

with $\bar{\hat{\epsilon}}_n = \frac{1}{n} \sum_{k=1}^{n} \hat{\epsilon}_k$. As $\hat{\epsilon}_k = X_k - \hat{\Psi}_n (X_{k-1})$,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t \otimes \hat{\epsilon}_t &= \frac{1}{n} \sum_{t=1}^{n} \left(X_t - \hat{\Psi}_n (X_{t-1}) \right) \otimes \left(X_t - \hat{\Psi}_n (X_{t-1}) \right) \\ &= \hat{\Gamma}_n + \frac{1}{n} (X_n \otimes X_n - X_0 \otimes X_0) - \frac{1}{n} \sum_{t=1}^{n} \hat{\Psi}_n (X_{t-1}) \otimes X_t \\ &\quad - \frac{1}{n} \sum_{t=1}^{n} X_t \otimes \hat{\Psi}_n (X_{t-1}) + \hat{\Psi}_n \hat{\Gamma}_n \hat{\Psi}_n^T \end{aligned}$$

As for the linear operator A and $y, z \in L^2$, we have $A(y) \otimes z = (y \otimes z) A^T$ and $y \otimes A(z) = A (y \otimes z)$ from the definition of $\otimes$, the second and third terms are $\hat{C}_n \hat{\Psi}_n^T$ and $\hat{\Psi}_n \hat{C}_n^T$ respectively, such that, as $\hat{C}_n = \hat{\Psi}_n \hat{\Gamma}_n$

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t \otimes \hat{\epsilon}_t &= \hat{\Gamma}_n - \hat{\Psi}_n \hat{\Gamma}_n \hat{\Psi}_n^T + \frac{1}{n} (X_n \otimes X_n - X_0 \otimes X_0) \\ &= \hat{\Gamma}_n - \hat{\Psi}_n \hat{\Gamma}_n \hat{\Psi}_n^T + \mathcal{O}_p \left(\frac{1}{n} \right) \end{aligned}$$

Similarly, we have

$$\begin{aligned} \bar{\hat{\epsilon}}_n \otimes \bar{\hat{\epsilon}}_n &= \frac{1}{n^2} \sum_{k,l=1}^{n} \hat{\epsilon}_k \otimes \hat{\epsilon}_l = \frac{1}{n^2} \sum_{k,l=1}^{n} \left(X_k - \hat{\Psi}_n (X_{k-1}) \right) \otimes \left(X_l - \hat{\Psi}_n (X_{l-1}) \right) \\ &= \bar{X}_{1:n} \otimes \bar{X}_{1:n} - \hat{\Psi}_n \left(\bar{X}_{1:n} \otimes \bar{X}_{0:(n-1)} \right) - \left(\bar{X}_{0:(n-1)} \otimes \bar{X}_{1:n} \right) \hat{\Psi}_n^T \\ &\quad \hat{\Psi}_n \left(\bar{X}_{0:(n-1)} \otimes \bar{X}_{0:(n-1)} \right) \bar{\Psi}_n^T \end{aligned}$$

where $\bar{X}_{1:n}$, $\bar{X}_{0:(n-1)}$ denote the sample means of $X_1, \ldots, X_n$ respectively $X_0, \ldots, X_{n-1}$. As $\left\| \hat{\Psi}_n - \Psi \right\|_{\mathcal{L}} \underset{p}{\to} 0$ we have $\left\| \hat{\Psi}_n \right\|_{\mathcal{L}} = \mathcal{O}_p(1)$, and as $\bar{X}_{0:(n-1)}, \bar{X}_{1:n}$ are $\mathcal{O}_p \left(\frac{1}{\sqrt{n}} \right)$ from the law of large numbers of FAR(1)-processes (compare Theorem 3.7 of Bosq, [7]), we immediately get that $\bar{\hat{\epsilon}}_n \otimes \bar{\hat{\epsilon}}_n = \mathcal{O}_p \left(\frac{1}{n} \right)$. So we get

$$\mathbb{E}^* \epsilon_k^* \otimes \epsilon_k^* = \hat{\Gamma}_n - \hat{\Psi}_n \hat{\Gamma}_n \hat{\Psi}_n^T + \frac{1}{n} R_n$$

with $R_n = \mathcal{O}_p(1)$. Hence, we have for the dominant term in $\mathbb{E}^*\hat{\Gamma}_n^*$

$$\begin{aligned}
\mathbb{E}^* \sum_{k=1}^{t} \hat{\Psi}_n^{t-k} \epsilon_k^* \otimes \epsilon_k^* \left(\hat{\Psi}_n^{t-k}\right)^T &= \sum_{k=1}^{t} \hat{\Psi}_n^{t-k} \left(\hat{\Gamma}_n - \hat{\Psi}_n \hat{\Gamma}_n \hat{\Psi}_n^T\right) \left(\hat{\Psi}_n^{t-k}\right)^T + \sum_{k=1}^{t} \hat{\Psi}_n^{t-k} \frac{1}{n} R_n \left(\hat{\Psi}_n^{t-k}\right)^T \\
&= \sum_{k=1}^{t} \hat{\Psi}_n^{t-k} \hat{\Gamma}_n \left(\hat{\Psi}_n^{t-k}\right)^T - \sum_{l=0}^{t-1} \hat{\Psi}_n^{t-l} \hat{\Gamma}_n \left(\hat{\Psi}_n^{t-l}\right)^T + \mathcal{O}_p\left(\frac{1}{n}\right) \\
&= \hat{\Gamma}_n - \hat{\Psi}_n^t \hat{\Gamma}_n \left(\hat{\Psi}_n^t\right)^T + \mathcal{O}_p\left(\frac{1}{n}\right)
\end{aligned}$$

where we have used that $R_n = \mathcal{O}_p(1)$, $\left\|\hat{\Psi}_n^l\right\|_{\mathcal{L}} \leq \left\|\hat{\Psi}_n\right\|_{\mathcal{L}}^l \leq \hat{\delta}^l$ for some $\hat{\delta} < 1$ and large enough n as in the proof of Proposition 6.4.1 and $\sum_{k=1}^{t} \hat{\delta}^{2(t-k)} \leq \dfrac{1}{1-\hat{\delta}^2}$. Finally,

$$\begin{aligned}
\mathbb{E}^* \frac{1}{n} \sum_{t=0}^{n-1} \sum_{k=1}^{t} \hat{\Psi}_n^{t-k} \epsilon_k^* \otimes \epsilon_k^* \left(\hat{\Psi}_n^{t-k}\right)^T &= \frac{1}{n} \sum_{t=0}^{n-1} \left(\hat{\Gamma}_n - \hat{\Psi}_n^t \hat{\Gamma}_n \left(\hat{\Psi}_n^t\right)^T\right) + \mathcal{O}_p\left(\frac{1}{n}\right) \\
&= \hat{\Gamma}_n + \mathcal{O}_p\left(\frac{1}{n}\right)
\end{aligned}$$

as, using again the above argument that $\left\|\hat{\Psi}_n^l\right\|_{\mathcal{L}} \leq \hat{\delta}_l$

$$\left\| \sum_{t=0}^{n-1} \hat{\Psi}_n^t \hat{\Gamma}_n \left(\hat{\Psi}_n^t\right)^T \right\|_{\mathcal{L}} \leq \sum_{t=0}^{n-1} \hat{\delta}^{2t} \left\|\hat{\Gamma}_n\right\|_{\mathcal{L}} \leq \frac{1}{1-\hat{\delta}^2} \left\|\hat{\Gamma}_n\right\|_{\mathcal{L}} = \mathcal{O}_p(1)$$

Analogously, the expectation of the first term in 6.4 is $\mathcal{O}_p\left(\dfrac{1}{n}\right)$. □

Note that under our standard assumption, $\left\|\hat{\Psi}_n - \Psi\right\|_{\mathcal{L}} \underset{a.s.}{\to} 0$ holds by Lemma 12.

The following result just states a rule of calculation needed in the proof of the following main result.

Lemma 14. *If (U, U^*), (V, V^*) are i.i.d. L^2-valued random variables such that $d_2^2(U, U^*) = \mathbb{E}\|U - U^*\|^2$; then*

$$\mathbb{E}\left\|(U \otimes V - U^* \otimes V^*)(x)\right\|^2 \leq 2\left(\mathbb{E}\|U\|^2 + \mathbb{E}\|U^*\|^2\right) d_2^2(U, U^*) \|x\|^2$$

for any x in L^2.

Proof. Using independence of (U, U^*) and (V, V^*)

$$\begin{aligned}
\mathbb{E}\left\|(U \otimes V - U^* \otimes V^*)(x)\right\|^2 &= \mathbb{E}\left\|(U - U^*) \otimes V(x) + U^* \otimes (V - V^*)(x)\right\|^2 \\
&\leq 2\left(\mathbb{E}\langle U - U^*, x\rangle^2 \mathbb{E}\|V\|^2 + \mathbb{E}\langle U^*, x\rangle^2 \mathbb{E}\|V - V^*\|^2\right) \\
&\leq 2 d_2^2(U, U^*) \|x\|^2 \left(\mathbb{E}\|U\|^2 + \mathbb{E}\|U^*\|^2\right)
\end{aligned}$$

as $\mathbb{E}\left\|V-V^*\right\|^2=\mathbb{E}\left\|U-U^*\right\|^2$ and $\mathbb{E}\left\|V\right\|^2=\mathbb{E}\left\|U\right\|^2$

□

Theorem 6.5.1. *Under the assumptions of Theorem 6.3.1, we have uniformly in $x\in L^2$ for $n\to\infty$*

$$\frac{1}{\left\|x\right\|^2}nd_2^2\left(\hat{\Gamma}_n(x)-\Gamma(x),\hat{\Gamma}_n^*(x)-\hat{\Gamma}_n(x)\right)\underset{p}{\to}0$$

Proof. As in the proof of Proposition 6.4.1 we choose $(\epsilon_t',\epsilon_t^*)$ i.i.d. such that $\mathcal{L}\left(\epsilon_t^*\right)=\hat{F}_n,\ \mathcal{L}\left(\epsilon_t'\right)=F$ and $\mathbb{E}\left\|\epsilon_t'-\epsilon_t^*\right\|^2=d_2^2\left(F,\hat{F}_n\right)$. Moreover, we choose $X_0^*=X_0'$ independent of $\left(\epsilon_k',\epsilon_k^*\right),\ k\geq 1$, and $\mathcal{L}\left(X_0'\right)=\mathcal{L}\left(X_0\right)$. Due to stationarity of $\{X_t\},\mathbb{E}\hat{\Gamma}_n=\Gamma$, and from Lemmas 12 and 13 we have $\mathbb{E}^*\hat{\Gamma}_n^*=\hat{\Gamma}_n+\mathcal{O}_p\left(\frac{1}{n}\right)$. Hence, up to terms of order $\frac{1}{n}$, which we denote by $\sim$ to simplify notation:

$$\begin{aligned}\hat{\Gamma}_n'-\Gamma-\left(\hat{\Gamma}_n^*-\hat{\Gamma}_n\right) &\sim \frac{1}{n}\sum_{t=0}^{n-1}\left(X_t'\otimes X_t'-\mathbb{E}\left(X_t'\otimes X_t'\right)-\left(X_t^*\otimes X_t^*-\mathbb{E}^*\left(X_t^*\otimes X_t^*\right)\right)\right)\\ &= \frac{1}{n}\sum_{t=0}^{n-1}A_t\end{aligned}$$

Replacing X_t',X_t^* by their representation in terms of $X_0',\epsilon_1',\ldots,\epsilon_t'$ respectively $\epsilon_1^*,\ldots,\epsilon_t^*$, i.e.

$$X_t'=\Psi^t\left(X_0'\right)+\sum_{k=1}^{t}\Psi^{t-k}\left(\epsilon_k'\right),\ X_t^*=\hat{\Psi}_n^t\left(X_0'\right)+\sum_{k=1}^{t}\hat{\Psi}_n^{t-k}\left(\epsilon_k^*\right)$$

we have $A_t=a_t+b_t+b_t^T+c_t+d_t$ with

$$\begin{aligned}a_t &= \Psi^t\left[X_0'\otimes X_0'-\mathbb{E}\left(X_0'\otimes X_0'\right)\right]\left(\Psi^t\right)^T-\hat{\Psi}_n^t\left[X_0'\otimes X_0'-\mathbb{E}\left(X_0'\otimes X_0'\right)\right]\left(\hat{\Psi}_n^t\right)^T\\ b_t &= \sum_{k=1}^{t}\left[\Psi^{t-k}\left(\epsilon_k'\otimes X_0'\right)\left(\Psi^t\right)^T-\hat{\Psi}_n^{t-k}\left(\epsilon_k^*\otimes X_0'\right)\left(\hat{\Psi}_n^t\right)^T\right]\\ c_t &= \sum_{k\neq l=1}^{t}\left[\Psi^{t-k}\left(\epsilon_k'\otimes\epsilon_l'\right)\left(\Psi^{t-l}\right)^T-\hat{\Psi}_n^{t-k}\left(\epsilon_k^*\otimes\epsilon_l^*\right)\left(\hat{\Psi}_n^{t-l}\right)^T\right]\\ d_t &= \sum_{k=1}^{t}\left[\Psi^{t-k}\left(\epsilon_k'\otimes\epsilon_k'\right)-\mathbb{E}\left(\epsilon_k'\otimes\epsilon_k'\right)\left(\Psi^{t-k}\right)^T-\hat{\Psi}_n^{t-k}\left(\epsilon_k^*\otimes\epsilon_k^*\right)+\mathbb{E}^*\left(\epsilon_k^*\otimes\epsilon_k^*\right)\left(\hat{\Psi}_n^{t-k}\right)^T\right]\end{aligned}$$

where we have used that $(\epsilon_k',\epsilon_k^*)$ are i.i.d. with mean 0 to get, e.g., $\mathbb{E}\epsilon_k'\otimes\epsilon_l'=0$ for $k\neq l$. As we are interested in

$$\mathbb{E}\left\|\left[\hat{\Gamma}_n'-\Gamma-\left(\hat{\Gamma}_n^*-\hat{\Gamma}_n\right)\right](x)\right\|^2\sim\frac{1}{n^2}\sum_{s,t=0}^{n-1}\mathbb{E}\left\langle A_t(x),A_s(x)\right\rangle$$

we have to study terms like
$\mathbb{E}\sum_{s,t=0}^{n-1}\langle a_t(x), b_s(x)\rangle$, $\mathbb{E}\sum_{s,t=0}^{n-1}\langle c_t(x), c_s(x)\rangle$, $\mathbb{E}\sum_{s,t=0}^{n-1}\langle d_t(x), d_s(x)\rangle$.

a) We start with $\sum_{s,t=0}^{n-1}\mathbb{E}\langle c_t(x), c_s(x)\rangle = \sum_{s,t=0}^{n-1}\sum_{k\neq l=1}^{t}\sum_{i\neq j=1}^{s}\mathbb{E}B_{klij}^{(s,t)}(x)$ where

$$B_{klij}^{(s,t)}(x) = \Big\langle \Psi^{t-k}\left(\epsilon'_k\otimes\epsilon'_l\right)\left(\Psi^{t-l}\right)^T(x) - \hat{\Psi}_n^{t-k}\left(\epsilon_k^*\otimes\epsilon_l^*\right)\left(\hat{\Psi}_n^{t-l}\right)^T(x),$$
$$\Psi^{s-i}\left(\epsilon'_i\otimes\epsilon'_j\right)\left(\Psi^{s-j}\right)^T(x) - \hat{\Psi}_n^{s-i}\left(\epsilon_i^*\otimes\epsilon_j^*\right)\left(\hat{\Psi}_n^{s-j}\right)^T(x)\Big\rangle$$

As $k\neq l$, we have $\mathbb{E}\epsilon'_k\otimes\epsilon'_l(z) = \mathbb{E}\langle\epsilon'_k, z\rangle\epsilon'_l = \mathbb{E}\langle\epsilon'_k, z\rangle\mathbb{E}\epsilon'_l = 0$ and, analogously, $\mathbb{E}^*\epsilon_k^*\otimes\epsilon_l^*(z) = 0$ for all z. Moreover, if e.g. $j\neq k, l$, we have

$$\begin{aligned}\mathbb{E}\left\langle\Psi^{t-k}\left(\epsilon'_k\otimes\epsilon'_l\right)(z), \Psi^{s-i}\left(\epsilon'_i\otimes\epsilon'_j\right)(y)\right\rangle &= \mathbb{E}\langle\epsilon'_k, z\rangle\left\langle\Psi^{t-k}\epsilon'_l, \Psi^{s-i}\epsilon'_j\right\rangle\langle\epsilon'_i, y\rangle\\ &= \left\langle\mathbb{E}\left\{\langle\epsilon'_k, z\rangle\langle\epsilon'_i, y\rangle\Psi^{t-k}\epsilon'_l\right\}, \mathbb{E}\Psi^{s-i}\epsilon'_j\right\rangle = 0\end{aligned}$$

as $\mathbb{E}\Psi^{s-i}\epsilon'_j = \Psi^{s-i}\left(\mathbb{E}\epsilon'_j\right)$. Analogously, the expectations of the other terms are vanishing, such that for $k\neq l, i\neq j$, $\mathbb{E}B_{klij}^{(s,t)}(x) = 0$ except for $k=i\neq l=j$ or $k=j\neq l=i$. To get the expectations of the remaining terms, we decompose

$$\begin{aligned}&\Psi^{t-k}\left(\epsilon'_k\otimes\epsilon'_l\right)\left(\Psi^{t-l}\right)^T - \hat{\Psi}_n^{t-k}\left(\epsilon_k^*\otimes\epsilon_l^*\right)\left(\hat{\Psi}_n^{t-l}\right)^T\\ &= \left(\Psi^{t-k} - \hat{\Psi}_n^{t-k}\right)\left(\epsilon'_k\otimes\epsilon'_l\right)\left(\Psi^{t-l}\right)^T + \hat{\Psi}_n^{t-k}\left(\epsilon'_k\otimes\epsilon'_l\right)\left(\Psi^{t-l} - \hat{\Psi}_n^{t-l}\right)^T\\ &\quad + \hat{\Psi}_n^{t-k}\left(\epsilon'_k\otimes\epsilon'_l - \epsilon_k^*\otimes\epsilon_l^*\right)\left(\hat{\Psi}_n^{t-l}\right)^T = \beta_{1,t} + \beta_{2,t} + \beta_{3,t}\end{aligned}$$

and using $\|\Psi^j\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{L}}^j \leq \hat{\delta}^j$ for some $\hat{\delta} < 1$.

$$\begin{aligned}\|\beta_{1,t}(x)\| &\leq \left\|\Psi^{t-k} - \hat{\Psi}_n^{t-k}\right\|_{\mathcal{L}}\|\epsilon'_k\otimes\epsilon'_l\|_{\mathcal{L}}\left\|\Psi^{t-l}\right\|_{\mathcal{L}}\|x\|\\ &\leq D\hat{\delta}^{2t-k-l}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}\|\epsilon'_k\otimes\epsilon'_l\|_{\mathcal{L}}\|x\|\end{aligned}$$

for some generic constant D from the proof of Proposition 6.4.1. Analogously,

$$\begin{aligned}\|\beta_{2,t}(x)\| &\leq D\hat{\delta}^{2t-k-l}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}\|\epsilon'_k\otimes\epsilon'_l\|_{\mathcal{L}}\|x\|\\ \|\beta_{3,t}(x)\| &\leq \hat{\delta}^{2t-k-l}\|\epsilon'_k\otimes\epsilon'_l - \epsilon_k^*\otimes\epsilon_l^*\|_{\mathcal{L}}\|x\|\end{aligned}$$

where we use $\left\|\hat{\Psi}_n^j\right\|_{\mathcal{L}} \leq \left\|\hat{\Psi}_n\right\|_{\mathcal{L}}^j$ and $\left\|\hat{\Psi}_n\right\|_{\mathcal{L}} \leq \hat{\delta}$ for large enough n again from the proof of Proposition 6.4.1. Using $\|U\otimes V\|_{\mathcal{L}} \leq \|U\|\,\|V\|$, which follows from the definition of the operator norm and of $\otimes$ and from the Cauchy-Schwarz inequality, we have

$$\|\beta_{i,t}(x)\| \leq D\hat{\delta}^{2t-k-l}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}\|\epsilon'_k\|\,\|\epsilon'_l\|\,\|x\|, \ i = 1, 2$$

For β_{3t}, we use

$$\begin{aligned}\|\epsilon'_k\otimes\epsilon'_l - \epsilon_k^*\otimes\epsilon_l^*\|_{\mathcal{L}} &\leq \|(\epsilon'_k - \epsilon_k^*)\otimes\epsilon'_l\|_{\mathcal{L}} + \|\epsilon_k^*\otimes(\epsilon'_l - \epsilon_l^*)\|_{\mathcal{L}}\\ &\leq \|\epsilon'_k - \epsilon_k^*\|\,\|\epsilon'_l\| + \|\epsilon_k^*\|\,\|\epsilon'_l - \epsilon_l^*\|\end{aligned}$$

to get
$\|\beta_{3t}(x)\| \leq \hat{\delta}^{2t-k-l} \|x\| \{\|\epsilon_l'\| \|\epsilon_k' - \epsilon_k^*\| + \|\epsilon_k^*\| \|\epsilon_l' - \epsilon_l^*\|\}$.
Now, as $k \neq l$,

$$\begin{aligned}
\left|\mathbb{E}B_{klkl}^{(s,t)}(x)\right| &\leq \mathbb{E}\left|\langle \beta_{1t}(x) + \beta_{2t}(x) + \beta_{3t}(x), \beta_{1s}(x) + \beta_{2s}(x) + \beta_{3s}(x)\rangle\right| \\
&\leq 4D^2\hat{\delta}^{2(t+s-k-l)} \left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}^2 \mathbb{E}\|\epsilon_k'\|^2 \mathbb{E}\|\epsilon_l'\|^2 \|x\|^2 \\
&\quad +4D\hat{\delta}^{2(t+s-k-l)} \left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}} \Big\{\mathbb{E}\|\epsilon_l'\|^2 \mathbb{E}\left(\|\epsilon_k'\| \|\epsilon_k' - \epsilon_k^*\|\right) \\
&\qquad +\mathbb{E}\left(\|\epsilon_k'\| \|\epsilon_k^*\|\right) \mathbb{E}\left(\|\epsilon_l'\| \|\epsilon_l' - \epsilon_l^*\|\right)\Big\} \|x\|^2 \\
&\quad +\hat{\delta}^{2(t+s-k-l)} \mathbb{E}\left\{\|\epsilon_l'\| \|\epsilon_k' - \epsilon_k^*\| + \|\epsilon_k^*\| \|\epsilon_l' - \epsilon_l^*\|\right\}^2 \|x\|^2
\end{aligned}$$

Note that the expectation in the last term may be written as

$$\begin{aligned}
&\mathbb{E}\|\epsilon_l'\|^2 \mathbb{E}\|\epsilon_k' - \epsilon_k^*\|^2 + 2\mathbb{E}\left(\|\epsilon_l'\| \|\epsilon_l' - \epsilon_l^*\|\right) \mathbb{E}\left(\|\epsilon_k^*\| \|\epsilon_k' - \epsilon_k^*\|\right) + \mathbb{E}^*\|\epsilon_k^*\|^2 \mathbb{E}^*\|\epsilon_l' - \epsilon_l^*\|^2 \\
&\leq \left(\mathbb{E}\|\epsilon_l'\|^2 + 2\sqrt{\mathbb{E}\|\epsilon_l'\|^2}\sqrt{\mathbb{E}^*\|\epsilon_k^*\|^2} + \mathbb{E}^*\|\epsilon_k^*\|^2\right) d_2^2\left(F, \hat{F}_n\right) \\
&\leq 2\left(\mathbb{E}\|\epsilon_l'\|^2 + \mathbb{E}^*\|\epsilon_k^*\|^2\right) d_2^2\left(F, \hat{F}_n\right)
\end{aligned}$$

due to our particular choice of $(\epsilon_k', \epsilon_k^*)$. Analogously, we get for the term involving expectation in the second to last term that it is bounded by, using that ϵ_l', ϵ_k' are identically distributed,

$$\begin{aligned}
&\mathbb{E}\|\epsilon_l'\|^2 \sqrt{\mathbb{E}\|\epsilon_k'\|^2} d_2\left(F, \hat{F}_n\right) + \sqrt{\mathbb{E}\|\epsilon_k'\|^2}\sqrt{\mathbb{E}^*\|\epsilon_k^*\|^2}\sqrt{\mathbb{E}\|\epsilon_l'\|^2} d_2\left(F, \hat{F}_n\right) \\
&= \mathbb{E}\|\epsilon_k'\|^2 \left(\sqrt{\mathbb{E}\|\epsilon_k'\|^2} + \sqrt{\mathbb{E}^*\|\epsilon_k^*\|^2}\right) d_2\left(F, \hat{F}_n\right)
\end{aligned}$$

From Theorem 6.3.1, we have $d_2^2\left(F, \hat{F}_n\right) = o_p(1)$ and, using Lemma 8.3 of Bickel and Freedman, [6], $\mathbb{E}^*\|\epsilon_t^*\|^2 \underset{p}{\to} \mathbb{E}\|\epsilon_t\|^2$, i.e. $\mathbb{E}^*\|\epsilon_t^*\|^2 = \mathcal{O}_p(1)$. From Lemma 12, $\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}} \underset{a.s.}{\to} 0$. So, we have with some generic constant D
$\left|\mathbb{E}B_{klkl}^{(s,t)}(x)\right| \leq D\hat{\delta}^{2(t+s-k-l)} \|x\|^2 o_p(1)$
uniformly in k, l, s, t and x. Analogously, we have the same upper bound for $\left|\mathbb{E}B_{kllk}^{(s,t)}(x)\right|$ too. Finally, we conclude, using that $k = i, l = j$ or $k = j, l = i$ is only possible for $k, l \leq \min(s, t)$

$$\begin{aligned}
\left|\sum_{s,t=0}^{n-1} \mathbb{E}\left\langle c_t(x), c_s(x)\right\rangle\right| &\leq \sum_{s,t=0}^{n-1} \sum_{k\neq l=1}^{min(s,t)} \hat{\delta}^{2(t+s-k-l)} D\|x\|^2 o_p(1) \\
&= \|x\|^2 o_p(n)
\end{aligned}$$

as the first factor is $\mathcal{O}(n)$ by the same kind of calculations as in the proof of Proposition 5.3.3, part b).

b) As the next term, we consider

$$\sum_{s,t=0}^{n-1} \mathbb{E}\left\langle d_t(x), d_s(x)\right\rangle = \sum_{s,t=0}^{n-1}\sum_{k=1}^{t}\sum_{l=1}^{s} \mathbb{E}B_{kl}^{(s,t)}(x)$$

where

$$\begin{aligned} B_{kl}^{(s,t)}(x) &= \left\langle \left[\Psi^{t-k}\left(\epsilon'_k\otimes\epsilon'_k - \mathbb{E}\epsilon'_k\otimes\epsilon'_k\right)\left(\Psi^{t-k}\right)^T - \hat{\Psi}_n^{t-k}\left(\epsilon_k^*\otimes\epsilon_k^* - \mathbb{E}^*\epsilon_k^*\otimes\epsilon_k^*\right)\left(\hat{\Psi}_n^{t-k}\right)^T\right](x), \right. \\ &\quad \left. \left[\Psi^{s-l}\left(\epsilon'_l\otimes\epsilon'_l - \mathbb{E}\epsilon'_l\otimes\epsilon'_l\right)\left(\Psi^{s-l}\right)^T - \hat{\Psi}_n^{s-l}\left(\epsilon_l^*\otimes\epsilon_l^* - \mathbb{E}^*\epsilon_l^*\otimes\epsilon_l^*\right)\left(\hat{\Psi}_n^{s-l}\right)^T\right](x)\right\rangle \end{aligned}$$

Due to independence of $(\epsilon'_k, \epsilon_k^*)$, $(\epsilon'_l, \epsilon_l^*)$ and linearity of the operators involved, we have $\mathbb{E}B_{kl}^{(s,t)} = 0$ for $k \neq l$. For the remaining case, as in a), we decompose the left factor of the scalar product into 3 terms, where now

$$\begin{aligned} \beta_{1t} &= \left(\Psi^{t-k} - \hat{\Psi}_n^{t-k}\right)\left(\epsilon'_k\otimes\epsilon'_k - \mathbb{E}\epsilon'_k\otimes\epsilon'_k\right)\left(\Psi^{t-k}\right)^T \\ \beta_{2t} &= \hat{\Psi}_n^{t-k}\left(\epsilon'_k\otimes\epsilon'_k - \mathbb{E}\epsilon'_k\otimes\epsilon'_k\right)\left(\Psi^{t-k} - \hat{\Psi}_n^{t-k}\right)^T \\ \beta_{3t} &= \hat{\Psi}_n^{t-k}\left(\epsilon'_k\otimes\epsilon'_k - \mathbb{E}\epsilon'_k\otimes\epsilon'_k - \epsilon_k^*\otimes\epsilon_k^* + \mathbb{E}^*\epsilon_k^*\otimes\epsilon_k^*\right)\left(\hat{\Psi}_n^{t-k}\right)^T \end{aligned}$$

such that
$B_{kk}^{(s,t)}(x) = \langle \beta_{1t}(x) + \beta_{2t}(x) + \beta_{3t}(x), \beta_{1s}(x) + \beta_{2s}(x) + \beta_{3s}(x)\rangle$
For the first two terms, we use

$$\begin{aligned} \left\|\epsilon'_k\otimes\epsilon'_k - \mathbb{E}\left(\epsilon'_k\otimes\epsilon'_k\right)\right\|_{\mathcal{L}} &\leq \left\|\epsilon'_k\otimes\epsilon'_k\right\|_{\mathcal{L}} + \mathbb{E}\left\|\epsilon'_k\otimes\epsilon'_k\right\|_{\mathcal{L}} \\ &\leq \left\|\epsilon'_k\right\|^2 + \mathbb{E}\left\|\epsilon'_k\right\|^2 \end{aligned}$$

and we conclude as in a), with $\mathbb{E}\mathcal{O}_p(1) = \mathcal{O}(1)$ uniformly in k, t
$\|\beta_{it}(x)\| \leq \mathcal{O}_p(1)\hat{\delta}^{2(t-k)}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}\|x\|, \ i = 1, 2.$
For the third term, we abbreviate $\Delta_k = \epsilon'_k\otimes\epsilon'_k - \epsilon_k^*\otimes\epsilon_k^*$ such that
$\|\beta_{3t}(x)\| \leq \hat{\delta}^{2(t-k)}\left\|\Delta_k - \mathbb{E}\Delta_k\right\|_{\mathcal{L}}\|x\|$
Using Cauchy-Schwarz, we have for some generic constant D

$$\begin{aligned} \left|\mathbb{E}B_{kk}^{(s,t)}(x)\right| &\leq \mathbb{E}\left|\langle \beta_{1t}(x) + \beta_{2t}(x) + \beta_{3t}(x), \beta_{1s}(x) + \beta_{2s}(x) + \beta_{3s}(x)\rangle\right| \\ &\leq 4D\hat{\delta}^{2(t+s-2k)}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}^2\|x\|^2 \\ &\quad + 4D\hat{\delta}^{2(t+s-2k)}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}\mathbb{E}\left\|\Delta_k - \mathbb{E}\Delta_k\right\|_{\mathcal{L}}\|x\|^2 \\ &\quad + \hat{\delta}^{2(t+s-2k)}\mathbb{E}\left\|\Delta_k - \mathbb{E}\Delta_k\right\|_{\mathcal{L}}^2\|x\|^2 \end{aligned}$$

As $\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}} \underset{a.s.}{\to} 0$, and as, from Lemma 15 below

$$\mathbb{E}\left\|\Delta_k - \mathbb{E}\Delta_k\right\|_{\mathcal{L}} \leq \sqrt{\mathbb{E}\left\|\Delta_k - \mathbb{E}\Delta_k\right\|_{\mathcal{L}}^2} \text{ and } \mathbb{E}\left\|\Delta_k - \mathbb{E}\Delta_k\right\|_{\mathcal{L}}^2 \leq \mathbb{E}\left\|\Delta_k\right\|_{\mathcal{L}}^2 \underset{p}{\to} 0$$

uniformly in k as $(\epsilon'_k, \epsilon^*_k)$ are identically distributed, we have

$$\left|\mathbb{E}B^{(s,t)}_{kk}(x)\right| \leq D\left\|x\right\|^2 \hat{\delta}^{2(t+s-2k)} o_p(1)$$

uniformly in k, s, t, x. Hence, as for $k = l$, we have $k \leq \min(s,t)$

$$\begin{aligned}\left|\sum_{s,t=0}^{n-1} \mathbb{E}\left\langle d_t(x), d_s(x)\right\rangle\right| &\leq \sum_{s,t=0}^{n-1}\sum_{k=1}^{\min(s,t)} \hat{\delta}^{2(t+s-2k)} D\left\|x\right\|^2 o_p(1) \\ &= \left\|x\right\|^2 o_p(n)\end{aligned}$$

as the first factor is $\mathcal{O}(n)$ by the same calculations as in the proof of Proposition 5.3.3, part b).

c) We consider a third case in detail below. The other components of $\sum_{s,t=0}^{n-1}\mathbb{E}\left\langle A_t(x), A_s(x)\right\rangle$ can be shown to be of order $\left\|x\right\|^2 o_p(n)$ in the same manner, and we finally conclude, as $\mathbb{E}^*\hat{\Gamma}^*_n = \hat{\Gamma}_n + \mathcal{O}_p\left(\frac{1}{n}\right)$,

$$\begin{aligned}n\mathbb{E}\left\|\hat{\Gamma}'_n(x) - \Gamma(x) - \left[\hat{\Gamma}^*_n(x) - \hat{\Gamma}_n(x)\right]\right\|^2 &\sim \frac{1}{n}\sum_{s,t=0}^{n-1}\left\langle A_t(x), A_s(x)\right\rangle \\ &= \left\|x\right\|^2 o_p(1)\end{aligned}$$

uniformly in x and our assertion follows.

As the final case, we study

$$\sum_{s,t=0}^{n-1}\mathbb{E}\left\langle a_t(x), b_s(x)\right\rangle = \sum_{s,t=0}^{n-1}\sum_{k=1}^{s}\mathbb{E}B^{(s,t)}_k(x)$$

where

$$\begin{aligned}B^{(s,t)}_k(x) = \Big\langle &\Psi^t\left[X'_0\otimes X'_0 - \mathbb{E}X'_0\otimes X'_0\right]\left(\Psi^t\right)^T(x) - \hat{\Psi}^t_n\left[X'_0\otimes X'_0 - \mathbb{E}X'_0\otimes X'_0\right]\left(\hat{\Psi}^t_n\right)^T(x), \\ &\Psi^{s-k}\left(\epsilon'_k\otimes X'_0\right)\left(\Psi^s\right)^T(x) - \hat{\Psi}^{s-k}_n\left(\epsilon^*_k\otimes X'_0\right)\left(\hat{\Psi}^s_n\right)^T(x)\Big\rangle\end{aligned}$$

we decompose the left factor of the scalar product into $\gamma_{1t}(x) + \gamma_{2t}(x)$ with $\xi_0 = X'_0\otimes X'_0 - \mathbb{E}X'_0\otimes X'_0$,

$$\begin{aligned}\gamma_{1t}(x) &= \left(\Psi^t - \hat{\Psi}^t_n\right)\xi_0\left(\Psi^t\right)^T(x) \\ \gamma_{2t}(x) &= \left(\Psi^t\right)^T\xi_0\left(\Psi^t - \hat{\Psi}^t_n\right)^T(x)\end{aligned}$$

Analogously, the second factor is $\beta_{1s}(x) + \beta_{2s}(x) + \beta_{3s}(x)$ with

$$\begin{aligned}
\beta_{1s}(x) &= \left(\Psi^{s-k} - \hat{\Psi}_n^{s-k}\right)(\epsilon_k' \otimes X_0')\left(\Psi^s\right)^T(x) \\
\beta_{2s}(x) &= \hat{\Psi}_n^{s-k}(\epsilon_k' \otimes X_0')\left(\Psi^s - \hat{\Psi}_n^s\right)^T(x) \\
\beta_{3s}(x) &= \hat{\Psi}_n^{s-k}(\epsilon_k' \otimes X_0' - \epsilon_k^* \otimes X_0')\left(\hat{\Psi}_n^s\right)^T(x) = \hat{\Psi}_n^{s-k}[(\epsilon_k' - \epsilon_k^*) \otimes X_0']\left(\hat{\Psi}_n^s\right)^T(x)
\end{aligned}$$

As in part a) and b) of the proof, we have for some constant D

$$\begin{aligned}
\|\gamma_{it}(x)\| &\leq D\hat{\delta}^{2t}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}\|\xi_0\|_{\mathcal{L}}\|x\|, \ i = 1, 2 \\
\|\beta_{is}(x)\| &\leq D\hat{\delta}^{2s-k}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}\|\epsilon_k' \otimes X_0'\|_{\mathcal{L}}\|x\|, \ i = 1, 2 \\
\|\beta_{3s}(x)\| &\leq \hat{\delta}^{2s-k}\|(\epsilon_k' - \epsilon_k^*) \otimes X_0'\|_{\mathcal{L}}\|x\|
\end{aligned}$$

We use

$$\begin{aligned}
&\|\xi_0\|_{\mathcal{L}} \leq \|X_0' \otimes X_0'\|_{\mathcal{L}} + \mathbb{E}\|X_0' \otimes X_0'\|_{\mathcal{L}} \leq \|X_0'\|^2 + \mathbb{E}\|X_0'\|^2 \\
&\|\epsilon_k' \otimes X_0'\|_{\mathcal{L}} \leq \|\epsilon_k'\|\|X_0'\| \\
&\|(\epsilon_k' - \epsilon_k^*) \otimes X_0'\|_{\mathcal{L}} \leq \|\epsilon_k' - \epsilon_k^*\|\|X_0'\|
\end{aligned}$$

Using Cauchy-Schwarz, we have for some suitable constant D, as, due to independence of X_0' and $(\epsilon_k', \epsilon_k^*)$

$$\begin{aligned}
\mathbb{E}\|\xi_0\|_{\mathcal{L}}\|\epsilon_k' \otimes X_0'\|_{\mathcal{L}} &\leq \mathbb{E}\left(\|X_0'\|^2 + \mathbb{E}\|X_0'\|^2\right)\|\epsilon_k'\|\|X_0'\| \\
&= \mathbb{E}\|\epsilon_k'\| 2\mathbb{E}\|X_0'\|^3 < \infty \\
\mathbb{E}\|\xi_0\|_{\mathcal{L}}\|\epsilon_k' - \epsilon_k^*\|\|X_0'\| &= \mathbb{E}\|\epsilon_k' - \epsilon_k^*\|\mathbb{E}\|\xi_0\|_{\mathcal{L}}\|X_0'\| \\
&\leq d_2\left(F, \hat{F}_n\right) 2\mathbb{E}\|X_0'\|^3
\end{aligned}$$

So that

$$\begin{aligned}
\left|\mathbb{E}B_k^{(s,t)}(x)\right| &\leq \mathbb{E}\left|\langle\gamma_{1t}(x) + \gamma_{2t}(x), \beta_{1s}(x) + \beta_{2s}(x) + \beta_{3s}(x)\rangle\right| \\
&\leq 4D\hat{\delta}^{2(t+s)-k}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}}^2\|x\|^2 \\
&\quad + 2D\hat{\delta}^{2(t+s)-k}\left\|\Psi - \hat{\Psi}_n\right\|_{\mathcal{L}} d_2\left(F, \hat{F}_n\right)\|x\|^2 \\
&= \hat{\delta}^{2(t+s)-k}\|x\|^2 o_p(1)
\end{aligned}$$

uniformly in t, s, k, x. Therefore,

$$\begin{aligned}\sum_{s,t=0}^{n-1} \mathbb{E}\left\langle a_t(x), b_s(x)\right\rangle &\leq \sum_{s,t=0}^{n-1}\sum_{k=1}^{s} \hat{\delta}^{2(t+s)-k}\left\|x\right\|^2 o_p(1)\\ &\leq \frac{1}{1-\hat{\delta}}\sum_{s,t=0}^{n-1} \hat{\delta}^{2t+s}\left\|x\right\|^2 o_p(1)\\ &\leq \frac{1}{\left(1-\hat{\delta}\right)^2}\frac{1}{1-\hat{\delta}^2}\left\|x\right\|^2 o_p(1)\end{aligned}$$

Hence, this term is of even smaller order $\left\|x\right\|^2 o_p(1)$ compared to $\left\|x\right\|^2 o_p(n)$ of the terms considered in a) and b). □

Lemma 15. *Let $(\epsilon_t', \epsilon_t^*)$, $t = 1, \ldots, n$, be defined as in the proof of Theorem 6.5.1. Then, under the assumptions of that theorem*

$$\mathbb{E}\left\|\epsilon_k' \otimes \epsilon_k' - \epsilon_k^* \otimes \epsilon_k^*\right\|_{\mathcal{L}}^2 \underset{p}{\to} 0,\ n \to \infty$$

Proof. As in the proof of Theorem 6.5.1, part a), we have (for k=l here)

$$\left\|\epsilon_k' \otimes \epsilon_k' - \epsilon_k^* \otimes \epsilon_k^*\right\|_{\mathcal{L}}^2 \leq \left\|\epsilon_k' - \epsilon_k^*\right\|^2 \left(\left\|\epsilon_k'\right| + \left\|\epsilon_k^*\right\|\right)^2$$

Similar to the proof of Proposition 5.3.3, part c), we use that $\left\|\epsilon_k' - \epsilon_k^*\right\| \underset{p}{\to} 0$, which follows from $d_2^2\left(F, \hat{F}_n\right) = \mathbb{E}\left\|\epsilon_k' - \epsilon_k^*\right\|^2 \underset{p}{\to} 0$ and a dominated convergence argument i.e. we have to find a real random variable W with $\mathbb{E}W^4 < \infty$ and $\left\|\epsilon_k' - \epsilon_k^*\right\|^2 \left(\left\|\epsilon_k'\right| + \left\|\epsilon_k^*\right\|\right)^2 \leq W^4$. As $\left\|\epsilon_k' - \epsilon_k^*\right\| \leq \left\|\epsilon_k'\right\| + \left\|\epsilon_k^*\right\|$, we choose $W = \left\|\epsilon_k'\right\| + U$ for some $U \geq 0$ with $\mathbb{E}U^4 < \infty$ and $\left\|\epsilon_k^*\right\| \leq U$. Note that we have $\mathbb{E}\left\|\epsilon_k'\right\|^4 = \mathbb{E}\left\|\epsilon_k\right\|^4 < \infty$ by assumption. Recall that ϵ_k^* can be written as $\tilde{\epsilon}_J$ with J being a Laplace variable in $\{1, \ldots, n\}$, i.e. $pr\left(J = k\right) = \frac{1}{n},\ k = 1, \ldots, n$. Hence,

$$\epsilon_k^* = \tilde{\epsilon}_J = \hat{\epsilon}_J - \frac{1}{n}\sum_{k=1}^{n}\hat{\epsilon}_k = X_J - \hat{\Psi}_n\left(X_{J-1}\right) - \frac{1}{n}\sum_{k=1}^{n} X_k + \frac{1}{n}\sum_{k=1}^{n}\hat{\Psi}_n\left(X_{k-1}\right)$$

and using $\left\|\hat{\Psi}_n\right\| \leq \hat{\delta}$ for large enough n from the proof of Proposition 6.4.1,

$$\left\|\epsilon_k^*\right\| \leq \left\|X_J\right\| + \hat{\delta}\left\|X_{J-1}\right\| + \left\|\frac{1}{n}\sum_{k=1}^{n} X_k\right\| + \hat{\delta}\left\|\frac{1}{n}\sum_{k=1}^{n} X_{k-1}\right\| = U$$

We have $\mathbb{E}U^4 < \infty$ (recall that all expectations are conditional w.r.t. $X_0, \ldots, X_n$ if considering $(\epsilon_t', \epsilon_t^*)$), as, e.g.,

$$\mathbb{E}\left\|X_J\right\|^4 = \frac{1}{n}\sum_{j=1}^{n}\left\|X_J\right\|^4 \leq C$$

for any $C > \mathbb{E}\left\|X_J\right\|^4$ and all large enough n by the strong law of large numbers for strictly stationary real time series. □

Corollary 6.5.1. *Under the assumptions of Theorem 6.5.1, we have for any random variable W independent of $X_0, \epsilon_1, \ldots, \epsilon_n$*

$$\frac{1}{\mathbb{E}\left\|W\right\|^2} n d_2^2 \left(\hat{\Gamma}_n(W) - \Gamma(W), \hat{\Gamma}_n^*(W) - \hat{\Gamma}_n(W)\right) \underset{p}{\to} 0$$

uniformly in $\mathbb{E}\left\|W\right\|^2$

Proof. The result follows from replacing x by W in the proof of Proposition 6.2.2 and using that, due to independence, $\mathbb{E}\left\|W\right\|^2$ shows up as a factor instead of $\mathbb{E}\left\|x\right\|^2$. □

6.6 Some remarks on bootstrapping the autoregressive operator

We would like to show that the bootstrap holds for the estimate of the autoregressive parameter Ψ. In view of the asymptotic results of Mas, [44], we can only expect that the prediction error restricted to the subspace generated by $(\hat{\nu}_1, \ldots, \hat{\nu}_{k_n})$ can be approximated by the bootstrap, e.g. that

$$\sqrt{n}\left(\hat{\Psi}_n - \Psi\hat{\Pi}_{k_n}\right)\left(\tilde{X}_t\right) \text{ and } \sqrt{n}\left(\hat{\Psi}_n^* - \hat{\Psi}_n\hat{\Pi}_{k_n}^*\right)\left(\tilde{X}_t\right)$$

asymptotically have the same distribution. Similar to the approach of Shibata, [51], for investigating the prediction performance of autoregressive models with data-adaptive order, we consider the prediction error of predicting $\tilde{X}_{t+1}$ from $\tilde{X}_t$ for an independent realisation of the time series $\{X_t\}$, i.e. we separate estimation based on $X_0, \ldots, X_n$ and prediction which is done for $\{\tilde{X}_t\}$. Now, from Lemma 10,

$$\begin{aligned}
\sqrt{n}\left(\hat{\Psi}_n - \Psi\hat{\Pi}_{k_n}\right)\left(\tilde{X}_t\right) &= \frac{1}{\sqrt{n}} S_n \hat{\Gamma}_n^{\dagger}\left(\tilde{X}_t\right) \\
\sqrt{n}\left(\hat{\Psi}_n^* - \hat{\Psi}_n\hat{\Pi}_{k_n}^*\right)\left(\tilde{X}_t\right) &= \frac{1}{\sqrt{n}} S_n^* \hat{\Gamma}_n^{*\dagger}\left(\tilde{X}_t\right)
\end{aligned}$$

However, showing that the bootstrap approximation holds for these terms, we cannot just combine the arguments leading to Propositions 6.2.2 and 6.4.1. The main issue is that

$\hat{\Gamma}_n(x) = \sum_{j=1}^{\infty} \hat{\lambda}_j \left\langle x, \hat{\nu}_j \right\rangle \hat{\nu}_j = \frac{1}{n}\sum_{t=0}^{n-1} X_t \otimes X_t$

has an explicit representation in terms of the data whereas we do not have an easy similar property for

$\hat{\Gamma}_n^\dagger(x) = \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j} \langle x, \hat{\nu}_j \rangle \hat{\nu}_j$
or even for the eigenvectors $\hat{\nu}_j$. In proving his asymptotic Theorem 3.1, Mas [44] uses pertubation theory of operators to get representations like

$$\hat{\nu}_k \otimes \hat{\nu}_k = \frac{1}{2\pi i} \int_{\mathcal{C}_n} \left(zI - \hat{\Gamma}_n \right)^{-1} dz$$

where $\mathcal{C}_n$ is the boundary of an open subset of $\mathbb{C}$ containing λ_k, but no other eigenvalue, e.g. the boundary of a small circle with center λ_k. This leads to a corresponding representation

$$\hat{\Gamma}_k^\dagger = \sum_{j=1}^{k_n} \int_{\mathcal{C}_j} \frac{1}{z} \left(zI - \hat{\Gamma}_n \right)^{-1} dz$$

of the truncated inverse of $\hat{\Gamma}_n$ in terms of $\hat{\Gamma}_n$. Under appropriate assumptions Mas shows as his Proposition 5.1

$$\frac{1}{\sqrt{nk_n}} S_n \left(\hat{\Gamma}_n^\dagger - \Gamma^\dagger \right) (X_{n+1}) \underset{p}{\to} 0 \text{ with } \Gamma^\dagger = \sum_{j=1}^{k_n} \frac{1}{\lambda_j} \nu_j \otimes \nu_j$$

It is easy to check, that this result also holds with $\tilde{X}_t$ replacing X_{n+1}.

Note that in the definition of S_n on p.1242 of Mas, [44], a factor $\frac{1}{n}$ is missing. Otherwise, equation (13) of Mas would not hold - compare also our Lemma 10. As $\mathbb{E} \sum_{t=1}^n X_{t-1} \otimes \epsilon_t(x) = 0$ and $\mathbb{E} \left\| \sum_{t=1}^n X_{t-1} \otimes \epsilon_t \right\|^2 = \mathbb{E} \langle X, x \rangle^2 \mathbb{E} \left\| \epsilon_1 \right\|^2$, we have to scale with $\frac{1}{\sqrt{n}}$, not with $\sqrt{n}$ as in the proposition 5.1 of Mas.

If we consider a fixed $k_n = p$, writing $\hat{\Gamma}_{n,p}^\dagger, \hat{\Psi}_{n,p}$ etc. to make this visible, we even have, as this is just a special case

$$\frac{1}{\sqrt{n}} S_n \left(\hat{\Gamma}_{n,p}^\dagger - \Gamma_p^\dagger \right) (X_{n+1}) \underset{p}{\to} 0 \text{ with } \Gamma_p^\dagger = \sum_{j=1}^{p} \frac{1}{\lambda_j} \nu_j \otimes \nu_j \tag{6.5}$$

To use this result we need the additional assumptions 3.1-3.4 of Mas, [44], from our section 3.2 above. Assumption 3.1 is part of our usual assumptions in this chapter. It is easily checked that it holds for the bootstrap process X_t^* too for all large enough n, i.e. $ker\left(\hat{\Gamma}_n\right) = 0$ a.s., $\mathbb{E}^* \left\| \epsilon_t^* \right\|^2 \leq C < \infty$ uniformly in n and $\left\| \hat{\Psi}_n \right\|_{\mathcal{L}} < 1$ a.s.

However, the other assumptions and, in particular 3.3 about the Karhunen-Loéve expansion, are hard to check for X_t^* such that we do not easily have the analogue of 6.5

in the bootstrap world, though, in view of Propositions 6.2.1 and 6.2.2 it is likely true too. If we assume this property, then we can show immediately that the bootstrap works for the predictor $\hat{\Psi}_{n,p}$, for fixed $k_n = p$. Note, that for constructing the bootstrap data we still use the estimate $\hat{\Psi}_n$ with increasing k_n as, otherwise, the residuals $\hat{\epsilon}_t$ would not approximate the innovations ϵ_t.

Proposition 6.6.1. *Under the conditions of Theorem 6.3.1 and assuming additionally 3.2-3.4 and*

$$\frac{1}{\sqrt{n}} S_n^* \left(\hat{\Gamma}_{n,p}^{*\dagger} - \Gamma_p^\dagger \right) \left(\tilde{X}_t \right) \underset{p}{\to} 0 \tag{6.6}$$

we have for $\{\tilde{X}_t\}$ from an independent realisation of $\{X_t\}$
a) $\sqrt{n} \left(\hat{\Psi}_{n,p} - \Psi \hat{\Pi}_p \right) \left(\tilde{X}_t \right) \underset{w}{\to} S_\infty \Gamma_p^\dagger(\tilde{X}_t)$, for $n \to \infty$, where $S_\infty, \Gamma_p^\dagger$ are as in Theorem 6.2.1.
b) $\sqrt{n} \left(\hat{\Psi}_{n,p}^* - \hat{\Psi}_n \hat{\Pi}_p^* \right) \left(\tilde{X}_t \right) \underset{w}{\to} S_\infty \Gamma_p^\dagger(\tilde{X}_t)$, for $n \to \infty$, too, i.e. the bootstrap approximation holds.

Proof. a) follows from Theorem 6.2.1. As in the proofs of Propositions 6.2.1 and 6.2.2, we choose $X_0', (\epsilon_t', \epsilon_t^*)$, $t = 1, \ldots, n$, where $X_0, \ldots, X_n$ are treated as given. Then, $\sqrt{n} \left(\hat{\Psi}_{n,p}' - \Psi \hat{\Pi}_p' \right) \left(\tilde{X}_t \right)$ has the same distribution as $\sqrt{n} \left(\hat{\Psi}_{n,p} - \Psi \hat{\Pi}_p \right) \left(\tilde{X}_t \right)$. From Lemma 10,

$$\sqrt{n} \left(\hat{\Psi}_{n,p}' - \Psi \hat{\Pi}_p' \right) = \frac{1}{\sqrt{n}} S_n' \hat{\Gamma}_{n,p}'^\dagger, \quad \sqrt{n} \left(\hat{\Psi}_{n,p}^* - \hat{\Psi}_n \hat{\Pi}_p^* \right) = \frac{1}{\sqrt{n}} S_n^* \hat{\Gamma}_{n,p}^{*\dagger}$$

We decompose, writing here $\sim$ for equal distribution for $n \to \infty$

$$\begin{aligned} \frac{1}{\sqrt{n}} S_n' \hat{\Gamma}_{n,p}'^\dagger \left(\tilde{X}_t \right) &= \frac{1}{\sqrt{n}} S_n' \left(\Gamma_p^\dagger \left(\tilde{X}_t \right) \right) + \frac{1}{\sqrt{n}} S_n' \left(\hat{\Gamma}_{n,p}'^\dagger - \Gamma_p^\dagger \right) \left(\tilde{X}_t \right) \\ &\sim \frac{1}{\sqrt{n}} S_n' \left(\Gamma_p^\dagger \left(\tilde{X}_t \right) \right) \end{aligned}$$

as the second term on the right-hand side of the first line converges to 0 in probability from Proposition 5.1 of Mas, [44].

From assumption 6.6 we have the analogue for the bootstrap data

$$\frac{1}{\sqrt{n}} S_n^* \hat{\Gamma}_{n,p}^{*\dagger} \left(\tilde{X}_t \right) \sim \frac{1}{\sqrt{n}} S_n^* \left(\Gamma_p^\dagger \left(\tilde{X}_t \right) \right)$$

Then, from Proposition 6.4.1

$$d_2^2 \left(\frac{1}{\sqrt{n}} S_n \left(\Gamma_p^\dagger \left(\tilde{X}_t \right) \right), \frac{1}{\sqrt{n}} S_n^* \left(\Gamma_p^\dagger \left(\tilde{X}_t \right) \right) \right) \underset{p}{\to} 0$$

such that b) follows from a) and Lemma 8.3 of Bickel and Freedman, [6]. □

Chapter 7

Simulation Studies and Results

In this chapter results for the simulations carried out are reported. The goal of the simulation study is to assess the theoretical results obtained in the previous chapter. In particular, we carry out a simulation to test the goodness of fit of the model in which case the critical value is obtained by bootstrap technique. We also carry out a simulation to estimate the model parameter using the bootstrap technique and also to check whether the distribution obtained by bootstrapping is close to that obtained asymptotically i.e. to investigate numerically the weak convergence of the predictions.

7.1 Weak Convergence of the FAR(1) Process

The FAR(1) series is generated according to the model

$$X_{n+1}(t) = \int_0^1 \psi(t,s)X_n(s)ds + \epsilon_{n+1}(t) \tag{7.1}$$

The sample consists of 252 curves $X_n(t)$, including a burn-in phase of 50 observations which we do not use for estimation, each curve with 100 data points between [0,1] which are generated from an initial observation X_0 of zeros. Residuals (ϵ) are generated from a Wiener process, chosen for its simplicity, randomness and constant mean zero.

The Gaussian kernel is chosen for the process and numerical integration employed to estimate the value of the constant C such that $\|\Psi\|= 0.5$ or $\|\Psi\| = 0.8$. It should be quickly noted that the constant C is required and chosen so as to ensure the observed process will have a stationary solution. This follows directly from the scalar autoregressive

process where to attain stationarity the condition $\psi < 1$ should be fulfilled. The Gaussian kernel is given by

$$\psi(t,s) = Cexp\{-(t^2+s^2)/2\} \tag{7.2}$$

The constant kernel $\psi(t,s) = C$ and the sloping kernel $\psi(t,s) = Ct$ were also considered and their constants too were chosen such that the condition $\|\Psi\| = 0.5$ was fulfilled.

A summary of the simulation process is given below;

1. Start with $X_0 = 0$.

2. Generate a sample $X_1, ..., X_{n+2}$. Allow a burn-in period of 50 observations, such that the actual sample size used for estimation is $N = 200$. The first N functions are used in the estimation of $\hat{\Psi}_n$ which is then used to predict $\hat{X}_{n+2}$ i.e $\hat{X}_{n+2} = \hat{\Psi}_n(X_{n+1})$.

3. From the relationship $\hat{\Psi}_n = \Psi\hat{\Pi}_{k_N} + \frac{1}{N}S_n\hat{\Gamma}_n^{\dagger}$, we calculate $\frac{1}{N}S_n\Gamma_n^{\dagger}$ and subtract this value from $\hat{\Psi}_n$ to obtain $\Psi\hat{\Pi}_{k_N}$. $S_n\hat{\Gamma}_n^{\dagger}$ is defined as

$$S_n\Gamma_n^{\dagger}(x) = \sum_{k=1}^{n}\sum_{l\leq k_N}\sum_{j\leq k_N}\hat{\lambda}_l^{-1}\left\langle x,\hat{e}_l\right\rangle\left\langle X_{k-1},\hat{e}_l\right\rangle\left\langle \epsilon_k,\hat{e}_j\right\rangle\epsilon_j \tag{7.3}$$

4. Finally we obtain

$$\sqrt{\frac{N}{k_N}}\left(\hat{\Psi}_n(X_{n+1}) - \Psi\hat{\Pi}_{k_N}(X_{n+1})\right) \tag{7.4}$$

5. For the bootstrap process, we obtain the residuals by subtracting $\hat{\Psi}(X_{n+1})$ from $\Psi\hat{\Pi}_{k_N}(X_n)$ and center them.

6. We generate B bootstrap samples $X_{n+1}^* = \hat{\Psi}_n(X_n^*) + \epsilon_{n+1}^*$ from a fixed starting point say $X_0^* = X_0$ and as above allow for a burn-in of 50 observations.

7. $\hat{\Psi}_n^*$ is calculated in the same manner as above for each bootstrap sample and we use the average for prediction.

8. Finally we compute the quantity

$$\sqrt{\frac{N}{k_N}}\left(\hat{\Psi}_n^*(X_{n+1}) - \hat{\Psi}_n(X_{n+1})\right) \tag{7.5}$$

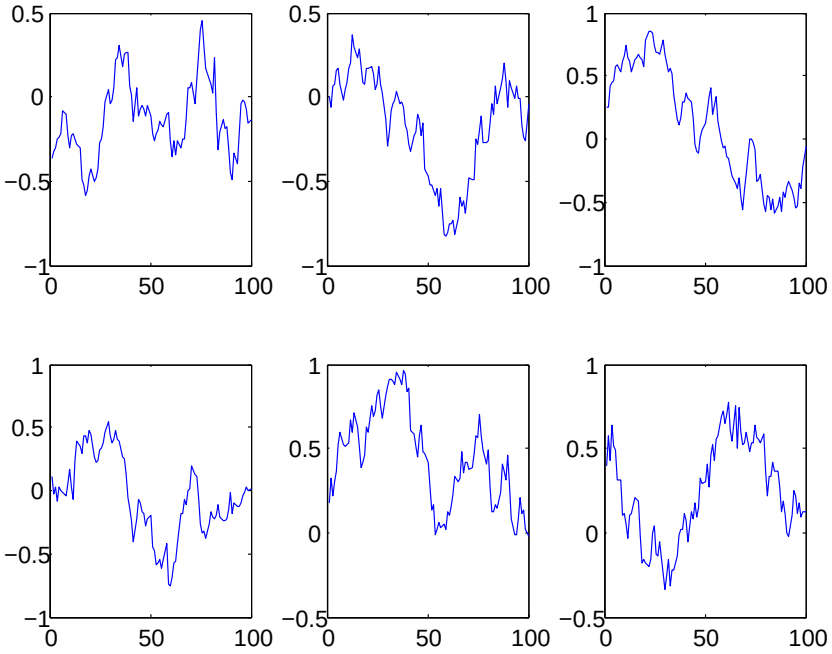


Figure 7.1: Sample of functional observations

9. The whole process is repeated M=1000 times and the distances between the bootstrap and empirical approximations are compared.

A sample of six consecutive curves computed using the Gaussian kernel with $\|\Psi\| = 0.5$ can be seen in Figure 7.1. As already mentioned in the preceding Chapters, due to an ill posed inverse problem, we employ a truncated version of the population parameter. We illustrate the effect of this truncation diagrammatically in Figure 7.2. The figures were computed using $B = 200$ replications, and a Gaussian kernel with $\|\Psi\| = 0.5$.

It is obvious from Figure 7.2 that the bootstrap estimator of the function is very close to the empirical predictor, a feature that is highly desirable since we aim at employing the bootstrap predictor in place of the empirical predictor in the event that the sample size is small so that we cannot rely on asymptotic estimations. Another feature that is noticeable is the fact that both the bootstrap and empirical estimates of the observation are smoother (shorter). This is obviously due to the fact that the empirical and bootstrap estimators are projected onto the first p principal components that represent most of the variability. However, although the graphical features are not quite satisfactory, $\hat{\Psi}_n$ still retains its position as the best estimator of the unknown underlying linear operator, see Kokoszka, [28] for more details on the comparison of the estimators of Ψ.

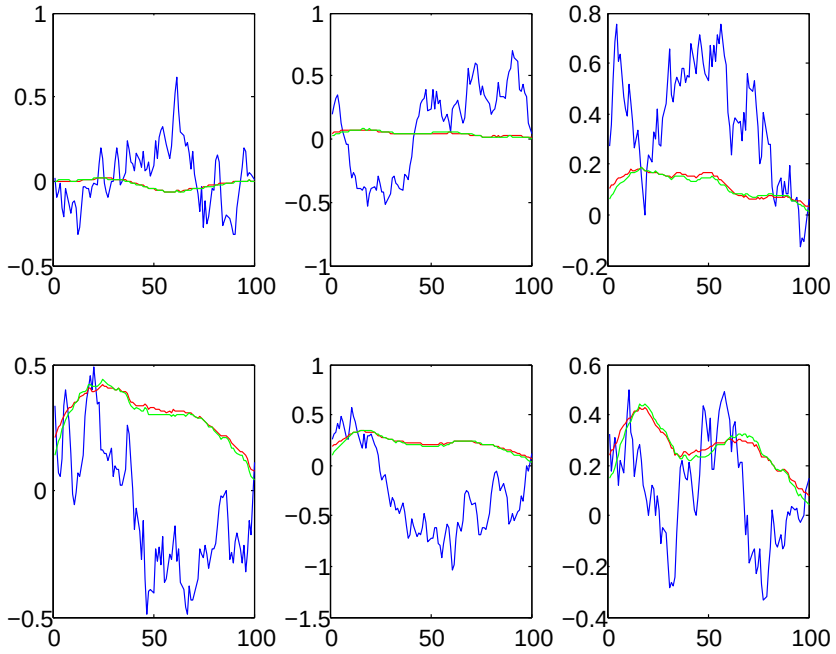


Figure 7.2: Sample of functional observations with their empirical and bootstrap predictions

Figure 7.3 shows the observation together with its bootstrap and empirical estimators of the second curve in Figure 7.2 on the left while on the right, with a different scale, we have only the empirical and bootstrap predictions. This is to have a clearer picture of their form.

As noted in Mas, [44], the linear operator $\hat{\Psi}_n$ does not converge in distribution to Ψ and a truncation is necessary in order to ensure convergence. Due to this truncation, we have that the corresponding truncated observations are smooth in appearance as compared to the true observations and they therefore are closer to the empirical and bootstrap observations in terms of graphical appearance. An illustration of this is given in Figure 7.4. The Figure 7.4 shows six consecutive randomly chosen truncated observations (blue) together with their empirical (red) and bootstrap (green) predictions. The bootstrap observations were generated from a naive bootstrap procedure with 200 observations and Gaussian kernel with $\|\Psi\| = 0.5$.

We seek to compare the sequences 7.4 and 7.5. When the sequences are close, then we can infer that the bootstrap predictions are close to the empirical predictions. This in turn implies that in the event that we have a small sample then we can employ the bootstrap for estimating the operator and carrying out predictions. Apart from this, using the

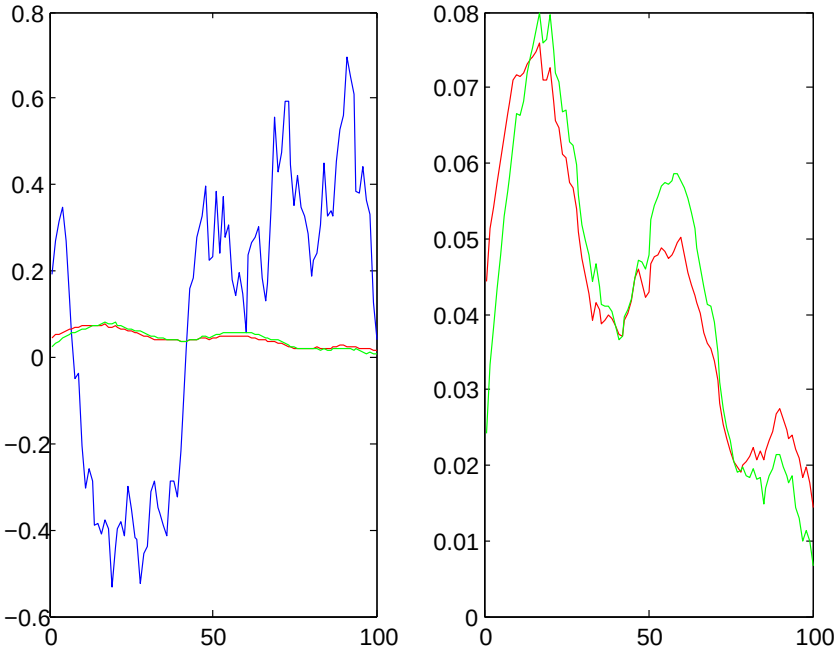


Figure 7.3: Empirical and bootstrap predictions

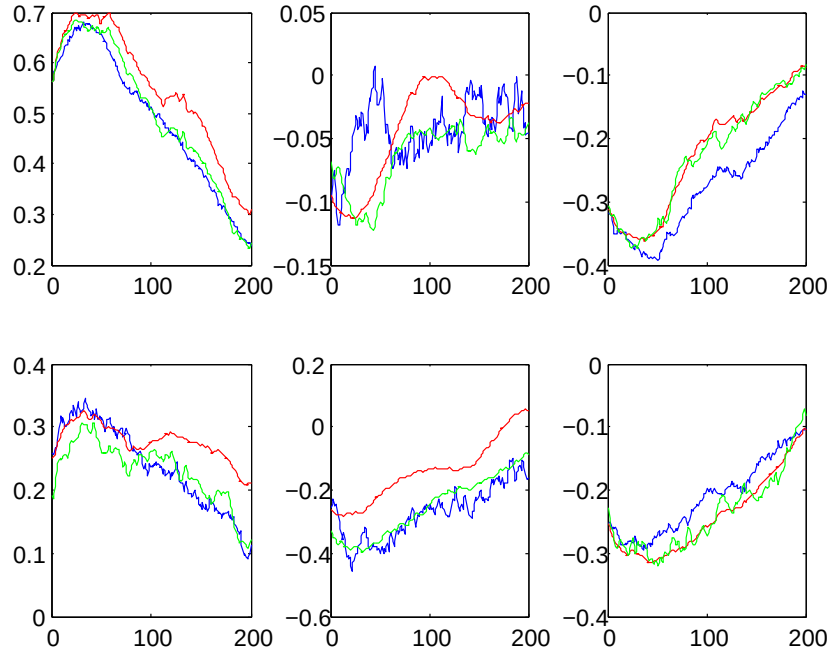


Figure 7.4: Truncated observations, Empirical and bootstrap predictions

Table 7.1: Prediction errors; Brownian motion innovations, Gaussian kernel $\|\Psi\|= 0.8$, Naive bootstrap

B=1000	E_n	R_n	BE_n	BR_n
Average	0.0766	0.0658	0.0760	0.0657

Table 7.2: Prediction errors; Brownian motion innovations, Gaussian kernel $\|\Psi\|= 0.8$, Block bootstrap

B=1000	E_n	R_n	BE_n	BR_n
Average	0.0873	0.0745	0.0708	0.06886

bootstrap to estimate the asymptotic distribution of the parameter of interest (operator Ψ), will also enable to compute other statistics of interest for instance confidence bands or to carry out hypothesis tests among other things. We compare the sequences in two ways, the distance between the predictions E_n and R_n which are defined below and diagrammatically by projecting the differences between the curves into the most relevant directions (further details below) as in Ferraty *et. al.*, [18].

To estimate the prediction error at time n, we consider the quantities E_n and R_n from Didericksen *et. al.*, [16], which are defined as

$$E_n = \sqrt{\int_0^1 \left(X_n(t) - \hat{X}_n(t)\right)^2 dt} \quad \text{and} \quad R_n = \int_0^1 \left|X_n(t) - \hat{X}_n(t)\right| dt \tag{7.6}$$

Their bootstrap counterparts are given by

$$BE_n^* = \sqrt{\int_0^1 \left(\hat{X}_n(t) - \hat{X}_n^*(t)\right)^2 dt} \quad \text{and} \quad BR_n^* = \int_0^1 \left|\hat{X}_n(t) - \hat{X}_n^*(t)\right| dt$$

The quantities above allow us to assess the distance between the empirical and bootstrap approximations of the predictions. A summary of the prediction errors for the real and bootstrap cases is given in Tables 7.1 and 7.2.

To obtain a graphical comparison, we consider the method employed by Ferraty *et. al.* which involves comparing the density of the componentwise bootstrapped error $\left\langle \hat{\Psi}_n^*(X) - \hat{\Psi}_n(X), \hat{u}_j \right\rangle$, with that of the true error $\left\langle \hat{\Psi}_n(X) - \Psi\hat{\Pi}_{k_N}(X_{n+1}), \hat{u}_j \right\rangle$, $j = 1,,,4$. $\hat{\nu}_j$ is the jth orthonormal eigenfunction of

$$C_{\hat{\Psi}_n(X)} = \frac{1}{n}\sum_{i=1}^{n} \left\langle \hat{\Psi}_n(X_i), . \right\rangle \hat{\Psi}_n(X_i) \tag{7.7}$$

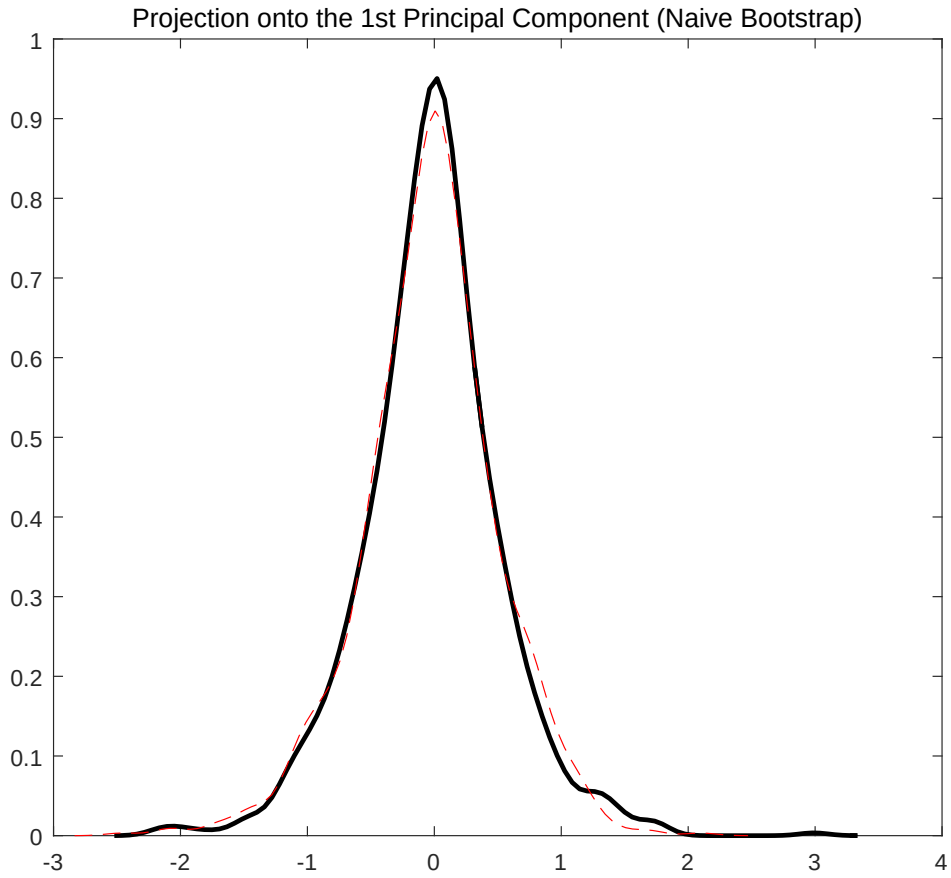


Figure 7.5: Empirical and Naive bootstrap predictions

These particular eigenfunctions are chosen because they are relevant directions for the variable $\hat{\Psi}_n(X)$, and the method in general helps in checking whether we have asymptotic normality in the relevant directions, i.e the first p most important principal components.

The Bspline basis of order four is used over the interval $[0, 1]$. Figures **??** and **??** give the projections onto the first principal component for the naive and block (length=25) bootstrap cases. It is evident that the naive bootstrap gives a better approximation as far as the graph is concerned. In terms of the distances E_n and R_n we observe the same behaviour with the naive bootstrap posting slightly better results. We quickly note however that with better tuning of the block length the performance of the block bootstrap may be improved. In this case the results obtained may not be so reliable since a block of length 25 ensures that there are only 10 blocks to resample from which may be considered too small.

We do not consider the 2nd, 3rd and 4th principal components, as for this model the

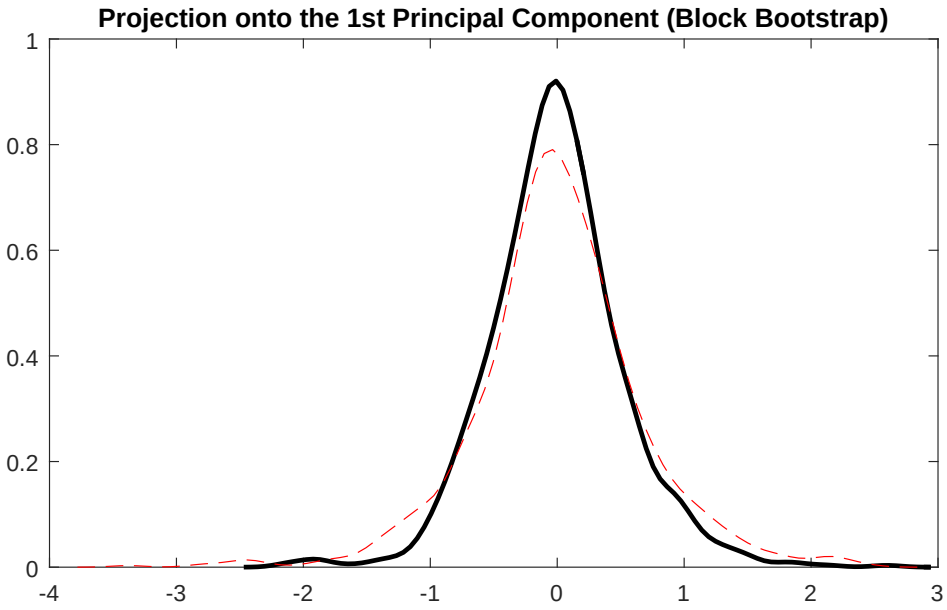


Figure 7.6: Empirical and Block bootstrap predictions

projection onto the 1st principle component already explains approximately 95% of variability.

7.2 Goodness of Fit

To test the goodness of fit of the FAR(1) model, we test the null hypothesis

$$H_0 : \Psi = 0 \quad versus \quad H_1 : \Psi \neq 0 \tag{7.8}$$

where Ψ is as defined in Equation 2.1.

From the relation $C = \Psi\Gamma$, (see equations 2.4 for a definition) it is evident that when $\Psi = 0$, then $C = 0$ too. The test statistic employed is therefore of the form

$$T_N(p,p) = N \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\lambda}_i^{-1} \hat{\lambda}_j^{-1} \left\langle C\left(\hat{\nu}_i\right), \hat{\nu}_j \right\rangle^2 \tag{7.9}$$

where $\left\langle C\left(\hat{\nu}_i\right), \hat{\nu}_j \right\rangle = \dfrac{1}{N-1} \sum_{n=1}^{N-1} \left\langle X_n, \hat{\nu}_i \right\rangle \left\langle X_{n+1}, \hat{\nu}_j \right\rangle$ and we reject or fail to reject the null hypothesis by comparing the value of the test statistic (kernel) for fixed p with the critical value obtained by bootstrapping. This test statistic has been chosen since it is clear that if H_0 fails then $\Psi(\hat{\nu}_j) \neq 0$ for some $j \in 1, ...p$, since we are projecting onto the subspace consisting of the first p principle components. If $\Psi(\hat{\nu}_j) \neq 0$, then from the test

statistics definition we have that the power goes to 1 as $n \to \infty$, thus the test is consistent. The same test statistic was also employed for a scalar response model (functional linear model) and it was found to give good results, see [50]. It is also employed by [7] for the FAR(p) of which the FAR(1) is a special case. A summary of the testing procedure is given below.

Under the null hypothesis, we simulate $M = 1000$ time series $X_{t+1} = \epsilon_{t+1}$. For each $m \in 1, ..., M$;

1. Perform functional principal component analysis and compute the value of the test statistic T_N. Estimate the kernel operator $\hat{\Psi}_n$.

2. Compute the residuals $\hat{\epsilon}_{n+1} = X_{n+1} - \hat{\Psi}(X_n)$ and center them.

3. Using the centered residuals generate $B = 1000$ bootstrap samples each time computing the test statistic T^*. Compute the critical value C^*, which is the 95% quantile of $T_b^*, b = 1, ..., B$.

4. Reject H_0 if $T_b^* > C^*$ and compute the probability value $\dfrac{\hat{p} = \#\{b : T_b > C^*\}}{M}$.

Under the alternative, we simulate $M = 1000$ time series $X_{t+1} = \Psi(X_t) + \epsilon_{t+1}$. We considered three kernels ψ, namely Gaussian, sloping kernel and the constant kernel. For each $m \in 1, ..., M$, steps $1, ..., 4$ above were repeated.

7.2.1 Results

When the Gaussian kernel with $p = 4$ and $B = M = 1000$ was employed, the probability-value (p-value) under the null hypothesis was 0.044 compared to the nominal level0.05 while under the alternative hypothesis we obtained 1. This can be tied to the fact that for functional data each curve contains a lot of information and therefore it is simpler to differentiate between white noise and an FAR(1) process. The same results were obtained for the sloping and constant kernels, whose constants C, like that in the Gaussian case were chosen so as to fulfill the condition $\|\Psi\| = 0.55$.

7.3 Kernel estimation using the bootstrap

Following Didericksen *et. al.* [16], we compare the kernel estimated by means of the bootstrap with that estimated from the sample. The main aim of this work is to establish whether the bootstrap estimated kernel is close to the empirically estimated kernel as is desired.

To check the accuracy of the estimation, we compute the root mean squared error(RMSE) and the Averaged Distance (AD), defined as

$$RMSE = \sqrt{\int_0^1 \int_0^1 \left(\hat{\psi}(t,s) - \psi(t,s)\right)^2 dsdt} \tag{7.10}$$

$$AD = \int_0^1 \int_0^1 \left|\hat{\psi}(t,s) - \psi(t,s)\right| dsdt \tag{7.11}$$

As in [16], we compare different kernels to see whether significant differences can arise as a result of the kernel employed in generating the data. Kernels considered include

1. Gaussian: $\psi(t,s) = Cexp\left\{-0.5\left(t^2+s^2\right)\right\}$
2. Identity: $\psi(t,s) = C$
3. Sloping plane (t): $\psi(t,s) = Ct$

where the constant C is chosen such that $\|\Psi\| = 0.5$ and $\|\Psi\| = 0.8$.

We consider the first 4 principal components which should explain most of the variability in the sample. The value 85% is the cumulative percentage of variability that is considered standard and is the acceptable value in most studies, see for instance [28]. It is evident from [16] that the distances decrease with increasing N. The estimated kernel ψ however has peculiar behaviour in that its mean squared error or mean absolute error increase with increasing number of principal components which is counter intuitive considering the opposite should occur. This can be attributed to the fact that the values of λ_j are very small and therefore a small error in their estimation leads to a larger error in their inverse.

Our aim is to compute the bootstrap estimate of the kernel and the distances between the bootstrap estimates $\hat{\psi}^*(t,s)$ and the estimator of the true kernel $\hat{\psi}(t,s)$. Two types

of bootstrap were employed; the block bootstrap and the naive bootstrap which differ in the manner of resampling. For the innovations we consider the Brownian bridge and Brownian motion.

When considering the block bootstrap, the resampling procedure is as follows;

1. Generate a FAR(1) series according to equation 2.1. In this case the Gaussian kernel was employed and for the error we employed the Brownian bridge defined as

$$BB(t) = W(t) - tW(1) \tag{7.12}$$

where $W(.)$ is the standard Wiener process. A burn-in period of 50 observations is allowed. $\hat{\psi}(t,s)$ is estimated as

$$\hat{\psi}_p(t,s) = \frac{1}{N-1}\sum_{k,l=1}^{p} \hat{\psi}_{kl}\hat{\nu}_k(t)\hat{\nu}_l(s) \tag{7.13}$$

where $\hat{\psi}_{ji} = \hat{\lambda}_i^{-1}(N-1)^{-1}\sum_{n=1}^{N-1}\langle X_n, \hat{\nu}_i\rangle\langle X_{n+1}, \hat{\nu}_j\rangle$.

2. For each sample generated, a block bootstrap with B=50 was carried out each time computing $\hat{\psi}_b^*(t,s),\ b = 1, \cdots, B$. The bootstrap kernel estimators are added together before being averaged to produce the final bootstrap estimate.

3. The root mean squared error (RMSE) and averaged distance (AD) are then computed.

The procedure is repeated 50 times.

Although we tried several different block lengths, in all cases the perfomance of the block bootstrap was not very good. It is observed that the distance between the bootstrap and empirical kernels decreases with increasing N and increases with increasing p.

The resampling procedure for the naive bootstrap is given as

1. Generate a FAR(1) series according to equation 2.1 with given kernel (Gaussian, identity or sloping) and innovation (Brownian motion or Brownian Bridge).

2. Estimate the kernel $\hat{\psi}(t,s)$ and $\hat{X}_n$, the predictions obtained by using the operator estimated from the sample.

3. Compute the residuals which are given by $\hat{\epsilon}_n = X_n - \hat{X}_n$ and center them to obtain $\tilde{\epsilon}$.

4. Resample the residuals with replacement and generate an FAR (1) process with the kernel operator estimated in Step 2 above.

5. Estimate the bootstrap kernels $\hat{\psi}_b^*(t,s),\ b = 1, \cdots, B$ and compute their average to obtain the estimate to be used in comparing distances.

6. Repeat the whole procedure M times.

As in the block bootstrap above, we used $M = 50$ replications.

In the naive bootstrap case, we have that the distance between the bootstrap and true estimator also reduces with increasing sample size N. We notice too that 50 bootstrap replications are enough to provide a good estimate and increasing the number of bootstrap replications does not improve the estimates.

The naive bootstrap proved to be better at estimating the distance between the bootstrap and empirical kernels as compared to the block bootstrap. As in the empirical case, [16], the distances reduced with increasing N but as observed in the true case increasing p from $p = 2$ to $p = 4$ instead of reducing the distance increased it. The same reason is given for the bootstrap case i.e the eigenvalues estimated decrease with increasing p and therefore a small error in their estimation results in a large error when they are inverted. Didericksen *et. al.*, [16] suggest remedying this situation by adding a baseline $\hat{b}$ to $\lambda_i,\ i > 2$, where $\hat{b} = 1.5\left(\hat{\lambda}_1 + \hat{\lambda}_2\right)$. This ensures that the MSE and AD do not increase with p. In the bootstrap case we do not employ this remedy, we simply note that the bootstrap kernel is very close to the empirical kernel which was our main aim.

Results of simulations are given in Tables [7.3-7.14]. CPV is an abbreviation for cumulative percentage of variance which is a method that is employed inorder to decide the number of principal components to be included in the estimation. A CPV of 85% and above is considered sufficient in most cases in literature. The performance improves with increasing sample size.

Table 7.3: Kernel estimation errors; Brownian bridge innovations, Gaussian kernel $\|\Psi\|=0.5$, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.3880 (0.0204)	0.0998 (0.0055)	0.3321 (0.0166)	0.0830 (0.0043)
CPV	76.9%	77.73%	76.9%	77.73%
$p=3$	0.5651 (0.0268)	0.1560 (0.0207)	0.4538 (0.0075)	0.1243 (0.0057)
CPV	83.87%	84.85%	83.87%	84.85%
$p=4$	0.8522 (0.0386)	0.2596 (0.0308)	0.6686 (0.0139)	0.2025 (0.0107)
CPV	88.79%	90.56%	88.79%	90.56%
$N=100$				
$p=2$	0.3328 (0.0109)	0.0706 (0.0096)	0.2860 (0.0039)	0.0595 (0.0032)
CPV	77.15%	77.57%	77.15%	77.57%
$p=3$	0.3879 (0.0141)	0.0964 (0.0112)	0.3209 (0.0036)	0.0784 (0.0029)
CPV	83.87%	84.34%	83.87%	84.34%
$p=4$	0.5968 (0.0232)	0.1580 (0.0078)	0.4741 (0.0175)	0.1251 (0.0057)
CPV	87.38%	87.97%	87.38%	87.97%
$N=200$				
$p=2$	0.2884 (0.0090)	0.0611 (0.0079)	0.2511 (0.0022)	0.0515 (0.0019)
CPV	76.51%	76.91%	76.51%	76.91%
$p=3$	0.3108 (0.0158)	0.0745 (0.0129)	0.2566 (0.0028)	0.0600 (0.0021)
CPV	84.98%	85.84%	84.98%	85.84%
$p=4$	0.4563 (0.0231)	0.1072 (0.0179)	0.3676 (0.0079)	0.0859 (0.0062)
CPV	87%	87.39%	87%	87.39%

Table 7.4: Kernel estimation errors; Brownian bridge innovations, Gaussian kernel $\|\Psi\|=0.8$, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.3661 (0.0186)	0.1561 (0.0157)	0.3062 (0.0067)	0.1315 (0.0055)
CPV	81.53%	81.33%	81.53%	81.33%
$p=3$	0.5219 (0.0255)	0.1551 (0.0072)	0.4175 (0.0200)	0.1257 (0.0060)
CPV	87.88%	87.7%	87.88%	87.7%
$p=4$	0.8037 (0.0366)	0.2720 (0.0162)	0.6275 (0.0286)	0.2118 (0.0127)
CPV	90.68%	90.87%	90.68%	90.87%
$N=100$				
$p=2$	0.2774 (0.0118)	0.1293 (0.0102)	0.2344 (0.0035)	0.1091 (0.0030)
CPV	83.2%	83.1%	83.2%	83.1%
$p=3$	0.3524 (0.0152)	0.1036 (0.0031)	0.2877 (0.0122)	0.0831 (0.0025)
CPV	88.97%	88.66%	88.97%	88.66%
$p=4$	0.5457 (0.0236)	0.1473 (0.0058)	0.4309 (0.0184)	0.1178 (0.0046)
CPV	91.37%	91.30%	91.37%	91.30%
$N=200$				
$p=2$	0.2343 (0.0105)	0.1203 (0.0092)	0.1999 (0.0025)	0.1023 (0.0022)
CPV	83.33%	83.62%	83.33%	83.62%
$p=3$	0.2625 (0.0146)	0.0772 (0.0117)	0.2152 (0.0021)	0.0613 (0.0016)
CPV	89.16%	89.09%	83.33%	83.62%
$p=4$	0.4279 (0.0221)	0.1027 (0.0172)	0.3384 (0.0047)	0.0828 (0.0036)
CPV	91.43%	91.44%	91.43%	91.44%

Table 7.5: Kernel estimation errors; Brownian bridge innovations, Identity kernel $\|\Psi\|=0.5$, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.4006 (0.0210)	0.1005 (0.0173)	0.3484 (0.0052)	0.0842 (0.0041)
CPV	76.62%	77.49%	76.62%	81.33%
$p=3$	0.5625 (0.0267)	0.1564 (0.0210)	0.4554 (0.0073)	0.1252 (0.0055)
CPV	83.79%	84.78%	83.79%	84.78%
$p=4$	0.8542 (0.0374)	0.2627 (0.0140)	0.6733 (0.0296)	0.2054 (0.0106)
CPV	87.54%	88.67%	87.54%	88.67%
$N=100$				
$p=2$	0.3483 (0.0109)	0.0711 (0.0095)	0.3041 (0.0039)	0.0602 (0.0031)
CPV	76.9%	77.36%	76.9%	77.36%
$p=3$	0.3898 (0.0140)	0.0978 (0.0112)	0.3257 (0.0035)	0.0803 (0.0029)
CPV	83.77%	84.27%	83.77%	84.27%
$p=4$	0.5957 (0.0225)	0.1579 (0.0169)	0.4755 (0.0075)	0.1263 (0.0057)
CPV	87.28%	87.90%	87.28%	87.90%
$N=200$				
$p=2$	0.3038 (0.0086)	0.0593 (0.0077)	0.2679 (0.0019)	0.0508 (0.0017)
CPV	76.21%	76.64%	76.21%	76.64%
$p=3$	0.3149 (0.0162)	0.0781 (0.0134)	0.2625 (0.0029)	0.0634 (0.0023)
CPV	83.46%	83.82%	83.46%	83.82%
$p=4$	0.4596 (0.0202)	0.1086 (0.0153)	0.3718 (0.0051)	0.0873 (0.0039)
CPV	86.88%	87.30%	86.88%	87.30%

Table 7.6: Kernel estimation errors; Brownian bridge innovations, Identity kernel $\|\Psi\|=$ 0.8, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.3706 (0.0183)	0.1536 (0.0155)	0.3151 (0.0043)	0.1306 (0.0037)
CPV	81%	80.97%	81%	80.97%
$p=3$	0.5155 (0.0255)	0.1539 (0.0203)	0.4171 (0.0071)	0.1251 (0.0051)
CPV	83.79%	84.78%	83.79%	84.78%
$p=4$	0.8100 (0.0349)	0.2702 (0.0269)	0.6342 (0.0125)	0.2093 (0.0093)
CPV	90.63%	90.8%	90.63%	90.8%
$N=100$				
$p=2$	0.3002 (0.0117)	0.1370 (0.0105)	0.2554 (0.0036)	0.1164 (0.0031)
CPV	82.64%	82.65%	82.64%	82.65%
$p=3$	0.3503 (0.0157)	0.1038 (0.0030)	0.2874 (0.0126)	0.0840 (0.0025)
CPV	88.86%	88.57%	88.86%	88.57%
$p=4$	0.5385 (0.0232)	0.1485 (0.0058)	0.4260 (0.0180)	0.1194 (0.0047)
CPV	91.26%	91.22%	91.26%	91.22%
$N=200$				
$p=2$	0.2535 (0.0101)	0.1325 (0.0091)	0.2181 (0.0033)	0.1142 (0.0030)
CPV	82.7%	83.18%	82.7%	83.18%
$p=3$	0.2632 (0.0146)	0.0785 (0.0117)	0.2168 (0.0022)	0.0629 (0.0017)
CPV	89.05%	89%	89.05%	89%
$p=4$	0.4260 (0.0212)	0.1080 (0.0161)	0.3378 (0.0044)	0.0869 (0.0036)
CPV	91.3%	91.34%	91.3%	91.34%

Table 7.7: Kernel estimation errors; Brownian bridge innovations, Sloping kernel $\|\Psi\|=0.5$, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.3951 (0.0186)	0.1108 (0.0038)	0.3248 (0.0164)	0.0862 (0.0027)
CPV	76.26%	77.32%	76.26%	77.32%
$p=3$	0.5589 (0.0285	0.1728 (0.0095)	0.4481 (0.0228)	0.1292 (0.0062)
CPV	83.33%	84.52%	83.33%	84.52%
$p=4$	0.8557 (0.0343)	0.2536 (0.0115)	0.6700 (0.0266)	0.1928 (0.0083)
CPV	87.37%	88.57%	87.37%	88.57%
$N=100$				
$p=2$	0.3490 (0.0096)	0.0830 (0.0029)	0.2815 (0.0086)	0.0616 (0.0024)
CPV	76.45%	77.14%	76.45%	77.14%
$p=3$	0.3996 (0.0144)	0.1093 (0.0039)	0.3234 (0.0122)	0.0830 (0.0030)
CPV	83.12%	83.85%	83.12%	83.85%
$p=4$	0.5877 (0.0243)	0.1537 (0.0068)	0.4653 (0.0190)	0.1172 (0.0050)
CPV	87.04%	87.74%	87.04%	87.74%
$N=200$				
$p=2$	0.3038 (0.0078)	0.0736 (0.0019)	0.2418 (0.0073)	0.0536 (0.0015)
CPV	75.81%	76.45%	75.81%	76.45%
$p=3$	0.3354 (0.0175)	0.0968 (0.0048)	0.2721 (0.0151)	0.0698 (0.0034)
CPV	82.79%	83.35%	82.79%	83.35%
$p=4$	0.4566 (0.0204)	0.1112 (0.0047)	0.3668 (0.0161)	0.0815 (0.0030)
CPV	86.65%	87.12%	86.65%	87.12%

Table 7.8: Kernel estimation errors; Brownian bridge innovations, Sloping kernel $\|\Psi\|=$ 0.8, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.4047 (0.0142)	0.1446 (0.0036)	0.3142 (0.0132)	0.1084 (0.0026)
CPV	80.22%	80.5%	80.22%	80.5%
$p=3$	0.5357 (0.0245)	0.1526 (0.0062)	0.4241 (0.02)	0.1154 (0.0045)
CPV	86.62%	86.92%	86.62%	86.92%
$p=4$	0.8042 (0.0326)	0.2226 (0.01)	0.6284 (0.0258)	0.1698 (0.0076)
CPV	89.98%	90.45%	89.98%	90.45%
$N=100$				
$p=2$	0.3615 (0.0089)	0.1177 (0.0027)	0.2771 (0.0086)	0.0848 (0.0021)
CPV	81.04%	81.34%	81.04%	81.34%
$p=3$	0.4023 (0.0155)	0.1082 (0.0033)	0.3152 (0.0133)	0.0802 (0.0025)
CPV	86.93%	87.07%	86.93%	87.07%
$p=4$	0.5561 (0.0245)	0.1351 (0.0059)	0.4407 (0.0204)	0.1042 (0.0045)
CPV	90.13%	90.31%	90.13%	90.31%
$N=200$				
$p=2$	0.3242 (0.0053)	0.1037 (0.0021)	0.2389 (0.0060)	0.0754 (0.0016)
CPV	81.03%	81.48%	81.03%	81.48%
$p=3$	0.3340 (0.0140)	0.0890 (0.0023)	0.2564 (0.0124)	0.0635 (0.0016)
CPV	86.93%	87.14%	86.93%	87.14%
$p=4$	0.4342 (0.0170)	0.0867 (0.0027)	0.3424 (0.0137)	0.0658 (0.0020)
CPV	90.03%	90.18%	90.03%	90.18%

Table 7.9: Kernel estimation errors; Brownian motion innovations, Gaussian kernel $\|\Psi\|=0.5$, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.4332 (0.0285)	0.1357 (0.0064)	0.3618 (0.0242)	0.1093 (0.0048)
CPV	90.51%	90.88%	90.51%	90.88%
$p=3$	0.7502 (0.0456)	0.2156 (0.0122)	0.6054 (0.0376)	0.1643 (0.0081)
CPV	93.79%	94.18%	93.79%	94.18%
$p=4$	1.1924 (0.0607)	0.3276 (0.0184)	0.9272 (0.0484)	0.2476 (0.0130)
CPV	95.5%	95.86%	95.5%	95.86%
$N=100$				
$p=2$	0.3386 (0.019)	0.0883 (0.0037)	0.2850 (0.0161)	0.0715 (0.0029)
CPV	90.09%	90.39%	90.09%	90.39%
$p=3$	0.5288 (0.0314)	0.1263 (0.0058)	0.4310 (0.0253)	0.0962 (0.0036)
CPV	93.47%	93.74%	93.47%	93.74%
$p=4$	0.9364 (0.0584)	0.1850 (0.0111)	0.4741 (0.0453)	0.3200 (0.0076)
CPV	95.24%	95.45%	95.24%	95.45%
$N=200$				
$p=2$	0.2217 (0.0103)	0.0779 (0.0023)	0.1845 (0.0089)	0.0614 (0.0018)
CPV	90.49%	90.80%	90.49%	90.80%
$p=3$	0.3814 (0.0235)	0.0973 (0.0047)	0.3119 (0.0195)	0.0756 (0.0037)
CPV	93.69%	93.91%	93.69%	93.91%
$p=4$	0.6339 (0.0374)	0.1250 (0.0049)	0.5034 (0.0298)	0.0969 (0.0038)
CPV	95.36%	95.50%	95.36%	95.50%

Table 7.10: Kernel estimation errors; Brownian motion innovations, Gaussian kernel $\|\Psi\|= 0.8$, Naive bootstrap

	MSE		AD	
	$\hat{\psi} - \psi$	$\hat{\psi}^* - \hat{\psi}$	$\hat{\psi} - \psi$	$\hat{\psi}^* - \hat{\psi}$
$N = 50$				
$p = 2$	0.3477 (0.0264)	0.1336 (0.0048)	0.2826 (0.0217)	0.1058 (0.0035)
CPV	92.98%	92.61%	92.98%	92.61%
$p = 3$	0.6528 (0.0375)	0.1872 (0.008)	0.5175 (0.0291)	0.1436 (0.0061)
CPV	95.52%	95.38%	95.52%	95.38%
$p = 4$	1.1332 (0.0599)	0.2960 (0.0166)	0.8755 (0.0465)	0.2242 (0.0118)
CPV	96.8%	96.78%	96.8%	96.78%
$N = 100$				
$p = 2$	0.2509 (0.0180)	0.0851 (0.0031)	0.2035 (0.0149)	0.0666 (0.0023)
CPV	93.17%	92.95%	93.17%	92.95%
$p = 3$	0.4628 (0.0238)	0.1054 (0.0046)	0.3699 (0.02)	0.0811 (0.0036)
CPV	95.61%	95.49%	95.61%	95.49%
$p = 4$	0.8510 (0.0532)	0.1464 (0.0063)	0.6695 (0.0419)	0.1116 (0.0046)
CPV	96.87%	96.8%	96.87%	96.8%
$N = 200$				
$p = 2$	0.1487 (0.0078)	0.0654 (0.0019)	0.1205 (0.0062)	0.0513 (0.0016)
CPV	93.88%	93.87%	93.88%	93.87%
$p = 3$	0.3414 (0.0245)	0.0670 (0.0033)	0.2716 (0.0209)	0.0526 (0.0027)
CPV	96.07%	96.06%	96.07%	96.06%
$p = 4$	0.6090 (0.0371)	0.0954 (0.0047)	0.4814 (0.0301)	0.0754 (0.0036)
CPV	97.15%	97.15%	97.15%	97.15%

Table 7.11: Kernel estimation errors; Brownian motion innovations, Identity kernel $\|\Psi\|=0.5$, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.4503 (0.0307)	0.1312 (0.0067)	0.3768 (0.0259)	0.1062 (0.0052)
CPV	90.73%	91.03%	90.73%	91.03%
$p=3$	0.7647 (0.0463)	0.2167 (0.0126)	0.6184 (0.0384)	0.1633 (0.0083)
CPV	93.97%	94.31%	93.97%	94.31%
$p=4$	1.2107 (0.0615)	0.3319 (0.019)	0.9405 (0.0497)	0.2503 (0.0134)
CPV	95.61%	95.96%	95.61%	95.96%
$N=100$				
$p=2$	0.3489 (0.0203)	0.0806 (0.0035)	0.2959 (0.0173)	0.0659 (0.0029)
CPV	90.34%	90.59%	90.34%	90.59%
$p=3$	0.5396 (0.0335)	0.1263 (0.0062)	0.4413 (0.0271)	0.0963 (0.0040)
CPV	93.66%	93.90%	93.66%	93.90%
$p=4$	0.9524 (0.0579)	0.1927 (0.0126)	0.7502 (0.0447)	0.1440 (0.0085)
CPV	87.28%	87.90%	87.28%	87.90%
$N=200$				
$p=2$	0.2304 (0.0115)	0.0698 (0.0024)	0.1939 (0.0101)	0.0561 (0.0019)
CPV	90.73%	91.02%	90.73%	91.02%
$p=3$	0.3885 (0.0235)	0.0950 (0.0048)	0.3180 (0.0194)	0.0734 (0.0037)
CPV	93.89%	94.09%	93.89%	94.09%
$p=4$	0.6398 (0.0366)	0.1280 (0.0051)	0.5077 (0.0291)	0.0986 (0.0038)
CPV	95.47%	95.62%	95.47%	95.62%

Table 7.12: Kernel estimation errors; Brownian motion innovations, Identity kernel $\|\Psi\|=$ 0.8, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.3795 (0.0297)	0.1364 (0.0055)	0.3093 (0.0243)	0.1089 (0.0038)
CPV	93.27%	92.89%	93.27%	92.89%
$p=3$	0.6711 (0.0387)	0.1873 (0.0077)	0.5337 (0.0298)	0.1429 (0.0055)
CPV	95.79%	95.60%	95.79%	95.60%
$p=4$	1.1523 (0.0619)	0.3018 (0.0182)	0.8882 (0.0480)	0.2289 (0.013)
CPV	96.98%	96.93%	96.98%	96.93%
$N=100$				
$p=2$	0.2795 (0.0212)	0.0868 (0.0033)	0.2291 (0.0178)	0.0688 (0.0025)
CPV	93.52%	93.33%	93.52%	93.33%
$p=3$	0.4758 (0.0250)	0.1040 (0.0041)	0.3818 (0.0209)	0.0803 (0.0031)
CPV	95.90%	95.78%	95.90%	95.78%
$p=4$	0.8701 (0.0527)	0.1527 (0.0072)	0.6818 (0.0414)	0.1160 (0.0053)
CPV	97.06%	96.98%	97.06%	96.98%
$N=200$				
$p=2$	0.1641 (0.0092)	0.0689 (0.0022)	0.1337 (0.0074)	0.0545 (0.0017)
CPV	94.20%	94.24%	94.20%	94.24%
$p=3$	0.3503 (0.0241)	0.0677 (0.0034)	0.2797 (0.0205)	0.0529 (0.0027)
CPV	96.35%	96.34%	96.35%	96.34%
$p=4$	0.6145 (0.0366)	0.0989 (0.0053)	0.4849 (0.0297)	0.0779 (0.0040)
CPV	97.34%	97.34%	97.34%	97.34%

Table 7.13: Kernel estimation errors; Brownian motion innovations, Sloping kernel $\|\Psi\|=0.5$, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.4342 (0.0238)	0.0942 (0.0053)	0.3445 (0.0204)	0.0719 (0.0040)
CPV	91.67%	91.73%	91.67%	91.73%
$p=3$	0.8201 (0.0554)	0.1981 (0.0124)	0.6413 (0.0471)	0.1506 (0.0089)
CPV	94.46%	94.65%	94.46%	94.65%
$p=4$	1.3519 (0.0742)	0.3598 (0.0289)	1.0401 (0.0594)	0.2715 (0.0211)
CPV	95.87%	96.15%	95.87%	96.15%
$N=100$				
$p=2$	0.3475 (0.0220)	0.0529 (0.0028)	0.2751 (0.0191)	0.0408 (0.0023)
CPV	91.23%	91.24%	91.23%	91.24%
$p=3$	0.6202 (0.0502)	0.0987 (0.0049)	0.4925 (0.0411)	0.0757 (0.0040)
CPV	83.12%	83.85%	83.12%	83.85%
$p=4$	0.9982 (0.0597)	0.1906 (0.0133)	0.7747 (0.0469)	0.1449 (0.0101)
CPV	95.59%	95.72%	95.59%	95.72%
$N=200$				
$p=2$	0.2792 (0.0162)	0.0292 (0.0017)	0.2172 (0.0148)	0.0225 (0.0014)
CPV	91.66%	91.68%	91.66%	91.68%
$p=3$	0.4478 (0.0274)	0.0624 (0.0041)	0.3548 (0.0219)	0.0491 (0.0034)
CPV	94.36%	94.41%	94.36%	94.41%
$p=4$	0.6975 (0.0363)	0.1009 (0.0053)	0.5438 (0.0276)	0.0782 (0.0043)
CPV	95.72%	95.79%	95.72%	95.79%

Table 7.14: Kernel estimation errors; Brownian motion innovations, Sloping kernel $\|\Psi\|=$ 0.8, Naive bootstrap

	MSE		AD	
	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$	$\hat{\psi}-\psi$	$\hat{\psi}^*-\hat{\psi}$
$N=50$				
$p=2$	0.4639 (0.0220)	0.1025 (0.0049)	0.3573 (0.0197)	0.0783 (0.0038)
CPV	93.95%	93.61%	93.95%	93.61%
$p=3$	0.8305 (0.0534)	0.2001 (0.0110)	0.6494 (0.0456)	0.1520 (0.0080)
CPV	95.99%	95.87%	95.99%	95.87%
$p=4$	1.3655 (0.0773)	0.3666 (0.0352)	1.0532 (0.0615)	0.2781 (0.026)
CPV	97.03%	97.04%	97.03%	97.04%
$N=100$				
$p=2$	0.3932 (0.0184)	0.0571 (0.0027)	0.2948 (0.0172)	0.0435 (0.0021)
CPV	93.85%	93.64%	93.85%	93.64%
$p=3$	0.6383 (0.0487)	0.1022 (0.0051)	0.5029 (0.0404)	0.0782 (0.0042)
CPV	95.89%	95.82%	95.89%	95.82%
$p=4$	0.9910 (0.0586)	0.1900 (0.0129)	0.7726 (0.0468)	0.1443 (0.0098)
CPV	96.92%	96.9%	96.92%	96.9%
$N=200$				
$p=2$	0.3481 (0.0139)	0.0309 (0.0016)	0.2560 (0.0138)	0.0238 (0.0013)
CPV	94.41%	94.35%	94.41%	94.35%
$p=3$	0.4834 (0.0253)	0.0624 (0.0040)	0.3773 (0.0206)	0.0488 (0.0033)
CPV	96.22%	96.22%	96.22%	96.22%
$p=4$	0.7087 (0.0352)	0.1011 (0.0053)	0.5524 (0.0271)	0.0780 (0.0043)
CPV	97.14%	97.15%	97.14%	97.15%

Bibliography

[1] Antoniadis, A. and Sapatinas, T., (2003) *Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes*, *Journal of Multivariate Analysis*, 87:133-158.

[2] Aston, A. D. J. and Kirch C. (2012) *Evaluating Stationarity via Change-Point Alternatives with Application to FMRI Data*, *The Annals of Applied Statistics* 6:4:1906-1948.

[3] Aue, A. and Horváth, L. (2013) *Structural breaks in time series*, *Journal of Time Series Analysis* 34:1-16.

[4] Berkes, I., Gabrys, R., Horváth, L. and Kokoszka P. (2009) *Detecting changes in the mean of functional observations*, *Journal of the Royal Statistical Society*, 71:5:927-946.

[5] Besse, P., Cardot, H. and Stephenson, D., (2000) *Autoregressive forecasting of some functional climatic variations*, *Scandinavian Journal of Statistics*, 24:467-487.

[6] Bickel, P.J., Freedman, D.A., (1981) *Some asymptotic theory for the bootstrap*, *Annals Statistics*, 9:1196-1217.

[7] Bosq D. (2000) *Linear processes in function spaces*, Springer, New York.

[8] Brockwell, P. J. and Davis, R. A., (1991) *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag, Berlin-Heidelberg-New York.

[9] Bühlmann, P., (2002) *Bootstraps for Time Series*,*Statistical Science* Vol.17, No.1,52-72.

[10] Cardot, H., Mas, A. and Sarda, P., (2007) *CLT in functional linear regression models*, *Probability Theory and Related Fields*, 138:325-361.

[11] Carlstein, E. (1986) *The use of subseries values for estimating the variance of a general statistics from a stationary time series*, *The Annals of Statistics*, 14:1171-1179.

[12] Fernández de Castro, B., Guillas, S. and González Manteiga, W., (2005) *Functional samples and bootstrap for predicting sulfur dioxide levels*, *Technometrics*, 47:2.

[13] Cuevas, A.,Febrero, M. and Fraiman R., (2006) *On the use of the bootstrap for estimating functions with functional data*, *Computational Statistics and Data Analysis*, 51:1063-1074.

[14] Damon, J. and Guillas, S., (2002) *The inclusion of exogeneous variables in functional autoregressive ozone forecasting*, *Environmetrics*, 13:759-774.

[15] Dehling, H., Sharipov, S.O. and Wendler, M., (2015) *Bootstrap for dependent Hilbert space-valued random variables with application to von Mises statistics*, *Journal of Multivariate Analysis*, 133:200-215.

[16] Didericksen, D., Kokoszka P., and Zhang X., (2012) *Emprical properties of forecasts with the functional autoregressive model,Computational statistics* 27:285-298.

[17] Efron, B. (1979) *Bootstrap methods: another look at the jacknife*, *The Annals of Statistics*, 7:1-26.

[18] Ferraty, F., Van Keilegom, I. and Vieu P. (2010) *On the validity of the bootstrap in non-parametric functional regression*, *Scandinavian Journal of Statistics*, Vol. 37:286-306.

[19] Freedman, D. A., (1981) *Bootstrapping regression models*, *The Annals of Statistics*, Vol.9, No. 6, 1218-1228.

[20] Kreiss, J. P. and Franke, J., (1992) *Bootstrapping stationary autoregressive moving average models*, *Journal of Time Series Analysis*, 13:297-317.

[21] Gabrys, R. and Kokoszka P., (2007) *Portmanteau test of independence for functional observations*, *American Statistical Association*, Vol. 102, No. 480, Theory and Methods.

[22] Guillas, S., (2001) *Rates of convergence of auto correlation estimates for autoregressive Hilbertian processes*, *Statistics and Probability Letters*, 55:281-291.

[23] Gohberg, I., Golberg. S. and Kaashoek, M. A. (1990) *Classes of Linear Operators, Vol. 1. Operator Theory: Advances and Applications*, Birkhäuser, Basel,Switzerland.

[24] Härdle, W., Horowitz, J. and Kreiss, J.-P., (2003) *Bootstrap methods for time series*, *International Statistical Review*, 71,2:435-459.

[25] Hörmann, S. and Kokoszka, P., (2010) *Weakly dependent functional data*, *The Annals of Statistics*, 38,3:1845-1884.

[26] Horváth, L., Hušková, M. amd Kokoszka, P. (2010) Testing the stability of the functional autoregressive process,*Journal of Multivariate Analysis*, 101:352-357.

[27] Kiefer, J. (1959) *K-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests*, *The Annals of Mathematical Statistics* 30:420-447.

[28] Horváth, L. and Kokoszka P. (2010) *Inference for Functional Data with Applications*, Springer, New York.

[29] Ibragimov, I. A., and Linnik Ju. V. (1971) *Independent and Stationary Sequences of Random Variables*, Springer, New York.

[30] Jiofack, J. G. A. and Nkiet, G. M., (2010) *Testing for lack of effect between functional variables*, *Statistics and Probability Letters*, 80:1210-1217.

[31] Kargin, V. and Onatski, A., (2008) *Curve forecasting by functional autoregression*, *Journal of Multivariate Analysis*, 99:2508-2526.

[32] Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L., (2008) *Testing for lack of dependence in the functional linear model*, *The Canadian Journal of Statistics*, 36:2:1-16.

[33] Kreiss, J.-P. and Paparoditis, E., (2011) *Bootstrap methods for dependent data: A review*, *Journal of the Korean Statistical Society*, 40:357-378.

[34] Kreiss, J.-P. and Lahiri, S. N., (2012) *Bootstrap methods for time series*, *Time Series Analysis: Methods and Applications*, 30, ISSN:0169-7161.

[35] Künsch, H. R., (1989) *The Jacknife And The Bootstrap For General Stationary Observations*, *The Annals of Statistics*, Vol.17, No. 3, 1217-1241.

[36] Lahiri, N. S. (2003) *Resampling Methods for Dependent Data*, Springer, New York.

[37] Laukaitis, A. and Račkauskas, A. (2002) Functional data analysis of payment systems, *Nonlinear Analysis:Modelling and Control*, 7:53-68.

[38] Laukaitis, A., (2008) *Functional data analysis for cash flow and transactions intensity continuous-time prediction using Hilbert-valued autoregressive processes*, *European Journal of Operational Research*, 185:1607-1614.

[39] McLeish, D. L., (1974) *Dependent Central Limit Theorems and Invariance Principles*, *The Annals of Probability*, Vol.2, No. 4., pp.620-628.

[40] Mammen, E. (1992) *When Does Bootstrap Work? Asymptotic Results and Simulations*, New York: Springer-Verlag.

[41] Mas, A., (2004) *Consistence du prédicteur dans le modéle ARH(1)*, *Annales de l'Isup*, 48, No.3,39-48.

[42] Mas, A. and Pumo, B., (2009) *Linear processes for functional data*, arXiv:0901.2503v1[math.ST].

[43] Mas, A. and Menneteau, L., (2003) *Pertubation approach applied to the asymptotic study of random operators*, *Progress in Probability*, 55:127-134.

[44] Mas, A., (2007) *Weak convergence in the functional autoregressive model*, *Journal of Multivariate Analysis*, 98:1231-1261.

[45] Menneteau, L., (2005) *Some laws of the iterated logarithm in Hilbertian autoregressive models*, *Journal of Multivariate Analysis*, 92,2:405-425.

[46] Paparoditis, E. and Sapatinas, T., (2015) *Bootstrap based K-sample testing for functional data*, arXiv:1409.4317v2

[47] Politis, D. and Romano, J. (1992) *Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Application to the Stationary Bootstrap*, Technical Report 407, Stanford University.

[48] Priestley, M. B., (2004) *Spectral Analysis and Time Series*, London: Academic Press.

[49] Ramsey, J.,Hooker, G. and Graves, S. (2009) *Functional Data Analysis with R and MATLAB*, Springer.

[50] Cardot, H., Ferraty, F., Mas, A. and Sarda, P., (2003) *Testing hypotheses in the functional linear model*, *Scandinavian Journal of Statistics*, 30:241-255.

[51] Shibata, R. (1976) *Selection of order of an autoregressive model by Akaike's information criterion*, *Biometrika* 63:117-126.

[52] Ramsey, J. O., Silverman B. W. (2005) *Functional Data Analysis*, Springer.

[53] Singh, K. (1981) *On the Asymptotic Accuracy of Efron's Bootstrap*, *The Annals of Statistics*, 9:1187-1195, No.6 (Nov., 1981).

[54] Efron, B. and Tibshirani, R. J. (1994) *An Introduction to the Bootstrap*, Chapmann and Hall.

[55] Van der Vaart, A. W. (2000) *Asymptotic Statistics*, Cambridge University Press, Cambridge.

[56] Van der Vaart, A. W. and Wellner J. A. (1996) *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer, New York.

[57] Varadarajan, V., (1958) *Weak convergence of measures on separable metric spaces*, *Sankhy: The Indian Journal of Statistics (1933-1960)*, 19:15-22, No. 1/2 (Feb., 1958).

[58] Yao, F., Müller, H. and Wang, J., (2005) *Functional data analysis for sparse longitudinal data*, *Journal of the American Statistical Association*, 13:759-774.

[59] Fan, J. and Yao, Q., (2005) *Nonlinear Time Series, Nonparametric and Parametric Methods*, Springer, New York.

Curriculum Vitae

Euna Gesare Nyarige

2006-2009 Bachelor of Science in Mathematics and Computer Science

2010-2011 Master of Science in Applied Statistics

2012-Date PhD in Mathematics

Wissenschaftlicher Werdegang

Euna Gesare Nyarige

2006-2009 Bachelor of Science in Mathematik und Informatik

2010-2011 Master of Science in Angewandte Statistik

2012-Date PhD in Mathematik