Linear Prediction of Long-Range Dependent Time Series

Fanny Godet*

Laboratoire de Mathématiques Jean Leray, UMR CNRS 6629
Université de Nantes 2 rue de la Houssinière - BP 92208 F-44322 Nantes Cedex 3

13th March 2007

Abstract

We present two approaches for next step linear prediction of long memory time series. The first is based on the truncation of the Wiener-Kolmogorov predictor by restricting the observations to the last k terms, which are the only available values in practice. Part of the mean squared prediction error comes from the truncation, and another part comes from the parametric estimation of the parameters of the predictor. By contrast, the second approach is non-parametric. An AR(k) model is fitted to the long memory time series and we study the error made with this misspecified model.

Keywords: Long memory, linear model, autoregressive process, forecast error

ARMA (autoregressive moving-average) processes are often called short-memory processes because their covariances decay rapidly (i.e. their covariance decay exponentially). By contrast, a long-memory process is characterised by the following feature: the autocovariance function σ decays more slowly i.e. it is not absolutely summable. They are so-named because of the strong association between observations widely separated in time. The long-memory time series models have attracted much attention lately and there is now a growing realisation that time series possessing long-memory characteristics arise in subject areas as diverse as Economics, Geophysics, Hydrology or telecom traffic (see, e.g., Mandelbrot and Wallis (1969) and Granger and Joyeux (1980)). Although there exists substantial literature on the prediction of short-memory processes(see Bhansali (1978) for the univariate case or Lewis and Reinsel (1985) for the multivariate case), there are less results for long-memory time series. In this paper, we consider the question of the prediction of the latter.

More precisely, we will compare two prediction methods for long-memory process. Our goal is a linear predictor $\widetilde{X}_{k+1}$ from observed values which is optimal in the sense that it minimizes the mean-squared error $\mathbb{E}\left[\left(X_{k+1}-\widetilde{X}_{k+1}\right)^2\right]$. This paper is organized as follows. First we will introduce our model and our main assumptions. Then in section 2, we study the best linear predictor i.e. the Wiener-Kolmogorov predictor proposed by Whittle (1963) and by Bhansali and Kokoszka (2001) for the long-memory time series. In practice, only the last k values of the process are available. Therefore we need to truncate the infinite series which defines the predictor and derive

*fanny.godet@math.univ-nantes.fr

the asymptotic behaviour as $k \to +\infty$ of the mean-squared error. Then we propose an estimator of the coefficients of the infinite autoregressive representation based on a realisation of length T. Under the simplifying assumption that the series used for estimation and the series used for prediction are generated from two independent process which have the same stochastic structure, we obtain an approximation of the mean-squared prediction error when $T \to +\infty$ and then $k \to +\infty$.

In Section 3, we discuss the asymptotic properties of the forecast error if we fit a misspecified AR(k) model to a long-memory time series. This approach has been proposed by Ray (1993) for fractional noise series F(d). His simulations show that high-order AR m-models forecast fractional integrated noise very well. In that case we also study the consequences of the estimation of the forecast coefficients. Therefore we shall rewrite the heuristic proof of Theorem 1 of Ray (1993) and develop a generalization of this result to a larger class of long-memory models. We conclude by comparing our asymptotic approximation for the global prediction error of long-memory processes and that of Berk (1974) and Bhansali (1978) in the case of short memory time series. Subsidiary proofs are given in the Appendix.

1 Model

Let $(X_n)_{n\in\mathbb{Z}}$ be a discrete-time (weakly) stationary process in L^2 with mean 0 and σ its autocovariance function. We assume that the process $(X_n)_{n\in\mathbb{Z}}$ is a long-memory process i.e.:

$$\sum_{k=-\infty}^{\infty} |\sigma(k)| = \infty.$$

The process $(X_n)_{n\in\mathbb{Z}}$ admits an infinite moving average representation as follows:

$$X_n = \sum_{j=0}^{\infty} b_j \varepsilon_{n-j} \tag{1}$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ is a white-noise series consisting of uncorrelated random variables, each with mean 0 and variance σ_ε^2 and $(b_j)_{j\in\mathbb{N}}$ are square-summable. We shall further assume that $(X_n)_{n\in\mathbb{Z}}$ admits an infinite autoregressive representation:

$$\varepsilon_n = \sum_{j=0}^{\infty} a_j X_{n-j}, \tag{2}$$

where the $(a_j)_{j\in\mathbb{N}}$ are absolutely summable. We assume also that $(a_j)_{j\in\mathbb{N}}$ and $(b_j)_{j\in\mathbb{N}}$, occurring respectively in (2) and (1), satisfy the following conditions for all $\delta > 0$:

$$|a_j| \leq C_1 j^{-d-1+\delta} \tag{3}$$

$$|b_j| \leq C_2 j^{d-1+\delta}. \tag{4}$$

where C_1 and C_2 are constants and d is a parameter verifying $d \in]0, 1/2[$. For example, a FARIMA process $(X_n)_{n\in\mathbb{Z}}$ is the stationary solution to the difference equations:

$$\phi(B)(1-B)^d X_n = \theta(B)\varepsilon_n$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ is a white noise series, B is the backward shift operator and ϕ et θ are polynomials with no zeroes on the unit disk. Its coefficients verify equations (3) and (4). In particular, if $\phi = \theta = 1$ then the process $(X_n)_{n\in\mathbb{Z}}$ is called *fractionally integrated noise* and denoted F(d). More generally, series like:

$$\begin{array}{rcl} |a_j| & \underset{+\infty}{\sim} & L(j^{-1})j^{-d-1} \\ |b_j| & \underset{+\infty}{\sim} & L'(j^{-1})j^{d-1} \end{array}$$

where L and L' are slowly varying functions and therefore verify conditions (3) and (4). A positive L will be called a slowly varying function in the sense of Zygmund (1968) if, for any $\delta > 0$, $x \mapsto x^{-\delta}L(x)$ is decreasing and $x \mapsto x^{\delta}L(x)$ is increasing.

The condition (4) implies that the autocovariance function σ of the process $(X_n)_{n\in\mathbb{Z}}$ verifies:

$$\forall \delta > 0, \exists C_3 \in \mathbb{R}, \quad |\sigma(j)| \leq C_3 j^{2d-1+\delta}. \tag{5}$$

Since, if $\delta < \frac{1-2d}{2}$:

$$\begin{aligned} \sigma(k) &= \sum_{j=0}^{+\infty} b_j b_{j+k} \\ |\sigma(k)| &\leq \sum_{j=0}^{+\infty} |b_j b_{j+k}| \\ &\leq C_2^2 \sum_{j=0}^{+\infty} j^{d-1+\delta}(k+j)^{d-1+\delta} \\ &\leq C_2^2 \int_{-1}^{+\infty} j^{d-1+\delta}(k+j)^{d-1+\delta}\mathrm{d}j \\ &\leq C_2^2 k^{2d-1+2\delta} \int_{-1}^{+\infty} j^{d-1+\delta}(1+j)^{d-1+\delta}\mathrm{d}j \\ &\leq C_3 k^{2d-1+2\delta} \end{aligned}$$

Notice that it suffices to prove (5) for δ near 0 in order to verify (5) for $\delta > 0$ arbitrarily chosen. More accurately, Inoue (1997) has proved than if:

$$b_j \sim L\left(j^{-1}\right) j^{d-1}$$

then

$$\sigma(j) \sim j^{2d-1} \left[L\left(j^{-1}\right)\right]^2 \beta(1-2d, d)$$

where L is a slowly varying function and β is the beta function. The converse is not true, we must have more assumptions about the series $(b_j)_{j\in\mathbb{N}}$ in order to get an asymptotic equivalent for $(\sigma(j))_{j\in\mathbb{N}}$ (see Inoue (2000)).

2 Wiener-Kolmogorov Prediction Theory

The aim of this part is to compute the best linear one-step predictor (with minimum mean-square distance from the true random variable) knowing all the past $\{X_{k+1-j}, j \leqslant 1\}$. Our predictor is therefore an infinite linear combination of the infinite past:

$$\widetilde{X_k}(1) = \sum_{j=0}^{\infty} \lambda(j) X_{k-j}$$

where $(\lambda(j))_{j\in\mathbb{N}}$ are chosen to ensure that the mean squared prediction error:

$$\mathbb{E}\big[\big(\widetilde{X_k}(1) - X_{k+1}\big)^2\big]$$

is as small as possible. Following Whittle (1963), and in view of the moving average representation of $(X_n)_{n\in\mathbb{Z}}$, we may rewrite our predictor $\widetilde{X_k}(1)$ as:

$$\widetilde{X_k}(1) = \sum_{j=0}^{\infty} \phi(j)\varepsilon_{k-j}.$$

where $(\phi(j))_{j\in\mathbb{N}}$ depends only on $(\lambda(j))_{j\in\mathbb{N}}$ and $(a_j)_{j\in\mathbb{N}}$ defined in (2). From the infinite moving average representation of $(X_n)_{n\in\mathbb{Z}}$ given below in (1), we can rewrite the mean-squared prediction error as:

$$\begin{aligned}
\mathbb{E}\big[\big(\widetilde{X_k}(1) - X_{k+1}\big)^2\big] &= \mathbb{E}\left[\left(\sum_{j=0}^{\infty} \phi(j)\varepsilon_{k-j} - \sum_{j=0}^{\infty} b(j)\varepsilon_{k+1-j}\right)^2\right] \\
&= \mathbb{E}\left[\left(\varepsilon_{k+1} - \sum_{j=0}^{\infty} \left(\phi(j) - b(j+1)\right)\varepsilon_{k-j}\right)^2\right] \\
&= \left(1 + \sum_{j=0}^{\infty} \big(b_{j+1} - \phi(j)\big)^2\right) {\sigma_\varepsilon}^2
\end{aligned}$$

since the random variables $(\varepsilon_n)_{n\in\mathbb{Z}}$ are uncorrelated with variance ${\sigma_\varepsilon}^2$. The smallest mean-squared prediction error is obtained when setting $\phi(j) = b_{j+1}$ for $j \geq 0$.

The smallest prediction error of $(X_n)_{n\in\mathbb{Z}}$ is ${\sigma_\varepsilon}^2$ within the class of linear predictors. Furthermore, if

$$A(z) = \sum_{j=0}^{+\infty} a_j z^j,$$

denotes the characteristic polynomial of the $(a(j))_{j\in\mathbb{Z}}$ and

$$B(z) = \sum_{j=0}^{+\infty} b_j z^j,$$

that of the $(a(j))_{j\in\mathbb{Z}}$, then in view of the identity, $A(z) = B(z)^{-1}$, $|z| \leq 1$, we may write:

$$\widetilde{X_k}(1) = -\sum_{j=1}^{\infty} a_j X_{k+1-j}. \tag{6}$$

2.1 Mean Squared Prediction Error when the Predictor is Truncated

In practice, we only know a finite part of the past, the one which we have observed. So the predictor should only depend on the observations. Assume that we only know the set $\{X_0, \dots, X_k\}$ and that we replace the unknown values by 0, then we have the following new predictor:

$$\widetilde{X'_k}(1) = -\sum_{j=1}^{k} a_j X_{k+1-j}. \tag{7}$$

It is equivalent to say that we have truncated the infinite series (6) to k terms. The following proposition provides us the asymptotic properties of the mean squared prediction error as a function of k.

Proposition 2.1.1. *Let $(X_n)_{n\in\mathbb{Z}}$ be a linear stationary process defined by* (1), (2) *and possessing the features* (3) *and* (4)*. We can approximate the mean-squared prediction error of $\widetilde{X'_k}(1)$ by:*

$$\forall \delta > 0, \quad \mathbb{E}\big(\big[X_{k+1} - \widetilde{X'_k}(1)\big]^2\big) = {\sigma_\varepsilon}^2 + \mathrm{O}(k^{-1+\delta}).$$

Furthermore, this rate of convergence $\mathrm{O}(k^{-1})$ is optimal since for fractionally integrated noise, we have the following asymptotic equivalent:

$$\mathbb{E}\big(\big[X_{k+1} - \widetilde{X'_k}(1)\big]^2\big) = {\sigma_\varepsilon}^2 + Ck^{-1} + \mathrm{o}\left(k^{-1}\right).$$

We note that the prediction error is the sum of ${\sigma_\varepsilon}^2$, the error of Wiener-Kolmogorov model and the error due to the truncation to k terms which is bounded by $\mathrm{O}(k^{-1+\delta})$ for all $\delta > 0$.

Proof.

$$\begin{aligned} X_{k+1} - \widetilde{X'_k}(1) &= X_{k+1} - \widetilde{X_k}(1) + \widetilde{X_k}(1) - \widetilde{X'_k}(1) \\ &= X_{k+1} - \sum_{j=0}^{+\infty} b_{j+1}\varepsilon_{k-j} - \sum_{j=k+1}^{+\infty} a_j X_{k+1-j} \\ &= \varepsilon_{k+1} - \sum_{j=k+1}^{+\infty} a_j X_{k+1-j}. \end{aligned} \tag{8}$$

The two parts of the sum (8) are orthogonal for the inner product associated with the mean square norm. Consequently:

$$\mathbb{E}\big(\big[X_{k+1} - \widetilde{X'_k}(1)\big]^2\big) = {\sigma_\varepsilon}^2 + \sum_{j=k+1}^{\infty}\sum_{l=k+1}^{\infty} a_j a_l \sigma(l-j).$$

For the second term of the sum we have:

$$\begin{aligned} \left|\sum_{j=k+1}^{+\infty}\sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j)\right| &= \left|2\sum_{j=k+1}^{+\infty} a_j \sum_{l=j+1}^{+\infty} a_l\sigma(l-j) + \sum_{j=k+1}^{+\infty} a_j^2\sigma(0)\right| \\ &\leq 2\sum_{j=k+1}^{+\infty} |a_j|\,|a_{j+1}|\,|\sigma(1)| + \sum_{j=k+1}^{+\infty} a_j^2\sigma(0) \\ &\quad +2\sum_{j=k+1}^{+\infty} |a_j| \sum_{l=j+2}^{+\infty} |a_l||\sigma(l-j)| \end{aligned}$$

from the triangle inequality, it follows that:

$$\begin{aligned}
&\Big|\sum_{j=k+1}^{+\infty}\sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j)\Big| \\
\leq\ & C_1^2 C_3 \left(2\sum_{j=k+1}^{+\infty} j^{-d-1+\delta}(j+1)^{-d-1+\delta} + \sum_{j=k+1}^{+\infty}\left(j^{-d-1+\delta}\right)^2\right) && (9)\\
+\ & 2C_1^2 C_3 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta} \sum_{l=j+2}^{+\infty} l^{-d-1+\delta}|l-j|^{2d-1+\delta} && (10)
\end{aligned}$$

for all $\delta > 0$ from inequalities (3) and (5). Assume now that $\delta < 1/2 - d$. For the terms (9), since $j \mapsto j^{-d-1+\delta}(j+1)^{-d-1+\delta}$ is a positive and decreasing function on $\mathbb{R}^+$, we have the following approximations:

$$\begin{aligned}
2C_1^2 C_3 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta}(j+1)^{-d-1+\delta} &\sim 2C_1^2 C_3 \int_k^{+\infty} j^{-d-1+\delta}(j+1)^{-d-1+\delta}\mathrm{d}j \\
&\sim \frac{2C_1^2 C_3}{1+2d-2\delta} k^{-2d-1+2\delta}
\end{aligned}$$

Since the function $j \mapsto \left(j^{-d-1+\delta}\right)^2$ is also positive and decreasing, we can establish in a similar way that:

$$\begin{aligned}
C_1^2 C_3 \sum_{j=k+1}^{+\infty}\left(j^{-d-1+\delta}\right)^2 &\sim C_1^2 C_3 \int_k^{+\infty}\left(j^{-d-1+\delta}\right)^2 \mathrm{d}j \\
&\sim \frac{C_1^2 C_3}{1+2d-2\delta} k^{-2d-1+2\delta}.
\end{aligned}$$

For the infinite double series (10), we will similarly compare the series with an integral. In the next Lemma, we establish the necessary result for this comparison:

Lemma 2.1.1. *Let g the function $(l,j) \mapsto j^{-d-1+\delta}\, l^{-d-1+\delta}\,|l-j|^{2d-1+\delta}$. Let m and n be two positive integers. We assume that $\delta < 1-2d$ and $m \geq \frac{\delta-d-1}{\delta+2d-1}$ for all $\delta \in \left]0, \frac{\delta-d-1}{\delta+2d-1}\right[$. We will call $A_{n,m}$ the square $[n, n+1]\times[m, m+1]$. If $n \geq m+1$ then*

$$\int_{A_{n,m}} g(l,j)\,\mathrm{d}j\,\mathrm{d}l \geq g(n+1, m).$$

Proof. see the appendix 4.1 □

Assume now that $\delta < 1-2d$ without loss of generality. Thanks to the previous Lemma and the asymptotic equivalents of (9), there exists $K \in \mathbb{N}$ such that if $k > K$:

$$\Big|\sum_{j=k+1}^{+\infty}\sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j)\Big| \leq C\int_{k+1}^{+\infty} j^{-d-1+\delta}\left[\int_j^{+\infty} l^{-d-1+\delta}(l-j)^{2d-1+\delta}\mathrm{d}l\right]\mathrm{d}j + \mathrm{O}\left(k^{-2d-1+2\delta}\right)$$

In the integral over l by using the substitution $jl' = l$, we obtain:

$$\left|\sum_{j=k+1}^{+\infty}\sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j)\right| \leq C' \int_{k+1}^{+\infty} j^{-2+3\delta} \int_{1}^{+\infty} l^{-d-1+\delta}(l-1)^{2d-1+\delta} \mathrm{d}l \mathrm{d}j + \mathrm{O}\left(k^{-2d-1}\right).$$

Since if $\delta < (1-d)/2$

$$\int_{1}^{+\infty} l^{-d-1+\delta}(l-1)^{2d-1+\delta} \mathrm{d}l < +\infty,$$

it follows:

$$\begin{aligned}\left|\sum_{j=k+1}^{+\infty}\sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j)\right| &\leq \mathrm{O}\left(k^{-1+3\delta}\right) + \mathrm{O}\left(k^{-2d-1}\right) \\ &\leq \mathrm{O}\left(k^{-1+3\delta}\right). \end{aligned} \tag{11}$$

If $\delta > 0$, $\delta < 1-2d$ and $\delta < (1-d)/2$, we have:

$$\left|\sum_{j=k+1}^{+\infty}\sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j)\right| = \mathrm{O}\left(k^{-1+3\delta}\right).$$

Notice that if the equality is true under the assumptions $\delta > 0$, $\delta < 1-2d$ and $\delta < (1-d)/2$, it is also true for any $\delta > 0$. Therefore we have proven the first part of the theorem.
We prove now that there exists long-memory processes whose prediction error attains the rate of convergence k^{-1}. Assume now that $(X_n)_{n\in\mathbb{Z}}$ is fractionally integrated noise $\mathrm{F}(d)$, which is the stationary solution of the difference equation:

$$X_n = (1-B)^{-d}\varepsilon_n \tag{12}$$

with B the usual backward shift operator, $(\varepsilon_n)_{n\in\mathbb{Z}}$ is a white-noise series and $d \in]0, 1/2[$ (see for example Brockwell and Davis (1991)). We can compute the coefficients and obtain that:

$$\forall j > 0, \quad a_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \text{ and } \forall j \geq 0, \quad \sigma(j) = \frac{(-1)^j \Gamma(1-2d)}{\Gamma(j-d+1)\Gamma(1-j-d)}\sigma_\varepsilon^2$$

then we have:

$$\forall j > 0, \quad a_j < 0 \text{ and } \forall j \geq 0, \quad \sigma(j) > 0$$

and

$$a_j \sim \frac{j^{-d-1}}{\Gamma(-d)} \text{ and } \sigma(j) \sim \frac{j^{2d-1}\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \quad \text{when } j \to \infty.$$

In this particular case, we can estimate the prediction error more precisely:

$$\begin{aligned}\sum_{k+1}^{+\infty}\sum_{k+1}^{+\infty} a_j a_l \sigma(l-j) &= \sum_{k+1}^{+\infty} |a_j| \sum_{j+1}^{+\infty} |a_l||\sigma(l-j)| + \sum_{k+1}^{+\infty} a_j^2 \sigma(0) \\ &\sim \frac{\Gamma(1-2d)}{\Gamma(-d)^2\Gamma(d)\Gamma(1-d)} \int_{k+1}^{+\infty} j^{-2} \int_{1/j+1}^{+\infty} l^{-d-1}(l-1)^{2d-1} \mathrm{d}l \mathrm{d}j + \mathrm{O}\left(k^{-2d-1}\right) \\ \sum_{k+1}^{+\infty}\sum_{k+1}^{+\infty} a_j a_l \sigma(l-j) &\sim \frac{\Gamma(1-2d)\Gamma(2d)}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)} k^{-1} \end{aligned} \tag{13}$$

The asymptotic bound $\mathrm{O}(k^{-1})$ is therefore as small as possible. □

In the specific case of fractionally integrated noise, we may write the prediction error as:

$$\mathbb{E}\big(\big[X_{k+1} - \widetilde{X_k'}(1)\big]^2\big) = {\sigma_\varepsilon}^2 + C(d)k^{-1} + \mathrm{o}\left(k^{-1}\right)$$

and we can express $C(d)$ as a function of d:

$$C(d) = \frac{\Gamma(1-2d)\Gamma(2d)}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)}. \tag{14}$$

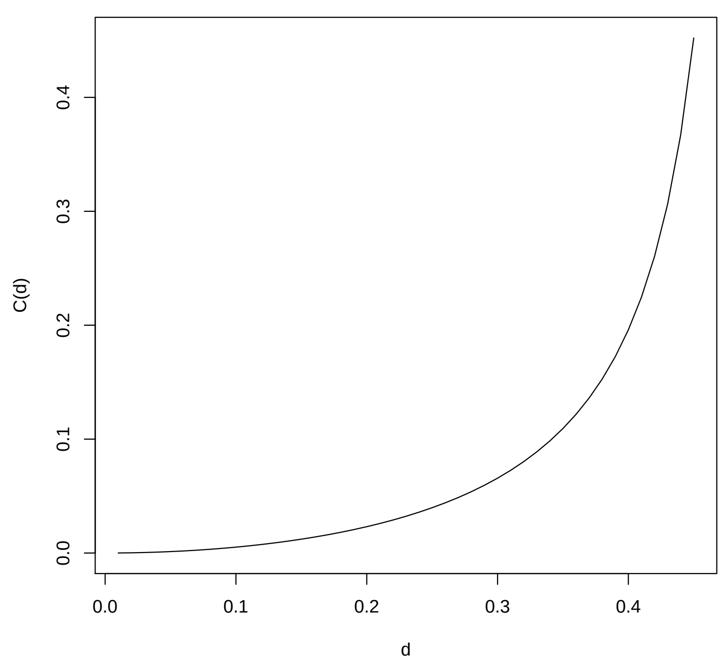
It is easy to prove that $C(d) \to +\infty$ as $d \to 1/2$ and we may write the following asymptotic equivalent as $d \to 1/2$:

$$C(d) \sim \frac{1}{(1-2d)\Gamma(-1/2)^2\Gamma(1/2)\Gamma(3/2)}. \tag{15}$$

As $d \to 0$, $C(d) \to 0$ and we have the following equivalent as $d \to 0$:

$$C(d) \sim d^2.$$

Figure 2.1: The Constant $C(d)$, $d \in [0, 1/2[$, defined in (14)



As the figure 2.1 suggests and the asymptotic equivalent given in (15) proves, the mean-squared error tends to $+\infty$ as $d \to 1/2$. By contrast, the constant $C(d)$ takes small values for d in a large interval of $[0, 1/2[$.Although the rate of convergence has a constant order k^{-1}, the forecast error is bigger when $d \to 1/2$. This result is not surprising since the correlation between the random variable, which we want to predict, and the random variables, which we take equal to 0, increases when $d \to 1/2$.

2.2 Estimates of Forecast Coefficients and the Associated Mean Square Error

We will now estimate the mean-squared error between the predictor $\widetilde{X'_k}(1)$ defined on (7) and the predictor $\widetilde{X'_{T,k}}(1)$ defined as:

$$\widetilde{X'_{T,k}}(1) := -\sum_{j=1}^{k} \widehat{a}_j X_{k+1-j}$$

where $\widehat{a}_j$ are estimates of a_j computed using a length T realisation of the process. More precisely, we consider a parametric approach and we assume that:

$$a_j = a_j(\theta) \text{ with } \theta \text{ an unknown vector in } \Theta$$

where Θ is a compact subset of $\mathbb{R}^p$. Assume that the process $(Y_n)_{n\in\mathbb{Z}}$ is Gaussian. Let θ_0 be the true value of the parameter. We assume the realisation $(Y_n)_{1\leq n\leq T}$ to be known. We estimate the $(a_j)_{1\leq j\leq k}$ by $\widehat{a}_j := a_j(\widehat{\theta}_T)$ where $\widehat{\theta}_T$ is an estimate of θ_0, for example the Whittle estimate. In order to use the Whittle estimate and follow the approach suggested in Fox and Taqqu (1986), we assume from now on that all the processes in the parametric class have a spectral density denoted by $f(.,\theta)$.

We define the Whittle estimate by (see Fox and Taqqu (1986)):

$$\hat{\theta}_T = \underset{\theta\in\Theta}{\operatorname{argmin}} \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} [f(\lambda,\theta)]^{-1} I_T(\lambda)\mathrm{d}\lambda\right] \tag{16}$$

where I_T is the periodogram:

$$I_T(\lambda) = \frac{|\sum_{j=1}^{T} \mathrm{e}^{ij\lambda}(Y_j - \overline{Y_T})|^2}{2\pi T}.$$

Before we state the theorem, we will give assumptions on the regularity of the spectral densities in our parametric class. Under those standard conditions, the estimated vector converges to the true parameter if the process is a Gaussian long-memory time series (see Fox and Taqqu (1986)). We will refer to the following assumptions.

We say that $f(x,\theta)$ satisfies conditions A0-A6 if there exists $0 < \alpha(\theta) < 1$ such that for each $\delta > 0$,

A0. $f(\lambda,\theta_0) = |\lambda|^{2\alpha(\theta_0)} L(\lambda,\theta_0)$ with $L(.,\theta_0)$ bounded. $L(.,\theta_0)$ is differentiable at 0 and $L(.,\theta_0) \neq 0$.

A1. $\theta \mapsto \int_{-\pi}^{\pi} f(\theta,\lambda)\mathrm{d}\lambda < +\infty$ can be twice differentiated under the integral sign.

A2. $f(\theta,\lambda)$ is continuous at all (θ,λ), $\lambda\neq 0$, $f^{-1}(\theta,\lambda)$ is continuous at all (θ,λ) and,

$$f(\theta,\lambda) = \mathrm{O}(|\lambda|^{-\alpha(\theta)-\delta}) \quad \text{as } \lambda \to 0.$$

A3. $(\partial/\partial\theta_j) f^{-1}(\theta,\lambda)$ and $(\partial^2/\partial\theta_j\partial\theta_l) f^{-1}(\theta,\lambda)$ are continuous at all (θ,λ),

$$\forall 1\leq j\leq p, \quad \frac{\partial}{\partial\theta_j} f^{-1}(\theta,\lambda) = \mathrm{O}(|\lambda|^{\alpha(\theta)-\delta}) \quad \text{as } \lambda\to 0$$

and

$$\forall 1\leq j,l\leq p \quad \frac{\partial^2}{\partial\theta_j\partial\theta_l} f^{-1}(\theta,\lambda) = \mathrm{O}(|\lambda|^{\alpha(\theta)-\delta}) \quad \text{as } \lambda\to 0.$$

A4. $(\partial/\partial\lambda)f(\theta,\lambda)$ is continuous at all (θ,λ), $\lambda \neq 0$, and

$$\frac{\partial}{\partial\lambda}f(\theta,\lambda) = \mathrm{O}(|\lambda|^{-\alpha(\theta)-1-\delta}) \quad \text{as } \lambda \to 0$$

A5. $(\partial^2/\partial\theta_j\partial\lambda)f^{-1}(\theta,\lambda)$ are continuous at all (θ,λ), $\lambda \neq 0$, and

$$\forall 1 \leq j \leq p, \quad \frac{\partial^2}{\partial\theta_j\partial\lambda}f^{-1}(\theta,\lambda) = \mathrm{O}(|\lambda|^{\alpha(\theta)-1-\delta}) \quad \text{as } \lambda \to 0.$$

A6. $(\partial^3/\partial\theta_j\partial^2\lambda)f^{-1}(\theta,\lambda)$ are continuous at all (θ,λ), $\lambda \neq 0$, and

$$\forall 1 \leq j \leq p, \quad \frac{\partial^3}{\partial\theta_j\partial^2\lambda}f^{-1}(\theta,\lambda) = \mathrm{O}(|\lambda|^{\alpha(\theta)-2-\delta})$$

We can now express the asymptotic behavior of the mean-squared prediction error due to the estimation of the forecast coefficients. We assume in this Section that the process is Gaussian. Let $(X_j)_{j\in\mathbb{Z}}$ be a stochastic process, which verifies the assumptions of section 1, and let $(Y_j)_{j\in\mathbb{Z}}$ be a process which is independent of $(X_j)_{j\in\mathbb{Z}}$, but has the same stochastic structure. We want to predict X_{k+1} knowing $(X_j)_{j\in[\![1,k]\!]}$ and we assume that the parameter θ and so the forecast coefficients are estimated based on a realisation $(Y_j)_{j\in[\![1,T]\!]}$.

Theorem 2.2.1. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary Gaussian long-memory sequence with mean 0 and spectral density $f(\theta,\lambda)$ and $\theta \in \Theta$ is an unknown parameter. The set Θ is assumed to be compact. We assume also that θ_0 is in the interior of Θ and that $\forall\theta \in \mathring{\Theta}$, the conditions A1-A6 hold. Moreover we assume that each process $(Z_n)_{n\in\mathbb{Z}}$ in our parametric class with $\theta \in \Theta$ admits an autoregressive representation:*

$$\varepsilon_n = \sum_{j=0}^{\infty} a_j(\theta) Z_{n-j}$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ is a Gaussian white noise. Let θ_0 be the true value of the parameter and assume that $\theta_0 \in \mathring{\Theta}$. Assume also that $f(\theta_0,\lambda)$ verifies A0 and that for any $j \in \mathbb{N}$, a_j verifies:

(i) a_j is uniformly bounded on a neighbourhood of θ_0;

(ii) the first and second derivatives of a_j are continuous and bounded on a neighbourhood of θ_0.

and that:

$$\forall\delta > 0, \exists C_l, \forall j \in \mathbb{N}^* \quad \left|\frac{\partial a_j}{\partial\theta_l}(\theta_0)\right| \leq C_l j^{-1+\delta}. \tag{17}$$

We have then the following result:

$$\mathbb{E}\left(\widetilde{X'_{T,k}}(1) - \widetilde{X'_k}(1)\right)^2 = \mathrm{O}\left(\frac{k^{2d}}{T}\right).$$

An example to which our theorem applies is the fractionally integrated processes. In this case, the parameter θ is scalar and corresponds to the long-memory parameter d. Assumptions A0-A6 hold for fractionally processes. We define d_0 by $d_0 := \theta_0$ and then we have a_j:

$$a_j(d) := \left(\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}\right)$$

Since the gamma function Γ is analytic on $\{\mathbb{C} \setminus \mathbb{N}\}$, there exists a neighbourhood of d_0 on which the function a_j and its first and second derivatives are bounded. Finally when $j \to +\infty$,

$$\left| \frac{\partial a_j}{\partial d}(d_0) \right| \sim C j^{-1}$$

where C is a constant. As a consequence our Theroem can be applied on the class of fractionally integrated noise because all the assumptions hold. Similarly we can also show that the class of FARIMA time series verify the assumptions of this Theorem.

Proof. We first define the following vector:

$$\alpha_k^* := \left(a_1(\widehat{\theta}_T) - a_1(\theta_0), \ldots, a_k(\widehat{\theta}_T) - a_k(\theta_0) \right)$$

where v^* is the transpose vector of v and

$$\left(\mathbf{X}_1^k \right)^* := (X_k, \ldots, X_1).$$

$$\begin{aligned}
\mathbb{E}\left(\widetilde{X'_{T,k}}(1) - \widetilde{X'_k}(1) \right)^2 &= \mathbb{E}\left(\left(a_1(\widehat{\theta}_T) - a_1, \ldots, a_k(\widehat{\theta}_T) - a_k \right) \begin{pmatrix} X_k \\ \vdots \\ X_1 \end{pmatrix} \right)^2 \\
&= \mathbb{E}\left[\left(\alpha_k^* \mathbf{X}_1^k \right)^2 \right] \\
&= \operatorname{trace}\left(\mathbb{E}\left(\alpha_k^* \mathbf{X}_1^k \right)^2 \right) \\
&= \mathbb{E}\left(\operatorname{trace}\left(\alpha_k \alpha_k^* \mathbf{X}_1^k \left(\mathbf{X}_1^k \right)^* \right) \right) \\
&= \operatorname{trace}\left(\mathbb{E}\left(\alpha_k \alpha_k^* \right) \Sigma_k \right)
\end{aligned}$$

with

$$\Sigma_k := \mathbb{E}\left(\mathbf{X}_1^k \left(\mathbf{X}_1^k \right)^* \right)$$

Let us first study the covariance matrix of the estimated coefficients $\mathbb{E}(\alpha_k \alpha_k^*)$. We can write $(\alpha_k \alpha_k^*)_{i,j} = \mathbb{E}\left(g_{i,j}(\hat{\theta}) \right)$ when $g_{i,j}$ is defined by $g_{i,j} : \theta \mapsto (a_i(\theta) - a_i(\theta_0))(a_j(\theta) - a_j(\theta_0))$. We then use an order 2 Taylor series expansion of $g_{i,j}$ and apply Theorem 5.4.3 form Fuller (1976).We will refer to the following version.
If the following assumptions hold

(i) $\forall m \in [\![1, p]\!], \quad \mathbb{E}\left(|\widehat{\theta}_{T,m} - \theta_{0,m}|^3 \right) = \mathrm{O}(\eta(T))$ where $\widehat{\theta}_{T,m}$ is the m^{th} entry of $\widehat{\theta}_T$;

(ii) $\widehat{\theta}_T \to \theta_0, \quad \mathbb{P}-$a.s.;

(iii) $g_{i,j}$ is uniformly bounded on a neighbourhood of θ_0;

(iv) the first and the second derivatives $g_{i,j}$ are continuous and bounded on a neighbourhood of θ_0

then

$$\begin{aligned}\mathbb{E}\left(g_{i,j}(\widehat{\theta}_T)\right) &= g_{i,j}(\theta_0)+\sum_{l=1}^{p}\mathbb{E}(\widehat{\theta}_{T,l}-\theta_{0,l})\frac{\partial g_{i,j}}{\partial\theta_l}(\theta_0)\\ &+\frac{1}{2}\sum_{l=1}^{p}\sum_{n=1}^{p}\frac{\partial^2 g_{i,j}}{\partial\theta_l\partial\theta_n}(\theta_0)\mathbb{E}\left(\left(\widehat{\theta}_{T,l}-\theta_{0,l}\right)\left(\widehat{\theta}_{T,n}-\theta_{0,n}\right)\right)+\mathrm{O}(\eta(T)).\end{aligned}$$

By assumption, conditions (ii) et (iv) hold. We note also that:

$$g_{i,j}(\theta_0)=0 \quad \text{et} \quad \forall l\in[\![1,p]\!], \frac{\partial g_{i,j}}{\partial\theta_l}(\theta_0)=0$$

Next we compute the fourth order moments of $\widehat{\theta}_T-\theta_0$ in order to estimate the second and the third moments. We define:

$$\begin{aligned}\sigma_T(\theta) &:= \left[\frac{1}{2\pi}\int_{-\pi}^{\pi}[f(\lambda,\theta)]^{-1}I_T(\lambda)\mathrm{d}\lambda\right]\\ &= \frac{\mathbf{Y}'A_T(\theta)\mathbf{Y}}{T}\end{aligned}$$

where $(\mathbf{Y})^*=(Y_1,\dots,Y_T)$ and

$$(A_T(\theta))_{j,l}:=\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\mathrm{e}^{i(j-l)\lambda}[f(\lambda,\theta)]^{-1}\mathrm{d}\lambda.$$

We follows now the proof of Fox and Taqqu (1986). Since $\widehat{\theta}_T=\underset{\theta}{\operatorname{argmin}}\{\sigma_T(\theta)\}$ and according to the mean-value theorem, we have:

$$\exists\theta^* \text{ such that } |\theta^*-\theta_0|\le|\hat{\theta}-\theta_0| \text{ and } \quad \hat{\theta}-\theta_0=-\left[\left(\frac{\partial^2}{\partial\theta_i\theta_j}\sigma_T(\theta^*)\right)_{1\le i,j\le p}\right]^{-1}\frac{\partial}{\partial\theta}\sigma_T(\theta_0).$$

It is justified because $\theta\mapsto[f(\lambda,\theta)]^{-1}$ is twice differentiable with respect to θ and all the partial derivatives are integrable on $[-\pi,\pi]$ with respect to λ by assumption A3. It follows from Fox and Taqqu (1986) that:

$$\frac{\partial^2}{\partial\theta_i\theta_j}\sigma_T(\theta^*)\overset{\mathbb{P}-\text{a.s.}}{\longrightarrow}\frac{\sigma_\varepsilon^2}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial}{\partial\theta_i}f^{-1}(\lambda,\theta_0)\right)\left(\frac{\partial}{\partial\theta_j}f^{-1}(\lambda,\theta_0)\right)f^2(\lambda,\theta_0)\mathrm{d}\lambda:=W_{i,j} \tag{18}$$

where $W:=(W_{i,j})_{1\le i,j\le p}$ is a positive definite matrix. Since the matrix norm $x\mapsto\|x\|_4$ is continuous, there exists $C>0$ such that $\|W\|_4>C$ and:

$$\exists M\in\mathbb{N},\quad T>M \quad ,\left\|\widehat{\theta}_k-\theta_0\right\|_4 \ \le\ C\left\|\frac{\partial}{\partial\theta}\sigma_T(\theta_0)\right\|_4 \quad \mathbb{P}-\text{a.s.} \tag{19}$$

Using this inequality, we can now estimate the fourth moments for any $m\in[\![1,p]\!]$:

$$\mathbb{E}\left[\left|\frac{\partial}{\partial\theta_m}\sigma_T(\theta_0)\right|^4\right] = \mathbb{E}\left[\left(\frac{\mathbf{Y}'\frac{\partial A_T(\theta_0)}{\partial\theta_m}\mathbf{Y}}{T}\right)^4\right].$$

Let $m \in [\![1,p]\!]$. We define the matrix Δ_m with (j,l)-th entries:

$$\delta_{j,l} := \int_{-\pi}^{\pi} \mathrm{e}^{i(j-l)\lambda} \frac{\partial}{\partial \theta_m} f^{-1}(\lambda, \theta_0) \mathrm{d}\lambda.$$

Next we rewrite this expression as:

$$\begin{aligned} \mathbb{E}\left[\left(\frac{\partial}{\partial \theta_m}\sigma_T(\theta_0)\right)^4\right] &= T^{-4}\mathbb{E}\left(\left(\sum_{j=1}^{T}\sum_{l=1}^{T} Y_j Y_l \delta_{j,l}\right)^4\right) \\ &= T^{-4} \sum_{j_1,j_2,\ldots,j_8=1}^{T} \delta_{j_1,j_2}\delta_{j_3,j_4}\delta_{j_5,j_6}\delta_{j_7,j_8}\mathbb{E}\left(Y_{j_1}Y_{j_2}\ldots Y_{j_8}\right) \end{aligned} \tag{20}$$

The process is Gaussian then all the moments are a function of the autocovariances (see Triantafyllopoulos (2003)). In equation (20), we can rewrite each fourth moment in the sum as a linear combination of product of 4 covariances. We then count how many covariances belongs to the set $S = \{\mathbb{E}\left(Y_{j_1}Y_{j_2}\right), \mathbb{E}\left(Y_{j_3}Y_{j_4}\right), \mathbb{E}\left(Y_{j_5}Y_{j_6}\right), \mathbb{E}\left(Y_{j_7}Y_{j_8}\right)\}$:

1. either we have $\mathbb{E}(Y_{j_1}Y_{j_2}) \times C$ and we can distinguish the following possibilities:
   - $C = \mathbb{E}(Y_{j_3}Y_{j_4})\mathbb{E}\left(Y_{j_5}Y_{j_6}\right)\mathbb{E}\left(Y_{j_7}Y_{j_8}\right)$ only one possibility or;
   - C has one element in S and no other which makes 6 possibilities =(3 choices in S)×(2 choices for the other covariances) or;
   - C has no elements in S which makes 8 possibilities. First choose a complement for Y_{j_3} (4 possibilities) then a complement for Y_{j_4} (only 2 possibilities because the pairs in S are excluded);
2. or Y_{j_1} is with Y_{j_l}, $l > 2$, which makes 5 possibilities. Let us assume that Y_{j_1} is associated with Y_{j_3}. We can then distinguish the following cases:
   - we obtain 2 pairs in S which are consequently (Y_{j_5}, Y_{j_6}) and (Y_{j_7}, Y_{j_8}) or;
   - we have only one couple in S which makes (2 choices in S) $\times((\mathrm{C}_4^2 - 1)$ choices for the other covariances) or;
   - we have non elements in S: either Y_{j_2} is the complement of Y_{j_4} and then we have only 2 possibilities, or we have 4 choices for the complement of Y_{j_2} and only 2 for Y_{j_4}. Finally we have 10 possibilities.

Therefore we obtain:

$$
\begin{aligned}
& \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_m}\sigma_T(\theta_0)\right)^4\right] \\
= \quad & T^{-4}\left(\sum_{j=1}^{T}\sum_{l=1}^{T}\delta_{j,l}\sigma(j-l)\right)^4 \\
+ \quad & T^{-4}6\left(\sum_{j=1}^{T}\sum_{l=1}^{T}\delta_{j,l}\sigma(j-l)\right)^2 \sum_{j_1,j_2,j_3,j_4=1}^{T}\delta_{j_1,j_2}\delta_{j_3,j_4}\sigma(j_1-j_3)\sigma(j_2-j_4) \\
+ \quad & T^{-4}8\left(\sum_{j=1}^{T}\sum_{l=1}^{T}\delta_{j,l}\sigma(j-l)\right) \sum_{j_1,j_2,j_3,j_4,j_5,j_6=1}^{T}\delta_{j_1,j_2}\delta_{j_3,j_4}\delta_{j_5,j_6}\sigma(j_1-j_3)\sigma(j_4-j_5)\sigma(j_6-j_2) \\
+ \quad & 5T^{-4}\left(\sum_{j=1}^{T}\sum_{l=1}^{T}\delta_{j,l}\sigma(j-l)\right)^2 \sum_{j_1,j_2,j_3,j_4=1}^{T}\delta_{j_1,j_2}\delta_{j_3,j_4}\sigma(j_1-j_3)\sigma(j_2-j_4) \\
+ \quad & 5T^{-4}10\left(\sum_{j=1}^{T}\sum_{l=1}^{T}\delta_{j,l}\sigma(j-l)\right) \sum_{j_1,j_2,j_3,j_4,j_5,j_6=1}^{T}\delta_{j_1,j_2}\delta_{j_3,j_4}\delta_{j_5,j_6}\sigma(j_1-j_3)\sigma(j_4-j_5)\sigma(j_6-j_2) \\
+ \quad & 5T^{-4}10 \sum_{j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8=1}^{T}\delta_{j_1,j_2}\delta_{j_3,j_4}\delta_{j_5,j_6}\delta_{j_7,j_8}\sigma(j_1-j_3)\sigma(j_4-j_5)\sigma(j_6-j_7)\sigma(j_8-j_2)
\end{aligned}
$$

All the terms of this sum are like:

$$
\sum_{j_1,\ldots,j_{2p}=1}^{T}\delta_{j_1,j_2}\ldots\delta_{j_{2p-1},j_{2p}}\sigma(j_1-j_3)\ldots\sigma(j_{2p}-j_2) := S_{p,T}. \tag{21}
$$

Note that $S_{p,T} = \operatorname{trace}\left((\Sigma_T\Delta_m)^p\right)$ and that Δ_m is the covariance matrix defined by the spectral density:

$$
\frac{\partial}{\partial\theta_m}f^{-1}(\lambda,\theta_0) = \mathrm{O}\left(\lambda^{\alpha(\theta_0)-\delta}\right) \quad \text{as} \quad \lambda\to 0, \quad \text{for any } \delta>0
$$

by assumption A3. By applying the Theorem 1 of Fox and Taqqu (1987), we prove that:

$$
\begin{aligned}
& \lim_{T\to+\infty}\frac{1}{T}\sum_{j_1,\ldots,j_{2p}=1}^{T}\delta_{j_1,j_2}\ldots\delta_{j_{2p-1},j_{2p}}\sigma(j_2-j_3)\ldots\sigma(j_{2p}-j_1) \\
= \quad & (2\pi)^{2p-1}\int_{-\pi}^{\pi}\left(\frac{\partial}{\partial\theta_m}f^{-1}(\lambda,\theta_0)f(\lambda,\theta_0)\right)^p \mathrm{d}\lambda \\
= \quad & \mathrm{O}(1).
\end{aligned} \tag{22}
$$

It follows from assumptions A2 and A3 that this integral is always finite. We need a more precise result for the term:

$$
\sum_{j=1}^{T}\sum_{l=1}^{T}\delta_{j,l}\sigma(j-l).
$$

Despite this can also be expressed like (21), the estimate given below is not sufficient to conclude. By Fox and Taqqu (1986)[proof of Theorem 2], we have:

$$\forall \delta > 0, \quad \sum_{j=1}^{T}\sum_{l=1}^{T} \delta_{j,l}\sigma(j-l) = \mathrm{O}\left(T^{\delta}\right). \tag{23}$$

By (22) et (23), we may conclude that

$$\forall \delta > 0, \quad \mathbb{E}\left[\left(\frac{\partial}{\partial \theta_m}\sigma_T(\theta_0)\right)^4\right] = \mathrm{O}(T^{-3+\delta}). \tag{24}$$

Next using the asymptotic estimate of the fourth moments, we can now obtain asymptotic properties for the second moments:

$$\mathbb{E}\left[\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right)\right].$$

First we have to prove the uniform integrability of $\sqrt{T}\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)\sqrt{T}\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right)$:

$$\begin{aligned}
T^2\mathbb{E}\left(\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)^2\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right)^2\right) &\leq T^2\sqrt{\mathbb{E}\left(\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)^4\right)\mathbb{E}\left(\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right)^4\right)} \\
&\leq T^2\mathbb{E}\left(\left\|\widehat{\theta}_T - \theta_0\right\|_4^4\right) \\
&\leq T^2C^4\mathbb{E}\left(\left\|\frac{\partial}{\partial \theta}\sigma_T(\theta_0)\right\|_4^4\right)
\end{aligned}$$

if $T > M$ from (19). By applying result (24), we conclude that:

$$\begin{aligned}
\forall \delta > 0, \quad T^2\mathbb{E}\left(\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)^2\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right)^2\right) &= \mathrm{O}\left(T^{2-3+\delta}\right) \\
&= \mathrm{O}(1)
\end{aligned}$$

We have proved the uniform integrability of $\sqrt{T}\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)\sqrt{T}\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right)$ since if $\mathbb{E}\left(X_T^2\right)$ is finite for any T,then the collection (X_T) is uniformly integrable. Moreover according to Fox and Taqqu (1986) [Theorem 2]:

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\pi W^{-1}\right)$$

where W is the matrix defined in (18) and we have also the following convergence in law:

$$h_{j,l}\left(\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right)\right) := T\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right) \xrightarrow{\mathcal{L}} h_{j,l}(Z)$$

with $Z \sim \mathcal{N}\left(0, 4\pi W^{-1}\right)$. By the convergence in law and the uniform integrability we apply Theorem 5.4 in Billingsley (1968) that:

$$\mathbb{E}\left(T\left(\widehat{\theta}_{T,j} - \theta_{0,j}\right)\left(\widehat{\theta}_{T,l} - \theta_{0,l}\right)\right) \to 4\pi W_{j,l}^{-1} \text{ as } T \to +\infty.$$

Now we give an asymptotic bound for the third order moment by applying the Cauchy-Schwarz inequality. Using the inequalities (19) and (24), we conclude that:

$$\begin{aligned}\exists C>0,\ \mathbb{E}\left(\left|\widehat{\theta}_{T,j}-\theta_{0,j}\right|^3\right) &\leq \sqrt{\mathbb{E}\left(\widehat{\theta}_{T,j}-\theta_{0,j}\right)^2\mathbb{E}\left(\widehat{\theta}_{T,j}-\theta_{0,j}\right)^4}\\ &\leq \sqrt{CT^{-1}T^{-3+\delta}},\quad \forall\delta>0\\ &= \mathrm{O}\left(T^{-2+\delta}\right),\quad \forall\delta>0\end{aligned}$$

We obtain the following Taylor series for any $\delta>0$:

$$\begin{aligned}\mathbb{E}\left((a_i(\widehat{\theta}_T)-a_i(\theta_0))(a_j(\widehat{\theta}_T)-a_j(\theta_0))\right) &= 2\pi T^{-1}\sum_{l=1}^m\sum_{n=1}^m\frac{\partial^2 g_{i,j}}{\partial\theta_l\partial\theta_n}(\theta_0)W^{-1}_{l,n}+\mathrm{O}\left(T^{-2+\delta}\right)\\ &\sim 2\pi T^{-1}\sum_{l=1}^m\sum_{n=1}^m\left[\frac{\partial a_i(\theta_0)}{\partial\theta_l}\frac{\partial a_j(\theta_0)}{\partial\theta_n}+\frac{\partial a_j(\theta_0)}{\partial\theta_l}\frac{\partial a_i(\theta_0)}{\partial\theta_n}\right]W^{-1}_{l,n}.\end{aligned}$$

We can now conclude and find an asymptotic equivalent of $\mathbb{E}\left(\alpha_k\alpha_k^*\right)$. Since W^{-1} is symmetric:

$$\mathbb{E}\left(\alpha_k\alpha_k^*\right)\sim 4\pi T^{-1}DW^{-1}D^* \tag{25}$$

with

$$D:=\begin{pmatrix}\frac{\partial a_1(\theta_0)}{\partial\theta_1} & \dots & \frac{\partial a_1(\theta_0)}{\partial\theta_m}\\ \vdots & \ddots & \vdots\\ \frac{\partial a_k(\theta_0)}{\partial\theta_1} & \dots & \frac{\partial a_k(\theta_0)}{\partial\theta_m}\end{pmatrix}.$$

W^{-1} is a positive definite matrix because W is too. So it can be expressed as :

$$W^{-1}=P^*\begin{pmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \vdots\\ 0 & \dots & 0 & \lambda_m\end{pmatrix}P$$

where $P=(p_{ij})_{1\leq i,j\leq m}$ is an orthogonal matrix and the $(\lambda_i)_{1\leq i\leq m}$ are the positive eigenvalues of W^{-1}. We may rewrite our expression as:

$$\begin{aligned}DW^{-1}D^* &= \sum_{r=1}^m\left(\left[\sqrt{\lambda_r}\sum_{l=1}^m p_{rl}\frac{\partial a_i}{\partial\theta_l}(\theta_0)\right]\left[\sqrt{\lambda_r}\sum_{l=1}^m p_{rl}\frac{\partial a_j}{\partial\theta_l}(\theta_0)\right]\right)_{1\leq i,j\leq k}\\ &= \sum_{r=1}^m\beta_r^*\beta_r \end{aligned}\tag{26}$$

where β_r is the vector $\left(\sqrt{\lambda_r}\sum_{l=1}^m p_{rl}\frac{\partial a_i}{\partial\theta_l}(\theta_0)\right)_{1\leq i\leq k}$.

$$\begin{aligned}\mathbb{E}\left(\left(a_1(\widehat{\theta}_T)-a_1(\theta_0),\dots,a_k(\widehat{\theta}_T)-a_k(\theta_0)\right)\begin{pmatrix}X_0\\ \vdots\\ X_{-k+1}\end{pmatrix}\right)^2 &= \mathrm{trace}\left(\mathbb{E}\left(\alpha_k\alpha_k^*\right)\Sigma_k\right)\\ &\sim 4\pi T^{-1}\sum_{r=1}^m\mathrm{trace}\left(\beta_r^*\beta_r\Sigma_k\right)\end{aligned}$$

from (25) and (26). Therefore we obtain:

$$\begin{aligned}
\operatorname{trace}\left(\mathbb{E}\left(\alpha_k \alpha_k^*\right) \Sigma_k\right) &\sim 4\pi T^{-1} \sum_{r=1}^{m} \beta_r \Sigma_k \beta_r^* \\
&\leq 4\pi T^{-1} \sum_{r=1}^{m} \Lambda_k \|\beta_r\|_2^2
\end{aligned}$$

where Λ_k is the greatest eigenvalue of Σ_k. The last inequality is a consequence of Σ_k being symmetric matrix. Following assumption (17), we have:

$$\forall \delta > 0, \exists C_l, \forall j \in \mathbb{N}^* \quad \left|\frac{\partial a_j}{\partial \theta_l}(\theta_0)\right| \leq C_l j^{-1+\delta}$$

and we can hence estimate $\|\beta_r\|_2^2$. Let $\delta = 1/2$, there exists $C_1, \ldots, C_m$ such that:

$$\begin{aligned}
\|\beta_r\|_2^2 &= \sum_{j=1}^{k} \lambda_r \sum_{l_1=1}^{m} \sum_{l_2=1}^{m} p_{rl_1} p_{rl_2} \frac{\partial a_j}{\partial \theta_{l_1}}(\theta_0) \frac{\partial a_j}{\partial \theta_{l_2}}(\theta_0) \\
&\leq \lambda_r \sum_{l_1=1}^{m} \sum_{l_2=1}^{m} |p_{rl_1} p_{rl_2}| \sum_{j=1}^{k} C_{l_1} C_{l_2} j^{-3} \\
&\leq \lambda_r \sum_{l_1=1}^{m} \sum_{l_2=1}^{m} |p_{rl_1} p_{rl_2}| \sum_{j=1}^{+\infty} C_{l_1} C_{l_2} j^{-3} := C_r(\theta_0)
\end{aligned}$$

where $C_r(\theta_0)$ does not depend on k.

From Boettcher and Virtanen (2006), the spectral norm of a Toeplitz matrix (its spectral norm), whose symbol has the form $\lambda \mapsto \lambda^{-\alpha} L(\lambda)$ with L is a bounded, continuous at 0 function and does not vanish at 0, is equivalent to Ck^{α} with C constant. We conclude the proof:

$$\mathbb{E}\left(\left(a_1(\widehat{\theta}_T) - a_1(\theta_0), \ldots, a_k(\widehat{\theta}_T) - a_k(\theta_0)\right) \begin{pmatrix} X_0 \\ \vdots \\ X_{-k+1} \end{pmatrix}\right)^2 \leq C 4\pi \sum_{r=1}^{m} C_r(\theta_0) \frac{k^{\alpha(\theta_0)}}{T}$$

with C constant. □

2.3 Conclusion

Prediction with the Wiener-Kolmogorov predictor involves two mean-squared error components: the first is due to the truncation to k terms and this is bounded by $\mathrm{O}(k^{-1})$, the second is due to the estimation of the coefficients a_j from a realisation of the process of length T and is bounded by $\mathrm{O}(k^{2d}/T)$. The mean-squared difference between the best linear predictor $\widetilde{X_k}(1)$ and our predictor is given by:

$$\begin{aligned}
\left\|\widetilde{\widetilde{X_{T,k}}}(1) - \widetilde{X_k}(1)\right\|_2 &\leq \left\|\widetilde{\widetilde{X_{T,k}}}(1) - \widetilde{X'_k}(1)\right\|_2 + \left\|\widetilde{X'_k}(1) - \widetilde{X_k}(1)\right\|_2 \\
&\leq \mathrm{O}(k^{-1/2}) + \mathrm{O}(k^d/\sqrt{T}).
\end{aligned}$$

If we want to compare the two types of prediction errors, we need a relation between the rate of convergence of T and k to $+\infty$. For example, if $T = \mathrm{o}(k^{2d+1})$, the error due to the estimation of the coefficients is predominant and gives the bound for the general error.

Truncating to k terms the series which defines the Wiener-Kolmogorov predictor amounts to using an AR(k) model for predicting. Therefore in the following section we look for the AR(k) which minimizes the forecast error.

3 The Autoregressive Models Fitting Approach

In this section we shall develop a generalisation of the "autoregressive model fitting" approach developed by Ray (1993) in the case of fractionally integrated noise F(d) (defined in (12)). We study asymptotic properties of the forecast mean-squared error when we fit a misspecified AR(k) model to the long-memory time series $(X_n)_{n\in\mathbb{Z}}$.

3.1 Rationale

Let Φ a k^{th} degree polynomial defined by:

$$\Phi(z) = 1 - a_{1,k}z - \ldots - a_{k,k}z^k.$$

We assume that Φ has no zeroes on the unit disk. We define the process $(\eta_n)_{n\in\mathbb{Z}}$ by:

$$\forall n \in \mathbb{Z},\ \eta_n = \Phi(B)X_n$$

where B is the backward shift operator. Note that $(\eta_n)_{n\in\mathbb{Z}}$ is not a white noise series because $(X_n)_{n\in\mathbb{Z}}$ is a long-memory process and hence does not belong to the class of autoregressive processes. Since Φ has no root on the unit disk, $(X_n)_{n\in\mathbb{Z}}$ admits a moving-average representation as the fitted AR(k) model in terms of $(\eta_n)_{n\in\mathbb{Z}}$:

$$X_n = \sum_{j=0}^{\infty} c(j)\eta_{n-j}.$$

If $(X_n)_{n\in\mathbb{Z}}$ was an AR(k) associated with the polynomial Φ, the best next step linear predictor would be:

$$\begin{aligned}\widehat{X_n}(1) &= \sum_{j=1}^{\infty} c(i)\eta_{t+1-i} \\ &= a_{1,k}X_n + \ldots + a_{k,k}X_{n+1-k} \text{ si } n \geqslant k.\end{aligned}$$

Here $(X_n)_{n\in\mathbb{Z}}$ is a long-memory process which verifies the assumptions of Section 1. Our goal is to express the polynomial Φ which minimizes the forecast error and to estimate this error.

3.2 Mean-Squared Error

There exists two approaches in order to define the coefficients of the k^{th} degree polynomial Φ: the spectral approach and the time approach.

In the time approach, we choose to define the predictor as the projection mapping on to the closed span of the subset $\{X_n, \ldots, X_{n+1-k}\}$ of the Hilbert space $\mathrm{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ with inner product $< X, Y >= \mathbb{E}(XY)$. Consequently the coefficients of Φ verify the equations, which are called the k^{th} order Yule-Walker equations:

$$\forall j \in [\![1, k]\!], \quad \sum_{i=1}^{k} a_{i,k}\sigma(i-j) = \sigma(j) \tag{27}$$

The mean-squared prediction error is:

$$\mathbb{E}\big[\big(\widehat{X_n}(1) - X_{n+1}\big)^2\big] = c(0)^2\mathbb{E}(\eta_{n+1}^2) = \mathbb{E}(\eta_{n+1}^2).$$

We may write the moving average representation of $(\eta_n)_{n\in\mathbb{N}}$ in terms of $(\varepsilon_n)_{n\in\mathbb{N}}$:

$$\begin{aligned}\eta_n &= \sum_{j=0}^{\infty}\sum_{k=0}^{min(j,p)} \Phi_k b(j-k)\varepsilon_{n-j} \\ &= \sum_{j=0}^{\infty} t(j)\varepsilon_{n-j}\end{aligned}$$

with

$$\forall j \in \mathbb{N}, \quad t(j) = \sum_{k=0}^{min(j,p)} \Phi_k b(j-k).$$

Finally we obtain:

$$\mathbb{E}\big[\big(\widehat{X_n}(1) - X_{n+1}\big)^2\big] = \sum_{j=0}^{\infty} t(j)^2\sigma_\varepsilon^2.$$

In the spectral approach, minimizing the prediction error is equivalent to minimizing a contrast between two spectral densities:

$$\int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda,\Phi)}\mathrm{d}\lambda$$

where f is the spectral density of X_n and $g(.,\Phi)$ is the spectral density of the AR(p) process defined by the polynomial Φ (see for example Yajima (1993)),so:

$$\begin{aligned}\int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda,\Phi)}\mathrm{d}\lambda &= \int_{-\pi}^{\pi} \Big|\sum_{j=0}^{\infty} b(j)e^{-ij\lambda}\Big|^2\Big|\Phi(e^{-i\lambda})\Big|^2\mathrm{d}\lambda \\ &= \int_{-\pi}^{\pi} |\sum_{j=0}^{\infty} t(j)e^{-ij\lambda}|^2\mathrm{d}\lambda \\ &= 2\pi\sum_{j=0}^{\infty} t(j)^2.\end{aligned}$$

In both approaches we nedd to minimize $\sum_{j=0}^{\infty} t(j)$.

3.3 Rate of Convergence of the Error by AR(k) Model Fitting

In the next theorem we derive an asymptotic expression for the prediction error by fitting autoregressive models to the series:

Theorem 3.3.1. *Assume that $(X_n)_{n\in\mathbb{Z}}$ is a long-memory process which verifies the assumptions of Section 1. If $0 < d < \frac{1}{2}$:*

$$\mathbb{E}\big[\big(\widehat{X_k}(1) - X_{k+1}\big)^2\big] - \sigma_\varepsilon^2 = \mathrm{O}(k^{-1})$$

Proof. Since fitting an AR(k) model minimizes the forecast error using k observations, the error by using truncation way is bigger. So, since the truncation method involves an error bounded by $\mathrm{O}\left(k^{-1}\right)$, we obtain:

$$\mathbb{E}\big[\big(\widehat{X_k}(1) - X_{k+1}\big)^2\big] - \sigma_\varepsilon^2 = \mathrm{O}(k^{-1}).$$

Consequently we only need to prove that this rate of convergence is attained . This is the case for the fractionally integrated processes defined in (12). We want the error made when fitting an AR(k) model in terms of the Wiener-Kolmogorov truncation error. Note first that the variance of the white noise series is equal to:

$$\sigma_\varepsilon^2 = \int_{-\pi}^{\pi} f(\lambda) \left| \sum_{j=0}^{+\infty} a_j e^{ij\lambda} \right|^2 \mathrm{d}\lambda.$$

Therefore in the case of a fractionally integrated process $\mathrm{F}(d)$ we need only show that:

$$\int_{-\pi}^{\pi} f(\lambda) \left| \sum_{j=0}^{+\infty} a_j e^{ij\lambda} \right|^2 \mathrm{d}\lambda - \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda, \Phi_k)} \mathrm{d}\lambda \sim C(k^{-1}).$$

$$\begin{aligned}
\int_{-\pi}^{\pi} f(\lambda) \left| \sum_{j=0}^{+\infty} a_j e^{ij\lambda} \right|^2 \mathrm{d}\lambda - \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda, \Phi_k)} \mathrm{d}\lambda &= \int_{-\pi}^{\pi} f(\lambda) \left(\left| \sum_{j=0}^{+\infty} a_j e^{ij\lambda} \right|^2 - \left| \sum_{j=0}^{k} a_{j,k} e^{ij\lambda} \right|^2 \right) \mathrm{d}\lambda \\
&= \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} \left(a_j a_l - a_{j,k} a_{l,k}\right) \sigma(j-l)
\end{aligned}$$

we set $a_{j,k} = 0$ if $j > k$.

$$\begin{aligned}
& \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} \left(a_j a_l - a_{j,k} a_{l,k}\right) \sigma(j-l) && (28) \\
= & \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} (a_j a_l - a_{j,k} a_l) \sigma(j-l) + \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} (a_{j,k} a_l - a_{j,k} a_{l,k}) \sigma(j-l) \\
= & \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sum_{l=0}^{+\infty} a_l \sigma(l-j) + \sum_{j=0}^{k} a_{j,k} \sum_{l=0}^{+\infty} (a_l - a_{l,k}) \sigma(j-l) && (29)
\end{aligned}$$

We first study the first term of the sum (29). For any $j > 0$, we have $\sum_{l=0}^{+\infty} a_l \sigma(l-j) = 0$:

$$\begin{aligned}
\varepsilon_n &= \sum_{j=0}^{\infty} a_l X_{n-l} \\
X_{n-j} \varepsilon_n &= \sum_{l=0}^{\infty} a_l X_{n-l} X_{n-j} \\
\mathbb{E}\left(X_{n-j} \varepsilon_n\right) &= \sum_{l=0}^{\infty} a_l \sigma(l-j) \\
\mathbb{E}\left(\sum_{l=0}^{\infty} b_l \varepsilon_{n-j-l} \varepsilon_n \right) &= \sum_{l=0}^{\infty} a_l \sigma(l-j)
\end{aligned}$$

and we conclude that $\sum_{l=0}^{+\infty} a_l \sigma(l-j) = 0$ because $(\varepsilon_n)_{n\in\mathbb{Z}}$ is an uncorrelated white noise. We can thus rewrite the first term of (29) like:

$$\begin{aligned}\sum_{j=0}^{+\infty}(a_j - a_{j,k})\sum_{l=0}^{+\infty} a_l\sigma(l-j) &= (a_0 - a_{0,k})\sum_{l=0}^{+\infty} a_l\sigma(l) \\ &= 0\end{aligned}$$

since $a_0 = a_{0,k} = 1$ according to definition. Next we study the second term of the sum (29):

$$\sum_{j=0}^{k} a_{j,k}\sum_{l=0}^{+\infty}(a_l - a_{l,k})\sigma(j-l).$$

And we obtain that:

$$\begin{aligned}\sum_{j=0}^{k} a_{j,k}\sum_{l=0}^{+\infty}(a_l - a_{l,k})\sigma(j-l) &= \sum_{j=1}^{k}(a_{j,k}-a_j)\sum_{l=1}^{k}(a_l - a_{l,k})\sigma(j-l) \\ &+\sum_{j=1}^{k}(a_{j,k}-a_j)\sum_{l=k+1}^{+\infty} a_l\sigma(j-l) \qquad (30) \\ &+\sum_{j=0}^{k} a_j\sum_{l=1}^{k}(a_l - a_{l,k})\sigma(j-l) \qquad (31) \\ &+\sum_{j=0}^{k} a_j\sum_{l=k+1}^{+\infty} a_l\sigma(j-l)\end{aligned}$$

Similarly we rewrite the term (30) using the Yule-Walker equations:

$$\sum_{j=1}^{k}(a_{j,k}-a_j)\sum_{l=k+1}^{+\infty} a_l\sigma(j-l) = -\sum_{j=1}^{k}(a_{j,k}-a_j)\sum_{l=0}^{k} a_l\sigma(j-l)$$

We then remark that this is equal to (31). Hence it follows that:

$$\begin{aligned}\sum_{j=0}^{k} a_{j,k}\sum_{l=0}^{+\infty}(a_l - a_{l,k})\sigma(j-l) &= \sum_{j=1}^{k}(a_{j,k}-a_j)\sum_{l=1}^{k}(a_l - a_{l,k})\sigma(j-l) \\ &+2\sum_{j=1}^{k}(a_{j,k}-a_j)\sum_{l=k+1}^{+\infty} a_l\sigma(j-l) \\ &+\sum_{j=0}^{k} a_j\sum_{l=k+1}^{+\infty} a_l\sigma(j-l) \qquad (32)\end{aligned}$$

On a similar way we can rewrite the third term of the sum (32) using Fubini Theorem:

$$\sum_{j=0}^{k} a_j\sum_{l=k+1}^{+\infty} a_l\sigma(j-l) = -\sum_{j=k+1}^{+\infty}\sum_{l=k+1}^{+\infty} a_j a_l\sigma(j-l).$$

This third term is therefore equal to the forecast error in the method of prediction by truncation.

In order to compare the prediction error by truncating the Wiener-Kolmogorov predictor and by fitting an autoregressive model to a fractionally integrated process $\mathrm{F}(d)$, we need the sign of all the components of the sum (32). For a fractionally integrated noise, we know the explicit formula for a_j and $\sigma(j)$:

$$\forall j > 0, \quad a_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} < 0 \text{ and } \forall j \geq 0, \quad \sigma(j) = \frac{(-1)^j \Gamma(1-2d)}{\Gamma(j-d+1)\Gamma(1-j-d)} \sigma_\varepsilon^2 > 0.$$

In order to get the sign of $a_{j,k} - a_j$ we use the explicit formule given in Brockwell and Davis (1988) and we easily obtain that $a_{j,k} - a_j$ is negative for all $j \in [\![1, k]\!]$.

$$\begin{aligned} a_j - a_{j,k} &= \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} - \frac{\Gamma(k+1)\Gamma(j-d)\Gamma(k-d-j+1)}{\Gamma(k-j+1)\Gamma(j+1)\Gamma(-d)\Gamma(k-d+1)} \\ &= -a_j \left(-1 + \frac{\Gamma(k+1)\Gamma(k-d-j+1)}{\Gamma(k-j+1)\Gamma(k-d+1)}\right) \\ &= -a_j \left(\frac{k...(k-j+1)}{(k-d)...(k-d-j+1)} - 1\right) \\ &> 0 \end{aligned}$$

since $\forall j \in \mathbb{N}^* \ a_j < 0$. To give an asymptotic equivalent for the prediction error, we use the sum given in (32). We have the sign of the three terms: the first is negative, the second is positive and the last is negative. Moreover the third is equal to the forecast error by truncation and we have proved that this asymptotic equivalent has order $\mathrm{O}(k^{-1})$. The prediction error by fitting an autoregressive model converges faster to 0 than the error by truncation only if the second term is equivalent to Ck^{-1}, with C constant. Consequently, we search for a bound for $a_j - a_{j,k}$ given the explicit formula for these coefficients (see for example Brockwell and Davis (1988)):

$$\begin{aligned} a_j - a_{j,k} &= \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} - \frac{\Gamma(k+1)\Gamma(j-d)\Gamma(k-d-j+1)}{\Gamma(k-j+1)\Gamma(j+1)\Gamma(-d)\Gamma(k-d+1)} \\ &= -a_j \left(-1 + \frac{\Gamma(k+1)\Gamma(k-d-j+1)}{\Gamma(k-j+1)\Gamma(k-d+1)}\right) \\ &= -a_j \left(\frac{k...(k-j+1)}{(k-d)...(k-d-j+1)} - 1\right) \\ &= -a_j \left(\prod_{m=0}^{j-1} \left(\frac{1-\frac{l}{k}}{1-\frac{l+d}{k}}\right) - 1\right) \\ &= -a_j \left(\prod_{m=0}^{j-1} \left(1 + \frac{\frac{d}{k}}{1-\frac{d+l}{k}}\right) - 1\right). \end{aligned}$$

Then we use the following inequality:

$$\forall x \in \mathbb{R}, \quad 1 + x \leq \exp(x)$$

which gives us:

$$
\begin{aligned}
a_j - a_{j,k} &\leq -a_j \left(\exp\left(\sum_{m=0}^{j-1} \frac{\frac{d}{k}}{1-\frac{d+l}{k}} \right) - 1 \right) \\
&\leq -a_j \left(\exp\left(d \sum_{m=0}^{j-1} \frac{1}{k-d-l} \right) - 1 \right) \\
&\leq -a_j \exp\left(d \sum_{m=0}^{j-1} \frac{1}{k-d-l} \right)
\end{aligned}
$$

According to the previous inequality, we have:

$$
\begin{aligned}
\sum_{j=1}^{k} (a_j - a_{j,k}) \sum_{l=k+1}^{+\infty} -a_l \sigma(j-l) &= \sum_{j=1}^{k-1} (a_j - a_{j,k}) \sum_{l=k+1}^{+\infty} -a_l \sigma(j-l) \\
&\quad + (a_k - a_{k,k}) \sum_{l=k+1}^{+\infty} -a_l \sigma(k-l) \\
&\leq \sum_{j=1}^{k-1} -a_j \exp\left(d \sum_{m=0}^{j-1} \frac{1}{k-d-m} \right) \sum_{l=k+1}^{+\infty} -a_l \sigma(j-l) \\
&\quad + (-a_k) \exp\left(d \sum_{m=0}^{k-1} \frac{1}{k-d-m} \right) \sum_{l=k+1}^{+\infty} -a_l \sigma(k-l) \\
&\leq \sum_{j=1}^{k-1} -a_j \exp\left(d \int_0^j \frac{1}{k-d-m} \mathrm{d}m \right) \sum_{l=k+1}^{+\infty} -a_l \sigma(j-l) \\
&\quad + (-a_k) k^{\frac{3}{2}d} \sum_{l=k+1}^{+\infty} -a_l \sigma(k-l)
\end{aligned}
$$

As the function $x \mapsto \frac{1}{k-d-x}$ is increasing, we use the Integral Test Theorem. The inequality on the second term follows from:

$$
\begin{aligned}
\sum_{m=0}^{k-1} \frac{1}{k-d-m} &\sim \ln(k) \\
&\leq \frac{3}{2} \ln(k)
\end{aligned}
$$

for k large enough. Therefore there exists K such that for all $k \geq K$:

$$\begin{aligned}
\sum_{j=1}^{k}(a_j - a_{j,k}) \sum_{l=k+1}^{+\infty} -a_l\sigma(j-l) \quad &\leq \quad \sum_{j=1}^{k-1} -a_j \exp\left(d \ln\left(\frac{k-d}{k-d-j}\right)\right) \sum_{l=k+1}^{+\infty} -a_l\sigma(j-l) \\
&\quad +(-a_k)k^{\frac{3}{2}d} \sum_{l=k+1}^{+\infty} -a_l\sigma(0) \\
&\leq \quad C(k-d)^d \sum_{j=1}^{k-1} j^{-d-1}(k-d-j)^{-d} \sum_{l=k+1}^{+\infty} l^{-d-1}(l-j)^{2d-1} \\
&\quad +Ck^{-d-1}k^{\frac{3}{2}d}k^{-d} \\
&\leq \quad \frac{C}{(k-d)^2} \int_{1/(k-d)}^{1} j^{-d-1}(1-j)^{-d} \int_{1}^{+\infty} l^{-d-1}(l-1)^{2d-1} \mathrm{d}l\mathrm{d}j \\
&\quad +Ck^{-\frac{1}{2}d-1} \\
&\leq \quad C'(k-d)^{-2+d} + Ck^{-\frac{1}{2}d-1}
\end{aligned}$$

and so the positive term has a smaller asymptotic order than the forecast error made by truncating. Therefore we have proved that in the particular case of F(d) processes, the two prediction errors are equivalent to Ck^{-1} with C constant. □

The two approaches to next-step prediction, by truncation to k terms or by fitting an autoregressive model AR(k) have consequently a prediction error with the same rate of convergence k^{-1}. So it is interesting to study how the second approach improves the prediction The following quotient:

$$r(k) := \frac{\sum_{j=1}^{k}(a_{j,k} - a_j)\sum_{l=1}^{k}(a_l - a_{l,k})\sigma(j-l) + 2\sum_{j=1}^{k}(a_{j,k} - a_j)\sum_{l=k+1}^{+\infty} a_l\sigma(j-l)}{\sum_{j=0}^{k} a_j \sum_{l=k+1}^{+\infty} a_l\sigma(j-l)} \tag{33}$$

is the ratio of the difference between the two prediction errors and the prediction error by truncatingn in the particular case of a fractionally integrated noise F(d). The figure 3.1 shows that the prediction by truncation incurs a larger performance loss when $d \to 1/2$. The improvement reaches 50 per cent when $d > 0.3$ and $k > 20$.

3.4 Error due to Estimation of the Forecast Coefficients

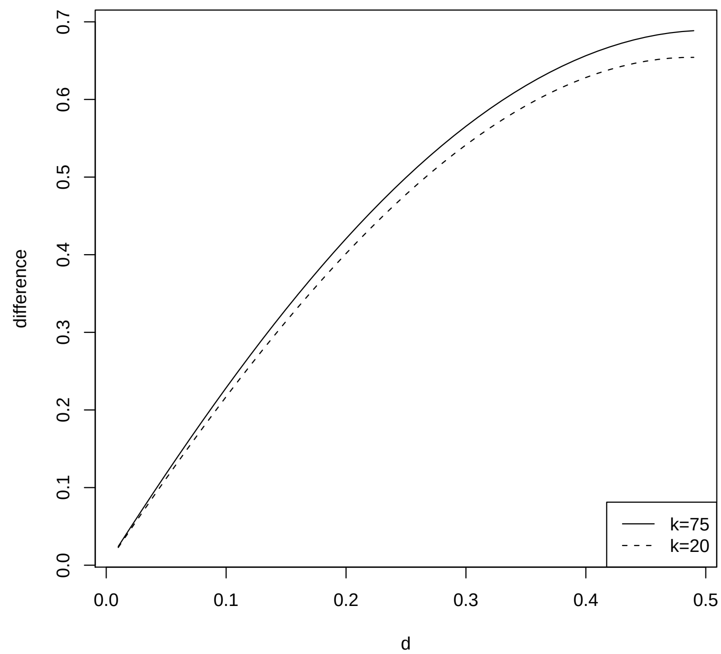
Let $(X_j)_{j\in\mathbb{Z}}$ be a stochastic process, which verifies the assumptions of section 1, and let $(Y_j)_{j\in\mathbb{Z}}$ be a process which is independent of $(X_j)_{j\in\mathbb{Z}}$, but which has the same stochastic structure. We want to predict X_{k+1} knowing $(X_j)_{j\in[\![1,k]\!]}$ and we assume that forecast coefficients are estimated based on a realisation $(Y_j)_{j\in[\![1,T]\!]}$.

We estimate the forecast coefficients using the Yule-Walker equations (27) where we replace the true covariances by the empirical covariances computed from the realisation $(Y_j)_{j\in[\![1,T]\!]}$:

$$\widehat{\sigma(k)} = \frac{1}{T}\sum_{t=1}^{T-k} Y_t Y_{t+k} \tag{34}$$

There exists a recursive scheme for computing the forecast coefficients. It is known as the Durbin-Levinson or innovation algorithm and it is described for example in Brockwell and Davis (1991).

Figure 3.1: Ratio $r(k)$, $d \in]0, 1/2[$ defined in (33)



Let $(Y_j)_{j\in\mathbb{Z}}$ be a zero-mean process with autocovariance function σ. The coefficients $(a_{i,k})_{i\in[\![1,k]\!]}$ satisfy the Yule-Walker k^{th} equations:

$$\forall j \in [\![1, k]\!], \quad \sigma(j) = \sum_{u=1}^{k} \sigma(u-j) a_{u,k}.$$

If we let $v(0) = \sigma(0)$ and $a_{1,1} = \sigma(1)/\sigma(0)$, then we have for any integer n:

$$\begin{aligned} a_{n,n} &= \left[\sigma(n) - \sum_{j=1}^{n-1} a_{j,n-1}\sigma(n-j)\right] \frac{1}{v(n-1)} \\ \begin{pmatrix} a_{1,n} \\ \vdots \\ a_{n-1,n} \end{pmatrix} &= \begin{pmatrix} a_{1,n-1} \\ \vdots \\ a_{n-1,n-1} \end{pmatrix} - a_{n,n} \begin{pmatrix} a_{n-1,n-1} \\ \vdots \\ a_{1,n-1} \end{pmatrix} \\ v(n) &= v(n-1)(1 - a_{n,n}^2). \end{aligned}$$

We denote by $(\widehat{a_{1,k}}, \ldots, \widehat{a_{k,k}})$ the respective solutions to the Yule-Walker equations obtained by replacing the covariances by theirs estimates the empirical covariances defined in (34). Contrary to Section 2.2, the estimation of the forecast coefficients is non-parametric.

Another way to estimate the coefficients has been considered by Yajima (1993). Our method borrow the idea (see section 3.2) that the coefficients of the AR(k) minimize:

$$\int_{-\pi}^{\pi} f(\lambda)/g(\lambda,\Phi)\mathrm{d}\lambda. \tag{35}$$

If we replace in (35) the spectral density by the periodogram I_T :

$$I_T(\lambda) = \frac{\left|\sum_{t=1}^{T} X_t e^{it\lambda}\right|^2}{2\pi T},$$

then:

$$(\widehat{a_{1,k}},\ldots,\widehat{a_{k,k}}) = \operatorname*{argmin}_{\Phi} \int_{-\pi}^{\pi} I_T(\lambda)/g(\lambda,\Phi)\mathrm{d}\lambda. \tag{36}$$

From now on we incorporate the effects of estimation of the AR(k) coefficients using a realisation of length T, as $T \to +\infty$ and study the mean-squared prediction error due to this estimation. We define $\widehat{X_{T,k}}(1)$ the predictor with all the coefficients $a_{j,k}$ replaced by their estimates:

$$\widehat{X_{T,k}}(1) := \sum_{j=1}^{k} \widehat{a_{j,k}} X_{k+1-j}$$

More precisely, we study the mean-squared difference between the predictor with the estimated coefficients $\widehat{a_{j,k}}$ and the predictor with the true coefficients $a_{j,k}$:

$$\begin{aligned}
&\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X}_k(1)\right)^2\right] \\
=\;& \mathbb{E}\left((\widehat{a_{1,k}} - a_{1,k},\ldots,\widehat{a_{k,k}} - a_{k,k}) \begin{pmatrix} X_k \\ \vdots \\ X_1 \end{pmatrix} \right)^2 \\
=\;& \operatorname{trace}\left(\mathbb{E}\left(\begin{pmatrix} \widehat{a_{1,k}} - a_{1,k} \\ \vdots \\ \widehat{a_{k,k}} - a_{k,k} \end{pmatrix} (\widehat{a_{1,k}} - a_{1,k},\ldots,\widehat{a_{k,k}} - a_{k,k}) \right) \mathbb{E}\left(\begin{pmatrix} X_k \\ \vdots \\ X_1 \end{pmatrix} (X_k,\ldots,X_1) \right) \right) \\
=\;& \operatorname{trace}\left(\mathbb{E}\left(\begin{pmatrix} \widehat{a_{1,k}} - a_{1,k} \\ \vdots \\ \widehat{a_{k,k}} - a_{k,k} \end{pmatrix} (\widehat{a_{1,k}} - a_{1,k},\ldots,\widehat{a_{k,k}} - a_{k,k}) \right) \Sigma_k \right).
\end{aligned}$$

First we estimate the covariance matrix:

$$\mathbb{E}\left(\begin{pmatrix} \widehat{a_{1,k}} - a_{1,k} \\ \vdots \\ \widehat{a_{k,k}} - a_{k,k} \end{pmatrix} (\widehat{a_{1,k}} - a_{1,k},\ldots,\widehat{a_{k,k}} - a_{k,k}) \right).$$

For later convenience, we now introduce the vector:

$$\mathbf{1}_k^* := \underbrace{(1,\ldots,1)}_{k}$$

and the $(k \times k)$ matrix:

$$\mathbf{1}_{k,k} := \mathbf{1}_k \mathbf{1}_k^*.$$

We now state the theorem which allows us to conclude.

Theorem 3.4.1. *We assume that the process $(Y_n)_{n\in\mathbb{Z}}$ is Gaussian, that its autocovariance function σ verifies:*

$$\sigma(j) \sim \lambda j^{2d-1} \text{ with } \lambda > 0,$$

that the coefficients of its infinite moving average representation b_j verify:

$$b_j \sim \delta j^{d-1} \text{ with } \delta > 0,$$

and finally that the white noise process $(\varepsilon_n)_{n\in\mathbb{Z}}$ is such that $\forall n \in \mathbb{Z}$, $\mathbb{E}(\varepsilon_n^4) < +\infty$. We will denote by $g_{i,j}$ the function:

$$\begin{array}{cccc} g_{i,j}: & \mathbb{R}^{k+1} & \to & \mathbb{R} \\ & (x_0, \ldots, x_k) & \mapsto & (y_i - a_{i,k})(y_j - a_{j,k}) \end{array}$$

where

$$\begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} x_0 & x_1 & \ldots & x_k \\ x_1 & x_0 & \ddots & x_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ x_k & x_{k-1} & \ldots & x_0 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

Then

$$\begin{aligned} &\mathbb{E}(g_{i,j}(\widehat{\sigma(0)}, \widehat{\sigma(1)}, \ldots, \widehat{\sigma(k)})) \\ &= \begin{cases} \left(1 - \sum_{r=1}^k a_{r,k}\right)^2 C n^{4d-2} \left(\Sigma_k^{-1} \boldsymbol{1}_{k,k} \Sigma_k^{-1}\right)_{(i,j)} + \mathrm{O}(n^{6d-3}) & \text{if } \frac{1}{4} < d < \frac{1}{2} \\ \left(1 - \sum_{r=1}^k a_{r,k}\right)^2 D \frac{\ln(n)}{n} \left(\Sigma_k^{-1} \boldsymbol{1}_{k,k} \Sigma_k^{-1}\right)_{(i,j)} + \mathrm{O}(n^{-3/2}) & \text{if } d = 1/4 \\ n^{-1} 4 \left(\Sigma_k^{-1} H \Sigma_k^{-1}\right)_{(i,j)} + \mathrm{O}(n^{-3/2}) & \text{if } 0 < d < \frac{1}{4} \end{cases} \end{aligned}$$

where C and D are constants independent of n and k. The definition of the matrix H follows. We define h as $h(\lambda) = |1 - \sum_{r=1}^k a_{r,k} \mathrm{e}^{ir\lambda}|^2$ and we denote by $h^{(r)}$ the derivative of the function h with respect to $a_{r,k}$. The (i,j)-th entry of the matrix H is given by:

$$H_{i,j} := \int_{-\pi}^{\pi} h^{(i)}(\lambda) h^{(j)}(\lambda) f^2(\lambda) \mathrm{d}\lambda.$$

Proof. The proof is given in Appendix. □

Next we estimate the asymptotic behaviour of $\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X}_k(1)\right)^2\right]$ and we state the following Theorem which gives an estimation of the mean-squared error when $d \geq 1/4$:

Theorem 3.4.2. *We assume that the assumptions of the Theorem 3.4.1 hold. We assume also that the spectral density of the process is such that:*

$$\forall x \in [-\pi, \pi], \quad f(x) = f_d(x) L(x)$$

with f_d defined by:

$$\forall x \in [-\pi,\pi], \quad f_d(x) = 2^{-2d-1}\pi^{-1}\left(\sin^2(x/2)\right)^{-d}$$

and L a positive, integrable on $[-\pi,\pi]$, continuous at 0 and bounded below by a positive constant. If $d = 1/4$ then

$$\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X_k}(1)\right)^2\right] = \mathrm{O}\left(\frac{\log(T)\sqrt{k}}{T}\right)$$

and if $d \in]1/4, 1/2[$, we thus get

$$\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X_k}(1)\right)^2\right] = \mathrm{O}\left(\frac{k^{1-2d}}{T^{2-4d}}\right).$$

Remark The assumption that L and so f are bounded below by a positive constant is not a new very restrictive assumption. Since we have assumed that the process admits an infinite autoregressive representation:

$$\varepsilon_n = \sum_{j=0}^{\infty} a_j X_{n-j},$$

where the coefficients a_j are absolutely summable, we have that the spectral density can be written as:

$$f(\lambda) = \frac{1}{\left|\sum_{j=0}^{\infty} a_j \mathrm{e}^{ij\lambda}\right|^2}$$

and consequently the spectral density can not vanish on $[-\pi,\pi[$.

Proof. Applying the last theorem, we obtain that if $d = 1/4$ then

$$\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X_k}(1)\right)^2\right] = \mathrm{O}\left(\mathrm{trace}\left(\frac{\log(T)}{T}\left(\sum_{j=0}^{k} a_{j,k}\right)^2 \Sigma_k^{-1}\mathbf{1}_{k,k}\right)\right)$$

and if $d \in]1/4, 1/2[$ then

$$\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X_k}(1)\right)^2\right] = \mathrm{O}\left(\mathrm{trace}\left(\frac{1}{T^{2-4d}}\left(\sum_{j=0}^{k} a_{j,k}\right)^2 \Sigma_k^{-1}\mathbf{1}_{k,k}\right)\right).$$

First we estimate $\left|\sum_{j=0}^{k} a_{j,k}\right|$. We write this like:

$$\left|\sum_{j=0}^{k} a_{j,k}\right| \quad \leq \quad \sqrt{k}\sqrt{\sum_{j=1}^{k} |a_{j,k} - a_j|} + \sum_{j=0}^{k} |a_j|.$$

We follow the proof of Theorem 3.3 of Inoue and Kasahara (2006) about the convergence of the sequence of the misspecified AR(k) model coefficients to the R(∞) representation coefficients. We shall remark that there exists C_1, C_2 and K such that if:

$$k \geq K, \quad k\left(a_{j,k} - a_j\right) \leq C_1 \sum_{u=k-j}^{+\infty} |a_u| + C_2 \sum_{u=j}^{+\infty} |a_u|$$

and C is a generic constant:

$$\text{si} \quad k \geq K, \quad k\left(a_{j,k} - a_j\right) \leq C\left(\sum_{u=k-j}^{+\infty} |a_u| + \sum_{u=j}^{+\infty} |a_u|\right)$$

We thus get that if $k \geq K$:

$$\begin{aligned}
\sum_{j=1}^{k} (a_{j,k} - a_j)^2 &\leq \frac{C}{k^2} \sum_{j=1}^{k} \left[\left(\sum_{u=k-j}^{+\infty} |a_u|\right)^2 + \left(\sum_{u=j}^{+\infty} |a_u|\right)\right]^2 \\
&\leq \frac{C}{k^2} \sum_{j=1}^{k} (k-j+1)^{-2d} + j^{-2d} \\
&\leq \mathrm{O}\left(k^{-2d-1}\right). \qquad (37)
\end{aligned}$$

So we may conclude that:

$$\begin{aligned}
\left|\sum_{j=0}^{k} a_{j,k}\right| &\leq \mathrm{O}(k^{-d}) + \sum_{j=0}^{+\infty} |a_j| \\
&= \mathrm{O}(1). \qquad (38)
\end{aligned}$$

Next we have to study the asymptotic properties of:

$$\operatorname{trace}\left(\Sigma_k^{-1} \mathbf{1}_{k,k}\right) = (1 \ldots 1) \ \Sigma_k^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Then applying Theorem 6.1 of Adenstedt (1974) under the assumptions of theorem 3.4.2 we obtain the following asymptotic equivalent:

$$(1 \ldots 1) \ \Sigma_k^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \sim \left(\frac{k^{1-2d}\Gamma(-2d+1)L(0)}{\beta(-d+1, -d+1)}\right)^{-1}$$

where Γ and β are respectively the gamma function and the beta function. The result follows.

□

The last case is when $0 < d < 1/4$:

Theorem 3.4.3. *We assume that the assumptions of 3.4.1 hold. We assume also that the spectral density f is bounded above by a constant positive. If $0 < d < 1/4$ then*

$$\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X}_k(1)\right)^2\right] = \mathrm{O}\left(\frac{k}{T}\right).$$

Proof. We call $(\Phi_i := \Phi_{i,1} + \ldots + \Phi_{i,i}x^{i-1})_{i\in\mathbb{N}^*}$ the orthonormal polynomials associated with the spectral density f, that is to say Φ_i is a $(i-1)^{\text{th}}$ degree polynomial such that

$$\forall j,l \in \mathbb{N}^*, \quad \int_{-\pi}^{\pi} f(\lambda)\Phi_j\left(\mathrm{e}^{i\lambda}\right)\Phi_l\left(\mathrm{e}^{-i\lambda}\right)\mathrm{d}\lambda = \delta_{j,l}$$

where δ is the Kronecker delta. We then define the matrix T_k by:

$$T_k = \begin{pmatrix} \Phi_{1,1} & 0 & \ldots & 0 \\ \Phi_{2,1} & \Phi_{2,2} & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{k,1} & \Phi_{k,2} & \ldots & \Phi_{k,k} \end{pmatrix}$$

T_k verifies the following conditions:

$$T_k\Sigma_k T_k^* = Id_k$$

and so

$$\Sigma_k^{-1} = T_k^* T_k. \tag{39}$$

Using (39), we obtain that:

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X_k}(1)\right)^2\right] &= \frac{1}{T}\mathrm{trace}\left(\Sigma_k^{-1}H\right) \\ &= \frac{1}{T}\mathrm{trace}\left(T_k H T_k^*\right) \end{aligned} \tag{40}$$

with H defined in Theorem 3.4.1. We define $G_k : \lambda \mapsto \sum_{j=0}^{k} a_{j,k}\mathrm{e}^{ij\lambda}$.
$\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X_k}(1)\right)^2\right]$ can therefore be rewritten like:

$$\begin{aligned} & \frac{1}{T}\mathrm{trace}\left(\left(\int_{-\pi}^{\pi} f^2(\lambda)\mathrm{Re}\left(G_k(\lambda)\Phi_j(\mathrm{e}^{ij\lambda})\right)\mathrm{Re}\left(G_k(-\lambda)\Phi_l(\mathrm{e}^{-il\lambda})\right)\mathrm{d}\lambda\right)_{j,l\in[\![1,k]\!]}\right) \\ = & \frac{1}{T}\mathrm{trace}\left(\left(\int_{-\pi}^{\pi} f^2(\lambda)\mathrm{Re}\left(G_k(\lambda)\Phi_j(\mathrm{e}^{ij\lambda})G_k(-\lambda)\Phi_l(\mathrm{e}^{-il\lambda})\right)\mathrm{d}\lambda\right)_{j,l\in[\![1,k]\!]}\right) \\ + & \frac{1}{T}\mathrm{trace}\left(\left(\int_{-\pi}^{\pi} f^2(\lambda)\mathrm{Im}\left(G_k(\lambda)\Phi_j(\mathrm{e}^{ij\lambda})\right)\mathrm{Im}\left(G_k(-\lambda)\Phi_l(\mathrm{e}^{-il\lambda})\right)\mathrm{d}\lambda\right)_{j,l\in[\![1,k]\!]}\right) \end{aligned}$$

because $\mathrm{Re}(ab) = \mathrm{Re}(a)\mathrm{Re}(b) - \mathrm{Im}(a)\mathrm{Im}(b)$. For later convenience, we note:

$$\begin{aligned} A &:= \left(\int_{-\pi}^{\pi} f^2(\lambda)\mathrm{Re}\left(G_k(\lambda)\Phi_j(\mathrm{e}^{i\lambda})\right)\mathrm{Re}\left(G_k(-\lambda)\Phi_l(\mathrm{e}^{-i\lambda})\right)\mathrm{d}\lambda\right)_{j,l\in[\![1,k]\!]} \\ B &:= \left(\int_{-\pi}^{\pi} f^2(\lambda)\mathrm{Re}\left(G_k(\lambda)\Phi_j(\mathrm{e}^{i\lambda})G_k(-\lambda)\Phi_l(\mathrm{e}^{-i\lambda})\right)\mathrm{d}\lambda\right)_{j,l\in[\![1,k]\!]} \\ &= \left(\int_{-\pi}^{\pi} f^2(\lambda)|G_k(\lambda)|^2\Phi_j(\mathrm{e}^{i\lambda})\Phi_l(\mathrm{e}^{-i\lambda})\mathrm{d}\lambda\right)_{j,l\in[\![1,k]\!]} \\ C &:= \left(\int_{-\pi}^{\pi} f^2(\lambda)\mathrm{Im}\left(G_k(\lambda)\Phi_j(\mathrm{e}^{i\lambda})\right)\mathrm{Im}\left(G_k(-\lambda)\Phi_l(\mathrm{e}^{-i\lambda})\right)\mathrm{d}\lambda\right)_{j,l\in[\![1,k]\!]} \end{aligned}$$

Then we have $A = B + C$. We will prove that A, B and $-C$ are symmetric and positive matrices, which implies that $0 \leq \text{trace}(A) \leq \text{trace}(B)$. First we study the symmetry: A is symmetric because the real part of a complex is equal to that of its conjugate, B is symmetric because $\lambda \mapsto f^2(\lambda)|G_k(\lambda)|^2$ is a symmetric function and C is symmetric because the imaginary part is equal to the negative of the imaginary part of its conjugate. Next we study the positivity. Let $q := (q_1, \ldots, q_k)$ be a vector. We have:

$$\begin{aligned}
qAq^* &= \int_{-\pi}^{\pi} f^2(\lambda)\text{Re}\left(\sum_{j=1}^{k} G_k(\lambda)q_j\Phi_j(\mathrm{e}^{i\lambda})\right)\text{Re}\left(\sum_{l=1}^{k} G_k(-\lambda)q_l\Phi_l(\mathrm{e}^{-i\lambda})\right)\mathrm{d}\lambda \geq 0 \\
qBq^* &= \int_{-\pi}^{\pi} f^2(\lambda)|G_k(\lambda)|^2 \sum_{j=1}^{k} q_j\Phi_j(\mathrm{e}^{i\lambda}) \sum_{l=1}^{k} q_l\Phi_l(\mathrm{e}^{-i\lambda})\mathrm{d}\lambda \geq 0 \\
qCq^* &= \int_{-\pi}^{\pi} f^2(\lambda)\text{Im}\left(\sum_{j=1}^{k} G_k(\lambda)q_j\Phi_j(\mathrm{e}^{i\lambda})\right)\text{Im}\left(\sum_{j=1}^{k} G_k(-\lambda)q_l\Phi_l(\mathrm{e}^{-i\lambda})\right)\mathrm{d}\lambda \leq 0.
\end{aligned}$$

The traces of these matrices A , B et $-C$ are equal to the sum of theirs eigenvalues since they are symmetric and thus diagonalizable. Because these matrices are positive, all their eigenvalues are positive and the traces are also positive. Therefore we obtain that:

$$0 \leq \text{trace}(A) \leq \text{trace}(B). \tag{41}$$

To find a bound for trace(A), it is sufficient to find a bound for trace(B):

$$\begin{aligned}
\text{trace}(B) &= \sum_{j=1}^{k} \int_{-\pi}^{\pi} f^2(\lambda)|G_k(\lambda)|^2\Phi_j(\mathrm{e}^{i\lambda})\Phi_j(\mathrm{e}^{-i\lambda})\mathrm{d}\lambda \\
&= \int_{-\pi}^{\pi} f^2(\lambda)|G_k(\lambda)|^2 K_k(\mathrm{e}^{i\lambda}, \mathrm{e}^{i\lambda})\mathrm{d}\lambda
\end{aligned}$$

where K_k is the reproducing kernel defined by:

$$\forall x, y \in \mathbb{C}, K_k(x, y) = \sum_{j=1}^{k} \Phi_j(x)\Phi_j(\bar{y}).$$

We have assumed that the spectral density f is bounded from below by a positive constant c, so we can apply the Theorem 2.2.4 of Simon (2005) and we get:

$$\forall \lambda \in [-\pi, \pi], K_k(\mathrm{e}^{i\lambda}, \mathrm{e}^{i\lambda}) \leq k\frac{2\pi}{c}.$$

We look for a bound for $|G_k(\lambda)|^2$:

$$\begin{aligned}
\forall \lambda \in [-\pi, \pi], |G_k(\lambda)|^2 &\leq \left(\sum_{j=0}^{k} |a_{j,k}|\right)^2 \\
&= \mathrm{O}(1)
\end{aligned}$$

as we have proven in (38). This bound is independent of λ. We finally notice that if $0 < d < \frac{1}{4}$ then f is square integrable. So we obtain that:

$$\text{trace}(B) = \mathrm{O}(k)$$

and we conclude using (40) and (41) that:

$$\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X}_k(1)\right)^2\right] = \mathrm{O}\left(\frac{k}{T}\right).$$

□

3.5 Conclusion

Fitting an AR(k) model also involves two mean-squared error components: the first is due to fitting a misspecified model and is bounded by $\mathrm{O}(k^{-1})$ and the second is due to the estimation of the Yule-Walker coefficients $a_{j,k}$ from a independent realisation of length T and is bounded by $\mathrm{O}(k/T)$ if $0 < d < 1/4$ (Bhansali (1978) has the same asymptotic equivalent for short memory processes), bounded by $\mathrm{O}(k^{1/2}\log(T)/T)$ if $d = 1/4$ and bounded by $\mathrm{O}(k^{1-2d}/T^{2-4d})$ if $1/4 < d < 1/2$. As in Section 2.3, if we want to compare the two types of forecast error, we need to state a relation between k and T and moreover distinguish 3 cases for the value of d.

In both methods by truncating to k terms the Wiener-Kolmogorov predictor or by fitting an AR(k) model, the mean-squared error of prediction due to the method is bounded by $\mathrm{O}(k^{-1})$. Nevertheless, the factor of k^{-1} in this equivalent depends on d. We have shown that the factor tends to infinity when d tends to 1/2 in the method by truncation in the special case of fractionally integrated noise (Section 2.1) so that the error increases for d near 1/2. Moreover for this value of d, figure (3.1) show that fitting an AR(k) model greatly reduces the error. For the errors due to the estimation of the forecast coefficients, the method by truncation is optimal since if we assume that T/k tends to infinity (necessary condition to have some mean-squared error which converges to 0), then for all d in $]0, 1/2[\setminus\{1/4\}$:

$$\mathbb{E}\left(\widetilde{X'_{T,k}}(1) - \widetilde{X}'_k(1)\right)^2 = \mathrm{o}\left(\mathbb{E}\left[\left(\widehat{X_{T,k}}(1) - \widehat{X}_k(1)\right)^2\right]\right).$$

In the end, we have so to consider the value of long-memory parameter d, the length of the series k and T to decide on a prediction method.

4 Appendix

4.1 Proof of Lemma 2.1.1

Let g be the function $(l, j) \mapsto j^{-d-1+\delta} l^{-d-1+\delta} |l-j|^{2d-1+\delta}$. Let m and n be two integers. We assume that $\delta < 1 - 2d$ and that $m \geq \frac{\delta-d-1}{\delta+2d-1}$ for all $\delta \in \left]0, \frac{\delta-d-1}{\delta+2d-1}\right[$. We introduce $A_{n,m}$ the square $[n, n+1] \times [m, m+1]$. If $n \geq m+1$ then

$$\int_{A_{n,m}} g(l, j)\mathrm{d}j\mathrm{d}l \geq g(n+1, m).$$

Proof. We restrict the domain of g to the square $A_{n,m}$. First we will show that $g(.,j)$ is a decreasing and then we compute its derivative:

$$\begin{aligned}(g(j,.))'(l) &= \left[(-d-1+\delta)l^{-1}+(2d-1+\delta)(l-j)^{-1}\right]j^{-d-1+\delta}l^{-d-1+\delta}(l-j)^{2d-1+\delta} \\ &\leq 0\end{aligned}$$

since $\delta < 1-2d$. We show then that $g(l,.)$ is increasing:

$$\begin{aligned}(g(.,l))'(j) &= \left[(-d-1+\delta)j^{-1}-(2d-1+\delta)(l-j)^{-1}\right]j^{-d-1+\delta}l^{-d-1+\delta}(l-j)^{2d-1+\delta} \\ &\geq 0\end{aligned}$$

because

$$j \geq \frac{\delta-d-1}{\delta+2d-1}.$$

Then the function g attains its minimum at $(n+1,m)$ and we have

$$\begin{aligned}\forall(l,j)\in A_{n,m}, g(l,j) &\geq g(n+1,m) \\ \int_{A_{n,m}} g(l,j)\mathrm{d}j\mathrm{d}l &\geq \int_{A_{n,m}} g(n+1,m)\mathrm{d}j\mathrm{d}l \\ \int_{A_{n,m}} g(l,j)\mathrm{d}j\mathrm{d}l &\geq g(n+1,m).\end{aligned}$$

The results follows.

□

4.2 Proof of Theorem 3.4.1

By assumption, the process $(Y_n)_{n\in\mathbb{Z}}$ is Gaussian. We also assume that its autocovariance function σ verifies:

$$\sigma(j) \sim \lambda j^{2d-1} \text{ with } \lambda > 0,$$

that the coefficients of its moving average representation b_j are such that:

$$b_j \sim \delta j^{d-1} \text{ with } \delta > 0,$$

and that the white-noise series $(\varepsilon_n)_{n\in\mathbb{Z}}$ has finite fourth moments. Let $g_{i,j}$ be the function:

$$\begin{array}{cccc} g_{i,j}: & \mathbb{R}^{k+1} & \to & \mathbb{R} \\ & (x_0,\ldots,x_k) & \mapsto & (y_i-a_{i,k})(y_j-a_{j,k}) \end{array}$$

with

$$\begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} x_0 & x_1 & \ldots & x_k \\ x_1 & x_0 & \ddots & x_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ x_k & x_{k-1} & \ldots & x_0 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

Therefore

$$\mathbb{E}(g_{i,j}(\widehat{\sigma(0)},\widehat{\sigma(1)},\dots,\widehat{\sigma(k)}))$$
$$= \begin{cases} \left(1-\sum_{r=1}^{k} a_{r,k}\right)^2 Cn^{4d-2}\left(\Sigma_k^{-1}\mathbf{1}_{k,k}\Sigma_k^{-1}\right)_{(i,j)} + \mathrm{O}(n^{6d-3}) & \text{if } \frac{1}{4}<d<\frac{1}{2} \\ \left(1-\sum_{r=1}^{k} a_{r,k}\right)^2 D\frac{\ln(n)}{n}\left(\Sigma_k^{-1}\mathbf{1}_{k,k}\Sigma_k^{-1}\right)_{(i,j)} + \mathrm{O}(n^{-3/2}) & \text{if } d=1/4 \\ n^{-1}4\left(\Sigma_k^{-1}H\Sigma_k^{-1}\right)_{(i,j)} + \mathrm{O}(n^{-3/2}) & \text{if } 0<d<\frac{1}{4} \end{cases}$$

where C and D are constants and independent of n and k. The definition of matrix H follows. We define h as $h(\lambda)=|1-\sum_{r=1}^{k} a_{r,k}\mathrm{e}^{ir\lambda}|^2$ and we denote by $h^{(r)}$ the derivative of the function h with respect to $a_{r,k}$. The (i,j)-th entry of the matrix H is given by:

$$H_{i,j} := \int_{-\pi}^{\pi} h^{(i)}(\lambda)h^{(j)}(\lambda)f^2(\lambda)\mathrm{d}\lambda \tag{42}$$

Proof. We write a 2^{nd} order Taylor expansion of the function $g_{i,j}$ applying Theorem 5.4.3 in Fuller (1976) as in the Section 2.2. We will refer to the following version.:

(i) If $\mathbb{E}\left(|\widehat{\sigma(k)}-\sigma(k)|^3\right)=\mathrm{O}(a_n)$;

(ii) if $g_{i,j}$ is uniformly bounded;

(iii) if the first and the second derivatives of $g_{i,j}$ are continuous and bounded functions on a neighbourhood of $(\sigma(0),\dots,\sigma(k))$

then

$$\begin{aligned} &\mathbb{E}\left(g_{i,j}(\widehat{\sigma(0)},\widehat{\sigma(1)},\dots,\widehat{\sigma(k)})\right) \\ = \;& g_{i,j}(\sigma(0),\dots,\sigma(k)) + \sum_{l=0}^{k}\mathbb{E}(\widehat{\sigma(l)}-\sigma(l))\frac{\partial g_{i,j}}{\partial x_l}(\sigma(0),\dots,\sigma(k)) \\ &+\frac{1}{2}\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial^2 g_{i,j}}{\partial x_l\partial x_m}(\sigma(0),\dots,\sigma(k))\mathbb{E}\left((\widehat{\sigma(l)}-\sigma(l))(\widehat{\sigma(m)}-\sigma(m))\right)+\mathrm{O}(a_n). \end{aligned}$$

We first verify that the assumptions hold. We need a bound for the third order moments of the empirical covariances.

Lemma 4.2.1.

$$\mathbb{E}\left[\Big|\frac{1}{n}\sum_{t=0}^{n-k}X_tX_{t+k}-\sigma(k)\Big|^3\right] = \begin{cases}\mathrm{O}(n^{-3/2}) & \text{if } d\le 1/4 \\ \mathrm{O}(n^{6d-3}) & \text{if } d>1/4\end{cases} \tag{43}$$

Proof. Lemma 4.2.1 is proven in Section 4.3. □

In this way we obtain a bound for the rest of the Taylor series. Moreover $g_{i,j}$ is an uniformly bounded function since its results are the coefficients of the autoregressive process. Since the derivatives of $g_{i,j}$ at $(\sigma(0),\dots,\sigma(k))$ are finite, there exits a neighbourhood of $(\sigma(0),\dots,\sigma(k))$ such that on this all the derivatives are uniformly bounded. So we apply the Theorem 5.4.3 in Fuller (1976). First we note that:

$$g_{i,j}(\sigma(0),\dots,\sigma(k))=0$$

and

$$
\begin{aligned}
\forall l \in [\![0,k]\!], \quad \frac{\partial g_{i,j}}{\partial x_l}(\sigma(0),\dots,\sigma(k)) &= \frac{\partial y_i}{\partial x_l}(\sigma(0),\dots,\sigma(k))(y_j - a_{j,k}) \\
&\quad + (y_i - a_{i,k})\frac{\partial y_j}{\partial x_l}(\sigma(0),\dots,\sigma(k)) \\
\frac{\partial g_{i,j}}{\partial x_l}(\sigma(0),\dots,\sigma(k)) &= 0 \qquad (44)
\end{aligned}
$$

because

$$\forall i \in [\![1,k]\!], \quad y_i(\sigma(0),\dots,\sigma(k)) - a_{i,k} = 0.$$

From the Taylor series and Lemma 4.2.1, it follows that:

$$
\begin{aligned}
&\mathbb{E}(g_{i,j}(\widehat{\sigma(0)},\widehat{\sigma(1)},\dots,\widehat{\sigma(k)})) \\
&= \begin{cases} \sum_{l=0}^{k}\sum_{m=0}^{k} \frac{\partial g_{i,j}}{\partial x_l \partial x_m}(\sigma(0),\dots,\sigma(k))\mathbb{E}((\widehat{\sigma(l)}-\sigma(l))(\widehat{\sigma(m)}-\sigma(m))) + \mathrm{O}(n^{-3/2}) \\ \text{if } 0 < d \leq 1/4 \\ \sum_{l=0}^{k}\sum_{m=0}^{k} \frac{\partial g_{i,j}}{\partial x_l \partial x_m}(\sigma(0),\dots,\sigma(k))\mathbb{E}((\widehat{\sigma(l)}-\sigma(l))(\widehat{\sigma(m)}-\sigma(m))) + \mathrm{O}(n^{6d-3}) \\ \text{if } 1/4 < d < 1/2 \end{cases}
\end{aligned}
$$

According to the results from Hosking (1996), we shall compute the second term of the Taylor series. First we note that:

$$\frac{\partial g_{i,j}}{\partial x_l \partial x_m}(\sigma(0),\dots,\sigma(k)) = \frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}(\sigma(0),\dots,\sigma(k)) + \frac{\partial y_i}{\partial x_m}\frac{\partial y_j}{\partial x_l}(\sigma(0),\dots,\sigma(k))$$

because the other terms vanish at $(\sigma(0),\dots,\sigma(k))$ by (44). Moreover we can apply Hosking (1996):

$$
\begin{aligned}
&\mathbb{E}((\widehat{\sigma(l)}-\sigma(l))(\widehat{\sigma(m)}-\sigma(m))) \\
&\underset{n\to+\infty}{\sim} \begin{cases} Cn^{4d-2} & \text{if } \frac{1}{4} < d < \frac{1}{2} \\ Dn^{-1}\ln(n) & \text{if } d = \frac{1}{4} \\ n^{-1}\left(\sum_{s=-\infty}^{\infty}(\sigma(s)\sigma(s+l-m)+\sigma(s)\sigma(s+l+m)) + F\sigma(l)\sigma(m)\right) & \text{if } 0 < d < \frac{1}{4} \end{cases}
\end{aligned}
$$

where C, D and F are constants and independent of l and m. Consequently, we can compute:

$$\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial g_{i,j}}{\partial x_l \partial x_m}(\sigma(0),\dots,\sigma(k))\mathbb{E}((\widehat{\sigma(l)}-\sigma(l))(\widehat{\sigma(m)}-\sigma(m))).$$

First we study the case $d \geq 1/4$ and we prove that:

$$\sum_{l=0}^{k}\frac{\partial y_i}{\partial x_l}(\sigma(0),\dots,\sigma(k)) = \left(1 - \sum_{r=1}^{k} a_{r,k}\right)\left(\Sigma_k^{-1}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\right)_i \qquad (45)$$

since if we define $\sigma_0^k := (\sigma(0),\ldots,\sigma(k))$, we may write the partial derivative as:

$$\begin{aligned}\frac{\partial y_i}{\partial x_l}(\sigma_0^k) &= \left(\Sigma_k^{-1}\left[\frac{\partial}{\partial x_l}\begin{pmatrix} x_1 \\ \vdots \\ x_k\end{pmatrix}\right](\sigma_0^k)\right)_i \\ &- \left(\Sigma_k^{-1}\left[\frac{\partial}{\partial x_l}\begin{pmatrix} x_0 & x_1 & \dots & x_k \\ x_1 & x_0 & \ddots & x_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ x_k & x_{k-1} & \dots & x_0\end{pmatrix}\right](\sigma_0^k)\Sigma_k^{-1}\begin{pmatrix}\sigma(1) \\ \vdots \\ \sigma(k)\end{pmatrix}\right)_i . \end{aligned} \tag{46}$$

Using (45), the result follows because:

$$\begin{aligned}&\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial g_{i,j}}{\partial x_l \partial x_m}(\sigma(0),\ldots,\sigma(k))\mathbb{E}((\widehat{\sigma(l)}-\sigma(l))(\widehat{\sigma(m)}-\sigma(m))) \\ = &\begin{cases}\left(1-\sum_{r=1}^k a_{r,k}\right)^2 C n^{4d-2}\left(\Sigma_k^{-1}\mathbf{1}_{k,k}\Sigma_k^{-1}\right)_{(i,j)} & \text{if} \quad \frac{1}{4}<d<\frac{1}{2} \\ \left(1-\sum_{r=1}^k a_{r,k}\right)^2 D n^{-1}\ln(n)\left(\Sigma_k^{-1}\mathbf{1}_{k,k}\Sigma_k^{-1}\right)_{(i,j)} & \text{if} \quad d=\frac{1}{4}.\end{cases}\end{aligned}$$

When $d < 1/4$, we first notice by using (46) that:

$$\sum_{l=0}^{k}\frac{\partial y_i}{\partial x_l}\sigma(l) = \left(-\Sigma_k^{-1}\Sigma_k\Sigma_k^{-1}\begin{pmatrix}\sigma(1)\\ \vdots \\ \sigma(k)\end{pmatrix} + \Sigma_k^{-1}\begin{pmatrix}\sigma(1)\\ \vdots \\ \sigma(k)\end{pmatrix}\right)_i = 0$$

Then it follows that:

$$\begin{aligned}&\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial g_{i,j}}{\partial x_l \partial x_m}(\sigma(0),\ldots,\sigma(k))\mathbb{E}((\widehat{\sigma(l)}-\sigma(l))(\widehat{\sigma(m)}-\sigma(m))) \\ = &\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\sum_{s=-\infty}^{\infty}(\sigma(s)\sigma(s+l-m)+\sigma(s)\sigma(s+l+m)) \\ = &\frac{1}{2}\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\sum_{s=-\infty}^{\infty}(\sigma(s)\sigma(s+l-m)+\sigma(s)\sigma(s+m-l)+\sigma(s)\sigma(s+l+m)+\sigma(s)\sigma(s-l-m)) \\ = &\frac{1}{2}\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\sum_{s=-\infty}^{\infty}\sigma(s)\int_{-\pi}^{\pi}f(\lambda)\mathrm{e}^{is\lambda}(\mathrm{e}^{i(l-m)\lambda}+\mathrm{e}^{i(m-l)\lambda}+\mathrm{e}^{i(m+l)\lambda}+\mathrm{e}^{i(-m-l)\lambda})\mathrm{d}\lambda \\ = &2\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\sum_{s=-\infty}^{\infty}\sigma(s)\int_{-\pi}^{\pi}f(\lambda)\mathrm{e}^{is\lambda}\cos(l\lambda)\cos(m\lambda)\mathrm{d}\lambda \\ = &2\int_{-\pi}^{\pi}\sum_{s=-\infty}^{\infty}\mathrm{e}^{is\lambda}\sigma(s)f(\lambda)\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\cos(l\lambda)\cos(m\lambda)\mathrm{d}\lambda \\ = &2\int_{-\pi}^{\pi}f(\lambda)^2\sum_{l=0}^{k}\sum_{m=0}^{k}\frac{\partial y_i}{\partial x_l}\frac{\partial y_j}{\partial x_m}\cos(l\lambda)\cos(m\lambda)\mathrm{d}\lambda \\ = &2\left(\Sigma_k^{-1}H\Sigma_k^{-1}\right)\end{aligned}$$

with H defined in (42). $\square$

4.3 Proof of Lemma 4.2.1

Show that:

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\right|^3\right] = \begin{cases} \mathrm{O}(n^{-3/2}) & \text{if } d \leq 1/4 \\ \mathrm{O}(n^{6d-3}) & \text{if } d > 1/4 \end{cases} \tag{47}$$

Proof.

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\right|^3\right] \leq \sqrt{\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\right|^2\right]\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\right|^4\right]}$$

We will separately consider the two terms. First we have:

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\right|^2\right] = \sigma(k)^2 - 2\sigma(k)\frac{1}{n}\mathbb{E}\left(\sum_{t=1}^{n-k} X_t X_{t+k}\right) + \frac{1}{n^2}\mathbb{E}\left(\sum_{t=1}^{n-k} X_t X_{t+k}\sum_{s=1}^{n-k} X_s X_{s+k}\right).$$

Since the process is Gaussian, we have (see Triantafyllopoulos (2003)):

$$\mathbb{E}\left(X_t X_{t+k} X_s X_{s+k}\right) = \mathbb{E}\left(X_t X_{t+k}\right)\mathbb{E}\left(X_s X_{s+k}\right) + \mathbb{E}\left(X_t X_s\right)\mathbb{E}\left(X_{t+k} X_{s+k}\right) + \mathbb{E}\left(X_t X_{s+k}\right)\mathbb{E}\left(X_{t+k} X_s\right);$$

and thus:

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\right|^2\right] &= \left(\frac{(n-k)^2}{n^2} - 2\frac{n-k}{n} + 1\right)\sigma(k)^2 \\
&\quad + \frac{1}{n^2}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sigma(t-s)^2 + \sigma(t+k-s)\sigma(s+k-t) \\
&= \frac{k^2}{n^2}\sigma(k)^2 + \frac{1}{n^2}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\left(\sigma(t-s)^2 + \sigma(t+k-s)\sigma(s+k-t)\right)
\end{aligned}$$

We note that

$$\begin{aligned}
\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sigma(t-s)^2 &= (n-k)\sigma(0)^2 + 2\sum_{t=1}^{n-k}(n-k-t)\sigma(t)^2 \\
&= \mathrm{O}(n) + (n-k)\sum_{t=1}^{n-k}\sigma(t)^2 - 2\sum_{t=1}^{n-k} t\sigma(t)^2 \\
&= \mathrm{O}(n) + \mathrm{O}(n^{4d})
\end{aligned}$$

In a similar way for:

$$\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sigma(t+k-s)\sigma(s+k-t)$$

we obtain that:

$$\sqrt{\mathbb{E}\left[\Big|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\Big|^2\right]} = \begin{cases} \mathrm{O}(n^{-1/2}) & \text{if } d \le 1/4 \\ \mathrm{O}(n^{2d-1}) & \text{if } d > 1/4 \end{cases} \tag{48}$$

For the second term, we have:

$$\begin{aligned} \mathbb{E}\left[\Big|\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k)\Big|^4\right] &= \sigma(k)^4 - 4\sigma(k)\mathbb{E}\left[\left(\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k}\right)^3\right] + \frac{6\sigma(k)^2}{n^2}\mathbb{E}\left[\left(\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k}\right)^2\right] \\ & \quad -\frac{4\sigma(k)^3}{n}\mathbb{E}\left(\sum_{t=1}^{n-k} X_t X_{t+k}\right) + \mathbb{E}\left[\left(\frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k}\right)^4\right] \end{aligned}$$

Since the process is Gaussian, we can apply the result in Triantafyllopoulos (2003) and develop the moments as functions which depend only on the covariances of the process. Then we count the order of $\sigma(k)$ in each term of the sum. The coefficient of $\sigma(k)^4$ is:

$$1 - \frac{4(n-k)^3}{n^3} + \frac{6(n-k)^2}{n^2} - \frac{4(n-k)}{n} + \frac{(n-k)^4}{n^4} = \frac{k^4}{n^4};$$

the coefficient of $\sigma(k)^2$ is:

$$\begin{aligned} & \left(\sum_{t=1}^{n-k}\sum_{s=1}^{n-k} \sigma(t-s)^2 + \sigma(t+k-s)\sigma(s+k-t)\right)\left(\frac{-12(n-k)}{n^3} + \frac{6}{n^2} + \frac{6(n-k)^2}{n^4}\right) \\ = \; & \frac{6k^2}{n^4}\left(\sum_{t=1}^{n-k}\sum_{s=1}^{n-k} \sigma(t-s)^2 + \sigma(t+k-s)\sigma(s+k-t)\right) \\ = \; & \begin{cases} \mathrm{O}(n^{-3}) & \text{if } d \le 1/4 \\ \mathrm{O}(n^{-4+4d}) & \text{if } d > 1/4 \end{cases} \end{aligned}$$

and the coefficient of $\sigma(k)$ is:

$$\begin{aligned} & \left(\frac{1}{n^3}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sum_{r=1}^{n-k} 6\sigma(t-s)\sigma(r-s)\sigma(r-t+k) + \sigma(t+k-r)\sigma(s+k-r)\sigma(r+k-s)\right) \\ \times \; & \left(\frac{-4}{n^3} + \frac{4(n-k)}{n^4}\right). \end{aligned}$$

We study this asymptotic behaviour as follows:

$$\begin{aligned}
& \frac{6}{n^3}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sum_{r=1}^{n-k}\sigma(t-s)\sigma(r-s)\sigma(r-t+k) \\
\leq\ & \frac{6}{n^3}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sum_{r=1}^{n-k}|\sigma(t-s)\sigma(r-s)\sigma(r-t+k)| \\
\sim\ & \frac{6}{n^3}\int_1^{n-k}\int_1^{n-k}\int_0^{n-k}|t-s|^{2d-1}|r-s|^{2d-1}|r-t+k|^{2d-1}\mathrm{d}t\mathrm{d}s\mathrm{d}r \\
\leq\ & \frac{6}{n^3}\int_1^{n}\int_1^{n}\int_1^{n}\int_1^{n}|t-s|^{2d-1}|r-s|^{2d-1}|r-t|^{2d-1}\mathrm{d}t\mathrm{d}s\mathrm{d}r \\
\sim\ & 6n^{6d-3}\int_0^{1}\int_0^{1}\int_0^{1}|t-s|^{2d-1}|r-s|^{2d-1}|r-t|^{2d-1}\mathrm{d}t\mathrm{d}s\mathrm{d}r \\
=\ & \mathrm{O}(n^{6d-3})
\end{aligned}$$

The factor of $\sigma(k)$ is bounded by $\mathrm{O}(n^{6d-4})$. The constant terms are either like:

$$\frac{1}{n^4}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sum_{r=1}^{n-k}\sum_{v=1}^{n-k}\sigma(t-s)\sigma(t-r)\sigma(s-v)\sigma(r-v)$$

According to a comparison with an integral, they are bounded by $\mathrm{O}(n^{8d-4})$, or they are like:

$$\frac{1}{n^4}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sum_{r=1}^{n-k}\sum_{v=1}^{n-k}\sigma(t-s)^2\sigma(r-v)^2.$$

We separate the two sums and using the previous results we obtain that:

$$\frac{1}{n^4}\sum_{t=1}^{n-k}\sum_{s=1}^{n-k}\sum_{r=1}^{n-k}\sum_{v=0}^{n-k}\sigma(t-s)^2\sigma(r-v)^2 = \begin{cases}\mathrm{O}(n^{-2}) & \text{if } d \leq 1/4 \\ \mathrm{O}(n^{8d-4}) & \text{if } d > 1/4\end{cases} \tag{49}$$

When we sum the different components, we obtain that:

$$\sqrt{\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k}X_tX_{t+k}-\sigma(k)\right|^4\right]} = \begin{cases}\mathrm{O}(n^{-1}) & \text{if } d \leq 1/4 \\ \mathrm{O}(n^{4d-2}) & \text{if } d > 1/4\end{cases} \tag{50}$$

Finally, we have obtained that:

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^{n-k}X_tX_{t+k}-\sigma(k)\right|^3\right] = \begin{cases}\mathrm{O}(n^{-3/2}) & \text{if } d \leq 1/4 \\ \mathrm{O}(n^{6d-3}) & \text{if } d > 1/4\end{cases} \tag{51}$$

□

References

Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Stat.*, 2:1095–1107.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Stat.*, 2:489–502.

Bhansali, R. (1978). Linear prediction by autoregressive model fitting in the time domain. *Ann. Stat.*, 6:224–231.

Bhansali, R. and Kokoszka, P. (2001). Prediction of long-memory time series: An overview. *Estadística 53, No.160-161, 41-96.*

Billingsley, P. (1968). *Convergence of probability measures. 2nd ed.*

Boettcher, A. and Virtanen, J. (2006). Norms of Toeplitz Matrices with Fisher-Hartwig Symbols. Technical report, arXiv:math.FA/0606016.

Brockwell, P. and Davis, R. (1988). Simple consistent estimation of the coefficients of a linear filter. *Stochastic Processes and their Applications.*

Brockwell, P. and Davis, R. (1991). *Time Series : Theory and Methods.* Springer Series in Statistics.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Stat.*, 14:517–532.

Fox, R. and Taqqu, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Relat. Fields*, 74:213–240.

Fuller, W. A. (1976). *Introduction to statistical time series.* Wiley Series in Probability and Mathematical Statistics. New York.

Granger, C. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1:15–29.

Hosking, J. R. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econom.*, 73(1):261–284.

Inoue, A. (1997). Regularly varying correlation functions and KMO-Langevin equations. *Hokkaido Math. J.*, 26(2):457–482.

Inoue, A. (2000). Asymptotics for the partial autocorrelation function of a stationary process. *J. Anal. Math.*, 81:65–109.

Inoue, A. and Kasahara, Y. (2006). Explicit representation of finite predictor coefficients and its applications. *The Annals of Statistics*, 34.

Lewis, R. and Reinsel, G. (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of multivariate analysis.*

Mandelbrot, B. and Wallis, J. (1969). Some long-run properties of geophysical records. *Water Resource Research*, 5:321–340.

Ray, B. K. (1993). Modeling long-memory processes for optimal long-range prediction. *J. Time Ser. Anal.*, 14(5):511–525.

Simon, B. (2005). *Orthogonal polynomials on the unit circle. Part 1: Classical theory.*

Triantafyllopoulos, K. (2003). On the central moments of the multidimensional Gaussian distribution. *Math. Sci.*, 28(2):125–128.

Whittle, P. (1963). *Prediction and regulation by linear least-square methods. 2nd ed.*

Yajima, Y. (1993). Asymptotic properties of estimates in incorrect ARMA models for long- memory time series. In *New directions in time series analysis. Part II. Proc. Workshop, Minneapolis/MN (USA) 1990, IMA Volumes in Mathematics and Its Applications 46, 375-382.*

Zygmund, A. (1968). *Trigonometric series.* Cambridge University Press.