



HAL
open science

Locally stationary processes prediction by auto-regression

François Roueff, Andres Sanchez-Perez

► **To cite this version:**

François Roueff, Andres Sanchez-Perez. Locally stationary processes prediction by auto-regression. 2016. hal-01269137v1

HAL Id: hal-01269137

<https://hal.science/hal-01269137v1>

Preprint submitted on 5 Feb 2016 (v1), last revised 12 Jan 2018 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Locally stationary processes prediction by auto-regression

François Roueff*¹ and Andres Sanchez-Perez^{†1}

¹LTCI, CNRS, Université Paris-Saclay

February 5, 2016

Abstract

In this contribution we introduce locally stationary time series through the local approximation of the non-stationary covariance structure by a stationary one. This allows us to define autoregression coefficients in a non-stationary context, which, in the particular case of a locally stationary Time Varying Autoregressive (TVAR) process, coincide with the generating coefficients. We provide and study an estimator of the time varying autoregression coefficients in a general setting. The proposed estimator of these coefficients enjoys an optimal minimax convergence rate under limited smoothness conditions. In a second step, using a bias reduction technique, we derive a minimax-rate estimator for arbitrarily smooth time-evolving coefficients, which outperforms the previous one for large data sets. For TVAR, the predictor naturally obtained from the estimator also exhibits an optimal minimax convergence rate.

1 Introduction

In many applications, one is interested in predicting the next values of an observed time series. It is the case in various areas like finance (stock market, volatility on prices), social sciences (population studies), epidemiology, meteorology and network systems (Internet traffic). Autoregressive processes have been used successfully in a stationary context for several decades. On the other hand, in a context where the number of observations can be very large, the usual stationarity assumption has to be weakened to take into account some smooth evolution of the environment.

Many prediction methods developed in signal processing are well known to adapt to a changing environment. This is the case of the wide spread recursive least square algorithms. The initial goal of these methods is to provide an online algorithm for estimating a regression vector with low numerical cost. Such methods usually rely

*francois.roueff@telecom-paristech.fr

†andres.sanchez-perez@telecom-paristech.fr

on a forgetting factor or a gradient step size γ and they can be shown to be consistent in a stationary environment when γ decreases adequately to zero (see e.g. [14]). However when the environment is changing, that is, when the regression parameter evolves along the time, a “small enough” γ often yields a good tracking of the evolving regression parameter. In order to have a sound and comprehensive understanding of this phenomenon, an interesting approach is to consider a local stationarity assumption, as successfully initiated in [11] by relying on a non-stationary spectral representation introduced in [23]; see also [12] and the references therein for a recent overview. The basic idea is to provide an asymptotic analysis for the statistical inference of non-stationary time series such as time varying autoregressive (TVAR) processes by relying on local stationary approximations. The analysis of the NLMS algorithm for tracking a moving autoregression parameter in this framework is tackled in [22]. Such an analysis is based on the usual tools of non-parametric statistics. The TVAR parameter θ is seen as the regular samples of a smooth \mathbb{R}^d -valued function. An in-fill asymptotic allows one to derive minimax rates of convergence for estimating this function on a fixed interval $[0, 1]$ within particular smoothness classes of functions. As shown in [22], it turns out that the NLMS algorithm provides an optimal minimax rate for estimating the TVAR parameters with Hölder smoothness index $\beta \in (0, 1]$ but is no longer optimal for $\beta > 1$, that is when the TVAR parameters are smoother than a continuously differentiable function. An improvement of the NLMS is proposed in [22] to cope with the case $\beta \in (0, 2]$ but, to the best of our knowledge, there is no available method neither for the θ minimax-rate estimation nor for the minimax-rate prediction when $\beta > 2$, that is when the TVAR parameters are smoother than a two-times continuously differentiable function.

In the present work, our main contribution is twofold. First we introduce the concept of time-varying linear prediction coefficients to a general class of locally stationary processes. This general class extends the class of locally stationary processes as introduced in [11] in a way that we believe is more natural and appropriate to the signal processing community. In the specific case of a TVAR process, these coefficients correspond to the time-varying autoregression parameters. Second, we show that the Yule-Walker estimator introduced in [13] for TVAR processes also applies to this general class and is minimax-rate for Hölder index $\beta = 2$. Moreover, by applying a bias reduction technique, we derive a new estimator which is minimax-rate for any Hölder index $\beta \geq 2$.

The paper is organized as follows. In Section 2, we introduce the locally stationary time series and define the regression problem investigated in this work. The Yule-Walker estimator is detailed in Section 4. In Section 3, we explain why and how minimax estimation is crucial for deriving practical predictors. Main results are presented in Section 5 relying on Hölder smoothness assumptions on the local spectral density of the locally stationary time series. The particular case of TVAR processes is treated in Section 6. Numerical experiments complete our study in Section 7, confirming the benefits of our approach when the length of the data set becomes very large.

Four appendices complete this paper. Appendix A contains useful results on locally stationary time series needed for showing the main theorems of Section 5. The proof of the main theorems of Section 5 are provided in Appendix B. Some useful technical results can be found in Appendix C. As a support of Appendix A, we refer to the basic

tool-kit on weakly stationary processes presented in Appendix D.

2 General setting

In the following, vectors are denoted using boldface symbols, $\|\mathbf{x}\|$ denotes the Euclidean norm of \mathbf{x} , $\|\mathbf{x}\| = (\sum_i |x_i|^2)^{1/2}$, and $\|\mathbf{x}\|_1$ its ℓ_1 norm, $\|\mathbf{x}\|_1 = \sum_i |x_i|$. If f is a function, $\|f\|_\infty = \sup_x |f(x)|$ corresponds to its sup norm.

2.1 Main definitions

We consider a doubly indexed time series $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$, which we assume to be centred for convenience. Here t refers to a discrete time index and T is an additional index indicating the sharpness of the *local approximation* of the time series $(X_{t,T})_{t \in \mathbb{Z}}$ by a stationary one. Coarsely speaking, $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ is considered to be *locally stationary* if, for T large, given a set S_T of sample indices such that $t/T \approx u$ over $t \in S_T$, the sample $(X_{t,T})_{t \in S_T}$ can be approximately viewed as the sample of a stationary time series which depends on the *rescaled location* u . Note that u is a continuous time parameter, sometimes referred to as the *rescaled time index*. Following [11], it is classical to set T as the number of available observations, in which case all the definitions are restricted to $1 \leq t \leq T$ and $u \in [0, 1]$. However this is not essential in the mathematical derivations and it is more convenient to set $t \in \mathbb{Z}$ and $u \in \mathbb{R}$ for presenting our setting.

We first introduce definitions for the time varying covariance and the local covariance functions.

Definition 1 (Time varying covariance function). *Let $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ be an array of random variables with finite variances. The local time varying covariance function γ^* is defined for all $t \in \mathbb{Z}, T \in \mathbb{N}^*$ and $\ell \in \mathbb{Z}$ as*

$$\gamma^*(t, T, \ell) = \text{cov}(X_{t,T}, X_{t-\ell,T}) . \quad (2.1)$$

Definition 2 (Local covariance function). *A local spectral density f is a $\mathbb{R}^2 \rightarrow \mathbb{R}_+$ function, (2π) -periodic and locally integrable with respect to the second variable. The local covariance function γ associated with the time varying spectral density f is defined on $\mathbb{R} \times \mathbb{Z}$ by*

$$\gamma(u, \ell) = \int_{-\pi}^{\pi} \exp(i\ell\lambda) f(u, \lambda) d\lambda . \quad (2.2)$$

In (2.2), the variable u should be seen as *rescaled time index* (in \mathbb{R}), ℓ as a (non-rescaled) time index and λ as a frequency (in $[-\pi, \pi]$). Recall that, by the Herglotz theorem (see [4, Theorem 4.3.1]), Equation (2.2) guaranties that for any $u \in \mathbb{R}$, $(\gamma(u, \ell))_{\ell \in \mathbb{Z}}$ is indeed the autocovariance function of a stationary time series. Now, we can state the definition of locally stationary processes that we use here.

Definition 3 (Locally stationary processes). *Let $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ be an array of random variables with finite variances. We say that $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ is locally stationary with local*

spectral density f if the time varying covariance function γ^* of $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ and the local covariance function γ associated with f satisfy

$$\left| \gamma^*(t, T, \ell) - \gamma\left(\frac{t}{T}, \ell\right) \right| \leq \frac{C}{T}, \quad (2.3)$$

where C is a constant.

Let us give some examples fulfilling this definition.

Example 1. Locally stationary processes were first introduced by [11] using the spectral representation

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(it\omega) A_{t,T}^0(\omega) \xi(d\omega), \quad (2.4)$$

where $\xi(d\omega)$ is the spectral representation of a white noise and $(A_{t,T}^0)_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ is a collection of transfer functions such that there exist a constant K and a (unique) 2π -periodic function $A : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{C}$ with $A(u, -\omega) = \overline{A(u, \omega)}$ such that for all $T \geq 1$,

$$\sup_{t, \omega} \left| A_{t,T}^0(\omega) - A\left(\frac{t}{T}, \omega\right) \right| \leq \frac{K}{T}. \quad (2.5)$$

This class of locally stationary processes satisfies Definition 3 (see [10, Section 1]) with $f(u, \lambda) = |A(u, \lambda)|^2$.

Example 2 (TVAR(p) model). Under suitable assumptions, the TVAR process is a particular case of Example 1 (see [11, Theorem 2.3]). It is defined by the recursive equation

$$X_{t,T} = \sum_{j=1}^p \theta_j\left(\frac{t}{T}\right) X_{t-j,T} + \sigma\left(\frac{t}{T}\right) \xi_t,$$

where $\boldsymbol{\theta} = [\theta_1 \dots \theta_p]' : \mathbb{R} \rightarrow \mathbb{R}^p$ are the time varying autoregressive coefficients and $(\xi_t)_{t \in \mathbb{Z}}$ are i.i.d. centred and with variance 1.

Example 3 (Non-stationary Causal Bernoulli Shift). Let $p > 0$ and $\varphi : \mathbb{R}^{p+2} \rightarrow \mathbb{R}$. Consider

$$X_{t,T} = \varphi\left(\frac{t}{T}, \xi_t, \dots, \xi_{t-p}\right),$$

where $(\xi_t)_{t \in \mathbb{Z}}$ are i.i.d. such that $\mathbb{E}[|\xi_0|^q] < \infty$ for all $q \geq 1$, $\mathbb{E}[\varphi(u, \xi_0, \dots, \xi_p)] = 0$ for all $u \in \mathbb{R}$ and there exist $K, C, r > 0$ such that, for all $u, u' \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^{p+1}$

$$\begin{aligned} |\varphi(u, \mathbf{x})| &\leq C \left(1 + \sum_{i=0}^p |x_i|^r \right), \\ |\varphi(u, \mathbf{x}) - \varphi(u', \mathbf{x})| &\leq K |u - u'| \left(1 + \sum_{i=0}^p |x_i|^r \right). \end{aligned}$$

In contrast to Examples 1 and 2, Example 3 do not rely on a linear representation of the process.

2.2 Statement of the problem

Let $d \in \mathbb{N}^*$. For each $t = 1, \dots, T$, define the prediction vector of order d by

$$\boldsymbol{\theta}_{t,T}^* = \arg \min_{\boldsymbol{\theta} = [\theta_1 \dots \theta_d]' \in \mathbb{R}^d} \mathbb{E} \left[\left(X_{t,T} - \sum_{k=1}^d \theta_k X_{t-k,T} \right)^2 \right] = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^d} \mathbb{E} \left[(X_{t,T} - \boldsymbol{\theta}' \mathbf{X}_{t-1,T})^2 \right], \quad (2.6)$$

where A' denotes the transpose of matrix A and $\mathbf{X}_{s,T} = [X_{s,T} \dots X_{s-(d-1),T}]'$. Provided that $\Gamma_{t,T}^*$ is invertible, the solution is given by

$$\boldsymbol{\theta}_{t,T}^* = \left(\Gamma_{t,T}^* \right)^{-1} \boldsymbol{\gamma}_{t,T}^*, \quad (2.7)$$

where $\boldsymbol{\gamma}_{t,T}^* = [\gamma^*(t, T, 1) \dots \gamma^*(t, T, 1)]'$, $\Gamma_{t,T}^*$ is the time varying covariances matrix $\Gamma_{t,T}^* = (\gamma^*(t-i, T, j-i); i, j = 1, \dots, d)$ and γ^* is the time varying covariance function as defined in (2.1). Analogously to (2.7), and with the aim of approximating the local solution of the stationary Yule-Walker equations, we set

$$\boldsymbol{\theta}_u = \Gamma_u^{-1} \boldsymbol{\gamma}_u, \quad (2.8)$$

where $\boldsymbol{\gamma}_u = [\gamma(u, 1) \dots \gamma(u, d)]'$, Γ_u is the covariances matrix $\Gamma_u = (\gamma(u, i-j); i, j = 1, \dots, d)$ and γ is the local covariance function as defined in (2.2).

Assuming particular regularity conditions on $\boldsymbol{\theta}$, an estimator $\widehat{\boldsymbol{\theta}}$ of it is studied in [13] for the model of Example 1. In the following we improve these results by deriving minimax rate properties of the estimator of [13] and extensions of it. We will use the following smoothness class of functions. For $\alpha \in (0, 1]$ the α -Hölder semi-norm of a function $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{C}^d$ is defined by

$$|\mathbf{f}|_{\alpha,0} = \sup_{0 < |s-s'| < 1} \frac{\|\mathbf{f}(s) - \mathbf{f}(s')\|}{|s - s'|^\alpha}.$$

This semi-norm is used to build a norm for any $\beta > 0$ as it follows. Let $k \in \mathbb{N}$ and $\alpha \in (0, 1]$ be such that $\beta = k + \alpha$. If \mathbf{f} is k times differentiable on \mathbb{R} , we define

$$|\mathbf{f}|_\beta = |\mathbf{f}^{(k)}|_{\alpha,0} + \max_{0 \leq s \leq k} \|\mathbf{f}^{(s)}\|_\infty,$$

and $|\mathbf{f}|_\beta = \infty$ otherwise. For $R > 0$ and $\beta > 0$, the (β, R) -Hölder ball of dimension d is denoted by

$$\Lambda_d(\beta, R) = \left\{ \mathbf{f} : \mathbb{R} \rightarrow \mathbb{C}^d, \text{ such that } |\mathbf{f}|_\beta \leq R \right\}.$$

We can now derive the main assumption used on the model which depends on some positive constants β, R and f_- .

- (M-1) The sequence $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ is a locally stationary process in the sense of Definition 3 such that $\mathbb{P}(X_{t,T} = 0) = 0$ for any t . The spectral density $f(\cdot, \lambda)$ belongs to $\Lambda_1(\beta, R)$ for any $\lambda \in \mathbb{R}$, and satisfies $f(u, \lambda) \geq f_-$ for all $u, \lambda \in \mathbb{R}$. The constant C in (2.3) depends continuously and at most on $\|f\|_\infty$ and $\sup_{u, \lambda \in \mathbb{R}} |\partial f(u, \lambda) / \partial u|$.

Note in particular that for $\beta > 1$, **(M-1)** implies that f is continuously differentiable in its first component.

The problem that we are interested in is to derive a minimax rate estimator $\tilde{\theta}$ for any $\beta \geq 2$, which means, that for such a β , the estimation risk, say the quadratic risk $\mathbb{E}[\|\tilde{\theta}_{i,T} - \theta_{i,T}^*\|^2]$ can be bounded uniformly over all processes satisfying **(M-1)** (among with some additional assumptions), and that the corresponding rate of convergence as $T \rightarrow \infty$ cannot be improved by any other estimator. The case $\beta \leq 2$ is solved in [22] for a particular subclass.

3 Minimax estimation for adaptive prediction

Let $\widehat{X}_{d,t,T}^*$ denote the best linear predictor of order d of $X_{i,T}$, which as a consequence of (2.6), reads

$$\widehat{X}_{d,t,T}^* = (\theta_{i,T}^*)' \mathbf{X}_{t-1,T},$$

We denote by $\widehat{X}_{i,T}^*$ the best predictor of $X_{i,T}$ given its past, which corresponds to the conditional expectation

$$\widehat{X}_{i,T}^* = \mathbb{E}[X_{i,T} | X_{s,T}, s \leq t-1]. \quad (3.1)$$

As explained before, the goal of this paper is to derive estimators, say $\tilde{\theta}_{i,T} \in \mathbb{R}^d$, of $\theta_{i,T}$, which is a local approximation of $\theta_{i,T}^*$. In this section, we assume that $\tilde{\theta}_{i,T}$ is a function of the past $X_{s,T}, s \leq t-1$. Then $\tilde{\theta}_{i,T}' \mathbf{X}_{t-1,T}$ is a legitimate predictor of $X_{i,T}$ and we have the following decomposition of the corresponding prediction quadratic risk

$$\mathbb{E} \left[(X_{i,T} - \tilde{\theta}_{i,T}' \mathbf{X}_{t-1,T})^2 \right] = \mathbb{E} \left[(X_{i,T} - \widehat{X}_{i,T}^*)^2 \right] + \mathbb{E} \left[(\tilde{\theta}_{i,T}' \mathbf{X}_{t-1,T} - \widehat{X}_{i,T}^*)^2 \right].$$

The first term is the minimal prediction error that one would achieve with the conditional expectation (which requires the true distribution of the whole process). Furthermore, inserting $\widehat{X}_{d,t,T}^* = (\theta_{i,T}^*)' \mathbf{X}_{t-1,T}$ and using the Minkowski and Cauchy-Schwartz inequality, the square root of the second term can be bounded as

$$\begin{aligned} \left(\mathbb{E} \left[(\tilde{\theta}_{i,T}' \mathbf{X}_{t-1,T} - \widehat{X}_{i,T}^*)^2 \right] \right)^{1/2} &\leq \left(\mathbb{E} \left[(\widehat{X}_{d,t,T}^* - \widehat{X}_{i,T}^*)^2 \right] \right)^{1/2} \\ &\quad + \left(\mathbb{E} \left[\|\mathbf{X}_{t-1,T}\|^4 \right] \right)^{1/4} \left(\mathbb{E} \left[\|\tilde{\theta}_{i,T} - \theta_{i,T}^*\|^4 \right] \right)^{1/4}. \end{aligned}$$

The first term in the upper bound is due to the approximation of the best predictor by the best linear predictor of order d and can only be improved by increasing d . Note that, in the case of the TVAR(p) model with $p \leq d$, this error term vanishes. The quantity $(\mathbb{E}[\|\mathbf{X}_{t-1,T}\|^2])^{1/2}$ is typically bounded by a universal constant independent of (t, T) over the class of processes under consideration. Hence, for a given d , the control of the prediction risk boils down to the control of the quadratic estimation risk $\mathbb{E}[\|\tilde{\theta}_{i,T} - \theta_{i,T}^*\|^2]$.

To do so, we can further decompose the quadratic loss as

$$\left\| \widetilde{\boldsymbol{\theta}}_{i,T} - \boldsymbol{\theta}_{i,T}^* \right\| \leq \left\| \widetilde{\boldsymbol{\theta}}_{i,T} - \boldsymbol{\theta}_{i/T} \right\| + \left\| \boldsymbol{\theta}_{i/T} - \boldsymbol{\theta}_{i,T}^* \right\| ,$$

and note that the second term is a deterministic error basically accounting for the approximation precision of the non-stationary model by a stationary one, which, under appropriate assumptions, will be shown to be of order T^{-1} . As a result of these successive decompositions, our effort in the following focus on controlling the estimation risk $\mathbb{E}[\|\widetilde{\boldsymbol{\theta}}_{i,T} - \boldsymbol{\theta}_{i/T}\|^2]$ uniformly over a class of locally stationary processes with given smoothness index $\beta \geq 2$.

By achieving this goal, we will provide a theoretical justification of the intuitive fact that, in a non-stationary context, any predictor should be adapted to how smoothly the time varying parameter evolves along the time. On the other hand, in practical situations, one may not have a strong *a priori* on the smoothness index β and one should rely on data driven methods that are therefore called *adaptive*. This problem was recently tackled in [16]. More precisely, using aggregation techniques introduced in the context of individual sequences prediction (see [25, 20, 8, 1]) and statistical learning ([3, 6, 7, 18, 27, 28, 19]), one can aggregate sufficiently many predictors in order to build a minimax predictor which adapts to the unknown smoothness β of the time varying parameter. However, a crucial requirement in [16] is to dispose of β -minimax-rate sequences of predictors for any $\beta > 0$. Hence, following [16] and [22], where minimax estimators are derived only for $\beta \leq 2$, our results will pave the way for adaptive minimax-rate forecasting at any (unknown) smoothness rate.

4 Tapered Yule-Walker estimate

Following [13], a local empirical covariance function is defined as follows. It relies on a real data taper function h and a bandwidth M which may depend on T .

Definition 4 (Empirical local covariance function). *Consider a function $h : [0, 1] \rightarrow \mathbb{R}$ and $M \in 2\mathbb{N}^*$. The empirical local covariance function $\widehat{\gamma}_M$ with taper h is defined in $\mathbb{R} \times \mathbb{Z}$ as*

$$\widehat{\gamma}_M(u, \ell) = \frac{1}{H_M} \sum_{\substack{t_1, t_2=1 \\ t_1-t_2=\ell}}^M h\left(\frac{t_1}{M}\right) h\left(\frac{t_2}{M}\right) X_{\lfloor uT \rfloor + t_1 - M/2, T} X_{\lfloor uT \rfloor + t_2 - M/2, T} ,$$

where $H_M = \sum_{k=1}^M h^2(k/M) \sim M \int_0^1 h^2(x) dx$ is the normalizing factor. We assume that $H_M > 0$.

For $h \equiv 1$ in Definition 4 we obtain the classical covariance estimate for a centred sample $\{X_s, \lfloor uT \rfloor - M/2 \leq s \leq \lfloor uT \rfloor + \ell + M/2\}$. Taking into account the interval $[t - M/2 + 1, t + M/2]$, and with the help of the data taper function h , the following empirical Yule-Walker equations are then derived

$$\widehat{\boldsymbol{\theta}}_{i,T}(M) = \widehat{\Gamma}_{i,T,M}^{-1} \widehat{\boldsymbol{\gamma}}_{i,T,M} , \quad (4.1)$$

where $\widehat{\boldsymbol{\gamma}}_{t,T,M} = [\widehat{\gamma}_M(t/T, 1) \dots \widehat{\gamma}_M(t/T, d)]'$, $\widehat{\Gamma}_{t,T,M}$ is the matrix of empirical covariances $\widehat{\Gamma}_{t,T,M} = (\widehat{\gamma}_M(t/T, i - j); i, j = 1, \dots, k)$ and $\widehat{\gamma}_M$ is the empirical covariance function as in Definition 4.

5 Main results in the general framework

5.1 Additional assumptions

For convenience, we introduce the following notation. Let $p > 0$, $q, r, s \in \mathbb{N}^*$, $u : \mathbb{R} \rightarrow \mathbb{R}$, $a, b : \mathbb{R}^r \rightarrow \mathbb{R}$, $\mathbf{c} \in \mathbb{R}^q$ and a collection of random matrices $\{U_M \in \mathbb{R}^{r \times s}, M \in \mathbb{N}^*\}$. We write

- (i) $U_M = O_{L^p, \mathbf{c}}(u(M))$ if there exists $C_{p, \mathbf{c}} > 0$, depending continuously and at most on (p, \mathbf{c}') , such that for all $M \in \mathbb{N}^*$

$$\max_{1 \leq i \leq r, 1 \leq j \leq s} \left(\mathbb{E} \left[|U_{M,i,j}|^p \right] \right)^{1/p} \leq C_{p, \mathbf{c}} |u(M)|, \quad (5.1)$$

where $U_{M,i,j}$ is the (i, j) -th entry of the matrix U_M .

- (ii) $U_M = O_{L^*, \mathbf{c}}(u(M))$ if $U_M = O_{L^p, \mathbf{c}}(u(M))$ for all $p > 0$.
- (iii) $a(\mathbf{x}) = O_{\mathbf{c}}(b(\mathbf{x}))$ if and only if there exists a constant $C_{\mathbf{c}}$ depending continuously and at most on the index \mathbf{c} , such that for all $\mathbf{x} \in \mathbb{R}^r$

$$|a(\mathbf{x})| \leq C_{\mathbf{c}} |b(\mathbf{x})|.$$

Concerning the function h we have the following assumption.

- (H) The function $h : [0, 1] \rightarrow \mathbb{R}$ is piecewise continuously differentiable, that is, for $0 = u_0 < u_1 < \dots < u_N = 1$, h is C^1 on $(u_{i-1}, u_i]$, $i = 1, \dots, N$. Moreover we denote $\|h\|_{\infty} = \sup_{u \in [0, 1]} |h(u)|$ and $\|h'\|_{\infty} = \max_{1 \leq i \leq N} \sup_{u \in (u_{i-1}, u_i]} |h'(u)|$.

Provided a piecewise continuously differentiable function h (as in (H)) and a local spectral density function f continuously differentiable on its first component, we also consider the following assumption.

- (C) For all $q > 0$, $M_q := \sup_{t,T} \mathbb{E} \left[|X_{t,T}|^q \right] < \infty$ and for all $\ell \in \mathbb{Z}$ the empirical covariance function satisfy

$$\gamma_M(u, \ell) - \mathbb{E}[\gamma_M(u, \ell)] = O_{L^*, \ell, f, \|h\|_{\infty}, \|h'\|_{\infty}, \|f\|_{\infty}, \|\partial f / \partial u\|_{\infty}} \left(M^{-1/2} \right).$$

At first glance Assumption (C) may seem restrictive but it is not. Locally stationary processes of Example 1 satisfy it (see [13, Theorem 4.1]) and also m-dependent sequences as those in Example 3.

5.2 Bound of the estimation risk

Our first result provides an equality satisfied by the estimation error of $\widehat{\boldsymbol{\theta}}_{t,T}(M)$.

Theorem 5.1. *Let $\beta \geq 2, R, f_- > 0$ and $h : [0, 1] \rightarrow \mathbb{R}$. Let $k \in \mathbb{N}$ and $\alpha \in (0, 1]$ be uniquely defined such that $\beta = k + \alpha$ and consider $M \in 2\mathbb{N}^*$. Suppose that Assumptions **(M-1)**, **(H)** and **(C)** hold. Let $\widehat{\boldsymbol{\theta}}_{t,T}(M)$ be obtained from Equation (4.1). The following relation is satisfied*

$$\widehat{\boldsymbol{\theta}}_{t,T}(M) - \boldsymbol{\theta}_{t/T} = \sum_{\ell=1}^k \mathbf{a}_{h,f,\ell} \left(\frac{M}{T}\right)^\ell + O_{d,f,-,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\frac{1}{M} + \left(\frac{M}{T}\right)^\beta\right) + \mathbf{v}_M, \quad (5.2)$$

where $\mathbf{a}_{h,f,\ell}$ depends only on h , the spectral density f and ℓ and $\mathbf{v}_M = O_{L^*,d,f,-,\|h\|_\infty,\|h'\|_\infty,\beta,R}(M^{-1/2})$. Moreover, $\mathbf{a}_{h,f,1} = 0$ if $h(x) = h(1-x)$ for $x \in [0, 1]$.

The proof can be found in Appendix B.1. Theorem 5.1 suggests to combine several $\widehat{\boldsymbol{\theta}}_{t,T}(M)$ to obtain a more accurate estimation by cancelling out the first k bias terms in (5.2). The technique was already used for eliminate one term of bias in [22, Theorem 8] for example. It is inspired by the Romberg's method in numerical analysis (see [2]). Let $\boldsymbol{\alpha} = [\alpha_0 \dots \alpha_k]' \in \mathbb{R}^{k+1}$, be the solution of the equation

$$A\boldsymbol{\alpha} = \mathbf{e}_1, \quad (5.3)$$

where $\mathbf{e}_1 = [1 \ 0 \ \dots \ 0]'$ is the \mathbb{R}^{k+1} - vector having a 1 in the first position and zero everywhere else and A is a $(k+1) \times (k+1)$ matrix with entries $A_{i,j} = 2^{-ij}$ for $0 \leq i, j \leq k$.

Theorem 5.2. *Let $\beta \geq 2, R, f_- > 0$ and $h : [0, 1] \rightarrow \mathbb{R}$. Let $k \in \mathbb{N}$ and $\alpha \in (0, 1]$ be uniquely defined such that $\beta = k + \alpha$ and consider $M \in 2^{k+1}\mathbb{N}^*$. Suppose that Assumptions **(M-1)**, **(H)** and **(C)** hold. Let $\widetilde{\boldsymbol{\theta}}_{t,T}(M)$ be obtained from Equation (4.1). Then, $\boldsymbol{\theta}_{t,T}(M) = \sum_{\ell=0}^k \alpha_\ell \widetilde{\boldsymbol{\theta}}_{t,T}(M/2^\ell)$ with $\boldsymbol{\alpha}$ defined by (5.3) satisfies*

$$\widetilde{\boldsymbol{\theta}}_{t,T}(M) - \boldsymbol{\theta}_{t/T} = O_{d,f,-,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\frac{1}{M} + \left(\frac{M}{T}\right)^\beta\right) + \mathbf{v}_M, \quad (5.4)$$

where $\mathbf{v}_M = O_{L^*,d,f,-,\|h\|_\infty,\|h'\|_\infty,\beta,R}(M^{-1/2})$.

The proof is postponed to Appendix B.2. It is straightforward to check that the optimal bandwidth for minimizing the order of the right term of Equation (5.4) is $M \propto T^{2\beta/(2\beta+1)}$. The next result is a direct consequence of Lemma 3, Theorem 5.2 and this observation.

Corollary 1. *Let $\beta \geq 2, R, f_- > 0$ and $h : [0, 1] \rightarrow \mathbb{R}$. Let $k \in \mathbb{N}$ and $\alpha \in (0, 1]$ be uniquely defined such that $\beta = k + \alpha$ and consider $M = 2^{k+1} \lfloor T^{2\beta/(2\beta+1)} \rfloor$. Suppose that Assumptions **(M-1)**, **(H)** and **(C)** hold. Let $\widetilde{\boldsymbol{\theta}}_{t,T}(M)$ be obtained as in Theorem 5.2. Then, for any $q > 0$ there exist a constant C only depending on h, q, d, f_-, R and continuously on β and a $T_0 > 0$ only depending on d, R and f_- such that, if $T \geq T_0$ we have, for all $t \in \mathbb{Z}$,*

$$\mathbb{E} \left[\left\| \widetilde{\boldsymbol{\theta}}_{t,T}(M) - \boldsymbol{\theta}_{t,T}^* \right\|^q \right] \leq \frac{C}{T^{q\beta/(2\beta+1)}}. \quad (5.5)$$

6 Application to TVAR processes

Time varying autoregressive processes (see Example 2) are a handful model to illustrate our results.

The index β sets the regularity of the functions we are interested in (the TVAR parameter θ). The following concepts are related to standard stability conditions on them.

For $\theta : \mathbb{R} \rightarrow \mathbb{R}^p$, we define the time varying autoregressive polynomial by $\theta(z; u) = 1 - \sum_{j=1}^p \theta_j(u)z^j$.

Let us denote, for any $\delta > 0$, $s_p(\delta) = \{\theta : \mathbb{R} \rightarrow \mathbb{R}^p, \theta(z; u) \neq 0, \forall |z| < \delta^{-1}, u \in [0, 1]\}$.

Define, for $\beta > 0, R > 0, \delta \in (0, 1), \rho \in [0, 1]$ and $\sigma_+ > 0$, the class of parameters

$$C(\beta, R, \delta, \rho, \sigma_+) = \left\{ (\theta, \sigma) : \mathbb{R} \rightarrow \mathbb{R}^p \times [\rho\sigma_+, \sigma_+] : \theta \in \Lambda_p(\beta, R) \cap s_p(\delta) \right\}.$$

Given an i.i.d. sequence $(\xi_t)_{t \in \mathbb{Z}}$ and constants $\delta \in (0, 1), \rho \in [0, 1], \sigma_+ > 0, \beta > 0$ and $R > 0$, we consider the following assumption.

(M-2) The sequence $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ is a TVAR process with time varying standard deviation σ , time varying AR coefficients $\theta_1, \dots, \theta_p$ and innovations $(\xi_t)_{t \in \mathbb{Z}}$, and $(\theta, \sigma) \in C(\beta, R, \delta, \rho, \sigma_+)$.

A TVAR process admits a linear representation with respect to the innovations (see [16, Proposition 1]). It is convenient to introduce the assumption below.

(I) For all $q > 0$ the innovations $(\xi_t)_{t \in \mathbb{Z}}$ satisfy $m_q := \mathbb{E}[|\xi|^q] < \infty$.

Time varying autoregressive processes are locally stationary under certain conditions on their parameters and moments. The next result is consequence of [11, Theorem 2.3].

Theorem 6.1. *Let $\delta \in (0, 1), \beta > 0, R > 0$ and $\rho \in [0, 1]$. Suppose that Assumptions **(M-2)** and **(I)** hold. Then, the process is locally stationary in the sense of Definition 3 with*

$$f(u, \lambda) = \frac{\sigma^2(u)}{2\pi} \left(1 - \sum_{j=1}^p \theta_j(u) \exp(-ij\lambda) \right)^{-2}. \quad (6.1)$$

Moreover, $\theta_{t,T}^* \in \mathbb{R}^d$ as defined by Equation (2.8) coincides with $\theta(t/T)$ when $p = d$.

To apply the results of Section 5 to the TVAR fulfilling Assumption **(M-2)** and **(I)**, we should take care of the regularity of the spectral density and also of its bounds. The analysis of Appendix C points in that direction. From that, we conclude that the conditions of Corollary 1 are fulfilled.

Corollary 2. *Let $\delta \in (0, 1), \beta \geq 2, R > 0$ and $\rho \in [0, 1]$. Let $k \in \mathbb{Z}$ and $\alpha \in (0, 1]$ be uniquely defined such that $\beta = k + \alpha$ and consider $M = 2^{k+1} \lfloor T^{2\beta/(2\beta+1)} \rfloor$. Suppose that Assumptions **(M-2)** and **(I)** hold and that $\mathbb{P}(X_{t,T} = 0) = 0$ for any t . Assume moreover that $\sigma \in \Lambda_1(\beta, R)$. Let $\tilde{\theta}_{t,T}(M)$ be a p dimensional vector obtained as in Theorem 5.2 (i.e. $p = d$). Then, for any $q \in \mathbb{N}$ there exists a constant C only depending*

on $q, h, p, \delta, \rho, \sigma_+, R$ and continuously on β , and $T_0 > 0$ depending only on $p, \delta, \rho, \sigma_+, R$ and β such that, for $T \geq T_0$ we have

$$\mathbb{E} \left[\left\| \tilde{\theta}_{t,T}(M) - \theta \left(\frac{t}{T} \right) \right\|^q \right] \leq \frac{C}{T^{q\beta/(2\beta+1)}}. \quad (6.2)$$

The estimator $\tilde{\theta}$ proposed in Corollary 2 is β -minimax-rate for TVAR processes according to [22, Theorem 4]. Hence, it is also β -minimax-rate in the class of locally stationary processes satisfying Assumption (M-1). [16, Section A.1] explains how to construct minimax-rate predictors from minimax-rate estimators of θ . Applying their approach, Corollary 2 also provides a crucial ingredient in building β -minimax-rate predictors for any $\beta \geq 2$.

7 Numerical work

We test both methods on data simulated according to a TVAR process with $p = 3$. The smooth parameter function $t \mapsto \theta(t)$ within $s_p(\delta)$ for some $\delta \in (0, 1)$ is chosen as follows. First we pick randomly some smoothly time varying partial autocorrelation functions up to the order p that are bounded between -1 and 1 , $\theta_{k,k}(u) = \delta^k \sum_{j=1}^{F-1} a_{j,k} j^2 \cos(ju) / [F(F-1)(2F-1)/6]$, where $a_{j,k}$ are random numbers in $[-1, 1]$, the same ones for all u . Then we use Algorithm 1 and set $\theta = -[\theta_{1,p} \dots \theta_{p,p}]$. From the classical Levinson-Durbin recurrence (i.e. Algorithm 1 with $\delta = 1$) we obtain a function in $s_p(1)$ (see for example [21]), it is straightforward to check that the θ produced by Algorithm 1 with $\delta \in (0, 1)$ is in $s_p(\delta)$. The three components of our $\theta(t)$ are displayed in Figure 1. The generated θ is, in theory, C^∞ . We can

Algorithm 1: Adapted Levinson-Durbin algorithm.

parameters the stability parameter $\delta > 0$ and the time varying partial autocorrelation functions $\theta_{k,k}, k = 1, \dots, p$;

for $k = 2$ **to** p **do**

for $j = 2$ **to** $p - 1$ **do**

$\theta_{j,k} = \theta_{j,k-1} + \delta^{2j-2k} \theta_{k,k} \theta_{k-j,k-1}$;

then ensure that for any $\beta > 0$, it is in $\Lambda_p(\beta, R)$ for some $R > 0$. For convenience we build $\tilde{\theta}$ with $k = 1$.

For each $T \in \{2^{2j}, j = 5, \dots, 15\}$ we generate 100 realizations of a TVAR process from innovation sequences $(\xi_t)_{t \in \mathbb{Z}}$ of i.i.d. centred Gaussian random variables with unit variance by sampling θ at a rate T^{-1} . Then we compare $\tilde{\theta}$ and $\hat{\theta}$ for estimating $\theta(1/2)$ using $h \equiv 1$ and different values of M . Recall that $\theta(1/2) = \theta_{T/2,T}^*$. Figure 2 shows the boxplots corresponding to this evaluation for two different T s.

In Figure 2 we observe that for $T = 2^{20}$ the error of $\hat{\theta}$ is minimized in $M = 2^{15}$ while that of $\tilde{\theta}$ reaches its minimum in $M = 2^{17}$. The estimator $\tilde{\theta}$ beats $\hat{\theta}$ for the two biggest values of M . In the case $T = 2^{30}$, the error of $\hat{\theta}$ reaches its minimum in $M = 2^{23} = T^{4/5}/2$ and that of $\tilde{\theta}$ in $M = 2^{26} = 2^2 T^{4/5}$. The estimator

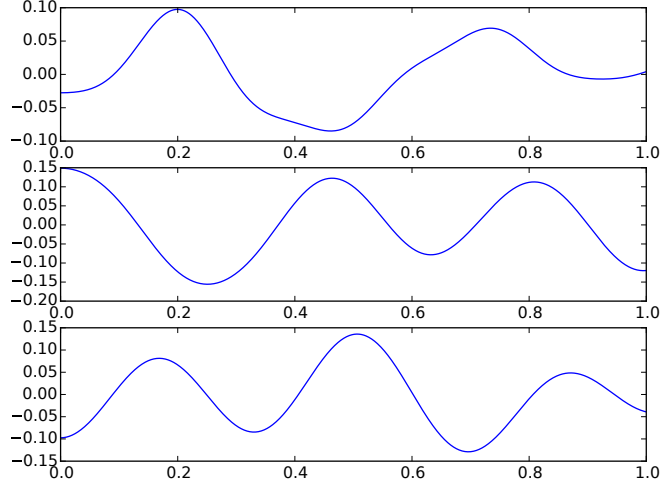


Figure 1: Plots of $\theta_1(t)$ (top), $\theta_2(t)$ (middle) and $\theta_3(t)$ (bottom) on the interval $t \in [0, 1]$.

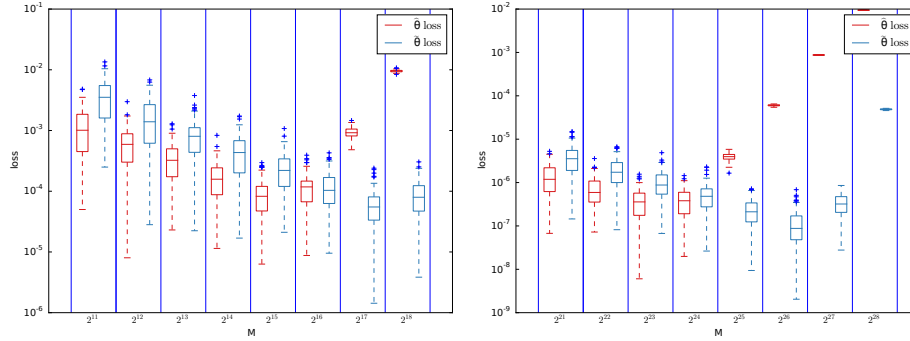


Figure 2: Box plots of the quadratic losses for estimating $\theta(1/2)$ using $\widehat{\theta}_{T/2,T}(M)$ (red boxes) and $\widetilde{\theta}_{T/2,T}(M)$ (blue boxes) for various bandwidths M , when $T = 2^{20}$ (left) and $T = 2^{30}$ (right).

$\widetilde{\theta}$ beats $\widehat{\theta}$ for the four biggest values of M . These experiences illustrate the theoretical result established in [13, Theorem 2.2] (where an optimal rate for $\widehat{\theta}$ estimation is obtained with $M \propto T^{4/5}$) and Corollary 2 (exhibiting the optimal rate for $\widetilde{\theta}$ estimation in $M \propto T^{4/5}$, if $\beta = 2$). Figure 3 (left graph) displays the oracle errors $\min_M \|\widehat{\theta}_{T/2,T}(M) - \theta(1/2)\|$ and $\min_M \|\widetilde{\theta}_{T/2,T}(M) - \theta(1/2)\|$ for all $T \in \{2^{2^j}, j = 5, \dots, 15\}$. The slope corresponding to $\widetilde{\theta}$ (in blue) is steeper than the one corresponding to $\widehat{\theta}$ (in red), meaning that, in average, $\widetilde{\theta}$ outperforms $\widehat{\theta}$ by an increasing order of magnitude as T increases. This corroborates what is expected from our theoretical

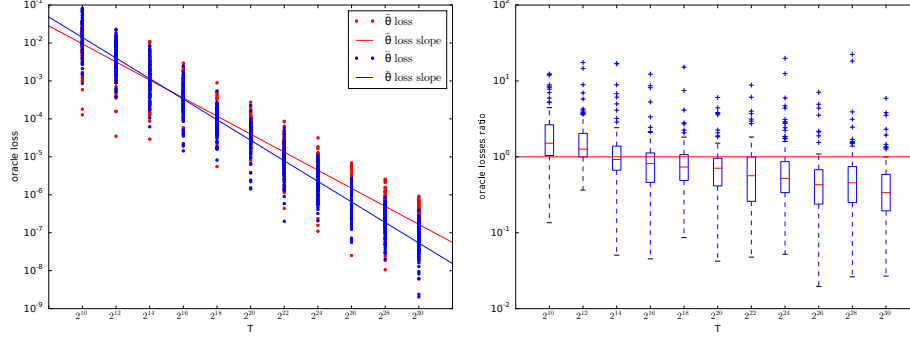


Figure 3: Oracle losses (using the best choice for the bandwidth M) for estimating $\theta(1/2)$ using $\widehat{\theta}_{T/2,T}(M)$ (red points) and $\widetilde{\theta}_{T/2,T}(M)$ (blue points) for various values of T . The left plot displays the losses over all the Monte Carlo simulations and the two resulting log-log regression lines. The right plot displays boxplots of the corresponding losses ratio.

analysis (see Corollary 2). The boxplots of Figure 3 (right graph) represent the ratios $\min_M \|\widehat{\theta}_{T/2,T}(M) - \theta(1/2)\| / \min_M \|\widetilde{\theta}_{T/2,T}(M) - \theta(1/2)\|$ computed for each T and realization of the TVAR process. Observe that for $2^{14} \leq T \leq 2^{18}$ the estimator $\widetilde{\theta}$ beats $\widehat{\theta}$ in at least half of the cases. For $T \geq 2^{20}$, it happens in at least 75% of the cases.

A Useful results on locally stationary time series

This section provides the background necessary to support the proof of our main results about locally stationary processes. The next two lemmas allow to control the norms of $\widehat{\theta}_{t,T}$ and $\theta_{t/T}$.

Lemma 1. *Let $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ be a locally stationary process in the sense of Definition 3 such that $\mathbb{P}(X_{t,T} = 0) = 0$ for any $t \leq T$. The Yule-Walker estimate $\widehat{\theta}_{t,T}(M)$ defined by Equation (4.1) satisfies $\|\widehat{\theta}_{t,T}(M)\| \leq 2^d$ almost surely.*

Proof. This proof is an adaptation of that of [13, Lemma 4.2].

We start by showing that $\widehat{\Gamma}_{t,T,M}$, with entries defined as in (4.1), is non-singular almost surely. Suppose on the contrary that $\mathbb{P}(\det(\widehat{\Gamma}_{t,T,M}) = 0) > 0$. This means that there is an $\mathbf{x} \in (\mathbb{R}^d)^*$ such that $\widehat{\Gamma}_{t,T,M}\mathbf{x} = 0$ and therefore $\mathbf{x}'\widehat{\Gamma}_{t,T,M}\mathbf{x} = \int_{-\pi}^{\pi} \widehat{f}_M(t/T, \lambda) |\sum_{j=1}^d x_j \exp(ij\lambda)|^2 d\lambda = 0$. The expression inside the modulus vanishes at most for $d-1$ values of λ , otherwise $\mathbf{x} = 0$ because the obtained Vandermonde determinant is non-zero. Then $\widehat{f}_M(t/T, \lambda) = 0$ for almost all $\lambda \in [-\pi, \pi]$. Since $\{\exp(-i\lambda s), s = 0, \dots, M-1\}$ is a subset of an orthogonal basis of $L^2([0, 1])$ we get that $h(s/M)X_{t-M/2+s+1,T} = 0$ for $s = 0, \dots, M-1$, but then $\mathbb{P}(X_{t,T} = 0) > 0$ for some $1 \leq t \leq T$.

Observe that for any s , $\widehat{\gamma}_M(s, \cdot)$ defined by (4.1) is an autocovariance function. Setting $s = t/T$, the corresponding covariance matrix $\widehat{\Gamma}_{t,T,M}$ is positive-definite almost

surely. As consequence of Lemma 9 (Appendix D) we have that z_1, \dots, z_d , the roots of the polynomial $\widehat{\theta}_{t,T}(z) = 1 - \sum_{j=1}^d \widehat{\theta}_{j,t,T} z^j$ satisfy $|z_j| > 1$ for any j . Then,

$$\left\| \widehat{\theta}_{t,T}(M) \right\|^2 + 1 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| 1 - \sum_{j=1}^d \widehat{\theta}_{j,t,T} \exp(ij\lambda) \right|^2 d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \widehat{\theta}_{t,T}(\exp(i\lambda)) \right|^2 d\lambda. \quad (\text{A.1})$$

Note that $\prod_{j=1}^d (-z_j) = 1$. Therefore

$$\widehat{\theta}_{t,T}(z) = \prod_{j=1}^d (z - z_j) = \prod_{j=1}^d (1 - z z_j^{-1}). \quad (\text{A.2})$$

If $|z| = 1$, Equation (A.2) implies that $|\widehat{\theta}_{t,T}(z)| \leq 2^d$. Putting this into (A.1) the proof is completed. \square

Lemma 2. *Let $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ be a locally stationary process in the sense of Definition 3. Assume that $f(u, \lambda) > 0$ for all $u, \lambda \in \mathbb{R}$. The vector θ_u defined by Equation (2.8) satisfies $\|\theta_u\| \leq 2^d$.*

Proof. The proof follows the same scheme of that of Lemma 1 up to simplifications. Here the contradiction $f(u, \lambda) = 0$ for almost all $\lambda \in [-\pi, \pi]$ raises immediately from the assumptions. Observe that, instead of an almost sure result, this is a deterministic one. \square

Lemma 2 is necessary to prove the following.

Lemma 3. *Let $(X_{t,T})_{t \in \mathbb{Z}, T \in \mathbb{N}^*}$ be a locally stationary process in the sense of Definition 3 where the spectral density f satisfies $f(u, \lambda) \geq f_-$ for all $u, \lambda \in \mathbb{R}$. Then, there exist two constants $C_1, T_0 > 0$ depending only on d, C (see Inequality (2.3)) and f_- such that, for $T \geq T_0$ we have*

$$\|\theta_{t,T}^* - \theta_{t/T}\| \leq \frac{C_1}{T}, \quad (\text{A.3})$$

Proof. From equations (2.7) and (2.8) we obtain that

$$\theta_{t,T}^* - \theta_{t/T} = (\Gamma_{t,T}^*)^{-1} \left[(\Gamma_{t/T} - \Gamma_{t,T}^*) \theta_{t/T} + \gamma_{t,T}^* - \gamma_{t/T} \right].$$

Applying matrix inequalities (specifically with the spectral norm) we get

$$\|\theta_{t,T}^* - \theta_{t/T}\| \leq \left\| (\Gamma_{t,T}^*)^{-1} \right\| \left(\|\Gamma_{t/T} - \Gamma_{t,T}^*\| \|\theta_{t/T}\| + \|\gamma_{t,T}^* - \gamma_{t/T}\| \right).$$

Inequality (2.3) implies that $\|\Gamma_{t/T} - \Gamma_{t,T}^*\| \leq d^{3/2}C/T$ and that $\|\gamma_{t,T}^* - \gamma_{t/T}\| \leq d^{1/2}C/T$. The smallest eigenvalue of the matrix $\Gamma_{t/T}$ is positive, at least $2\pi f_-$ (see [5, Proposition 4.5.3]). Observe that

$$\begin{aligned} \inf_t \inf_{\|a\|=1} a' \Gamma_{t,T}^* a &= \inf_t \inf_{\|a\|=1} \left\{ a' (\Gamma_{t,T}^* - \Gamma_{t/T}) a + a' \Gamma_{t/T} a \right\} \\ &\geq \inf_t \inf_{\|a\|=1} a' (\Gamma_{t,T}^* - \Gamma_{t/T}) a + \inf_t \inf_{\|a\|=1} a' \Gamma_{t/T} a \geq 2\pi f_- - \frac{d^{3/2}C}{T}. \end{aligned}$$

Then, for $T \geq T_0 = Cd^{3/2}(\pi f_-)^{-1}$ we have $\|(\Gamma_{t,T}^*)^{-1}\| \leq (\pi f_-)^{-1}$. Lemma 2 ensures that $\|\theta_{t/T}\| \leq 2^d$ and the result follows with $C_1 = Cd^{1/2}(\pi f_-)^{-1}(d2^d + 1)$. \square

Theorem A.1. *Let $d \in \mathbb{N}^*$, $\beta \geq 2$, $R > 0$, $f_- = 0$ and $h : [0, 1] \rightarrow \mathbb{R}$. Let $k \in \mathbb{N}$ and $\alpha \in (0, 1]$ be uniquely defined such that $\beta = k + \alpha$ and consider $M \in 2\mathbb{N}^*$ with $M > d$. Suppose that Assumptions **(M-1)** and **(H)** hold. Then, for all $-d \leq j \leq d$ we have*

$$\mathbb{E} \left[\widehat{\gamma}_M \left(\frac{t}{T}, j \right) \right] = \gamma \left(\frac{t}{T}, j \right) + \sum_{\ell=1}^k c_{h,f,j,\ell} \left(\frac{M}{T} \right)^\ell + O_{d,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\frac{1}{M} + \left(\frac{M}{T} \right)^\beta \right),$$

where $c_{h,f,j,\ell}$ only depends on h , the spectral density f , j and ℓ . If $h(x) = h(1-x)$ for all $x \in [0, 1]$, then $c_{h,f,j,1} = 0$.

Our proof of Theorem A.1 can be found in Appendix A.1. It uses the following lemma.

Lemma 4. *Let $\beta > 0$ and $R > 0$. Consider $f : \mathbb{R} \rightarrow \mathbb{R}$, a function in $\Lambda_1(\beta, R)$ and $a \in \mathbb{R}$. Let $k \in \mathbb{N}$ and $\alpha \in (0, 1]$ be uniquely defined such that $\beta = k + \alpha$. The function f admits the representation*

$$f(x) = \sum_{\ell=0}^k \frac{f^{(\ell)}(a)}{\ell!} (x-a)^\ell + f_k(x), \quad (\text{A.4})$$

where $f_k(x) = O_{\beta,R}((x-a)^\beta)$.

Proof. The expression (A.4) corresponds to the Taylor expansion of the function f . Without loss of generality, let $x > a$. We just need to prove that the remainder term is bounded by $(x-a)^\beta$ up to a constant. Using the definition of the norm $|\cdot|_\beta$ we have $f_k^{(k)}(x) \leq R(x-a)^\alpha$. The result follows by integrating k times the previous inequality. \square

A.1 Proof of Theorem A.1

Without loss of generality let $j \geq 0$. We start by expressing $\widehat{\gamma}_M$ in function of γ^*

$$\begin{aligned} \mathbb{E} \left[\widehat{\gamma}_M \left(\frac{t}{T}, j \right) \right] &= \frac{1}{H_M} \sum_{\substack{t_1, t_2=1 \\ t_1-t_2=\ell}}^M h \left(\frac{t_1}{M} \right) h \left(\frac{t_2}{M} \right) \mathbb{E} [X_{t+t_1-M/2,T} X_{t+t_2-M/2,T}] , \\ &= \frac{1}{H_M} \sum_{s=j+1}^M h \left(\frac{s}{M} \right) h \left(\frac{s-j}{M} \right) \gamma^* \left(t+s-\frac{M}{2}, T, j \right) . \end{aligned}$$

Since Inequality (2.3) guaranties that

$$\left| \gamma^* \left(t+s-\frac{M}{2}, T, j \right) - \gamma \left(\frac{t+s-M/2}{T}, j \right) \right| = O_R \left(\frac{1}{T} \right), \quad (\text{A.5})$$

we evaluate

$$\gamma_{M,j} = \frac{1}{H_M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) \gamma\left(\frac{t+s-M/2}{T}, T, j\right),$$

and then use the expression of $\gamma_{M,j}$ for computing $\mathbb{E}[\widehat{\gamma}_M(t/T, j)]$.

We apply Lemma 4 on the first component of f . The corresponding ℓ -th derivative is denoted by ∂_1^ℓ .

$$f\left(\frac{t-M/2+s}{T}, \lambda\right) = \sum_{\ell=0}^k \frac{\partial_1^\ell f(t/T, \lambda)}{\ell!} \left(\frac{-M/2+s}{T}\right)^\ell + f_k\left(\frac{t-M/2+s}{T}, \lambda\right),$$

with $f_k((t-M/2+s)/T, \lambda) = O_{\beta,R}((M/T)^\beta)$. Then

$$\begin{aligned} \gamma_{M,j} &= \frac{1}{H_M} \int_{-\pi}^{\pi} \exp(ij\lambda) \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) f\left(\frac{t-M/2+s}{T}, \lambda\right) d\lambda = \\ &= \sum_{\ell=0}^k \int_{-\pi}^{\pi} \frac{\partial_1^\ell f(t/T, \lambda)}{\ell!} \exp(ij\lambda) \frac{1}{H_M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) \left(\frac{-M/2+s}{T}\right)^\ell d\lambda \\ &\quad + \int_{-\pi}^{\pi} \exp(ij\lambda) \frac{1}{H_M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) f_k\left(\frac{t-M/2+s}{T}, \lambda\right) d\lambda. \quad (\text{A.6}) \end{aligned}$$

Note that for all $\ell = 1, \dots, k$

$$\begin{aligned} \frac{1}{H_M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) \left(\frac{-M/2+s}{T}\right)^\ell &= \\ &= \left(\frac{M}{T}\right)^\ell \frac{M}{H_M} \frac{1}{M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) \left(-\frac{1}{2} + \frac{s}{M}\right)^\ell. \quad (\text{A.7}) \end{aligned}$$

Since h is piecewise C^1 , maybe except for N values of s in $j+1, \dots, M$ we have

$$h\left(\frac{s-j}{M}\right) = h\left(\frac{s}{M}\right) + O_{\|h'\|_\infty} \left(\frac{d}{M}\right),$$

and we express the right-hand side of (A.7) as two right Riemann sums

$$\begin{aligned}
\frac{1}{M} \sum_{s=j+1}^M h^2\left(\frac{s}{M}\right) \left(-\frac{1}{2} + \frac{s}{M}\right)^\ell &= \int_0^1 h^2(u) \left(u - \frac{1}{2}\right)^\ell du + \frac{\|h\|_\infty (\|h'\|_\infty + \ell \|h\|_\infty)}{2^\ell M} o_{1,M,\ell} \\
&\quad + \frac{d \|h\|_\infty^2}{M} o_{2,M,\ell}, \\
\frac{1}{M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) \left(-\frac{1}{2} + \frac{s}{M}\right)^\ell &= \int_0^1 h(u) \left(u - \frac{1}{2}\right)^\ell du + \frac{\|h'\|_\infty + 2\ell \|h\|_\infty}{2^{\ell+1} M} o_{3,M,\ell} \\
&\quad + \frac{d \|h\|_\infty}{M} o_{4,M,\ell},
\end{aligned}$$

with $|o_{i,M,\ell}| \leq 1$ for $i = 1, \dots, 4$. Analogously

$$\frac{M}{H_M} = \left(\int_0^1 h^2(u) du \right)^{-1} \left(1 + \frac{\|h\|_\infty \|h'\|_\infty}{2^\ell M} o_{M,\ell} \right), \quad (\text{A.8})$$

with $|o_{M,\ell}| \leq 1$. Hence

$$\frac{1}{H_M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) \left(\frac{-M/2+s}{T}\right)^\ell = c_{h,\ell} \left(\frac{M}{T}\right)^\ell + O_{d,\|h\|_\infty,\|h'\|_\infty} \left(\frac{1}{M}\right) \left(\frac{M}{T}\right)^\ell. \quad (\text{A.9})$$

Observe that $c_{h,0} = 1$ and $c_{h,1} = 0$ if $h(x) = h(1-x)$ for all $x \in [0, 1]$. Using (A.9) and the upper bound on f_k , we express the terms of the second and third lines of (A.6) as follows

$$\begin{aligned}
\int_{-\pi}^{\pi} \frac{\partial_1^\ell f(t/T, \lambda)}{\ell!} \exp(ij\lambda) \frac{1}{H_M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) \left(\frac{-M/2+s}{T}\right)^\ell d\lambda &= c_{h,f,j,\ell} \left(\frac{M}{T}\right)^\ell \\
&\quad + O_{d,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\frac{1}{M}\right), \\
\int_{-\pi}^{\pi} \phi(\lambda) \frac{1}{H_M} \sum_{s=j+1}^M h\left(\frac{s}{M}\right) h\left(\frac{s-j}{M}\right) f_k\left(\frac{t-M/2+s}{T}, \lambda\right) d\lambda &= O_{d,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\left(\frac{M}{T}\right)^\beta\right),
\end{aligned}$$

where, in particular $c_{h,f,j,0} = \gamma(t/T, j)$. This implies that

$$\gamma_{M,j} = \gamma\left(\frac{t}{T}, j\right) + \sum_{\ell=1}^k c_{h,f,j,\ell} \left(\frac{M}{T}\right)^\ell + O_{d,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\frac{1}{M} + \left(\frac{M}{T}\right)^\beta\right).$$

Note that the relation (A.5) together with (A.9) evaluated in $\ell = 0$ allow to conclude the proof.

B Proof of bounds of the estimation risk

B.1 Proof of Theorem 5.1

We start by enunciating and proving the following.

Lemma 5. *Let d be a positive integer. Consider the $d \times d$ real non singular matrices Γ and $\widehat{\Gamma}$ and the vectors $\boldsymbol{\gamma}, \widehat{\boldsymbol{\gamma}}, \boldsymbol{\theta}, \widehat{\boldsymbol{\theta}} \in \mathbb{R}^d$ satisfying the relations*

$$\Gamma \boldsymbol{\theta} = \boldsymbol{\gamma}, \quad (\text{B.1})$$

$$\widehat{\Gamma} \widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\gamma}}. \quad (\text{B.2})$$

Then, for any $k \in \mathbb{N}$ we have

$$\begin{aligned} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} &= \left(\Gamma^{-1} + \sum_{\ell=1}^k (\Gamma^{-1} (\Gamma - \widehat{\Gamma}))^\ell \right) (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\ &\quad + \sum_{\ell=1}^{k+1} (\Gamma^{-1} (\Gamma - \widehat{\Gamma}))^\ell \boldsymbol{\theta} + (\Gamma^{-1} (\Gamma - \widehat{\Gamma}))^{k+1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \end{aligned} \quad (\text{B.3})$$

Proof. From Equations (B.1) and (B.2) we get

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \Gamma^{-1} [(\Gamma - \widehat{\Gamma}) \widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}].$$

The result follows by applying recursion. \square

Gathering together Assumption (C) and Theorem A.1 yields

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_M \left(\frac{t}{T}, j \right) &= \boldsymbol{\gamma} \left(\frac{t}{T}, j \right) + \sum_{\ell=1}^k c_{h,f,j,\ell} \left(\frac{M}{T} \right)^\ell \\ &\quad + O_{d, \|h\|_\infty, \|h'\|_\infty, \beta, R} \left(\frac{1}{M} + \left(\frac{M}{T} \right)^\beta \right) + u_M \left(\frac{t}{T}, j \right), \end{aligned} \quad (\text{B.4})$$

where $u_M(t/T, j) = O_{L^*, \|h\|_\infty, \|h'\|_\infty, R, j}(M^{-1/2})$ and $c_{h,f,j,1} = 0$ if $h(x) = h(1-x)$ for all $x \in [0, 1]$.

For the sake of simplicity, we drop t, T in the notation and set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_{t/T}, \widehat{\boldsymbol{\gamma}}_M = \widehat{\boldsymbol{\gamma}}_{t,T,M}, \Gamma \equiv \Gamma_{t/T}$ and $\widehat{\Gamma}_M \equiv \widehat{\Gamma}_{t,T,M}$. Using the expression (B.4), we choose $j = 0, \dots, d$ and obtain

$$\Gamma - \widehat{\Gamma}_M = \sum_{\ell=1}^k C_{h,f,\ell} \left(\frac{M}{T} \right)^\ell + O_{d, \|h\|_\infty, \|h'\|_\infty, \beta, R} \left(\frac{1}{M} + \left(\frac{M}{T} \right)^\beta \right) + U_M, \quad (\text{B.5})$$

$$\widehat{\boldsymbol{\gamma}}_M - \boldsymbol{\gamma} = \sum_{\ell=1}^k \mathbf{c}_{h,f,\ell} \left(\frac{M}{T} \right)^\ell + O_{d, \|h\|_\infty, \|h'\|_\infty, \beta, R} \left(\frac{1}{M} + \left(\frac{M}{T} \right)^\beta \right) + \mathbf{u}_M, \quad (\text{B.6})$$

where the matrices $C_{h,f,\ell} \in \mathbb{R}^{d \times d}$ and the vectors $\mathbf{c}_{h,f,\ell} \in \mathbb{R}^d$ only depend on h, f and ℓ . Furthermore $U_M = O_{L^*, d, \|h\|_\infty, \|h'\|_\infty, \beta, R}(M^{-1/2})$ and $\mathbf{u}_M = O_{L^*, d, \|h\|_\infty, \|h'\|_\infty, \beta, R}(M^{-1/2})$. Again $C_{h,f,1} = 0$ and $\mathbf{c}_{h,f,1} = 0$ if $h(x) = h(1-x)$ for all $x \in [0, 1]$.

Note that the product of $q > 0$ expressions for the form

$$\sum_{\ell=1}^k C_{h,\ell} \left(\frac{M}{T}\right)^\ell + O_{d,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\frac{1}{M} + \left(\frac{M}{T}\right)^\beta\right) + V_M,$$

with $V_M = O_{L^*,d,\|h\|_\infty,\|h'\|_\infty,\beta,R}(M^{-1/2})$, has the form

$$\sum_{\ell=q}^k D_{h,\ell} \left(\frac{M}{T}\right)^\ell + O_{d,\|h\|_\infty,\|h'\|_\infty,\beta,R} \left(\frac{1}{M} + \left(\frac{M}{T}\right)^\beta\right) + W_M,$$

with $W_M = O_{L^*,d,\|h\|_\infty,\|h'\|_\infty,\beta,R}(M^{-1/2})$ and $D_{h,\ell} = 0$ for $\ell \in [q, 2q]$ if all the $C_{h,1}$ of the factors vanish. This remark, together with (B.3) of Lemma 5, (B.5), (B.6), and the bounds provided by Lemma 1, Lemma 2 and Lemma 8 imply what is claimed in (5.2).

B.2 Proof of Theorem 5.2

For each $\ell = 0, \dots, k$ plug $M/2^\ell$ instead of M into Equation (5.2), multiply the resulting expression by α_ℓ and sum. Matrix A (definition below Equation (5.3)) is a non singular Vandermonde matrix and α is well defined. If $h(x) = h(1-x)$ for $x \in [0, 1]$ we can remove the second row of matrix A because the first order term of (5.2) is zero.

C Useful results on time varying autoregressive processes

Let δ be a positive real number. Consider the set

$$s_{(p)}(\delta) = \left\{ \theta \in \mathbb{R}^p : \theta(z) = 1 - \sum_{k=1}^p \theta_k z^k \neq 0, \text{ for } |z| < \delta^{-1} \right\}. \quad (\text{C.1})$$

As an immediate consequence of Hurwitz's theorem (see [9, Theorem 2.5] or [15, Section 3, Chapter VIII]) we obtain the following lemma.

Lemma 6. *For any $\delta > 0$ the set $s_{(p)}(\delta)$ defined by Equation (C.1) is closed and $\min_{\theta \in s_{(p)}(\delta)} \|\theta\|_\infty > 0$ only depends on p and δ .*

Since $\theta \in s_p(\delta)$ and $\sigma \in \Lambda_1(\beta, R)$, thank to Lemma 6 we have that for any $\lambda \in \mathbb{R}$ the spectral density $f(\cdot, \lambda)$ belongs to a $\Lambda_1(\beta, R')$ with R' depending only on R, δ and continuously on β . A direct consequence of Lemma 6 is given below.

Lemma 7. *Let $\delta \in (0, 1), \beta > 0, R > 0$ and $\rho \in [0, 1]$. Suppose that Assumptions (M-2) and (I) hold. There exist two constants f_-, f_+ , depending only on p, δ, ρ and σ_+ such that $0 < f_- \leq f(u, \lambda) \leq f_+$ for all $u, \lambda \in \mathbb{R}$.*

D Useful results on weakly stationary processes

In the context of real weakly stationary processes (see [4] and [24]), the autocovariance matrix of $(X_t)_{t \in \mathbb{Z}}$, that we call Γ_d , is Toeplitz and symmetric. Observe that

$$\Gamma_d = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(d-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(d-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(d-1) & \gamma(d-2) & \gamma(d-3) & \dots & \gamma(0) \end{bmatrix}. \quad (\text{D.1})$$

Proposition 1. *A complex-valued function defined on \mathbb{Z} is the autocovariance function of a weakly stationary process $(X_t)_{t \in \mathbb{Z}}$ taking values in \mathbb{C} if and only if the following two properties hold.*

(i) *Hermitian symmetry: for all $s \in \mathbb{Z}$,*

$$\gamma(-s) = \overline{\gamma(s)}.$$

(ii) *Nonnegativity: for all $d \in \mathbb{N}^*$ and $a_1, \dots, a_d \in \mathbb{C}$,*

$$\sum_{i=1}^d \sum_{j=1}^d \overline{a_i} \gamma(t_i - t_j) a_j \geq 0.$$

A crucial concept in the study of weakly stationary processes is the spectral measure, defined from the autocovariance function γ (see [4, Theorem 4.3.1]). We denote by $\mathcal{B}([-\pi, \pi])$ the Borel σ -algebra associated with $[-\pi, \pi]$.

Theorem D.1 (Herglotz). *A sequence γ is nonnegative definite and hermitian in the sense of Proposition 1 if and only if there exists a finite nonnegative measure ν on $([-\pi, \pi], \mathcal{B}([-\pi, \pi]))$ such that, for all $s \in \mathbb{Z}$:*

$$\gamma(s) = \int_{-\pi}^{\pi} \exp(ik\lambda) \nu(d\lambda). \quad (\text{D.2})$$

Furthermore, the measure ν is unique.

The next result links the spectral density function (when it exists) and the spectrum of the covariance matrix Γ_d .

Lemma 8. *Assume that the autocovariance function γ has a spectral density function $f \in [f_-, f_+]$ with $f_- \leq f_+$. For any $d \in \mathbb{N}^*$, the spectrum of the covariance matrix (Equation (D.1)) is contained in $[2\pi f_-, 2\pi f_+]$.*

Proof. Consider $\mathbf{a} = [a_1 \dots a_d]' \in \mathbb{R}^d$. If we express γ using the representation (D.2) we obtain

$$\mathbf{a}'\Gamma_d\mathbf{a} = \int_{-\pi}^{\pi} \left| \sum_{j=1}^d a_j \exp(ij\lambda) \right|^2 f(\lambda) d\lambda.$$

Therefore $2\pi f_+ \sum_{j=1}^d a_j^2 \geq \mathbf{a}'\Gamma_d\mathbf{a} \geq 2\pi f_- \sum_{j=1}^d a_j^2$. Choosing \mathbf{a} as any eigenvector of Γ_d the result follows. \square

The lemma below is similar in flavor to the statistical result of [26, Section 3]. It is also a classical property of orthogonal polynomials (see [17, Section 2.4]). We provide an elementary proof.

Lemma 9. *Let γ be a real autocovariance function (in the sense of Proposition 1) such that for any $d \in \mathbb{N}^*$, the covariance matrix Γ_d defined by Equation (D.1) is positive-definite. Denote the vector $\boldsymbol{\gamma}_d = [\gamma(1) \dots \gamma(d)]'$ and let $\widehat{\boldsymbol{\theta}} = [\widehat{\theta}_1 \dots \widehat{\theta}_d]' = \Gamma_d^{-1}\boldsymbol{\gamma}_d$. Then, all the roots of the polynomial $\widehat{\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^d \widehat{\theta}_j z^j$ are in the set $\{z \in \mathbb{C} : |z| > 1\}$.*

Proof. For $j = 1, \dots, d$, let $\mathbf{e}_j = [0 \dots 1 \dots 0]'$ be the \mathbb{R}^d -vector having a 1 in the j -th position and zero everywhere else. Consider also the matrix

$$A = \begin{bmatrix} \widehat{\theta}_1 & \widehat{\theta}_2 & \dots & \dots & \widehat{\theta}_d \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\theta}}' \\ \mathbf{e}'_1 \\ \vdots \\ \mathbf{e}'_{d-1} \end{bmatrix}.$$

Since the roots of $\widehat{\boldsymbol{\theta}}(z)$ are the inverses of the eigenvalues of A , we need to proof that these eigenvalues are inside the open unit disk. Observe that

$$\Gamma_d - A\Gamma_d A' = \Gamma_d - \begin{bmatrix} \widehat{\boldsymbol{\theta}}'\Gamma_d\widehat{\boldsymbol{\theta}} & \widehat{\boldsymbol{\theta}}'\Gamma_d\mathbf{e}_1 & \dots & \dots & \widehat{\boldsymbol{\theta}}'\Gamma_d\mathbf{e}_{d-1} \\ \mathbf{e}'_1\Gamma_d\widehat{\boldsymbol{\theta}} & \mathbf{e}'_1\Gamma_d\mathbf{e}_1 & \dots & \dots & \mathbf{e}'_1\Gamma_d\mathbf{e}_{d-1} \\ \mathbf{e}'_2\Gamma_d\widehat{\boldsymbol{\theta}} & \mathbf{e}'_2\Gamma_d\mathbf{e}_1 & \dots & \dots & \mathbf{e}'_2\Gamma_d\mathbf{e}_{d-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{e}'_{d-1}\Gamma_d\widehat{\boldsymbol{\theta}} & \mathbf{e}'_{d-1}\Gamma_d\mathbf{e}_1 & \dots & \dots & \mathbf{e}'_{d-1}\Gamma_d\mathbf{e}_{d-1} \end{bmatrix}.$$

Because Γ_d is a Toeplitz matrix, its (i, j) -th entries, and those of $A\Gamma_d A'$ are equal for $i, j \geq 2$. The definition of $\widehat{\boldsymbol{\theta}}$ implies also the equality of the (i, j) -th entries of both matrices when $i = 1, j \geq 2$ and $i \geq 2, j = 1$. Since $\widehat{\boldsymbol{\theta}}$ is the solution of $\Gamma_d\widehat{\boldsymbol{\theta}} = \boldsymbol{\gamma}_d$, we have that $\widehat{\theta}_j = -\Gamma_{d+1,d,j}/\det(\Gamma_d)$ where $\Gamma_{d,i,j}$ is the cofactor of the (i, j) -th entry of Γ_d . Finally, in the position $(1, 1)$ we have $\gamma(0) - \widehat{\boldsymbol{\theta}}'\boldsymbol{\gamma}_d = \sum_{j=0}^d \gamma(j)\Gamma_{d+1,d,j}/\det(\Gamma_d) = \det(\Gamma_{d+1})/\det(\Gamma_d) > 0$. Consider now λ , an eigenvalue of A and the corresponding eigenvector $\mathbf{v} \neq 0$. We verify that $\mathbf{v} = [\lambda^{d-1} \dots \lambda \ 1]'\mathbf{v}_d$. From the previous analysis we get $\mathbf{v}'(\Gamma_d - A\Gamma_d A')\mathbf{v} = |\lambda|^{2d-2}(1 - |\lambda|^2)\det(\Gamma_{d+1})/\det(\Gamma_d)|\mathbf{v}_d|^2 \geq 0$ and the inequality is strict if $\lambda \neq 0$. As claimed $|\lambda| < 1$. \square

Acknowledgements

This work has been partially supported by the Conseil régional d'Île-de-France under a doctoral allowance of its program Réseau de Recherche Doctoral en Mathématiques de l'Île de France (RDM-IdF) for the period 2012 - 2015 and by the Labex LMH (ANR-11-IDEX-003-02).

References

- [1] Oren Anava, Elad Hazan, Shie Mannor, and Ohad Shamir. Online learning for time series prediction. *J. Mach. Learn. Res.*, 30:172–184, 2013.
- [2] Jacques Baranger and Claude Brezinski. *Analyse numérique*, volume 38. Hermann, Paris, 1991. Collection Méthodes.
- [3] Andrew R. Barron. Are bayes rules consistent in information? In Thomas M. Cover and B. Gopinath, editors, *Open Problems in Communication and Computation*, pages 85–91. Springer New York, 1987.
- [4] Peter J. Brockwell and Richard A. Davis. *Introduction to time series and forecasting*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2002. With 1 CD-ROM (Windows).
- [5] Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer, New York, 2006. Reprint of the second (1991) edition.
- [6] Olivier Catoni. A mixture approach to universal model selection. Technical report, École Normale Supérieure, 1997.
- [7] Olivier Catoni. *Statistical learning theory and stochastic optimization*, volume 1851 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2004. Lecture notes from the 31st Summer School on Probability Theory held in Saint-Flour, July 8–25, 2001.
- [8] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, Cambridge, 2006.
- [9] John B. Conway. *Functions of one complex variable*. Springer-Verlag, New York-Heidelberg, 1973. Graduate Texts in Mathematics, 11.
- [10] R. Dahlhaus. Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonparametr. Statist.*, 6(2-3):171–191, 1996.
- [11] R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.*, 62(1):139–168, 1996.
- [12] Rainer Dahlhaus. Locally stationary processes. In T.S. Rao, S.S. Rao, and C.R. Rao, editors, *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*, pages 351–413. North Holland, 2012.

- [13] Rainer Dahlhaus and Liudas Giraitis. On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Ser. Anal.*, 19(6):629–655, 1998.
- [14] Marie Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.
- [15] Theodore W. Gamelin. *Complex analysis*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 2001.
- [16] Christophe Giraud, François Roueff, and Andres Sanchez-Perez. Aggregation of predictors for nonstationary sub-linear processes and online adaptive forecasting of time varying autoregressive processes. *Ann. Statist.*, 43(6):2412–2450, 2015.
- [17] Ulf Grenander and Gábor Szegő. *Toeplitz forms and their applications*. Chelsea Publishing Co., New York, second edition, 1984.
- [18] Anatoli Juditsky and Arkadii Nemirovski. Functional aggregation for nonparametric regression. *Ann. Statist.*, 28(3):681–712, 2000.
- [19] Gilbert Leung and Andrew R. Barron. Information theory and mixing least-squares regressions. *IEEE Trans. Inform. Theory*, 52(8):3396–3410, 2006.
- [20] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Inform. and Comput.*, 108(2):212–261, 1994.
- [21] John Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, april 1975.
- [22] Eric Moulines, Pierre Priouret, and François Roueff. On recursive estimation for time varying autoregressive processes. *Ann. Statist.*, 33(6):2610–2654, 2005.
- [23] M. B. Priestley. Evolutionary spectra and non-stationary processes.(With discussion). *J. Roy. Statist. Soc. Ser. B*, 27:204–237, 1965.
- [24] Robert H. Shumway and David S. Stoffer. *Time series analysis and its applications*. Springer Texts in Statistics. Springer, New York, third edition, 2011. With R examples.
- [25] Volodimir G Vovk. Aggregating strategies. In *Proc. Third Workshop on Computational Learning Theory*, pages 371–383, San Mateo, CA, 1990. Morgan Kaufmann.
- [26] P. Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50:129–134, 1963.
- [27] Yuhong Yang. Combining different procedures for adaptive regression. *J. Multivariate Anal.*, 74(1):135–161, 2000.

- [28] Yuhong Yang. Combining forecasting procedures: some theoretical results.
Econometric Theory, 20(1):176–222, 2004.