



**HAL**  
open science

## Prévision linéaire des processus à longue mémoire

Fanny Godet

► **To cite this version:**

Fanny Godet. Prévision linéaire des processus à longue mémoire. Mathématiques [math]. Université de Nantes, 2008. Français. NNT : . tel-00349384

**HAL Id: tel-00349384**

**<https://theses.hal.science/tel-00349384>**

Submitted on 30 Dec 2008

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

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

UNIVERSITÉ DE NANTES  
FACULTÉ DES SCIENCES ET TECHNIQUES

ÉCOLE DOCTORALE SCIENCES ET TECHNOLOGIES  
DE L'INFORMATION ET DES MATHÉMATIQUES

Année : 2008

N° B.U. : 

--	--	--	--	--	--	--	--	--	--

## PRÉVISION LINÉAIRE DES PROCESSUS À LONGUE MÉMOIRE

Thèse de Doctorat de l'Université de Nantes

Spécialité : MATHÉMATIQUES ET APPLICATIONS

*Présentée et soutenue publiquement par*

**Fanny GODET**

*le 5 décembre 2008 à l'Université de Nantes.*

*Président du jury* : Marie-Claude VIANO Professeur émérite (Université de Lille I)

*Rapporteurs* : Jean-Marc BARDET Professeur (Université de Paris I)  
Philippe SOULIER Professeur (Université de Paris X)

*Examineurs* : Frédéric LAVANCIER Maître de Conférence (Université de Nantes)  
Remigijus LEIPUS Professeur (Université de Vilnius)  
Anne PHILIPPE Professeur (Université de Nantes)

*Directrice de thèse* : Anne PHILIPPE  
*Laboratoire* : Laboratoire Jean Leray (UMR 6629 UN-CNRS-ECN)

N° E.D. : 0366-xxx



# Remerciements

Je tiens tout d'abord à remercier ma directrice de thèse, Anne Philippe, pour m'avoir proposé un sujet de thèse passionnant et m'avoir encadré pendant ces trois années. Elle a su guider par des problématiques pertinentes mes recherches. Sa disponibilité, son écoute, sa confiance ont rendu ce travail très agréable. J'ai aussi découvert chez elle une très bonne critique cinéma et littéraire, les sorties ou les lectures qu'elle m'a encouragé à faire resteront comme autant de bons souvenirs de mes années à Nantes.

Je remercie également les autres membres de mon jury : Jean-Marc Bardet, Frédéric Lavancier, Remis Leipus, Philippe Soulier, Marie-Claude Viano. Je remercie particulièrement mes deux rapporteurs Jean-Marc Bardet et Philippe Soulier pour la relecture attentive de mon manuscrit. Je remercie également Marie-Claude Viano qui m'a fait l'honneur de présider mon jury de thèse après avoir fait découvrir la statistique et avoir encadré mes premiers pas en recherche au cours de mon stage de DEA à Lille. Je tiens à témoigner ma reconnaissance à Remis Leipus pour s'être intéressé lors de chacun de ses séjours à Nantes aux progrès de ma thèse et pour avoir fait le déplacement pour assister à ma soutenance. Je remercie enfin Frédéric Lavancier pour les discussions que nous avons eu concernant la recherche ou l'enseignement.

Un grand merci également à tous les membres du laboratoire Jean Leray : à Brigitte, Annick, Martine et Colette pour m'avoir aidée à régler les questions administratives, à Saïd et à Guy pour avoir trouvé rapidement des solutions à mes si nombreux problèmes en informatique et à Claude pour avoir fait le tour de France téléphonique pour trouver la documentation dont j'avais besoin. Je n'oublie pas non plus le groupe des thésards et les nombreuses parties de taroinche que j'y ai disputées : Rodolphe champion international de taroinche, Alain mon co-bureau dont l'accent me manque sous les nuages du Nord, Aymeric qui lui non plus ne boit pas de café, Etienne (et l'histoire du chou), Antoine avec qui j'ai partagé mes premier pas d'enseignante, Vincent toujours prêt à refaire le monde, Simon docteur, Simon (qui le sera bientôt), Manu qui m'a fait succomber au démon des jeux de cartes, Julien nouveau membre du bureau 27, Ronan qui défend maintenant seul l'honneur de la statistique, Skander, Nicolas et enfin Alexandre dont je n'oublie pas la bonne humeur communicative.

La poursuite de telles études est à la la fois une question de rencontres et de moyens. Je tiens à remercier mes professeurs en classe préparatoire Isabelle Goubet et Jean Voedts qui ont suscité en moi la passion des mathématiques et m'ont soutenue pendant les moments difficiles. Cette thèse a ensuite été rendue possible par mon entrée à l'Ecole Nor-

male Supérieure de Cachan : en effet le statut d'élève normalien m'a permis d'envisager sereinement le prolongement des mes études au-delà des concours d'enseignement et de côtoyer de grands professeurs sachant rendre communicatif leur goût pour la recherche. Pour terminer je tiens à remercier l'équipe du DEA de Mathématiques Appliquées de Lille I de m'avoir accueillie et en particulier son responsable Bernhard Beckermann. Nos correspondances sur les matrices de Toeplitz m'ont ouvert une nouvelle approche sur la longue mémoire.

Un grand merci à ma famille notamment mon papa, mes amis et mes chats pour leur patience. Pour finir mes pensées vont à Sébastien, mon rayon de soleil réunionnais, qui m'a soutenue tout au long de ce travail en m'encourageant lorsque les résultats tardaient à venir ou encore en corrigeant avec patience mon anglais souvent défaillant.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Les Processus à longue mémoire . . . . .	7
1.1.1	Définitions et exemples . . . . .	7
1.1.2	Résultats asymptotiques . . . . .	9
1.2	Méthodes de prévision . . . . .	12
1.2.1	Le Prédicteur de Wiener-Kolmogorov . . . . .	12
1.2.2	Lorsque le passé infini n'est pas entièrement connu . . . . .	14
1.3	Estimation des coefficients des prédicteurs . . . . .	16
1.3.1	Cas du prédicteur tronqué . . . . .	16
1.3.2	Cas du prédicteur projeté . . . . .	18
1.3.3	Estimer et prévoir sur une même réalisation . . . . .	20
<b>2</b>	<b>Convergence des prédicteurs linéaires d'ordre fini dans <math>L^2</math></b>	<b>23</b>
2.1	Introduction . . . . .	24
2.2	Wiener-Kolmogorov Next Step Prediction Theory . . . . .	25
2.2.1	The Wiener-Kolmogorov predictor . . . . .	25
2.2.2	Mean-squared prediction error when the predictor is truncated . . . . .	26
2.3	Projection Method . . . . .	32
2.3.1	Definition of the predictor . . . . .	32
2.3.2	Rate of convergence of the error by projection . . . . .	33
2.4	The Multistep Predictors . . . . .	37
2.4.1	The Wiener-Kolmogorov predictor . . . . .	37
2.4.2	The truncated Wiener-Kolmogorov predictor and the linear least-squares predictor . . . . .	39
2.5	Appendix : Proof of Proposition 2.2.2 . . . . .	42
<b>3</b>	<b>Estimation des coefficients du prédicteur tronqué</b>	<b>46</b>
3.1	Introduction . . . . .	47
3.2	The Truncated Wiener-Kolmogorov Predictor . . . . .	48
3.2.1	Prediction Method . . . . .	48
3.2.2	Estimates of Forecast Coefficients and Associated Mean-Squared Error for the Next Step Predictor . . . . .	49
3.2.3	$h$ -step Prediction . . . . .	58

3.3	Simulations and Conclusion . . . . .	60
3.4	Appendix . . . . .	64
3.4.1	Proof of Lemma 3.2.1 . . . . .	64
3.4.2	Proof of Lemma 3.2.2 . . . . .	68
<b>4</b>	<b>Estimation des coefficients du prédicteur projeté</b>	<b>74</b>
4.1	Introduction . . . . .	74
4.2	Estimation de l'erreur quadratique due à l'estimation des coefficients . . .	75
4.3	Conclusion . . . . .	81
4.4	Annexe . . . . .	81
4.4.1	Preuve du théorème 4.2.1 . . . . .	81
4.4.2	Preuve du lemme 4.4.1 . . . . .	85
<b>5</b>	<b>Prédiction et estimation du modèle sur une même réalisation</b>	<b>89</b>
5.1	Introduction . . . . .	90
5.2	Moment bounds . . . . .	93
5.3	The mean-squared prediction error of the least-squares predictor . . . . .	99
5.4	Central limit theorem . . . . .	104
5.5	Appendix . . . . .	107
5.5.1	Preliminary lemmas . . . . .	107
5.5.2	Proof of Lemma 5.3.1 . . . . .	110
<b>6</b>	<b>Simulations</b>	<b>117</b>
6.1	Trajectoires simulées de processus fractionnaires . . . . .	117
6.1.1	Le processus fractionnaire . . . . .	117
6.1.2	Méthodes de simulation . . . . .	118
6.1.3	Prévision par le prédicteur de Wiener-Kolmogorov tronqué et le prédicteur projeté sur un passé fini . . . . .	118
6.2	Étude du niveau du lac majeur . . . . .	122
6.2.1	Les données . . . . .	122
6.2.2	Prévision . . . . .	123
6.2.3	Prévision par le prédicteur tronqué . . . . .	123
6.2.4	Prévision par le prédicteur projeté . . . . .	124
	<b>Bibliographie</b>	<b>129</b>

# Chapitre 1

## Introduction

### 1.1 Les Processus à longue mémoire

#### 1.1.1 Définitions et exemples

Nous nous intéresserons aux processus stationnaires d'ordre 2 à longue mémoire. Ces processus modélisent une grande variété de phénomènes que ce soit l'exemple classique du débit du Nil (voir Hosking (1984) et Beran (1994) chapitre 1) ou des exemples plus récents comme par exemple le trafic ethernet (voir Willinger *et al.* (1997)) ou des séries économétriques comme l'inflation de pays comme le Royaume-Uni et les États-Unis (voir Doornik et Ooms (2004)), certaines caractéristiques du New York Stock Exchange comme la volatilité ou le nombre d'opérations sur un intervalle fini (voir Deo *et al.* (2007)). Baillie (1996) cite de nombreux autres exemples en économie.

Dans le cadre  $L^2$ , une façon d'évaluer la longue mémoire d'un processus est de regarder la convergence vers 0 de ses autocovariances. On dit qu'un processus est à longue mémoire si sa fonction d'autocovariance converge lentement vers 0 au sens où elle n'est pas sommable. Cette caractérisation par les caractéristiques d'ordre 2 a toutefois des limites. D'une part si pour les processus gaussiens la fonction d'autocovariance caractérise entièrement la loi du processus, pour un processus non-gaussien la fonction d'autocovariance peut n'être qu'une information très faible sur le processus. D'autre part les processus ARCH( $\infty$ ) étudiés par Giraitis *et al.* (2000) ont des covariances absolument sommables mais sont pourtant utilisés pour modéliser des séries temporelles présentant une forte dépendance i.e. leurs covariances empiriques décroissent lentement. De plus cette définition ne suffit pas en général à valider les méthodes statistiques pour les processus à longue mémoire. Il faut donc définir plus précisément la longue mémoire soit dans le domaine temporel par l'expression de la fonction d'autocovariance, soit dans le domaine spectral avec l'expression de la densité spectrale. Les définitions suivantes sont les plus communes (voir Taqqu (2003)) pour ce que l'on appelle la longue mémoire régulière.

**Définition 1.** *On dit qu'un processus  $(X_n)_{n \in \mathbb{Z}}$  stationnaire d'ordre 2 est à longue mémoire régulière*



– si ses autocovariances  $\sigma$  vérifient

$$\sigma(j) \sim j^{-\beta} L_1(j) \text{ lorsque } j \rightarrow +\infty \text{ avec } 0 < \beta < 1 \quad (1.1)$$

– ou si sa densité spectrale  $f$  est continue sur le segment  $[-\pi, \pi]$  sauf en 0 où elle vérifie

$$f(\lambda) \sim |\lambda|^{-\gamma} L_2(\lambda^{-1}) \text{ lorsque } \lambda \rightarrow 0 \text{ avec } 0 < \gamma < 1 \quad (1.2)$$

où  $L_1$  et  $L_2$  sont des fonctions à variation lente au voisinage de l'infini.

Comme définition de fonction à variation lente, on choisit celle de Zygmund (1968) :  $L$  est une fonction à variation lente au voisinage de l'infini si et seulement si quelque soit  $\delta > 0$ ,  $x \mapsto x^{-\delta} L(x)$  est une fonction décroissante et  $x \mapsto x^\delta L(x)$  est une fonction croissante à partir d'un certain  $x_0$  suffisamment grand. Cette définition implique que pour tout  $s > 0$ ,  $L(ts)/L(t)$  tend vers 1 lorsque  $t$  tend vers l'infini. Cette propriété est parfois donnée comme définition de fonction à variation lente (voir Bingham *et al.* (1987)). Des exemples classiques de fonction à variation lente sont les fonctions constantes et la fonction logarithme.

Dans les équations (1.1) et (1.2), les paramètres  $\beta$  et  $\gamma$  mesurent l'intensité de la longue mémoire : plus  $\beta$  est proche de 0 ou plus  $\gamma$  est proche de 1 plus la dépendance du processus est forte. Par un théorème de type abélien, on peut montrer que (1.1) implique (1.2), mais pour la réciproque on a besoin d'un théorème taubérien et donc de l'hypothèse supplémentaire que la suite des autocovariances  $(\sigma(j))_{j \in \mathbb{N}}$  est décroissante à partir d'un certain rang. Sous cette hypothèse, on obtient que  $\gamma = 1 - \beta$  et on pose  $d = \gamma/2$  que l'on appelle paramètre de longue mémoire. L'exemple le plus célèbre de processus vérifiant les propriétés (1.1) et (1.2) est celui des processus FARIMA.

**Exemple 1.**  $(X_n)_{n \in \mathbb{Z}}$  est un processus FARIMA( $p, d, q$ ) s'il est solution stationnaire d'une équation de la forme :

$$(I - B)^d \phi(B) X_n = \psi(B) \varepsilon_n$$

où  $B$  est l'opérateur retard,  $(\varepsilon_n)_{n \in \mathbb{Z}}$  est un bruit blanc,  $d$  appartient à  $]0, 1/2[$  et  $\phi$  et  $\psi$  sont des polynômes de degrés respectifs  $p$  et  $q$  sans racine commune et qui ne s'annulent pas sur le disque unité fermé.

Un cas particulier des processus FARIMA est les processus fractionnaires pour lesquels les polynômes  $\phi$  et  $\psi$  sont choisis constants égaux à 1.

On peut aussi associer à la longue mémoire des effets saisonniers (voir Viano *et al.* (1995)), la singularité de la densité spectrale n'est alors plus nécessairement localisée en 0. Un exemple classique de tels processus est les processus de type GARMA introduits par Hosking (1981). Leur densité spectrale est de la forme :

$$f(\lambda) = C(2|\cos \lambda - \cos \nu|)^{-2d} \text{ avec } \nu \in [0, \pi[ \setminus \left\{ \frac{\pi}{2} \right\}$$

avec  $C$  constante positive et  $d \in ]0, 1/2[$ . Elle présente donc une singularité en  $-\nu$  et  $\nu$ . Et leur fonction d'autocovariance vérifie :

$$\sigma(j) = K \cos(j\nu) j^{2d-1} (1 + O(j^{-1})) \text{ lorsque } j \rightarrow +\infty$$

avec  $K$  constante indépendante de  $j$ .

Par opposition on parlera de processus à courte mémoire s'ils vérifient la définition suivante :

**Définition 2.** *On dit qu'un processus  $(X_n)_{n \in \mathbb{Z}}$  stationnaire d'ordre 2 est à courte mémoire*

– si sa fonction d'autocovariance  $\sigma$  vérifie

$$\sum_{j \in \mathbb{Z}} |\sigma(j)| < +\infty$$

– ou si sa densité spectrale  $f$  est continue presque partout et bornée inférieurement et supérieurement i.e. il existe  $m, M \in \mathbb{R}^{*+}$  tels que pour tout  $\lambda \in [-\pi, \pi]$ ,  $m < f(\lambda) < M$ .

Les processus ARMA sont un exemple courant de processus à courte mémoire :

**Exemple 2.**  $(X_n)_{n \in \mathbb{Z}}$  est un ARMA( $p, q$ ) s'il est solution stationnaire de

$$\phi(B)X_n = \psi(B)\varepsilon_n$$

avec  $(\varepsilon_n)_{n \in \mathbb{Z}}$  bruit blanc et  $\phi$  et  $\psi$  polynômes de degrés respectifs  $p$  et  $q$  sans racines communes et sans racine dans le disque unité.

Les processus à courte mémoire sont étudiés depuis plus longtemps que les processus à longue mémoire. La question de leur prévision a donc été très largement traitée (de nombreuses références seront données dans ce chapitre). Il sera intéressant de comparer les résultats que nous obtiendrons en longue mémoire à ceux obtenus pour des processus à courte mémoire. Mais avant de passer à la question de la prévision, nous allons rappeler quelques résultats dont on dispose pour les processus à longue mémoire.

### 1.1.2 Résultats asymptotiques

Pour prévoir une série temporelle, on a besoin d'estimer sa fonction de prévision. On rappelle dans cette partie quelques résultats asymptotiques utilisés dans ce travail pour estimer la fonction de prévision. Or la convergence de nombreux estimateurs dépend de celle de formes quadratiques du processus. Nous donnons donc ici quelques résultats sur le comportement quand  $n \rightarrow +\infty$  de quantités s'écrivant sous la forme :

$$\sum_{i=1}^n \sum_{j=1}^n a_{i-j} X_i X_j \quad (1.3)$$

où  $(X_n)_{n \in \mathbb{Z}}$  est un processus à longue mémoire stationnaire d'ordre 2. Si on note  $f$  la densité spectrale du processus et  $g$  la fonction qui a pour coefficients de Fourier la suite  $(a_j)_{j \in \mathbb{N}}$ , Fox et Taquq (1987) et Giraitis et Surgailis (1990) ont montré la normalité asymptotique de (1.3) sous des hypothèses de régularité de  $f$  et  $g$  respectivement pour des processus gaussiens et linéaires.

**Théorème 1.1.1** (Fox et Taqqu (1987)). *Soient  $f$  et  $g$  deux fonctions réelles symétriques dont l'ensemble des points de discontinuité forment un ensemble de mesure nulle. On suppose de plus qu'il existe  $\alpha < 1$  et  $\beta < 1$  tels que  $\alpha + \beta < \frac{1}{2}$  et tels que pour tout  $\delta > 0$  on a*

$$f(x) = O(|x|^{-\alpha-\delta}) \text{ lorsque } x \rightarrow 0$$

et

$$g(x) = O(|x|^{-\beta-\delta}) \text{ lorsque } x \rightarrow 0.$$

*Soit  $(X_n)_{n \in \mathbb{Z}}$  un processus gaussien stationnaire de moyenne nulle et de densité spectrale  $f$ . On suppose que :*

$$\int_{-\pi}^{\pi} \log(f(x)) dx = 0.$$

*On note les coefficients de Fourier de  $g$  :  $a_n = \int_{-\pi}^{\pi} \exp(inx)g(x)dx$ . On a alors*

$$\sqrt{n}^{-1} \left( \sum_{1 \leq i, j \leq n} a_{i-j} X_i X_j - \mathbb{E} \left[ \sum_{1 \leq i, j \leq n} a_{i-j} X_i X_j \right] \right)$$

*qui tend en loi lorsque  $n \rightarrow +\infty$  vers une gaussienne centrée de moyenne nulle et de variance  $16\pi^3 \int_{-\pi}^{\pi} [f(x)g(x)]^2 dx$ .*

Ces conditions de régularité pour  $f$  incluent de nombreux processus à courte mémoire et à des processus à longue mémoire dont le paramètre de longue mémoire est inférieur à une valeur dépendant de la suite  $(a_n)_{n \in \mathbb{N}}$ . Mais cette normalité asymptotique n'est pas toujours vérifiée pour les processus à longue mémoire. Rosenblatt (1979) a par exemple prouvé que les covariances empiriques (cas particulier de (1.3) avec  $a_j = \delta_{i-j=h}$  où  $\delta$  est la fonction delta de Kronecker) d'un processus à longue mémoire dont le paramètre de longue mémoire  $d$  est strictement supérieur à  $1/4$  convergent en loi à une vitesse  $n^{1-2d}$  vers une variables aléatoire appelée variable aléatoire de Rosenblatt. Fox et Taqqu (1985) ont généralisé ce résultat pour des formes quadratiques dont les coefficients  $(a_j)_{j \in \mathbb{N}}$  sont sommables pour la même région de  $d$ . Ils ont également prouvé que la limite en loi dépend du compromis entre la vitesse de décroissance des covariances du processus et les coefficients  $(a_j)_{j \in \mathbb{N}}$  considérés. Les résultats que nous venons de citer considèrent des formes quadratiques simples et aux coefficients constants, mais ces hypothèses peuvent également être relaxées. On peut par exemple faire dépendre les coefficients  $a_j$  de  $n$  (voir Bhansali *et al.* (2007)), ou considérer des formes quadratiques de polynômes de Hermite Fox et Taqqu (1985). La particularité des processus à longue mémoire est que la normalité asymptotique n'est pas toujours assurée comme pour les processus à courte mémoire car non seulement la limite peut-être différente mais surtout la vitesse de convergence de la forme quadratique peut-être altérée.

Nous rappelons maintenant quelques résultats sur les estimateurs des paramètres des processus à longue mémoire. Nous allons commencer par la classe des estimateurs paramétriques. On peut par exemple utiliser le maximum de vraisemblance (voir Beran

(1994)) et en particulier l'estimateur de Whittle qui minimise la fonction de contraste suivante :

$$\mathcal{L}(\theta) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I_n(\lambda)}{f_\theta(\lambda)} d\lambda \quad (1.4)$$

où  $I_n$  est le périodogramme calculé sur  $nn$  observations et  $f_\theta$  la densité spectrale associée au paramètre  $\theta$ . La convergence normale de cet estimateur a d'abord été montré dans le cas gaussien par Fox et Taquq (1986) puis dans le cas général par Giraitis et Surgailis (1990).

**Théorème 1.1.2** (Fox et Taquq (1986)). *Soit  $(X_n)_{n \in \mathbb{Z}}$  un processus gaussien stationnaire de moyenne nulle et de densité spectrale  $f(\cdot, \theta_0)$ . On suppose que  $f(\cdot, \theta_0)$  appartient à une famille de densités spectrales qui vérifient certaines conditions de régularité (que nous rappellerons au chapitre 3).*

*Alors si  $\hat{\theta}_n$  minimise le contraste (1.4),  $\hat{\theta}_n - \theta_0$  tend en loi vers une gaussienne centrée lorsque  $n \rightarrow +\infty$ .*

Lorsqu'on ne veut estimer que le paramètre de longue mémoire  $d$ , on peut utiliser des estimateurs semi-paramétriques locaux comme l'estimateur de Whittle local proposé par Künsch (1987) pour les processus gaussien, l'estimateur du log-périodogramme local de Geweke et Porter-Hudak (1983). L'estimateur local de Whittle consiste à remplacer dans la fonction de contraste  $\mathcal{L}$  défini en (1.4) l'intégrale sur  $[\pi, \pi]$  par une intégrale sur un voisinage de 0 et de restreindre la famille des densités spectrales à  $\{|1 - \exp(i\cdot)|^{2d}, d \in \Theta_0\}$ . L'estimateur GPH consiste lui à régresser la fonction  $-2d \log(\cdot) + C$  sur le log-périodogramme autour de 0. Il existe également des estimateurs semi-paramétriques globaux donnant à la fois un estimateur pour le paramètre de longue mémoire  $d$  et pour la densité spectrale  $f$ . Dans cette approche, on distingue la partie à longue mémoire du reste du signal. On suppose que la densité spectrale  $f$  du processus est de la forme :  $f(x) = |1 - \exp(ix)|^{-2d} f^*(x)$  avec  $f^*$  ne présentant pas de singularité et donc pouvant être considérée comme la densité spectrale d'un processus à courte mémoire. Par exemple, l'estimateur FAR consiste à ajuster un processus FARIMA( $p, d, 0$ ) en utilisant comme critère la fonction de contraste écrite en (1.4) de l'estimateur de Whittle. Dans ce cas, on ajuste donc au processus défini par la densité spectrale  $f^*$  un processus autorégressif d'ordre fini  $p$ . Il a été étudié par Bhansali *et al.* (2006). Sur le même principe l'estimateur FEXP proposé par Moulines et Soulier (1999) permet d'estimer  $d$  et les coefficients du développement tronqué de  $\log(f^*)$  dans la base des fonctions cosinus. Toutes ces méthodes semi-paramétriques suppose que le pôle de la densité spectrale se trouve en 0 et exclut donc les processus à longue mémoire saisonniers. Lorsque le pôle ne se situe pas en 0, il faut tout d'abord estimer sa fréquence. Yajima (1996) a par exemple proposé un estimateur très simple : la fréquence où le périodogramme est maximum. On peut ensuite les méthodes précédentes en supposant la fréquence du pôle égale à celle estimée. Hidalgo et Soulier (2004) ont démontré la pertinence de cette méthode pour l'estimateur GPH où on l'estime la fréquence du pôle par l'estimateur de Yajima (1996).

Nous allons maintenant introduire les méthodes de prévision que nous étudierons dans cette thèse.

## 1.2 Méthodes de prévision

On suppose dans cette partie que l'on connaît la loi de probabilité du processus étudié et que c'est un processus linéaire causal et inversible.  $(X_n)_{n \in \mathbb{Z}}$  admet donc une représentation moyenne mobile infinie :

$$X_n = \sum_{j=0}^{\infty} b_j \varepsilon_{n-j} \quad (1.5)$$

où  $(\varepsilon_n)_{n \in \mathbb{Z}}$  est un bruit blanc  $L^2$ , centré, de variance  $\sigma_\varepsilon^2$  et où la suite  $(b_j)_{j \in \mathbb{N}}$  vérifie  $\sum b_j^2 < +\infty$ . On suppose en outre que  $(X_n)_{n \in \mathbb{Z}}$  admet une représentation autorégressive infinie :

$$\varepsilon_n = \sum_{j=0}^{\infty} a_j X_{n-j}, \quad (1.6)$$

telle que les  $(a_j)_{j \in \mathbb{N}}$  vérifient  $\sum |a_j| < +\infty$ . Sous ces hypothèses le processus  $(X_n)_{n \in \mathbb{Z}}$  est linéaire mais on n'a pas d'informations sur sa fonction d'autocovariance ou sur la forme de sa densité spectrale. Un processus vérifiant ces hypothèses peut donc être à courte ou à longue mémoire.

Nous nous limiterons à l'étude de prédicteurs linéaires. Le critère que nous utiliserons pour mesurer la qualité d'un prédicteur  $\hat{X}$  pour la variable aléatoire  $X$  est l'erreur quadratique  $\mathbb{E}[(X - \hat{X})^2]$ . Pratiquement, il permet de pénaliser plus fortement les grandes erreurs que les petites. Mathématiquement, cela nous permettra de nous placer dans l'espace de Hilbert  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  et de pouvoir utiliser tous les outils afférents notamment pour expliciter  $\hat{X}$ . On obtiendra ainsi un prédicteur ponctuel et une estimation de l'erreur quadratique commise.

Mais la prédiction ponctuelle n'est pas toujours une information suffisante pour les utilisateurs des prévisions. Ils peuvent également avoir besoin d'un intervalle de confiance. Dans ce but on démontrera des théorèmes de normalité asymptotique pour les prédicteurs, qui impliquent des intervalles de confiance asymptotiques.

### 1.2.1 Le Prédicteur de Wiener-Kolmogorov

Nous allons commencer par définir un prédicteur de référence pour lequel l'erreur quadratique de prédiction est minimum. Plus précisément nous allons calculer le meilleur prédicteur linéaire  $\widehat{X}_k(h)$  à l'horizon  $h$  au sens  $L^2$  connaissant le passé infini  $\{X_{k+1-j}, j \leq 1\}$  que nous appellerons prédicteur de Wiener-Kolmogorov, c'est à dire que l'on va projeter  $X_{k+h}$  sur le complété du sous-espace engendré par la famille de variables aléatoires  $\{X_{k-i}, i \leq 0\}$  dans l'espace  $L^2$  des variables aléatoires réelles de carré intégrables comme Whittle (1983).

On note  $P_{\mathcal{F}}$  le projecteur sur le complété du sous-espace engendré par la famille  $\mathcal{F}$  et  $[X]_{\ell}^j$

le sous-espace vectoriel engendré par  $\{X_\ell, \dots, X_j\}$ . Le prédicteur de Wiener-Kolmogorov s'écrit alors :

$$\widetilde{X}_k(h) = P_{[X]_{-\infty}^k} (X_{k+h}) = \sum_{j=0}^{\infty} \lambda(j) X_{k-j}$$

où les  $(\lambda(j))_{j \in \mathbb{N}}$  sont obtenus en minimisant :

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

Comme  $(X_n)_{n \in \mathbb{Z}}$  est causal et inversible, les sous-espaces vectoriels  $[\varepsilon]_{-\infty}^k$  et  $[X]_{-\infty}^k$  sont égaux. Le prédicteur  $\widetilde{X}_k(h)$  s'écrit aussi sous la forme :

$$\widetilde{X}_k(h) = \sum_{j=0}^{\infty} \phi_h(j) \varepsilon_{k-j}.$$

où les  $(\phi_h(j))_{j \in \mathbb{N}}$  sont fonction des  $(\lambda(j))_{j \in \mathbb{N}}$  et des  $(a_j)_{j \in \mathbb{N}}$  définis en (1.6). A partir du développement en moyenne mobile infinie de  $(X_n)_{n \in \mathbb{Z}}$  donné en (1.5), l'erreur de prévision au sens  $L^2$  s'écrit :

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

car les variables aléatoires  $(\varepsilon_n)_{n \in \mathbb{Z}}$  sont non corrélées et de variance  $\sigma_\varepsilon^2$ . L'erreur quadratique de prévision minimale est égale à  $\sum_{j=0}^{h-1} b_j^2 \sigma_\varepsilon^2$  et elle est atteinte en prenant  $\phi(j) = b_{j+h}$  pour tout  $j$  positif.

Comme les sous-espaces vectoriels  $[\varepsilon]_{-\infty}^k$  et  $[X]_{-\infty}^k$  sont égaux, on a pour  $h = 1$

$$\widetilde{X}_k(1) = - \sum_{j=1}^{\infty} a_j X_{k+1-j} \quad (1.7)$$

et par induction pour  $h > 1$

$$\widetilde{X}_k(h) = - \sum_{j=1}^{h-1} a_j \widetilde{X}_k(h-j) - \sum_{j=h}^{\infty} a_j X_{k+h-j}. \quad (1.8)$$

L'erreur minimale de prévision à l'horizon  $h : \sum_{j=0}^{h-1} b_j^2 \sigma_\varepsilon^2$  tend vers  $\sigma(0)$  lorsque l'horizon de prévision  $h$  tend vers l'infini. Or  $\sigma(0)$  est l'erreur de prévision du prédicteur naïf qu'est la moyenne du processus. Si l'erreur commise est trop proche de  $\sigma(0)$ , le prédicteur de Wiener-Kolmogorov n'apporte plus d'information par rapport à une prédiction naïve. On montrera dans le chapitre 2 que lorsque l'horizon de prévision  $h$  tend vers l'infini le prédicteur de Wiener-Kolmogorov est plus performant pour les processus à longue mémoire que pour les processus à courte mémoire puisque la convergence de l'erreur de prévision vers  $\sigma(0)$  est plus lente dans ce cas.

### 1.2.2 Lorsque le passé infini n'est pas entièrement connu

En pratique on ne connaît jamais le passé infini d'une série, on ne peut donc pas utiliser le prédicteur de Wiener-Kolmogorov mais il s'agit du meilleur prédicteur linéaire au sens  $L^2$  et tous les autres prédicteurs linéaires lui sont comparés.

Lorsque le nombre de valeurs manquantes dans le passé est fini (cas de valeurs aberrantes retirées de la série par exemple), Bondon (2005) propose de projeter sur le passé infini privé des valeurs manquantes. Plus précisément si on note  $X_{k-t_1}, \dots, X_{k-t_N}$  les observations manquantes ( $t_1 < \dots < t_N$ ), le prédicteur de  $X_{k+1}$  sera son projeté sur le complété du sous-espace vectoriel engendré par la famille de variables aléatoires  $\{X_{k-i}, i \leq 0, i \neq t_1, \dots, t_N\}$ . Bondon a majoré l'erreur quadratique entre ce prédicteur et le prédicteur de Wiener-Kolmogorov en fonction de la vitesse de décroissance des coefficients de la représentation autorégressive  $(a_j)_{j \in \mathbb{N}}$  vers 0 et de l'indice de la première valeur manquante  $t_1$  lorsque  $t_1$  tend vers l'infini. Il a obtenu deux majorations différentes selon la vitesse de décroissance des coefficients  $(a_j)_{j \in \mathbb{N}}$  vers 0.

**Théorème 1.2.1** (Bondon (2005)). *Soit  $(X_n)_{n \in \mathbb{Z}}$  un processus stationnaire d'ordre 2 causal et inversible dont la représentation autorégressive infinie est donnée par (1.6). On a alors :*

- $\mathbb{E} \left[ \left( \mathbb{P}_{\{X_{k-i}, i \leq 0, i \neq t_1, \dots, t_N\}}(X_{k+1}) - X_{k+1} \right)^2 \right] = O(\alpha^{t_1})$  si  $|a_j| = O(\alpha^j)$  lorsque  $j \rightarrow +\infty$  avec  $\alpha \in ]0, 1[$ ;
- $\mathbb{E} \left[ \left( \mathbb{P}_{\{X_{k-i}, i \leq 0, i \neq t_1, \dots, t_N\}}(X_{k+1}) - X_{k+1} \right)^2 \right] \asymp t_1^\alpha$  si  $|a_j| \sim Cj^\alpha$  lorsque  $j \rightarrow +\infty$  avec  $C > 0$  et  $\alpha < -1/2$ .<sup>1</sup>

Les processus ARMA ont un développement autorégressif vérifiant  $|a_j| = O(\alpha^j)$ , et les processus FARIMA un développement vérifiant  $|a_j| \sim Cj^\alpha$ . La perte d'un nombre fini d'observations du passé dégrade donc plus l'erreur quadratique de prévision pour les processus à longue mémoire que pour les processus à courte mémoire.

En pratique, il ne manque pas seulement un nombre fini d'observations mais une infinité car le passé observé est fini. Notons  $k$  la longueur du passé observé. Dans ce cas, deux prédicteurs peuvent être utilisés.

1. On tronque la série infinie (1.7) qui définit le prédicteur de Wiener-Kolmogorov à l'ordre  $k$  et par la suite on appellera ce prédicteur le prédicteur tronqué à l'ordre  $k$ . A l'horizon 1, on obtient le prédicteur suivant :

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

En général parmi les éléments de  $[X]_1^k$ ,  $\widetilde{X}'_k(1)$  n'est pas optimal au sens  $L^2$  sauf dans le cas où  $[X]_{-\infty}^k = [X]_1^k$ .

---

<sup>1</sup> $f(t) \asymp g(t)$  si et seulement si il existe un réel  $t_0$  et deux constantes strictement positives  $C$  et  $C'$  telles que si  $t > t_0$  alors  $C < \frac{f(t)}{g(t)} < C'$

2. Comme dans le cas d'un nombre fini de valeurs manquantes, on projette  $X_{k+1}$  sur le passé fini observé  $(X_1, \dots, X_k)$ . On appelle ce prédicteur : prédicteur projeté d'ordre  $k$  et on le notera

$$\widehat{X}_k(1) = P_{[X]_1^k}(X_{k+1}) = - \sum_{j=1}^k a_{j,k} X_{k+1-j}.$$

Le prédicteur projeté est par définition optimal lorsqu'on ne dispose que des  $k$  dernières observations pour prédire, néanmoins le prédicteur tronqué présente l'avantage d'être plus simple à calculer car les coefficients du prédicteur ne doivent pas être recalculés entièrement si la valeur de  $k$  change.

Baxter (1962) a obtenu la vitesse de convergence du prédicteur optimal vers le prédicteur de Wiener-Kolmogorov pour les processus à courte mémoire unidimensionnels. Plus généralement, Pourahmadi (1989) a comparé les deux prédicteurs tronqué et optimal au prédicteur de Wiener-Kolmogorov pour un processus stationnaire multidimensionnel d'ordre 2 à courte mémoire. Dans ce cas, les deux prédicteurs convergent exponentiellement vite vers le prédicteur de Wiener-Kolmogorov lorsque le nombre d'observations disponibles  $k$  croît. Nous énonçons ici le théorème en unidimensionnel :

**Théorème 1.2.2** (version unidimensionnelle du théorème de 4.1 de Pourahmadi (1989)). *Soit  $(X_n)_{n \in \mathbb{Z}}$  un processus stationnaire  $L^2$  dont la densité spectrale  $f$  est égale presque partout à une fonction analytique et est bornée inférieurement et supérieurement par des constantes strictement positives. Alors il existe  $0 \leq \rho < 1$  tel que*

- pour le prédicteur tronqué

$$\mathbb{E} \left[ \left( \widetilde{X}_k(1) - \widetilde{X}'_k(1) \right)^2 \right] = O(\rho^k) \text{ lorsque } k \rightarrow +\infty$$

- pour le prédicteur projeté d'ordre  $k$

$$\mathbb{E} \left[ \left( \widetilde{X}_k(1) - \widehat{X}_k(1) \right)^2 \right] = O(\rho^k) \text{ lorsque } k \rightarrow +\infty.$$

Ces résultats ne s'appliquent pas aux processus à longue mémoire qui ont une densité spectrale par définition non bornée. Pourtant ces deux prédicteurs sont souvent employés pour prévoir des séries temporelles modélisées par des processus à longue mémoire.

Par exemple Barkoulas et Baum (2006) utilisent le prédicteur tronqué pour prévoir des indices monétaires américains modélisés par des processus FARIMA, Ray (1993a) pour prévoir les recettes de IBM.

Le prédicteur optimal est lui présenté comme l'ajustement d'un processus autorégressif fini d'ordre  $k$  par Ray (1993b). L'efficacité de ce prédicteur y est illustrée sur des séries simulées de processus FARIMA. Elle paraît surprenante quand on présente le prédicteur comme l'ajustement d'un modèle autorégressif fini d'ordre  $k$  donc à courte mémoire sur un processus à longue mémoire. Mais elle est naturelle car à l'horizon



1, ajuster un modèle autorégressif ou projeter sur les observations disponibles donne le même prédicteur. D'autres modèles à courte mémoire sont également ajustés à des processus à longue mémoire pour les prévoir. Tiao et Tsay (1994) proposent par exemple d'ajuster un modèle ARMA(1,1) à un processus fractionnaire. Dans ce cas, l'efficacité du prédicteur dépend beaucoup de la manière dont on estime ces paramètres (elle est mauvaise si on utilise un estimateur du maximum de vraisemblance mais meilleure avec une méthode adaptative minimisant la variance des erreurs commises pour chaque horizon de prévision).

**Résultats de la thèse** Dans le chapitre 2, nous donnerons la vitesse de convergence des erreurs quadratiques des prédicteurs projeté et tronqué vers le prédicteur de Wiener-Kolmogorov lorsque  $k$  tend vers l'infini pour des processus à longue mémoire généralisant ainsi les résultats de Pourahmadi (1989). Comme dans le contexte étudié par Bondon (2005), la vitesse de convergence va se dégrader par rapport aux vitesses obtenues en courte mémoire passant en  $O(k^{-1})$ . Cette vitesse s'explique par le fait que les observations manquantes sont plus corrélées avec la variable à prévoir en longue mémoire et il manque donc plus d'informations dans le prédicteur. Toutefois cette vitesse, qui est par exemple celle du théorème de la limite centrale classique, n'invalide pas l'utilisation de tels prédicteurs bien qu'elle ne pourra plus être considérée comme négligeable comme en courte mémoire.

### 1.3 Estimation des coefficients des prédicteurs

On suppose dans cette partie que la loi du processus n'est plus connue. On doit donc estimer les fonctions de prédiction et dans le cas des prédicteurs linéaires estimer les coefficients du prédicteur. Bhansali et Kokoszka (2001) distinguent deux types de prédicteurs dans ce contexte. Les prédicteurs de type I sont ceux pour lesquels il n'est pas nécessaire d'estimer le paramètre de longue mémoire  $d$ , et les prédicteurs de type II sont ceux pour lesquels il faut estimer ce paramètre. Ainsi on sépare les prédicteurs utilisant la structure à longue mémoire du processus et ceux ne l'utilisant pas. L'approche de type II nécessite l'estimation paramétrique du paramètre de longue mémoire  $d$ . En général, les coefficients du prédicteur tronqué sont estimés par des méthodes (semi-)paramétriques de type I et ceux du prédicteur projeté par des méthodes non-paramétriques de type II. Pour l'estimation des coefficients du prédicteur nous supposons dans un premier temps que l'on dispose de deux réalisations indépendantes : celle à partir de laquelle on doit prédire de longueur  $k$  et celle grâce à laquelle on va estimer les coefficients de longueur  $T$ .

#### 1.3.1 Cas du prédicteur tronqué

Pour le prédicteur tronqué soit on se place dans un cadre paramétrique c'est à dire que l'on écrit les coefficients des représentation autorégressive et moyenne mobile infinie

comme fonction connue d'un paramètre  $\theta$  et par méthode plug-in on remplace dans cette fonction le paramètre  $\theta$  par son estimé soit on utilise la factorisation canonique de la densité spectrale c'est à dire que l'on peut exprimer les coefficients du prédicteur de Wiener-Kolmogorov comme des fonctions de la densité spectrale (voir Whittle (1983) p26).

Dans le cadre paramétrique, le traitement des données commence par une identification du modèle avec des critères comme AIC ou BIC puis par une estimation des paramètres. Beran *et al.* (1998) utilise par exemple ces critères pour sélectionner un modèle dans la famille des processus FARIMA( $p, d, 0$ ). L'estimation des paramètres d'un processus en longue mémoire peut ensuite se faire grâce des estimateurs paramétriques comme l'estimateur de Whittle (voir Fox et Taquq (1986) pour les processus gaussiens et Giraitis et Surgailis (1990) sans cette hypothèse) ou le maximum de vraisemblance (voir Dahlhaus (1989)).

**Résultats de la thèse** Dans le chapitre 3, nous intéresserons à l'erreur quadratique de prévision commise lorsque les coefficients du prédicteur sont estimés à l'aide de l'estimateur de Whittle. Nous donnerons un encadrement de cette erreur en faisant tendre  $T$  puis  $k$  vers l'infini.

Toujours dans un cadre paramétrique une autre solution spécifique aux processus à longue mémoire est proposée par Crato et Ray (1996). Il s'agit de séparer l'estimation du paramètre de longue mémoire  $d$  de l'estimation des autres paramètres. On estime tout d'abord  $d$  par  $\hat{d}$  (voir la partie 1.1.2 pour des exemples de prédicteurs du paramètre de longue mémoire). Puis on filtre le processus à longue mémoire par l'opérateur  $(I - B)^{\hat{d}}$  où  $B$  est l'opérateur retard, et enfin on ajuste à la série obtenue un modèle classique à courte mémoire (par exemple un ARMA). Cette approche reste paramétrique, la seule différence est que le paramètre de longue mémoire est estimé dans une première étape avec une méthode non nécessairement paramétrique.

On peut enfin utiliser la factorisation canonique de la densité spectrale (voir Whittle (1983)). On écrit alors les coefficients de la représentation autorégressive  $a_j$  en fonction de la densité spectrale  $f$  :

$$a_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp \left\{ - \sum_{u=1}^{+\infty} d_u \exp(-iu\lambda) \right\} \exp(ij\lambda) d\lambda \quad (1.9)$$

où la suite  $(d_u)_{u \in \mathbb{N}^*}$  sont les coefficients de Fourier du logarithme de la densité spectrale. On remplace ensuite dans (1.9) la densité spectrale par le périodogramme pour estimer les coefficients de la représentation autoregressive. Cette méthode a été proposée par Bhan-sali (1974) pour des processus à courte mémoire et étudiée par Hidalgo et Yajima (2002) pour des processus à longue mémoire. Plus précisément ils ont démontré des théorèmes de convergence des coefficients et la convergence de la variance de l'erreur du prédicteur aux coefficients estimés vers la variance de l'erreur du prédicteur de Wiener-Kolmogorov. Pour obtenir ces convergences, Hidalgo et Yajima (2002) se placent toutefois dans un

contexte semi-paramétrique contrairement à Bhansali (1974), puisqu'ils supposent que la densité spectrale admet une singularité en une fréquence supposée connue qu'ils estiment. Dans le chapitre 3, nous comparerons empiriquement à l'aide de simulations numériques ce prédicteur avec celui dont les coefficients sont estimés grâce à l'estimateur de Whittle.

Toutes ces méthodes de prédiction sont donc caractérisées par Bhansali et Kokoszka (2001) comme des méthodes de type II car elles utilisent toutes le fait que le processus est à longue mémoire.

### 1.3.2 Cas du prédicteur projeté

Pour le prédicteur projeté, les coefficients du prédicteur sont solutions des équations de Yule-Walker et donc dépendent uniquement des autocovariances jusqu'à l'ordre  $k+h$  si  $k$  est le nombre d'observations dont on dispose et  $h$  l'horizon de la prévision. Pour estimer les coefficients du prédicteur, il suffit donc estimer les autocovariances jusqu'à l'ordre  $k+h$ . Pour cela soit on écrit les autocovariances comme fonction d'un paramètre que l'on estime, soit on se place dans un cadre non-paramétrique et on estime les covariances par les covariances empiriques. Nous nous restreindrons à la méthode non-paramétrique qui est celle étudiée en courte mémoire et présentée par Bhansali et Kokoszka (2001) comme méthode de type I.

L'erreur quadratique supplémentaire due à l'estimation des covariances a été estimée par Berk (1974), Bhansali (1978) et Lewis et Reinsel (1985) en courte mémoire (décroissance suffisamment rapide des coefficients de la représentation autorégressive). Ils en ont donné un équivalent asymptotique en fonction du nombre de données sur lequel on projette  $k$  et de la taille de l'échantillon pour prévoir  $T$ .

**Théorème 1.3.1** (Bhansali (1978)). *Soit  $(X_n)_{n \in \mathbb{Z}}$  un processus causal et inversible admettant donc les développements (1.6) et (1.5). On suppose que  $(\varepsilon_n)_{n \in \mathbb{Z}}$  a des moments d'ordre 4 finis que  $k = k_T$  est une fonction de  $T$  telle que :*

$$k_T \xrightarrow{T \rightarrow +\infty} +\infty, \quad \frac{k_T^3}{T} \xrightarrow{T \rightarrow +\infty} 0 \quad \text{et} \quad \sqrt{T} \sum_{j=k_T+1}^{+\infty} |a_j| \xrightarrow{T \rightarrow +\infty} 0. \quad (1.10)$$

*On suppose de plus qu'on dispose de deux réalisations indépendantes de  $(X_n)_{n \in \mathbb{Z}}$  :  $(X_j)_{j \in [1, k]}$  pour prédire et  $(Y_j)_{j \in [1, T]}$  pour estimer les  $a_{j, k}$  par  $\hat{a}_{j, k}$  en remplaçant les covariances par les covariances empiriques dans les équations de Yule-Walker.*

*On a alors la convergence en loi suivante :*

$$\sqrt{\frac{T}{k}} \sum_{j=1}^k (a_{j, k} - \hat{a}_{j, k}) X_{k+1-j} \xrightarrow{T \rightarrow +\infty} \mathcal{N}(0, \sigma_\varepsilon^2).$$

Les processus à longue mémoire ne satisfont pas la condition (1.10) en effet pour les processus FARIMA par exemple on a  $\sum_{j=k+1}^{+\infty} |a_j| \sim Ck^d$ .

Pour un processus  $(X_n)_{n \in \mathbb{Z}}$  à longue mémoire, la convergence des autocovariances empiriques définies par

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \bar{X}_n)(X_{i+h} - \bar{X}_n) \text{ avec } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

vers les autocovariances du processus a été démontrée par Hosking (1996).

**Théorème 1.3.2** (Hosking (1996)). *Soit  $(X_n)_{n \in \mathbb{Z}}$  un processus gaussien stationnaire causal vérifiant (1.1) avec  $L_1$  fonction constante.  $(X_n)_{n \in \mathbb{Z}}$  admet donc une représentation moyenne mobile (1.5) et on suppose que  $b_j \sim Cj^{d-1}$  lorsque  $j \rightarrow +\infty$ .*

- Si  $0 < d < \frac{1}{4}$ ,  $\sqrt{n}(\hat{\gamma}_n(h) - \gamma(h))$  tend en loi lorsque  $n$  tend vers l'infini vers une variable aléatoire gaussienne
- Si  $d = \frac{1}{4}$ ,  $\sqrt{\frac{n}{\log(n)}}(\hat{\gamma}_n(h) - \gamma(h))$  tend en loi lorsque  $n$  tend vers l'infini vers une variable aléatoire gaussienne
- Si  $\frac{1}{4} < d < \frac{1}{2}$ ,  $n^{1-2d}(\hat{\gamma}_n(h) - \gamma(h))$  tend en loi lorsque  $n$  tend vers l'infini vers une variable aléatoire de Rosenblatt.

Dans ces résultats on voit apparaître une difficulté par rapport à la courte mémoire car la vitesse de convergence des coefficients du prédicteur sera donnée par la vitesse de convergence des autocovariances or si  $d > 1/4$  la vitesse de convergence n'est plus  $\sqrt{T}$  mais  $\sqrt{T^{1-2d}}$  si  $T$  est le nombre d'observations. Des résultats de normalité asymptotique des estimateurs des coefficients du prédicteur ont aussi été directement prouvés en les considérant comme les coefficients d'un modèle autorégressif fini que l'on ajusterait à tort à des données présentant de la longue mémoire (voir Yajima (1993)). On y retrouve bien entendu la même dégradation de la vitesse de convergence lorsque  $d > 1/4$ . Ray (1993b) donne également un équivalent asymptotique de la matrice de covariance des coefficients estimés du prédicteur pour les processus fractionnaires. Les vitesses annoncées sont justes pour la convergence des coefficients mais pas les constantes. La démonstration est erronée car la formule permettant de calculer la matrice de covariance de ces coefficients est fautive. Le résultat exact est démontré par Yajima (1993).

**Résultats de la thèse** Puisque la vitesse de convergence des coefficients du prédicteur est plus faible en longue mémoire qu'en courte mémoire l'impact de l'estimation des coefficients du prédicteur sur la vitesse de convergence globale du prédicteur vers le prédicteur de Wiener-Kolmogorov sera donc plus fort dans notre cas. Dans le chapitre 4, on donnera une majoration de l'erreur quadratique supplémentaire due à l'estimation des coefficients du prédicteur tronqué lorsque le processus est à longue mémoire.

La résolution pratique des équations de Yule-Walker a donné lieu à de nombreux algorithmes comme celui de Durbin-Levinson (Brockwell et Davis (1991)) ou celui de Burg. L'algorithme d'innovation quant à lui permet de calculer récursivement le prédicteur projeté (Brockwell et Davis (1991)). Bondon (2001) a de plus proposé un algorithme de complexité moindre pour calculer ce prédicteur à l'horizon  $h$ .

### 1.3.3 Estimer et prévoir sur une même réalisation

Les travaux que nous avons décrits dans les deux paragraphes précédents sont faits sous l'hypothèse restrictive que l'on dispose de deux réalisations indépendantes du même processus : celle que l'on veut prédire et une autre servant à estimer les coefficients du prédicteur. Mais en pratique cette hypothèse est rarement vérifiée. En courte mémoire, on utilise le début de la série dont on dispose pour estimer la fonction de prédiction et la fin est considérée comme la série à prédire. Dans ce cas au moins la corrélation entre ces deux parties d'une même réalisation si elles sont suffisamment éloignées est faible, et on peut penser que les résultats de convergence des prédicteurs restent vrais dans cette situation. Ing et Wei (2003) se sont affranchis de cette hypothèse dans le cas du prédicteur projeté en courte mémoire et ils ont démontré que l'erreur quadratique due à l'estimation était du même ordre asymptotiquement que l'on travaille avec deux réalisations indépendantes comme par exemple Bhansali (1978) ou avec une seule comme eux.

**Théorème 1.3.3** (version pour les processus gaussiens du théorème 3 de Ing et Wei (2003)). *Soit  $(X_n)_{n \in \mathbb{Z}}$  un processus inversible et causal gaussien dont les coefficients de la représentation autorégressive donnée en (1.6) vérifient  $\sum_{j=1}^{+\infty} \sqrt{j}|a_j| < +\infty$ . On suppose de plus qu'il existe  $\delta > 0$  tel que  $K_T^{2+\delta} = O(T)$ . On dispose d'une série de longueur  $T$  : on utilise les  $k \in \llbracket 1, K_T \rrbracket$  dernières valeurs pour prédire et on les utilise toutes pour calculer les covariances empiriques et en déduire les  $\hat{a}_{j,k}$ . On a alors*

$$\lim_{T \rightarrow +\infty} \max_{1 \leq k \leq K_T} \left| \frac{\mathbb{E} \left( X_{k+1} - \left( - \sum_{j=1}^k \hat{a}_{j,k} X_{k+1-j} \right) \right)^2 - \sigma_\varepsilon^2}{L_T(k)} - 1 \right| = 0$$

où  $L_T(k) = \frac{k}{T} \sigma_\varepsilon^2 + E_k$  avec  $E_k$  l'erreur quadratique due à la projection (voir le chapitre (5) pour les détails).

**Résultats de la thèse** En longue mémoire, la justification empirique, qui utilise le fait que si l'on considère deux parties suffisamment éloignées de la même série elles sont faiblement corrélées, semble plus difficilement défendable. En effet les covariances dans ce cas décroissent lentement et il faudrait disposer d'une très longue série pour que cette approximation soit justifiée. Dans le chapitre 5, nous généraliserons les résultats de Ing et Wei (2003) à des processus à longue mémoire au prix d'hypothèses plus fortes sur la longueur des sous-séries utilisées pour estimer et prévoir.

#### Plan de la thèse

- **Chapitre 2** Convergence au sens  $L^2$  des prédicteurs linéaires finis (projeté et tronqué) vers le prédicteur de Wiener-Kolmogorov
- **Chapitre 3** Prédicteur de Wiener-Kolmogorov tronqué : erreur quadratique de prévision due à l'estimation des coefficients sur une réalisation indépendante

- **Chapitre 4** Prédicteur projeté sur un passé fini : erreur quadratique de prévision due à l'estimation des coefficients sur une réalisation indépendante et normalité asymptotique du prédicteur
- **Chapitre 5** Prédicteur projeté sur un passé fini : erreur quadratique de prévision due à l'estimation des coefficients sur la réalisation servant à la prévision et normalité asymptotique du prédicteur
- **Chapitre 6** Comparaison des prédicteurs précédemment étudiés sur des séries réelles et simulées

Cause Prédicteurs	observation d'un passé fini de longueur $k$	estimation de la fonction de prédiction estimée sur une réalisation indépendante de celle à prévoir de longueur $T$	estimation de la fonction de prédiction estimée sur la même réalisation avec $T$ observations
Wiener-Kolmogorov tronqué à l'ordre $k$	$k^{-1}$	$\frac{k^{2d}}{T}$	
Projeté sur $[X]_1^k$	$k^{-1}$	<ul style="list-style-type: none"> <li>• <math>\frac{k}{T}</math> pour <math>d &lt; \frac{1}{4}</math></li> <li>• <math>\frac{\sqrt{k} \log(T)}{T}</math> pour <math>d = \frac{1}{4}</math></li> <li>• <math>\frac{k^{1-2d}}{T^{2-4d}}</math> pour <math>d &gt; \frac{1}{4}</math></li> </ul>	$\frac{k}{T}$

TAB. 1.1 – Tableau récapitulatif des résultats : ordres de grandeur des erreurs quadratiques de prévision commises

## Chapitre 2

# Convergence des prédicteurs linéaires d'ordre fini dans $L^2$

En pratique pour faire de la prévision, on ne dispose jamais d'un nombre infini d'observations. On ne peut donc pas utiliser pour prévoir le prédicteur de Wiener-Kolmogorov. Comme nous l'avons expliqué dans l'introduction, on a alors deux possibilités : soit recourir au prédicteur de Wiener-Kolmogorov tronqué soit utiliser le prédicteur projeté. Dans ce chapitre nous allons tout d'abord répondre à la question suivante : lorsqu'on dispose d'un passé de longueur finie  $k$  quel est l'ordre de l'erreur quadratique de prévision due à cette absence de données ? En courte mémoire on sait que pour les deux prédicteurs la décroissance de cette erreur est géométrique (voir le théorème 1.2.2 démontré par Pourahmadi (1989)). On va ici montrer que en longue mémoire dans les deux cas elle est majorée par  $O(k^{-1})$  et que cette vitesse est atteinte par de nombreux processus à longue mémoire classiques par exemple les processus FARIMA ou GARMA. La dégradation de la vitesse de convergence entre le cas de la courte mémoire et celui de la longue mémoire était attendue car le passé qui est négligé est beaucoup plus dépendant de la variable aléatoire à prévoir en longue mémoire qu'en courte mémoire. Toutefois cette vitesse de convergence reste suffisante pour faire de la prévision, elle est par exemple du même ordre que la vitesse de convergence du théorème de la limite centrale.

Après avoir montré que les prédicteurs tronqué et projeté convergent avec une vitesse du même ordre vers le prédicteur de Wiener-Kolmogorov, on a ensuite comparé les vitesses réelles de convergence des deux prédicteurs c'est-à-dire les constantes que l'on trouve devant le facteur  $k^{-1}$ . Le prédicteur projeté, étant le prédicteur optimal connaissant un passé fini, est toujours plus efficace que le prédicteur tronqué. Toutefois plus la longue mémoire est importante c'est-à-dire plus le paramètre de longue mémoire  $d$  est proche de  $1/2$ , plus il est efficace par rapport au prédicteur tronqué dans le cas d'un processus fractionnaire où on peut calculer explicitement le développement asymptotique de l'erreur de prédiction.

Prévoir un processus à longue mémoire présente l'avantage de pouvoir prévoir à un horizon plus large. En effet en raison de la forte dépendance persistente entre les données,



une information substantielle sur le futur lointain est contenu dans les données. L'erreur quadratique du prédicteur de Wiener-Kolmogorov à l'horizon  $h$  tend plus lentement vers la variance du processus en longue mémoire qu'en courte mémoire. Lorsqu'on ne dispose pas du passé infini, il est donc intéressant de savoir jusqu'à quel horizon  $h$  on peut prédire connaissant un passé de longueur  $k$ . Nous démontrerons que si  $h = o(k)$ , le prédicteur de Wiener-Kolmogorov tronqué et donc le prédicteur projeté apporte une information par rapport à la prévision triviale par la moyenne du processus.

Les résultats de ce chapitre ont fait l'objet d'une publication sous le titre Linear Prediction of Long-Range Dependent Time Series dans E.S.A.I.M. Probabilités & Statistique (Godet (2008)).

## 2.1 Introduction

Let  $(X_n)_{n \in \mathbb{Z}}$  be a discrete-time (weakly) stationary process in  $L^2$  with mean 0 and  $\sigma$  its autocovariance function. We assume that the process  $(X_n)_{n \in \mathbb{Z}}$  admits an infinite moving average representation as follows :

$$X_n = \sum_{j=0}^{\infty} b_j \varepsilon_{n-j} \quad (2.1)$$

where  $(\varepsilon_n)_{n \in \mathbb{Z}}$  is a white noise series consisting of uncorrelated random variables, each with mean zero and variance  $\sigma_\varepsilon^2$ , 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 :

$$\varepsilon_n = \sum_{j=0}^{\infty} a_j X_{n-j}, \quad (2.2)$$

where the sequence  $(a_j)_{j \in \mathbb{N}}$  is absolutely summable and  $a_0 = 1$ . We also assume that  $(a_j)_{j \in \mathbb{N}}$  and  $(b_j)_{j \in \mathbb{N}}$ , occurring respectively in (2.2) and (2.1), satisfy the following conditions for all  $\delta > 0$  :

$$|a_j| \leq C_1 j^{-d-1+\delta} \quad (2.3)$$

$$|b_j| \leq C_2 j^{d-1+\delta} \quad (2.4)$$

where  $C_1$  and  $C_2$  are constants and  $d$  is a parameter verifying  $d \in ]0, 1/2[$ . The class of processes satisfying (2.3) and (2.4) includes long-memory processes i.e. processes with covariances that are not absolutely summable :

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

For example, FARIMA processes  $(X_n)_{n \in \mathbb{Z}}$ , the most studied long-memory processes, are stationary solutions to the difference equations :

$$\phi(B)(1-B)^d X_n = \theta(B)\varepsilon_n \quad (2.5)$$

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  $\phi$  and  $\theta$  are polynomials with no zeroes in the unit disk. The coefficients verify

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

and thus (2.3) and (2.4) hold.

More generally, conditions (2.3) and (2.4) hold when the coefficients are of the form :

$$|a_j| \underset{+\infty}{\sim} L(j)j^{-d-1} \quad (2.6)$$

$$|b_j| \underset{+\infty}{\sim} L'(j)j^{d-1} \quad (2.7)$$

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

(2.6) and (2.7) define again a class of long-memory process. Indeed Inoue (1997) proved that (2.7) implies that

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

where  $L$  is a slowly varying function and  $\beta$  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  $(\sigma(j))_{j \in \mathbb{N}}$  (see Inoue (2000)).

## 2.2 Wiener-Kolmogorov Next Step Prediction Theory

### 2.2.1 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 :

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

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

$$\widetilde{X}_k(1) = - \sum_{j=1}^{\infty} a_j X_{k+1-j}. \quad (2.8)$$

### 2.2.2 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 :

$$\widetilde{X}'_k(1) = - \sum_{j=1}^k a_j X_{k+1-j}. \quad (2.9)$$

It is equivalent to say that we have truncated the infinite series (2.8) to  $k$  terms. The following proposition provides the asymptotic properties of the mean-squared prediction error as a function of  $k$ .

**Theorem 2.2.1.** *Let  $(X_n)_{n \in \mathbb{Z}}$  be a linear stationary process defined by (2.1), (2.2) and verifying conditions (2.3) and (2.4). 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  $\sigma_\varepsilon^2$ , 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 hereafter denote  $\widetilde{\sigma}_k^2 := \mathbb{E}([X_{k+1} - \widetilde{X}'_k(1)]^2)$  the sum of these two errors. The proof of Theorem (2.2.1) hinges on the following lemma :

**Lemma 2.2.1.** *Under assumption (2.4), the autocovariance function  $\sigma$  of the process  $(X_n)_{n \in \mathbb{Z}}$  verifies :*

$$\forall \delta > 0, \exists C_3 \in \mathbb{R}, \quad |\sigma(j)| \leq C_3 j^{2d-1+\delta}. \quad (2.10)$$

*Proof.* Notice that it suffices to prove (2.10) for  $\delta$  near 0 in order to prove (2.10) for  $\delta > 0$  small enough. 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 2.2.1.

*Proof.*

$$X_{k+1} - \widetilde{X}'_k(1) = \varepsilon_{k+1} - \sum_{j=k+1}^{+\infty} a_j X_{k+1-j}. \quad (2.11)$$

The two parts of the sum (2.11) are orthogonal with respect to the inner product associated with the mean-squared norm. Consequently :

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

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

$$\widetilde{\sigma}_k^2 - \sigma_\varepsilon^2 = \sum_{j=k+1}^{\infty} \sum_{l=k+1}^{\infty} a_j a_l \sigma(l-j). \quad (2.12)$$

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 (2.3) and (2.10) and the triangular inequality, it follows that :

$$\begin{aligned} &\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) \\ &\leq C_1^2 C_3 \left( 2 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta} (j+1)^{-d-1+\delta} + \sum_{j=k+1}^{+\infty} (j^{-d-1+\delta})^2 \right) \end{aligned} \quad (2.13)$$

$$+ 2C_1^2 C_3 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta} \sum_{l=j+2}^{+\infty} l^{-d-1+\delta} |l-j|^{2d-1+\delta} \quad (2.14)$$

for all  $\delta > 0$ . Assume now that  $\delta < 1/2 - d$ . For the terms (2.13), 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 (2.14), we will similarly compare the series with an integral. In the next lemma, we establish a necessary result for this comparison :

**Lemma 2.2.2.** *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 positive integers. We assume that  $\delta < 1-2d$  and  $m \geq \frac{\delta-d-1}{\delta+2d-1}$  for all  $\delta \in \left] 0, \frac{\delta-d-1}{\delta+2d-1} \right[$ . 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 (2.13), 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 :

$$\begin{aligned} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) &\leq O(k^{-1+3\delta}) + O(k^{-2d-1}) \\ &\leq O(k^{-1+3\delta}). \end{aligned} \tag{2.15}$$

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.  $\square$

Theorem 2.2.1 is not interesting for short memory processes. Pourahmadi (1989) proves 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 2.2.1.** *Let  $(X_n)_{n \in \mathbb{Z}}$  be a linear stationary process defined by (2.1), (2.2) and verifying (2.6). Assume that the coefficients  $(a_j)_{j \in \mathbb{N}^*}$  are ultimately of constant sign. We assume also that the autocovariance function  $\sigma$  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  $\sigma$  is of a constant sign. Then the rate of convergence  $O(k^{-1})$  is optimal i.e. we have the following asymptotic equivalence :

$$\mathbb{E}([X_{k+1} - \tilde{X}_k(1)]^2) = \sigma_\varepsilon^2 + L''(k)k^{-1} + o(k^{-1}) \text{ as } k \rightarrow +\infty \quad (2.16)$$

where  $L''$  is a slowly varying function.

*Proof.* In this particular case, we can estimate the prediction error more precisely for large  $k$  :

$$\begin{aligned} & \left| \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) \right| \\ &= 2 \sum_{j=k+1}^{+\infty} |a_j| \sum_{l=j+1}^{+\infty} |a_l| |\sigma(l-j)| + \sum_{k+1}^{+\infty} a_j^2 \sigma(0) \\ &= \int_{k+1}^{+\infty} j^{-d-1} L(j) \int_{j+1}^{+\infty} l^{-d-1} L(l) (l-j)^{2d-1} L'(l-j) dl dj + \frac{k^{-2d-1} L^2(k)}{1+2d} + O(k^{-2d-1}) \\ &\sim \int_{k+1}^{+\infty} j^{-2} L^2(j) L'(j) \int_{1/j+1}^{+\infty} l^{-d-1} \frac{L(lj)}{L(j)} (l-1)^{2d-1} \frac{L'(j(l-1))}{L'(j)} dl dj \\ &\sim k^{-1} \frac{L^2(k) L'(k) \Gamma(1-d) \Gamma(2d)}{\Gamma(1+d)} \end{aligned} \quad (2.17)$$

$\square$

**Example 2.2.1.** *Assume now that  $(X_n)_{n \in \mathbb{Z}}$  is fractionally integrated noise, which is the stationary solution of the difference equation :*

$$X_n = (1 - B)^{-d} \varepsilon_n \quad (2.18)$$

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 et Davis (1991)). This is a particular case of the FARIMA model defined in (2.5) 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 (2.16) :

$$\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) \sim \frac{\Gamma(1-2d)\Gamma(2d)\sigma_\varepsilon^2}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)} k^{-1} \quad (2.19)$$

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  $d$  :

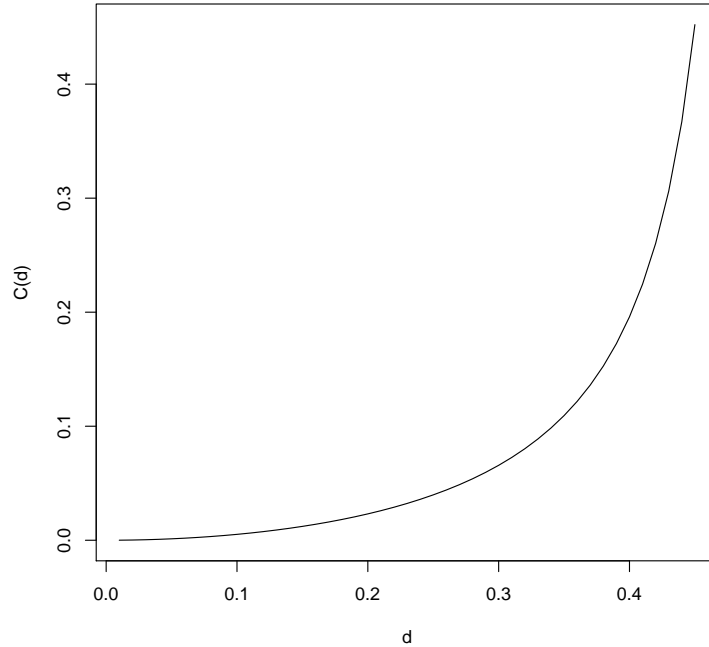
$$C(d) = \frac{\Gamma(1-2d)\Gamma(2d)\sigma_\varepsilon^2}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)}. \quad (2.20)$$

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

$$C(d) \sim \frac{\sigma_\varepsilon^2}{(1-2d)\Gamma(-1/2)^2\Gamma(1/2)\Gamma(3/2)}. \quad (2.21)$$

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

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

FIG. 2.1 – Behaviour of the constant  $C(d)$ ,  $d \in [0, 1/2[$ , defined in (2.20)

As Figure 2.1 suggests and the asymptotic equivalent given in (2.21) 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 2.2.2.** Long-memory processes with seasonal effects (see for example Viano et al. (1995)) provide another class of long-memory processes verifying assumptions of Theorem 2.2.1 but not of Proposition 2.2.1. 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 Hosking (1981) and studied for example by Gray et al. (1989). We first recall their definition.

**Definition 1.**  $(X_n)_{n \in \mathbb{Z}}$  is a GARMA process if it is the stationary solution of the equation :

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

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 in (2.2). The coefficients verify :

$$a_j = \frac{\cos[(j-d)\nu + (d\pi/2)]}{\Gamma(-d) \sin^{-d}(\nu)} \left(\frac{2}{j}\right)^{1+d} + O\left(j^{-5/2}\right) \quad (2.22)$$

The autocovariance function  $\sigma$  is such that :

$$\sigma(j) = K \cos(j\nu) j^{2d-1} (1 + O(j^{-1})) \quad (2.23)$$

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 2.2.2.** *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 the Appendix. □

After these results on the truncated Wiener-Kolmogorov predictor, we propose another predictor. Instead of truncating the series which defines the Wiener-Kolmogorov predictor to  $k$  terms, 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.

## 2.3 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 Ray (1993b) in the case of fractionally integrated noise (defined in (2.18)). Fitting a  $k$ -th order autoregressive process is equivalent to projecting onto the span of the subset  $\{X_k, \dots, X_1\}$ .

### 2.3.1 Definition of the predictor

Here  $(X_n)_{n \in \mathbb{Z}}$  is a long-memory process which verifies the assumptions (2.1), (2.2), (2.3) and (2.4) of Section 2.1. 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 equations, which are called the  $k^{\text{th}}$  order Yule-Walker equations :

$$\forall j \in \llbracket 1, k \rrbracket, \quad \sum_{i=1}^k a_{i,k} \sigma(i-j) = -\sigma(j) \quad (2.24)$$

The mean-squared prediction error is :

$$\mathbb{E}[(\widehat{X}_k(1) - X_{k+1})^2] = \sigma_\varepsilon^2 + \mathbb{E} \left[ \left( - \sum_{j=1}^{\infty} a_j X_{k+1-j} + \sum_{j=1}^k a_{j,k} X_{k+1-j} \right)^2 \right]. \quad (2.25)$$

### 2.3.2 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 2.3.1.** *Let  $(X_n)_{n \in \mathbb{Z}}$  be a linear stationary process defined by (2.1), (2.2) and verifying conditions (2.3) and (2.4) of Section 2.1. We can approximate the mean-squared prediction error of  $\widehat{X}_k(1)$  by :*

$$\mathbb{E}[(\widehat{X}_k(1) - X_{k+1})^2] - \sigma_\varepsilon^2 = O(k^{-1}) \text{ as } k \rightarrow +\infty. \quad (2.26)$$

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 when using  $k$  observations, the error by using truncation is larger and since the truncation method involves an error bounded by  $O(k^{-1})$ , we obtain (2.26).

Consequently we only need prove that this rate of convergence is attained . This is the case for the fractionally integrated processes defined in (2.18). 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 developing (2.25)

$$\widehat{\sigma}_k^2 - \sigma_\varepsilon^2 = \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} (a_j - a_{j,k})(a_l - a_{l,k})\sigma(l-j) \quad (2.27)$$

$$= \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sum_{l=0}^{+\infty} a_l \sigma(l-j) - \sum_{l=0}^k a_{l,k} \sum_{j=0}^{+\infty} (a_j - a_{j,k})\sigma(l-j) \quad (2.28)$$

We first study the first term of the sum (2.28). 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  $\varepsilon_n$  is orthogonal to  $X_{n-j}$  for  $j > 0$ . We can thus rewrite the first term of (2.28) :

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

since  $a_0 = a_{0,k} = 1$  according to the definition. Next we study the second term of the sum in (2.28). And we obtain :

$$\begin{aligned} \sum_{l=0}^k a_{l,k} \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sigma(l-j) &= \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=1}^k (a_j - a_{j,k}) \sigma(l-j) \\ &\quad + \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=k+1}^{+\infty} a_j \sigma(l-j) \end{aligned} \quad (2.29)$$

$$\begin{aligned} &\quad + \sum_{l=0}^k a_l \sum_{j=1}^k (a_j - a_{j,k}) \sigma(l-j) \quad (2.30) \\ &\quad + \sum_{l=0}^k a_l \sum_{j=k+1}^{+\infty} a_j \sigma(l-j). \end{aligned}$$

Similarly we rewrite the term (2.29) 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 (2.29) is equal to the opposite of (2.30). Hence it follows that :

$$\begin{aligned} \sum_{l=0}^k a_{l,k} \sum_{j=0}^{+\infty} (a_j - a_{j,k}) \sigma(l-j) &= \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=1}^k (a_j - a_{j,k}) \sigma(l-j) \\ &\quad + 2 \sum_{l=1}^k (a_{l,k} - a_l) \sum_{j=k+1}^{+\infty} a_j \sigma(l-j) \\ &\quad + \sum_{l=0}^k a_l \sum_{j=k+1}^{+\infty} a_j \sigma(l-j). \end{aligned} \quad (2.31)$$

In a similar way we can rewrite the third term of the sum in (2.31) 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 (2.12).

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 (2.31). For 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  $a_{l,k} - a_l$  we use the explicit formula given in Brockwell et Davis (1988) and we easily obtain that  $a_{l,k} - a_l$  is negative for all  $l \in \llbracket 1, k \rrbracket$ .

$$\begin{aligned} a_l - a_{l,k} &= \frac{\Gamma(l-d)}{\Gamma(l+1)\Gamma(-d)} - \frac{\Gamma(k+1)\Gamma(l-d)\Gamma(k-d-l+1)}{\Gamma(k-l+1)\Gamma(l+1)\Gamma(-d)\Gamma(k-d+1)} \\ &= -a_l \left( -1 + \frac{\Gamma(k+1)\Gamma(k-d-l+1)}{\Gamma(k-l+1)\Gamma(k-d+1)} \right) \\ &= -a_l \left( \frac{k \dots (k-j+1)}{(k-d) \dots (k-d-j+1)} - 1 \right) > 0 \end{aligned} \quad (2.32)$$

since  $\forall j \in \mathbb{N}^*$ ,  $a_j < 0$ . To give an asymptotic equivalent of the prediction error, we use the sum given in (2.31). 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  $a_l - a_{l,k}$  given the explicit formula in (2.32) :

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

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  $k \geq K$  :

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

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 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 consequently have 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 quantity :

$$r(k) := \frac{(\tilde{\sigma}_k^2 - \sigma_\varepsilon^2) - (\hat{\sigma}_k^2 - \sigma_\varepsilon^2)}{\tilde{\sigma}_k^2 - \sigma_\varepsilon^2} \quad (2.33)$$

which is the ratio of the difference between the two prediction errors (prediction error due to the truncation (2.12) and prediction error due to the projection (2.27)) and the prediction error due to the truncation. By (2.31), 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 (2.18). Figure 2.2 shows that the prediction by truncation incurs a larger performance loss when  $d \rightarrow 1/2$ . The improvement reaches 50 per cent when  $d > 0.3$  and  $k > 20$ .

## 2.4 The Multistep Predictors

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

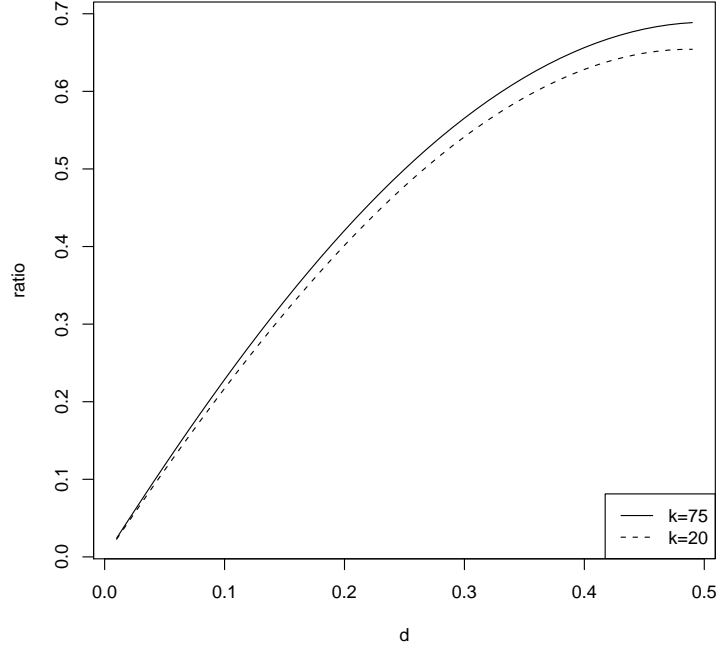
### 2.4.1 The Wiener-Kolmogorov predictor

Since we assume that the process  $(X_n)_{n \in \mathbb{Z}}$  has an autoregressive representation (2.2) and moving average representation (2.1), 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 :

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

(see for example Theorem 5.5.1 of Brockwell et Davis (1991)). The corresponding mean squared error of prediction is :

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

FIG. 2.2 – Ratio  $r(k)$ ,  $d \in ]0, 1/2[$  defined in (2.33)

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 (2.7) we can elucidate the asymptotic behaviour of the mean-squared error. We have as  $h$  tends to infinity :

$$\sigma(0) - \mathbb{E} \left[ \left( \widehat{X}_k(h) - X_{k+h} \right)^2 \right] = \sigma_\varepsilon^2 \sum_{j=h}^{+\infty} b_j^2 \quad (2.34)$$

$$\begin{aligned} &\sim \sum_{j=h}^{+\infty} j^{2d-2} L'^2(j) \\ &\sim \int_h^{+\infty} j^{2d-2} L'^2(j) dj \\ &\underset{h \rightarrow +\infty}{\sim} \frac{1}{1-2d} h^{2d-1} L'^2(h) \end{aligned} \quad (2.35)$$

according to Proposition 1.5.10 of Bingham *et al.* (1987).

On the contrary, for short memory processes we 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 Brockwell et Davis (1991)). The moving average coefficients  $b_j$  are bounded by :

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

where  $\rho$  verifies  $|\rho| < 1$  and  $m$  is a non negative integer (see for example Brockwell et Davis (1991) 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 truncation effect and the projection effect.

#### 2.4.2 Truncated Wiener-Kolmogorov predictor and the linear least-squares predictor

In practice, we only observe a finite number  $k$  of samples  $(X_1, \dots, X_k)$  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 of order  $k$  recursively as :

$$\widetilde{X}'_k(h) = - \sum_{j=1}^{h-1} a_j \widetilde{X}'_k(h-j) - \sum_{j=1}^k a_{h-1+j} X_{k+1-j}. \quad (2.36)$$

The following proposition describes the asymptotic behaviour of the mean-squared error of the predictor (2.36).

**Proposition 2.4.1.** *Let  $(X_n)_{n \in \mathbb{Z}}$  be a linear stationary process defined by (2.1), (2.2) and verifying (2.3) and (2.4). We can approximate the mean-squared prediction error of  $\widetilde{X}'_k(h)$  by :*

$$\forall \delta > 0, \mathbb{E} \left[ \widetilde{X}'_k(h) - X_{k+h} \right]^2 = \sum_{l=0}^{h-1} b_l^2 \sigma_\varepsilon^2 + O \left( h^{2d+\delta} k^{-1+\delta} \right) \text{ as } k \rightarrow +\infty. \quad (2.37)$$



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

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

We proceed by induction on  $h$  to show that

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

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 :

$$\widetilde{X}'_k(h) - X_{k+h} = -\sum_{l=0}^{h-1} b_l \varepsilon_{k+h-l} + \sum_{j=k+1}^{+\infty} \sum_{m=0}^{h-1} a_{j+m} b_{h-1-m} X_{k+1-j}. \quad (2.38)$$

Since the process  $(\varepsilon_n)_{n \in \mathbb{Z}}$  is uncorrelated, the two terms of the sum (2.38) are orthogonal. We can rewrite the mean-squared error :

$$\mathbb{E} \left[ \widetilde{X}'_k(h) - X_{k+h} \right]^2 = \sum_{l=0}^{h-1} b_l^2 \sigma_\varepsilon^2 \quad (2.39)$$

$$+ \mathbb{E} \left[ \sum_{j=k+1}^{+\infty} \left( \sum_{m=0}^{h-1} a_{j+h-1-m} b_m \right) X_{k+1-j} \right]^2. \quad (2.40)$$

The first part of the error (2.39) is due to the prediction method and the second (2.40) due to the truncation the predictor. We now approximate the error term (2.40) by using (2.3) and (2.4). We obtain the following upper bound :

$$\begin{aligned} \forall \delta > 0, \left| \sum_{m=0}^{h-1} a_{j+h-1-m} b_m \right| &\leq \sum_{m=1}^{h-1} |a_{j+h-1-m} b_m| + |b_0 a_{j+h-1}| \\ &\leq C_1 C_2 \int_0^h (j+h-1-l)^{-d-1+\delta} l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ &\leq C_1 C_2 h^{-1+2\delta} \int_0^1 \left( \frac{j}{h} + 1 - l \right)^{-d-1+\delta} l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ &\leq C_1 C_2 h^{-1+2\delta} j^{-d-1+\delta} \int_0^1 \left( \frac{1}{h} + \frac{1-l}{j} \right)^{-d-1+\delta} l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ &\leq C_1 C_2 h^{d+2\delta} j^{-d-1+\delta} \int_0^1 l^{d-1+\delta} dl + C_1 (j+h)^{-d-1} \\ \left| \sum_{m=0}^h a_{j+h-m} b_m \right| &\leq C_1 C_2 \frac{h^{d+2\delta}}{d} j^{-d-1+\delta}. \end{aligned} \quad (2.41)$$

Using Theorem 2.2.1 for one-step prediction, we conclude the proof.  $\square$

Having  $k$  observations, we search for a sufficient condition to have the variance of the  $h$ -step predictor bounded by  $\sigma(0)$ . More precisely, we want to select  $h$  so that the prediction error is negligible with respect to the information (2.34) given by the linear least-squares predictor given the infinite past. We now consider that the moving average coefficients  $b_j$  verify (2.7) and the autoregressive coefficients  $a_j$  verify (2.3). In this case, the prediction error has the asymptotic bound  $O(h^{2d} k^{-1})$  and the information (2.34) given by the least-squares predictor is equivalent to  $h^{2d-1} L'(h)$  with  $L'$  a slowly varying function (see (2.35)). It is then sufficient to verify :

$$h^{2d} k^{-1} = o(h^{2d-1}) \quad (2.42)$$

and then  $h = o(k)$  to obtain a prediction error negligible with respect to the available information (2.34). With the truncated Wiener-Kolmogorov predictor, it is interesting

to compute the  $h$ -step predictor if we have  $k$  observations  $h = o(k)$ .

We now 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 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 :

$$(c_{j,k})_{1 \leq j \leq k} = -\Sigma_k^{-1}(\sigma_{h-1+j})_{1 \leq j \leq k} \quad (2.43)$$

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

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

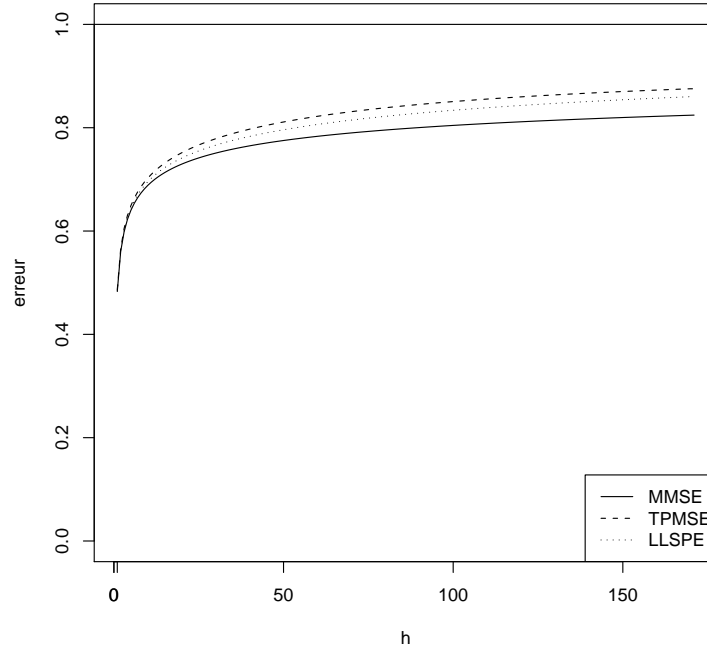
By construction of  $\widehat{X}_k(h)$ , the mean-squared error of prediction is lower than  $\sigma(0)$  for any  $k$  and  $h$ . However it can be interesting to look for a condition on  $k$  and  $h$  to have the prediction error due to the projection negligible with respect to the information (2.34) 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 Figure 2.3). 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 2.4.1 and condition (2.42), the prediction error due to the projection is negligible with respect to the information (2.34) given by the Wiener-Kolmogorov predictor. On Figure 2.3, we show that the behaviours of the truncated predictor and 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 (2.34) to obtain a good predictor in the case of fractionally integrated noise defined in (2.18) with  $k = 80$  and  $h \in [1, 150]$ .

## 2.5 Appendix : Proof of Proposition 2.2.2

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 (2.22)

FIG. 2.3 – 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 (2.18)



and (2.23). We obtain :

$$\begin{aligned}
 & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \sigma(l-j) \\
 = & \sum_{j=k+1}^{+\infty} a_j^2 \sigma(0) + 2 \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} a_j a_l \sigma(l-j) \\
 = & \sum_{j=k+1}^{+\infty} \left[ \left( \frac{\cos[(j-d)\nu + (d\pi/2)]}{\Gamma(-d) \sin^{-d}(\nu)} \right)^2 \left( \frac{2}{j} \right)^{2+2d} + O\left(j^{-7/2-d}\right) \right] \sigma(0) \quad (2.44)
 \end{aligned}$$

$$+ \frac{2^4 \sigma(0) \sigma_\varepsilon^2 \sin(d\pi) \Gamma(1-2d)}{\pi (\Gamma(-d))^2} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} B_{j,l} \quad (2.45)$$

where  $B_{j,l}$  is :

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

The first term (2.44) is bounded by  $O(k^{-2d-1})$  and thus negligible with respect to  $k^{-1}$ . For the second term (2.45) :

$$\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} B_{j,l} \\ = \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[ (j-d)\nu + \frac{d\pi}{2} \right] \cos \left[ (l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \quad (2.46)$$

$$+ \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[ (j-d)\nu + \frac{d\pi}{2} \right] \cos \left[ (l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O(j^{d-1/2}) \quad (2.47)$$

$$+ \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[ (j-d)\nu + \frac{d\pi}{2} \right] \cos \left[ (l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O(l^{d-1/2}) \quad (2.48)$$

$$+ \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[ (j-d)\nu + \frac{d\pi}{2} \right] \cos \left[ (l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O((j-l)^{-1}) \quad (2.49)$$

We first prove that the terms (2.47), (2.48) and (2.49) are bounded by  $O(k^{d-3/2})$  and thus negligible with respect to  $k^{-1}$ . For example, we study (2.47) :

$$\left| \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[ (j-d)\nu + \frac{d\pi}{2} \right] \cos \left[ (l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} O(j^{d-1/2}) \right| \\ \leq C \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} j^{-3/2} l^{-1-d} (j-l)^{2d-1} \\ = O\left(k^{d-3/2}\right).$$

The calculations for (2.48) and (2.49) have the same form. The dominant term of (2.45) is then (2.46). (2.46) can be split into three parts :

$$\begin{aligned} & \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos \left[ (j-d)\nu + \frac{d\pi}{2} \right] \cos \left[ (l-d)\nu + \frac{d\pi}{2} \right] \cos((j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \\ = & \frac{1}{2} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \end{aligned} \quad (2.50)$$

$$+ \frac{1}{2} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} \cos(2(j-l)\nu) (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \quad (2.51)$$

$$+ \frac{1}{4} \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{j-1} [\cos[(2j-2d)\nu + d\pi] + \cos[(2l-2d)\nu + d\pi]] (jl)^{-1-d} \frac{\Gamma(j-l+2d)}{\Gamma(j-l+1)} \quad (2.52)$$

We have proved in Proposition 2.2.1 that (2.50) is equivalent to  $Ck^{-1}$ . And it is easy to show (by Abel transformations), that (2.51) and (2.52) are negligible with respect to (2.50). This concludes the proof.

## Chapitre 3

# Erreur quadratique due à l'estimation paramétrique des coefficients du prédicteur tronqué

Dans le chapitre précédent, nous avons proposé deux prédicteurs lorsqu'on dispose d'une série de longueur  $k$  pour prédire et que l'on connaît la loi du processus : le prédicteur tronqué et le prédicteur projeté. Nous allons ici restreindre notre étude au prédicteur tronqué et étudier l'erreur quadratique commise lorsqu'on ne connaît pas la structure stochastique du processus et qu'il faut donc estimer les coefficients du prédicteur.

On suppose ici que l'on veut prédire les processus  $(X_n)_{n \in \mathbb{Z}}$  en disposant d'une réalisation de longueur  $k$ . Pour l'estimation de la fonction de prédiction, on observe une série indépendante  $(Y_n)_{n \in \mathbb{Z}}$  de longueur  $T$  suivant la même loi que  $(X_n)_{n \in \mathbb{Z}}$ . On suppose de plus que le processus  $(X_n)_{n \in \mathbb{Z}}$  appartient à une famille paramétrique et plus précisément que sa densité spectrale  $f(\cdot, \theta_0)$  appartient à la famille  $\mathcal{F}_\Theta := \{f(\cdot, \theta), \theta \in \Theta \text{ compact}\}$ . Pour estimer le paramètre, on utilise l'estimateur de Whittle (voir Fox et Taquq (1986)) :

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

où  $I_T$  est le périodogramme

$$I_T(\lambda) = \frac{|\sum_{j=1}^T e^{ij\lambda} (Y_j - \bar{Y}_T)|^2}{2\pi T} ..$$

On remplace ensuite dans le prédicteur tronqué :

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

les coefficients par leurs estimations. On obtient comme nouveau prédicteur :

$$\widetilde{X}_{T,k}(1) = - \sum_{j=1}^k a_j(\hat{\theta}_T) X_{k+1-j}.$$

On se place sous les hypothèses de Fox et Taqqu (1986) qui assurent la convergence presque sûre et en loi de l'estimateur, plus quelques hypothèses de régularité supplémentaires pour la famille de densités spectrales. On encadre alors l'erreur quadratique de prévision supplémentaire due à la prévision des coefficients  $a_j$  par

$$\lim_{T \rightarrow +\infty} T \mathbb{E} \left( \widetilde{X}_{T,k}(1) - \widetilde{X}_k(1) \right)^2 \leq C k^{2d}$$

avec  $C$  constante indépendante de  $k$ . Ce résultat (voir le théorème 3.2.1) est d'abord montré pour le prédicteur à l'horizon 1 puis généralisé pour le prédicteur à l'horizon  $h$ . Dans le cas particulier des processus FARIMA, on démontre un résultat de normalité asymptotique entre le prédicteur de Wiener-Kolmogorov et le prédicteur de Wiener-Kolmogorov tronqué dont les coefficients sont estimés. Ce résultat permettra notamment d'obtenir des intervalles de confiance asymptotiques pour ce prédicteur. L'étude de ce prédicteur se conclut par des simulations numériques. Notre prédicteur est enfin comparé au prédicteur optimal aux coefficients estimés (voir le chapitre 5) et au prédicteur tronqué dont les coefficients sont estimés par une méthode semi-paramétrique (voir Hidalgo et Yajima (2002)).

### 3.1 Introduction

The Wiener-Kolmogorov prediction theory is one of the most popular methods to predict stationary time series as introduced simultaneously in Wiener (1949) and Kolmogorov (1941) and then described in Whittle (1983). Its principle is to minimise the mean-squared error of prediction by projecting the future on the known infinite past. In practice there are two restrictions to this predictor : we never have access to the infinite past and we do not know precisely the stochastic structure of the process. Two methods of linear prediction are often proposed when we only know a finite past of length  $k$  : the truncation of the infinite series in the definition of the Wiener-Kolmogorov predictor and the projection onto the available data. When the stochastic structure of the process is unknown, we also need to estimate the forecast coefficients. Here we only study the truncated Wiener-Kolmogorov predictor.

In the short memory case, the additional errors related to the truncation and the estimation are well documented. On the first hand, the prediction error due to the truncation is computed in Pourahmadi (1989) for processes with a spectral density which is bounded and with a strictly positive lower bound. The mean-squared error of prediction decays exponentially fast with the length of the known past. On the other hand, the estimation of the model has been investigated in Akaike (1970) in the case of finite autoregressive processes and in Bhansali (1977) for processes with a spectral density



satisfying a uniform Lipschitz condition. They both worked in a non-parametric context and proposed an approximation of the mean-squared error of prediction. They also minimised the error of prediction under conditions on the length of the series used to estimate the forecast coefficients.

However the aforementioned works do not tackle the long-memory case. In Godet (2008), we estimated the mean-squared error of prediction due to the truncation for long-memory time series. It decays only hyperbolically fast as  $k$  goes to infinity. In a semi-parametric context, Hidalgo et Yajima (2002) proved that the mean-squared error of the truncated predictor with estimated coefficients converges to the mean-squared error of the Wiener Kolmogorov predictor but do not estimate the rate of convergence.

In this paper we assume a parametric model for long-memory time series and estimate the forecast coefficients on a realisation of length  $T$ . We consider the effect of the estimation on the prediction error of the truncated Wiener-Kolmogorov predictor. We estimate the forecast coefficients with the Whittle estimator (see Fox et Taqqu (1986)). Under the simplifying assumption that the series used for estimation and the series used for prediction are generated from two independent processes with the same stochastic structure, we propose an approximation of the mean-squared prediction error when the number  $T$  of observations used to estimate the forecast coefficients, and the number  $k$  of observations used in the actual predictor, both tend to infinity. This paper is organised as follows. In Section 3.2, in a parametric context, we estimate the forecast coefficients of the truncated Wiener-Kolmogorov predictor and discuss the asymptotic properties of the associated mean-squared error. In Section 3.3, we discuss the results and obtain a central limit theorem for FARIMA processes. We illustrate the quality of the predictor for fractionally integrated noise with numerical simulations. Subsidiary proofs are given in the Appendix.

## 3.2 The Truncated Wiener-Kolmogorov Predictor

### 3.2.1 Prediction Method

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

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

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

$$X_n = \sum_{j=0}^{\infty} b_j \varepsilon_{n-j} \tag{3.1}$$

where  $(\varepsilon_n)_{n \in \mathbb{Z}}$  is a white noise series consisting of uncorrelated random variables, each with mean zero and variance  $\sigma_\varepsilon^2$ . The sequence  $(b_j)_{j \in \mathbb{N}}$  is square-summable and  $b_0 = 1$ .

We shall further assume that  $(X_n)_{n \in \mathbb{Z}}$  admits an infinite autoregressive representation :

$$\varepsilon_n = \sum_{j=0}^{\infty} a_j X_{n-j}, \quad (3.2)$$

where the  $(a_j)_{j \in \mathbb{N}}$  are absolutely summable and  $a_0 = 1$ .

Since we have the representations (3.1) and (3.2), we can use the classical Wiener-Kolmogorov theory. The Wiener-Kolmogorov predictor of  $X_{k+h}$ , denoted by  $X_k^{\text{WK}}(h)$ , is the best linear predictor in the least-squares sense with full knowledge of the past  $(X_j, -\infty \leq j \leq k)$ . Since  $X_k^{\text{WK}}(h)$  is the projection of  $X_{k+h}$  on the infinite past, we can write  $X_k^{\text{WK}}(h)$  as :

$$X_k^{\text{WK}}(h) = \sum_{j=1}^{+\infty} c_j(h) X_{k+1-j} \quad (3.3)$$

where  $c_j(h) = -\sum_{m=0}^{h-1} a_{j+m} b_{h-1-m}$  (see Whittle (1983)). In the particular case of one-step prediction ( $h = 1$ ) the expression of  $c_j(1)$  is simplified and we have for all  $j \in \mathbb{N}^*$ ,  $c_j(1) = -a_j$ . In practice, we only observe a finite number of samples  $(X_1, \dots, X_k)$ . The infinite sum defined in (3.3) is truncated at order  $k$  and we have the following new predictor :

$$\widetilde{X}_k(h) = \sum_{j=1}^k c_j(h) X_{k+1-j}. \quad (3.4)$$

The asymptotic error due to the truncation has been studied in Godet (2008).

We thereafter study the asymptotic behaviour of the mean-squared error of prediction that is related to the estimation of the forecast coefficients  $c_j(h)$ . Section 3.2.2 tackles the particular case  $h = 1$ . We then generalise our result for  $h$ -step prediction in Section 3.2.3.

### 3.2.2 Estimates of Forecast Coefficients and Associated Mean-Squared Error for the Next Step Predictor

Under the assumptions of Section 3.2.1 we estimate the mean-squared error between the predictor  $\widetilde{X}_k(1)$  defined in (3.4) 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  $(Y_j)_{j \in [1, T]}$  that is independent of the process  $(X_n)_{n \in \mathbb{Z}}$  but shares the same stochastic structure.

More precisely, in a parametric approach, we assume that the process  $(X_n)_{n \in \mathbb{Z}}$  has a spectral density  $f(\cdot, \theta)$  which belongs to a class of functions parametrised by  $\theta$ .

Let  $\theta_0$  be the true value of the parameter. We estimate the autoregressive coefficients  $(a_j(\theta_0))_{1 \leq j \leq k}$  with  $\widehat{a}_j := a_j(\widehat{\theta}_T)$  where  $\widehat{\theta}_T$  is the Whittle estimate of  $\theta_0$ . We define the

Whittle estimate by (see Fox et Taqqu (1986)) :

$$\hat{\theta}_T = \operatorname{argmin}_{\theta \in \Theta} \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} [f(\lambda, \theta)]^{-1} I_T(\lambda) d\lambda \right] \quad (3.5)$$

where  $I_T$  is the periodogram :

$$I_T(\lambda) = \frac{|\sum_{j=1}^T e^{ij\lambda} (Y_j - \bar{Y}_T)|^2}{2\pi T}. \quad (3.6)$$

Consequently our predictor is given by

$$\widetilde{X_{T,k}}(1) = - \sum_{j=1}^k a_j(\hat{\theta}_T) X_{k+1-j} \quad (3.7)$$

where  $\hat{\theta}_T$  is defined in (3.5).

When the process is a Gaussian long-memory time series (see Fox et Taqqu (1986)), standard sufficient conditions on the parametric family of spectral densities ensure that the estimated vector  $\hat{\theta}_T$  converges almost surely to the true parameter  $\theta_0$ . These are recalled below. We define the parametric family of spectral density  $\mathcal{F}_\Theta := \{f(\cdot, \theta), \theta \in \Theta \text{ compact}\}$ . If  $\theta$  and  $\theta'$  are distinct elements of  $\Theta$ , we suppose that the set  $\{\lambda : f(\lambda, \theta) \neq f(\lambda, \theta')\}$  has positive Lebesgue measure. We say that  $\mathcal{F}_\Theta$  satisfies conditions B0-B3 if for all  $f(\cdot, \theta) \in \mathcal{F}_\Theta$  there exists a continuous function  $0 < \alpha(\theta) < 1$  and constants  $C(\delta)$  and  $C_0(\delta)$  such that for each  $\delta > 0$ ,

B0.  $f(\lambda, \theta) = |\lambda|^{-\alpha(\theta)} L(\lambda, \theta)$  with  $L(\cdot, \theta)$  bounded, bounded away from 0 and differentiable at 0. Then we have :

$$\int_{-\pi}^{\pi} |\log f(\lambda, \theta)| d\lambda < +\infty. \quad (3.8)$$

This condition guarantees that the processes with spectral density in  $\mathcal{F}_\Theta$  admit an infinite autoregressive and an infinite moving average representation. We denote by  $a_j(\theta)$  the coefficients of the autoregressive representation. Moreover, we suppose that they verify

$$|a_j(\theta)| \leq C(\delta) j^{-\alpha(\theta)/2-1+\delta}. \quad (3.9)$$

We assume also that the functions  $f(\lambda, \theta)$  are normalised so that

$$\int_{-\pi}^{\pi} \log f(\lambda, \theta) d\lambda = 0.$$

B1.  $f(\lambda, \theta)$  is continuous at all points  $(\lambda, \theta) \in [-\pi, \pi] \setminus \{0\} \times \Theta$  and

$$C_0(\delta) |\lambda|^{-\alpha(\theta)+\delta} \leq f(\lambda, \theta) \leq C(\delta) |\lambda|^{-\alpha(\theta)-\delta}.$$

B2.  $\frac{\partial}{\partial\theta_j}f(\lambda, \theta)$ ,  $\frac{\partial^2}{\partial\theta_j\partial\theta_l}f(\lambda, \theta)$  and  $\frac{\partial^3}{\partial\theta_j\partial\theta_l\partial\theta_k}f(\lambda, \theta)$  are continuous at all  $(\lambda, \theta) \in [-\pi, \pi] \setminus \{0\} \times \Theta$  and

$$\forall 1 \leq j \leq p, \quad \left| \frac{\partial}{\partial\theta_j}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-\delta},$$

$$\forall 1 \leq j, l \leq p \quad \left| \frac{\partial^2}{\partial\theta_j\partial\theta_l}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-\delta}$$

and

$$\forall 1 \leq j, l, k \leq p \quad \left| \frac{\partial^3}{\partial\theta_j\partial\theta_l\partial\theta_k}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-\delta}$$

B3.  $\frac{\partial}{\partial\lambda}f(\lambda, \theta)$ ,  $\frac{\partial^2}{\partial\lambda\partial\theta_j}f(\lambda, \theta)$ ,  $\frac{\partial^3}{\partial\lambda\partial\theta_j\partial\theta_l}f(\lambda, \theta)$ ,  $\frac{\partial^4}{\partial\lambda\partial\theta_j\partial\theta_l\partial\theta_k}f(\lambda, \theta)$  and  $\frac{\partial^3}{\partial^2\lambda\partial\theta_j}f(\lambda, \theta)$  are continuous at all points  $(\lambda, \theta) \in [-\pi, \pi] \setminus \{0\} \times \Theta$  and

$$\left| \frac{\partial}{\partial\lambda}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-1-\delta},$$

$$\forall 1 \leq j \leq p, \quad \left| \frac{\partial^2}{\partial\lambda\partial\theta_j}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-1-\delta},$$

$$\forall 1 \leq j, l \leq p, \quad \left| \frac{\partial^3}{\partial\lambda\partial\theta_j\partial\theta_l}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-1-\delta},$$

$$\forall 1 \leq j, l, k \leq p, \quad \left| \frac{\partial^4}{\partial\lambda\partial\theta_j\partial\theta_l\partial\theta_k}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-1-\delta}$$

and

$$\forall 1 \leq j \leq p, \quad \left| \frac{\partial^3}{\partial^2\lambda\partial\theta_j}f(\lambda, \theta) \right| \leq C(\delta)|\lambda|^{-\alpha(\theta)-2-\delta}.$$

The constants  $C(\delta)$  and  $C_0(\delta)$  which appear in the conditions B0-B3 are required to be independent of  $\theta$ .

**Example** A fractionally integrated process  $(X_n)_{n \in \mathbb{Z}}$  is the stationary solution of the difference equation :

$$(I - B)^d X_n = \varepsilon_n \tag{3.10}$$

where  $(\varepsilon_n)_{n \in \mathbb{Z}}$  is a white noise series. In this case, the parameter  $\theta$  is given by  $(d, \sigma_\varepsilon^2)$  and corresponds to the long-memory parameter  $d$  and the variance of the white noise  $\sigma_\varepsilon^2$ . In the following we denote by  $d_0$  the corresponding true value of the parameter  $d$ . Assumptions B0-B3 hold for fractionally integrated processes since the spectral density  $f(\cdot, d, \sigma_\varepsilon^2)$  of this process is given by  $f(\lambda, d, \sigma_\varepsilon^2) = |1 - \exp(i\lambda)|^{-2d} \sigma_\varepsilon^2 / 2\pi$ . The coefficients  $a_j$  are given by :

$$a_j(d, \sigma_\varepsilon^2) := \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \leq \frac{\Gamma(2-d)}{\Gamma(2)\Gamma(-d)} j^{-d-1}.$$

As a consequence, Theorem 3.2.1 can be applied on the class of fractionally integrated noise because the assumptions B0-B3 hold. Similarly a FARIMA process  $(X_n)_{n \in \mathbb{Z}}$  is the stationary solution of the difference equations :

$$(I - B)^d \phi(B) X_n = \psi(B) \varepsilon_n \quad (3.11)$$

with  $\phi$  and  $\psi$  polynomials with no zeroes on the closed unit circle (see Brockwell et Davis (1988)). The class of FARIMA time series also verifies the assumptions B0-B3.

We can now express the asymptotic behaviour of the mean-squared prediction error due to the estimation of the forecast coefficients.

**Theorem 3.2.1.** *Let  $\mathcal{F}_\Theta$  a family of spectral densities which verifies assumptions B0-B3. Let  $(X_n)_{n \in \mathbb{Z}}$  be a stationary Gaussian long-memory sequence with zero mean and spectral density  $f(\theta_0, \lambda) \in \mathcal{F}_\Theta$  with  $\theta_0$  in the interior of  $\Theta$ . Let  $(Y_n)_{n \in \mathbb{Z}}$  be a realisation of an independent copy of  $(X_n)_{n \in \mathbb{Z}}$ .  $\hat{\theta}_T$  is the estimation of  $\theta_0$  computed on  $(Y_n)_{n \in \mathbb{Z}}$  according to (3.5).*

*Then the additional error due to the estimation of the forecast coefficients verifies that there exists  $C$  a constant independent of  $k$  such that*

$$\forall k \in \mathbb{N}^*, \lim_{T \rightarrow +\infty} T \mathbb{E} \left( \widetilde{X}_{T,k}(1) - \widetilde{X}_k(1) \right)^2 \leq C k^{2\alpha(\theta_0)}$$

where  $\widetilde{X}_{T,k}(1)$  is defined in (3.7).

In the proof of Theorem 3.2.1, we need the following lemma.

**Lemma 3.2.1.** *Under the assumptions of Theorem 3.2.1, the coefficients of the infinite autoregressive representation  $a_j$  defined in (3.2) verify :*

(i)  $a_j(\theta)$  is uniformly bounded i.e. there exists  $C$  such that for all  $j \in \mathbb{N}$  and all  $\theta \in \Theta$

$$|a_j(\theta)| \leq C;$$

(ii) for all  $\delta > 0$ , there exists  $C$  such that for all  $j \in \mathbb{N}^*$ , for all  $l \in \llbracket 1, m \rrbracket$  and for all  $\theta$  in  $\Theta$  :

$$\left| \frac{\partial a_j}{\partial \theta_l}(\theta) \right| \leq C j^{-1+\delta}; \quad (3.12)$$

(iii) the second and third derivatives of  $a_j$  with respect to the parameter  $\theta$  are uniformly bounded : there exists  $C$  such that for all  $j \in \mathbb{N}$ , for all  $k, l, n \in \llbracket 1, m \rrbracket^3$  and for all  $\theta \in \Theta$ ,

$$\left| \frac{\partial^2 a_j}{\partial \theta_k \partial \theta_l}(\theta) \right| \leq C \text{ and } \left| \frac{\partial^3 a_j}{\partial \theta_k \partial \theta_l \partial \theta_n}(\theta) \right| \leq C; \quad (3.13)$$

(iv) the first and second derivatives of  $a_j$  with respect to the parameter  $\theta$  are continuous.

*Proof.* The proof is given in Appendix 3.4.1. □

*Proof of Theorem 3.2.1.* We define the vectors :

$$\alpha_{k,T} := \left( a_1(\hat{\theta}_T) - a_1(\theta_0), \dots, a_k(\hat{\theta}_T) - a_k(\theta_0) \right)' \quad \text{and} \quad \mathbf{X}_1^k := (X_k, \dots, X_1)'$$

where  $v'$  is the transpose vector of  $v$ . Simple algebraic manipulations yield :

$$\begin{aligned} \mathbb{E} \left( \widetilde{X}_{T,k}(1) - \widetilde{X}_k(1) \right)^2 &= \mathbb{E} \left( - \sum_{j=1}^k a_j(\hat{\theta}_T) X_{k+1-j} + \sum_{j=1}^k a_j X_{k+1-j} \right)^2 \\ &= \mathbb{E} \left( \left( a_1(\hat{\theta}_T) - a_1, \dots, a_k(\hat{\theta}_T) - a_k \right) \begin{pmatrix} X_k \\ \vdots \\ X_1 \end{pmatrix} \right)^2 \\ &= \mathbb{E} \left[ \left( \alpha'_{k,T} \mathbf{X}_1^k \right)^2 \right] \\ &= \text{trace} \left( \mathbb{E} \left( \alpha'_{k,T} \mathbf{X}_1^k \right)^2 \right) \\ &= \mathbb{E} \left( \text{trace} \left( \alpha_{k,T} \alpha'_{k,T} \mathbf{X}_1^k \left( \mathbf{X}_1^k \right)' \right) \right) \\ &= \text{trace} \left( \mathbb{E} \left( \alpha_{k,T} \alpha'_{k,T} \mathbf{X}_1^k \left( \mathbf{X}_1^k \right)' \right) \right). \end{aligned}$$

Since the realisations  $(X_j)_{j \in \llbracket 1, k \rrbracket}$  and  $(Y_j)_{j \in \llbracket 1, T \rrbracket}$  are independent,

$$\begin{aligned} \mathbb{E} \left( \widetilde{X}_{T,k}(1) - \widetilde{X}_k(1) \right)^2 &= \text{trace} \left( \mathbb{E} \left( \alpha_{k,T} \alpha'_{k,T} \right) \mathbb{E} \left( \mathbf{X}_1^k \left( \mathbf{X}_1^k \right)' \right) \right) \\ &= \text{trace} \left( \mathbb{E} \left( \alpha_{k,T} \alpha'_{k,T} \right) \Sigma_k \right) \end{aligned}$$

where  $\Sigma_k$  is the common covariance matrix of the processes  $(X_n)_{n \in \mathbb{Z}}$  and  $(Y_n)_{n \in \mathbb{Z}}$

$$\Sigma_k := (\sigma(i-j))_{1 \leq i, j \leq k}. \quad (3.14)$$

Our proof is divided into the two following steps :

Step 1 we give an approximation to the sequence of the mean-squared error of each entry of  $\mathbb{E} \left( \alpha_{k,T} \alpha'_{k,T} \right)$  as a function of  $T$  ;

Step 2 we rewrite the matrix  $\mathbb{E} \left( \alpha_{k,T} \alpha'_{k,T} \right)$  using the approximation obtained in Step 1 then we study the limit as  $T$  tends to infinity as a function of  $k$ .

**Step 1** Let us study the entries of the covariance matrix of the estimated coefficients  $\mathbb{E}(\alpha_{k,T} \alpha'_{k,T})$ . We can write the entries  $\left( \mathbb{E}(\alpha_{k,T} \alpha'_{k,T}) \right)_{i,j}$  as  $\mathbb{E} \left( g_{i,j}(\hat{\theta}_T) \right)$  where

$$g_{i,j}(\theta) = (a_i(\theta) - a_i(\theta_0))(a_j(\theta) - a_j(\theta_0)). \quad (3.15)$$

We then use an order 2 Taylor series expansion of  $g_{i,j}$  and apply Theorem 5.4.3 from Fuller (1976). We will refer to the following version.

Under the following assumptions :

- (1)  $\forall l \in \llbracket 1, m \rrbracket, \quad \mathbb{E} \left( |\widehat{\theta}_{T,l} - \theta_{0,l}|^3 \right) = O(\eta(T))$  where  $\widehat{\theta}_{T,l}$  is the  $l^{\text{th}}$  entry of  $\widehat{\theta}_T$  and  $\eta(T)$  tends to 0 as  $T$  tends to infinity ;
- (2)  $\widehat{\theta}_T \rightarrow \theta_0, \quad \mathbb{P}\text{-a.s. as } T \rightarrow +\infty ;$
- (3)  $g_{i,j}$  is uniformly bounded on a neighbourhood of  $\theta_0 ;$
- (4) the first and the second derivatives  $g_{i,j}$  are continuous and bounded on a neighbourhood of  $\theta_0$

we have

$$\begin{aligned} \mathbb{E} \left( g_{i,j}(\widehat{\theta}_T) \right) &= g_{i,j}(\theta_0) + \sum_{l=1}^m \mathbb{E}(\widehat{\theta}_{T,l} - \theta_{0,l}) \frac{\partial g_{i,j}}{\partial \theta_l}(\theta_0) \\ &+ \frac{1}{2} \sum_{l=1}^m \sum_{n=1}^m \frac{\partial^2 g_{i,j}}{\partial \theta_l \partial \theta_n}(\theta_0) \mathbb{E} \left( (\widehat{\theta}_{T,l} - \theta_{0,l}) (\widehat{\theta}_{T,n} - \theta_{0,n}) \right) \\ &+ \frac{1}{6} \sum_{k=1}^m \sum_{l=1}^m \sum_{n=1}^m \mathbb{E} \left( \frac{\partial^3 g_{i,j}}{\partial \theta_l \partial \theta_n \partial \theta_k}(\theta^*) (\widehat{\theta}_{T,l} - \theta_{0,l}) (\widehat{\theta}_{T,n} - \theta_{0,n}) (\widehat{\theta}_{T,k} - \theta_{0,k}) \right). \end{aligned} \quad (3.16)$$

where  $\theta^* = t_T \theta_0 + (1 - t_T) \widehat{\theta}_T$  with  $t_T \in [0, 1]$ . First we verify that the assumptions (2)-(4) hold. Under assumptions on the spectral density and its derivatives, Theorem 1 in Fox et Taqqu (1986) is valid and ensures the convergence (2). By the propositions (i) and (ii) of Lemma 3.2.1, conditions (3) and (4) hold.

We now state a lemma which permits us to apply the Fuller Theorem and to prove (1).

**Lemma 3.2.2.** *Under the assumptions on Theorem 3.2.1, we have :*

a) for all  $j, l \in \llbracket 1, m \rrbracket$

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

where the matrix  $W$  is defined by :

$$W_{i,j} := \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left( \frac{\partial}{\partial \theta_i} f^{-1}(\lambda, \theta_0) \right) \left( \frac{\partial}{\partial \theta_j} f^{-1}(\lambda, \theta_0) \right) f^2(\lambda, \theta_0) d\lambda; \quad (3.17)$$

b) for all  $j \in \llbracket 1, m \rrbracket$  ,

$$\mathbb{E} \left( \left| \widehat{\theta}_{T,j} - \theta_{0,j} \right|^3 \right) = O \left( T^{-3/2} \right) \text{ as } T \rightarrow +\infty.$$

*Proof.* The proof is given in Appendix 3.4.2. □

By Lemma 3.2.2 b), we obtain that the condition (1) holds by taking  $\eta(T) = T^{-3/2}$ . According to the Fuller theorem, the Taylor expansion (3.16) holds. The expression (3.16) can be simplified. Indeed we remark by the definition of the function  $g_{i,j}$  in (3.15) that :

$$g_{i,j}(\theta_0) = 0 \quad \text{and} \quad \forall l \in \llbracket 1, p \rrbracket, \quad \frac{\partial g_{i,j}}{\partial \theta_l}(\theta_0) = 0.$$

And by Lemma 3.2.2 a), we have :

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

Moreover by (i), (iii) and (iv) of Lemma 3.2.1 and b) of Lemma 3.2.2

$$\begin{aligned} & \sum_{k=1}^m \sum_{l=1}^m \sum_{n=1}^m \mathbb{E} \left( \frac{\partial^3 g_{i,j}}{\partial \theta_l \partial \theta_n \partial \theta_k}(\theta^*) \left( \widehat{\theta}_{T,l} - \theta_{0,l} \right) \left( \widehat{\theta}_{T,n} - \theta_{0,n} \right) \left( \widehat{\theta}_{T,k} - \theta_{0,k} \right) \right) \\ & \leq 8C^2 m^3 \mathcal{O}(\eta(T)). \end{aligned} \quad (3.18)$$

Finally we obtain the following approximation : for any  $\delta > 0$ ,

$$\begin{aligned} & \lim_{T \rightarrow +\infty} T \mathbb{E} \left( g_{i,j}(\widehat{\theta}_T) \right) \quad (3.19) \\ & = \lim_{T \rightarrow +\infty} T \mathbb{E} \left( (a_i(\widehat{\theta}_T) - a_i(\theta_0))(a_j(\widehat{\theta}_T) - a_j(\theta_0)) \right) \\ & = \lim_{T \rightarrow +\infty} T \sum_{l=1}^m \sum_{n=1}^m \frac{\partial^2 g_{i,j}}{\partial \theta_l \partial \theta_n}(\theta_0) \mathbb{E} \left( \left( \widehat{\theta}_{T,l} - \theta_{0,l} \right) \left( \widehat{\theta}_{T,n} - \theta_{0,n} \right) \right) + \lim_{T \rightarrow +\infty} T \mathcal{O} \left( T^{-2+\delta} \right) \\ & = 4\pi \sum_{l=1}^m \sum_{n=1}^m \frac{\partial^2 g_{i,j}}{\partial \theta_l \partial \theta_n}(\theta_0) (W^{-1})_{l,n} \quad (3.20) \end{aligned}$$

**Step 2** We now replace the approximation of each term of  $\mathbb{E} \left( \alpha_{k,T} \alpha'_{k,T} \right)$ .

We first study the limit given in (3.20). Since  $W^{-1}$  is symmetric :

$$\begin{aligned} & \left( \sum_{l=1}^m \sum_{n=1}^m \left[ \frac{\partial a_i(\theta_0)}{\partial \theta_l} \frac{\partial a_j(\theta_0)}{\partial \theta_n} + \frac{\partial a_j(\theta_0)}{\partial \theta_l} \frac{\partial a_i(\theta_0)}{\partial \theta_n} \right] (W^{-1})_{l,n} \right)_{i,j} \\ & = \left( \sum_{l=1}^m \sum_{n=1}^m \left[ \frac{\partial a_i(\theta_0)}{\partial \theta_l} \frac{\partial a_j(\theta_0)}{\partial \theta_n} (W^{-1})_{l,n} + \frac{\partial a_j(\theta_0)}{\partial \theta_l} \frac{\partial a_i(\theta_0)}{\partial \theta_n} (W^{-1})_{n,l} \right] \right)_{i,j} \\ & = 2 (DW^{-1}D')_{i,j} \end{aligned}$$

where

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



$W$  defined in (3.17) is positive definite thus  $W^{-1}$  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 eigenvalues of  $W^{-1}$ , which are positive. We may rewrite our expression as :

$$\begin{aligned} DW^{-1}D' &= \sum_{r=1}^m \left( \left[ \sqrt{\lambda_r} \sum_{l=1}^m p_{rl} \frac{\partial a_i}{\partial \theta_l}(\theta_0) \right] \left[ \sqrt{\lambda_r} \sum_{l=1}^m p_{rl} \frac{\partial a_j}{\partial \theta_l}(\theta_0) \right] \right)_{1 \leq i, j \leq k} \\ &= \sum_{r=1}^m \beta'_r \beta_r, \end{aligned} \quad (3.21)$$

where  $\beta_r$  is the row vector  $\left( \sqrt{\lambda_r} \sum_{l=1}^m p_{rl} \frac{\partial a_i}{\partial \theta_l}(\theta_0) \right)_{1 \leq i \leq k}$ . Then using the approximation obtained in (3.20) and we find the following expression of  $\mathbb{E}(\alpha_{k,T} \alpha'_{k,T})$ , for all  $\delta > 0$  and for all  $k, T > 0$  :

$$\lim_{T \rightarrow +\infty} T \mathbb{E}(\alpha_{k,T} \alpha'_{k,T}) = 8\pi \sum_{r=1}^m \beta'_r \beta_r. \quad (3.22)$$

Now we will study the asymptotic behaviour as  $T$  tends to infinity. We have for all  $\delta > 0$  :

$$\lim_{T \rightarrow +\infty} T \text{trace}(\mathbb{E}(\alpha_{k,T} \alpha'_{k,T}) \Sigma_k) = 8\pi \text{trace} \left( \sum_{r=1}^m \beta'_r \beta_r \Sigma_k \right). \quad (3.23)$$

We study the limit defined in (3.23) and search a lower and an upper bound of :

$$\text{trace} \left( \sum_{r=1}^m \beta'_r \beta_r \Sigma_k \right) = \sum_{r=1}^m \beta_r \Sigma_k \beta'_r. \quad (3.24)$$

We have :

$$\Lambda_k^- \sum_{r=1}^m \|\beta_r\|_2^2 \leq \sum_{r=1}^m \beta_r \Sigma_k \beta'_r \leq \Lambda_k^+ \sum_{r=1}^m \|\beta_r\|_2^2$$

where  $\Lambda_k^-$  is the smallest eigenvalue of  $\Sigma_k$  and  $\Lambda_k^+$  the largest eigenvalue.

We now estimate the upper bound. Using (3.12) proved in Lemma 3.2.1, we can estimate

$\|\beta_r\|_2^2$ . Let  $\delta < 1/2$ , there exists  $C_1, \dots, C_m$  such that :

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

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

From Boettcher et Virtanen (2006), the spectral norm of a covariance matrix associated to a spectral density of the form  $\lambda \mapsto \lambda^{-\alpha} L(\lambda)$  where  $L$  is bounded, continuous at 0 and does not vanish at 0, is equivalent to  $Ck^\alpha$  with  $C$  constant. Therefore  $\Lambda_k^+ \leq Ck^\alpha(\theta_0)$  with  $C$  constant. We then obtain that there exists a constant  $C$  such that :

$$\text{trace} \left( \sum_{r=1}^m \beta_r' \beta_r \Sigma_k \right) \leq Ck^\alpha(\theta_0).$$

We now search a lower bound for (3.24). Since the spectral density  $f(\cdot, \theta_0)$  is bounded away from 0 (from assumption B0), the lower eigenvalue  $\Lambda_k^-$  has a uniform lower bound (see for example Grenander et Szegö (1958)). Then there exists a positive  $C'$  such that for all  $k$ ,

$$\Lambda_k^- \sum_{r=1}^m \|\beta_r\|_2^2 \geq C' \sum_{r=1}^m \|\beta_r\|_2^2.$$

Moreover for all  $r \in \llbracket 1, m \rrbracket$ ,

$$\begin{aligned} \|\beta_r\|_2^2 &= \lambda_r \sum_{i=1}^k \left( \sum_{l=1}^m p_{r,l} \frac{\partial a_i}{\partial \theta_l}(\theta_0) \right)^2 \\ &\geq \lambda_r \left( \sum_{l=1}^m p_{r,l} \frac{\partial a_1}{\partial \theta_l}(\theta_0) \right)^2. \end{aligned}$$

Finally we obtain that there exists a positive constant  $C'$ , which does not depend on  $k$ , such that :

$$\text{trace} \left( \sum_{r=1}^m \beta_r' \beta_r \Sigma_k \right) \geq C'.$$

We obtain that the following inequality for the first term of (3.23) :

$$\frac{C'}{T} \leq \frac{\text{trace}(\sum_{r=1}^m \beta_r' \beta_r \Sigma_k)}{T} \leq C \frac{k^\alpha(\theta_0)}{T} \sum_{r=1}^m C_r(\theta_0). \quad (3.25)$$

Finally we conclude using (3.23) and the second inequality in (3.25) :

$$\lim_{T \rightarrow +\infty} \mathbb{E} \left( \left( a_1(\hat{\theta}_T) - a_1(\theta_0), \dots, a_k(\hat{\theta}_T) - a_k(\theta_0) \right) \begin{pmatrix} X_0 \\ \vdots \\ X_{-k+1} \end{pmatrix} \right)^2 \leq Ck^{\alpha(\theta_0)}.$$

This concludes the proof.  $\square$

### 3.2.3 $h$ -step Prediction

More generally, for  $h$ -step prediction the forecast coefficients are given by :

$$c_j(h) = - \sum_{m=0}^{h-1} a_{j+m} b_{h-1-m}. \quad (3.26)$$

Using a parametric approach, we may consider  $c_j(h)$  as a function of the parameter  $\theta$  for each  $h$  and we will estimate  $\theta$  by the Whittle estimator. Under similar assumptions, we can prove the same result for  $h$ -step prediction as for next step prediction.

**Theorem 3.2.2.** *We use the notations in Theorem 3.2.1 and its assumptions. Moreover we suppose that the coefficients of the moving average representation are uniformly bounded i.e. there exists  $C$  such that for all  $(\theta, j) \in \Theta \times \mathbb{N}$  we have  $|b_j(\theta)| \leq C$ . Then the additional error due to the estimation of the forecast coefficients verifies that there exists  $C$  a constant independent of  $k$  such that :*

$$\forall k \in \mathbb{N}^*, \quad \lim_{T \rightarrow +\infty} T \mathbb{E} \left( \widetilde{X}_{T,k}(h) - \widetilde{X}_k(h) \right)^2 \leq Ck^{2d}$$

where

$$\widetilde{X}_{T,k}(h) = - \sum_{j=1}^k c_j(h, \hat{\theta}_T) X_{k+h-j}.$$

The proof of this theorem is based on the fact that the coefficients  $c_j(h)$  also verify Lemma 3.2.1 under the assumptions of Theorem 3.2.1 and if the coefficients  $b_j$  are uniformly bounded.

These conditions of regularity for the spectral density and the coefficients  $c_j(h)$  are for example verified by the class of FARIMA processes defined in (3.11).

With Theorem 3.2.1, we know the asymptotic behaviour of the mean-squared error due to the estimation of the forecast coefficients i.e.  $O\left(\frac{k^{2d}}{T}\right)$  as  $T$ , the number of observations for estimation, tends to infinity. But we have proved in Godet (2008), that for FARIMA processes

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

Then for FARIMA processes, we can prove the convergence in distribution of the difference between our predictor  $\widetilde{X}_{T,k}(1)$  and the Wiener-Kolmogorov predictor  $X_k^{\text{WK}}(1) = -\sum_{j=1}^{+\infty} a_j X_{n+1-j}$ , which is the linear least-squares predictor based on all the past.

**Theorem 3.2.3.** *Assume that the process  $(X_n)_{n \in \mathbb{Z}}$  is a FARIMA process of order  $p$  and  $q$ .*

Let  $\mathcal{F}_\Theta = \{f(\cdot, \sigma_\varepsilon^2, \phi, \psi), (\sigma_\varepsilon^2, \phi, \psi, d) \in \Theta_0\}$  be a family of spectral densities which verifies

$$f(\lambda, \sigma_\varepsilon^2, \phi, \psi, d) = \frac{\sigma_\varepsilon^2}{2\pi} |1 - \exp(i\lambda)|^{-2d} \left| \frac{\psi(\exp(i\lambda))}{\phi(\exp(i\lambda))} \right|^2$$

with  $\phi$  and  $\psi$  polynomials of order respectively  $p$  and  $q$  with no zeroes on the closed unit circle and with  $\phi(0) = \psi(0) = 1$ . We denote the polynomials  $\psi$  and  $\phi$  by  $\phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \dots + \phi_p z^p$  and  $\psi(z) = 1 + \psi_1 z + \psi_2 z^2 + \dots + \psi_q z^q$ . In order to simplify and to have the same notations as in Theorems 3.2.1 and 3.2.2, we denote by  $\theta$  the vector  $(\sigma_\varepsilon^2, \phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q)$  and by  $\theta_0$  the true value of the parameter assumed to be in the interior of the compact  $\Theta$ .

Let  $(Y_n)_{n \in \mathbb{Z}}$  be an independent copy of  $(X_n)_{n \in \mathbb{Z}}$  and  $\widehat{\theta}_T$  the estimate of  $\theta_0$  defined in (3.5) and computed on  $(Y_n)_{n \in \mathbb{Z}}$ . Then there exists  $\sigma > 0$  such that

$$\lim_{k \rightarrow +\infty} \sqrt{k} \lim_{T \rightarrow +\infty} \left( \widetilde{X}_{T,k}(1) - X_k^{\text{WK}}(1) \right) = \mathcal{N}(0, \sigma^2) \quad (3.27)$$

where  $\lim$  in 3.27 denotes the limit in the sense of convergence in distribution.

*Proof.* From the definitions of  $\widetilde{X}_{T,k}(1)$  and  $X_k^{\text{WK}}(1)$ , their difference is

$$\widetilde{X}_{T,k}(1) - X_k^{\text{WK}}(1) = -\sum_{j=1}^k (a_j(\widehat{\theta}_T) - a_j(\theta_0)) X_{k+1-j} + \sum_{j=k+1}^{+\infty} a_j(\theta_0) X_{k+1-j}. \quad (3.28)$$

First we remark that by Theorem 3.2.1, we have :

$$\mathbb{E} \left[ \left( \sum_{j=1}^k (a_j(\widehat{\theta}_T) - a_j(\theta_0)) X_{k+1-j} \right)^2 \right] \rightarrow 0 \quad (3.29)$$

as  $T$  tends to infinity. We obtain with (3.28) and (3.29) :

$$\forall k \in \mathbb{N}^*, \left( \widetilde{X}_{T,k}(1) - X_k^{\text{WK}}(1) \right) = o_{\mathbb{P}}(1) + \sum_{j=k+1}^{+\infty} a_j(\theta_0) X_{k+1-j}$$

as  $T$  tends to infinity. Then

$$\forall k \in \mathbb{N}^*, \lim_{T \rightarrow +\infty} \left( \widetilde{X}_{T,k}(1) - X_k^{\text{WK}}(1) \right) = \sum_{j=k+1}^{+\infty} a_j(\theta_0) X_{k+1-j}.$$

In Godet (2008), we have proved that if  $(X_n)_{n \in \mathbb{Z}}$  is a FARIMA processes, there exists  $\sigma > 0$  such that

$$\mathbb{E} \left[ \left( \sqrt{k} \sum_{j=k+1}^{+\infty} a_j(\theta_0) X_{k+1-j} \right)^2 \right] \rightarrow \sigma^2.$$

as  $k$  tends to infinity. Since the process  $(\sqrt{k} \sum_{j=k+1}^{+\infty} a_j(\theta_0) X_{k+1-j})_{k \in \mathbb{N}}$  is a centered Gaussian process, we have also obtained the convergence in distribution and then the result.  $\square$

### 3.3 Simulations and Conclusion

Theorems 3.2.1 and 3.2.2 only give bounds for the mean-squared errors due to the estimation of the forecast coefficients. In this section, simulations are used to evaluate the quality of the bounds given in these theorems. We use fractionally integrated noise defined in (3.10) with the parameter of longmemory  $d$  equal to 0.2. The simulation is based on the moving average representation of the fractionally integrated process (see Bardet *et al.* (2003)). We compute 600 independent realisations of  $\widetilde{X}_{T,k}(1) - \widetilde{X}_k(1)$  with  $k$  varying in  $[100, 170]$  and  $k$  is equal to  $T^{1/2.5}$ . We plot on Figure 3.1 the empirical variance of this sample versus  $k$  in a logarithm scale on the both axes. In fact we have proved in Theorem 3.2.1 that the variance of  $\widetilde{X}_{T,k}(1) - \widetilde{X}_k(1)$  is bounded by  $C \frac{k^{2d}}{T}$  as  $T$  tends to infinity. Then the logarithm of this bound is equal to  $\log(C) + (2d - 2.5) \log(k)$  when  $k$  is equal to  $T^{1/2.5}$ . On Figure 3.1 we plot a line with a slope of 2.1 (equal in your case to  $2d - 2.5$ ) and we verify that this line is compatible with our simulations.

To conclude we compare our predictor  $\widetilde{X}_{T,k}(1)$  with others. On the first hand, to observe the negligibility of the prediction error due to the estimation of the forecast coefficients, we plot for fractionally integrated noise ( $d = 0.4$ ) the empirical distribution of the error made by  $\widetilde{X}_{T,k}(1)$  for  $k = 40$  and  $T = 1000$ . We compare it to the distribution of the innovation which is the error made by the Wiener-Kolmogorov predictor i.e. the innovation process and to the distribution of the process which is the error done when we predict by the mean of the process here equal to 0. We see on Figure 3.2 that the distribution of the error of our predictor is close to the Wiener-Kolmogorov predictor and then our predictor is efficient.

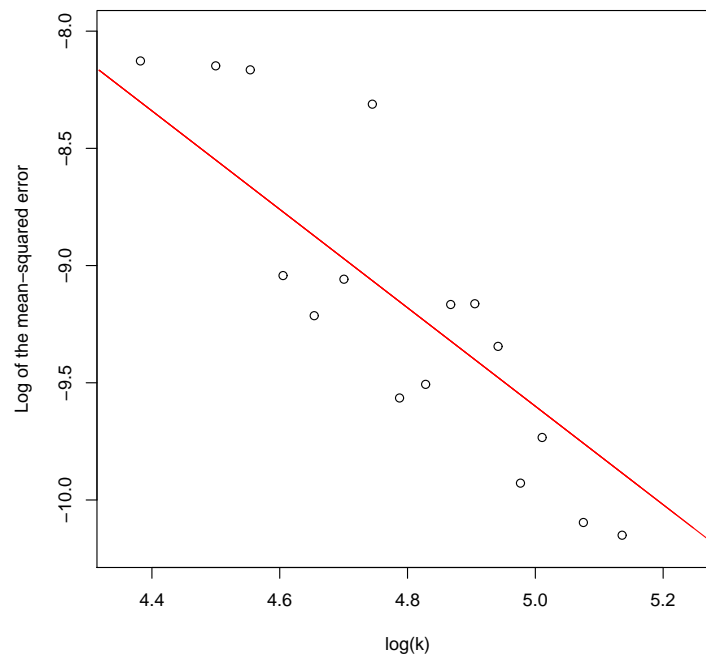


FIG. 3.1 – Empirical variance of  $\widetilde{X}_{T,k}(1) - \widetilde{X}_k(1)$  for a fractionally integrated noise ( $d = 0.2$ ) as a function of  $k$  in log-log scale.

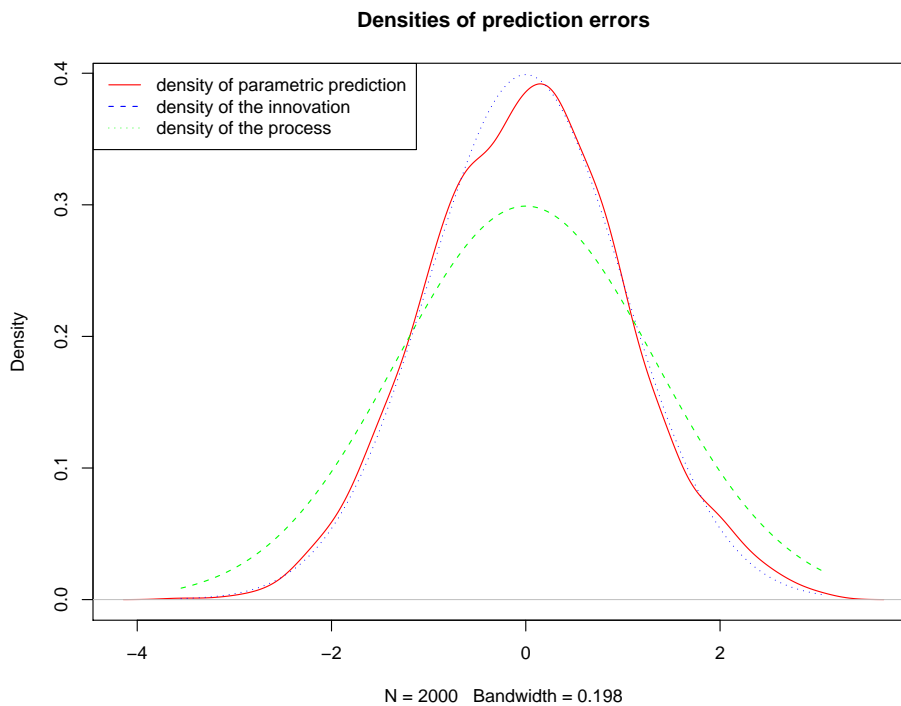


FIG. 3.2 – Distributions of prediction errors : empirical error made by  $\widetilde{X}_{T,k}(1)$ , theoretical error of the Wiener-Kolmogorov predictor and theoretical error made by the naive predictor which is the mean of the process

On the other hand, we compare our predictor with other solutions which have been proposed for long-memory processes. The first one is the predictor of Hidalgo et Yajima (2002). The theoretical predictor is also the truncated Wiener-Kolmogorov predictor but the estimation of the forecast coefficients is based on the canonical factorisation of the spectral density. The forecast coefficients of the one step predictor are given by

$$a_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp \left\{ - \sum_{u=1}^{+\infty} d_u \exp(-iu\lambda) \right\} \exp(ij\lambda) d\lambda \quad (3.30)$$

where  $(d_u)_{u \in \mathbb{N}^*}$  are the Fourier coefficients of the logarithm of the spectral density of the process. And to estimate the forecast coefficients, Hidalgo et Yajima (2002) plug a semi-parametric estimation of the spectral density in equation (3.30).

We again consider two independent realisations of a fractionally integrated process with  $d = 0.4$ . We use a past of length  $k = 40$  to predict and  $T = 10000$  to estimate the forecast coefficients for the both predictors. We can observe on Figure 3.3 that the two prediction errors have a Gaussian distribution but it is difficult to see which is the more efficient. We then computed on 1800 realisations the empirical variance of these two distributions. For our predictor we found an empirical variance equal to 1.009 and for the predictor of Hidalgo and Yajima a variance equal to 1.240. The truncated Wiener-Kolmogorov is then more efficient than the Hidalgo-Yajima predictor. This comes from the fact that our predictor is parametric and theirs is only semi-parametric.

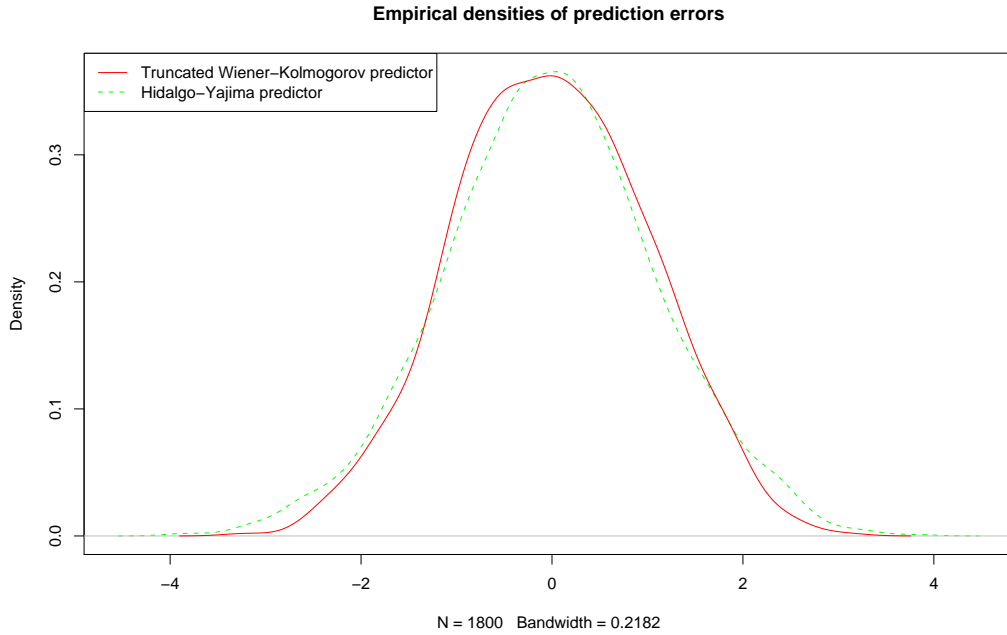


FIG. 3.3 – Empirical distributions of prediction error for  $\widetilde{X}_{T,k}(1)$  defined in (3.7) and for the Hidalgo-Yajima predictor defined by (3.30)

We compare the predictor with the least-squares predictor studied in Godet (2007b). Instead of truncating the Wiener-Kolmogorov, the predicted random variable is projected onto the available finite past. The forecast coefficients are then the solution of the Yule-Walker equations (see Brockwell et Davis (1991)) and then function of the covariances of the processes. As in Ing et Wei (2003) and Godet (2007b), we plug in the Yule-Walker equations the empirical autocovariances, estimated on the same realisation i.e. we only have a time series to predict and to estimate the forecast coefficients. The convergence of mean-squared error of the least-squares predictor to that of the Wiener-Kolmogorov predictor was obtained in Godet (2007b) for long-memory processes. We compare these two predictors using 1800 independent realisations of fractionally integrated processes with  $d = 0.4$ . With a past of length  $k = 40$  to predict and  $T = 10000$  to estimate the forecast coefficients, we can not observe a difference between the least-squares predictor and our predictor. But if we reduce  $T$  to 1000, we observe on Figure 3.4 that the variance of our predictor (1.083) is lower than the variance of the least-squares predictor equal to 1.539. In this case, the least-squares predictor does not make better than a prevision by 0, since the variance of its error is equivalent to the variance of the process. This bad performance comes from the fact that the rate of convergence of the forecast coefficients depends on the convergence of the empirical covariance which is slow when  $d = 0.4$  (see Hosking (1996) and Godet (2007a)).

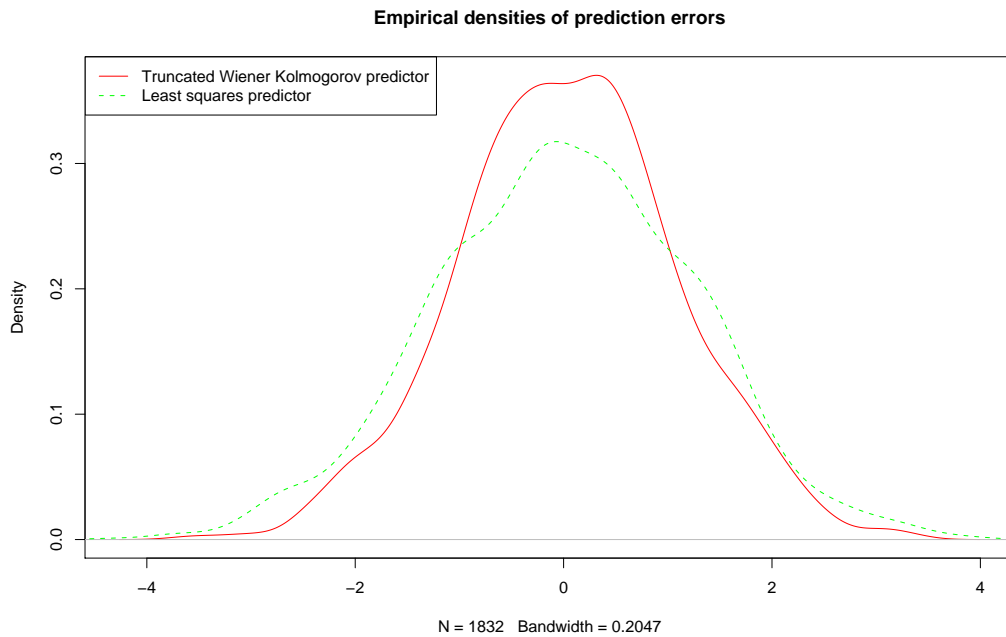


FIG. 3.4 – Empirical distributions of prediction errors for  $\widetilde{X}_{T,k}(1)$  defined in (3.7) and for the least-squares predictor



## 3.4 Appendix

### 3.4.1 Proof of Lemma 3.2.1

We first remark that the boundedness of  $a_j$  required in (i) comes directly from (3.9). Under the assumption (3.8), the family of processes with spectral density in  $\mathcal{F}_\Theta$  is a family of non-deterministic processes. As Pourahmadi (2001) obtain in Theorem VII a relation between the moving average coefficients  $(b_j(\theta))_{j \in \mathbb{N}}$  and  $\int_{-\pi}^{\pi} \exp(ij\lambda) \log(f(\lambda, \theta)) d\lambda / 2\pi)_{j \in \mathbb{N}}$ , if we define  $(d_j(\theta)) := \int_{-\pi}^{\pi} \exp(ij\lambda) \log(f(\lambda, \theta)^{-1}) d\lambda / 2\pi)_{j \in \mathbb{N}^*}$  and  $d_0(\theta) := -1/2 \int_{-\pi}^{\pi} \log(f(\lambda, \theta)^{-1}) d\lambda / 2\pi$  we obtain that the autoregressive coefficients verify for  $|z| < 1$  :

$$\sum_{j=0}^{+\infty} a_j(\theta) z^j = \exp \left( \sum_{j=0}^{+\infty} d_j(\theta) z^j \right). \quad (3.31)$$

*Proof of (ii)* Differentiating both sides of the equation (3.31) with respect to  $\theta_l$  we obtain :

$$\frac{\partial}{\partial \theta_l} \sum_{j=0}^{+\infty} a_j(\theta) z^j = \frac{\partial}{\partial \theta_l} \left[ \sum_{j=0}^{+\infty} d_j(\theta) z^j \right] \sum_{j=0}^{+\infty} a_j(\theta) z^j. \quad (3.32)$$

Now, we look for an asymptotic expression of  $\frac{\partial}{\partial \theta_l} d_j(\theta)$  as  $j$  goes to infinity.

$$\begin{aligned} \frac{\partial}{\partial \theta_l} d_j(\theta) &= \frac{1}{2\pi} \frac{\partial}{\partial \theta_l} \int_{-\pi}^{\pi} \exp(ij\lambda) \log(f(\lambda, \theta)^{-1}) d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(ij\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} d\lambda \end{aligned}$$

We can differentiate under the integral sign since  $\frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l}$  is absolutely integrable by assumptions B1-B2 since for all  $(\lambda, \theta)$  and for all  $\delta > 0$ ,

$$\left| \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} \right| \leq \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta}.$$

Moreover we obtain by Theorem 2.22 of Zygmund (1968) Chapter V that for all  $\delta > 0$  as  $j$  goes to infinity

$$\int_{-\pi}^{\pi} \cos(j\lambda) |\lambda|^{-2\delta} d\lambda \sim C_1(\delta) j^{2\delta-1}. \quad (3.33)$$

Then since  $\frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l}$  is even

$$\begin{aligned} \int_{-\pi}^{\pi} \exp(ij\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} d\lambda &= \int_{-\pi}^{\pi} \cos(j\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} d\lambda \\ &= 2 \int_0^{\frac{\pi}{2j}} \cos(j\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} \frac{1}{|\lambda|^{-2\delta}} |\lambda|^{-2\delta} d\lambda \\ &\quad + 2 \int_{\frac{\pi}{2j}}^{\pi} \cos(j\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} d\lambda. \end{aligned} \quad (3.34)$$

By assumptions B1 and B2, we have

$$\left| \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} \right| \leq \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \quad (3.35)$$

and then if  $\lambda \in \left[0, \frac{\pi}{2j}\right]$ ,

$$-\frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \cos(j\lambda) \leq \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} \cos(j\lambda) \leq \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \cos(j\lambda). \quad (3.36)$$

Using (3.34) and (3.36), we can obtain a upper bound of :

$$\begin{aligned} \int_{-\pi}^{\pi} \exp(ij\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} d\lambda &\leq \frac{2C(\delta)}{C_0(\delta)} \int_0^{\frac{\pi}{2j}} \cos(j\lambda) |\lambda|^{-2\delta} d\lambda \\ &\quad + 2 \int_{\frac{\pi}{2j}}^{\pi} \cos(j\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} d\lambda \\ &\leq \frac{2C(\delta)}{C_0(\delta)} \int_0^{\pi} \cos(j\lambda) |\lambda|^{-2\delta} d\lambda \\ &\quad + 2 \int_{\frac{\pi}{2j}}^{\pi} \cos(j\lambda) \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} - \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) d\lambda. \end{aligned}$$

Using (3.33), we have

$$\frac{2C(\delta)}{C_0(\delta)} \int_0^{\pi} \cos(j\lambda) |\lambda|^{-2\delta} d\lambda \sim C_2(\delta) j^{-2\delta-1}.$$

Moreover integrating by parts the second term, we obtain

$$\begin{aligned} &\int_{\frac{\pi}{2j}}^{\pi} \cos(j\lambda) \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} - \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) d\lambda \\ &= \left[ \frac{\sin(j\lambda)}{j} \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} - \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) \right]_{\frac{\pi}{2j}}^{\pi} \\ &\quad - \int_{\frac{\pi}{2j}}^{\pi} \frac{\sin(j\lambda)}{j} \frac{\partial}{\partial \lambda} \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} - \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) d\lambda. \end{aligned}$$

Using (3.35) the integrated part is bounded by

$$\begin{aligned} \left| \left[ \frac{\sin(j\lambda)}{j} \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} - \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) \right]_{\frac{\pi}{2j}}^{\pi} \right| &\leq \frac{1}{j} \left| \frac{\partial \log(f^{-1})}{\partial \theta_l}(\pi, \theta) - \frac{C(\delta)}{C_0(\delta)} |\pi|^{-2\delta} \right| \\ &+ \frac{2}{j} \frac{C(\delta)}{C_0(\delta)} \left| \frac{\pi}{2j} \right|^{-2\delta} \\ &\leq C_3(\delta) j^{2\delta-1}. \end{aligned}$$

Using assumptions B1, B2 and B3 we obtain that the integral is bounded by

$$\begin{aligned} &\left| \int_{\frac{\pi}{2j}}^{\pi} \frac{\sin(j\lambda)}{j} \frac{\partial}{\partial \lambda} \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} - \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) d\lambda \right| \\ &\leq \frac{1}{j} \left( \frac{C(\delta)}{C_0(\delta)} \int_{\frac{\pi}{2j}}^{\pi} \lambda^{-2\delta-1} d\lambda + \frac{C(\delta)^2}{C_0(\delta)^2} \int_{\frac{\pi}{2j}}^{\pi} \lambda^{-4\delta-1} d\lambda + \frac{2\delta C(\delta)}{C_0(\delta)} \int_{\frac{\pi}{2j}}^{\pi} \lambda^{-2\delta-1} d\lambda \right) \\ &\leq C_3(\delta) j^{4\delta-1}. \end{aligned}$$

Finally we obtain

$$\int_{\frac{\pi}{2j}}^{\pi} \cos(j\lambda) \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} - \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) d\lambda \leq C_3(\delta) \left( j^{2\delta-1} + j^{4\delta-1} \right).$$

To conclude of the asymptotic expression of  $\frac{\partial}{\partial \theta_l} d_j(\theta)$ , we search a lower bound using also (3.34) and (3.36),

$$\begin{aligned} \int_{-\pi}^{\pi} \exp(ij\lambda) \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} d\lambda &\geq \frac{-2C(\delta)}{C_0(\delta)} \int_0^{\pi} \cos(j\lambda) |\lambda|^{-2\delta} d\lambda \\ &+ 2 \int_{\frac{\pi}{2j}}^{\pi} \cos(j\lambda) \left( \frac{\partial \log(f(\lambda, \theta)^{-1})}{\partial \theta_l} + \frac{C(\delta)}{C_0(\delta)} |\lambda|^{-2\delta} \right) d\lambda. \end{aligned}$$

By the same procedure, we obtain that this lower bound is also absolutely bounded by  $j^{4\delta-1}$  up to a multiplicative constant which only depends on  $\delta$ . We then can conclude that

$$\frac{\partial}{\partial \theta_l} d_j(\theta) \leq C_5(\delta) j^{\delta-1}.$$

Then for all  $z$  such that  $|z| < 1$ , we have

$$\frac{\partial}{\partial \theta_l} \left[ \sum_{j=0}^{+\infty} d_j(\theta) z^j \right] = \sum_{j=0}^{+\infty} \frac{\partial d_j(\theta)}{\partial \theta_l} z^j.$$

The coefficients of the powers of  $z$  of the right side of (3.32) are given by the discrete convolution of the sequences  $a_j(\theta)$  and  $\partial d_j(\theta)/\partial \theta_l$ . Since  $a_j \leq C(\delta)(j^{-d-1+\delta})$  and

$\partial d_j(\theta)/\partial\theta_l \leq C_5(\delta)(j^{\delta-1})$  the Cauchy product of order  $j$  is bounded by

$$\begin{aligned} \sum_{m=0}^j \frac{\partial}{\partial\theta_l} d_m(\theta) a_{j-m}(\theta) &\leq C(\delta)C_5(\delta) \sum_{m=1}^{j-1} m^{\delta-1}(j-m)^{-d-1+\delta} + d_0 C(\delta)j^{-d-1+\delta} + C_5(\delta)j^{\delta-1} \\ &\leq C_6(\delta)j^{\delta-1}. \end{aligned}$$

We now study the left term of (3.31). We want to prove the absolute summability of  $\frac{\partial a_j(\theta)}{\partial\theta_l} z^j$ . to obtain that for  $|z| < 1$  :

$$\frac{\partial}{\partial\theta_l} \left[ \sum_{j=0}^{+\infty} a_j(\theta) z^j \right] = \sum_{j=0}^{+\infty} \frac{\partial a_j(\theta)}{\partial\theta_l} z^j.$$

The autoregressive coefficients  $a_j(\theta)$  can be computed by the following recursive formula (see Pourahmadi (1989)) :

$$a_{j+1}(\theta) = \sum_{m=0}^j \left( \frac{j+1-m}{j+1} \right) d_{j+1-m}(\theta) a_j(\theta). \quad (3.37)$$

Differentiating with respect to  $\theta_l$ , we obtain :

$$\begin{aligned} \frac{\partial a_{j+1}(\theta)}{\partial\theta_l} &= \frac{\partial}{\partial\theta_l} \left[ \sum_{m=0}^j \left( \frac{j+1-m}{j+1} \right) d_{j+1-m}(\theta) a_j(\theta) \right] \\ &= \sum_{m=0}^j \left( \frac{j+1-m}{j+1} \right) \left[ \frac{\partial d_{j+1-m}(\theta)}{\partial\theta_l} a_j(\theta) + d_{j+1-m}(\theta) \frac{\partial a_j(\theta)}{\partial\theta_l} \right]. \end{aligned}$$

But this recursive formula is also the formula which gives the coefficients of the powers of  $z$  in the power series expansion of  $\sum_{j=0}^{+\infty} \frac{\partial d_j(\theta)}{\partial\theta_l} z^j \sum_{j=0}^{+\infty} a_j(\theta) z^j$ . Then

$$\frac{\partial a_j(\theta)}{\partial\theta_l} = \sum_{m=0}^j \frac{\partial}{\partial\theta_l} d_m(\theta) a_{j-m}(\theta) \quad (3.38)$$

which is absolutely summable if  $|z| < 1$ .

Matching the powers of  $z$  in (3.32), one obtains that for all  $\delta > 0$  as  $j$  goes to infinity

$$\frac{\partial a_j(\theta)}{\partial\theta_l} = O(j^{\delta-1}).$$

*Proof of (ii)* By a similar procedure using the assumptions B1-B3, we obtain (3.13).

*Proof of (iv)* To show the continuity of the first derivatives of  $a_j(\theta)$ , it is sufficient using (3.38) to prove the continuity of the functions  $a_j(\theta)$  and the first derivatives of  $d_j(\theta)$ .

Moreover the autoregressive parameters  $a_j(\theta)$  can be computed by the recursive formula (3.37). The second derivatives of  $a_j(\theta)$  can be expressed in the same way as a function of  $d_j(\theta)$  and their first and second derivatives. It is then sufficient to prove the continuity of the functions  $d_j(\theta)$  and their first and second derivatives to prove (ii).

We only prove the continuity of  $d_j(\theta)$  since the results on the other derivatives can be obtained by the same procedure. Let  $\varepsilon \in ]0, 1[$ . By assumption B1, the function  $\log(f(\cdot, \cdot))$  is continuous at all  $(\lambda, \theta)$  with  $\lambda \neq 0$  then it is uniformly continuous on the compact  $[\varepsilon, \pi] \times \Theta$ . Then there exists  $\eta > 0$  such that for all  $\lambda \in [\varepsilon, \pi]$  if  $\|\theta_1 - \theta_2\| \leq \eta$  then  $|\log(f(\lambda, \theta_1)) - \log(f(\lambda, \theta_2))| \leq \varepsilon$ . For all  $(\theta_1, \theta_2) \in \Theta^2$  such that  $\|\theta_1 - \theta_2\| \leq \eta$ ,

$$\begin{aligned} |d_j(\theta_1) - d_j(\theta_2)| &= \left| 2 \int_0^\pi (\log(f(\lambda, \theta_1)) - \log(f(\lambda, \theta_2))) \cos(j\lambda) d\lambda \right| \\ &\leq 2 \left( \int_0^\varepsilon |\log(f(\lambda, \theta_1)) - \log(f(\lambda, \theta_2))| d\lambda + \int_\varepsilon^\pi \varepsilon d\lambda \right). \end{aligned}$$

By assumption B1, we have

$$\begin{aligned} \int_0^\varepsilon |\log(f(\lambda, \theta_1))| d\lambda &\leq \varepsilon \max(|\log(C_0(\delta))|, |\log(C(\delta))|) + (\alpha(\theta_1) + \delta) \int_0^\varepsilon |\log(\lambda)| d\lambda \\ &\leq \varepsilon \max(|\log(C_0(\delta))|, |\log(C(\delta))|) + (1 + \delta)\varepsilon(1 - \log(\varepsilon)). \end{aligned}$$

Finally we obtain that as  $\|\theta_1 - \theta_2\| \leq \eta$

$$|d_j(\theta_1) - d_j(\theta_2)| \leq 2[2\varepsilon \max(|\log(C_0(\delta))|, |\log(C(\delta))|) + 2(1 + \delta)\varepsilon(1 - \log(\varepsilon)) + \pi\varepsilon].$$

When  $\varepsilon$  goes to 0, the bound goes also to 0, then we have obtained the continuity of  $d_j(\theta)$ .

### 3.4.2 Proof of Lemma 3.2.2

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

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

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

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

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

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

It is justified because  $\theta \mapsto [f(\lambda, \theta)]^{-1}$  is twice differentiable with respect to  $\theta$  and all the partial derivatives are integrable on  $[-\pi, \pi]$  with respect to  $\lambda$  by assumption B3. It follows from Fox et Taqqu (1986) that :

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

where  $W = (W_{i,j})_{1 \leq i, j \leq p}$  defined in (3.17) 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 :

$$\exists M \in \mathbb{N}, \quad T > M, \quad \left\| \hat{\theta}_k - \theta_0 \right\|_4 \leq C \left\| \frac{\partial}{\partial \theta} \sigma_T(\theta_0) \right\|_4 \quad \mathbb{P}\text{-a.s.} \quad (3.39)$$

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

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

Let  $m \in \llbracket 1, p \rrbracket$ . We define the matrix  $\Delta_m$  with  $(j, l)$ -th entries :

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

Next we rewrite this expression as :

$$\begin{aligned} \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta_m} \sigma_T(\theta_0) \right)^4 \right] &= T^{-4} \mathbb{E} \left( \left( \sum_{j=1}^T \sum_{l=1}^T Y_j Y_l \delta_{j,l} \right)^4 \right) \\ &= T^{-4} \sum_{j_1, j_2, \dots, j_8=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \delta_{j_5, j_6} \delta_{j_7, j_8} \mathbb{E}(Y_{j_1} Y_{j_2} \dots Y_{j_8}) \end{aligned} \quad (3.40)$$

Since we have a Gaussian process, all the moments are functions of the autocovariances (see Major (1981)). Therefore in equation (3.40) we can rewrite each fourth moment in the sum as a linear combination of product of 4 covariances. The aim of the following decomposition is to factorise terms like  $\delta_{i,j} \mathbb{E}(Y_i Y_j)$  to extract the sum :  $\sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l)$ . We then count how many covariances belong to the set  $S = \{\mathbb{E}(Y_{j_1} Y_{j_2}), \mathbb{E}(Y_{j_3} Y_{j_4}), \mathbb{E}(Y_{j_5} Y_{j_6}), \mathbb{E}(Y_{j_7} Y_{j_8})\}$  in the product of 4 covariances :

1. either we have  $\mathbb{E}(Y_{j_1} Y_{j_2}) \times Z$  and we can distinguish the following possibilities :
  - $Z = \mathbb{E}(Y_{j_3} Y_{j_4}) \mathbb{E}(Y_{j_5} Y_{j_6}) \mathbb{E}(Y_{j_7} Y_{j_8})$  only one possibility. The corresponding term is

$$T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^4$$

or ;

- $Z$  has one element in  $S$  and no other which makes 6 possibilities = (3 choices in  $S$ )  $\times$  (2 choices for the other covariances). The corresponding term is

$$6T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^2 \sum_{j_1, j_2, j_3, j_4=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \sigma(j_1 - j_3) \sigma(j_2 - j_4)$$

or ;

- $Z$  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). The corresponding term is

$$8T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right) \sum_{j_1, j_2, j_3, j_4, j_5, j_6=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \delta_{j_5, j_6} \sigma(j_1 - j_3) \sigma(j_4 - j_5) \sigma(j_6 - j_2);$$

2. or  $Y_{j_1}$  is with  $Y_{j_l}$ ,  $l > 2$ , which makes 6 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  $(Y_{j_7}, Y_{j_8})$ . The corresponding term is

$$6T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^2 \sum_{j_1, j_2, j_3, j_4=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \sigma(j_1 - j_3) \sigma(j_2 - j_4)$$

or ;

- we have only one couple in  $S$  (which makes 4 choices for the other covariances) The corresponding term is

$$24T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right) \sum_{j_1, j_2, j_3, j_4, j_5, j_6=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \delta_{j_5, j_6} \sigma(j_1 - j_3) \sigma(j_4 - j_5) \sigma(j_6 - j_2)$$

or ;

- we have no 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  $Y_{j_4}$ . The corresponding terms are

$$12T^{-4} \left( \sum_{j_1, j_2, j_3, j_4=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \sigma(j_1 - j_3) \sigma(j_2 - j_4) \right)^2$$

$$+ 48T^{-4} \sum_{j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \delta_{j_5, j_6} \delta_{j_7, j_8} \sigma(j_1 - j_3) \sigma(j_4 - j_5) \sigma(j_6 - j_7) \sigma(j_8 - j_2).$$

Therefore we obtain :

$$\begin{aligned}
& \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta_m} \sigma_T(\theta_0) \right)^4 \right] \\
&= T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^4 \\
&+ 6T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^2 \sum_{j_1, j_2, j_3, j_4=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \sigma(j_1 - j_3) \sigma(j_2 - j_4) \\
&+ 8T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^2 \sum_{j_1, j_2, j_3, j_4, j_5, j_6=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \delta_{j_5, j_6} \sigma(j_1 - j_3) \sigma(j_4 - j_5) \sigma(j_6 - j_2) \\
&+ 6T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^2 \sum_{j_1, j_2, j_3, j_4=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \sigma(j_1 - j_3) \sigma(j_2 - j_4) \\
&+ 24T^{-4} \left( \sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) \right)^2 \sum_{j_1, j_2, j_3, j_4, j_5, j_6=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \delta_{j_5, j_6} \sigma(j_1 - j_3) \sigma(j_4 - j_5) \sigma(j_6 - j_2) \\
&+ 12T^{-4} \left( \sum_{j_1, j_2, j_3, j_4=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \sigma(j_1 - j_3) \sigma(j_2 - j_4) \right)^2 \\
&+ 48T^{-4} \sum_{j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8=1}^T \delta_{j_1, j_2} \delta_{j_3, j_4} \delta_{j_5, j_6} \delta_{j_7, j_8} \sigma(j_1 - j_3) \sigma(j_4 - j_5) \sigma(j_6 - j_7) \sigma(j_8 - j_2).
\end{aligned}$$

All the terms of this sum are of the form :

$$\sum_{j_1, \dots, j_{2p}=1}^T \delta_{j_1, j_2} \dots \delta_{j_{2p-1}, j_{2p}} \sigma(j_1 - j_3) \sigma(j_4 - j_5) \dots \sigma(j_{2p-2} - j_{2p-1}) \sigma(j_{2p} - j_2) := S_{p,T}. \quad (3.41)$$

Note that  $S_{p,T} = \text{trace}((\Sigma_T \Delta_m)^p)$  where  $\Delta_m$  is the Toeplitz matrix defined by the symbol :

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

by assumption B2. By applying Theorem 1 of Fox et Taquq (1987), we prove that :

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



It follows from assumptions B1-B2 that this integral is always finite and then we have

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{j_1, \dots, j_{2p}=1}^T \delta_{j_1, j_2} \cdots \delta_{j_{2p-1}, j_{2p}} \sigma(j_2 - j_3) \cdots \sigma(j_{2p} - j_1) = O(1). \quad (3.42)$$

We need a more precise result for the term :

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

Although this can also be expressed like (3.41), the estimate given above is not sufficient to conclude. By Fox et Taquq (1986) [proof of Theorem 2], we have :

$$\sum_{j=1}^T \sum_{l=1}^T \delta_{j,l} \sigma(j-l) = O(T^\delta). \quad (3.43)$$

By (3.42) and (3.43), we conclude that

$$\forall \delta > 0, \quad \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta_m} \sigma_T(\theta_0) \right)^4 \right] = O(T^{-2}). \quad (3.44)$$

Next using the asymptotic estimate of the fourth moments, we can now obtain asymptotic properties for the second moments for fixed  $j$  and  $l$  :

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

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

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

The last inequality is due to (3.39). By applying result (3.44), we conclude that :

$$T^2 \mathbb{E} \left( \left( \widehat{\theta}_{T,j} - \theta_{0,j} \right)^2 \left( \widehat{\theta}_{T,l} - \theta_{0,l} \right)^2 \right) = O(T^{2-2}) = O(1).$$

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 sufficiently large  $T$ , then the collection  $(X_T)$  is uniformly integrable. Moreover according to Fox et 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 (3.17) 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 (1999) that implies :

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

Now we give an asymptotic bound for the third order moment by applying the Cauchy-Schwarz inequality. Using the inequalities (3.39) and (3.44), we conclude that there exists  $C > 0$  such that :

$$\begin{aligned} \mathbb{E} \left( \left| \widehat{\theta}_{T,j} - \theta_{0,j} \right|^3 \right) &\leq \sqrt{\mathbb{E} \left( \widehat{\theta}_{T,j} - \theta_{0,j} \right)^2 \mathbb{E} \left( \widehat{\theta}_{T,j} - \theta_{0,j} \right)^4} \\ &\leq \sqrt{CT^{-1}T^{-2}} \\ &= \mathcal{O} \left( T^{-3/2} \right). \end{aligned}$$

## Chapitre 4

# Erreur quadratique due à l'estimation des coefficients du prédicteur projeté

### 4.1 Introduction

Nous allons maintenant nous intéresser au prédicteur projeté sur les  $k$  dernières observations disponibles :

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

et comme dans le chapitre précédent, estimer l'erreur quadratique de prévision supplémentaire due à l'estimation des coefficients.

On suppose ici que l'on dispose de deux copies indépendantes du même processus :  $(X_j)_{j \in \llbracket 1, k \rrbracket}$  la version que l'on veut prédire et  $(Y_j)_{j \in \llbracket 1, T \rrbracket}$ , la version qui va nous permettre d'estimer les coefficients du prédicteur.

Pour estimer les coefficients, on remplace dans les équations de Yule-Walker

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

les covariances par les covariances empiriques calculées sur les données  $(Y_j)_{j \in \llbracket 1, T \rrbracket}$  :

$$\widehat{\sigma}(k) = \frac{1}{T} \sum_{t=1}^{T-k} Y_t Y_{t+k} \quad (4.1)$$

On utilise ensuite l'algorithme de Durbin-Levinson ou algorithme d'innovation (Brockwell et Davis (1991)) pour résoudre le système.

On note  $(\widehat{a}_{1,k}, \dots, \widehat{a}_{k,k})$  les coefficients obtenus en résolvant les équations de Yule-Walker en remplaçant les covariances par les covariances empiriques définies en (4.1). Contrairement au prédicteur de Wiener-Kolmogorov tronqué (voir la section 3.2.2), l'estimation des coefficients est non-paramétrique.

## 4.2 Estimation de l'erreur quadratique due à l'estimation des coefficients

Étudions maintenant l'estimation de l'erreur quadratique de prévision due à l'estimation des coefficients de l'AR( $k$ ) lorsque  $T$  tend vers l'infini. On note  $\widehat{X}_{T,k}(1)$  le prédicteur dont les coefficients de Yule-Walker sont estimés :

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

On s'intéresse à l'erreur quadratique de prévision suivante :

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

Il nous faut donc évaluer tout d'abord :

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

Pour cela on définit le vecteur :

$$\mathbf{1}_k^* := \underbrace{(1, \dots, 1)}_k$$

et la matrice de taille  $k \times k$  :

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

et on énonce le théorème suivant.

**Théorème 4.2.1.** *Le processus  $(Y_n)_{n \in \mathbb{Z}}$  est supposé gaussien. On suppose en plus que sa fonction de covariance  $\sigma$  vérifie :*

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

que les coefficients de sa représentation en moyenne mobile infinie  $b_j$  vérifient :

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

et que le bruit blanc qui lui est associé  $(\varepsilon_n)_{n \in \mathbb{Z}}$  a des moments d'ordre 4 finis. On considère  $g_{i,j}$  la fonction définie par :

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

où

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

Alors

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

où  $C$  et  $D$  sont des constantes indépendantes de  $n$  et de  $k$  et  $H$  est la matrice définie comme suit. On pose  $h$  la fonction définie par  $h(\lambda) = |1 - \sum_{r=1}^k a_{r,k} e^{ir\lambda}|^2$  et on note  $h^{(r)}$  la dérivée de la fonction  $h$  par rapport à la variable  $a_{r,k}$ . Le terme de la  $i$ ème ligne et de la  $j$ ème colonne de la matrice est alors donné par la formule suivante :

$$\forall i, j \in \llbracket 1, k, \rrbracket, \quad H_{i,j} := \int_{-\pi}^{\pi} h^{(i)}(\lambda) h^{(j)}(\lambda) f^2(\lambda) d\lambda.$$

*Démonstration.* La démonstration est donnée en annexe 4.4.1. □

On revient maintenant à l'estimation du comportement de  $\mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right]$  et on énonce le théorème suivant qui traite les cas  $d \geq 1/4$  :

**Théorème 4.2.2.** *On se place sous les hypothèses du théorème 4.2.1. On suppose de plus que la densité spectrale du processus est telle que :*

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

où  $f_d$  la densité spectrale du processus fractionnaire  $F(d)$  est définie par :

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

## 4.2. Estimation de l'erreur quadratique due à l'estimation des coefficients 77

et  $L$  est une fonction positive, intégrable sur  $[-\pi, \pi]$ , continue en 0 et ayant une borne inférieure strictement positive.

Si  $d = 1/4$  alors pour tout réel  $k$ ,

$$\lim_{T \rightarrow +\infty} \left( \frac{T}{\log(T)} \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right) \leq C\sqrt{k}$$

et si  $d \in ]1/4, 1/2[$  alors pour tout réel  $k$ ,

$$\lim_{T \rightarrow +\infty} \left( T^{2-4d} \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right) \leq Ck^{1-2d}$$

où  $C$  est une constante indépendante de  $k$ .

*Démonstration.* En appliquant le théorème précédent, on obtient que si  $d = 1/4$  alors

$$\lim_{T \rightarrow +\infty} \left( \frac{T}{\log(T)} \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right) \leq C \left( \text{trace} \left( \left( \sum_{j=0}^k a_{j,k} \right)^2 \Sigma_k^{-1} \mathbf{1}_{k,k} \right) \right)$$

et que si  $d \in ]1/4, 1/2[$  alors

$$\lim_{T \rightarrow +\infty} \left( T^{2-4d} \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right) \leq C \left( \text{trace} \left( \frac{1}{T^{2-4d}} \left( \sum_{j=0}^k a_{j,k} \right)^2 \Sigma_k^{-1} \mathbf{1}_{k,k} \right) \right).$$

On commence par contrôler  $\left| \sum_{j=0}^k a_{j,k} \right|$  que l'on décompose de la façon suivante :

$$\left| \sum_{j=0}^k a_{j,k} \right| \leq \sqrt{k} \sqrt{\sum_{j=1}^k |a_{j,k} - a_j|^2} + \sum_{j=0}^k |a_j|.$$

On reprend alors la démonstration du théorème 3.3 de Inoue et Kasahara (2006) sur la convergence des coefficients de l'AR( $k$ ) vers les coefficients de l'AR( $\infty$ ). On note qu'il existe  $C_1$ ,  $C_2$  et  $K$  tels que si :

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

et si on note  $C$  une constante générique :

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

On a alors si  $k \geq K$  :

$$\begin{aligned}
\sum_{j=1}^k (a_{j,k} - a_j)^2 &\leq \frac{C}{k^2} \sum_{j=1}^k \left[ \left( \sum_{u=k-j}^{+\infty} |a_u| \right)^2 + \left( \sum_{u=j}^{+\infty} |a_u| \right)^2 \right] \\
&\leq \frac{C}{k^2} \sum_{j=1}^k (k-j+1)^{-2d} + j^{-2d} \\
&\leq O(k^{-2d-1}).
\end{aligned} \tag{4.2}$$

Ceci nous permet de conclure que :

$$\begin{aligned}
\left| \sum_{j=0}^k a_{j,k} \right| &\leq O(k^{-d}) + \sum_{j=0}^{+\infty} |a_j| \\
&= O(1).
\end{aligned} \tag{4.3}$$

Il nous reste donc à étudier le comportement de :

$$\text{trace}(\Sigma_k^{-1} \mathbf{1}_{k,k}) = (1 \dots 1) \Sigma_k^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

On applique alors le théorème 6.1 de Adenstedt (1974) qui sous les hypothèses du théorème 4.2.2 nous donne l'équivalent suivant :

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

où  $\Gamma$  et  $\beta$  sont les fonctions Gamma et Beta. Ceci nous permet de conclure. □

**Remarque** Le théorème reste vrai dans le cas des processus à longue mémoire saisonniers car on peut autoriser des densités spectrales  $f$  de la forme :

$$\begin{aligned}
&\exists r \in \mathbb{N}, \exists \lambda_1, \dots, \lambda_r \in [-\pi, \pi], \exists \alpha_1, \dots, \alpha_r \in \mathbb{R}^+, \\
&\forall x \in [-\pi, \pi], f(x) = f_d(x) L(x) |x - \lambda_1|^{\alpha_1} \dots |x - \lambda_r|^{\alpha_r}
\end{aligned}$$

avec  $f_d$  et  $L$  ayant les mêmes propriétés que précédemment. En effet, on n'utilise la forme de la densité spectrale que lorsqu'on applique le théorème de Adenstedt (1974).

Il reste à traiter le cas  $0 < d < 1/4$ , dans ce cas on a le théorème suivant :

## 4.2. Estimation de l'erreur quadratique due à l'estimation des coefficients 79

**Théorème 4.2.3.** *On se place sous les hypothèses du théorème 4.2.1. On suppose de plus que la densité spectrale est bornée inférieurement par une constante strictement positive. Si  $0 < d < 1/4$  alors*

$$\lim_{T \rightarrow +\infty} \left( T \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right) \leq Ck$$

où  $C$  est une constante indépendante de  $k$ .

*Démonstration.* On note  $(\Phi_i := \Phi_{i,1} + \dots + \Phi_{i,i}x^{i-1})_{i \in \mathbb{N}^*}$  les polynômes orthogonaux associés à la densité spectrale  $f$  c'est-à-dire qu'ils sont tels que  $\Phi_i$  est de degré  $i - 1$  et

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

où  $\delta$  est le symbole de Kronecker. On construit à partir d'eux la matrice  $T_k$  :

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

$T_k$  ainsi définie vérifie les égalités suivantes :

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

et donc

$$\Sigma_k^{-1} = T_k^* T_k. \quad (4.4)$$

En utilisant l'égalité (4.4), on est ramené à l'étude de la limite suivante :

$$\begin{aligned} \lim_{T \rightarrow +\infty} \left( T \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right) &= \text{trace}(\Sigma_k^{-1} H) \\ &= \text{trace}(T_k H T_k^*) \end{aligned} \quad (4.5)$$

où  $H$  est la matrice définie dans le théorème 4.2.1. On note  $G_k : \lambda \mapsto \sum_{j=0}^k a_{j,k} e^{ij\lambda}$  et  $\lim_{T \rightarrow +\infty} \left( T \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right)$  se réécrit alors :

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



car  $\operatorname{Re}(ab) = \operatorname{Re}(a)\operatorname{Re}(b) - \operatorname{Im}(a)\operatorname{Im}(b)$ . Introduisons quelques notations :

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

On a donc  $A = B + C$ . Tout d'abord nous montrerons que  $A$ ,  $B$  et  $-C$  sont symétriques et positives ce qui implique que  $0 \leq \operatorname{trace}(A) \leq \operatorname{trace}(B)$ . Commençons par la symétrie :  $A$  est symétrique car la partie réelle d'un complexe conjugué est égale à la partie réelle du complexe,  $B$  est symétrique car  $\lambda \mapsto f^2(\lambda) |G_k(\lambda)|^2$  est une fonction paire et  $C$  est symétrique car la partie imaginaire d'un complexe conjugué est égale à l'opposé de la partie imaginaire du complexe. Passons à la positivité. On considère un vecteur  $q := (q_1, \dots, q_k)$  et on regarde le signe de :

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

Les traces de ces matrices  $A$ ,  $B$  et  $-C$  sont donc égales à la somme de leurs valeurs propres puisqu'elles sont diagonalisables car symétriques et sont positives puisque les matrices sont positives et donc toutes leurs valeurs propres le sont aussi. On en déduit que :

$$0 \leq \operatorname{trace}(A) \leq \operatorname{trace}(B). \quad (4.6)$$

On s'intéresse donc maintenant à l'étude de  $\operatorname{trace}(B)$  :

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

où  $K_k$  est le noyau reproduisant défini par :

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

On a supposé précédemment que la densité spectrale  $f$  était bornée inférieurement par une constante strictement positive  $c$ , on peut donc appliquer le théorème 2.2.4 de Simon (2005) et on obtient que :

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

On majore ensuite  $|G_k(\lambda)|^2$  :

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

comme on l'a démontré en (4.3). Cette majoration est indépendante de  $\lambda$ . Il reste enfin à remarquer que comme  $0 < d < \frac{1}{4}$ ,  $f$  est de carré intégrable. On a donc :

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

et on conclut de (4.5) et de (4.6) que :

$$\lim_{T \rightarrow +\infty} \left( T \mathbb{E} \left[ \left( \widehat{X}_{T,k}(1) - \widehat{X}_k(1) \right)^2 \right] \right) \leq Ck.$$

□

## 4.3 Conclusion

Les résultats obtenus dans les théorèmes 4.2.3 et 4.2.2 sont moins précis que la majoration de l'erreur quadratique due à l'estimation pour le prédicteur de Wiener Kolmogorov tronqué obtenue au chapitre 3. En effet ici on fait d'abord tendre le nombre d'observations pour estimer  $T$  vers l'infini puis on regarde comment se comporte cette limite en fonction de  $k$  alors que pour le prédicteur de Wiener-Kolmogorov tronqué on fait tendre simultanément  $k$  et  $T$  vers l'infini. Au chapitre 5, nous améliorons doublement les résultats obtenus dans les théorèmes 4.2.3 et 4.2.2 non seulement nous allons faire dépendre  $k$  de  $T$  mais en plus nous nous affranchissons de l'hypothèse selon laquelle les coefficients du prédicteur sont estimés sur une réalisation indépendante de celle à prédire.

## 4.4 Annexe

### 4.4.1 Preuve du théorème 4.2.1

Le processus  $(Y_n)_{n \in \mathbb{Z}}$  est supposé gaussien. On suppose en plus que sa fonction de covariance  $\sigma$  vérifie :

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

que les coefficients de sa représentation en moyenne mobile infinie  $b_j$  vérifient :

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

et que le bruit blanc qui lui est associé  $(\varepsilon_n)_{n \in \mathbb{Z}}$  a des moments d'ordre 4 finis. On considère  $g_{i,j}$  la fonction définie par :

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

où

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

Alors

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

où  $C$  et  $D$  sont des constantes indépendantes de  $n$  et de  $k$  et  $H$  est la matrice définie comme suit. On pose  $h$  la fonction définie par  $h(\lambda) = |1 - \sum_{r=1}^k a_{r,k} e^{ir\lambda}|^2$  et on note  $h^{(r)}$  la dérivée de la fonction  $h$  par rapport à la variable  $a_{r,k}$ . Le terme de la  $i$ ème ligne et de la  $j$ ème colonne de la matrice est alors donné par la formule suivante :

$$H_{i,j} := \int_{-\pi}^{\pi} h^{(i)}(\lambda) h^{(j)}(\lambda) f^2(\lambda) d\lambda \quad (4.7)$$

*Démonstration.* Pour cela nous allons faire un développement limité à l'ordre 2 de la fonction  $g_{i,j}$  et pour cela appliquer le théorème 5.4.3 de Fuller (1976) comme dans la section 3.2.2. Nous en appliquerons donc la version suivante :

- (i) Si  $\mathbb{E}(|\widehat{\sigma(k)} - \sigma(k)|^3) = O(a_n)$ ;
- (ii) si  $g_{i,j}$  est uniformément borné;
- (iii) si les dérivées premières et secondes de  $g_{i,j}$  sont continues et bornées sur une sphère  $S$  qui contient le vecteur  $(\sigma(0), \dots, \sigma(k))$
- (iv) si les dérivées troisièmes de  $g_{i,j}$  sont uniformément bornées sur la sphère  $S$  précédemment définie

alors

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

Nous allons commencer par en vérifier les hypothèses. Il nous faut majorer les moments d'ordre 3 des covariances empiriques.

**Lemma 4.4.1.**

$$\mathbb{E} \left[ \left| \frac{1}{n} \sum_{t=0}^{n-k} X_t X_{t+k} - \sigma(k) \right|^3 \right] = \begin{cases} O(n^{-3/2}) & \text{si } d \leq 1/4 \\ O(n^{6d-3}) & \text{si } d > 1/4 \end{cases} \quad (4.8)$$

*Démonstration.* La démonstration est donnée dans la section 4.4.2.  $\square$

On obtient ainsi la majoration du reste du développement limité. On sait de plus que la fonction  $g_{i,j}$  est une fonction uniformément bornée puisqu'elle renvoie les coefficients d'un processus autorégressif. Il existe une sphère de  $\mathbb{R}^{k+1}$  qui contient  $(\sigma(0), \dots, \sigma(k))$  sur laquelle  $g_{i,j}$  est une fonction  $C^3$  et sur laquelle toutes ses dérivées sont uniformément bornées. On peut donc appliquer le théorème 5.4.3 de Fuller (1976). On remarque tout d'abord que :

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

et

$$\begin{aligned} \forall l \in \llbracket 0, k \rrbracket, \quad \frac{\partial g_{i,j}}{\partial x_l}(\sigma(0), \dots, \sigma(k)) &= \frac{\partial y_i}{\partial x_l}(\sigma(0), \dots, \sigma(k))(y_j - a_{j,k}) \\ & \quad + (y_i - a_{i,k}) \frac{\partial y_j}{\partial x_l}(\sigma(0), \dots, \sigma(k)) \\ \frac{\partial g_{i,j}}{\partial x_l}(\sigma(0), \dots, \sigma(k)) &= 0 \end{aligned} \quad (4.9)$$

car

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

Le développement limité et le lemme 4.4.1 nous donnent donc que :

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

En utilisant les résultats de Hosking (1996), on peut calculer le terme d'ordre 2 du développement limité. Tout d'abord on remarque que

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

car les autres termes s'annulent au point  $(\sigma(0), \dots, \sigma(k))$  d'après (4.9). De plus Hosking (1996) a prouvé que :

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

où  $C$ ,  $D$  et  $F$  sont des constantes indépendantes de  $l$  et  $m$ . On peut maintenant passer au calcul de :

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

On commence par étudier le cas  $d \geq 1/4$  et on montre que :

$$\sum_{l=0}^k \frac{\partial y_i}{\partial x_l}(\sigma(0), \dots, \sigma(k)) = \left(1 - \sum_{r=1}^k a_{r,k}\right) \left(\Sigma_k^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right)_i \quad (4.10)$$

car si on note  $\sigma_0^k := (\sigma(0), \dots, \sigma(k))$ , les dérivées partielles s'écrivent

$$\begin{aligned} \frac{\partial y_i}{\partial x_l}(\sigma_0^k) &= \left(\Sigma_k^{-1} \left[ \frac{\partial}{\partial x_l} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \right] (\sigma_0^k)\right)_i \\ &- \left(\Sigma_k^{-1} \left[ \frac{\partial}{\partial x_l} \begin{pmatrix} x_0 & x_1 & \dots & x_k \\ x_1 & x_0 & \ddots & x_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ x_k & x_{k-1} & \dots & x_0 \end{pmatrix} \right] (\sigma_0^k) \Sigma_k^{-1} \begin{pmatrix} \sigma(1) \\ \vdots \\ \sigma(k) \end{pmatrix}\right)_i \end{aligned} \quad (4.11)$$

En utilisant (4.10), on obtient le résultat annoncé car :

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

Dans le cas  $d < 1/4$ , on remarque tout d'abord en utilisant (4.11) que :

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

On obtient alors que :

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

où la matrice  $H$  a été définie en (4.7). □

#### 4.4.2 Preuve du lemme 4.4.1

Montrons que :

$$\mathbb{E} \left[ \left| \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k) \right|^3 \right] = \begin{cases} O(n^{-3/2}) & \text{si } d \leq 1/4 \\ O(n^{6d-3}) & \text{si } d > 1/4 \end{cases} \quad (4.12)$$

*Démonstration.*

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

On étudie séparément les deux termes. On développe tout d'abord le produit :

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

Comme le processus est gaussien, on sait (voir Triantafyllopoulos (2003)) que :

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

et on obtient alors que :

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

Or

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

On traite de la même façon le terme :

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

Et on obtient donc que :

$$\sqrt{\mathbb{E} \left[ \left| \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k) \right|^2 \right]} = \begin{cases} O(n^{-1/2}) & \text{si } d \leq 1/4 \\ O(n^{2d-1}) & \text{si } d > 1/4 \end{cases} \quad (4.13)$$

On passe maintenant au deuxième terme :

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

On développe ensuite sachant que le processus est gaussien et en utilisant l'égalité énoncée par Triantafyllopoulos (2003) qui permet de réduire les moments pairs d'un vecteur gaussien à une expression n'utilisant que la fonction de covariance. On classe ensuite les termes selon la puissance de  $\sigma(k)$  que l'on peut mettre en facteur. En facteur de  $\sigma(k)^4$ , on a :

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

en facteur de  $\sigma(k)^2$ , on a :

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

et en facteur de  $\sigma(k)$ , on a :

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

On étudie le comportement asymptotique de cette quantité de la façon suivante :

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

Le terme en facteur de  $\sigma(k)$  est un  $O(n^{6d-4})$ . Les termes sans  $\sigma(k)$  en facteur sont de la forme :

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



qui se traite en repassant sous la forme intégrale comme précédemment et on trouve que ces termes sont des  $O(n^{8d-4})$ , où ils sont de la forme :

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

on sépare alors les deux sommes et on utilise les résultats précédemment démontrés pour montrer que :

$$\frac{1}{n^4} \sum_{t=1}^{n-k} \sum_{s=1}^{n-k} \sum_{r=1}^{n-k} \sum_{v=0}^{n-k} \sigma(t-s)^2 \sigma(r-v)^2 = \begin{cases} O(n^{-2}) & \text{si } d \leq 1/4 \\ O(n^{8d-4}) & \text{si } d > 1/4 \end{cases} \quad (4.14)$$

Il reste à sommer tous ces résultats et on obtient :

$$\sqrt{\mathbb{E} \left[ \left| \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k) \right|^4 \right]} = \begin{cases} O(n^{-1}) & \text{si } d \leq 1/4 \\ O(n^{4d-2}) & \text{si } d > 1/4 \end{cases} \quad (4.15)$$

Pour conclure on a obtenu que :

$$\mathbb{E} \left[ \left| \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k} - \sigma(k) \right|^3 \right] = \begin{cases} O(n^{-3/2}) & \text{si } d \leq 1/4 \\ O(n^{6d-3}) & \text{si } d > 1/4 \end{cases} \quad (4.16)$$

□

## Chapitre 5

# Prédiction et estimation du modèle sur une même réalisation

Dans les deux chapitres précédents, l'estimation des coefficients du prédicteur était faite sur une réalisation indépendante de celle que l'on souhaitait prévoir. Or en pratique on dispose rarement de deux réalisations indépendantes du même processus. Ing et Wei (2003) se sont affranchis de cette hypothèse contraignante pour le prédicteur projeté d'un processus à courte mémoire (densité spectrale bornée inférieurement et supérieurement). Nous allons nous inspirer de leur travail pour généraliser leurs résultats pour des processus à longue mémoire.

Nous allons commencer par démontrer des inégalités sur les moments des estimées des matrices d'autocovariance des processus à longue mémoire. En effet les coefficients du prédicteur projeté sont calculés en utilisant les équations de Yule-Walker (2.24). La vitesse de convergence des coefficients estimés dépend donc de la vitesse de convergence de l'inverse de la matrice des autocovariances estimée. On bornera ainsi les moments de la matrice de covariance empirique de taille  $k \times k$  :

$$\widehat{\Sigma}_n(k) := \frac{1}{n - K_n + 1} \sum_{j=K_n}^n \mathbf{X}_j(k) \mathbf{X}_j'(k)$$

avec  $\mathbf{X}_j'(k) := (X_j, \dots, X_{j-k+1})$  et sa vitesse de convergence vers la matrice de covariance du processus  $(X_n)_{n \in \mathbb{Z}}$ .

Ensuite nous passerons à l'étude de l'erreur quadratique du prédicteur projeté. Elle se décompose en trois termes : l'erreur minimale de prédiction (la variance du bruit blanc  $\sigma_\varepsilon^2$ ), l'erreur due à la projection plus un terme supplémentaire dû à l'estimation des coefficients du prédicteur. L'erreur due à la projection, nous avons donné son comportement asymptotique  $O(k^{-1})$  au chapitre 1. Si on note  $n$  la longueur du passé utilisé pour prévoir et  $K_n$  la longueur du passé utilisé pour estimer les coefficients, l'erreur supplémentaire a comme équivalent asymptotique  $\frac{K_n}{n}$  s'il existe  $\delta > 0$  tel que  $K_n^4 = o(n^{1-2d-\delta})$ . Cette erreur due à l'estimation a le même équivalent asymptotique que dans le théorème prouvé par Ing et Wei (2003) en courte mémoire mais sous des conditions plus restrictives. Il est

à remarquer que pour minimiser l'erreur quadratique totale de prévision, il faudra faire un compromis entre l'erreur due à la projection et celle due à l'estimation des coefficients car la première décroît avec  $K_n$  alors que l'autre croît. En effet plus on projette sur un nombre important de données  $k$ , plus il y a de coefficients à estimer et plus les erreurs dues à leurs estimations se cumulent.

Enfin on déduit de cette expression de l'erreur quadratique un théorème de la limite centrale de notre prédicteur vers le prédicteur de Wiener-Kolmogorov. La normalisation est donnée par la vitesse de convergence du prédicteur tronqué et non comme en courte mémoire par celle de l'estimation (voir Bhansali (1978) et Lewis et Reinsel (1985)). Comme l'erreur due à la projection est plus importante en longue mémoire qu'en courte mémoire, ce changement de normalisation est naturel.

Ce chapitre a été soumis pour publication à Journal of Statistical Planning Inference sous le titre Prevision of long-memory processes on same-realisation.

## 5.1 Introduction

Consider  $(X_t)_{t \in \mathbb{Z}}$  a stationary process with zero mean and finite variance. We wish to predict  $X_{n+1}$  from the observed past  $(X_1, \dots, X_n)$  using a linear predictor i.e. a linear combination of the observed data. First we define the coefficients of the optimal predictor in the least-squares sense assuming that the covariance function is known. Then we need to estimate the replace coefficients. This second step is often realised under the following restrictive hypothesis : we predict another future independent series with exactly the same probabilistic structure ; the observed series is only used to estimate the forecast coefficients (see for example Bhansali (1978), Lewis et Reinsel (1985) or Godet (2007a)). This assumption makes the mathematical analysis easier since the prediction problem can be reduced to an estimation problem of the forecast coefficients by conditioning on the process, which we forecast. But the practitioner rarely has two independent series : one to estimate the model and one to predict. He has to estimate the forecast coefficients on the same realisation as the forecast one. In the following we concentrate on this case called same-realisation prediction.

The performance of the predictor depends on two parameters : the dimension of the subspace on which we project and the number of available data to estimate the forecast coefficients. To reduce the prediction error, it is reasonable to increase the dimension of the space onto which we project, as more and more observations become available. But when the dimension and then the number of forecast coefficients increase, the estimation of these coefficients becomes more difficult and can affect the mean-squared error.

When the spectral density of the process  $(X_t)_{t \in \mathbb{Z}}$  exists, is bounded and bounded away from 0 (this is typical of the short memory case) Ing et Wei (2003) and Kunitomo et Yamamoto (1985) have studied the mean-squared prediction error for same-realisation prediction. The mean squared error for same-realisation prediction can be approximated by the sum of two terms : one due to the goodness of fit and one due to the model complexity. In the short-memory case, it is interesting to remark that the approximations of

the mean-squared error for same and independent realisation prediction are the same. The performance of the least-squares predictor of long-memory time series is still left unanswered. And in this case, the asymptotic equivalence between the mean-squared error in same and independent realisation should not be taken for granted since the autocovariance function decays more slowly than in the short memory case.

The paper is organised as follows. In Sections 5.2 and 5.3 we generalise the results of Ing et Wei (2003) to find an asymptotic expression of the mean-squared error for long-memory time series. The mean-squared error is approximated by the same function as in the short memory case but under more restrictive conditions on the number of available observations and on the model complexity. In the last section, we prove a central limit theorem. More precisely, we prove the convergence in distribution of the normalised difference between our predictor and the Wiener-Kolmogorov predictor, which is the least-squares predictor knowing all the past. The normalisation is different from the short memory case since it is given by the goodness of fit of the projection.

**Definition of the Predictor** Let  $(X_t)_{t \in \mathbb{Z}}$  a stationary process with zero mean and finite variance. We assume that the autocovariance function  $\sigma$  of the process is known. Our goal is to predict  $X_{n+1}$ , using the  $k$  previous observed data. The optimal linear predictor is defined as the projection mapping onto the closed span of the subset  $\{X_n, \dots, X_{n-k+1}\}$  in the Hilbert space  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  with inner product  $\langle X, Y \rangle = \mathbb{E}(X'Y)$  where  $X'$  denotes the transpose of the vector  $X$ . It is the least-squares predictor knowing  $(X_{n-k+1}, \dots, X_n)$ . We denote by  $\tilde{X}_{n+1}(k)$  this predictor and by  $-a_{j,k}$  the theoretical prediction coefficients i.e. :

$$\tilde{X}_{n+1}(k) = \sum_{j=1}^k (-a_{j,k}) X_{n+1-j}. \quad (5.1)$$

They are given by (see Brockwell et Davis (1988) Section 5.1) :

$$\begin{pmatrix} a_{1,k} \\ \vdots \\ a_{k,k} \end{pmatrix} = -\Sigma(k)^{-1} \begin{pmatrix} \sigma(1) \\ \vdots \\ \sigma(k) \end{pmatrix} \quad (5.2)$$

where  $\Sigma(k)$  is the covariance matrix of the vector  $(X_1, \dots, X_k)$ .

**Estimation of the Forecast Coefficients** When the autocovariance function  $\sigma$  of the process  $(X_t)_{t \in \mathbb{Z}}$  is unknown, we can plug-in an estimate of the prediction coefficients  $(-a_{j,k})$  in (5.1). The estimate is constructed from the last  $n$  observations  $(X_n, \dots, X_1)$  and our predictor is the projection of the last  $k$  observations ( $k \leq n$ ). The covariance matrix is estimated by :

$$\hat{\Sigma}_n(k) := \frac{1}{n - K_n + 1} \sum_{j=K_n}^n \mathbf{X}_j(k) \mathbf{X}_j'(k) \quad (5.3)$$

where

$$\mathbf{X}'_j(k) := (X_j, \dots, X_{j-k+1}) \quad (5.4)$$

and where  $K_n$  is the maximum dimension of the subspace, onto which we project i.e. we will study the family of predictors  $(\widehat{X}_{n+1}(k))_{1 \leq k \leq K_n}$ .  $K_n$  will be an increasing sequence of integers which can be bounded or can go to infinity.

The prediction coefficients  $a_{i,k}$  are estimated from  $(X_1, \dots, X_n)$  by :

$$\widehat{\mathbf{a}}'(k) = (-\widehat{a}_{1,k}, \dots, -\widehat{a}_{k,k}) = \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) X_{j+1}.$$

The resulting one-step predictor is :

$$\widehat{X}_{n+1}(k) = \mathbf{X}'_n(k) \widehat{\mathbf{a}}(k) \quad (5.5)$$

In this paper we use  $C$  to denote generic positive constants that are independent of the sample size  $n$  but may depend on the distributional properties of the process  $(X_t)_{t \in \mathbb{Z}}$ . Moreover  $C$  may also stand for different values in different equations.

The following assumptions on the process  $(X_t)_{t \in \mathbb{Z}}$  are essential to the results presented in the paper. There exists  $d \in ]0, 1/2[$  such that :

H.1 The stationary process  $(X_t)_{t \in \mathbb{Z}}$  is Gaussian and admits an infinite moving average representation and an infinite autoregressive representation as follows :

$$\varepsilon_t = \sum_{j=0}^{+\infty} a_j X_{t-j} \quad \text{and} \quad X_t = \sum_{j=0}^{+\infty} b_j \varepsilon_{t-j} \quad (5.6)$$

with  $a_0 = b_0 = 1$ , for any  $j \geq 1$  and for any  $\delta > 0$ ,  $|a_j| \leq C j^{-d-1+\delta}$  and  $|b_j| \leq C j^{d-1+\delta}$  and  $(\varepsilon_t)_{t \in \mathbb{Z}}$  is a white noise process. These assumptions on the coefficients are verified by both long-memory and short memory processes ;

H.2 The covariance  $\sigma(k)$  is equivalent to  $L(k)k^{2d-1}$  as  $k$  goes to infinity, where  $L$  is a slowly varying function (i.e. for every  $\alpha > 0$ ,  $x^\alpha L(x)$  is ultimately increasing and  $x^{-\alpha} L(x)$  is ultimately decreasing). Under this assumption the autocovariances are not absolutely summable and thus the process is long-memory process ;

H.3 The spectral density of the process  $(X_t)_{t \in \mathbb{Z}}$  exists and has a strictly positive lower bound ;

H.4 The coefficients  $(a_j)_{j \in \mathbb{N}}$  verify :

$$a_j \underset{j \rightarrow +\infty}{\sim} L(j)j^{-d-1} \quad (5.7)$$

with  $L$  a slowly varying function.

For example, the assumptions H.1-H.4 hold for the most studied long-memory process, the Gaussian FARIMA process, which is the stationary solution to the difference equations :

$$\phi(B)(1 - B)^d X_n = \theta(B)\varepsilon_n \quad (5.8)$$

where  $(\varepsilon_n)_{n \in \mathbb{Z}}$  is a white noise series with mean zero,  $B$  is the backward shift operator and  $\phi$  and  $\theta$  are polynomials with no zeroes in the unit disk.

We only use assumptions H.1-H.3 to give an asymptotic expression of the mean-squared error of the predictor. Assumption H.4 is a more restrictive assumption used to prove a central limit theorem for our predictor.

Assumption H.2 does not imply the bound on the coefficients  $(a_j)_{j \in \mathbb{N}}$  and  $(b_j)_{j \in \mathbb{N}}$  given in assumption H.1. Inoue (2000) has proved that the asymptotic expression of the autocovariance  $\sigma(k) \sim L(k)k^{2d-1}$  implies :

$$b_j \sim j^{d-1} \sqrt{\frac{L(j)}{B(d, 1-2d)}} \text{ as } j \rightarrow +\infty$$

and

$$a_j \sim j^{-d-1} \left( \sqrt{\frac{L(j)}{B(d, 1-2d)}} \right)^{-1} \frac{d \sin(\pi d)}{\pi} \text{ as } j \rightarrow +\infty$$

if we assume that the sequences  $(b_j)_{j \in \mathbb{N}}$  and  $(a_j)_{j \in \mathbb{N}}$  are eventually decreasing to zero and  $b_j \geq 0$  for all  $j \in \mathbb{N}$ . Such assumptions on the sign of the sequence  $(b_j)_{j \in \mathbb{N}}$  or its monotonicity are not necessary for example to prove Lemma 5.2.1 and to find moment bounds for the inverse sample covariance matrix.

## 5.2 Moment bounds

In this section, we establish moment bounds for the inverse sample covariance matrix and apply these results to obtain the rate of convergence of  $\widehat{\Sigma}_n(k)$  to  $\Sigma(k)$ .

Throughout the paper,  $\lambda_{\min}(Y)$  and  $\lambda_{\max}(Y)$  are respectively the smallest and the largest eigenvalues of the matrix  $Y$ . We equip the set of matrices with the norm

$$\|Y\|^2 = \lambda_{\max}(Y'Y) \quad (5.9)$$

(see for example Dahlhaus (1989)). For a symmetric matrix, this norm is equal to the spectral radius and for a vector  $(X_1, \dots, X_n)$ , it is equal to  $\sqrt{\sum_{i=1}^n X_i^2}$ . This norm is a matrix norm that verifies for any matrices  $A$  and  $B$  :

$$\|AB\| \leq \|A\| \|B\|. \quad (5.10)$$

**Lemma 5.2.1.** *Let  $(K_n)_{n \in \mathbb{N}}$  be an increasing sequence of positive integers satisfying  $K_n = o(\sqrt{n})$ . Assume (H.1). Then, for any  $q > 0$ , for any  $\theta > 0$  and for any  $1 \leq k \leq K_n$ ,*

$$\mathbb{E} \left[ \frac{\lambda_{\min}^{-q}(\widehat{\Sigma}_n(k))}{k^{(2+\theta)q}} \right] = O(1)$$

where  $\widehat{\Sigma}_n(k)$  is defined in (5.3).

*Proof.* The sketch of the proof is the same as that of Lemma 1 of Ing et Wei (2003). The arguments are the following :

1.  $\sum_{j=1}^{+\infty} |a_j|$  converges ;
2. the cumulative distribution function of the random variable  $\varepsilon_t$  is a Lipschitz function and we may choose a Lipschitz constant independent of  $t$ . For any integer  $t$  and for any real numbers  $x$  and  $y$ , there exists  $C$  independent of  $t$  such that :

$$|\mathbb{P}(\varepsilon_t < x) - \mathbb{P}(\varepsilon_t < y)| \leq C|x - y|.$$

In our context these two conditions are satisfied. The sequence  $(a_j)_{j \in \mathbb{N}}$  is summable under assumption H.1. Since we have assumed that the process  $(X_t)_{t \in \mathbb{Z}}$  is Gaussian,  $(\varepsilon_t)_{t \in \mathbb{Z}}$  is a sequence of independent and identically distributed Gaussian random variables. The distribution function of the process  $\varepsilon_t$  is independent of  $t$  and is a Lipschitz function.  $\square$

For  $n$  sufficiently large Lemma 5.2.1 guarantees that  $\widehat{\Sigma}_n^{-1}(k)$  almost surely exists as the minimum eigenvalue of  $\widehat{\Sigma}_n(k)$  is almost surely positive. We also obtain an upper bound for the mean of the maximum eigenvalue of  $\widehat{\Sigma}_n^{-1}(k)$ . But this upper bound is not uniform as  $k \rightarrow +\infty$  and therefore does not provide an asymptotic equivalent of the prediction error. Nevertheless the bound given in Lemma 5.2.1 is the basis for the following theorem.

**Theorem 5.2.1.** *Assume that the process  $(X_t)_{t \in \mathbb{Z}}$  verifies the hypotheses H.1-H.3*  
– *if  $d \in ]0, 1/4[$  and if there exists  $\delta > 0$  such that  $K_n^{2+\delta} = O(n)$  then for all  $q > 0$  and for all  $1 \leq k \leq K_n$  :*

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^q = O(1) \tag{5.11}$$

and

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \leq C \left( \frac{K_n^2}{n - K_n + 1} \right)^{q/4} \tag{5.12}$$

for sufficiently large  $n$  ;

- *if  $d \in ]1/4, 1/2[$  and if there exists  $\delta > 0$  and  $\delta' > 0$  such that  $K_n^{2+\delta} = O(n^{2-4d-\delta'})$  then for all  $q > 0$  and for all  $1 \leq k \leq K_n$  :*

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^q = O(1) \tag{5.13}$$

and

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \leq C \left( \frac{K_n^2 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right)^{q/4} \tag{5.14}$$

for sufficiently large  $n$  ;

- *if  $d = 1/4$  and if there exists  $\delta > 0$  and  $\delta' > 0$  such that  $K_n^{2+\delta} = O(n^{1-\delta'})$  then for all  $q > 0$  and for all  $1 \leq k \leq K_n$  :*

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^q = O(1) \tag{5.15}$$

and

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \leq C \left( \frac{K_n^2 L^2 (n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)} \right)^{q/4} \quad (5.16)$$

for sufficiently large  $n$ .

In the proof of Theorem 5.2.1, we need the following lemma.

**Lemma 5.2.2.** *If the process  $(X_t)_{t \in \mathbb{Z}}$  verifies (H.2), if  $1 \leq k \leq K_n$  and*

- *if  $d \in ]0, 1/4[$ , then for all  $q > 0$ ,*

$$\mathbb{E}\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q \leq C \left( \frac{K_n^2}{n - K_n + 1} \right)^{\frac{q}{2}}; \quad (5.17)$$

- *if  $d \in ]1/4, 1/2[$ , then for all  $q > 0$ ,*

$$\mathbb{E}\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q \leq C \left( \frac{K_n^2 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right)^{\frac{q}{2}}; \quad (5.18)$$

- *if  $d = 1/4$ , then for all  $q > 0$ ,*

$$\mathbb{E}\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q \leq C \left( \frac{K_n^2 L^2 (n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)} \right)^{\frac{q}{2}}. \quad (5.19)$$

*Proof.* We prove the inequalities (5.17), (5.18) and (5.19) only for  $q > 2$ . The general case ( $q > 0$ ) easily follows from Jensen's inequality. We consider the matrix norm  $\|\cdot\|_E$  (see Ciarlet (1982)) defined for any matrix  $Y = (y_{i,j})_{1 \leq i,j \leq k}$  by

$$\|Y\|_E = \sqrt{\sum_{i=1}^k \sum_{j=1}^k y_{i,j}^2}.$$

Since the matrix  $\widehat{\Sigma}_n(k) - \Sigma(k)$  is symmetric, we have

$$\|\widehat{\Sigma}_n(k) - \Sigma(k)\| \leq \|\widehat{\Sigma}_n(k) - \Sigma(k)\|_E.$$

We obtain :

$$\begin{aligned} \|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q &\leq \|\widehat{\Sigma}_n(k) - \Sigma(k)\|_E^q \\ &\leq \left( \sum_{i=1}^k \sum_{j=1}^k (\widehat{\sigma}_{i,j} - \sigma(i-j))^2 \right)^{q/2} \end{aligned} \quad (5.20)$$

where  $\widehat{\sigma}_{i,j}$  and  $\sigma(i-j)$  denote respectively the  $(i, j)$  entries of the matrices  $\widehat{\Sigma}_n(k)$  and  $\Sigma(k)$ .

Applying Jensen's inequality to (5.20) because  $q/2 > 1$ , we have :

$$\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q \leq \frac{k^q}{k^2} \sum_{i=1}^k \sum_{j=1}^k |\widehat{\sigma}_{i,j} - \sigma(i-j)|^q.$$



It follows that :

$$\mathbb{E} \|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q \leq k^{q-2} \sum_{i=1}^k \sum_{j=1}^k \mathbb{E} |\hat{\sigma}_{i,j} - \sigma(i-j)|^q \quad (5.21)$$

Now we derive the limiting distribution of  $\hat{\sigma}_{i,j} - \sigma(i-j)$  to find an asymptotic expression of  $\mathbb{E} |\hat{\sigma}_{i,j} - \sigma(i-j)|^q$ . We shall work with the definition of the empirical covariances. By (5.3), we have :

$$\hat{\sigma}_{i,j} = \frac{1}{n - K_n + 1} \sum_{l=K_n}^n X_{l+i-1} X_{l+j-1} \stackrel{\mathcal{L}}{=} \frac{1}{n - K_n + 1} \sum_{l=1}^{n-K_n+1} X_l X_{l+j-i}, \quad (5.22)$$

where the second equality is ensured by the strict stationarity of the process. Without loss of generality, we assume  $j \geq i$ . The right term of (5.22) can be written :

$$\frac{1}{n - K_n + 1} \sum_{l=1}^{n-K_n+1} X_l X_{l+j-i} = \frac{1}{n - K_n + 1} \mathbf{X}'_1(n - K_n + 1 + j - i) T_{i,j} \mathbf{X}_1(n - K_n + 1 + j - i) \quad (5.23)$$

where  $\mathbf{X}_1(n - K_n + 1 + j - i)$  is defined in (5.4) and the entries of the matrix  $T_{i,j}$  verify

$$t_{i,j}(s, t) = \begin{cases} 1/2 & \text{if } |s - t| = j - i \\ 0 & \text{otherwise.} \end{cases}$$

$T_{i,j}$  is a Toeplitz matrix because it has symbol  $g_{i,j}(x) = \cos((j-i)x)$  i.e.  $t_{i,j}(s, t) = \int_{-\pi}^{\pi} g_{i,j}(x) \cos(t-s)x dx$ .

Under Assumption H.2 with  $d \in ]0, 1/4[$ , the spectral density verifies in a neighbourhood of 0 :

$$f(x) = O\left(x^{-1/2}\right)$$

(see Zygmund (1968) Chap. 5 Theorem 2.6). By applying Theorem 2 of Fox et Taqqu (1987) to (5.23), we obtain the following convergence :

$$\frac{(n - K_n + 1) (\hat{\sigma}_{i,j} - \sigma(i-j))}{\sqrt{n - K_n + 1 + j - i}} \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}\left(0, 4\pi \int_{-\pi}^{\pi} f^2(\lambda) \cos^2((i-j)\lambda) d\lambda\right) \quad (5.24)$$

where  $\implies$  denotes the convergence in distribution. This convergence in distribution follows from the convergence of all the cross-cumulants and hence the convergence of all the moments of the left term of (5.24). From this convergence in distribution we can deduce an asymptotic expression of the moments of  $\hat{\sigma}_{i,j} - \sigma(i-j)$ . If  $q$  is even, we have an asymptotic equivalent as  $n \rightarrow +\infty$  :

$$\mathbb{E} |\hat{\sigma}_{i,j} - \sigma(i-j)|^q \underset{n \rightarrow +\infty}{\sim} \left( \frac{\sqrt{n - K_n + 1 + j - i}}{n - K_n + 1} \right)^q \mathbb{E} [|Y|^q], \quad (5.25)$$

where  $Y$  is a Gaussian random variable which has the probability distribution given on the right hand side of (5.24). The  $q$ th-order absolute moment has the form :

$$E[|Y|^q] = \frac{q!}{2^{q/2}(q/2)!} \sigma_Y^q. \quad (5.26)$$

Moreover notice that for all  $(i, j)$  :

$$\sigma_Y \leq \sqrt{4\pi \int_{-\pi}^{\pi} f^2(\lambda) d\lambda} := M. \quad (5.27)$$

Thus (5.25), (5.26) and (5.27) imply for sufficiently large  $n$  :

$$\begin{aligned} \mathbb{E}|\hat{\sigma}_{i,j} - \sigma(i-j)|^q &\leq \left( \frac{\sqrt{n - K_n + 1 + j - i}}{n - K_n + 1} \right)^q \frac{q!}{(q/2)!} M^q \\ &\leq C \left( \frac{1}{\sqrt{n - K_n + 1}} \right)^q \end{aligned}$$

with  $C$  independent of  $(i, j)$ . The result (5.17) follows from the previous inequality and inequality (5.21) for  $q > 2$ .

For any  $d \in ]1/4, 1/2[$ , we apply the proposition of Rosenblatt (1979) which gives the following convergence in distribution :

$$\frac{(n - K_n + 1)(\hat{\sigma}_{i,j} - \sigma(i-j))}{L(n - K_n + 1)(n - K_n + 1 + j - i)^{2d}} \xrightarrow[n \rightarrow +\infty]{\Rightarrow} \mathcal{R}(1)$$

where  $\mathcal{R}$  is a Rosenblatt process. This convergence in distribution is obtained by proving the convergence of the cumulants of any order and hence the convergence of the moments of any order. This limit does not depend on the difference  $(j - i)$ . Similarly to the proof of (5.25) we show for any even integer  $q$  :

$$\mathbb{E}|\hat{\sigma}_{i,j} - \sigma(i-j)|^q \leq C \left( \frac{L(n - K_n + 1)}{(n - K_n + 1)^{1-2d}} \right)^q$$

where  $C$  does not depend on  $(j - i)$ . This inequality and (5.21) yield the desired result.

Finally for  $d = 1/4$ , we apply Theorem 4 of Hosking (1996), which gives the following convergence in distribution :

$$\sqrt{\frac{(n - K_n + 1)}{\log(n - K_n + 1)L^2(n - K_n + 1)}} (\hat{\sigma}_{i,j} - \sigma(i-j)) \xrightarrow[n \rightarrow +\infty]{\Rightarrow} \mathcal{N}(0, \sigma^2)$$

where  $\sigma$  depends on the process  $(X_t)_{t \in \mathbb{Z}}$  but not on  $(i, j)$ . This convergence in distribution is obtained by proving the convergence of the cumulants of any order and hence the convergence of the moments. We then obtain that for any even integer  $q$  :

$$\mathbb{E}|\hat{\sigma}_{i,j} - \sigma(i-j)|^q \leq C \left( \frac{L^2(n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)} \right)^{q/2}$$

with  $C$  independent of  $(i, j)$ . The result (5.19) follows from this inequality and (5.21).  $\square$

We now prove Theorem 5.2.1.

*Proof of Theorem 5.2.1.* Since  $\|\cdot\|$  is a matrix norm (see (5.10)),

$$\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq \|\widehat{\Sigma}_n^{-1}(k)\|^q \|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q \|\Sigma^{-1}(k)\|^q.$$

Furthermore, by assumption H.3, the spectral density of the process  $(X_t)_{t \in \mathbb{Z}}$  has a strictly positive lower bound. Thus from Grenander et Szegö (1958), there exists a constant  $C$  such that for all  $n \geq 0$  :

$$\|\Sigma^{-1}(k)\|^q \leq C. \quad (5.28)$$

Using Hölder's inequality with  $1/p' + 1/q' = 1$ , we obtain :

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C \left( \mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^{qq'} \right)^{1/q'} \left( \mathbb{E}\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^{qp'} \right)^{1/p'}.$$

By Lemma 5.2.1, we have for all  $\theta > 0$  and for large  $n$  :

$$\left( \mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^{qq'} \right)^{1/q'} \leq C(k^{2+\theta})^q$$

then

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C(k^{2+\theta})^q \left( \mathbb{E}\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^{qp'} \right)^{1/p'}. \quad (5.29)$$

We now apply Lemma 5.2.2. In order to treat together the three situations  $d \in ]0, 1/4[$ ,  $d \in ]1/4, 1/2[$  and  $d = 1/4$ , we define  $h(n)$  by :

$$h(n) = \begin{cases} \frac{K_n^2}{n - K_n + 1} & \text{if } d \in ]0, 1/4[ \\ \frac{K_n^2 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} & \text{if } d \in ]1/4, 1/2[ \\ \frac{K_n^2 L^2 (n - K_n + 1) \log(n - K_n + 1)}{n - K_n + 1} & \text{if } d = 1/4 \end{cases}$$

For large  $n$ , we then obtain :

$$\left( \mathbb{E}\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^{qp'} \right)^{1/p'} \leq C(h(n))^{q/2}. \quad (5.30)$$

From inequality (5.29) and the bound (5.30), we obtain that there exists  $\theta > 0$  such that for sufficiently large  $n$  :

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C(k^{4+\theta} h(n))^{q/2}. \quad (5.31)$$

By inequalities (5.31) and (5.28), we have :

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^q \leq C \left( 1 + \left( k^{4+\theta} h(n) \right)^{q/2} \right). \quad (5.32)$$

This inequality is not sufficient to obtain (5.11) and (5.13) since under the assumptions of Theorem 5.2.1,  $(k^{4+\theta}h(n))^{q/2}$  is not necessarily bounded. We have to improve the intermediate inequality (5.32).

The Cauchy-Schwarz inequality and (5.28) give :

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \leq C \left( \mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^q \right)^{1/2} \left( \mathbb{E}\|\widehat{\Sigma}_n(k) - \Sigma(k)\|^q \right)^{1/2}. \quad (5.33)$$

And there exists  $C > 0$  independent of  $q$  such that :

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^{q/2} \leq C \left( \mathbb{E}\|\Sigma^{-1}(k)\|^{q/2} + \mathbb{E}\|\widehat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \right). \quad (5.34)$$

Inequalities (5.33), (5.28), (5.32) and Lemma 5.2.2 imply that :

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^{q/2} \leq C \left( 1 + \left( k^{4+\theta}h(n)^2 \right)^{q/4} \right). \quad (5.35)$$

Repeating  $s - 1$  times this argument (i.e. using inequalities (5.34), (5.33), (5.28), (5.35) and Lemma 5.2.2), one has for large  $n$  :

$$\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^{q2^{-s}} \leq C \left( 1 + \left( k^{4+\theta}h(n)^{(1+s)} \right)^{q/2^{-(s+1)}} \right). \quad (5.36)$$

By assumption, there exists  $\delta > 0$  such that  $h(n)k^\delta$  converges to 0 as  $n$  tends to infinity, therefore there exists  $s$ , such that  $\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^{q2^{-s}}$  is bounded. Since  $q$  in (5.36) is arbitrary, (5.11) and (5.13) are proved. Inequalities (5.12) and (5.14) follow from (5.33) and from Lemma 5.2.2.  $\square$

In the following section, we establish an asymptotic expression for the mean-squared prediction error of the least-squares predictor using the tight upper bound for  $\mathbb{E}\|\widehat{\Sigma}_n^{-1}(k)\|^q$  given in Theorem 5.2.1.

### 5.3 The mean-squared prediction error of the least-squares predictor

In this section, our goal is to give an asymptotic expression of the mean-squared prediction error of the predictor defined in (5.5). First we decompose the forecast error in :

$$X_{n+1} - \widehat{X}_{n+1}(k) = \varepsilon_{n+1} + f(k) + \mathcal{S}_n(k) \quad (5.37)$$

where  $\varepsilon_{n+1}$  is the innovation white noise at time  $n + 1$  and cannot be forecast,  $\mathcal{S}_n(k)$  is the error due to the projection onto the closed span of the subset  $X_n, \dots, X_{n-k+1}$ , and  $f(k)$  is the error due to the estimation of the prediction coefficients. More precisely if we set  $a_{i,k} = 0$  for  $i > k$  (for  $j \leq k$ , the coefficients  $a_{j,k}$  are defined in (5.2)), we have

$$\mathcal{S}_j(k) = - \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \quad (5.38)$$

and with  $\varepsilon_{j+1,k}$  equal to the forecast error of  $X_{j+1}$  due to the projection onto  $(X_j, \dots, X_{j-k+1})$  i.e.

$$\varepsilon_{j+1,k} = X_{j+1} - P_{[X_{j-k}, \dots, X_j]}(X_{j+1}) = X_{j+1} + \sum_{l=1}^k a_{l,k} X_{j+1-l},$$

we have

$$f(k) = -\mathbf{X}'_n(k) \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1,k}$$

where  $\mathbf{X}'_n(k)$  is defined in (5.4).

In view of (5.37), we obtain the decomposition of the mean-squared prediction error as the sum of the variance  $\sigma_\varepsilon^2$  of the white noise and the error due to the prediction method  $\mathbb{E}(f(k) + \mathcal{S}_n(k))^2$  :

$$\mathbb{E}\left(X_{n+1} - \widehat{X}_{n+1}(k)\right)^2 = \sigma_\varepsilon^2 + \mathbb{E}(f(k) + \mathcal{S}_n(k))^2.$$

**Theorem 5.3.1.** *Under assumptions H.1-H.3, if we choose the sequence  $(K_n)_{n \in \mathbb{N}}$  such that for some  $\delta > 0$  :*

$$K_n^4 = o(n^{1-2d-\delta}), \quad (5.39)$$

then

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \left| \frac{\mathbb{E}\left(X_{n+1} - \widehat{X}_{n+1}(k)\right)^2 - \sigma_\varepsilon^2}{L_n(k)} - 1 \right| = 0$$

where

$$L_n(k) = \mathbb{E}(\mathcal{S}_n(k))^2 + \frac{k}{n - K_n + 1} \sigma_\varepsilon^2 \quad (5.40)$$

$\mathcal{S}_n(k)$  being defined in (5.38).

**Remark** If we fit a misspecified AR( $k$ ) model to the long-memory time series  $(X_n)_{n \in \mathbb{Z}}$  to forecast it, we find the same predictor as (5.1). Consequently  $L_n(k)$  can be viewed as the quality of prediction by an AR model. From (5.40), this quality is the sum of the model complexity  $\frac{k}{n - K_n + 1} \sigma_\varepsilon^2$  and the goodness of fit  $\mathbb{E}(\mathcal{S}_n(k))^2$ .

*Proof.* By (5.37), we have :

$$\left| \frac{\mathbb{E}\left(X_{n+1} - \widehat{X}_{n+1}(k)\right)^2 - \sigma_\varepsilon^2}{L_n(k)} - 1 \right| = \left| \frac{\mathbb{E}(f(k) + \mathcal{S}_n(k))^2}{L_n(k)} - 1 \right|.$$

Our proof is divided into three steps :

1. we provide an approximation of  $\mathbb{E}(f(k))^2$  which is easier to estimate. This approximation denoted by  $\mathbb{E}(f_1(k))^2$  will be defined in (5.41) ;

2. we show that the asymptotic equivalent of  $\mathbb{E}(f_1(k))^2$  is  $\frac{k}{n-K_n+1}\sigma_\varepsilon^2$ ;
3. we prove that the cross-product term  $\mathbb{E}(f(k)\mathcal{S}_n(k))$  is negligible with respect to  $L_n(k)$ .

**First step** We introduce

$$f_1(k) := -\mathbf{X}_n^{*\prime}(k)\Sigma^{-1}(k)\frac{1}{n-K_n+1}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{X}_j(k)\varepsilon_{j+1}, \quad (5.41)$$

with

$$\mathbf{X}_n^{*\prime}(k) = \left( \sum_{j=0}^{\sqrt{n}/2-K_n} b_j\varepsilon_{n-j}, \dots, \sum_{j=0}^{\sqrt{n}/2-K_n} b_j\varepsilon_{n-k+1-j} \right).$$

**Lemma 5.3.1.** *If the assumptions of Theorem 5.3.1 hold, then*

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \mathbb{E} \left( \sqrt{\frac{1}{L_n(k)}} (f(k) - f_1(k)) \right)^2 = 0. \quad (5.42)$$

*Proof.* See the appendix. □

**Second step** We prove that

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \left| \mathbb{E} \left( \frac{n-K_n+1}{k\sigma_\varepsilon^2} f_1^2(k) \right) - 1 \right| = 0. \quad (5.43)$$

First observe that

$$\begin{aligned} & \mathbb{E} \left( \frac{n-K_n+1}{k\sigma_\varepsilon^2} f_1^2(k) \right) \\ &= \frac{n-K_n+1}{k\sigma_\varepsilon^2} \mathbb{E} \left( \mathbf{X}_n^{*\prime}(k)\Sigma^{-1}(k)\frac{1}{n-K_n+1}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{X}_j(k)\varepsilon_{j+1} \right)^2 \\ &= \frac{n-K_n+1}{k\sigma_\varepsilon^2} \mathbb{E} \left[ \text{trace} \left( \mathbf{X}_n^{*\prime}(k)\Sigma^{-1}(k)\frac{1}{n-K_n+1}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{X}_j(k)\varepsilon_{j+1} \right)^2 \right] \\ &= \frac{n-K_n+1}{k\sigma_\varepsilon^2} \text{trace} \left[ \mathbb{E} \left( \Sigma^{-1}(k)\frac{1}{(n-K_n+1)^2}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{X}_j(k)\varepsilon_{j+1}\sum_{l=K_n}^{n-\sqrt{n}-1}\mathbf{X}'_l(k)\varepsilon_{l+1}\Sigma^{-1}(k)\mathbf{X}_n^*\mathbf{X}_n^{*\prime} \right) \right] \end{aligned}$$

Since the vectors  $\mathbf{X}_n^{*'}$  and  $\sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(k)\varepsilon_{j+1}$  are uncorrelated because  $k \leq K_n$  :

$$\begin{aligned} & \mathbb{E} \left( \frac{n - K_n + 1}{k\sigma_\varepsilon^2} f_1^2(k) \right) \\ &= \frac{n - K_n + 1}{k\sigma_\varepsilon^2(n - K_n + 1)^2} \text{trace} \left( \Sigma^{-1}(k)(n - K_n + 1 - \sqrt{n})\sigma_\varepsilon^2 \Sigma(k)\Sigma^{-1}(k)\Sigma^*(k) \right) \\ &= \text{trace} \left( \Sigma^{-1}(k)\Sigma^*(k)k^{-1} \right) (n - K_n + 1 - \sqrt{n})(n - K_n + 1)^{-1}, \end{aligned}$$

where  $\Sigma^*(k)$  is the covariance matrix of the vector  $\mathbf{X}_n^*(k)$ . We note that :

$$(n - K_n + 1 - \sqrt{n})(n - K_n + 1)^{-1} \rightarrow 1 \text{ as } n \rightarrow +\infty.$$

So we only have to study the trace of  $(\Sigma^{-1}(k)\Sigma^*(k)k^{-1})$ . We will use the following inequality : for all  $k \times k$  matrices  $A$  and  $B$

$$\begin{aligned} |\text{trace}(AB)| &\leq \sqrt{\text{trace}(AA')} \sqrt{\text{trace}(BB')} \\ &\leq k \|A\| \|B\|. \end{aligned}$$

We obtain :

$$\begin{aligned} \max_{1 \leq k \leq K_n} \left| \text{trace} \left( \Sigma^{-1}(k)\Sigma^*(k)k^{-1} \right) - 1 \right| &= \max_{1 \leq k \leq K_n} \left| \text{trace} \left( \Sigma^{-1}(k)(\Sigma^*(k) - \Sigma(k))k^{-1} \right) \right| \\ &\leq \max_{1 \leq k \leq K_n} \left\| \Sigma^{-1}(k) \right\| \left\| (\Sigma^*(k) - \Sigma(k)) \right\| \\ &\leq \max_{1 \leq k \leq K_n} \left\| \Sigma^{-1}(k) \right\| \max_{1 \leq k \leq K_n} \left\| (\Sigma^*(k) - \Sigma(k)) \right\| \end{aligned}$$

$\Sigma(k) - \Sigma^*(k)$  is symmetric because  $\Sigma(k)$  and  $\Sigma^*(k)$  are two symmetric matrices, and its spectral norm is lower than every other matrix norm. We use the subordinate norm defined for all matrix  $Y = (y_{i,j})_{1 \leq i,j \leq k}$  by :

$$\|Y\|_1 = \max_j \sum_{i=1}^k |y_{i,j}|.$$

For large  $n$ , we obtain

$$\begin{aligned} \max_{1 \leq k \leq K_n} \left\| \Sigma^{-1}(k) \right\| \max_{1 \leq k \leq K_n} \left\| \Sigma^*(k) - \Sigma(k) \right\| &\leq \max_{1 \leq k \leq K_n} \left\| \Sigma^{-1}(k) \right\| \max_{1 \leq k \leq K_n} \left\| \Sigma(k) - \Sigma^*(k) \right\|_1 \\ &\leq \max_{1 \leq k \leq K_n} \left\| \Sigma^{-1}(k) \right\| \max_{1 \leq k \leq K_n} k \max_{0 \leq j \leq k-1} \sum_{l=\sqrt{n}/2-K_n+1}^{+\infty} |b_l b_{l+j}| \\ &= O \left( \frac{K_n}{(\sqrt{n})^{1-2d-\delta}} \right) \end{aligned}$$

for all  $\delta > 0$ . Then

$$\max_{1 \leq k \leq K_n} \left\| \Sigma^{-1}(k) \right\| \max_{1 \leq k \leq K_n} \left\| \Sigma^*(k) - \Sigma(k) \right\| = o(1)$$

follows from condition (5.39).

**Third step** We consider the cross-product term  $\mathbb{E}(f(k)\mathcal{S}_n(k)L_n^{-1}(k))$  and show that it is negligible. Ing et Wei (2003) proved that :

$$|\mathbb{E}(f(k)\mathcal{S}_n(k)L_n^{-1}(k))| = |\mathbb{E}((f(k) - f_1(k))\mathcal{S}_n(k)L_n^{-1}(k))|.$$

By the Cauchy-Schwarz inequality :

$$\begin{aligned} & \max_{1 \leq k \leq K_n} |\mathbb{E}((f(k) - f_1(k))\mathcal{S}_n(k)L_n^{-1}(k))| \\ & \leq \left[ \max_{1 \leq k \leq K_n} \mathbb{E}((f(k) - f_1(k))^2 L_n^{-1}(k)) \max_{1 \leq k \leq K_n} \mathbb{E}(\mathcal{S}_n^2(k)L_n^{-1}(k)) \right]^{1/2}. \end{aligned}$$

By (5.42), we obtain :

$$\max_{1 \leq k \leq K_n} \mathbb{E}((f(k) - f_1(k))^2 L_n^{-1}(k)) = o(1)$$

and by using the definition in (5.40) of  $L_n(k)$ , we have

$$\max_{1 \leq k \leq K_n} \mathbb{E}(\mathcal{S}_n^2(k)L_n^{-1}(k)) = O(1).$$

Finally we have

$$\max_{1 \leq k \leq K_n} |\mathbb{E}((f(k) - f_1(k))\mathcal{S}_n(k)L_n^{-1}(k))| = o(1).$$

□

In this theorem, we have obtained an asymptotic expression of the mean-squared prediction error of  $\widehat{X}_{n+1}(k)$ , which holds uniformly for all  $1 \leq k \leq K_n$ . In the short memory case i.e. assuming that the process  $(X_t)_{t \in \mathbb{Z}}$  is Gaussian, admits infinite moving average and autoregressive representations defined in (5.6), that the coefficients  $(a_j)_{j \in \mathbb{N}}$  verify  $\sum_{j=1}^{+\infty} \sqrt{j}|a_j| < \infty$  and that the coefficients  $(b_j)_{j \in \mathbb{N}}$  are absolutely summable, Ing et Wei (2003) proved that if  $K_n^{2+\delta} = O(n)$  for some  $\delta > 0$  :

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \left| \frac{\mathbb{E}\left(X_{n+1} - \widehat{X}_{n+1}(k)\right)^2 - \sigma_\varepsilon^2}{L_n(k)} - 1 \right| = 0$$

with  $L_n(k)$  defined as in (5.40).

The term  $L_n(k)$  has the same expression in the short memory case as in the long-memory case. It is the sum of two terms : the first term  $(k/n)\sigma_\varepsilon^2$  is proportional to the order of the model and is a measure of the complexity of the predictor, the second term  $\mathcal{S}_n^2(k)$  corresponds to the goodness of fit of the model. This second term has a different asymptotic behaviour in the short and long-memory case : for short memory time series it decays exponentially fast as a function of  $k$ , whereas for long-memory time series it has a Riemannian decay.

In the following section, we will use the proof of Theorem 5.3.1 to obtain a central limit theorem for our predictor.



## 5.4 Central limit theorem

Like Bhansali (1978) and Lewis et Reinsel (1985) for short memory processes, we look for a normalisation factor to obtain a convergence in distribution of the difference between our predictor  $\widehat{X}_{n+1}(K_n)$  and the Wiener-Kolmogorov predictor  $\widetilde{X}_{n+1} = -\sum_{j=1}^{+\infty} a_j X_{n+1-j}$ , which is the linear least-squares predictor based on all the past.

**Theorem 5.4.1.** *Under assumptions H.1-H.4, if we choose the sequence  $(K_n)_{n \in \mathbb{Z}}$  such that :*

$$K_n^4 = O(n) \quad \text{and} \quad K_n^{1+2d} = o\left(n^{1-2d}\right), \quad (5.44)$$

then

$$\frac{1}{\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}} \left( \widetilde{X}_{n+1} - \widehat{X}_{n+1}(K_n) \right) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 1).$$

*Proof.* The difference between our predictor  $\widehat{X}_{n+1}(K_n)$  and the Wiener-Kolmogorov predictor  $\widetilde{X}_{n+1}$  is equal to :

$$\widetilde{X}_{n+1} - \widehat{X}_{n+1}(K_n) = f(K_n) + \mathcal{S}_n(K_n). \quad (5.45)$$

Since  $(X_t)_{t \in \mathbb{Z}}$  is a Gaussian process with mean 0,  $(\sum_{i=1}^l (a_i - a_{i,K_n}) X_{t+1-i})_{t \in \mathbb{Z}}$  is a Gaussian random variable with mean 0 for any integer  $l$ . But  $(\sum_{i=1}^l (a_i - a_{i,K_n}) X_{t+1-i})_{t \in \mathbb{Z}}$  converges in mean-squared sense and thus in distribution to  $\mathcal{S}_n(K_n)$  as  $l$  tends to infinity. Then  $\mathcal{S}_n(K_n)$  is a Gaussian random variable with mean 0. Consequently it is enough to prove that

$$\frac{1}{\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}} f(K_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

First we search for a bound for  $1/\mathbb{E}[\mathcal{S}_n^2(K_n)]$ . For all integer  $l$ ,

$$\mathbb{E} \left( \sum_{i=1}^l (a_i - a_{i,K_n}) X_{n+1-i} \right)^2 \geq 2\pi \underline{f} \sum_{i=1}^l (a_i - a_{i,K_n})^2$$

because the spectral density  $f$ , which is the Toeplitz symbol of the covariance matrix, is bounded below by a positive constant  $\underline{f}$  (see Grenander et Szegö (1958)). By taking the limit as  $l \rightarrow +\infty$ , we obtain :

$$\begin{aligned} E \left( \sum_{i=1}^{+\infty} (a_i - a_{i,K_n}) X_{j+1-i} \right)^2 &\geq 2\pi \underline{f} \sum_{i=1}^{+\infty} (a_i - a_{i,K_n})^2 \\ &\geq 2\pi \underline{f} \sum_{i=K_n+1}^{+\infty} a_i^2 \end{aligned}$$

since  $a_{i,K_n} = 0$  when  $i > K_n$ . Under assumption H.4,

$$\sum_{i=K_n+1}^{+\infty} a_i^2 \underset{n \rightarrow +\infty}{\sim} \frac{1}{1+2d} K_n^{-2d-1} L^2(K_n)$$

(see Proposition 1.5.10 of Bingham *et al.* (1987)). Then for any  $\delta > 0$ , there exists  $C > 0$  such that :

$$\frac{1}{\mathbb{E}[\mathcal{S}_n^2(K_n)]} \leq C K_n^{2d+1+\delta} \quad (5.46)$$

By introducing  $f_1$  defined in the proof of Theorem 5.3.1, we decompose the proof of the mean-squared convergence in two parts. We will first show that :

$$\frac{1}{\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}} (f(K_n) - f_1(K_n)) \xrightarrow[n \rightarrow +\infty]{L^2} 0 \quad (5.47)$$

then

$$\frac{1}{\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}} f_1(K_n) \xrightarrow[n \rightarrow +\infty]{L^2} 0. \quad (5.48)$$

More precisely we will prove the mean-squared convergence (5.47), using the decomposition in four terms (5.53), (5.54), (5.55) and (5.56) of proof of Lemma 5.3.1 (see appendix). Using (5.57) and (5.46), the term (5.53) verifies for any  $\delta > 0$  :

$$\begin{aligned} & \mathbb{E} \left( \frac{1}{\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}} \mathbf{X}_n^{*\prime}(K_n) \Sigma^{-1}(K_n) \frac{1}{n - K_n + 1} \left[ \sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(K_n) \varepsilon_{j+1} - \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) \varepsilon_{j+1} \right] \right)^2 \\ &= \mathcal{O} \left( \frac{K_n^{3+2d+\delta}}{n^{5/4}} \right). \end{aligned} \quad (5.49)$$

Under assumption (5.44), the mean (5.49) converges to 0.

Similarly for the term (5.54) using (5.60) and (5.46) we obtain for any  $\delta > 0$  :

$$\mathbb{E} \left( \frac{1}{\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}} \mathbf{X}_n^{*\prime}(K_n) \left[ \Sigma^{-1}(K_n) - \widehat{\Sigma}_n^{-1}(K_n) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) \varepsilon_{j+1} \right)^2 = \mathcal{O} \left( \frac{K_n^{5+2d+\delta}}{(n - K_n + 1)^2} \right)$$

which converges to 0 under assumption (5.44) for sufficiently small  $\delta$ .

For the third term (5.55), by (5.65) and (5.46) we obtain :

$$\mathbb{E} \left( \frac{1}{\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}} \left[ \mathbf{X}_n^{*\prime}(K_n) - \mathbf{X}_n^{\prime}(K_n) \right] \widehat{\Sigma}_n^{-1}(K_n) \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) \varepsilon_{j+1} \right)^2 = \mathcal{O} \left( \frac{K_n^{3+2d}}{(n - K_n + 1)^{\frac{3-2d}{2}}} \right)$$

which converges to 0 under assumption (5.44).

Finally the estimation of the fourth term (5.56) is directly given in (5.68) :

$$\begin{aligned} & \mathbb{E} \left( \sqrt{\frac{1}{\mathbb{E}(\mathcal{S}_n(K_n))^2}} \mathbf{X}'_n(K_n) \widehat{\Sigma}_n^{-1}(K_n) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) [\varepsilon_{j+1, K_n} - \varepsilon_{j+1}] \right)^2 \\ &= O \left( \frac{K_n^{1+2d+\delta}}{(n - K_n + 1)^{1-2d-\delta}} \right) \end{aligned}$$

which converges to 0 under condition (5.44).

Now we will prove the mean-squared convergence (5.48).

By (5.43) :

$$\mathbb{E} \left( \frac{n - K_n + 1}{K_n \sigma_\varepsilon^2} f_1^2(K_n) \right) \underset{n \rightarrow +\infty}{\sim} \frac{K_n \sigma_\varepsilon^2}{n - K_n + 1}.$$

Under condition (5.44), bound (5.46) implies :

$$\frac{1}{\mathbb{E}[\mathcal{S}_n^2(K_n)]} \frac{K_n \sigma_\varepsilon^2}{n - K_n + 1} \xrightarrow{n \rightarrow +\infty} 0.$$

Then we have :

$$\lim_{n \rightarrow +\infty} \frac{1}{\mathbb{E}[\mathcal{S}_n^2(K_n)]} \mathbb{E}(f_1^2(K_n)) = 0.$$

□

**Remark 1** The normalisation in Theorem 5.4.1 is not an explicit function of  $K_n$ . Nevertheless we have a good idea of the rate of decay of  $K_n^{-1}$  to 0. We have shown in (5.46) that under assumption H.4 for all  $\delta > 0$  :

$$\exists C, CK_n^{-2d-1-\delta} \leq \mathbb{E}[\mathcal{S}_n^2(K_n)].$$

In Godet (2008)[Theorem 3.1], an upper bound for the rate of convergence is proved assuming H.1-H.2 :

$$\exists C, \mathbb{E}[\mathcal{S}_n^2(K_n)] \leq CK_n^{-1}.$$

For some processes, we even have an equivalent of  $\mathbb{E}[\mathcal{S}_n^2(K_n)]$ . Consider a fractionally integrated noise  $(X_t)_{t \in \mathbb{Z}}$ , which is the stationary solution of the difference equation :

$$(I - B)^d X_t = \varepsilon_t$$

where  $(\varepsilon_t)_{t \in \mathbb{Z}}$  is a white noise with mean 0 and constant finite variance  $\sigma_\varepsilon^2$  and  $B$  is the backward shift operator. In this case, the rate of convergence is given by :

$$\exists C, \mathbb{E}[\mathcal{S}_n^2(K_n)] \sim CK_n^{-1}.$$

**Remark 2** In both the short and the long-memory case, the prediction error between our predictor and the Wiener-Kolmogorov predictor has the same expression  $L_n(K_n) = \frac{K_n}{n} \sigma_\varepsilon^2 + \mathbb{E}[\mathcal{S}_n^2(K_n)]$ . But we do not use the same normalisation for central limit theorems.

In the central limit theorem for short memory processes, we only know results for independent realisation prediction i.e. when the aim is to predict an independent series which has exactly the same probabilistic structure as the observed one. Bhansali (1978) and Lewis et Reinsel (1985) proved a convergence in distribution of  $(\tilde{X}_{n+1} - \hat{X}_{n+1}(K_n))$  normalised by  $\sqrt{\frac{n}{K_n \sigma_\varepsilon^2}}$  respectively in the univariate case and in the multivariate case. This normalisation corresponds to the complexity of the estimation of the projection coefficients.

In the long-memory case the normalisation  $\sqrt{\mathbb{E}[\mathcal{S}_n^2(K_n)]}$  is given by the rate of convergence of the predictor knowing a finite past to the linear least-squares predictor knowing the infinite past. In the long-memory case the rate of convergence due to the projection decays hyperbolically and is the main term of the global error of prediction  $L_n(K_n)$ . On the contrary in the short memory case, the rate of convergence due to the projection decays exponentially fast and is negligible with respect to the rate of convergence due to the estimation of the projection coefficients.

## 5.5 Appendix

### 5.5.1 Preliminary lemmas

In the following lemmas we prove subsidiary asymptotic results, which we need in the proof of Theorem 5.3.1.

**Lemma 5.5.1.** *Assume H.2. If  $q \geq 1$ , then for all  $\delta > 0$ , there exists  $C$  constant such that for all  $1 \leq k \leq K_n$  :*

$$\mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) (\varepsilon_{j+1,k} - \varepsilon_{j+1}) \right\|^q \leq C \left( k(n - K_n + 1)^{2d+\delta} \mathbb{E} (\mathcal{S}_n(k))^2 \right)^{q/2} \quad (5.50)$$

where the norm  $\|\cdot\|$  is defined in (5.9).

*Proof.* We have

$$\frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) (\varepsilon_{j+1,k} - \varepsilon_{j+1}) = \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i}.$$

Without loss of generality, we assume that  $q > 2$  since the result for  $q > 1$  can be

obtained from the result for  $q > 2$  and Jensen's inequality. Observe that :

$$\begin{aligned} & \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right\|^q \\ &= (n - K_n + 1)^{-q/2} \left( \sum_{l=0}^{k-1} \left( \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right)^2 \right)^{q/2} \end{aligned}$$

Since the function  $x \mapsto x^{q/2}$  is convex on  $\mathbb{R}^+$  if  $q > 2$ , we obtain by Jensen's inequality :

$$\left( \sum_{l=0}^{k-1} \left( \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right)^2 \right)^{q/2} \leq k^{-1} \sum_{l=0}^{k-1} k^{q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right|^q.$$

Consequently

$$\begin{aligned} & \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) (\varepsilon_{j+1,k} - \varepsilon_{j+1}) \right\|^q \\ & \leq k^{q/2-1} \sum_{l=0}^{k-1} \mathbb{E} \left( (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right|^q \right). \quad (5.51) \end{aligned}$$

Furthermore

$$\begin{aligned} \mathbb{E} \left( X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right) &= \mathbb{E} \left( X_{j-l} \left[ \varepsilon_{j+1} - X_{j+1} - \sum_{i=1}^k a_{i,k} X_{j+1-i} \right] \right) \quad (5.52) \\ &= -\sigma_{l+1} - \sum_{i=1}^k a_{i,k} \sigma_{l+1-i} \\ &= 0 \end{aligned}$$

for any integer  $l \in [1, k]$  by definition of  $(a_{i,k})_{1 \leq i \leq k}$ . Since the mean defined in (5.52) is equal to 0,

$$\begin{aligned} & \mathbb{E} \left( (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right|^q \right) \\ &= \mathbb{E} \left( (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} - \mathbb{E} \left( X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right) \right|^q \right). \end{aligned}$$

And then by applying Theorem 1 of Ing et Wei (2003) to the random variable

$Q = \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i}$ , we obtain :

$$\begin{aligned} & \mathbb{E} \left( (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right|^q \right) \\ & \leq C \left( \frac{1}{n - K_n + 1} \sum_{s=K_n}^{n-1} \sum_{t=K_n}^{n-1} \sigma(s-t) \sigma^*(s-t) \right)^{q/2}, \end{aligned}$$

where  $\sigma^*(\cdot)$  is the autocovariance function of the process  $(\sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{t+1-i})_{t \in \mathbb{Z}}$  i.e.

$$\sigma^*(s-t) = \mathbb{E} \left[ \left( \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{s+1-i} \right) \left( \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{t+1-i} \right) \right].$$

As  $|\sigma^*(s-t)| \leq \sigma^*(0)$ ,

$$\begin{aligned} & \mathbb{E} \left( (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right|^q \right) \\ & \leq C \left( \frac{1}{n - K_n + 1} \sigma^*(0) \sum_{s=K_n}^{n-1} \sum_{t=K_n}^{n-1} |\sigma(s-t)| \right)^{q/2} \\ & \leq C \left( \frac{1}{n - K_n + 1} \sigma^*(0) \sum_{s=1}^{n-K_n+1} \sum_{t=1}^{n-K_n+1} |\sigma(s-t)| \right)^{q/2}. \end{aligned}$$

Under assumption H.2, we have for any  $\delta > 0$

$$\sum_{s=1}^{n-K_n+1} \sum_{t=1}^{n-K_n+1} |\sigma(s-t)| \leq C(n - K_n + 1)^{2d+\delta+1}.$$

Then we have shown that for all  $\delta > 0$ ,

$$\mathbb{E} \left( (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \right|^q \right) \leq C \left( (n - K_n + 1)^{2d+\delta} \sigma^*(0) \right)^{q/2}.$$

Notice that :

$$\sigma^*(0) = \mathbb{E} (\mathcal{S}_n(k))^2.$$

And this remark allows us to conclude.  $\square$

**Lemma 5.5.2.** *Assume that the assumptions of Theorem 5.3.1 hold. If  $q > 1$ , then for any  $1 \leq k \leq K_n$*

$$\mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^q \leq C k^{q/2}$$

with  $C$  independent of  $n$  and then of  $k$ .

*Proof.* The arguments are similar to those used for verifying Lemma 5.5.1. Without loss of generality we assume that  $q > 2$ , since this result and Jensen's inequality allow to conclude for  $q > 1$ . Reasoning as for (5.51), we have by convexity :

$$\mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^q \leq k^{q/2} k^{-1} \sum_{l=0}^{k-1} \mathbb{E} \left[ (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \varepsilon_{j+1} \right|^q \right].$$

Applying again Theorem 1 of Ing et Wei (2003) :

$$\mathbb{E} \left[ (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \varepsilon_{j+1} \right|^q \right] \leq C \left( \frac{1}{n - K_n + 1} \sum_{s=K_n}^{n-1} \sum_{t=K_n}^{n-1} \sigma(s-t) \sigma_\varepsilon(s-t) \right)^{q/2}$$

where  $\sigma_\varepsilon(\cdot)$  is the autocovariance function of the process  $(\varepsilon_t)_{t \in \mathbb{Z}}$  i.e.

$$\sigma_\varepsilon(s-t) = \mathbb{E}(\varepsilon_t \varepsilon_s) = \begin{cases} 0 & \text{if } s \neq t \\ 1 & \text{otherwise} \end{cases}.$$

We obtain :

$$\begin{aligned} \mathbb{E} \left[ (n - K_n + 1)^{-q/2} \left| \sum_{j=K_n}^{n-1} X_{j-l} \varepsilon_{j+1} \right|^q \right] &\leq C \left( \frac{1}{n - K_n + 1} (n - K_n + 1) \sigma(0) \right)^{q/2} \\ &= O(1). \end{aligned}$$

This concludes the proof.  $\square$

### 5.5.2 Proof of Lemma 5.3.1

We recall that the constant  $C$  may have different values in the different equations but is always independent of  $n$  and of  $k$  since we want a convergence for all  $1 \leq k \leq K_n$ . We decompose  $f(k) - f_1(k)$  into 4 parts, which we estimate separately :

$$\begin{aligned} &f(k) - f_1(k) \\ &= \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{n - K_n + 1} \left( \sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(k) \varepsilon_{j+1} - \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right) \end{aligned} \quad (5.53)$$

$$+ \mathbf{X}_n^{*'}(k) \left( \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \quad (5.54)$$

$$+ \left( \mathbf{X}_n^{*'} - \mathbf{X}_n'(k) \right) \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \quad (5.55)$$

$$+ \mathbf{X}_n'(k) \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) (\varepsilon_{j+1} - \varepsilon_{j+1,k}) \quad (5.56)$$

**Study of the term given in (5.53)** In this part we want to prove the mean-squared convergence to 0 of :

$$\begin{aligned} & \sqrt{\frac{n - K_n + 1}{k}} \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{n - K_n + 1} \left[ \sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(k) \varepsilon_{j+1} - \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right] \\ = & \sqrt{\frac{1}{k}} \mathbf{X}_n^{*'} \Sigma^{-1}(k) \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=n-\sqrt{n}-1}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1}. \end{aligned}$$

Hölder's inequality applied twice with  $1/p + 1/q = 1$  and  $1/p' + 1/q' = 1$  gives :

$$\begin{aligned} & \mathbb{E} \left( \sqrt{\frac{1}{k}} \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=n-\sqrt{n}-1}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ \leq & \left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \right\|^{2q} \right)^{1/q} \left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=n-\sqrt{n}-1}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \\ \leq & \left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} \left( \mathbb{E} \|\Sigma^{-1}(k)\|^{2p'q} \right)^{1/(p'q)} \left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=n-\sqrt{n}-1}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p}. \end{aligned}$$

Under assumption H.3 for all  $p'$  and  $q'$  :

$$\left( \mathbb{E} \|\Sigma^{-1}(k)\|^{2p'q} \right)^{1/(p'q)} = \|\Sigma^{-1}(k)\|^2 = O(1) \quad (5.57)$$

since the spectral density of the process  $(X_t)_{t \in \mathbb{Z}}$  admits a positive lower bound and then the largest eigenvalue of  $\Sigma^{-1}(k)$  is bounded. Furthermore by the convexity of the function  $x \mapsto x^{2q'q}$  and the stationarity of the process  $(\varepsilon_t)_{t \in \mathbb{Z}}$ ,

$$\left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} \leq \left( \mathbb{E} \left[ \sum_{j=0}^{\sqrt{n}/2 - K_n} b_j \varepsilon_{n-j} \right]^{2q'q} \right)^{1/(q'q)}$$

and by Lemma 2 of Wei (1987) :

$$\begin{aligned} \left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} & \leq C' \left( \sum_{j=0}^{\sqrt{n}/2 - K_n} b_j^2 \right) \\ & \leq C \end{aligned} \quad (5.58)$$



because the sequence  $(b_j^2)_{j \in \mathbb{N}}$  is summable.

Finally by Lemma 5.5.2 :

$$\begin{aligned} & \left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=n-\sqrt{n}-1}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \\ & \leq \frac{(n+1)^{1/4}}{\sqrt{n - K_n}} \left( \mathbb{E} \left\| \frac{1}{(n+1)^{1/4}} \sum_{j=n-\sqrt{n}-1}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \\ & \leq C \left( \frac{k}{n^{1/4}} \right) \end{aligned} \quad (5.59)$$

By inequalities (5.57), (5.58) and (5.59) :

$$\mathbb{E} \left( \sqrt{\frac{1}{k}} \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=n-\sqrt{n}-1}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \leq C \left( \frac{k}{n^{1/4}} \right) \leq C \left( \frac{K_n}{n^{1/4}} \right)$$

which converges to 0 as  $n$  tends to infinity under assumption (5.39).

**Study of the term given in (5.54)** Prove that :

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \mathbf{X}_n^{*'}(k) \left[ \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 = 0. \quad (5.60)$$

Applying twice Hölder's inequality, we have :

$$\begin{aligned} & \mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \mathbf{X}_n^{*'}(k) \left[ \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq \left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} \left( \mathbb{E} \left\| \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \\ & \quad \left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p}. \end{aligned}$$

Applying Lemma 5.5.2 we obtain that :

$$\left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \leq Ck \leq CK_n \quad (5.61)$$

since  $k \leq K_n$ . Now we derive the mean-squared convergence to 0 when  $d \in ]0, 1/2[$ .

For  $d \in ]0, 1/4[$ , we apply Theorem 5.2.1 and we get :

$$\left( \mathbb{E} \left\| \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \leq C \left( \frac{K_n^2}{n - K_n + 1} \right). \quad (5.62)$$

Then it follows from (5.58), (5.61) and (5.62) that :

$$\mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \mathbf{X}_n^{*'}(k) \left[ \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \leq C \left( \frac{K_n^3}{n - K_n + 1} \right)$$

which converges to 0 if condition (5.39) holds.

If  $d \in ]1/4, 1/2[$ , we obtain by Theorem 5.2.1 that

$$\left( \mathbb{E} \left\| \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \leq C \left( \frac{K_n^2 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right). \quad (5.63)$$

The inequalities (5.58), (5.61) and (5.63) allow us to conclude that :

$$\mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \mathbf{X}_n^{*'}(k) \left[ \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \leq C \left( \frac{K_n^3 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right)$$

which converges to 0 under assumption (5.39).

For  $d = 1/4$ , applying Theorem (5.2.1) we obtain that

$$\left( \mathbb{E} \left\| \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \leq C \left( \frac{K_n^2 L^2 (n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)} \right). \quad (5.64)$$

The inequalities (5.58), (5.61) and (5.64) allow us to conclude that :

$$\mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \mathbf{X}_n^{*'}(k) \left[ \Sigma^{-1}(k) - \widehat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \leq C \left( \frac{K_n^3 L^2 (n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)} \right)$$

**Study of the term given in (5.55)** Prove that :

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \left[ \mathbf{X}_n^{*'}(k) - \mathbf{X}'_n(k) \right] \widehat{\Sigma}_n^{-1}(k) \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 = 0. \quad (5.65)$$

Using Holder's inequality twice, we have :

$$\begin{aligned} & \mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \left[ \mathbf{X}_n^{*'}(k) - \mathbf{X}'_n(k) \right] \widehat{\Sigma}_n^{-1}(k) \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq \left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \left[ \mathbf{X}_n^{*'}(k) - \mathbf{X}'_n(k) \right] \right\|^{2q'q} \right)^{1/(q'q)} \left( \mathbb{E} \left\| \widehat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \\ & \quad \left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p}. \end{aligned}$$

In view of the convexity of the function  $x \mapsto x^{qq'}$  and of the stationarity of the process  $(\varepsilon_t)_{t \in \mathbb{Z}}$ , we have :

$$\max_{1 \leq k \leq K_n} \left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \left[ \mathbf{X}_n^{*'}(k) - \mathbf{X}'_n(k) \right] \right\|^{2q'q} \right)^{1/(q'q)} \leq \left( \mathbb{E} \left( \sum_{j=\sqrt{n}/2 - K_n + 1}^{+\infty} b_j \varepsilon_{n-j-l} \right)^{2q'q} \right)^{1/(q'q)}.$$

And by Lemma 2 of Wei (1987), we obtain

$$\begin{aligned} \max_{1 \leq k \leq K_n} \left( \mathbb{E} \left\| \frac{1}{\sqrt{k}} \left[ \mathbf{X}_n^{*'}(k) - \mathbf{X}'_n(k) \right] \right\|^{2q'q} \right)^{1/(q'q)} & \leq C \left( \sum_{j=\sqrt{n}/2 - K_n + 1}^{+\infty} b_j^2 \right) \\ & \leq C n^{\frac{2d-1}{2}}. \end{aligned} \quad (5.66)$$

Then by Theorem 5.2.1 :

$$\left( \mathbb{E} \left\| \widehat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \leq C. \quad (5.67)$$

By inequalities (5.61), (5.66) and (5.67), we then obtain :

$$\mathbb{E} \left( \sqrt{\frac{n - K_n + 1}{k}} \left[ \mathbf{X}_n^{*'}(k) - \mathbf{X}'_n(k) \right] \widehat{\Sigma}_n^{-1}(k) \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \leq C \left( K_n n^{\frac{2d-1}{2}} \right)$$

which converges to 0 as  $n$  tends to infinity if condition (5.39) holds.

**Study of the term given in (5.56)** We want to prove that :

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left( \sqrt{\frac{1}{\mathbb{E}(\mathcal{S}_n(k))^2}} \mathbf{X}'_n(k) \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 = 0. \quad (5.68)$$

Using Hölder's inequality twice, we have :

$$\begin{aligned} & \mathbb{E} \left( \sqrt{\frac{1}{\mathbb{E}(\mathcal{S}_n(k))^2}} \mathbf{X}'_n(k) \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 \\ & \leq \frac{1}{(n - K_n + 1) \mathbb{E}(\mathcal{S}_n(k))^2} \left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right\|^{2q'q} \right)^{1/(q'q)} \\ & \quad \left( \mathbb{E} \left\| \widehat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \left( \mathbb{E} \|\mathbf{X}'_n\|^{2p} \right)^{1/p}. \end{aligned}$$

Applying Lemma 5.5.1, we obtain for every  $\delta > 0$  :

$$\left( \mathbb{E} \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right\|^{2q'q} \right)^{1/(q'q)} \leq C \left( k(n - K_n + 1)^{2d+\delta} \mathbb{E}(\mathcal{S}_n(k))^2 \right). \quad (5.69)$$

Finally we choose  $p = 2$  and we have :

$$\begin{aligned} \left( \mathbb{E} \|\mathbf{X}'_n(k)\|^4 \right)^{1/2} &= \sqrt{\mathbb{E} \left( \sum_{j=1}^k X_j^2 \right)^2} \\ &= \sqrt{\left( \sum_{j=1}^k \sigma(j) \right)^2 + 2 \sum_{j=1}^k \sum_{l=1}^k \sigma(j-l)^2} \end{aligned}$$

since the process  $(X_t)_{t \in \mathbb{Z}}$  is Gaussian. Using the assumption H.2 on the covariances, we verify that for all  $\delta > 0$  :

$$\left( \mathbb{E} \|\mathbf{X}'_n(k)\|^4 \right)^{1/2} \leq C \sqrt{k^{4d+\delta}} \leq C \sqrt{k} \quad (5.70)$$

if  $d \in ]0, 1/4[$ . With inequalities (5.67), (5.69), (5.70) and  $1 \leq k \leq K_n$ , we conclude that :

$$\begin{aligned} \forall \delta > 0, \quad & \mathbb{E} \left( \sqrt{\frac{1}{\mathbb{E}(\mathcal{S}_n(k))^2}} \mathbf{X}'_n(k) \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) (\varepsilon_{j+1,k} - \varepsilon_{j+1}) \right)^2 \\ & \leq C \frac{1}{(n - K_n + 1) \mathbb{E}(\mathcal{S}_n(k))^2} \left( \sqrt{k} k(n - K_n + 1)^{2d+\delta} \mathbb{E}(\mathcal{S}_n(k))^2 \right) \\ & \leq C \frac{K_n^{3/2}}{(n - K_n + 1)^{1-2d-\delta}}. \end{aligned}$$

which converges to 0 under condition (5.39).

On the other hand if  $d \in [1/4, 1/2[$ , inequality (5.70) becomes :

$$\forall \delta > 0, \left( \mathbb{E} \|\mathbf{X}'_n(k)\|^4 \right)^{1/2} \leq C k^{4d+\delta} \quad (5.71)$$

Using inequalities (5.67), (5.69) and (5.71), we have for all  $\delta > 0$

$$\begin{aligned} & \mathbb{E} \left( \sqrt{\frac{1}{\mathbb{E}(\mathcal{S}_n(k))^2}} \mathbf{X}'_n(k) \widehat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 \\ & \leq C \left( \frac{K_n^{1+2d+\delta}}{(n - K_n + 1)^{1-2d-\delta}} \right) \end{aligned}$$

which converges to 0 under condition (5.39).

We have proved that for all  $d \in ]0, \frac{1}{2}[$

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \mathbb{E} \left( \sqrt{\frac{1}{L_n(k)}} (f(k) - f_1(k)) \right)^2 = 0.$$

## Chapitre 6

# Comparaison des prédicteurs sur des séries réelles ou simulées

Nous allons sur des simulations numériques<sup>1</sup> tester l'efficacité des deux prédicteurs linéaires que nous avons étudiés dans cette thèse : le prédicteur de Wiener-Kolmogorov tronqué et le prédicteur projeté sur un passé fini. Nous étudierons deux situations : une série simulée et une série réelle.

### 6.1 Trajectoires simulées de processus fractionnaires

#### 6.1.1 Le processus fractionnaire

Le choix des processus fractionnaires s'est fait sur deux critères : ils forment une classe de modélisation très courante des phénomènes à longue mémoire et de plus leurs coefficients caractéristiques sont explicitement calculés (coefficients des développements autorégressif et moyenne mobile, autocovariances voir Brockwell et Davis (1991) chapitre 13.2) ce qui donne accès à certaines méthodes de simulations (voir la partie 6.1.2) et facilite le calcul des différents prédicteurs. En effet pour calculer les coefficients du prédicteur de Wiener-Kolmogorov (tronqué ou non), on a besoin de connaître les coefficients de la représentation autorégressive infinie notés  $a_j$  (voir (1.8)). Pour un processus fractionnaire de paramètre  $d$ , ils sont égaux à

$$a_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}.$$

Pour le prédicteur projeté sur le passé fini  $[X]_1^k$  à l'horizon 1, les coefficients  $a_{j,k}$  peuvent être calculés explicitement. Ils sont donnés par

$$a_{j,k} = -C_k^j \frac{\Gamma(j-d)\Gamma(k-d-j+1)}{\Gamma(-d)\Gamma(k-d+1)}.$$

---

<sup>1</sup>Les simulations numériques ont été réalisées sur la plateforme de calcul financée par l'opération "Modélisation statistique et probabiliste pour aide à la décision" du CPER 2007-2013 projet n°7 STIC et Calculs

### 6.1.2 Méthodes de simulation

Il existe de nombreuses méthodes permettant de simuler les processus fractionnaires gaussiens. Dans Bardet *et al.* (2003), deux méthodes se sont révélées performantes pour simuler un processus fractionnaire : celle par matrice circulante de Wood et Chan (1994) et celle par filtrage d'un bruit blanc lorsqu'on connaît la représentation en moyenne mobile du processus (ce qui est le cas pour un processus fractionnaire). En effet les trajectoires simulées par ces deux méthodes sont acceptées par les tests effectués sur la densité spectrale ou sur l'estimation du paramètre de longue mémoire par l'estimateur de Whittle. Nous avons utilisé celle par filtrage d'un bruit blanc car elle est plus rapide que la méthode utilisant une matrice circulante pour des longues trajectoires.

### 6.1.3 Préviation par le prédicteur de Wiener-Kolmogorov tronqué et le prédicteur projeté sur un passé fini

Dans cette partie nous allons étudier l'erreur quadratique de prévision des prédicteurs que nous avons étudiés dans cette thèse lorsque l'on n'a observé qu'un passé fini et que la fonction de prédiction est estimée.

#### Le prédicteur de Wiener-Kolmogorov tronqué

Le premier prédicteur que nous avons calculé sur des séries simulées est le prédicteur de Wiener-Kolmogorov tronqué dont les coefficients sont estimés grâce à l'estimateur de Whittle sur une série de longueur  $T$  :

$$\widetilde{X}_{T,k_T}(1) = - \sum_{j=1}^{k_T} a_j(\hat{\theta}_T) X_{k_T+1-j}$$

où  $\hat{\theta}_T$  est l'estimateur de Whittle défini en au chapitre 3 équation (3.5). On estime via une procédure de Monte-Carlo

$$\mathbb{E} \left[ \left( X_{k_T+1} - \widetilde{X}_{T,k_T}(1) \right)^2 \right].$$

On simule  $n = 10000$  répliquions indépendantes de la variable aléatoire  $X_{k_T+1} - \widetilde{X}_{T,k_T}(1)$ . Puis on estime sa variance par la variance empirique  $S^2$  pour laquelle l'intervalle de confiance de niveau  $1 - \alpha$  s'écrit :

$$\left[ \frac{(n-1)S^2}{k((n-1), \alpha/2)}, \frac{(n-1)S^2}{k((n-1), 1-\alpha/2)} \right]$$

où  $k(n, \alpha)$  est le quantile supérieure d'ordre  $\alpha$  de la loi du  $\chi^2$  à  $n$  degrés de liberté. On fait varier la longueur  $k$  de la série utilisée pour prédire, la longueur  $T$  de la série utilisée pour estimer et  $d$  le paramètre de longue mémoire. On calcule deux jeux

T	100			1000		
d	0.1	0.25	0.4	0.1	0.25	0.4
k=50	1.02 ±0.03	1.06 ±0.03	1.18 ±0.03	1.03 ±0.03	1.01±0.03	1.02±0.03
k=100	1.03 ±0.03	1.06 ±0.03	1.16±0.03	0.98±0.03	1.04±0.03	1.01±0.03

T	10000		
d	0.1	0.25	0.4
k=50	1.01 ±0.03	1.04±0.03	1.00±0.03
k=100	1.01 ±0.03	1.01±0.03	1.00 ±0.03

TAB. 6.1 – Erreurs quadratiques empiriques pour le prédicteur de Wiener-Kolmogorov tronqué dont les coefficients sont estimés sur une série indépendante par plug-in de l'estimateur de Whittle et leurs intervalles de confiance à 5%

T	100			1000		
d	0.1	0.25	0.4	0.1	0.25	0.4
k=50	1.03±0.03	1.06±0.03	1.15 ±0.03	1.01±0.03	1.06±0.03	1.03±0.03
k=100	1.04±0.03	1.06±0.03	1.15±0.03	1.01±0.03	1.02±0.03	1.04±0.03

T	10000		
d	0.1	0.25	0.4
k=50	1.00±0.03	1.01 ±0.03	1.01±0.03
k=100	1.03 ±0.03	0.99 ±0.03	0.98 ±0.03

TAB. 6.2 – Erreurs quadratiques empiriques pour le prédicteur de Wiener-Kolmogorov tronqué dont les coefficients sont estimés par plug-in de l'estimateur de Whittle calculé sur la même série et leurs intervalles de confiance à 5%

d'expériences en estimant le paramètre du modèle soit sur une série indépendante comme au chapitre 3 (voir le tableau 6.1), ou sur la série que l'on doit prédire (voir tableau 6.2).

En comparant les tableaux 6.1 et 6.2, on remarque que les résultats sont similaires que l'on dispose d'une série indépendante pour estimer les coefficients ou non. Lorsqu'on augmente le nombre  $k$  d'observations utilisées pour prédire, on peut s'attendre à deux effets comme on tronque à un ordre plus élevé l'erreur de prévision théorique doit diminuer mais comme on estime plus de coefficients les erreurs commises sur chaque estimation peuvent se cumuler et augmente l'erreur finale sur le prédicteur (voir le chapitre 3). Ici pour  $k = 50$  et  $k = 100$ , on ne distingue pas de différence ce qui montrerait que le deuxième aspect est négligeable par rapport au premier. Par contre la variance de l'erreur de prédiction est sensible au choix de  $T$  le nombre d'observations utilisées pour estimer la fonction de prédiction. Disposer de  $T = 1000$  observations diminue par exemple beaucoup l'erreur quadratique pour  $d = 0.4$ .

Nous allons maintenant comparer les erreurs quadratiques estimées à deux erreurs de référence : l'erreur minimale et l'erreur maximale réalisée par un prédicteur naïf. L'erreur



de prédiction minimale est réalisée par le prédicteur de Wiener-Kolmogorov et égale à  $\sigma_\varepsilon^2 = 1$ . Le prédicteur naïf est le plus simple que l'on puisse proposer, c'est la moyenne du processus. En particulier, il n'utilise pas la structure de dépendance du processus et c'est donc le prédicteur optimal pour le bruit blanc. Son erreur quadratique de prévision, égale à la variance du processus, dépend de  $d$ . Pour le processus fractionnaire, on obtient les valeurs suivantes

$d$	0.1	0.25	0.4
variance du processus	1.02	1.17	1.78

TAB. 6.3 – Variance d'un processus fractionnaire  $F(d)$

Pour  $d = 0.1$ , nous ne pouvons pas juger de l'efficacité du prédicteur de Wiener-Kolmogorov tronqué puisque nos intervalles de confiances à 5 % contiennent à la fois l'erreur de prévision minimale et l'erreur du prédicteur naïf. Pour  $d = 0.25$  et  $d = 0.4$  même pour de faibles valeurs de  $k$  nombres d'observations utilisées pour prédire et  $T$  celui pour estimer les coefficients (comme  $k = 50$  et  $T = 100$ ) le prédicteur de Wiener-Kolmogorov tronqué est plus efficace que la moyenne du processus. Dès que  $T = 1000$  l'intervalle de confiance de la variance du prédicteur contient la variance minimale du prédicteur de Wiener-Kolmogorov. Le prédicteur tronqué en est donc indistinguishable.

### Le prédicteur projeté

Pour estimer l'erreur quadratique de prévision pour le prédicteur projeté à l'horizon 1 étudié au chapitre 5 :

$$\widehat{X}_{T+1}(k) = \sum_{i=1}^k \widehat{a}_{i,k} X_{T+1-i},$$

nous avons simulé 10000 prédicteurs indépendants d'un processus fractionnaire et nous avons calculé la variance empirique. Nous avons de plus précisé l'intervalle de confiance à 95% des résultats obtenus. Plus précisément nous avons effectué deux jeux de simulations : on a tout d'abord estimé les coefficients du prédicteur sur une série indépendante de celle que l'on doit prédire comme on l'a supposé au chapitre 4 (voir tableau 6.4) puis on a estimé sur la même série que celle utilisée pour prédire comme au chapitre 5 (voir tableau 6.5). Dans les deux cas, on a fait varier trois paramètres : le paramètre de longue mémoire  $d$ , le nombre de données utilisées pour prévoir  $k$  et le nombre de données utilisées pour estimer  $T$ .

Commençons par quelques remarques communes aux deux jeux de simulation dont les résultats sont résumés dans les tableaux 6.4 et 6.5. Lorsque le paramètre de longue mémoire  $d$  augmente, les coefficients estimés du prédicteur convergent moins rapidement vers les vrais coefficients comme nous l'avons rappelé au chapitre 4 (voir Yajima (1993)). Cette dégradation de l'estimation des coefficients et de la prédiction se voit bien lorsque  $T$  est égal à 100 (voir les tableaux 6.4 et 6.5). Lorsque  $k$  le nombre de données utilisées pour prévoir augmente à  $T$  fixé, l'erreur de prévision

augmente puisque plus le nombre de coefficients à estimer est important plus les erreurs peuvent se cumuler. Cela observe notamment dans les colonnes  $T = 1000$ . Enfin plus on dispose de données pour estimer les coefficients, meilleure est la prédiction.

T	100			1000		
d	0.1	0.25	0.4	0.1	0.25	0.4
k=50	1.48 $\pm$ 0.04	1.58 $\pm$ 0.04	1.75 $\pm$ 0.04	1.05 $\pm$ 0.03	1.05 $\pm$ 0.03	1.07 $\pm$ 0.03
k=100				1.08 $\pm$ 0.03	1.10 $\pm$ 0.03	1.11 $\pm$ 0.03

T	10000		
d	0.1	0.25	0.4
k=50	1.00 $\pm$ 0.03	1.02 $\pm$ 0.03	1.02 $\pm$ 0.03
k=100	1.02 $\pm$ 0.03	1.03 $\pm$ 0.03	1.01 $\pm$ 0.03

TAB. 6.4 – Erreurs quadratiques empiriques pour le prédicteur projeté dont les coefficients sont estimés à partir des covariances empiriques du processus calculées sur une série indépendante et intervalle de confiance à 5% <sup>2</sup>

T	100			1000		
d	0.1	0.25	0.4	0.1	0.25	0.4
k=50	1.20 $\pm$ 0.03	1.32 $\pm$ 0.04	1.54 $\pm$ 0.04	1.04 $\pm$ 0.03	1.05 $\pm$ 0.03	1.07 $\pm$ 0.03
k=100				1.06 $\pm$ 0.03	1.09 $\pm$ 0.03	1.10 $\pm$ 0.03

T	10000		
d	0.1	0.25	0.4
k=50	1.00 $\pm$ 0.03	1.01 $\pm$ 0.03	1.02 $\pm$ 0.03
k=100	1.02 $\pm$ 0.03	1.02 $\pm$ 0.03	1.02 $\pm$ 0.03

TAB. 6.5 – Erreur quadratique empirique pour le prédicteur projeté dont les coefficients sont estimés à partir des covariances empiriques du processus calculées sur la même série et intervalle de confiance à 5% <sup>2</sup>

Ensuite nous allons comparer les deux jeux de simulations selon que l'on estime les coefficients du prédicteur sur la même série ou sur une série indépendante. La comparaison est en faveur de l'estimation sur la même série dans tous les cas que nous avons testés. Dans nos résultats, rien ne laissait envisager ce résultat empirique. En courte mémoire, Ing et Wei (2003) avaient même prouvé que l'erreur due à l'estimation des coefficients avaient le même équivalent asymptotique pour une estimation sur la même série avec une estimation sur une série indépendante (voir Bhansali (1978)).

Comparons maintenant les résultats obtenus à l'erreur de prévision minimale qui est égale à la variance du bruit blanc ici égale à 1 et à la variance du processus qui est

<sup>2</sup>Les cases grisées correspondent à des cas où le prédicteur n'est pas calculable.

l'erreur commise par le prédicteur naïf que constitue la moyenne du processus (voir le tableau 6.3). Que ce soit pour des trajectoires indépendantes ou pour une seule trajectoire, on converge assez vite vers l'erreur minimale qui vaut 1 si  $T$  est suffisamment grand ( $T = 10000$  convient ici). Si  $d = 0.1$  c'est à dire si la longue mémoire du processus est faible, la variance du processus est égale à 1.02 et la qualité du prédicteur ne peut être mesurée grâce aux simulations que nous avons effectuées puisque nos intervalles de confiance avec 10000 simulations sont trop longs. Par contre pour  $d = 0.25$  et  $d = 0.4$ , la variance du processus fractionnaire est respectivement de 1.17 et 1.78, le prédicteur projeté s'avère efficace dès que  $T$  le nombre d'observations pour estimer est suffisamment grand. En effet à partir de  $T = 1000$ , on dispose de suffisamment de données pour bien estimer les coefficients du prédicteur et donc avoir un prédicteur efficace. Si  $T = 100$ , si  $d = 0.4$  et si on estime les coefficients du prédicteur sur la série que l'on veut prédire, le prédicteur projeté est légèrement plus efficace que le prédicteur naïf mais généralement  $T = 100$  observations ne suffisent pas pour estimer les coefficients du prédicteur projeté.

### Comparaison des prédicteurs de Wiener-Kolmogorov tronqué et projeté

Au chapitre 2, nous avons montré que lorsqu'on connaît la loi du processus le prédicteur projeté sur les  $k$  dernières observations est plus efficace que le prédicteur de Wiener-Kolmogorov tronqué et cela par définition. De plus pour un processus fractionnaire, nous avons montré que l'écart entre les erreurs commises par les deux prédicteurs croît avec le paramètre de longue mémoire  $d$ . Il atteint par exemple 60% pour  $d = 0.4$  (voir le graphique 2.2).

Par contre lorsqu'on ne connaît pas la loi du processus et donc qu'on estime les coefficients des prédicteurs, le prédicteur de Wiener-Kolmogorov tronqué est plus efficace que le prédicteur projeté sauf lorsqu'on dispose de  $T = 10000$  observations pour estimer les coefficients où les résultats sont similaires. En effet l'estimation des coefficients du prédicteur de Wiener-Kolmogorov tronqué se fait sous une hypothèse paramétrique alors que pour le prédicteur projeté, on utilise une méthode non-paramétrique. La convergence des estimateurs de la fonction de prédiction est plus lente dans le cadre non-paramétrique si le paramètre de longue mémoire  $d$  est supérieur à  $1/4$  (voir Yajima (1993)).

## 6.2 Étude du niveau du lac majeur

### 6.2.1 Les données

Nous allons nous intéresser à la série correspondant à la quantité d'eau arrivée quotidiennement dans le lac Majeur entre le 1er janvier 1943 et le 30 avril 1994. Cette série de longueur 18748 a été étudiée dans Montanari *et al.* (1997).

Cette série hydrologique contient une partie périodique due à la saisonnalité des précipitations. Nous allons commencer par estimer puis soustraire la composante périodique annuelle des données par la méthode STL proposée par Cleveland *et al.* (1990) (voir le graphique 6.1). Sur le troisième graphique du graphe 6.1, on observe une tendance et donc la série ne peut-être considérée comme stationnaire mais à plus court

terme on peut envisager une telle modélisation. Pour détecter la longue mémoire dans la série désaisonnalisée, plusieurs méthodes sont proposées par Montanari *et al.* (1997) comme la statistique R/S, le tracé de la fonction d'autocovariance ou le tracé en échelle log-log du périodogramme.

Nous allons commencer par la statistique R/S, introduite par Hurst (1951) et dont les propriétés asymptotiques ont été démontrées par Mandelbrot, qui consiste à regarder le processus à une échelle réduite. On définit  $Y(n, t) = \sum_{j=1}^t (X_j - \bar{X}_n)$  avec  $\bar{X}_n$  la moyenne empirique calculées sur les  $n$  premières valeurs,  $R(n) = \max_{1 \leq t \leq n} Y(n, t) - \min_{1 \leq t \leq n} Y(n, t)$  et  $S(n) = \sqrt{\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2}$ . Si le processus  $(X_n)_{n \in \mathbb{Z}}$  est à longue mémoire, il existe  $H$  tel qu' on a la convergence en loi suivante :

$$n^{-H} \frac{R(n)}{S(n)} \rightarrow Q$$

où  $H$  est le paramètre de Hurst ( $H = d + 1/2$ ). En traçant cette statistique, en échelle log-log, on estime  $H$  et donc  $d$ .

On sait aussi que la moyenne empirique a une variance équivalente à  $Cn^{2d-1}$  si elle est calculée sur un processus à longue mémoire de paramètre  $d$  ( voir Hosking (1996) ). En traçant la variance des moyennes empiriques en échelle log-log sur le graphique, on peut ainsi estimer  $d$  par 0.105.

La densité spectrale  $f$  d'un processus à longue mémoire définie comme en (1.2) vérifie si on prend son logarithme :

$$\log(f(\lambda)) \sim -2d \log(\lambda) \quad \text{lorsque } \lambda \rightarrow 0.$$

Donc si les données sont la réalisation d'un processus à longue mémoire, le périodogramme, qui est un estimateur de la densité spectrale, tracé en échelle log-log doit se comporter comme une application affine de coefficient directeur négatif lorsque le logarithme des fréquences tend vers l'infini. Or le graphique 6.2 pour le périodogramme de la série désaisonnalisée est compatible avec cette propriété. Sur le graphique 6.2, on observe la décroissance lente de la fonction d'autocovariance du processus comme en longue mémoire. En longue mémoire, les autocovariances du processus vérifient :

$$\log(\gamma(k)) \sim (2d - 1) \log(k).$$

Sur les autocovariances de notre série tracées en échelle log-log sur le graphique 6.2, on observe cette relation linéaire qui nous permet d'estimer de plus  $d$ .

### 6.2.2 Prévion

### 6.2.3 Prévion par le prédicteur tronqué

Pour prévoir par le prédicteur de Wiener-Kolmogorov (tronqué), il faut connaître la représentation autorégressive du processus et donc la loi du processus. Montanari *et al.* (1997) proposent de modéliser la série par un modèle FARIMA(1, $d$ ,1). Pour estimer les trois paramètres du modèle, ils utilisent l'estimateur de Whittle comme nous l'avons fait

au chapitre 3. Si on utilise la méthode STL de Cleveland *et al.* (1990) pour isoler la partie saisonnière, l'estimateur de Whittle nous donne le modèle FARIMA suivant pour le résidu :

$$(I - 0.11B)(I - B)^{0.38}X_t = (I + 0.23B)\varepsilon_t. \quad (6.1)$$

Pour tester si ce modèle s'ajuste bien à nos données, on a utilisé le test de Fay et Philippe (2002) qui utilise une fonctionnelle non linéaire du périodogramme. Celui-ci accepte le modèle.

On peut alors calculer le développement autorégressif du processus solution stationnaire de l'équation (6.1) et donc calculer le prédicteur de Wiener-Kolmogorov tronqué. On va tronquer le développement à l'ordre  $k = 80$ . Si on calcule tous les prédicteurs à l'horizon 1 sur cette série pour tout les points dont on connaît le passé sur 80 observations, on obtient une variance empirique de tous ces prédicteurs égale à 3580. Les observations ne sont pas indépendantes donc on n'obtient pas une très bonne estimation de la variance du prédicteur mais elle nous donne tout de même une idée de sa qualité. On peut par exemple la comparer à la variance du bruit blanc  $(\varepsilon_n)_{n \in \mathbb{Z}}$  sous-jacent qui est l'erreur quadratique de prévision du prédicteur de Wiener-Kolmogorov c'est à dire l'erreur minimale pour un prédicteur linéaire. On estime la variance du bruit blanc par  $\hat{\sigma}_\varepsilon^2$  en utilisant la formule

$$\sigma_\varepsilon^2 = \exp \left( \int_{-\pi}^{\pi} \log(f(\lambda)) d\lambda / 2\pi \right)$$

et en y remplaçant la densité spectrale  $f$  par le périodogramme. On obtient  $\hat{\sigma}_\varepsilon^2 = 2960$ . L'autre valeur de référence est l'erreur de prévision du prédicteur naïf qu'est la moyenne du processus. Elle est sur cette série estimée à 77500. Le prédicteur de Wiener-Kolmogorov tronqué est donc assez proche du prédicteur de Wiener-Kolmogorov ce qui montre son efficacité.

#### 6.2.4 Prévision par le prédicteur projeté

Pour prédire la série désaisonnalisée, on va maintenant utiliser le prédicteur projeté sur un passé fini de longueur  $k$  du chapitre 5 dont les coefficients sont estimés sur la même série. On a choisi de prédire en projetant sur les  $k = 80$  dernières valeurs en estimant les covariances empiriques sur l'ensemble de la série ( 18748 observations). Sur le graphique 6.3, on a tracé les 80 dernières prévisions. Comme on dispose de beaucoup d'observations pour estimer la fonction de prévision, on peut considérer que l'erreur principale est due à la projection et non à l'estimation c'est-à-dire que l'on se trouve sous la condition (5.44) du théorème 5.4.1. On applique alors le théorème 5.4.1 pour obtenir un intervalle de confiance asymptotique. On va de plus supposer que nos observations sont comme dans la partie précédente la réalisation d'un processus FARIMA. Pour les processus FARIMA, Inoue et Kasahara (2006) a prouvé que l'erreur quadratique due à la projection est équivalente à  $\frac{\sigma_\varepsilon^2 d^2}{k}$ . En supposant que le processus est gaussien, on

obtient l'intervalle de confiance asymptotique suivant :

$$\mathbb{P} \left( \widehat{X}_{k+1} - P_{[X]_1^k}(X_{k+1}) \in \left[ -q \left( \frac{\beta}{2} \right) \sqrt{\frac{\widehat{\sigma}_\varepsilon^2 d^2}{k}}, q \left( \frac{\beta}{2} \right) \sqrt{\frac{\widehat{\sigma}_\varepsilon^2 d^2}{k}} \right] \right) = 1 - \beta$$

où  $\widehat{X}_{k+1}$  est le prédicteur de Wiener-Kolmogorov et  $q(\beta)$  est le quantile d'ordre  $\beta$  de  $\mathcal{N}(0, 1)$ . Pour nos  $k$  dernières observations, on a donc tracé en pointillé l'intervalle asymptotique de confiance à 5% ( voir le graphique 6.3 ). Sur l'ensemble de la série, on trouve que 90% des valeurs prises par la série sont dans l'intervalle de confiance que nous avons calculé.

Si on calcule la variance de tous les prédicteur à l'horizon 1 que l'on peut regarder sur cette série. Pour le prédicteur projeté, on trouve une variance 3265. La variance estimée du bruit blanc est de 2960 et celle du processus de 77500. Le prédicteur projeté sur un passé fini semble donc efficace. Sa variance estimée est inférieure à celle du prédicteur de Wiener-Kolmogorov tronqué. Comme on dispose de beaucoup d'observations ( $T = 18748$ ), la convergence des coefficients estimés du prédicteur projeté est bonne. Les résultats des prédicteurs tronqué et projeté devraient donc être similaires si on étudiait des séries simulées comme dans la partie 6.1 Mais ici, nos données ne sont pas simulées selon le modèle FARIMA(1,d,1) que nous avons ajusté et le prédicteur projeté non-paramétrique reprend alors l'avantage sur le tronqué paramétrique.

FIG. 6.1 – Extraction de la composante périodique des données : le premier graphique correspond au tracé des données, le deuxième à l'estimation de la composante périodique par la méthode STL, la somme des troisième et quatrième à la différence des deux premiers.

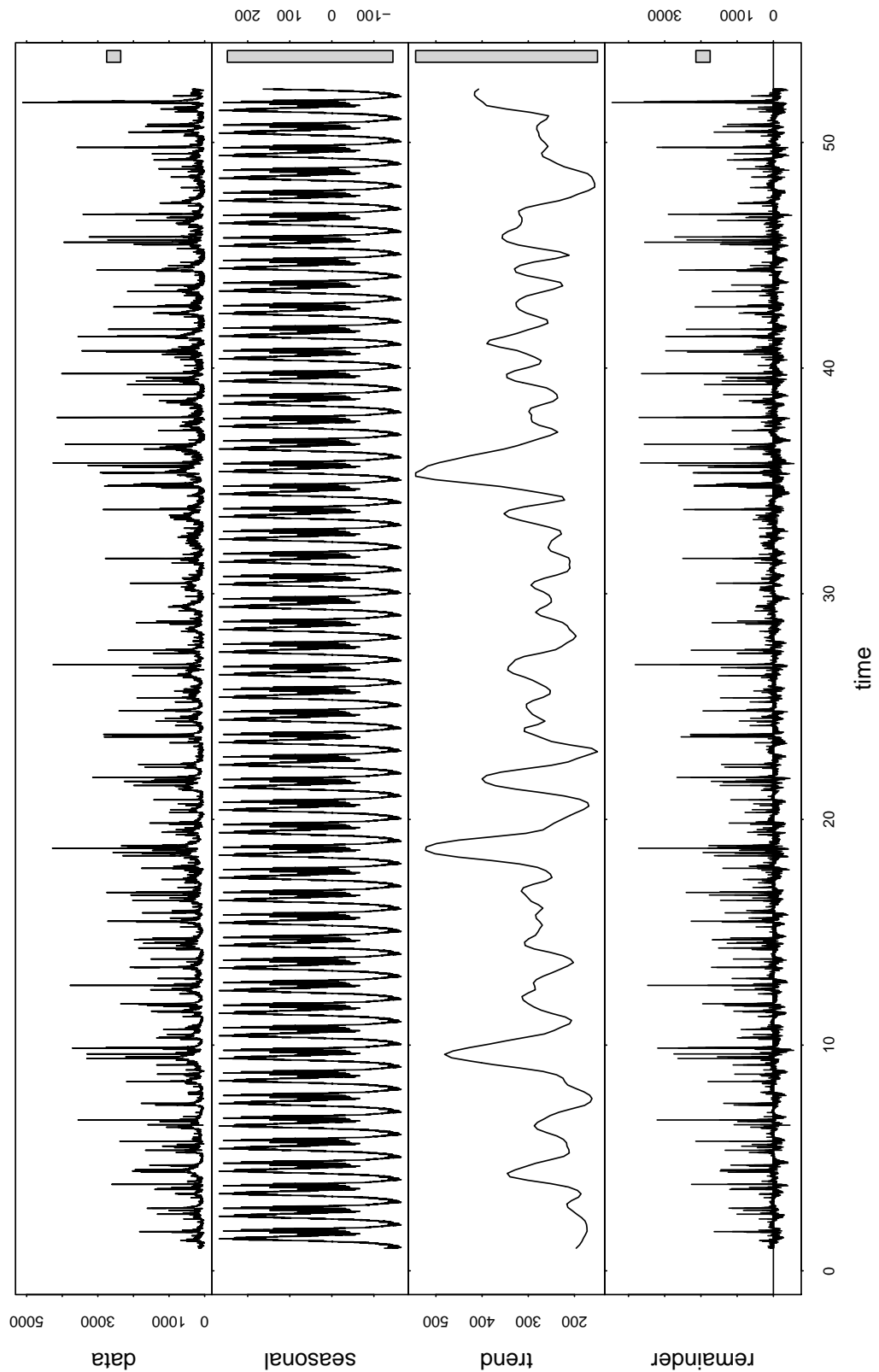


FIG. 6.2 – Statistiques pour tester la longue mémoire appliquées à la série désaisonnalisée (de droite à gauche et de haut en bas) : la statistique R/S, le périodogramme tracé en échelle log-log, la variance de la moyenne empirique du processus, les autocovariances empiriques et les autocovariances empiriques tracées en échelle log-log.

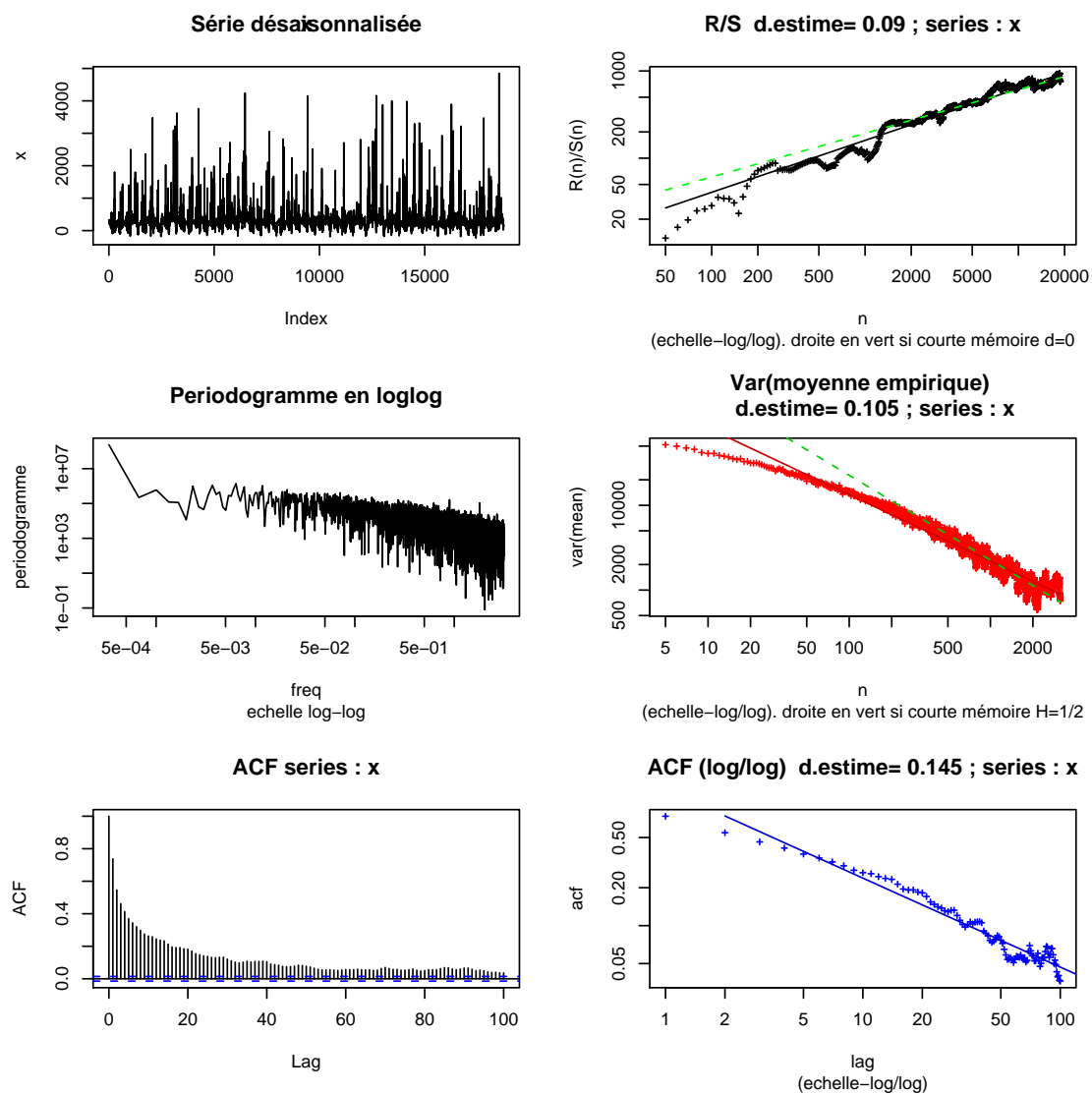
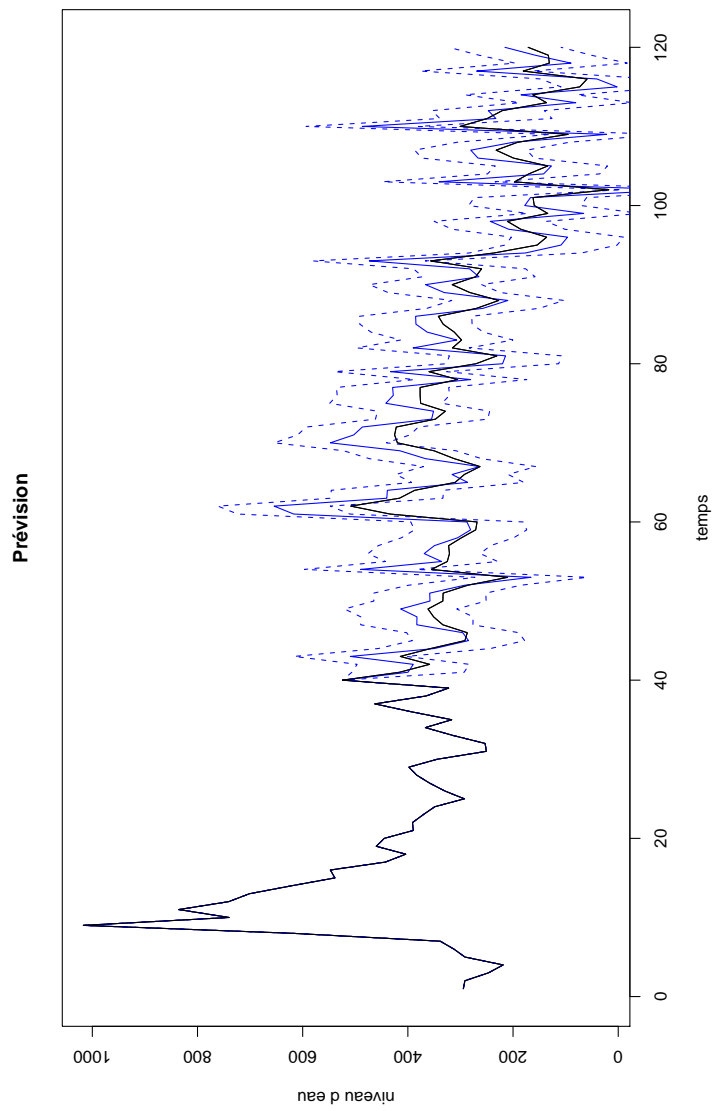




FIG. 6.3 – Prévisions à l’horizon 1 par le prédicteur projeté sur les  $k = 80$  dernières valeurs observées et intervalles de confiance asymptotiques de niveau 5% associés



# Bibliographie

- Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Stat.*, 2 :1095–1107.
- Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Stat. Math.*, 22 :203–217.
- Baillie, R. (1996). Long memory processes and fractional integration in econometrics. *J. Econometrics*, 73(1) :5–59.
- Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., et Taqqu, M. (2003). Generators of long-range dependent processes : a survey. In *Theory and applications of long-range dependence*, pages 579–623. Birkhäuser Boston, Boston, MA.
- Barkoulas, J. et Baum, C. (2006). Long-memory forecasting of US monetary indices. *J. Forecast.*, 25(4) :291–302.
- Baxter, G. (1962). An asymptotic result for the finite predictor. *Math. Scand.*, 10 :137–144.
- Beran, J. (1994). *Statistics for long-memory processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.
- Beran, J., Bhansali, R., et Ocker, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika*, 85(4) :921–934.
- Berk, K. (1974). Consistent autoregressive spectral estimates. *Ann. Stat.*, 2 :489–502.
- Bhansali, R. (1974). Asymptotic properties of the Wiener-Kolmogorov predictor. I. *J. Roy. Statist. Soc. Ser. B*, 36 :61–73.
- Bhansali, R. (1977). Asymptotic properties of the Wiener-Kolmogorov predictor. II. *J. Roy. Statist. Soc. Ser. B*, 39(1) :66–72.
- Bhansali, R. (1978). Linear prediction by autoregressive model fitting in the time domain. *Ann. Stat.*, 6 :224–231.

- Bhansali, R., Giraitis, L., et Kokoszka, P. (2007). Approximations and limit theory for quadratic forms of linear processes. *Stochastic Process. Appl.*, 117(1) :71–95.
- Bhansali, R. et Kokoszka, P. (2001). Prediction of long-memory time series : An overview. *Estadística 53, No.160-161*, 41-96.
- Bhansali, R. J., Giraitis, L., et Kokoszka, P. S. (2006). Estimation of the memory parameter by fitting fractionally differenced autoregressive models. *J. Multivariate Anal.*, 97(10) :2101–2130.
- Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, second edition.
- Bingham, N., Goldie, C., et Teugels, J. (1987). *Regular variation*. Encyclopedia of Mathematics and its applications, Vol. 27. Cambridge etc. : Cambridge University Press.
- Boettcher, A. et Virtanen, J. (2006). Norms of Toeplitz Matrices with Fisher-Hartwig Symbols. Technical report, arXiv :math.FA/0606016.
- Bondon, P. (2001). Recursive relations for multistep prediction of a stationary time series. *J. Time Ser. Anal.*, 22(4) :399–410.
- Bondon, P. (2005). Influence of missing values on the prediction of a stationary time series. *J. Time Ser. Anal.*, 26(4) :519–525.
- Brockwell, P. et Davis, R. (1988). Simple consistent estimation of the coefficients of a linear filter. *Stochastic Processes and their Applications*.
- Brockwell, P. et Davis, R. (1991). *Time Series : Theory and Methods*. Springer Series in Statistics.
- Ciarlet, P. (1982). *Introduction à l'analyse numérique matricielle et à l'optimisation*. Collection Mathématiques Appliquées pour la Maitrise. Paris.
- Cleveland, R., Cleveland, W., et McRae, J. (1990). Stl : A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6 :3–73.
- Crato, N. et Ray, B. (1996). Model selection and forecasting for long-range dependent processes. *J. Forecast.*, 15(2) :107–125.
- Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*.
- Deo, R., Hsieh, M., et Hurvich, C. (2007). Long memory in intertrade durations, counts and realized volatility of nyse stocks. *Preprint*.

- Doornik, J. et Ooms, M. (2004). Inference and forecasting for ARFIMA models with an application to US and UK inflation. *Stud. Nonlinear Dyn. Econom.*, 8(2).
- Fay, G. et Philippe, A. (2002). Goodness-of-fit test for long range dependent processes. *ESAIM Probab. Statist.*, 6 :239–258 (electronic). New directions in time series analysis (Luminy, 2001).
- Fox, R. et Taqqu, M. S. (1985). Noncentral limit theorems for quadratic forms in random variables having long-range dependence. *Ann. Probab.*, 13 :428–446.
- Fox, R. et Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Stat.*, 14 :517–532.
- Fox, R. et 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. (1976). *Introduction to statistical time series*. Wiley Series in Probability and Mathematical Statistics. New York.
- Geweke, J. et Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.*, 4(4) :221–238.
- Giraitis, L., Kokoszka, P., et Leipus, R. (2000). Stationary ARCH models : dependence structure and central limit theorem. *Econometric Theory*, 16(1) :3–22.
- Giraitis, L. et Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Relat. Fields*, 86(1) :87–104.
- Godet, F. (2007a). Linear Prediction of Long-Range Dependent Time Series. <http://arxiv.org/abs/math/0702485>.
- Godet, F. (2007b). Prediction of long memory processes on same-realisation. <http://arxiv.org/abs/0712.1922>.
- Godet, F. (2008). Linear Prediction of Long-Range Dependent Time Series. *ESAIM Probab. Statist.*
- Gray, H., Zhang, N.-F., et Woodward, W. (1989). On generalized fractional processes. *J. Time Ser. Anal.*, 10(3) :233–257.
- Grenander, U. et Szegő, G. (1958). *Toeplitz forms and their applications*. California Monographs in Mathematical Sciences. University of California Press, Berkeley.
- Hidalgo, J. et Soulier, P. (2004). Estimation of the location and exponent of the spectral singularity of a long memory process. *J. Time Ser. Anal.*, 25(1) :55–81.
- Hidalgo, J. et Yajima, Y. (2002). Prediction and signal extraction of strongly dependent processes in the frequency domain. *Econometric Theory*, 18(3) :584–624.

- Hosking, J. (1981). Fractional differencing. *Biometrika*, 68 :165–176.
- Hosking, J. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econom.*, 73(1) :261–284.
- Hosking, J. R. M. (1984). Modeling Persistence In Hydrological Time Series Using Fractional Differencing. *Water Resources Research*, 20 :1898–1908.
- Hurst, H. (1951). Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Eng.*, 116 :770–799.
- Ing, C.-K. et Wei, C.-Z. (2003). On same-realization prediction in an infinite-order autoregressive process. *J. Multivariate Anal.*, 85(1) :130–155.
- 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. et Kasahara, Y. (2006). Explicit representation of finite predictor coefficients and its applications. *Ann. Stat.*, 34(2) :973–993.
- Kolmogorov, A. (1941). Stationary sequences in Hilbert’s space. *Bull. Mosk. Gos. Univ. Mat.*, 2(6) :1–40.
- Kunitomo, N. et Yamamoto, T. (1985). Properties of predictors in misspecified autoregressive time series models. *J. Amer. Statist. Assoc.*, 80(392) :941–950.
- Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the 1st World Congress of the Bernoulli Society, Vol. 1 (Tashkent, 1986)*, pages 67–74, Utrecht. VNU Sci. Press.
- Lewis, R. et Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.*, 16 :393–411.
- Major, P. (1981). *Multiple Wiener-Itô integrals*, volume 849 of *Lecture Notes in Mathematics*. Springer, Berlin. With applications to limit theorems.
- Montanari, A., Rosso, R., et Taqqu, M. (1997). Fractionally differenced arima models applied to hydrologic time series :identification, estimation, and simulation. *Water resources research*, 33(5) :1035–1044.
- Moulines, E. et Soulier, P. (1999). Broadband log-periodogram regression of time series with long-range dependence. *Ann. Statist.*, 27(4) :1415–1439.
- Pourahmadi, M. (1989). On the convergence of finite linear predictors of stationary processes. *J. Multivariate Anal.*, 30(2) :167–180.

- Pourahmadi, M. (2001). *Foundations of time series analysis and prediction theory*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. Wiley-Interscience, New York.
- Ray, B. (1993a). Model selection and forecasting for long-range dependent processes. *International Journal of Forecasting*, 9(2) :255–269.
- Ray, B. (1993b). Modeling long-memory processes for optimal long-range prediction. *J. Time Ser. Anal.*, 14(5) :511–525.
- Rosenblatt, M. (1979). Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 49 :125–132.
- Simon, B. (2005). *Orthogonal polynomials on the unit circle. Part 1*, volume 54 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI. Classical theory.
- Taqqu, M. (2003). Fractional Brownian motion and long-range dependence. In *Theory and applications of long-range dependence*, pages 5–38. Birkhäuser Boston, Boston, MA.
- Tiao, G. et Tsay, R. (1994). Some advances in non-linear and adaptive modelling in time-series. *J. Forecasting*, 13(2) :109–131.
- Triantafyllopoulos, K. (2003). On the central moments of the multidimensional Gaussian distribution. *Math. Sci.*, 28(2) :125–128.
- Viano, M.-C., Deniau, C., et Oppenheim, G. (1995). Long-range dependence and mixing for discrete time fractional processes. *J. Time Ser. Anal.*, 16(3) :323–338.
- Wei, C.-Z. (1987). Adaptive prediction by least-squares predictors in stochastic regression models with applications to time series. *Ann. Stat.*, 15(4) :1667–1682.
- Whittle, P. (1983). *Prediction and regulation by linear least-square methods*. University of Minnesota Press, Minneapolis, MN, second edition. With a foreword by Thomas J. Sargent.
- Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications*. The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass.
- Willinger, W., Taqqu, M., Sherman, R., et Wilson, D. (1997). Self-similarity through high-variability : Statistical analysis of ethernet lan traffic at the spurce level. *Transactions on Networking*.
- Wood, A. et Chan, G. (1994). Simulation of stationary Gaussian processes in  $[0, 1]^d$ . *J. Comput. Graph. Statist.*, 3(4) :409–432.

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.

Yajima, Y. (1996). Estimation of the frequency of unbounded spectral density. In *Proceedings of the Business and Economic Statistical Section*. American Statistical Association.

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

**Résumé :** Nous étudions des méthodes de prévision pour les processus à longue mémoire. Ils sont supposés stationnaires du second ordre, linéaires, causals et inversibles. Nous supposons tout d'abord que l'on connaît la loi du processus mais que l'on ne dispose que d'un nombre fini d'observations pour le prédire. Nous proposons alors deux prédicteurs linéaires : celui de Wiener-Kolmogorov tronqué et celui construit par projection sur le passé fini observé. Nous étudions leur comportement lorsque le nombre d'observations disponibles tend vers l'infini. Dans un deuxième temps nous ne supposons plus la loi du processus connue, il nous faut alors estimer les fonctions de prévision obtenues dans la première partie. Pour le prédicteur de Wiener-Kolmogorov tronqué, nous utilisons une approche paramétrique en estimant les coefficients du prédicteur grâce à l'estimateur de Whittle calculé sur une série indépendante de la série à prédire. Pour le prédicteur obtenu par projection, on estime les coefficients du prédicteur en remplaçant dans les équations de Yule-Walker les covariances par les covariances empiriques calculé sur une série indépendante ou sur la série à prédire. Pour les deux prédicteurs, on estime les erreurs quadratiques due à l'estimation des coefficients et on prouve leurs normalités asymptotiques.

**Mots clés :** processus à longue mémoire, modèle linéaire, erreur quadratique de prévision, normalité asymptotique, intervalle de prévision.

**Summary :** This PhD thesis deals with predicting long-memory processes. We assume that the processes are weakly stationary, linear, causal and invertible, but only a finite subset of the past observations is available. We first present two approaches when the stochastic structure of the process is known : one is the truncation of the Wiener-Kolmogorov predictor, and the other is the projection of the forecast value on the observations, i.e. the least-squares predictor. We show that both predictors converge to the Wiener-Kolmogorov predictor. When the stochastic structure is not known, we have to estimate the coefficients of the predictors defined in the first part. For the truncated Wiener-Kolmogorov, we use a parametric approach and we plug in the forecast coefficients from the Whittle estimator, which is computed on an independent realisation of the series. For the least-squares predictor, we plug the empirical autocovariances (computed on the same realisation or on an independent realisation) into the Yule-Walker equations. For the two predictors, we estimate the mean-squared error and prove the asymptotic normality.

**Key words :** prediction theory, long-memory, linear model, mean-squared forecast error, central limit theorem, prediction interval.