

LINEAR PREDICTION OF LONG-RANGE DEPENDENT TIME SERIES

FANNY GODET¹

Abstract. We present two approaches for linear prediction of long-memory time series. The first approach consists in truncating the Wiener-Kolmogorov predictor by restricting the observations to the last k terms, which are the only available data in practice. We derive the asymptotic behaviour of the mean-squared error as k tends to $+\infty$. The second predictor is the finite linear least-squares predictor *i.e.* the projection of the forecast value on the last k observations. It is shown that these two predictors converge to the Wiener Kolmogorov predictor at the same rate k^{-1} .

Mathematics Subject Classification. 62M20; 62M10.

Received May 11, 2007. Revised December 31, 2007.

1. INTRODUCTION

ARMA (autoregressive moving-average) processes are often called short-memory processes because their covariances decay geometrically. On the other hand, a long-memory process is characterised by the following feature: the autocovariance function σ decays more slowly in so far as it is not absolutely summable. These processes 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 areas as diverse as Economics, Geophysics, Hydrology or telecom traffic (see, *e.g.*, [14] and [8]). Although there exists substantial literature on the prediction of short-memory processes (see [5] and [2] for the univariate case or [13] for the multivariate case), there are fewer results for long-memory time series. Most prediction methods are successfully applied on long memory data either simulated like in [7] or on real data like electricity demand (see [17]). In this paper, we consider the question of prediction in long memory models from a theoretical point of view.

More precisely, we compare two prediction methods for long-memory processes. We search for linear predictors of X_{k+h} knowing (X_1, \dots, X_k) which are optimal in the sense that they minimise the mean-squared error. We study the behaviour of the mean-squared errors as k tends to infinity as [15] does for short memory processes. The paper is organised as follows. First we introduce our model and our main assumptions. Then, in Section 2, we study the best linear predictor knowing all the past *i.e.* the Wiener-Kolmogorov predictor proposed by [19] and by [3] for long-memory time series. In practice, only the last k values of the process are available. Therefore we need to truncate the infinite series in the definition of the predictor and to derive the asymptotic behaviour of the mean-squared error as k tends to $+\infty$. This method is for example used by [1] to forecast US monetary indices.

Keywords and phrases. Long-memory, linear model, autoregressive process, forecast error.

¹ Laboratoire de Mathématiques Jean Leray, UMR CNRS 6629, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France; fanny.godet@math.univ-nantes.fr

In Section 3 we discuss the asymptotic properties of the forecast error if we project onto the closed span of the last k observations. This approach has been proposed by [16] for fractional noise series and by [7] for other simulated data. These authors view this approach as fitting a misspecified autoregressive model of order k . The simulations in [16] show that high-order AR-models predict fractional integrated noise very well.

Finally in Section 4 we compare the two previous methods for h -step prediction. We give some asymptotic properties of the mean-squared error of the linear least-squares predictor as h tends to $+\infty$ in the particular case of long-memory processes. Then we study our k th order predictors as k tends to $+\infty$.

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

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

where $(\varepsilon_n)_{n \in \mathbb{Z}}$ is a white-noise series consisting of uncorrelated random variables, each with mean zero and variance σ_ε^2 , and where $(b_j)_{j \in \mathbb{N}}$ is square-summable and $b_0 = 1$. We shall further assume that $(X_n)_{n \in \mathbb{Z}}$ admits an infinite autoregressive representation:

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

where the sequence $(a_j)_{j \in \mathbb{N}}$ is absolutely summable and $a_0 = 1$. 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 $\delta > 0$:

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

$$|b_j| \leq C_2 j^{d-1+\delta}, \quad (4)$$

where C_1 and C_2 are constants and d is a parameter verifying $d \in]0, 1/2[$. The class of processes satisfying (3) and (4) includes long memory processes *i.e.* processes with covariances not absolutely summable:

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

For example, the FARIMA process $(X_n)_{n \in \mathbb{Z}}$, the most studied long memory process, is the stationary solution to the difference equations:

$$\phi(B)(1-B)^d X_n = \theta(B)\varepsilon_n \quad (5)$$

where $(\varepsilon_n)_{n \in \mathbb{Z}}$ is a white noise series in L^2 with mean zero, B is the backward shift operator and ϕ and θ are polynomials with no zero in the closed unit disk. The coefficients verify as j goes to infinity

$$\begin{aligned} |a_j| &\underset{+\infty}{\sim} C_1 j^{-d-1} \\ |b_j| &\underset{+\infty}{\sim} C_2 j^{d-1} \end{aligned}$$

and thus (3) and (4) hold.

More generally, conditions (3) and (4) hold when the coefficients are of the form:

$$|a_j| \underset{+\infty}{\sim} L(j)j^{-d-1} \quad (6)$$

$$|b_j| \underset{+\infty}{\sim} L'(j)j^{d-1} \quad (7)$$

where L and L' are slowly varying functions. A positive function L is a slowly varying function in the sense of [20] if, for any $\delta > 0$, $x \mapsto x^{-\delta}L(x)$ is ultimately decreasing and $x \mapsto x^\delta L(x)$ is ultimately increasing.

Under conditions (6) and (7), $(X_n)_{n \in \mathbb{Z}}$ is a long memory process. Indeed [11] proved that (7) implies that:

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

where β is the beta function. Note that the converse is not true. More assumptions about the series $(b_j)_{j \in \mathbb{N}}$ are needed in order to get an asymptotic equivalent for $(b_j)_{j \in \mathbb{N}}$ using an asymptotic equivalent of $(\sigma(j))_{j \in \mathbb{N}}$ (see [12]).

2. WIENER-KOLMOGOROV NEXT STEP PREDICTION THEORY

2.1. The Wiener-Kolmogorov predictor

The aim of this part is to compute the best linear one-step predictor (with minimum mean-squared distance from the true random variable) knowing all the past $\{X_{k+1-j}, j \leq 1\}$. 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 minimise the mean squared prediction error:

$$\mathbb{E} \left[(\widetilde{X}_k(1) - X_{k+1})^2 \right].$$

Since ε_{k+1} is orthogonal to the past $[\varepsilon]_{-\infty}^k = [X]_{-\infty}^k$, the Wiener-Kolmogorov predictor is equal to

$$\widetilde{X}_k(1) = - \sum_{j=1}^{\infty} a_j X_{k+1-j}. \tag{8}$$

2.2. Mean-squared prediction error when the predictor is truncated

In practice, we only know a finite subset 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 $\{X_1, \dots, X_k\}$ and that we replace the unknown values by 0, then we have the following new predictor:

$$\widetilde{X}'_k(1) = - \sum_{j=1}^k a_j X_{k+1-j}. \tag{9}$$

It is equivalent to say that we have truncated the infinite series (8) to k terms. The following proposition provides the asymptotic properties of the mean-squared prediction error as a function of k .

Theorem 2.1. *Let $(X_n)_{n \in \mathbb{Z}}$ be a linear stationary process defined by (1), (2) and verifying conditions (3) and (4). We can approximate the mean-squared prediction error of $\widetilde{X}'_k(1)$ by:*

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

Note that the prediction error is the sum of σ_ε^2 , the error in the Wiener-Kolmogorov model, and the error due to the truncation to k terms which is bounded by $O(k^{-1+\delta})$ for all $\delta > 0$. We denote $\widetilde{\sigma}_k^2 := \mathbb{E}([X_{k+1} - \widetilde{X}'_k(1)]^2)$ hereafter the sum of these two errors.

The proof of Theorem (2.1) hinges on the following lemma:

Lemma 2.2. *Under assumption (4), the autocovariance function σ of the process $(X_n)_{n \in \mathbb{Z}}$ verifies:*

$$\forall \delta > 0, \exists C_3 \in \mathbb{R}, \quad |\sigma(j)| \leq C_3 j^{2d-1+\delta}. \tag{10}$$

Proof. Notice that it suffices to prove (10) for δ near 0 in order to prove (10) for all $\delta > 0$. Let $\delta < 1 - 2d$:

$$\begin{aligned}
\sigma(k) &= \sum_{j=0}^{+\infty} b_j b_{j+k} \\
|\sigma(k)| &\leq \sum_{j=1}^{+\infty} |b_j b_{j+k}| + |b_0 b_k| \\
&\leq C_2^2 \sum_{j=1}^{+\infty} j^{d-1+\delta/2} (k+j)^{d-1+\delta/2} + |b_0 b_k| \\
&\leq C_2^2 k^{2d-1+\delta} \int_0^{+\infty} j^{d-1+\delta/2} (1+j)^{d-1+\delta/2} dj + C_2 k^{d-1+\delta/2} \\
&\leq C_3 k^{2d-1+\delta}.
\end{aligned}$$

□

We now prove Theorem 2.1.

Proof.

$$X_{k+1} - \widetilde{X}'_k(1) = \varepsilon_{k+1} - \sum_{j=k+1}^{+\infty} a_j X_{k+1-j}. \quad (11)$$

The two parts of the sum (11) are orthogonal with respect to the inner product associated with the mean-squared norm. Consequently:

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

The error due to the truncation, which we have to study, is then equal to:

$$\widetilde{\sigma}_k^2 - \sigma_\varepsilon^2 = \sum_{j=k+1}^{\infty} \sum_{l=k+1}^{\infty} a_j a_l \sigma(l-j) \quad (12)$$

we have:

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

From (3) and (10) and the triangular inequality, it follows that:

$$\begin{aligned}
\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) &\leq C_1^2 C_3 \left(2 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta} (j+1)^{-d-1+\delta} + \sum_{j=k+1}^{+\infty} (j^{-d-1+\delta})^2 \right) \\
&\quad + 2C_1^2 C_3 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta} \sum_{l=j+2}^{+\infty} l^{-d-1+\delta} |l-j|^{2d-1+\delta}
\end{aligned} \quad (13)$$

for all $\delta > 0$. Assume now that $\delta < 1/2 - d$. For the terms (13), 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:

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

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

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

For the infinite double series (13), we will similarly compare the series with an integral. In the next lemma, we establish a necessary result for this comparison:

Lemma 2.3. *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 $\delta \in]0, \frac{\delta-d-1}{\delta+2d-1}[$. If $n \geq m + 1$ then*

$$\int_{A_{n,m}} g(l, j) dj dl \geq g(n+1, m)$$

where $A_{n,m}$ is the square domain $[n, n+1] \times [m, m+1]$.

Assume now that $\delta < 1 - 2d$ without loss of generality. Thanks to the previous Lemma and the asymptotic equivalents of (13), there exists $K \in \mathbb{N}$ such that if $k > K$:

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

By using the substitution $jl' = l$ in the integral over l we obtain:

$$\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) \leq C' \int_{k+1}^{+\infty} j^{-2+3\delta} \int_1^{+\infty} l^{-d-1+\delta} (l-1)^{2d-1+\delta} dl dj + O(k^{-2d-1}).$$

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

$$\int_1^{+\infty} l^{-d-1+\delta} (l-1)^{2d-1+\delta} dl < +\infty,$$

it follows:

$$\begin{aligned} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) &\leq O(k^{-1+3\delta}) + O(k^{-2d-1}) \\ &\leq O(k^{-1+3\delta}). \end{aligned} \tag{14}$$

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

$$\tilde{\sigma}_k^2 - \sigma_\varepsilon^2 = \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) = O(k^{-1+3\delta}).$$

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 $\delta > 0$. This concludes the proof. \square

The result of Theorem 2.1 is not interesting for short memory processes. It is proved in [15] that the error due to the truncation $(\tilde{\sigma}_k^2 - \sigma_\varepsilon^2)$ converges to 0 at an exponential rate for short memory processes. We now prove that there exist long-memory processes whose prediction error attains the rate of convergence k^{-1} .

When we specify the form of the autoregressive coefficients and the autocovariance function, we obtain an equivalent of the mean-squared prediction error.

Proposition 2.4. *Let $(X_n)_{n \in \mathbb{Z}}$ be a linear stationary process defined by (1), (2) and verifying (6). Assume that the coefficients $(a_j)_{j \in \mathbb{N}^*}$ are ultimately of constant sign. We assume also that the autocovariance function σ of the process $(X_n)_{n \in \mathbb{Z}}$ verifies:*

$$|\sigma(j)| \underset{+\infty}{\sim} L'(j)j^{2d-1}$$

with L' a slowly varying function and that the function σ is of a constant sign. Then the rate of convergence $O(k^{-1})$ is optimal i.e. we have the following asymptotic equivalent:

$$\mathbb{E}([X_{k+1} - \widetilde{X}'_k(1)]^2) = \sigma_\varepsilon^2 + L''(k)k^{-1} + o(k^{-1}) \text{ as } k \rightarrow +\infty \quad (15)$$

where L'' is a slowly varying function.

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

$$\begin{aligned} \left| \sum_{k+1}^{+\infty} \sum_{k+1}^{+\infty} a_j a_l \sigma(l-j) \right| &= \sum_{k+1}^{+\infty} |a_j| \sum_{j+1}^{+\infty} |a_l| |\sigma(l-j)| + \sum_{k+1}^{+\infty} a_j^2 \sigma(0) \\ &= \int_{k+1}^{+\infty} j^{-d-1} L(j) \int_{j+1}^{+\infty} l^{-d-1} L(l) (l-j)^{2d-1} L'(l-j) dl dj \\ &\quad + \frac{k^{-2d-1} L^2(k)}{1+2d} + O(k^{-2d-1}) \\ &\sim \int_{k+1}^{+\infty} j^{-2} L^2(j) L'(j) \int_{1/j+1}^{+\infty} l^{-d-1} \frac{L(lj)}{L(j)} (l-1)^{2d-1} \frac{L'(j(l-1))}{L'(j)} dl dj \\ &\sim k^{-1} \frac{L^2(k) L'(k) \Gamma(1-d) \Gamma(2d)}{\Gamma(1+d)}. \end{aligned} \quad (16)$$

\square

Example 2.5. Assume now that $(X_n)_{n \in \mathbb{Z}}$ is fractionally integrated noise, which is the stationary solution of the difference equation:

$$X_n = (1 - B)^{-d} \varepsilon_n \quad (17)$$

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 [6]). This is a particular case of FARIMA models defined in (5) with $\phi = \theta = 1$. We can compute the coefficients and obtain that:

$$\forall j > 0, \quad a_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} < 0$$

and

$$a_j \sim \frac{j^{-d-1}}{\Gamma(-d)} \quad \text{as } j \rightarrow \infty.$$

We can also compute the autocovariance function and we obtain:

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

and

$$\sigma(j) \sim \frac{j^{2d-1} \Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \sigma_\varepsilon^2 \quad \text{as } j \rightarrow \infty.$$

In this particular case, we can estimate the prediction error more precisely and give the explicit expression of L'' defined in (15):

$$\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) \sim \frac{\Gamma(1-2d)\Gamma(2d)\sigma_\varepsilon^2}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)} k^{-1} \quad (18)$$

since in this case the functions L and L' are constant and equal respectively to $\frac{1}{\Gamma(-d)}$ and to $\frac{\Gamma(1-2d)\sigma_\varepsilon^2}{\Gamma(d)\Gamma(1-d)}$.

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

$$\mathbb{E}([X_{k+1} - \widetilde{X}'_k(1)]^2) = \sigma_\varepsilon^2 + C(d)k^{-1} + o(k^{-1})$$

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

$$C(d) = \frac{\Gamma(1-2d)\Gamma(2d)\sigma_\varepsilon^2}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)}. \quad (19)$$

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

$$C(d) \sim \frac{\sigma_\varepsilon^2}{(1-2d)\Gamma(-1/2)^2\Gamma(1/2)\Gamma(3/2)}. \quad (20)$$

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

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

As Figure 1 suggests and the asymptotic equivalent given in (20) proves, the mean-squared error tends to $+\infty$ as $d \rightarrow 1/2$. 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 $d \rightarrow 1/2$. This result is not surprising since the correlation between the predicted variable and the unobserved variables increases when $d \rightarrow 1/2$.

Example 2.6. Long-memory processes with seasonal effects (see for example [18]) provide another class of long memory processes verifying assumptions of Theorem 2.1 but not those of Proposition 2.4. We can also prove that the rate of convergence k^{-1} is attained for some seasonal processes. We consider the particular case of GARMA processes introduced by [10] and studied for example by [9]. We first recall their definition. $(X_n)_{n \in \mathbb{Z}}$ is a GARMA process if it is the stationary solution of the equation:

$$\forall n \in \mathbb{Z}, X_n = (1 - 2 \cos \nu B + B^2)^{-d} \varepsilon_n$$

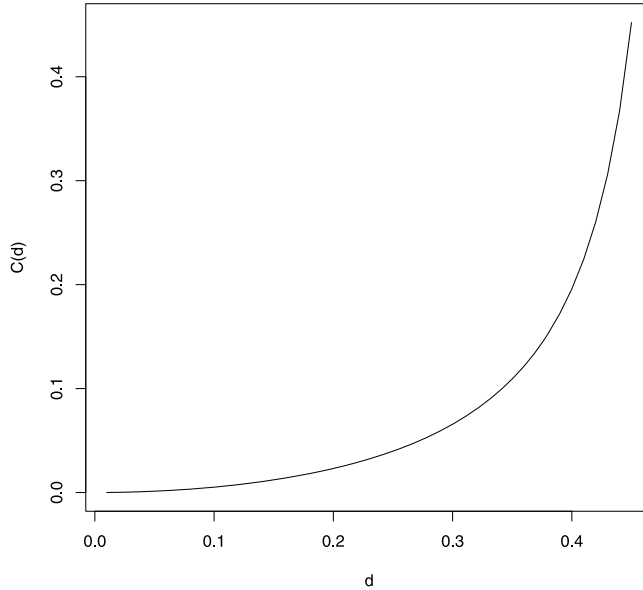


FIGURE 1. Behaviour of constant $C(d)$, $d \in [0, 1/2[$, defined in (19).

with $\cos \nu \neq 0$, $d \in]0, 1/2[$, $(\varepsilon_n)_{n \in \mathbb{Z}}$ a white noise and B the backward shift operator. $(X_n)_{n \in \mathbb{Z}}$ admits an infinite autoregressive representation like in (2). The coefficients verify:

$$a_j = \frac{\cos[(j-d)\nu + (d\pi/2)]}{\Gamma(-d) \sin^{-d}(\nu)} \left(\frac{2}{j}\right)^{1+d} + O\left(j^{-5/2}\right). \quad (21)$$

Its autocovariance function σ is such that:

$$\sigma(j) = K \cos(j\nu) j^{2d-1} (1 + O(j^{-1})) \quad (22)$$

where K is a constant independent of j .

We now give the asymptotic behaviour of the mean-squared error of prediction for these processes.

Proposition 2.7. *If $(X_n)_{n \in \mathbb{Z}}$ is a GARMA process, then*

$$\mathbb{E}([X_{k+1} - \widetilde{X}'_k(1)]^2) = \sigma_\varepsilon^2 + Ck^{-1} + o(k^{-1}) \quad \text{as } k \rightarrow +\infty.$$

Proof. The proof is given in Appendix. □

After these results on the truncated Wiener-Kolmogorov predictor, we will propose another predictor. Instead of truncating to k terms the series which defines the Wiener-Kolmogorov predictor it is more efficient to project directly onto the observations in order to minimise the forecast error. We study this method in the following section.

3. PROJECTION METHOD

In this section, we study the asymptotic properties of the forecast mean-squared error when we project the forecast random variable X_{k+1} onto the closed span of the subset $\{X_k, \dots, X_1\}$. This is a generalisation of the “autoregressive model fitting” approach developed by [16] in the case of fractionally integrated noise (defined in (17)). Fitting a k -th order autoregressive process is equivalent to projecting onto the span of the subset $\{X_k, \dots, X_1\}$.

3.1. Definition of the predictor

Here $(X_n)_{n \in \mathbb{Z}}$ is a long-memory process which verifies the assumptions (1)–(4) of Section 1. We choose to define the predictor as the projection mapping onto the closed span of the subset $\{X_k, \dots, X_1\}$ of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with inner product $\langle X, Y \rangle = \mathbb{E}(XY)$. Then the predictor can be written as:

$$\widehat{X}_k(1) = -a_{1,k}X_k - \dots - a_{k,k}X_1.$$

Consequently the coefficients $(-a_{i,k})_{1 \leq i \leq k}$ of the projection verify the k th order Yule-Walker equations:

$$\forall j \in \llbracket 1, k \rrbracket, \quad \sum_{i=1}^k a_{i,k} \sigma(i-j) = -\sigma(j). \tag{23}$$

The mean-squared prediction error is:

$$\mathbb{E} \left[(\widehat{X}_k(1) - X_{k+1})^2 \right] = \sigma_\varepsilon^2 + \mathbb{E} \left[\left(-\sum_{j=1}^{\infty} a_j X_{k+1-j} + \sum_{j=1}^k a_{j,k} X_{k+1-j} \right)^2 \right]. \tag{24}$$

3.2. Rate of convergence of the error by projection

In the next theorem we derive an asymptotic expression for the prediction error when projecting onto the finite past:

Theorem 3.1. *Let $(X_n)_{n \in \mathbb{Z}}$ be a linear stationary process defined by (1), (2) and verifying conditions (3) and (4) of Section 1. We can approximate the mean-squared prediction error of $\widehat{X}_k(1)$ by:*

$$\mathbb{E}[(\widehat{X}_k(1) - X_{k+1})^2] - \sigma_\varepsilon^2 = O(k^{-1}) \text{ as } k \rightarrow +\infty. \tag{25}$$

There exist long-memory processes such that:

$$\mathbb{E}[(\widehat{X}_k(1) - X_{k+1})^2] - \sigma_\varepsilon^2 \sim Ck^{-1} \text{ as } k \rightarrow +\infty$$

where C is a constant.

Proof. Hereafter we will denote $\widehat{\sigma}_k^2 := \mathbb{E}[(\widehat{X}_k(1) - X_{k+1})^2]$. Since the projection minimises the forecast error using k observations, the error by using truncation is bigger and since the truncation method involves an error bounded by $O(k^{-1})$, we obtain (25).

Consequently we only need prove that this rate of convergence is attained. This is the case for the fractionally integrated processes defined in (17). We want to express the error made by projection in terms of the Wiener-Kolmogorov truncation error. Therefore in the case of a fractionally integrated process we need only show that:

$$\widehat{\sigma}_k^2 - \sigma_\varepsilon^2 \sim Ck^{-1}.$$

Now, setting $a_{j,k} = 0$ if $j > k$ and developping (24)

$$\begin{aligned} \widehat{\sigma}_k^2 - \sigma_\varepsilon^2 &= \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} (a_j - a_{j,k})(a_l - a_{l,k}) \sigma(l-j) \\ &= \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sum_{l=0}^{+\infty} a_l \sigma(l-j) - \sum_{l=0}^k a_{l,k} \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sigma(l-j). \end{aligned} \tag{26}$$

We first study the first term of the sum (26). For any $j > 0$, we have $\sum_{l=0}^{+\infty} a_l \sigma(l-j) = 0$:

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

as ε_n is orthogonal to X_{n-j} for $j > 0$. We can thus rewrite the first term of (26):

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

since $a_0 = a_{0,k} = 1$ according to definition. Next we study the second term of the sum in (26). And we obtain:

$$\begin{aligned}\sum_{l=0}^k a_{l,k} \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sigma(l-j) &= \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=1}^k (a_j - a_{j,k}) \sigma(l-j) \\ &\quad + \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=k+1}^{+\infty} a_j \sigma(l-j) \\ &\quad + \sum_{l=0}^k a_l \sum_{j=1}^k (a_j - a_{j,k}) \sigma(l-j) \\ &\quad + \sum_{l=0}^k a_l \sum_{j=k+1}^{+\infty} a_j \sigma(l-j).\end{aligned}\tag{27}$$

Similarly we rewrite the term (27) using the Yule-Walker equations:

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

We then remark that (27) is equal to the opposite of (27). Hence it follows that:

$$\begin{aligned}\sum_{l=0}^k a_{l,k} \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sigma(l-j) &= \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=1}^k (a_j - a_{j,k}) \sigma(l-j) \\ &\quad + 2 \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=k+1}^{+\infty} a_j \sigma(l-j) \\ &\quad + \sum_{l=0}^k a_l \sum_{j=k+1}^{+\infty} a_j \sigma(l-j).\end{aligned}\tag{28}$$

In a similar way we can rewrite the third term of the sum in (28) using Fubini's Theorem and then the Yule-Walker equations:

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

This third term is therefore equal to the forecast error due to the truncation (12).

In order to compare the prediction error when truncating the Wiener-Kolmogorov predictor and when projecting onto the finite past for a fractionally integrated process, we need the sign of all the components of the sum in (28). 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 $a_l - a_{l,k}$ we use the explicit formula given in [5] and we easily obtain that $a_l - a_{l,k}$ is negative for all $l \in \llbracket 1, k \rrbracket$.

$$\begin{aligned} a_l - a_{l,k} &= \frac{\Gamma(l-d)}{\Gamma(l+1)\Gamma(-d)} - \frac{\Gamma(k+1)\Gamma(l-d)\Gamma(k-d-l+1)}{\Gamma(k-l+1)\Gamma(l+1)\Gamma(-d)\Gamma(k-d+1)} \\ &= -a_l \left(-1 + \frac{\Gamma(k+1)\Gamma(k-d-l+1)}{\Gamma(k-l+1)\Gamma(k-d+1)} \right) \\ &= -a_l \left(\frac{k \dots (k-j+1)}{(k-d) \dots (k-d-j+1)} - 1 \right) > 0 \end{aligned} \quad (29)$$

since $\forall j \in \mathbb{N}^*$, $a_j < 0$. To give an asymptotic equivalent for the prediction error, we use the sum given in (28). 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 $O(k^{-1})$. The prediction error when projecting 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 $a_l - a_{l,k}$ using (29):

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

Then we use the following inequality:

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

which gives:

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

According to the previous inequality, we have:

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

As the function $x \mapsto \frac{1}{k-d-x}$ is increasing, we can bound the series by the integrals.

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

We now search a bound for $\exp\left(d \int_{m=0}^{k-1} \frac{1}{k-d-m} dm\right)$. We have:

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

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

$$\begin{aligned}
\sum_{l=1}^k (a_l - a_{l,k}) \sum_{j=k+1}^{+\infty} -a_j \sigma(l-j) &\leq \sum_{l=1}^{k-1} -a_l \exp\left(d \ln\left(\frac{k-d}{k-d-l}\right)\right) \sum_{j=k+1}^{+\infty} -a_j \sigma(l-j) \\
&\quad - a_k k^{\frac{3}{2}d} \sum_{j=k+1}^{+\infty} -a_j \sigma(0) \\
&\leq C(k-d)^d \sum_{l=1}^{k-1} l^{-d-1} (k-d-l)^{-d} \sum_{j=k+1}^{+\infty} j^{-d-1} (j-l)^{2d-1} \\
&\quad + Ck^{-d-1} k^{\frac{3}{2}d} k^{-d} \\
&\leq \frac{C}{(k-d)^2} \int_{1/(k-d)}^1 l^{-d-1} (1-l)^{-d} \left(\int_1^{+\infty} j^{-d-1} (j-1)^{2d-1} dl\right) dj \\
&\quad + Ck^{-\frac{1}{2}d-1} \\
&\leq C'(k-d)^{-2+d} + Ck^{-\frac{1}{2}d-1}
\end{aligned}$$

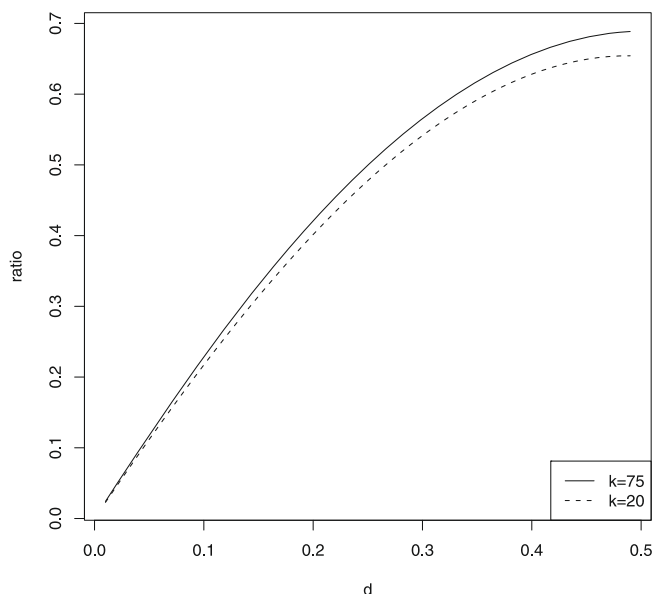


FIGURE 2. Ratio $r(k)$, $d \in]0, 1/2[$ defined in (30).

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 fractionally integrated processes, the two prediction errors are equivalent to Ck^{-1} with C constant. \square

The two approaches to next-step prediction, by truncation to k terms or by projection on the last k observations 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 and if the improvement is negligible with respect to the prediction error due to the truncation. In this aim, we define the following quotient:

$$r(k) := \frac{(\tilde{\sigma}_k^2 - \sigma_\varepsilon^2) - (\hat{\sigma}_k^2 - \sigma_\varepsilon^2)}{\tilde{\sigma}_k^2 - \sigma_\varepsilon^2} \tag{30}$$

which is the ratio of the difference between the two prediction errors (prediction error due to the truncation (12) and prediction error due to the projection (26)) and the prediction error due to the truncation. By (28), we obtain:

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

We evaluate this ratio in the particular case of a fractionally integrated noise defined in (17). Figure 2 shows that the prediction by truncation incurs a larger performance loss when $d \rightarrow 1/2$. The improvement reaches 50 percent when $d > 0.3$ and $k > 20$.

4. THE MULTISTEP PREDICTORS

After obtaining an asymptotic equivalent for the next step predictor, we will generalise the two methods to h -step prediction and aim to obtain their asymptotic behaviour as $k \rightarrow +\infty$ but also as $h \rightarrow +\infty$.

4.1. The Wiener-Kolmogorov predictor

Since we assume that the process $(X_n)_{n \in \mathbb{Z}}$ has an autoregressive representation (2) and moving average representation (1), the linear least-squares predictor, $\tilde{X}_k(h)$, of X_{k+h} based on the infinite past $(X_j, j \leq k)$ is

given by:

$$\widetilde{X}_k(h) = \sum_{j=h}^{+\infty} b_j \varepsilon_{k+h-j}$$

(see for example Th. 5.5.1 of [6]). The corresponding mean squared error of prediction is:

$$\mathbb{E} \left[\left(\widetilde{X}_k(h) - X_{k+h} \right)^2 \right] = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} b_j^2.$$

As the prediction step h tends to infinity, the mean-squared prediction error converges to $\sigma_\varepsilon^2 \sum_{j=0}^{+\infty} b_j^2 = \sigma(0)$, which is the variance of the process $(X_n)_{n \in \mathbb{Z}}$. But if the mean-squared prediction error is equal to $\sigma(0)$, we have no more interest in the prediction method since its error is equal to the error of predicting the future by 0. The convergence of the mean-squared error to $\sigma(0)$ is slower in the long-memory case than in the short-memory case since the sequence b_j decays more slowly to 0. More precisely in the case of a long-memory process under the assumption (7) we can explicit the asymptotic behaviour of the mean-squared error. We have as h tends to infinity:

$$\begin{aligned} \sigma(0) - \mathbb{E} \left[\left(\widetilde{X}_k(h) - X_{k+h} \right)^2 \right] &= \sigma_\varepsilon^2 \sum_{j=h}^{+\infty} b_j^2 \\ &\sim \sum_{j=h}^{+\infty} j^{2d-2} L'^2(j) \\ &\sim \int_h^{+\infty} j^{2d-2} L'^2(j) dj \\ &\underset{h \rightarrow +\infty}{\sim} \frac{1}{1-2d} h^{2d-1} L'^2(h) \end{aligned} \quad (31)$$

according to Proposition 1.5.10 of [4].

On the contrary, for short memory processes we can obtain an exponential rate of convergence of $\mathbb{E} \left[\left(\widetilde{X}_k(h) - X_{k+h} \right)^2 \right]$ to the variance of the process $\sigma(0)$. Consider the classical ARMA processes (see for example [6]). The moving average coefficients b_j are bounded by:

$$|b_j| \leq C j^{m-1} \rho^{-j}$$

where ρ verifies $|\rho| < 1$ and m is a non negative integer (see for example [6] p. 92). Thus the mean-squared prediction error is equivalent to:

$$\begin{aligned} \sigma(0) - \mathbb{E} \left[\left(\widetilde{X}_k(h) - X_{k+h} \right)^2 \right] &\underset{h \rightarrow +\infty}{\sim} \sigma_\varepsilon^2 C \sum_{j=h}^{+\infty} j^{2m-2} \rho^{-2j} \\ &\underset{h \rightarrow +\infty}{\sim} \sigma_\varepsilon^2 C (2 \log(\rho))^{1-2m} (2 \log(\rho) h)^{2m-2} \exp(2 \log(\rho) h). \end{aligned}$$

We obtain that the rate of convergence is exponential. The mean-squared prediction error goes faster to $\sigma(0)$ when the predicting process is ARMA than when the process is a long-memory process.

The h -step prediction is then more interesting for the long-memory process than for the short-memory process, having observed the infinite past. We now consider the truncating effect and the projection effect.

4.2. Truncated Wiener-Kolmogorov predictor and the linear least-squares predictor

In practice, we only observe a finite number k of data (X_1, \dots, X_k) so we cannot evaluate the Wiener-Kolmogorov predictor. We first study what happens when we truncate the series which defines this predictor.

We define the h -step truncated Wiener-Kolmogorov predictor of order k by induction as:

$$\widetilde{X}'_k(h) = - \sum_{j=1}^{h-1} a_j \widetilde{X}'_k(h-j) - \sum_{j=1}^k a_{h-1+j} X_{k+1-j}. \quad (32)$$

The following proposition describes the asymptotic behaviour of the mean-squared error of the predictor (32).

Proposition 4.1. *Let $(X_n)_{n \in \mathbb{Z}}$ be a linear stationary process defined by (1), (2) and verifying (3) and (4). We can approximate the mean-squared prediction error of $\widetilde{X}'_k(h)$ by:*

$$\forall \delta > 0, \mathbb{E} \left[\widetilde{X}'_k(h) - X_{k+h} \right]^2 = \sigma_\varepsilon^2 \sum_{l=0}^{h-1} b_l^2 + R(h, k) \quad (33)$$

where $R(h, k) \geq 0$ and where for all h and k ,

$$R(h, k) \leq Ch^{2d+\delta} k^{-1+\delta}$$

with C a constant independent of (h, k) .

Proof. First we write the difference between the predicted random variable and its predictor:

$$\begin{aligned} \widetilde{X}'_k(h) - X_{k+h} &= - \sum_{j=1}^{h-1} a_j \widetilde{X}'_k(h-j) - \sum_{l=1}^k a_{h-1+l} X_{k+1-j} - \varepsilon_{k+h} + \sum_{j=1}^{+\infty} a_j X_{k+h-j} \\ &= -\varepsilon_{k+h} + \sum_{j=1}^{h-1} a_j \left(X_{k+h-j} - \widetilde{X}'_k(h-j) \right) + \sum_{j=1}^k a_{h-1+j} (X_{k+1-j} - X_{k+1-j}) \\ &\quad + \sum_{j=k+1}^{+\infty} a_{h-1+j} X_{k+1-j} \\ &= -\varepsilon_{k+h} + \sum_{j=1}^{h-1} a_j \left(X_{k+h-j} - \widetilde{X}'_k(h-j) \right) + \sum_{j=k+1}^{+\infty} a_{h-1+j} X_{k+1-j}. \end{aligned}$$

We proceed by induction on h to show that

$$\begin{aligned} \widetilde{X}'_k(h) - X_{k+h} &= - \sum_{l=0}^{h-1} \left(\sum_{j_1+j_2+\dots+j_h=l} (-1)^{\text{card}(\{i, j_i \neq 0\})} a_{j_1} a_{j_2} \dots a_{j_h} \right) \varepsilon_{k+h-l} \\ &\quad + \sum_{j=k+1}^{+\infty} \left(\sum_{i_1+i_2+\dots+i_h=h-1} (-1)^{\text{card}(\{l>1, i_l \neq 0\})} a_{j+i_1} a_{i_2} \dots a_{i_h} \right) X_{k+1-j}. \end{aligned}$$

For $h = 2$, we have for example

$$\widetilde{X}'_k(2) - X_{k+2} = -(a_0 \varepsilon_{k+2} - a_1 \varepsilon_{k+1}) + \sum_{j=k+1}^{+\infty} (-a_1 a_j + a_{j+1}) X_{k+1-j}.$$

Denote $A(z) = 1 + \sum_{j=1}^{+\infty} a_j z^j$ and $B(z) = 1 + \sum_{j=1}^{+\infty} b_j z^j$. Since we have $A(z) = B(z)^{-1}$, we obtain the following conditions on the coefficients:

$$\begin{aligned} b_1 &= -a_1 \\ b_2 &= -a_2 + a_1^2 \\ b_3 &= -a_3 + 2a_1 a_2 - a_1^3 \\ &\dots \end{aligned}$$

So we obtain:

$$\widetilde{X}'_k(h) - X_{k+h} = - \sum_{l=0}^{h-1} b_l \varepsilon_{k+h-l} + \sum_{j=k+1}^{+\infty} \sum_{m=0}^{h-1} a_{j+m} b_{h-1-m} X_{k+1-j}. \quad (34)$$

Since the process $(\varepsilon_n)_{n \in \mathbb{Z}}$ is uncorrelated, the two terms of the sum (34) are orthogonal. We can rewrite the mean-squared error:

$$\mathbb{E} \left[\widetilde{X}'_k(h) - X_{k+h} \right]^2 = \sum_{l=0}^{h-1} b_l^2 \sigma_\varepsilon^2 + \mathbb{E} \left[\sum_{j=k+1}^{+\infty} \left(\sum_{m=0}^{h-1} a_{j+m} b_{h-1-m} \right) X_{k+1-j} \right]^2. \quad (35)$$

The first part of the error (35) is due to the prediction method and the second (35) due to truncating the predictor. We now approximate the error term (35) by using (3) and (4). We obtain the following upper bound:

$$\begin{aligned} \forall \delta > 0, \left| \sum_{m=0}^{h-1} a_{j+h-1-m} b_m \right| &\leq \sum_{m=1}^{h-1} |a_{j+h-1-m} b_m| + |b_0 a_{j+h-1}| \\ &\leq C_1 C_2 \int_0^h (j+h-1-l)^{-d-1+\delta} l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ &\leq C_1 C_2 h^{-1+2\delta} \int_0^1 \left(\frac{j}{h} + 1 - l \right)^{-d-1+\delta} l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ &\leq C_1 C_2 h^{-1+2\delta} j^{-d-1+\delta} \int_0^1 \left(\frac{1}{h} + \frac{1-l}{j} \right)^{-d-1+\delta} l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ &\leq C_1 C_2 h^{d+2\delta} j^{-d-1+\delta} \int_0^1 l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ \left| \sum_{m=0}^h a_{j+h-m} b_m \right| &\leq C_1 C_2 \frac{h^{d+2\delta}}{d} j^{-d-1+\delta} \end{aligned} \quad (36)$$

where the constants C_1 and C_2 are independent of (h, j) . We then obtain by (36):

$$\begin{aligned} \mathbb{E} \left[\sum_{j=k+1}^{+\infty} \left(\sum_{m=0}^{h-1} a_{j+h-1-m} b_m \right) X_{k+1-j} \right]^2 &\leq \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} \left(\sum_{m=0}^{h-1} a_{j+h-1-m} b_m \right) \left(\sum_{m=0}^{h-1} a_{l+h-1-m} b_m \right) \sigma(j-l) \\ &\leq \left(C_1 C_2 \frac{h^{d+2\delta}}{d} \right)^2 \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} j^{-d-1+\delta} l^{-d-1+\delta} \sigma(j-l). \end{aligned} \quad (37)$$

The series in (37) is independent of h and using Theorem 2.1 we have that it is bounded by $Ck^{-1+\delta}$ with C independent of k . This concludes the proof. \square

Remark 4.2. Gathering all the above results about the truncated h -step predictor $\widetilde{X}'_k(h)$, the variance of prediction error is

$$\mathbb{E}(X_{k+h} - \widetilde{X}'_k(h))^2 = \sigma(0) - \sigma_\varepsilon^2 \sum_{l=h}^{+\infty} b_l^2 + R(h, k)$$

where $\sigma(0) - \sigma_\varepsilon^2 \sum_{l=h}^{+\infty} b_l^2$ is the optimal mean-squared error of linear prediction while $R(h, k)$ quantifies the truncating effect. Truncation always increases the variance of forecasting error but it is possible to keep its effect as small as possible. Indeed, taking (31) into account, the condition

$$h^{2d} k^{-1} = o(h^{2d-1}), \tag{38}$$

which is satisfied if $h = o(k)$, insures that the truncating effect is negligible with respect to the quality of the optimal linear predictor $\widetilde{X}'_k(h)$.

In other words, long-step prediction in a long-memory context is interesting when the prediction is based on a sufficiently large sample (X_1, \dots, X_k) .

We can compare the prediction error of the truncated Wiener-Kolmogorov predictor with the least-squares predictor defined as the projection on the finite past (X_1, \dots, X_k) . Let $\widehat{X}_k(h)$ the predictor of X_{k+h} . It then can be written as:

$$\widehat{X}_k(h) = - \sum_{j=1}^k c_{j,k} X_{k+1-j},$$

where the vector $(c_{j,k})_{1 \leq j \leq k}$ minimises the mean-squared error:

$$\mathbb{E} \left[\widehat{X}_k(h) - X_{k+h} \right]^2.$$

Thus the coefficients $(c_{j,k})_{1 \leq j \leq k}$ satisfy the following relation:

$$(c_{j,k})_{1 \leq j \leq k} = -\Sigma_k^{-1} (\sigma_{h-1+j})_{1 \leq j \leq k} \tag{39}$$

where Σ_k is the autocovariance matrix of the vector (X_1, \dots, X_k) *i.e.*

$$\Sigma_k = \begin{pmatrix} \sigma(0) & \sigma(1) & \dots & \sigma(k-1) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(k-1) & \sigma(k-2) & \dots & \sigma(0) \end{pmatrix}.$$

By construction of $\widehat{X}_k(h)$, the mean-squared error of prediction is lower than $\sigma(0)$ for any k and h . However it can be interesting to search for a condition on k and h to have the prediction error due to the projection negligible with respect to the information (31) given by the Wiener-Kolmogorov predictor.

As $\widehat{X}_k(h)$ is the projection of X_{k+h} onto (X_1, \dots, X_k) , the mean-squared prediction error is also lower than the prediction error of the truncated Wiener-Kolmogorov predictor (see Fig. 3). The mean-squared error of prediction due to the projection onto the span of (X_1, \dots, X_k) tends at least as fast to zero as the mean-squared error due to truncation of the least-squares predictor. We obtain that under the assumptions of Proposition 4.1 and condition (38), the prediction error due to the projection is negligible with respect to the information (31) given by the Wiener-Kolmogorov predictor. In Figure 3, we show that the behaviours of the truncated predictor and of the least-squares predictor are similar. In both cases, the prediction error due to the prediction method is sufficiently small with respect to the available information (31) to obtain a good predictor in the case of fractionally integrated noise defined in (17) with $k = 80$ and $h \in [1, 150]$.

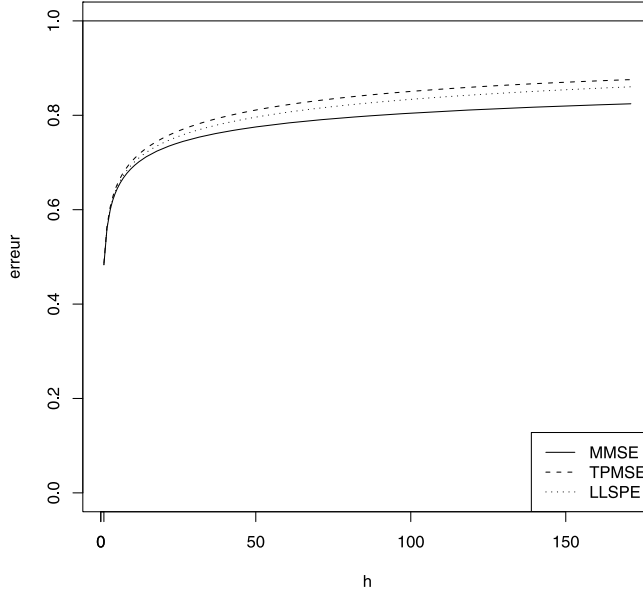


FIGURE 3. Mean-squared error of $\widetilde{X}_k(h)$ (MMSE), $\widetilde{X}'_k(h)$ (TPMSE) and $\widehat{X}_k(h)$ (LLSPE) for $d = 0.4$ and $k = 80$ for fractionally integrated noise defined in (17).

5. APPENDIX: PROOF OF PROPOSITION 2.7

We have to find an asymptotic equivalent of $\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j)$. First we replace the coefficients and the autocovariances by their asymptotic expressions given in (21) and (22). We obtain:

$$\begin{aligned}
\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) &= \sum_{j=k+1}^{+\infty} a_j^2 \sigma(0) + 2 \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} a_j a_l \sigma(l-j) \\
&= \sum_{j=k+1}^{+\infty} \left[\left(\frac{\cos[(j-d)\nu + (d\pi/2)]}{\Gamma(-d) \sin^{-d}(\nu)} \right)^2 \left(\frac{2}{j} \right)^{2+2d} + O(j^{-7/2-d}) \right] \sigma(0) \\
&\quad + \frac{2^4 \sigma(0) \sigma_\varepsilon^2 \sin(d\pi) \Gamma(1-2d)}{\pi (\Gamma(-d))^2} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} B_{j,l}
\end{aligned} \tag{40}$$

where $B_{j,l}$ is:

$$\begin{aligned}
B_{j,l} &= \cos \left[(j-d)\nu + \frac{d\pi}{2} \right] \cos \left[(l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \\
&\quad \times \left(1 + O(j^{d-1/2}) \right) \left(1 + O(l^{d-1/2}) \right) \left(1 + O((j-l)^{-1}) \right).
\end{aligned}$$

The first term (40) is bounded by $O(k^{-2d-1})$ and thus negligible with respect to k^{-1} . For the second term (40):

$$\begin{aligned}
 & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} B_{j,l} \\
 = & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[(j-d)\nu + \frac{d\pi}{2} \right] \cos \left[(l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \\
 + & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[(j-d)\nu + \frac{d\pi}{2} \right] \cos \left[(l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O(j^{d-1/2}) \\
 + & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[(j-d)\nu + \frac{d\pi}{2} \right] \cos \left[(l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O(l^{d-1/2}) \\
 + & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[(j-d)\nu + \frac{d\pi}{2} \right] \cos \left[(l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O((j-l)^{-1}). \quad (41)
 \end{aligned}$$

We first prove that the terms (41), (41) and (41) are bounded by $O(k^{d-3/2})$ and thus negligible with respect to k^{-1} . For example, we study (41):

$$\left| \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[(j-d)\nu + \frac{d\pi}{2} \right] \cos \left[(l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O(j^{d-1/2}) \right| \leq$$

$$C \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} j^{-3/2} l^{-1-d} (j-l)^{2d-1} = O(k^{d-3/2}).$$

The calculations for (41) and (41) have the same form. The dominant term of (40) is then (41). (41) can be broken into three parts:

$$\begin{aligned}
 & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[(j-d)\nu + \frac{d\pi}{2} \right] \cos \left[(l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} = \\
 & \frac{1}{2} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \\
 & + \frac{1}{2} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos(2(j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \\
 & + \frac{1}{4} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} [\cos[(2j-2d)\nu + d\pi] + \cos[(2l-2d)\nu + d\pi]] (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)}. \quad (42)
 \end{aligned}$$

We have proved in Proposition 2.4 that (42) is equivalent to Ck^{-1} . And it is easy to show (by Abel transformations), that (42) and (42) are negligible with respect to (42). This concludes the proof.

Acknowledgements. I would like to thank Anne Philippe, my Ph.D. advisor, who monitored my work. I am also grateful to the referees for their helpful comments.

REFERENCES

- [1] J. Barkoulas and C.F. Baum, Long-memory forecasting of US monetary indices. *J. Forecast.* **25** (2006) 291–302.
- [2] R.J. Bhansali, Linear prediction by autoregressive model fitting in the time domain. *Ann. Stat.* **6** (1978) 224–231.
- [3] R.J. Bhansali and P.S. Kokoszka, Prediction of long-memory time series: An overview. *Estadística* **53** No. 160–161 (2001) 41–96.
- [4] N.H. Bingham, C.M. Goldie and J.L. Teugels, *Regular variation* (1987).
- [5] P.J. Brockwell and R.A. Davis, Simple consistent estimation of the coefficients of a linear filter. *Stochastic Process. Appl.* (1988) 47–59.
- [6] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*. Springer Series in Statistics (1991).
- [7] N. Crato and B.K. Ray, Model selection and forecasting for long-range dependent processes. *J. Forecast.* **15** (1996) 107–125.
- [8] C.W.J. Granger and R. Joyeux, An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1** (1980) 15–29.
- [9] H.L. Gray, N.-F. Zhang and W.A. Woodward, On generalized fractional processes. *J. Time Ser. Anal.* **10** (1989) 233–257.
- [10] J.R.M. Hosking, Fractional differencing. *Biometrika* **68** (1981) 165–176.
- [11] A. Inoue, Regularly varying correlation functions and KMO-Langevin equations. *Hokkaido Math. J.* **26** (1997) 457–482.
- [12] A. Inoue, Asymptotics for the partial autocorrelation function of a stationary process. *J. Anal. Math.* **81** (2000) 65–109.
- [13] R. Lewis and G.C. Reinsel, Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.* **16** (1985) 393–411.
- [14] B. Mandelbrot and J.R. Wallis, Some long-run properties of geophysical records. *Water Resour. Res.* **5** (1969) 321–340.
- [15] M. Pourahmadi, On the convergence of finite linear predictors of stationary processes. *J. Multivariate Anal.* **30** (1989) 167–180.
- [16] B.K. Ray, Modeling long-memory processes for optimal long-range prediction. *J. Time Ser. Anal.* **14** (1993) 511–525.
- [17] L.J. Soares and L.R. Souza, Forecasting electricity demand using generalized long memory. *Int. J. Forecast.* **22** (2006) 17–28.
- [18] M.-C. Viano, Cl. Deniau and G. Oppenheim, Long-range dependence and mixing for discrete time fractional processes. *J. Time Ser. Anal.* **16** (1995) 323–338.
- [19] P. Whittle, Prediction and regulation by linear least-square methods. 2nd edn. (1963).
- [20] A. Zygmund, Trigonometric series. Cambridge University Press (1968).