

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE
SCIENTIFIQUE
UNIVERSITE MENTOURI CONSTANTINE
FACULTE DES SCIENCES EXACTES
DEPARTEMENT DE MATHEMATIQUES

N° d'ordre:

Série:

MEMOIRE
EN VUE DE L'OBTENTION DU DIPLOME DE
MAGISTER EN MATHEMATIQUES
OPTION : MATHEMATIQUES APPLIQUES
THEME
INFERENCE STATISTIQUE DANS LES MODELES LINEAIRES
PRESENTE PAR:
Kharoua Amina

Soutenu le:

Devant le jury:

Président	F. L. RAHMANI	M.C.	à l'Université Mentouri
Rapporteur	Z. MOHDEB	Prof.	à l'Université Mentouri
Examineur	F. MESSACI	M.C.	à l'Université Mentouri
Examineur	N. NEMOUCHI	M.C.	à l'Université Mentouri

Remerciements

Je tiens tout d'abord à remercier Pr Zahir Mohdeb pour m'avoir proposé un sujet passionnant et m'avoir encadré. Sa patience et sa confiance ont rendu ce travail très agréable.

Je remercie également les autres membres de mon jury : Dr Foued Rahmani, Dr Fatiha Messaci, Dr Nahima Nemouchi de l'honneur qu'ils m'ont fait en acceptant de siéger à mon jury de magistère.

Un grand merci à ma famille notamment ma mère, mon papa, mon frère et mes sœurs, sans oublier mes amies.

Enfin je remercie tous ceux qui m'ont aidé, car la poursuite de telles études est à la fois une question de rencontres et de moyens.

Amina KHAROUA

Table des matières

1	Introduction	5
1.1	Quelques notions et définitions	8
1.1.1	Rappels	8
1.1.2	Les modèles linéaires généraux	9
1.1.3	Les modèles autorégressifs et moyennes mobiles	13
1.1.4	Les modèles ARMA intégrés ou ARIMA(p, d, q)	15
1.1.5	Les modèles FARIMA(p, d, q)	16
1.1.6	Les modèles à longue mémoire	18
2	Prévision	19
2.1	Introduction	19
2.2	Prévision de Wiener-Kolmogorov	22
2.2.1	Prédiction d'un pas	22
2.2.2	Prédicteur de Wiener-Kolmogorov tronqué	23
2.3	Méthode de projection	29
2.3.1	Définition du prédicteur	29
2.3.2	Evaluation de l'erreur moyenne de prédiction	30
2.4	Prédiction à l'horizon h	35
2.4.1	Prédicteur de Wiener-Kolmogorov	36
2.4.2	Prédicteur de Wiener-Kolmogorov tronqué	37
2.4.3	Prédicteur linéaire des moindres carrés	40
3	Estimation des coefficients du prédicteur tronqué	42
3.1	Introduction	42
3.2	Prédicteur tronqué de Wiener-Kolmogorov	43
3.2.1	Méthode de prédiction	43
3.2.2	Estimation des coefficients et de l'erreur moyenne quadratique associées à la prédiction d'un pas	45
3.2.3	Prédiction à l'horizon h	54

4	Estimation des coefficients du prédicteur projeté	56
4.1	Introduction	56
4.2	Estimation de l'erreur quadratique due à l'estimation des coefficients	57
5	Prédiction et estimation du modèle sur une même réalisation	65
5.1	Introduction	65
5.2	Moment borné	68
5.3	L'erreur moyenne quadratique de prédiction du prédicteur de moindres carrés	72
5.4	Théorème central limite	77
6	Comparaison des prédicteurs sur des séries réelles et simulées	89
6.1	Trajectoires simulées de processus fractionnaires	89
6.1.1	Le processus fractionnaire	89
6.1.2	Méthodes de simulation	90
6.1.3	Comparaison des prédicteurs de Wiener-Kolmogorov tronqué et projeté	90
6.2	Etude de la rivière Nile	94
6.2.1	Modélisation	95
6.2.2	Prévision par le prédicteur projeté	95
6.3	Exemples	98
6.3.1	Etude de l'erreur quadratique due à la prévision d'un processus FARIMA(0,d,0)	98
6.3.2	Etude du niveau du lac majeur	101

Chapitre 1

Introduction

Pour un ensemble d'observations données nous allons associer un modèle approprié, en p articulier, choisir un modèle $ARMA(p, q)$ pour adapter une série de données donnée, et nous devons toujours non seulement découvrir les modèles cachés dans les données mais produire également les prévisions. Le terme modèle linéaire est parfois employé dans un sens plus général que ce que nous considérons ici, n'importe quel modèle qui relie des variables par un rapport linéaire est un modèle linéaire. Cette description adapte le modèle généralisé mentionné ici (c'est-à-dire : le modèle $ARMA(p, q)$).

Pourquoi le modèle linéaire ?

En raison de sa simplicité, il est mieux compris et plus facile à interpréter que la plupart des autres modèles, et les méthodes d'analyse et d'inférence sont mieux développés de plus il peut être employé au moins dans un premier temps car la formulation "modèle linéaire" est utile même pour les modèles non-linéaire qui peuvent être réduits à la forme linéaire au moyen de transformations.

Pour plus d'informations sur les modèles, nous allons étudier leurs comportements et leurs propriétés principales. Le chapitre 1 comporte alors en premier des rappels théoriques nécessaires sur les modèles, par exemple : stationnarité, inversibilité, expression des autocovariances, forme de la densité spectrale...etc.

Le modèle général autorégressif-moyenne mobile d'ordre (p, q) ou $ARMA(p, q)$ (*AutoRegressive Moving Average*) présenté par Box et Jenkins (1970) inclut deux classes de processus : processus $ARMA$ stationnaires dont les racines du polynôme autorégressif se trouvent dans le disque unité et le comportement non-stationnaire si le polynôme autorégressif admet de zéros à l'intérieur du cercle unité. La seconde classe des processus ainsi définie comprend deux cas particuliers importants : le modèle $ARMA$ non-stationnaire et non-homogène qui représente le comportement explosif ou

évolutionnaire, comme exemple nous avons "la croissance bactérienne" et le modèle ARMA non-stationnaire mais homogène, dont le polynôme autorégressif admet un zéro d'ordre d sur le cercle unité, qui peut être obtenu dans l'intégration du processus ARMA(p, q) stationnaire d fois, ce que nous appelons : processus autorégressif-moyenne mobile intégré noté ARIMA(p, d, q) (*AutoRegressive Integrated Moving Average*). La méthodologie d'ARIMA développée par Box et Jenkins (1976) (voir [12]) a gagné l'énorme popularité dans beaucoup de secteurs et la pratique en matière de recherche confirme sa puissance. Mais dans ce travail nous sommes intéressés par le modèle FARIMA(p, d, q) ou ARFIMA(p, d, q) (*AutoRegressive Fractionally Integrated Moving Average*); qui constituent une généralisation des processus ARIMA de Box et Jenkins dans lequel l'exposant de différenciation d était un entier positif. Dans le cas des processus FARIMA, d peut prendre des valeurs réelles, et non plus seulement des valeurs entières. Plus précisément nous sommes intéressés par le modèle FARIMA(p, d, q) stationnaire dont $d \in]0, 1/2[$.

Ce que nous allons étudier de façon plus détaillée; une classe des modèles paramétriques de série chronologique où on peut inclure les modèles ARIMA et FARIMA : les processus à longue mémoire généralisés. Ces processus peuvent tenir compte simultanément en modélisant, une dépendance de longue gamme aussi bien qu'un comportement cyclique périodique persistant. Ce genre de phénomène est souvent rencontré dans la pratique.

Dans le chapitre 2, 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 $\hat{X}_k(h)$ 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. Mais en pratique pour faire de la prévision, on ne dispose jamais d'un nombre infini d'observations. On ne peut pas donc utiliser pour prévoir le prédicteur de Wiener-Kolmogorov, on a alors deux possibilités : soit recourir au prédicteur de Wiener-Kolmogorov tronqué soit utiliser le prédicteur projeté. Ensuite nous donnerons la vitesse de convergence des erreurs quadratiques des prédicteurs projeté et tronqué vers le prédicteur de Wiener-Kolmogorov lorsque k , le nombre d'observations, tend vers l'infini pour des processus à longue mémoire.

Dans le chapitre 3, nous allons 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 que l'on veut prédire les processus $(X_n)_{n \in \mathbb{Z}}$ en dis-

posant d'une réalisation de longueur k . Pour l'estimation de la fonction de prédiction, on observe une série indépendante $(Y_n)_{n \in Z}$ de longueur T suivant la même loi que $(X_n)_{n \in Z}$. On suppose de plus que le processus $(X_n)_{n \in Z}$ appartient à une famille paramétrique et plus précisément que sa densité spectrale $f(\cdot, \theta_0)$ appartient à la famille $F_\theta = \{f(\cdot, \theta), \theta \in \Theta \text{ compact}\}$. Pour estimer le paramètre θ_0 , on utilise l'estimateur de Whittle (voir Fox et Taquq (1986) [16]). Puis nous 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 dans F aisant tendre T puis k vers l'infini.

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 vers le prédicteur de Wiener-Kolmogorov sera donc plus fort dans notre cas. Dans le chapitre 4, nous donnerons une majoration de l'erreur quadratique supplémentaire due à l'estimation des coefficients du prédicteur projeté 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 ou l'algorithme d'innovation quant à lui permet de calculer récursivement le prédicteur projeté (voir Brockwell (1991) [14]).

Dans les chapitres 3 et 4, l'estimation des coefficients du prédicteur est faite sur une réalisation indépendante de celle que l'on souhaite prévoir. Or en pratique on dispose rarement de deux réalisations indépendantes du même processus. Ing et Wei (2003) (voir [24]) se sont affranchis de cette hypothèse contraignante pour le prédicteur projeté d'un processus à courte mémoire. Godet (2008) (voir [21]) généralise leurs résultats pour des processus à longue mémoire. Nous intéresserons à cette généralisation, donc 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. 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 $k \times k$ et sa vitesse de convergence vers la matrice de covariance du processus $(X_n)_{n \in Z}$.

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 σ_ε^2), l'erreur due à la projection plus un terme supplémentaire dû à l'estimation des coefficients du prédicteur. Si on note n la longueur du passé utilisée 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 $K_n^4 = o(n^{1-2d-\delta})$. 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 centrale limite de notre prédicteur vers le prédicteur de Wiener-Kolmogorov.

A la fin dans le chapitre 6 nous allons comparer les prédicteurs étudiés aux chapitres précédents sur des séries réelles et simulées.

1.1 Quelques notions et définitions

1.1.1 Rappels

– Définition 1

Un processus stochastique est une famille de variables aléatoires (X_t) définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

– Définition 2

Une série chronologique est une série d'observations relevées dans le temps.

Fonction d'autocovariance

Si $(X_t)_{t \in Z}$ est un processus tel que $Var X_t < \infty, \forall t \in T \subset Z$, alors la fonction d'autocovariance γ_X de $(X_t)_{t \in Z}$ est définie par :

$$\gamma_X(r, s) = Cov(X_r, X_s) = E(X_r - E(X_r))(X_s - E(X_s)) \quad r, s \in T.$$

Stationnarité

La propriété de stationnarité suppose que le processus stochastique fluctue autour d'une valeur moyenne constante.

1. Stationnarité du second ordre :

La série temporelle $(X_t)_{t \in Z}$ est dite stationnaire du second ordre si :

- i) $E|X_t|^2 < \infty, \forall t \in Z$
- ii) $E(X_t) = m, \forall t \in Z$ (la moyenne ne dépend pas du temps t)
- iii) $\gamma_X(r, s) = \gamma_X(s + t, r + t), \forall t \in Z$.

Il serait alors convenable de redéfinir une fonction d'autocovariance d'un processus stationnaire du second ordre comme une fonction d'une seule variable :

$$\gamma_X(h) = \gamma_X(h, 0) = Cov(X_{t+h}, X_t), \quad \forall t, h \in Z.$$

2. Stationnarité au sens stricte :

La série temporelle $(X_t)_{t \in Z}$ est dite strictement stationnaire si les distributions de $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ et $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$ sont les mêmes pour tout entier h pour tout $t_1, t_2, \dots, t_n \in Z$

Inversibilité

Un processus ARMA(p, q) est dit inversible s'il existe une suite de coefficients (π_j) tels que $\sum_{j=0}^{\infty} |\pi_j| < \infty$ et :

$$\varepsilon_t = \sum_{j=0}^{+\infty} \pi_j X_{t-j} \quad t = 0, \pm 1, \dots$$

Densité spectrale

La densité spectrale d'un processus stationnaire est définie par :

$$\mathbf{f}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} e^{-ik\lambda} \gamma_k \quad -\pi \leq \lambda \leq \pi, \quad (1.1)$$

où γ est la fonction d'autocovariance qui doit être absolument sommable.

1.1.2 Les modèles linéaires généraux

Une question importante à laquelle on essaye de répondre par les modèles statistiques est la suivante : comment peut une quantité y être expliqué par un certain nombre des autres quantités x_1, x_2, \dots, x_n ?. Peut être le modèle le plus simple qui est employé pour répondre à cette question est le modèle linéaire :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon,$$

où $\beta_0, \beta_1, \dots, \beta_n$ sont des constantes et ε est un terme d'erreur qui explique les incertitudes.

On définit deux formes équivalentes de processus linéaires :

1. On décrit un modèle stochastique linéaire général qui suppose une série chronologique pour être produit par une agrégation linéaire des chocs aléatoires :

$$\begin{aligned}\tilde{X}_t &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \\ \tilde{X}_t &= \varepsilon_t + \sum_{j=1}^{+\infty} \psi_j \varepsilon_{t-j},\end{aligned}\tag{1.2}$$

où $\tilde{X}_t = X_t - \mu$ la déviation du processus X_t de sa moyenne et ε_t est un bruit blanc de moyenne nulle et de variance σ_ε^2 . Dans la représentation (1.2) de \tilde{X}_t , pour un processus stationnaire valide, il est nécessaire que les coefficients ψ_j soient absolument sommables c'est-à-dire : $\sum_{j=0}^{+\infty} |\psi_j| < \infty$

2. L'expression de \tilde{X}_t est une régression multiple de sa valeur courante sur ses valeurs passées $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \tilde{X}_{t-3}, \dots$:

$$\tilde{X}_t = \pi_1 \tilde{X}_{t-1} + \pi_2 \tilde{X}_{t-2} + \pi_3 \tilde{X}_{t-3} + \dots\tag{1.3}$$

La relation entre les coefficients ψ et π est obtenue en utilisant l'opérateur retard B défini par :

$$BX_t = X_{t-1} \quad \text{et} \quad B^j X_t = X_{t-j}.$$

Par exemple : nous considérons le modèle

$$\tilde{X}_t = \varepsilon_t - \theta \varepsilon_{t-1} = (1 - \theta B) \varepsilon_t,$$

avec $\psi_1 = -\theta, \psi_j = 0$ pour $j > 1$.

Si on décrit ε_t en terme de \tilde{X}_t on obtient

$$(1 - \theta B)^{-1} \tilde{X}_t = \varepsilon_t$$

et pour $|\theta| < 1$, on aura

$$(1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \dots) \tilde{X}_t = \varepsilon_t$$

d'où

$$\tilde{X}_t = -\theta \tilde{X}_{t-1} - \theta^2 \tilde{X}_{t-2} - \theta^3 \tilde{X}_{t-3} + \varepsilon_t$$

alors pour ce modèle $\pi_j = -\theta^j$.

En général, la représentation (1.2) sera écrite comme suit :

$$\tilde{X}_t = \left(1 + \sum_{j=1}^{+\infty} \psi_j B^j\right) \varepsilon_t = \Psi(B) \varepsilon_t, \quad (1.4)$$

où

$$\Psi(B) = 1 + \sum_{j=1}^{+\infty} \psi_j B^j.$$

Similairement, la représentation (1.3) sera :

$$\left(1 - \sum_{j=1}^{+\infty} \pi_j B^j\right) \tilde{X}_t = \Pi(B) \tilde{X}_t = \varepsilon_t, \quad (1.5)$$

où

$$\Pi(B) = 1 - \sum_{j=1}^{+\infty} \pi_j B^j.$$

Après avoir multiplié les deux membres de (1.5) par $\Psi(B)$ on obtient :

$$\begin{aligned} \Psi(B) \Pi(B) \tilde{X}_t &= \Psi(B) \varepsilon_t = \tilde{X}_t \\ \Psi(B) \Pi(B) &= 1 \\ \Pi(B) &= \Psi^{-1}(B). \end{aligned}$$

Fonction d'autocovariance d'un processus linéaire

La fonction d'autocovariance du processus linéaire représentée par (1.2) est donnée par :

$$\gamma_k = \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \psi_j \psi_{j+k} \quad (1.6)$$

Pour $k = 0$ on aura sa variance :

$$\gamma_0 = \sigma_\varepsilon^2 \sum_{j=0}^{+\infty} \psi_j^2. \quad (1.7)$$

La condition de stationnarité dont les coefficients ψ_j sont absolument sommables implique que le processus a une variance finie.

Une autre façon pour obtenir les autocovariances d'un processus linéaire est la fonction génératrice d'autocovariance :

$$G(z) = \sum_{k=-\infty}^{+\infty} \gamma_k z^k \quad (1.8)$$

dans laquelle la variance γ_0 du processus est le coefficient de $z^0 = 1$, et γ_k sont les coefficients de z et z^{-1} (car $\gamma_k = \gamma_{-k}$).

Conditions de stationnarité pour un processus linéaire

La convergence de la série (1.7) assure que le processus a une variance finie. Pour le processus linéaire (1.2) les conditions de stationnarités peuvent être garanties par la condition :

$$\sum_{j=0}^{+\infty} |\psi_j| < \infty.$$

Conditions d'inversibilité pour un processus linéaire

L'inversibilité est indépendante de la stationnarité et elle est aussi applicable pour les modèles linéaires non-stationnaires.

En général, le processus linéaire (1.2) est inversible s'il existe une suite de constantes $(\pi_j)_{j \geq 0}$, telles que :

$$\sum_{j=0}^{+\infty} |\pi_j| < \infty$$

et

$$\varepsilon_t = \sum_{j=0}^{+\infty} \pi_j X_{t-j}.$$

Spectre d'un processus linéaire

Si on remplace z dans la fonction génératrice d'autocovariance par $e^{-i2\pi\lambda}$ on obtient le spectre d'un processus linéaire :

$$\begin{aligned} \mathbf{f}(\lambda) &= 2\sigma_\varepsilon^2 \Psi(e^{-i2\pi\lambda}) \Psi(e^{i2\pi\lambda}) \\ \mathbf{f}(\lambda) &= 2\sigma_\varepsilon^2 |\Psi(e^{-i2\pi\lambda})|^2 \quad 0 \leq \lambda \leq \frac{1}{2}, \end{aligned} \quad (1.9)$$

où

$$\Psi(z) = 1 + \sum_{j=1}^{+\infty} \psi_j z^j.$$

1.1.3 Les modèles autorégressifs et moyennes mobiles

Les représentations (1.2) et (1.3) du processus linéaire général ne sont pas très applicables dans la pratique car elles contiennent un nombre infini de paramètres.

Une classe importante de modèles stochastiques utilisées pour décrire les séries chronologiques est la classe des modèles ARMA(p, q).

Les modèles autorégressifs d'ordre p ou AR(p)

On considère le cas particulier de la représentation (1.3), dans laquelle les p premiers coefficients seulement sont non nuls :

$$\begin{aligned}\tilde{X}_t &= \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \dots + \phi_p \tilde{X}_{t-p} + \varepsilon_t \\ (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{X}_t &= \varepsilon_t \\ \Phi(B) \tilde{X}_t &= \varepsilon_t,\end{aligned}\tag{1.10}$$

avec

$$\Phi(z) = 1 - \sum_{j=1}^p \phi_j z^j.$$

– Stationnarité des processus AR(p)

L'ensemble des paramètres $\phi_1, \phi_2, \dots, \phi_p$ d'un processus AR(p) doit satisfaire certaines conditions pour que le processus soit stationnaire :

$$\Phi(B) \tilde{X}_t = \varepsilon_t$$

donc

$$\tilde{X}_t = \Phi^{-1}(B) \varepsilon_t.$$

On obtient

$$\Phi(B) = (1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B),$$

où $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$ sont les racines de $\Phi(z) = 0$.

$$\tilde{X}_t = \Phi^{-1}(B) \varepsilon_t = \sum_{i=1}^p \frac{K_i}{1 - G_i B} \varepsilon_t,$$

avec

$$\Phi(B) = \prod_{i=1}^p (1 - G_i B).$$

Si $\Psi(B) = \Phi^{-1}(B)$ est une série convergente c'est-à-dire : les coefficients $\psi_j = \sum_{i=1}^p K_i G_i^j$ sont absolument sommables alors $\text{AR}(p)$ représente un processus stationnaire donc on doit avoir $|G_i| < 1$, pour $i = 1, 2, \dots, p$. On peut alors exprimer la condition de stationnarité que : $\Phi(z)$ n'ait pas de zéros dans le disque unité.

- Pour l'inversibilité aucune condition car le processus $\text{AR}(p)$ est toujours inversible.

Les modèles moyennes mobiles d'ordre q ou $\text{MA}(q)$

On considère de (1.2) où nous avons une infinité de paramètres que les q premiers coefficients seulement sont non nuls :

$$\tilde{X}_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

ou encore

$$\tilde{X}_t = \Theta(B)\varepsilon_t, \quad (1.11)$$

avec

$$\Theta(z) = 1 - \sum_{j=1}^q \theta_j z^j.$$

- **Inversibilité des processus $\text{MA}(q)$** : Les coefficients $\theta_1, \theta_2, \dots, \theta_q$ doivent satisfaire certaines conditions pour que le processus $\text{MA}(q)$ soit inversible :

$$\tilde{X}_t = \Theta(B)\varepsilon_t$$

donc

$$\Theta^{-1}(B)\tilde{X}_t = \varepsilon_t.$$

Si

$$\Theta(B) = \prod_{i=1}^q (1 - H_i B),$$

où $H_1^{-1}, H_2^{-1}, \dots, H_q^{-1}$ sont les racines de l'équation $\Theta(z) = 0$.

On obtient :

$$\Pi(B) = \Theta^{-1}(B) = \sum_{i=1}^q \frac{M_i}{1 - H_i B}.$$

La série $\Pi(B)$ converge si les coefficients $\pi_j = -\sum_{i=1}^q M_i H_i^j$ sont absolument sommables donc le processus $\text{MA}(q)$ est inversible si $|H_i| < 1$ pour $i = 1, 2, \dots, q$ c'est-à-dire : $\Theta(z)$ n'ait pas de zéros dans le disque unité.

- Pour la stationnarité aucune condition car le processus $\text{MA}(q)$ est toujours stationnaire.

Les modèles mixtes autorégressifs moyennes mobiles d'ordre (p, q) ou ARMA(p, q)

Pour permettre une meilleure flexibilité dans l'ajustement des séries chronologiques, il est souvent utile de combiner les formes autorégressives et moyennes mobiles. Ceci nous amène à définir les modèles mixtes autorégressifs moyennes mobiles d'ordre (p, q) , notés ARMA(p, q), vérifiant

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \dots + \phi_p \tilde{X}_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}.$$

On peut écrire encore :

$$\Phi(B)\tilde{X}_t = \Theta(B)\varepsilon_t, \quad (1.12)$$

avec

$$\Phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$$

$$\Theta(z) = 1 - \sum_{j=1}^q \theta_j z^j.$$

Notons que $\tilde{X}_t = X_t - \mu$ où $\mu = E(X_t)$ alors le processus ARMA(p, q) général peut être défini par :

$$\Phi(B)X_t = \theta_0 + \Theta(B)\varepsilon_t,$$

où $\theta_0 = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\mu$.

- **Conditions de stationnarité et d'inversibilité des processus ARMA** Les conditions de stationnarité et d'inversibilité du modèle mixte ARMA se déduisent de celles des AR(p) et MA(q) où on a :
 Processus ARMA(p, q) stationnaire : si $\Phi(z)$ n'ait pas de zéros dans le disque unité.
 Processus ARMA(p, q) inversible : si $\Theta(z)$ n'ait pas de zéros dans le disque unité.
 avec $\Phi(z)$ et $\Theta(z)$ n'aient pas de zéros communs.
- **Spectre des processus ARMA(p, q)** : La densité spectrale d'un processus ARMA(p, q) est définie par :

$$\mathbf{f}(\lambda) = 2\sigma_\varepsilon^2 \frac{|\Theta(e^{-i2\pi\lambda})|^2}{|\Phi(e^{-i2\pi\lambda})|^2} \quad 0 \leq \lambda \leq \frac{1}{2}. \quad (1.13)$$

1.1.4 Les modèles ARMA intégrés ou ARIMA(p, d, q)

Certaines séries chronologiques rencontrées dans la pratique, surtout dans les domaines d'industrie et d'économie, présentent des comportements non

stationnaires. Généralement ces séries n'admettent pas de moyenne fixe. Néanmoins, elles se caractérisent par un comportement homogène. En d'autres termes, bien que le niveau moyen autour duquel la série fluctue change d'une période à l'autre, la série chronologique garde la même allure générale moyennant de temps et d'échelle. Ce genre de comportement peut être représenté par un opérateur généralisé $\Phi^*(B)$ qui, considéré comme polynôme dans B , admet une ou plusieurs racines sur le cercle unité. Ce type d'opérateur peut s'écrire sous la forme :

$$\Phi^*(B) = \Phi(B)(1 - B)^d,$$

où $\Phi(B)$ est un opérateur autogréssif stationnaire et d un entier positif.

1.1.5 Les modèles FARIMA(p, d, q)

Le modèle général pour représenter des séries chronologiques stationnaires ou non stationnaires à comportement homogène peut être présenté sous la forme suivante :

$$\Phi^*(B)\tilde{X}_t = \Phi(B)(1 - B)^d\tilde{X}_t = \Theta(B)\varepsilon_t. \quad (1.14)$$

Ce modèle est alors appelé modèle autoregréssif moyenne mobile fractionnaire intégré (*AutoRegressive Fractionally Integrated Moving Average*) ou modèle FARIMA(p, d, q). Un modèle fractionnaire intégré a pour caractéristique une dépendance entre des observations éloignées.

– **Remarque**

L'expression (1.14) peut s'écrire sous la forme suivante :

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t, \quad (1.15)$$

où

$$Y_t = \nabla^d \tilde{X}_t.$$

Ainsi dans l'expression (1.15), $\Phi(B)$ est un opérateur autoregréssif stationnaire, donc le processus Y_t est un processus stationnaire.

Stationnarité des modèles FARIMA(p, d, q)

Il s'agit d'obtenir les conditions sur d pour qu'un processus FARIMA(p, d, q) soit stationnaire, et pour cela il suffit d'étudier le comportement d'un processus FARIMA d'ordre $(0, d, 0)$ qui s'appelle "bruit fractionnaire intégré" et qui satisfait la relation :

$$(1 - B)^d \tilde{X}_t = \varepsilon_t. \quad (1.16)$$

L'expression (1.16) est équivalente à :

$$\tilde{X}_t = (1 - B)^{-d} \varepsilon_t. \quad (1.17)$$

Le développement dans S érie de $(1 - B)^{-d}$ est donné par :

$$(1 - B)^{-d} = \sum_{j=0}^{+\infty} \frac{(-d)(-d-1)\dots(-d-j+1)}{j!} (-B)^j$$

$$(1 - B)^{-d} = \sum_{j=0}^{+\infty} \frac{(d)(d+1)\dots(d+j-1)}{j!} (B)^j.$$

Notons

$$\frac{\Gamma(d+j)}{\Gamma(d)} = (d)(d+1)\dots(d+j-1),$$

avec $\Gamma(n) = (n-1)!$.

L'expression (1.17) peut être représentée sous la forme :

$$\tilde{X}_t = \sum_{j=0}^{+\infty} b_j \varepsilon_{t-j} \quad (1.18)$$

et

$$b_j = \frac{\Gamma(d+j)}{\Gamma(d)\Gamma(j+1)}.$$

Le comportement asymptotique des coefficients (a_j) se déduit directement de la formule de Sheppard :

$$\frac{\Gamma(j+a)}{\Gamma(j+b)} \sim j^{a-b} \quad si \quad j \longrightarrow \infty$$

alors

$$b_j \sim \frac{j^{d-1}}{\Gamma(d)} \quad si \quad j \longrightarrow \infty$$

d'où

$$b_j^2 \sim \frac{j^{2d-2}}{\Gamma(d)^2} \quad si \quad j \longrightarrow \infty.$$

La suite des b_j est de carré sommable si et seulement si :

$$2d - 2 < -1 \Leftrightarrow d < \frac{1}{2}.$$

Donc on déduit qu'un processus FARIMA(p, d, q) est stationnaire si et seulement si $0 < d < \frac{1}{2}$.

1.1.6 Les modèles à longue mémoire

Les processus ARMA s'appellent souvent les processus à courte mémoire parce que leurs covariances décroissent rapidement. D'autre part, un processus à longue mémoire est caractérisé par la propriété suivante : la fonction d'autocovariance décroît plus lentement c'est-à-dire : n'est pas absolument sommable

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

Ils sont appelés ainsi dans R aison de la corrélation forte entre les observations largement séparées dans le temps. Ils permettent donc d'identifier les phénomènes de persistance. Les modèles de séries chronologiques à longue mémoire ont attirés beaucoup d'attention récemment car ces processus modélisent une grande variété de phénomènes et il y a maintenant une réalisation croissante que les caractéristiques de possession de longue mémoire de séries chronologiques surgissent dans des domaines aussi divers que des sciences économiques, géophysique ... etc.

Chapitre 2

Prévision

2.1 Introduction

Pour prévoir les valeurs futures d'une série d'observations donnée, on doit choisir un modèle approprié pour l'adapter. Pour que ce procédé soit significatif il doit être au moins plausible que les données soient dans F ait une réalisation d'un processus ARMA stationnaire, alors il est parfois nécessaire de faire une transformation afin de produire une nouvelle série qui est plus compatible avec la stationnarité d'un processus, on a par exemple, la transformation logarithmique qui est utilisée pour les séries dont l'écart type augmente linéairement avec la moyenne. Après avoir transformé les données, il devient nécessaire également d'identifier des valeurs appropriées pour les paramètres p et q . Il pourrait être évident à première vue que plus les valeurs de p et q choisis sont élevées plus le modèle adapté sera meilleur, mais pour obtenir le modèle ARMA le plus satisfaisant on doit minimiser certaines valeurs, par exemple, pour un processus autorégressif pur le modèle choisi est celui qui réduit au minimum la valeur FPE (*final prediction error*)

où

$$FPE = \hat{\sigma}^2 \frac{n+p}{n-p}$$

avec $\hat{\sigma}^2$ est l'estimateur de maximum de vraisemblance de σ^2 . On a encore d'autres valeurs ou critères comme le BIC, l'AIC et l'AICC.

Une partie essentielle du procédé doit examiner les résidus, si le modèle est satisfaisant, ont l'aspect d'une réalisation d'un bruit blanc.

Dans ce chapitre nous étudions des méthodes de prévision pour les processus à longue mémoire. Ils sont supposés stationnaires du second ordre, linéaires 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 prédire.

Nous présentons deux méthodes de prédiction linéaire des séries chronologiques à longue mémoire. La première méthode consiste à tronquer le prédicteur de Wiener-Kolmogorov en limitant les observations, c'est-à-dire que nous n'utilisons que les k dernières observations, qui sont les seules valeurs disponibles dans la pratique. Dans la seconde méthode, nous proposons le prédicteur construit par projection sur le passé fini observé.

Soit $(X_n)_{n \in \mathbb{Z}}$ un processus stationnaire (faiblement) à temps discret de moyenne nulle et γ sa fonction d'autocovariance. Nous supposons que le processus $(X_n)_{n \in \mathbb{Z}}$ est un processus à longue mémoire c'est-à-dire :

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

Le processus $(X_n)_{n \in \mathbb{Z}}$ admet une représentation moyenne mobile infinie comme suit :

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

où $(\varepsilon_n)_{n \in \mathbb{Z}}$ est un processus de bruit blanc composé des variables aléatoires non-corrélées, chacune de moyenne zéro et de variance σ_ε^2 et $(b_j)_{j \in \mathbb{N}}$ est une suite de coefficients qui sont de carrés sommables. Nous supposons plus loin que $(X_n)_{n \in \mathbb{Z}}$ admet une représentation autorégressive infinie :

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

où les $(a_j)_{j \in \mathbb{N}}$ sont absolument sommables. Nous supposons également que $(a_j)_{j \in \mathbb{N}}$ et $(b_j)_{j \in \mathbb{N}}$, donnés respectivement par (2.2) et (2.1), satisfont les conditions suivantes $\forall \delta > 0$:

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

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

où C_1 et C_2 sont des constantes et d un paramètre vérifiant $d \in]0, 1/2[$. Par exemple, le processus FARIMA $(X_n)_{n \in \mathbb{Z}}$, les processus à longue mémoire les plus étudiées, est la solution stationnaire de l'équation :

$$\Phi(B)(1 - B)^d X_n = \Theta(B)\varepsilon_n \quad (2.5)$$

où $(\varepsilon_n)_{n \in \mathbb{Z}}$ est un processus de bruit blanc de moyenne zéro. Les coefficients vérifient

$$|a_j| \sim C_1 j^{-d-1} \quad (2.6)$$

$$|b_j| \sim C_2 j^{d-1} \quad (2.7)$$

et ainsi pris de (2.3) et (2.4). Plus généralement, les conditions (2.3) et (2.4) sont vérifiées quand les coefficients sont de la forme :

$$|a_j| \sim L(j)j^{-d-1} \quad (2.8)$$

$$|b_j| \sim L'(j)j^{d-1}, \quad (2.9)$$

où L et L' sont des fonctions à variation lente. Une fonction positive L est une fonction à variation lente dans le sens de Zygmund (1986) (voir [39]) si, pour $\delta > 0$, $x \mapsto x^{-\delta}L(x)$ croît et $x \mapsto x^\delta L(x)$ décroît.

La condition (2.4) implique que la fonction d'autocovariance γ du processus $(X_n)_{n \in \mathbb{Z}}$ vérifie :

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

En effet,

pour $\delta < 1 - 2d$:

$$\begin{aligned} \gamma(k) &= \sum_{j=0}^{+\infty} b_j b_{j+k} \\ |\gamma(k)| &\leq \sum_{j=1}^{+\infty} |b_j b_{j+k}| + |b_0 b_k| \\ &\leq C_2^2 \sum_{j=1}^{+\infty} j^{d-1+\delta/2} (k+j)^{d-1+\delta/2} + |b_0 b_k| \\ &\leq C_2^2 k^{2d-1+\delta} \int_1^{+\infty} j^{d-1+\delta/2} (1+j)^{d-1+\delta/2} dj + C_2 k^{d-1+\delta/2} \\ &\leq C_3 k^{2d-1+\delta}. \end{aligned}$$

(2.8) et (2.9) définissent encore une classe de processus à longue mémoire. En effet, Inoue (1997) (voir [25]) a montré que (2.9) implique que :

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

où β est la fonction bêta. Noter que l'inverse n'est pas vrai, plus de conditions sur la série $(b_j)_{j \in \mathbb{N}}$ sont nécessaires afin d'obtenir un équivalent asymptotique pour $(\gamma_j)_{j \in \mathbb{Z}}$ (voir Inoue (2000) [26]).

2.2 Prédiction de Wiener-Kolmogorov

2.2.1 Prédiction d'un pas

Le but de la présente partie est de calculer la meilleure prédiction linéaire en une étape (avec la distance minimum de moyenne quadratique avec la vraie variable aléatoire) sachant tout le passé $\{X_{k+1-j}, j \leq 1\}$. Notre prédicteur est donc une combinaison linéaire infinie du passé

$$\hat{X}_k(1) = \sum_{j=0}^{+\infty} \lambda(j) X_{k-j},$$

où $(\lambda(j))_{j \in \mathbb{N}}$ sont choisis de telle sorte que l'erreur moyenne quadratique de prévision :

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

soit minimum. dans \mathbb{R} à l'aide de la représentation moyenne mobile de $(X_n)_{n \in \mathbb{N}}$, nous pouvons réécrire notre prédicteur $\hat{X}_k(1)$ comme suit :

$$\hat{X}_k(1) = \sum_{j=0}^{+\infty} \phi(j) \varepsilon_{k-j},$$

où $(\phi(j))_{j \in \mathbb{N}}$ sont fonction des $(\lambda(j))_{j \in \mathbb{N}}$ et des $(a_j)_{j \in \mathbb{N}}$ définis en (2.2).

Erreur moyenne quadratique

De la représentation moyenne mobile infinie de $(X_n)_{n \in \mathbb{N}}$ donnée en (2.1), nous pouvons réécrire l'erreur moyenne quadratique de la prédiction comme suit :

$$\begin{aligned} E[(\hat{X}_k(1) - X_{k+1})^2] &= E \left[\left(\sum_{j=0}^{+\infty} \phi(j) \varepsilon_{k-j} - \sum_{j=0}^{+\infty} b(j) \varepsilon_{k+1-j} \right)^2 \right] \\ &= E \left[\left(\varepsilon_{k+1} - \sum_{j=0}^{+\infty} (\phi(j) - b(j+1)) \varepsilon_{k-j} \right)^2 \right] \end{aligned}$$

$$= \left(1 + \sum_{j=0}^{+\infty} (b_{j+1} - \phi(j))^2 \right) \sigma_\varepsilon^2,$$

puisque les variables aléatoires $(\varepsilon_n)_{n \in \mathbb{Z}}$ sont non-corrélées et de variance σ_ε^2 . La plus petite erreur moyenne quadratique de la prédiction est obtenue dans \mathbb{R} en remplaçant les $\phi(j)$ par b_{j+1} pour tout $j \geq 0$.

La plus petite erreur de prévision de $(X_n)_{n \in \mathbb{Z}}$ est σ_ε^2 dans la classe des prédicteurs linéaires. En outre, si

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

représente le polynôme caractéristique de $(a(j))_{j \in \mathbb{N}}$ et

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

celui de $(b(j))_{j \in \mathbb{N}}$, alors de l'identité, $A(z) = B(z)^{-1}$, $|z| \leq 1$, nous pouvons écrire :

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

2.2.2 Prédicteur de Wiener-Kolmogorov tronqué

Dans la pratique, nous connaissons seulement un sous-ensemble fini du passé, celui que nous avons observé. Ainsi le prédicteur devrait seulement dépendre des observations. Supposons que nous connaissons seulement l'ensemble $\{X_1, X_2, \dots, X_k\}$ et que nous remplaçons les valeurs inconnues par zéro, alors nous avons le nouveau prédicteur suivant :

$$\hat{X}'_k(1) = - \sum_{j=1}^k a_j X_{k+1-j}. \quad (2.12)$$

Il est équivalent à dire que nous avons tronqué la série infinie (2.11) aux limites de k . La proposition suivante nous fournit les propriétés asymptotiques de l'erreur moyenne quadratique de prévision dans F onction de k .

Erreur moyenne quadratique quand le prédicteur est tronqué

Proposition 1 Soit $(X_n)_{n \in \mathbb{Z}}$ un processus linéaire stationnaire défini par (2.1), (2.2) et vérifiant les conditions (2.3) et (2.4). Nous pouvons approcher l'erreur moyenne quadratique de prévision de $\hat{X}'_k(1)$ par :

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

Notons que l'erreur de prévision est la somme de σ_ε^2 , l'erreur du modèle de Wiener-Kolmogorov et l'erreur due à la troncature à k termes qui est de l'ordre $O(k^{-1+\delta})$ pour tout $\delta > 0$.

Preuve. On a

$$\begin{aligned} & X_{k+1} - \hat{X}'_k(1) \\ = & X_{k+1} - \hat{X}_k(1) + \hat{X}_k(1) - \hat{X}'_k(1) \\ = & \sum_{j=0}^{+\infty} b_j \varepsilon_{k+1-j} - \sum_{j=0}^{+\infty} b_{j+1} \varepsilon_{k-j} - \sum_{j=1}^{+\infty} a_j X_{k+1-j} + \sum_{j=1}^k a_j X_{k+1-j} \\ = & \varepsilon_{k+1} - \sum_{j=k+1}^{+\infty} a_j X_{k+1-j}. \end{aligned} \quad (2.13)$$

Les deux parties de la somme (2.13) sont orthogonales pour le produit scalaire lié à la norme de moyenne quadratique. Donc

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

Si on note $\hat{\gamma}_k^2 = E[(X_{k+1} - \hat{X}'_k(1))^2]$ on aura :

$$\hat{\gamma}_k^2 - \sigma_\varepsilon^2 = \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \gamma(l-j) \quad (2.14)$$

Pour le second membre nous avons :

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

$$\leq 2 \sum_{j=k+1}^{+\infty} |a_j| |a_{j+1}| |\gamma(1)| + \sum_{l=k+1}^{+\infty} a_j^2 \gamma(0) + 2 \sum_{j=k+1}^{+\infty} |a_j| \sum_{l=j+2}^{+\infty} |a_l| |\gamma(l-j)|.$$

Pour tout $\delta > 0$ des inégalités (2.3) et (2.10), il s'ensuit que :

$$\left| \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \gamma(l-j) \right| \leq C_1^2 C_3 \left(2 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta} (j+1)^{-d-1+\delta} + \sum_{j=k+1}^{+\infty} (j^{-d-1+\delta})^2 \right) \quad (2.15)$$

$$+ 2C_1^2 C_3 \sum_{j=k+1}^{+\infty} j^{-d-1+\delta} \sum_{l=j+2}^{+\infty} l^{-d-1+\delta} |l-j|^{2d-1+\delta} \quad (2.16)$$

Supposons maintenant que $\delta < 1/2 - d$ pour le terme (2.15), puisque $j \mapsto j^{-d-1+\delta} (j+1)^{-d-1+\delta}$ est une fonction positive et décroissante sur R^+ , nous avons les approximations suivantes

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

Puisque la fonction $j \mapsto (j^{-d-1+\delta})^2$ est aussi positive et décroissante, nous pouvons établir d'une manière semblable que

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

Pour la double série infinie (2.16), nous pouvons comparer la série avec une intégrale.

Dans le prochain Lemme, nous établissons le résultat nécessaire pour cette comparaison

Lemme 1 Soient \mathbf{g} la fonction $(l, j) \mapsto j^{-d-1+\delta} l^{-d-1+\delta} |l-j|^{2d-1+\delta}$ et m et n deux entiers positifs. Supposons que $\delta < 1 - 2d$ et $m \geq \frac{\delta-d-1}{\delta+2d-1}$ pour

tout $\delta \in]0, \frac{\delta-d-1}{\delta+2d-1}[$. On pose $A_{n,m} = [n, n+1] \times [m, m+1]$. Si $n \geq m+1$ alors :

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

Preuve. (voir Godet (2007b) [20]). •

Supposons maintenant que $\delta < 1 - 2d$. Grâce au Lemme précédent et les équivalents asymptotiques de (2.15), il existe $K \in \mathbb{N}$ tel que si $k > K$, on a

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

En utilisant la substitution $jl' = l$ dans l'intégrale, on obtient :

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

Puisque si $\delta < (1-d)/2$,

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

il s'ensuit que

$$\begin{aligned} \left| \sum_{j=k+1}^{\infty} \sum_{l=k+1}^{\infty} a_j a_l \gamma(l-j) \right| &\leq O(k^{-1+3\delta}) + O(k^{-2d-1}) \\ &\leq O(k^{-1+3\delta}). \end{aligned} \quad (2.17)$$

Noter que si l'égalité est vraie avec les conditions $\delta > 0, \delta < 1 - 2d$ et $\delta < (1-d)/2$, elle est également vraie pour n'importe quel $\delta > 0$. Ceci conclut la preuve. •

Nous allons maintenant montrer qu'il existe des processus à longue mémoire dont l'erreur de prédiction est asymptotiquement de l'ordre $O(k^{-1})$. Lorsque nous précisons la forme des coefficients autorégressifs et la fonction d'autocovariance, on obtient un équivalent de l'erreur moyenne quadratique de prédiction.

Proposition 2 Soit $(X_n)_{n \in \mathbb{Z}}$ un processus linéaire stationnaire définie par (2.1), (2.2) et vérifiant (2.8). Supposons que les coefficients $(a_j)_{j \in \mathbb{N}^*}$ sont de signe constant. Nous supposons également que la fonction d'autocovariance γ du processus $(X_n)_{n \in \mathbb{Z}}$ vérifie :

$$|\gamma(j)| \sim L'(j)j^{2d-1},$$

avec L' une fonction variant lentement et que la fonction γ est d'un signe constant. Puis $O(k^{-1})$ est optimal c'est-à-dire : nous avons l'équivalence asymptotique suivante :

$$E \left(\left[X_{k+1} - \hat{X}'_k(1) \right]^2 \right) = \sigma_\varepsilon^2 + L''(k)k^{-1} + o(k^{-1}) \text{ quand } k \rightarrow +\infty \quad (2.18)$$

où L'' une fonction variant lentement.

Preuve. Dans ce cas particulier, on peut estimer l'erreur de prédiction plus précisément pour grand k :

$$\begin{aligned} & \left| \sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \gamma(l-j) \right| \\ &= 2 \sum_{j=k+1}^{+\infty} |a_j| \sum_{l=k+1}^{+\infty} |a_l| |\gamma(l-j)| \sum_{j=k+1}^{+\infty} a_j^2 \gamma(0) \\ &= \int_{k+1}^{+\infty} j^{-d-1} L(j) \int_{j+1}^{+\infty} l^{-d-1} L(l) (l-j)^{2d-1} L'(l-j) dl dj + \frac{k^{-2d-1} L^2(k)}{1+2d} \\ &\quad + O(k^{-2d-1}) \\ &\sim \int_{k+1}^{+\infty} j^2 L^2(j) L'(j) \int_{1/j+1}^{+\infty} l^{-d-1} \frac{L(lj)}{L(j)} (l-1)^{2d-1} \frac{L'(j(l-1))}{L'(j)} dl dj \\ &\sim k^{-1} \frac{L^2(k) L'(k) \Gamma(1-d) \Gamma(2d)}{\Gamma(1+d)}. \end{aligned} \quad (2.19)$$

•

Exemple 1 Supposons maintenant que $(X_n)_{n \in \mathbb{Z}}$ est un bruit fractionnaire intégré, qui est la solution stationnaire de

$$X_n = (1 - B)^{-d} \varepsilon_n \quad (2.20)$$

avec B l'opérateur retard, $(\varepsilon_n)_{n \in \mathbb{Z}}$ est un processus de bruit blanc et $d \in]0, 1/2[$ (voir par exemple Brockwell et Davis (1991) [14]). On peut calculer les coefficients pour obtenir

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

et

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

Nous pouvons également calculer la fonction d'autocovariance :

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

et

$$\gamma(j) \sim \frac{j^{2d-1} \Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \sigma_\varepsilon^2 \text{ quand } j \rightarrow +\infty.$$

Dans ce cas particulier, on peut estimer l'erreur de prévision plus précisément et donner l'expression explicite de L'' défini en (2.18) :

$$\sum_{j=k+1}^{+\infty} \sum_{l=k+1}^{+\infty} a_j a_l \gamma(l-j) \sim \frac{\Gamma(1-2d)\Gamma(2d)\sigma_\varepsilon^2}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)} k^{-1}. \quad (2.21)$$

Puisque dans ce cas les fonctions L et L' sont constantes et égales respectivement à $\frac{1}{\Gamma(-d)}$ et $\frac{\Gamma(1-2d)\sigma_\varepsilon^2}{\Gamma(d)\Gamma(1-d)}$.

Dans le cas spécifique du bruit fractionnaire intégré, nous pouvons écrire l'erreur de prédiction comme suit

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

et nous pouvons exprimer $C(d)$ dans fonction de d :

$$C(d) = \frac{\Gamma(1-2d)\Gamma(2d)\sigma_\varepsilon^2}{\Gamma(-d)^2\Gamma(d)\Gamma(1+d)}. \quad (2.22)$$

Il est facile de démontrer que $C(d) \rightarrow +\infty$ quand $d \rightarrow 1/2$ et nous pouvons écrire l'équivalent asymptotique de $C(d)$ quand $d \rightarrow 1/2$:

$$C(d) \sim \frac{\sigma_\varepsilon^2}{(1-2d)\Gamma(-1/2)^2\Gamma(1/2)\Gamma(3/2)}. \quad (2.23)$$

Quand $d \rightarrow 0$, $C(d) \rightarrow 0$ et nous avons l'équivalence suivante quand $d \rightarrow 0$:

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

Après ces résultats sur le prédicteur tronqué de Wiener-Kolmogorov, nous proposons un autre prédicteur. Au lieu de la troncature de la série qui définit le prédicteur de Wiener-Kolmogorov à k termes, il est plus efficace de projeter directement sur les observations afin de réduire au minimum l'erreur de prédiction. Nous étudions cette méthode dans la section suivante.

2.3 Méthode de projection

Dans cette section, nous étudions les propriétés asymptotiques de l'erreur moyenne quadratique de prédiction quand nous projetons la variable aléatoire X_{k+1} sur l'espace fermé engendré par le sous-ensemble $\{X_1, \dots, X_k\}$. Adapter un processus autorégressif d'ordre k est équivalent à la projection sur le sous espace fermé engendré par $\{X_1, \dots, X_k\}$.

2.3.1 Définition du prédicteur

Considérons le processus $(X_n)_{n \in \mathbb{Z}}$ à longue mémoire vérifiant les hypothèses (2.1),(2.2),(2.3) et (2.4). On définit le prédicteur comme étant la projection traçant sur le sous espace fermé de l'espace de Hilbert engendré par $\{X_1, \dots, X_k\}$ de l'espace de $L^2(\Omega, F, P)$ muni du produit scalaire $\langle X, Y \rangle = E(XY)$. Alors le prédicteur peut être écrit comme suit

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

Les coefficients $(-a_{i,k})_{1 \leq i \leq k}$ de la projection vérifient les équations, qui s'appellent les équations de Yule-Walker d'ordre k :

$$\forall j = \overline{1, k}, \quad \sum_{i=1}^k a_{i,k} \gamma(i-j) = -\gamma(j). \quad (2.24)$$

L'erreur moyenne quadratique de prédiction est :

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

2.3.2 Evaluation de l'erreur moyenne de prédiction

Dans le prochain théorème nous dérivons une expression asymptotique pour l'erreur de prévision en projetant sur le passé fini :

Théorème 1 Soit $(X_n)_{n \in \mathbb{Z}}$ un processus stationnaire linéaire défini par (2.1), (2.2) et vérifiant les conditions (2.3) et (2.4). Alors l'erreur moyenne quadratique de prédiction de $\tilde{X}_k(1)$ est donnée par

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

il existe des processus à longue mémoire tels que :

$$E[(X_{k+1} - \tilde{X}_k(1))^2] - \sigma_\varepsilon^2 \sim Ck^{-1} \quad k \rightarrow +\infty,$$

où C est une constante.

preuve. On pose $\tilde{\gamma}_k^2 = E[(X_{k+1} - \tilde{X}_k(1))^2]$. Puisque la projection donne la plus petite erreur de prédiction en utilisant k observations, l'erreur utilisant la troncature est plus grande et puisque la méthode de troncature implique une erreur liée par $O(k^{-1})$, nous obtenons (2.26). En conséquence nous devons seulement montrer que $O(k^{-1})$ est atteint. C'est le cas pour les processus fractionnaires intégrés. Nous voulons exprimer l'erreur due à la projection en termes d'erreur de troncature de Wiener-Kolmogorov. Par conséquent dans le cas d'un processus fractionnaire intégré nous avons besoin de montrer seulement que :

$$\tilde{\gamma}_k^2 - \sigma_\varepsilon^2 \sim Ck^{-1}.$$

Maintenant, remplaçons $a_{j,k}$ par 0 si $j > k$ et développons (2.25)

$$\begin{aligned} & \tilde{\gamma}_k^2 - \sigma_\varepsilon^2 \\ &= \sum_{j=0}^{+\infty} \sum_{l=0}^{+\infty} (a_j - a_{j,k})(a_l - a_{l,k})\gamma(l-j) \end{aligned} \quad (2.27)$$

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

Nous étudions d'abord le premier terme de la somme (2.28). Pour tout $j > 0$, nous avons $\sum_{l=0}^{+\infty} a_l \gamma(l-j) = 0$:

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

comme ε_n est orthogonal à X_{n-j} pour $j > 0$. Nous pouvons réécrire ainsi le premier terme de (2.28) :

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

car $a_0 = a_{0,k} = 1$ selon la définition. Après nous étudions le second terme de la somme dans (2.28). Nous avons :

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

$$+ \sum_{l=0}^k a_l \sum_{j=1}^k (a_j - a_{j,k}) \gamma(l-j) \quad (2.30)$$

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

De même, nous réécrivons le terme (2.29) en employant les équations de Yule-Walker :

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

Nous remarquons alors que (2.29) est égal à l'opposé de (2.30). Par conséquent il s'ensuit que :

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

D'une manière semblable, nous pouvons réécrire le troisième terme de la somme (2.31), en utilisant le Théorème de Fubini et puis les équations de Yule-Walker :

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

Ce dernier terme est donc égal à l'erreur de prédiction due à la troncature (2.14). Afin de comparer l'erreur de prédiction en tronquant le prédicteur de Wiener-Kolmogorov et en projetant sur le passé fini pour un processus fractionnaire intégré, nous avons besoin du signe de tous les composants de la somme dans (2.31). Pour le bruit fractionnaire intégré, nous savons la formule explicite pour a_j et $\gamma(j)$

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

et

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

Afin d'obtenir le signe de $(a_{l,k} - a_l)$ nous utilisons la formule explicite donnée dans Brockwell et Davis (1988) [13] et nous obtenons facilement que $(a_{l,k} - a_l)$ est négatif pour tout $l = 1, k$.

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

puisque $\forall j \in N^*, a_j < 0$. Pour donner un équivalent asymptotique de l'erreur de prédiction, nous utilisons la somme donnée en (2.31). Nous avons le signe des trois termes : le premier est négatif, le second est positif et le dernier est négatif. D'ailleurs, le troisième terme est égal à l'erreur de prédiction par troncature et nous avons montré que cette équivalence asymptotique de l'ordre $O(k^{-1})$. L'erreur de prédiction quand la projection converge plus rapidement vers zéro que l'erreur par troncature seulement si le deuxième terme est équivalent à Ck^{-1} , avec C une constante. Donc, nous recherchons une borne pour $(a_l - a_{l,k})$ donnée par la formule explicite en (2.32) :

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

En utilisant l'inégalité suivante ;

$$\forall x \in R, 1 + x \leq \exp(x),$$

on a

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

D'après l'inégalité précédente, on a

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

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

À mesure que la fonction $x \mapsto \frac{1}{k-d-m}$ croît, nous pouvons borner les séries par les intégrales.

Ainsi

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

Cherchons maintenant une borne pour $\exp\left(d \int_0^k \frac{1}{k-d-m} dm\right)$. Nous avons :

$$\begin{aligned} d \int_0^k \frac{1}{k-d-m} dm &\sim \ln(k) \\ &\leq \frac{3}{2} \ln(k) \end{aligned}$$

pour k assez grand. Par conséquent, il existe K tels que pour tout le $k \geq K$:

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

$$\begin{aligned}
&\leq C(k-d)^d \sum_{l=1}^{k-1} l^{-d-1} (k-d-l)^{-d} \sum_{j=k+1}^{+\infty} j^{-d-1} (j-l)^{2d-1} \\
&\quad + Ck^{-d-1} k^{\frac{3}{2}d} k^{-d} \\
&\leq \frac{C}{(k-d)^2} \int_{1/(k-d)}^1 l^{-d-1} (1-l)^{-d} \int_1^{+\infty} j^{-d-1} (j-1)^{2d-1} dl dj \\
&\quad + Ck^{-\frac{1}{2}d-1} \\
&\leq C'(k-d)^{-2+d} + Ck^{-\frac{1}{2}d-1}.
\end{aligned}$$

Ainsi le terme positif a un ordre asymptotique plus petit que l'erreur de prévision faite par troncature. C'est pourquoi nous avons prouvé que, dans le cas particulier de processus fractionnaire intégrée, les deux erreurs de prédiction sont équivalentes à Ck^{-1} avec C une constante. •

Les deux approches de prédiction à l'horizon 1, par troncature aux k termes ou par la projection sur les k dernières observations ont par conséquent une erreur de prédiction avec le même ordre de convergence k^{-1} . Ainsi il est intéressant d'étudier comment la deuxième approche améliore la prédiction et si l'amélioration est négligeable en ce qui concerne l'erreur de prédiction due à la troncature. Ainsi, nous définissons la quantité suivante :

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

qui est le rapport de la différence entre les deux erreurs de prévision (l'erreur de prévision due à la troncature (2.14) et l'erreur de prédiction due à la projection (2.27)). Par (2.31), nous obtenons :

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

Nous évaluons ce rapport dans le cas particulier d'un bruit fractionnaire intégrée. La figure (2.1) montre que la prédiction par troncature subit une perte plus grande de performance lorsque $d \rightarrow 1/2$. L'amélioration atteint 50 pour cent lorsque $d > 0.3$ et $k > 20$.

2.4 Prédiction à l'horizon h

Dans cette section, nous généralisons les deux méthodes de prévision à l'horizon h et nous allons essayer d'obtenir leur comportement asymptotique

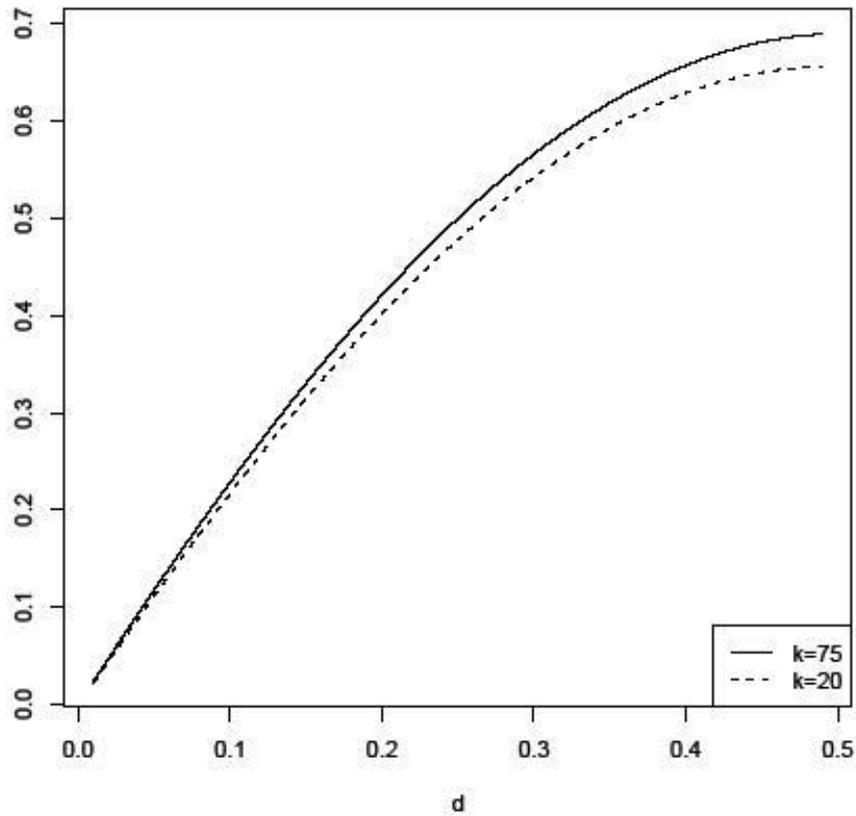


FIG. 2.1 – Le rapport $r(k)$ défini en (2.33) , $d \in]0, 1/2[$.

quand $k \rightarrow +\infty$, mais aussi quand $h \rightarrow +\infty$.

2.4.1 Prédicteur de Wiener-Kolmogorov

Puisque nous supposons que le processus $(X_n)_{n \in \mathbb{Z}}$ a une représentation autorégressive (2.2) et une représentation moyenne mobile (2.1), le prédicteur linéaire des moindres carrés, \hat{X}_{k+h} , de X_{k+h} basé sur le passé infini $(X_j, j \leq k)$ est donné par :

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

ou encore

$$\hat{X}_k(h) = - \sum_{j=1}^{h-1} a_j \hat{X}_k(h-j) - \sum_{j=h}^{+\infty} a_j X_{k+h-j}. \quad (2.34)$$

L'erreur moyenne quadratique correspondante à la prédiction est :

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

(voir théorème 5.5.1 de Brockwell et Davis (1991) [14]). Quand l'horizon h de prédiction tend vers l'infini, l'erreur moyenne quadratique de prédiction converge vers $\sigma_\varepsilon^2 \sum_{j=0}^{\infty} b_j^2$ qui est la variance du processus $(X_n)_{n \in \mathbb{Z}}$. Remarquons que l'erreur moyenne quadratique augmente plus lentement jusqu'à $\gamma(0)$ plus précisément, si nous supposons que :

$$b_j \sim j^{d-1} L'(j),$$

où L est une fonction qui varie lentement, nous pouvons exprimer le comportement asymptotique de l'erreur de prédiction. Car $j \mapsto L^2(j)$ est également une fonction qui varie lentement selon la définition de Zygmund (1986) (voir [39]), alors $b_j^2 \sim j^{2d-2} L^2(j)$ décroît. Le reste de la série et l'intégrale sont alors équivalents et nous pouvons écrire :

$$\gamma(0) - E[(\hat{X}_k(h) - X_{k+h})^2] = \sigma_\varepsilon^2 \sum_{j=h}^{+\infty} b_j^2 \quad (2.35)$$

$$\begin{aligned} &\sim \sum_{j=h}^{+\infty} j^{2d-2} L^2(j) \\ &\sim \int_h^{+\infty} j^{2d-2} L^2(j) dj \\ &\sim \frac{1}{1-2d} h^{2d-1} L^2(h). \end{aligned} \quad (2.36)$$

2.4.2 Prédicteur de Wiener-Kolmogorov tronqué

Dans la pratique, nous observons seulement un nombre fini d'échantillons. Nous supposons maintenant que nous disposons seulement de k observations (X_1, \dots, X_k) . Nous définissons alors le prédicteur de Wiener-Kolmogorov tronqué à l'horizon h d'ordre k par :

$$\hat{X}'_k(h) = - \sum_{j=1}^{h-1} a_j \hat{X}'_k(h-j) - \sum_{j=1}^k a_{h-1+j} X_{k+1-j}. \quad (2.37)$$

Erreur moyenne quadratique

Nous décrivons maintenant le comportement asymptotique de l'erreur moyenne quadratique du prédicteur défini en (2.37).

D'abord nous écrivons la différence entre la variable aléatoire de prédiction et son prédicteur :

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

Nous utilisons le processus de l'induction sur h pour montrer que :

$$\hat{X}'_k(h) - X_{k+h} = - \sum_{l=0}^{h-1} \left(\sum_{j_1+j_2+\dots+j_h=l} (-1)^{\text{card}(\{j_i, j_i \neq 0\})} a_{j_1} a_{j_2} \dots a_{j_h} \right) \varepsilon_{k+h-l}$$

$$+ \sum_{j=k+1}^{+\infty} \left(\sum_{i_1+i_2+\dots+i_h=h-1} (-1)^{\text{card}(\{i_l, i_l \neq 0, l > 1\})} a_{j+i_1} a_{i_2} \dots a_{i_h} \right) X_{k+1-j}$$

pour $h = 2$, nous avons par exemple

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

Soit $A(z)$ et $B(z)$ tels que $A(z) = 1 + \sum_{j=1}^{+\infty} a_j z^j$ et $B(z) = 1 + \sum_{j=1}^{+\infty} b_j z^j$. Puisque nous avons $A(z) = B(z)^{-1}$, nous obtenons les conditions suivantes sur les coefficients :

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

Ainsi nous obtenons :

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

Puisque le processus $(\varepsilon_n)_{n \in \mathbb{Z}}$ est non corrélé, alors les deux termes de la somme (2.38) sont orthogonales, nous pouvons réécrire l'erreur moyenne quadratique :

$$\begin{aligned} & E[(\hat{X}'_k(h) - X_{k+h})^2] \\ &= \sum_{l=0}^{h-1} b_l^2 \sigma_\varepsilon^2 + E \left[\left(\sum_{j=k+1}^{+\infty} \left(\sum_{m=0}^{h-1} a_{j+h-1-m} b_m \right) X_{k+1-j} \right)^2 \right] \quad (2.39) \end{aligned}$$

La première partie du membre droit de l'erreur (2.39) est due à la méthode de prévision et la deuxième est due à la troncature. Nous rapprochons maintenant le second terme de l'erreur en (2.39) en utilisant (2.3) et (2.4). Nous obtenons la borne supérieure suivante : $\forall \delta > 0$,

$$\left| \sum_{m=0}^{h-1} a_{j+h-1-m} b_m \right| \leq \sum_{m=1}^{h-1} |a_{j+h-1-m} b_m| + |b_0 a_{j+h-1}|$$

$$\begin{aligned}
&\leq C_1 C_2 \int_0^h (j+h-m)^{-d-1+\delta} m^{d-1+\delta} dm + C_1 (j+h)^{-d-1} \\
&\leq C_1 C_2 h^{-1+2\delta} \int_0^1 \left(\frac{j}{h} + 1 - m\right)^{-d-1+\delta} m^{d-1+\delta} dm + C_1 (j+h)^{-d-1} \\
&\leq C_1 C_2 h^{-1+2\delta} j^{-d-1+\delta} \int_0^1 \left(\frac{1}{h} + \frac{1-m}{j}\right)^{-d-1+\delta} m^{d-1+\delta} dm + C_1 (j+h)^{-d-1} \\
&\leq C_1 C_2 h^{d+\delta} j^{-d-1+\delta} \int_0^1 m^{d-1+\delta} dm + C_1 (j+h)^{-d-1} \\
&\leq C_1 C_2 \frac{h^{d+\delta}}{d} j^{-d-1+\delta}. \tag{2.40}
\end{aligned}$$

En utilisant la proposition 1, de la prédiction à l'horizon 1, on a la proposition suivante.

Proposition 3 *Soit $(X_n)_{n \in \mathbb{Z}}$ un processus linéaire stationnaire défini par (2.1), (2.2) et vérifiant les conditions (2.3) et (2.4). Alors, on a*

$$\forall \delta > 0, E[(\hat{X}'_k(h) - X_{k+h})^2] = \sum_{l=0}^{h-1} b_l^2 \sigma_\varepsilon^2 + O(h^{2d+\delta} k^{-1+\delta}). \tag{2.41}$$

2.4.3 Prédicteur linéaire des moindres carrés

Le prédicteur des moindres carrés est défini comme projection sur le passé fini (X_1, \dots, X_k) . Soit $\tilde{X}_k(h)$ le prédicteur de X_{k+h} . Il peut être écrit comme suit :

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

où le vecteur $(c_{j,k})_{1 \leq j \leq k}$ minimise l'erreur moyenne quadratique :

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

Ainsi, les coefficients $(c_{j,k})_{1 \leq j \leq k}$ satisfont la relation suivante :

$$(c_{j,k})_{1 \leq j \leq k} = -\Sigma_k^{-1} (\gamma_{h-1+j})_{1 \leq k \leq j}, \tag{2.42}$$

où Σ_k est la matrice d'autocovariance du vecteur (X_1, \dots, X_k) , c'est-à-dire :

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

Par construction de $\tilde{X}_k(h)$, l'erreur moyenne quadratique de prédiction est inférieure à $\gamma(0)$ pour tout k et h . Cependant, il peut être intéressant de chercher une condition sur k et h pour déterminer l'erreur de prédiction due à la projection négligeable par rapport à l'information (2.35) donné par le prédicteur de Wiener-Kolmogorov.

Chapitre 3

Estimation des coefficients du prédicteur tronqué

3.1 Introduction

La théorie de prévision de Wiener-Kolmogorov est l'une des méthodes les plus populaires pour prévoir la série chronologique stationnaire comme présentée simultanément dans les travaux de Wiener (1949) [35] et Kolmogorov (1941) [28] et après décrit dans Whittle (1983) [36]. Le principe est de réduire au minimum l'erreur moyenne quadratique de la prédiction en projetant le futur sur le passé infini. Dans la pratique il y a deux restrictions à ce prédicteur : nous n'avons jamais accès au passé infini et nous ne connaissons pas avec précision la structure stochastique du processus. On propose souvent deux méthodes de prédiction linéaire quand nous savons seulement un passé fini de longueur k : la troncature de la série infinie dans la définition du prédicteur de Wiener-Kolmogorov et la projection sur les données disponibles. Quand la structure stochastique du processus est inconnue, nous devons également estimer les coefficients de prédiction. Dans ce chapitre nous étudions seulement le prédicteur tronqué de Wiener-Kolmogorov.

Dans le cas de courte mémoire, les erreurs additionnelles liées à la troncature et l'estimation ont été largement étudiées. D'une part, l'erreur de prédiction due à la troncature est calculée en Pourahmadi (1989) [30] pour des processus avec une densité spectrale bornée et avec une borne inférieure strictement positive. L'erreur moyenne quadratique de la prédiction décroît exponentiellement avec la longueur du passé connu. D'autre part, l'estimation du modèle a été étudiée dans Akaike (1970) [2] dans le cas des processus autorégressifs finis et dans Bhansali (1977) [5] pour des processus avec une densité spectrale satisfaisant une condition uniforme de Lipschitz. Ils ont tra-

vallé dans un contexte non paramétrique et ont proposé une approximation de l'erreur moyenne quadratique de la prédiction. Ils ont également réduit au minimum l'erreur de prédiction sous les conditions sur la longueur de la série utilisée pour estimer les coefficients de prédiction.

Cependant les travaux mentionnés ci-dessus n'abordent pas le cas de longue mémoire. Godet (2008) a estimé l'erreur moyenne quadratique de prédiction due à la troncature pour les séries chronologiques à longue mémoire. Elle décroît seulement hyperboliquement rapidement quand k tend vers l'infini. Dans un contexte semi-paramétrique, Hidalgo et Yajima (2002) (voir [23]) ont montré que l'erreur moyenne quadratique de prédiction tronqué avec des coefficients estimés converge vers l'erreur moyenne quadratique du prédicteur de Wiener-Kolmogorov mais ne donne pas l'ordre de grandeur asymptotique.

Ici nous proposons un modèle paramétrique pour les séries chronologiques à longue mémoire et estimons les coefficients de prévision sur une réalisation de longueur T . Nous considérons l'effet de l'estimation sur l'erreur de prévision du prédicteur tronqué de Wiener-Kolmogorov. Nous estimons les coefficients de prévision avec l'estimateur de Whittle (voir Fox et Taqqu (1986) [16]). Sous la condition de simplification de laquelle les séries utilisées pour l'estimation et les séries utilisées pour la prévision sont produit de deux processus indépendants avec la même structure stochastique. Nous proposons une approximation de l'erreur moyenne quadratique de prévision quand le nombre d'observations T estime les coefficients de prévision et le nombre d'observations k utilisées dans le prédicteur réel, lorsque T et k tendent vers l'infini.

3.2 Prédicteur tronqué de Wiener-Kolmogorov

3.2.1 Méthode de prédiction

Soit $(X_n)_{n \in \mathbb{Z}}$ un processus gaussien stationnaire à temps discret dans L^2 , de moyenne nulle et de fonction d'autocovariance γ . Nous supposons que $(X_n)_{n \in \mathbb{Z}}$ est un processus à longue mémoire c'est-à-dire :

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

Nous supposons que le processus $(X_n)_{n \in \mathbb{Z}}$ admet une représentation moyenne mobile infinie comme suit

$$X_n = \sum_{j=0}^{+\infty} b_j \varepsilon_{n-j}, \quad (3.1)$$

où $(\varepsilon_n)_{n \in \mathbb{Z}}$ est une série de bruit blanc faible, de moyenne nulle et de variance σ_ε^2 . La suite $(b_j)_{j \in \mathbb{N}}$ est de carré sommable et $b_0 = 1$.

Nous supposons dans ce qui suit que $(X_n)_{n \in \mathbb{Z}}$ admet une représentation autorégressive infinie :

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

où $(a_j)_{j \in \mathbb{N}}$ sont absolument sommables et $a_0 = 1$.

Puisque nous avons les représentations (3.1) et (3.2), nous pouvons employer la théorie classique de Wiener-Kolmogorov. Le prédicteur de Wiener-Kolmogorov de X_{k+h} , noté $X_k^{WK}(h)$, est le meilleur prédicteur linéaire dans le sens des moindres carrés avec la connaissance du passé ($X_j, -\infty \leq j \leq k$). Puisque $X_k^{WK}(h)$ est la projection de X_{k+h} sur le passé infini, nous pouvons écrire $X_k^{WK}(h)$ comme

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

où $c_j(h) = -\sum_{m=0}^{h-1} a_{j+m} b_{h-1-m}$ (voir Whittle (1983) [36]). Dans le cas particulier d'une prévision d'un pas ($h = 1$) l'expression de $c_j(1)$ est simplifiée et nous prenons pour tout $j \in \mathbb{N}$, $c_j(1) = -a_j$. Dans la pratique, nous observons seulement un nombre fini d'échantillon (X_1, \dots, X_k) . La somme infinie définie en (3.3) est tronquée à l'ordre k et nous avons le nouveau prédicteur suivant

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

Nous étudions ensuite le comportement asymptotique de l'erreur moyenne quadratique de la prévision qui est liée à l'estimation des coefficients de prédiction $c_j(h)$. La section 3.2.2 aborde le cas particulier $h = 1$. Nous généralisons alors notre résultat pour la prévision à l'horizon h dans la section 3.2.3.

3.2.2 Estimation des coefficients et de l'erreur moyenne quadratique associées à la prédiction d'un pas

Sous les hypothèses de la section 3.2.1, nous estimons l'erreur moyenne quadratique entre le prédicteur $\hat{X}_k(1)$ définie en (3.4) et le prédicteur $\hat{X}_{T,k}(1)$ définie par

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

où \hat{a}_j sont des estimateurs des a_j calculés en utilisant une réalisation $(Y_j)_{1 \leq j \leq T}$ de longueur T qui est indépendant du processus $(X_n)_{n \in \mathbb{Z}}$ mais possède la même structure stochastique. Plus précisément, dans une approche paramétrique, nous supposons que le processus $(X_n)_{n \in \mathbb{Z}}$ a une densité spectrale $f(\cdot, \theta)$ qui appartient à une classe des fonctions paramétrées par θ .

Soit θ_0 la valeur vraie du paramètre. Nous estimons les coefficients autorégressifs $(a_j(\theta_0))_{1 \leq j \leq k}$ avec $\hat{a}_j = a_j(\hat{\theta}_T)$ où $\hat{\theta}_T$ est l'estimateur de Whittle de θ_0 . Nous définissons l'estimateur de Whittle (voir Fox et Taqqu (1986) [16]) :

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

où I_T est le périodogramme

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

En conséquence notre prédicteur est donné par

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

où $\hat{\theta}_T$ est défini en (3.5).

Quand le processus est une série chronologique gaussienne à longue mémoire (voir Fox et Taqqu (1986) [16]), les conditions suffisantes de norme sur la famille paramétrique des densités spectrales assurent convergence presque sûre de l'estimateur $\hat{\theta}_T$ vers le paramètre vrai θ_0 . Celles-ci sont rappelées ci-dessous. Nous définissons la famille paramétrique de la densité spectrale $F_{\Theta} = \{f(\cdot, \theta), \theta \in \Theta \text{ compact}\}$. Si θ et θ' sont des éléments distincts de Θ , nous supposons que l'ensemble $\{\lambda : f(\lambda, \theta) \neq f(\lambda, \theta')\}$ a la mesure de Lebesgue.

Pour tout $f(., \theta) \in F_\Theta$ il existe une fonction continue $0 < \alpha(\theta) < 1$ et des constantes $C(\delta)$ et $C_0(\delta)$ telles que pour chaque $\delta > 0$, on a

B0. $f(\lambda, \theta) = |\lambda|^{-\alpha(\theta)}L(\lambda, \theta)$ avec $L(\lambda, \theta)$ bornée, tend vers 0 et différentiable en 0. Alors nous avons :

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

Cette condition garantit que les processus avec la densité spectrale dans F_Θ admettent une représentation autorégressive infinie et moyenne mobile infinie. Nous notons par $a_j(\theta)$ les coefficients de la représentation autorégressive et nous supposons qu'ils vérifient

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

Nous supposons également que les fonctions $f(\lambda, \theta)$ sont normalisées de sorte que

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

B1. $f(\lambda, \theta)$ est continue en tout point $(\lambda, \theta) \in ([-\pi, \pi[-\{0\}] \times \Theta$ et

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

B2. $\frac{\partial}{\partial \theta_j} f(\lambda, \theta)$, $\frac{\partial^2}{\partial \theta_j \partial \theta_l} f(\lambda, \theta)$ et $\frac{\partial^3}{\partial \theta_j \partial \theta_l \partial \theta_k} f(\lambda, \theta)$ sont continues en tout $(\lambda, \theta) \in ([-\pi, \pi[-\{0\}] \times \Theta$ et

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

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

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

B3. $\frac{\partial}{\partial \lambda} f(\lambda, \theta)$, $\frac{\partial^2}{\partial \lambda \partial \theta_j} f(\lambda, \theta)$, $\frac{\partial^3}{\partial \lambda \partial \theta_j \partial \theta_l} f(\lambda, \theta)$, $\frac{\partial^4}{\partial \lambda \partial \theta_j \partial \theta_l \partial \theta_k} f(\lambda, \theta)$ et $\frac{\partial^3}{\partial \lambda^2 \partial \theta_j} f(\lambda, \theta)$ sont continues en tout point $(\lambda, \theta) \in [-\pi, \pi[-\{0\}] \times \Theta$ et

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

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

$$\begin{aligned} \forall 1 \leq j, l \leq p, & \left| \frac{\partial^3}{\partial \lambda \partial \theta_j \partial \theta_l} f(\lambda, \theta) \right| \leq C(\delta) |\lambda|^{-\alpha(\theta)-1-\delta}, \\ \forall 1 \leq j, l, k \leq p, & \left| \frac{\partial^4}{\partial \lambda \partial \theta_j \partial \theta_l \partial \theta_k} f(\lambda, \theta) \right| \leq C(\delta) |\lambda|^{-\alpha(\theta)-1-\delta}, \\ \forall 1 \leq j \leq p, & \left| \frac{\partial^3}{\partial^2 \lambda \partial \theta_j} f(\lambda, \theta) \right| \leq C(\delta) |\lambda|^{-\alpha(\theta)-2-\delta}. \end{aligned}$$

Les constantes $C(\delta)$ et $C_0(\delta)$ qui apparaissent dans les conditions B0-B3 sont exigées pour être indépendant de θ .

Nous pouvons maintenant exprimer le comportement asymptotique de l'erreur moyenne quadratique de prédiction due à l'estimation des coefficients de prévision.

Théorème 2 *Soit F_Θ une famille des densités spectrales qui vérifient les conditions B0-B3. Soit $(X_n)_{n \in \mathbb{Z}}$ une suite gaussienne à longue mémoire stationnaire de moyenne nulle et de densité spectrale $f(\lambda, \theta_0) \in F_\Theta$ avec θ_0 à l'intérieur de Θ . Soit $(Y_n)_{n \in \mathbb{Z}}$ une réalisation indépendante de $(X_n)_{n \in \mathbb{Z}}$. $\hat{\theta}_T$ est l'estimateur de θ_0 calculé sur $(Y_n)_{n \in \mathbb{Z}}$ selon (3.5).*

Alors l'erreur additionnelle due à l'estimation des coefficients de prédiction vérifie qu'il existe une constante C indépendante de k tel que

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

où $\hat{X}_{T,k}(1)$ est défini en (3.7).

Preuve. Pour démontrer le Théorème 2, nous avons besoin du Lemme suivant.

Lemme 2 *Sous les conditions du théorème 2, les coefficients de la représentation autorégressive infinie de a_j définis en (3.2) vérifient :*

- (i) $a_j(\theta)$ est uniformément borné c'est-à-dire : il existe C tel que $\forall j \in \mathbb{N}^*, \forall \theta \in \Theta$;

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

- (ii) Pour tout $\delta > 0$, il existe C tel que : $\forall j \in \mathbb{N}^*, \forall 1 \leq l \leq m$ et $\forall \theta \in \Theta$;

$$\left| \frac{\partial a_j}{\partial \theta_l}(\theta) \right| \leq C j^{-1+\delta}. \quad (3.10)$$

- (iii) La deuxième et la troisième dérivés de a_j en ce qui concerne le paramètre θ sont uniformément bornés c'est-à-dire : il existe C tel que $\forall j \in N, \forall 1 \leq k, l, n \leq m$ et $\forall \theta \in \Theta$;

$$\left| \frac{\partial^2 a_j}{\partial \theta_k \partial \theta_l}(\theta) \right| \leq C \text{ et } \left| \frac{\partial^3 a_j}{\partial \theta_k \partial \theta_l \partial \theta_n}(\theta) \right| \leq C. \quad (3.11)$$

- (iv) La première et la deuxième dérivés de a_j en ce qui concerne le paramètre θ sont continues.

Preuve. (voir Godet (2007b) [20]). •

Nous définissons les vecteurs :

$$\alpha_{k,T} = \left(a_1(\hat{\theta}_T) - a_1(\theta_0), \dots, a_k(\hat{\theta}_T) - a_k(\theta_0) \right)'_1 = (X_k, \dots, X_1)',$$

où v' est le vecteur transposé de v .

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

Puisque les réalisations $(X_j)_{j=\overline{1,k}}$ et $(Y_j)_{j=\overline{1,T}}$ sont indépendantes,

$$\begin{aligned} E\left(\hat{X}_{T,k}(1) - \hat{X}_k(1)\right)^2 &= \text{trace}\left(E\left(\alpha_{k,T} \alpha'_{k,T}\right) E\left(X_1^k (X_1^k)'\right)\right) \\ &= \text{trace}\left(E\left(\alpha_{k,T} \alpha'_{k,T} \Sigma_k\right)\right), \end{aligned}$$

où Σ_k est la matrice commune de covariance des processus $(X_n)_{n \in Z}$ et $(Y_n)_{n \in Z}$

$$\Sigma_k = (\gamma(i-j))_{1 \leq i, j \leq k}. \quad (3.12)$$

Notre preuve est composée de deux étapes suivantes :

Étape 1 Nous donnons une approximation à la suite de l'erreur moyenne quadratique de chaque élément de $E(\alpha_{k,T}\alpha'_{k,T})$ dans F onction de T .

Étape 2 Nous récrivons la matrice $E(\alpha_{k,T}\alpha'_{k,T})$ en utilisant l'approximation obtenue dans l'étape 1, après nous étudions la limite quand T tend vers l'infini dans F onction de k .

Étape 1 Nous allons étudier les éléments de la matrice de covariance des coefficients estimés $E(\alpha_{k,T}\alpha'_{k,T})$. Nous pouvons écrire les éléments $\left(E(\alpha_{k,T}\alpha'_{k,T})\right)_{i,j}$ comme $E(g_{i,j}\hat{\theta}(T))$ où

$$g_{i,j} = (a_i(\theta) - a_i(\theta_0))(a_j(\theta) - a_j(\theta_0)). \quad (3.13)$$

Alors nous utilisons le développement dans S éries de Taylor d'ordre 2 de $g_{i,j}$ et appliquons le Théorème 5.4.3 de Fuller (1976) (voir [18]). Nous avons, sous les conditions suivantes :

1. $\forall 1 \leq l \leq m, E(|\hat{\theta}_{T,l} - \theta_{0,l}|^3) = O(\eta(T))$ où $\hat{\theta}_{T,l}$ est le $l^{\text{ème}}$ élément de $\hat{\theta}(T)$ et $\eta(T)$ tend vers 0 quand T tend vers l'infini.
2. $\hat{\theta}(T) \rightarrow \theta_0, P$ et p.s si $T \rightarrow +\infty$.
3. $g_{i,j}$ est uniformément borné sur un voisinage de θ_0 .
4. La première et la deuxième dérivés de $g_{i,j}$ sont continues et bornées sur un voisinage de θ_0 .

Nous avons :

$$\begin{aligned} E\left(g_{i,j}(\hat{\theta}_T)\right) &= g_{i,j}(\theta_0) + \sum_{l=1}^m E(\hat{\theta}_{T,l} - \theta_{0,l}) \frac{\partial g_{i,j}}{\partial \theta_l}(\theta_0) \\ &+ \frac{1}{2} \sum_{l=1}^m \sum_{n=1}^m \frac{\partial^2 g_{i,j}}{\partial \theta_l \partial \theta_n}(\theta_0) E\left((\hat{\theta}_{T,l} - \theta_{0,l})(\hat{\theta}_{T,n} - \theta_{0,n})\right) \\ &+ \frac{1}{6} \sum_{k=1}^m \sum_{l=1}^m \sum_{n=1}^m E\left(\frac{\partial^3 g_{i,j}}{\partial \theta_l \partial \theta_n \partial \theta_k}(\theta^*)(\hat{\theta}_{T,l} - \theta_{0,l})(\hat{\theta}_{T,n} - \theta_{0,n})(\hat{\theta}_{T,k} - \theta_{0,k})\right), \end{aligned} \quad (3.14)$$

où $\theta^* = t_T \theta_0 + (1-t_T)\hat{\theta}_T$ avec $t_T \in [0, 1]$. D'abord nous vérifions les conditions (2)-(4). Sous les conditions sur la densité spectrale et ses dérivés, le Théorème 1 de Fox et Taquq (1986) (voir [16]) est vérifié et assure la convergence (2). Par les propositions (i) et (ii) du Lemme 2, les conditions (3) et (4) sont vérifiées.

Nous énonçons maintenant un Lemme qui nous permet d'appliquer le Théorème de Fuller et de prouver (1).

Lemme 3 *Sous les conditions du Théorème 2, nous avons :*

a) $\forall 1 \leq j, l \leq m$

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

où la matrice W est définie par :

$$W_{i,j} = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \theta_i} f^{-1}(\lambda, \theta_0) \right) \left(\frac{\partial}{\partial \theta_j} f^{-1}(\lambda, \theta_0) \right) f^2(\lambda, \theta_0) d\lambda. \quad (3.15)$$

b) $\forall 1 \leq j \leq m$

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

Preuve. (voir Godet (2007b) [20]). •

D'après le Lemme 3.b., on a la condition (1) qui est vérifiée en p en prenant $\eta(T) = T^{-3/2}$. Selon le théorème de Fuller, on peut appliquer le développement de Taylor à (3.14). L'expression (3.14) peut être donc simplifiée. En effet, nous remarquons par la définition de la fonction $g_{i,j}$ dans (3.13) que :

$$g_{i,j}(\theta_0) = 0$$

et

$$\forall 1 \leq l \leq p, \quad \frac{\partial g_{i,j}}{\partial \theta_l}(\theta_0) = 0.$$

D'après le Lemme 3.a., nous avons :

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

Ensuite d'après (i), (iii) et (iv) du Lemme 2 et (b.) du Lemme 3

$$\begin{aligned} \sum_{k=1}^m \sum_{l=1}^m \sum_{n=1}^m E\left(\frac{\partial^3 g_{i,j}}{\partial \theta_l \partial \theta_n \partial \theta_k}(\theta^*) (\hat{\theta}_{T,l} - \theta_{0,l})(\hat{\theta}_{T,n} - \theta_{0,n})(\hat{\theta}_{T,k} - \theta_{0,k}) \right) \\ \leq 8C^2 m^3 O(\eta(T)). \end{aligned} \quad (3.16)$$

Enfin nous obtenons l'approximation suivante : pour tout $\delta > 0$

$$\lim_{T \rightarrow +\infty} TE \left(g_{i,j}(\hat{\theta}(T)) \right) \quad (3.17)$$

$$\begin{aligned} &= \lim_{T \rightarrow +\infty} TE \left((a_i(\hat{\theta}_T) - a_i(\theta_0))(a_j(\hat{\theta}_T) - a_j(\theta_0)) \right) \\ &= \lim_{T \rightarrow +\infty} T \sum_{l=1}^m \sum_{n=1}^m \frac{\partial^2 g_{i,j}}{\partial \theta_l \partial \theta_n}(\theta_0) E \left((\hat{\theta}_{T,l} - \theta_{0,l})(\hat{\theta}_{T,n} - \theta_{0,n}) \right) \\ &\quad + \lim_{T \rightarrow +\infty} TO(T^{-2+\delta}) \\ &= 4\pi \sum_{l=1}^m \sum_{n=1}^m \frac{\partial^2 g_{i,j}}{\partial \theta_l \partial \theta_n}(\theta_0) (W_{l,n}^{-1}). \end{aligned} \quad (3.18)$$

Étape 2 Nous remplaçons maintenant l'approximation de chaque élément de $E(\alpha_{k,T} \alpha'_{k,T})$.

Nous étudions d'abord la limite donnée en (3.18). Puisque W^{-1} est symétrique :

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

où

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

W donnée par (3.15) est définie positive ainsi que W^{-1} , avec

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

où $P = (p_{ij})_{1 \leq i, j \leq m}$ est une matrice orthogonale et $(\lambda_i)_{1 \leq i \leq m}$ sont les valeurs propres de W^{-1} , qui sont positives. Nous pouvons réécrire notre expression comme :

$$\begin{aligned} DW^{-1}D' &= \sum_{r=1}^m \left(\left[\sqrt{\lambda_r} \sum_{l=1}^m p_{rl} \frac{\partial a_i}{\partial \theta_l}(\theta_0) \right] \left[\sqrt{\lambda_r} \sum_{l=1}^m p_{rl} \frac{a_j}{\partial \theta_l}(\theta_0) \right] \right)_{1 \leq i, j \leq k} \\ &= \sum_{r=1}^m \beta'_r \beta_r, \end{aligned} \quad (3.19)$$

avec β_r est le vecteur $(\sqrt{\lambda_r} \sum_{l=1}^m p_{rl} \frac{\partial a_i}{\partial \theta_l}(\theta_0))_{1 \leq i \leq k}$.

En utilisant (3.18) nous trouvons l'expression suivante $E(\alpha_{k,T} \alpha'_{k,T})$, pour tout $\delta > 0$ et pour tout $k, T > 0$

$$\lim_{T \rightarrow +\infty} [T E(\alpha_{k,T} \alpha'_{k,T})] = 8\pi \sum_{r=1}^m \beta'_r \beta_r. \quad (3.20)$$

Nous prenons pour tout $\delta > 0$:

$$\lim_{T \rightarrow +\infty} [T \text{trace}(E(\alpha_{k,T} \alpha'_{k,T}) \Sigma_k)] = 8\pi \text{trace} \left(\sum_{r=1}^m \beta'_r \beta_r \Sigma_k \right). \quad (3.21)$$

Nous étudions la limite définie en (3.21) et recherchons une borne inférieure et supérieure de

$$\text{trace} \left(\sum_{r=1}^m \beta'_r \beta_r \Sigma_k \right) = \sum_{r=1}^m \beta_r \Sigma_k \beta'_r. \quad (3.22)$$

Nous avons

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

où Λ_k^- est la plus petite valeur propre de Σ_k et Λ_k^+ sa plus grande valeur propre.

Nous estimons maintenant la borne supérieure. En utilisant (3.10) du Lemme 2, nous pouvons estimer $\|\beta_r\|_2^2$.

Soit $\delta < 1/2$, ils existent C_1, \dots, C_m tels que

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

où $C_r(\theta_0)$ ne dépend pas de k .

De Boettcher et Virtanen (2006) (voir [11]), la norme d'une matrice de covariance associée à une densité spectrale de la forme $\lambda \mapsto \lambda^{-\alpha} L(\lambda)$ où L est borné, continue en 0, est équivalente à Ck^α avec C une constante. Par conséquent $\Lambda_k^+ \leq Ck^{\alpha(\theta_0)}$. Nous montrons alors qu'il existe une constante C telle que

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

Essayons maintenant de trouver une borne inférieure pour (3.22). Alors la densité spectrale $f(\cdot, \theta_0)$ est bornée loin de 0 (de la condition B0), la valeur propre inférieure Λ_k^- a une borne inférieure uniforme (voir par exemple Grenander et Szegö (1958) [22]). Il existe alors une constante positive C' telle que pour tout k ,

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

De plus $\forall 1 \leq r \leq m$,

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

Ainsi il existe une constante positive C' , ne dépend pas de k , telle que

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

Nous avons alors l'inégalité suivante pour le premier terme de (3.21)

$$\frac{C'}{T} \leq \frac{\text{trace}\left(\sum_{r=1}^m \beta'_r \beta_r \Sigma_k\right)}{T} \leq C \frac{k^{\alpha(\theta_0)}}{T} \sum_{r=1}^m C_r(\theta_0). \quad (3.23)$$

Enfin nous concluons en utilisant (3.21) et la deuxième inégalité en (3.23)

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

Ceci conclut la preuve. •

3.2.3 Prédiction à l'horizon h

Plus généralement, pour la prévision à l'horizon h , les coefficients sont données par :

$$c_j(h) = - \sum_{m=0}^{h-1} a_{j+m} b_{h-1-m}. \quad (3.24)$$

En utilisant une approche paramétrique, nous pouvons considérer $c_j(h)$ dans F onction du paramètre θ , pour tout h et nous estimons θ par l'estimateur de Whittle. Sous les mêmes conditions, nous pouvons prouver le même résultat pour la prédiction à l'horizon h quant à la prédiction d'un pas.

Théorème 3 *Supposons que les conditions Théorème 2 sont satisfaites et que les coefficients de la représentation moyenne mobile sont uniformément bornés c'est-à-dire qu'il existe une constante C telle que $\forall(\theta, j) \in \Theta \times N$ nous avons $|b_j(\theta)| \leq C$.*

Alors l'erreur additionnelle due à l'estimation des coefficients de prédiction vérifie la condition

il existe une constante C indépendante de k telle que :

$$\forall k \in N^*, \lim_{T \rightarrow +\infty} \left[T E \left(\hat{X}_{T,k}(h) - \hat{X}_k(h) \right)^2 \right] \leq C k^{2d},$$

où

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

Preuve. La preuve de ce théorème est basée sur le fait que les coefficients $c_j(h)$ vérifient également le Lemme 2 sous les conditions du Théorème 2 et si les coefficients b_j sont uniformément bornés.

Avec le Théorème 2, nous savons le comportement asymptotique de l'erreur moyenne quadratique due à l'estimation des coefficients c'est-à-dire : $O(\frac{k^{2d}}{T})$ quand, le nombre d'observations pour l'estimation, T tend vers l'infini. •

Chapitre 4

Estimation des coefficients du prédicteur projeté

4.1 Introduction

Nous allons maintenant nous intéresser au prédicteur projeté sur le sous-espace linéaire fermé engendré par $\{X_1, \dots, X_k\}$. Ainsi la prédiction à l'horizon 1 s'écrit

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

Comme dans le chapitre précédent, nous allons également estimer l'erreur quadratique de prédiction supplémentaire due à l'estimation des coefficients. On suppose ici que l'on dispose de deux familles indépendantes du même processus : $(X_j)_{1 \leq j \leq k}$ la version que l'on veut prédire et $(Y_j)_{1 \leq j \leq T}$, la version qui va nous permettre d'estimer les coefficients du prédicteur.

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

$$\forall 1 \leq j \leq k, \sum_{i=1}^k a_{i,k} \gamma(i-j) = -\gamma(j),$$

les covariances par les covariances empiriques calculées sur la base des données $(Y_j)_{1 \leq j \leq T}$:

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

On utilise ensuite l'algorithme de Durbin-Levinson ou l'algorithme d'innovation (voir Brockwell et Davis (1991) [14]) pour résoudre le système.

On note $(\hat{a}_{1,k}, \dots, \hat{a}_{k,k})$ les coefficients obtenus dans R éolvant les équations de

Yule-Walker où les covariances sont remplacées par les covariances empiriques données par (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

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

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

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

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

avec

$$\Sigma_k = E \left(\begin{pmatrix} X_k \\ \vdots \\ X_1 \end{pmatrix} \begin{pmatrix} X_k, \dots, X_1 \end{pmatrix} \right).$$

Il nous faut donc évaluer tout d'abord

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

Pour cela on définit le vecteur

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

la matrice de taille $k \times k$

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

et on énonce le Théorème suivant.

Pour cela,

On considère $g_{i,j}$ la fonction définie par

$$\begin{aligned} g_{i,j} : R^{k+1} &\longrightarrow R \\ (x_0, \dots, x_k) &\longmapsto (y_i - a_{i,k})(y_j - a_{j,k}), \end{aligned} \quad (4.2)$$

où

$$\begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = - \begin{pmatrix} x_0 & x_1 & \dots & x_{k-1} \\ x_1 & x_0 & \ddots & x_{k-2} \\ \vdots & \ddots & \ddots & \vdots \\ x_{k-1} & x_{k-2} & \dots & x_0 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$$

Et H la matrice définie comme suit ;

on pose h la fonction définie par

$$h(\lambda) = \left| 1 - \sum_{r=1}^k a_{r,k} e^{ir\lambda} \right|^2$$

et on note par $h^{(r)}$ la dérivée de la fonction h par rapport à la variable $a_{r,k}$. L'élément de la i -ème ligne et la j -ème colonne de la matrice H est alors donné par la formule suivante :

$$\forall 1 \leq i, j \leq k, H_{i,j} = \int_{-\pi}^{\pi} h^{(i)}(\lambda) h^{(j)}(\lambda) f^2(\lambda) d\lambda. \quad (4.3)$$

Théorème 4 *Supposons que le processus $(Y_n)_{n \in \mathbb{Z}}$ est gaussien tel que sa fonction de covariance γ et les coefficients de sa représentation en moyenne mobile infinie b_j vérifient*

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

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

et que le bruit blanc qui lui est associé $(\varepsilon_n)_{n \in \mathbb{Z}}$ dans la représentation moyenne mobile a des moments d'ordre 4 finis.

Alors

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

où $g_{i,j}$ est définie par 4.2, C et D sont des constantes indépendantes de n et k et H est définie en 4.3.

Preuve. (voir Godet (2007b) [20]). •

On revient maintenant à l'estimation du comportement de

$E\left[\left(\tilde{X}_{T,k}(1) - \tilde{X}_k(1)\right)^2\right]$ et on énonce le Théorème suivant qui traite le cas $d \geq 1/4$

Théorème 5 *Supposons que les hypothèses du Théorème 4 sont vérifiées et que la densité spectrale du processus est de la forme*

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

où f_d la densité spectrale du processus fractionnaire définie par :

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

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 ,

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

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

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

où C est une constante indépendante de k .

Preuve. En appliquant le Théorème précédent, on obtient que si $d = 1/4$ alors

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

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

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

On commence par contrôler $|\sum_{j=0}^k a_{j,k}|$ 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) (voir [27]) sur la convergence des coefficients de l'AR(k) vers les coefficients de l'AR(∞). On note qu'il existe C_1, C_2 et K tels que si :

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

et si on note C une constante générique :

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

On a alors si $k \geq K$

$$\begin{aligned} \sum_{j=1}^k (a_{j,k} - a_j)^2 &\leq \frac{C}{k^2} \sum_{j=1}^k \left[\left(\sum_{u=k-j}^{+\infty} |a_u| \right)^2 + \left(\sum_{u=j}^{+\infty} |a_u| \right)^2 \right]^2 \\ &\leq \frac{C}{k^2} \sum_{j=1}^k (k-j+1)^{-2d} + j^{-2d} \\ &\leq O(k^{-2d-1}). \end{aligned} \tag{4.4}$$

Ceci nous permet de conclure que :

$$\begin{aligned} \left| \sum_{j=0}^k a_{j,k} \right| &\leq O(k^{-d}) + \sum_{j=0}^{+\infty} |a_j| \\ &= O(1). \end{aligned} \tag{4.5}$$

Il nous reste donc à étudier le comportement de

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

On applique alors le Théorème 6.1 de Adenstedt (1974) (voir [1]) qui sous les hypothèses du Théorème 5 nous donne l'équivalence suivante

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

où Γ et β sont respectivement les fonctions Gamma et Beta. Ceci nous permet de conclure.●

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

Théorème 6 *Supposons que les hypothèses du Théorème 4 sont vérifiées et que la densité spectrale est bornée inférieurement par une constante strictement positive. Si $0 < d < 1/4$ alors*

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

où C est une constante indépendante de k .

Preuve. On note $(\phi_i = \Phi_{i,1} + \dots + \Phi_{i,i} x^{i-1})_{i \in N^*}$ les polynômes orthogonaux associés à la densité spectrale f c'est-à-dire Φ_i est de degré $i - 1$ pour tout $i \in N^*$ et

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

où δ est le symbole de Kronecker. On construit la matrice T_k :

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

T_k ainsi définie vérifie les égalités suivantes

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

Donc, on a

$$\Sigma_k^{-1} = T_k^* T_k. \quad (4.6)$$

En utilisant l'égalité (4.6), on est ramené à l'étude de la limite suivante

$$\begin{aligned} \lim_{T \rightarrow +\infty} \left(T E \left[\left(\tilde{X}_{T,k}(1) - \tilde{X}_k(1) \right)^2 \right] \right) &= \text{trace}(\Sigma_k^{-1} H) \\ &= \text{trace}(T_k^* H T_k), \end{aligned} \quad (4.7)$$

où H est la matrice définie dans le Théorème 4. On note

$G_k : \lambda \mapsto \sum_{j=0}^k a_{j,k} e^{ij\lambda}$ et $\lim_{T \rightarrow +\infty} \left(T E \left[\left(\tilde{X}_{T,k}(1) - \tilde{X}_k(1) \right)^2 \right] \right)$ se réécrit alors :

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

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

$$\begin{aligned} A &= \left(\int_{-\pi}^{\pi} f^2(\lambda) \text{Re} \left(G_k(\lambda) \Phi_j(e^{ij\lambda}) \right) \text{Re} \left(G_k(-\lambda) \Phi_l(e^{-il\lambda}) \right) d\lambda \right)_{1 \leq j, l \leq k} \\ B &= \left(\int_{-\pi}^{\pi} f^2(\lambda) \text{Re} \left(G_k(\lambda) \Phi_j(e^{ij\lambda}) G_k(-\lambda) \Phi_l(e^{-il\lambda}) \right) d\lambda \right)_{1 \leq j, l \leq k} \end{aligned}$$

$$\begin{aligned}
&= \left(\int_{-\pi}^{\pi} f^2(\lambda) |G_k(\lambda)|^2 \Phi_j(e^{ij\lambda}) \Phi_l(e^{-il\lambda}) d\lambda \right)_{1 \leq j, l \leq k}, \\
C &= \left(\int_{-\pi}^{\pi} f^2(\lambda) \operatorname{Im} \left(G_k(\lambda) \Phi_j(e^{ij\lambda}) \right) \operatorname{Im} \left(G_k(-\lambda) \Phi_l(e^{-il\lambda}) \right) d\lambda \right)_{1 \leq j, l \leq k}.
\end{aligned}$$

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 des matrices A , B et C . 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 :

$$0 \leq \operatorname{trace}(A) \leq \operatorname{trace}(B). \quad (4.8)$$

On s'intéresse 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

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

On a supposé précédemment que la densité spectrale f est bornée inférieurement par une constante strictement positive c , on peut donc appliquer le Théorème 2.2.4 de Simon (2005) (voir [34]) 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 $|G_k(\lambda)|^2, \forall \lambda \in [-\pi, \pi]$,

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

comme on l'a démontré en (4.5). Cette majoration est indépendante de λ . 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 (4.7) et de (4.8) que :

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

•

Les résultats obtenus dans les Théorèmes 5 et 6 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 dans 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 5 et 6 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.

Chapitre 5

Prédiction et estimation du modèle sur une même réalisation

5.1 Introduction

Considérons $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de moyenne zéro et de variance finie. Nous souhaitons prédire la valeur de X_{n+1} à partir des observations de (X_1, \dots, X_n) en utilisant un prédicteur linéaire c'est-à-dire : une combinaison linéaire des données observées. D'abord nous définissons les coefficients du prédicteur optimal dans le sens des moindres carrés dans \mathcal{S} supposant que la fonction de covariance est connue et nous sommes amenés ainsi à estimer les coefficients de la combinaison linéaire de (X_1, \dots, X_n) . Cette deuxième étape est souvent réalisée sous l'hypothèse restrictive suivante : nous prévoyons d'autres séries indépendantes futures avec exactement la même structure probabiliste ; la série observée est seulement utilisée pour estimer les coefficients de prédiction (voir par exemple Bhansali (1987) [6] et Lewis et Reinsel (1985) [29] ou Godet (2007b) [20]). Cette hypothèse facilite l'analyse mathématique puisque le problème de prédiction peut être réduit à un problème d'estimation des coefficients de prédiction par le traitement sur le processus, que nous prévoyons. Mais le praticien a rarement deux séries indépendantes : une pour estimer le modèle et une pour prévoir. Il doit estimer les coefficients de prédiction sur la même réalisation. Dans la suite nous nous concentrons sur ce cas appelé la prédiction de même-réalisation. L'exécution du prédicteur dépend de deux paramètres : la dimension du sous-espace sur lequel nous projetons et le nombre de données disponibles pour estimer les coefficients de prédiction. Pour réduire l'erreur de prédiction, il est

raisonnable d'augmenter la dimension de l'espace sur lequel nous projetons, puisqu'on dispose dans ce cas d'un grand nombre d'observations. Mais quand la dimension et le nombre de coefficients estimés augmentent, l'estimation de ces coefficients devient plus difficile et peut affecter l'erreur moyenne quadratique.

Ce chapitre est organisé comme suit. Dans les sections 5.2 et 5.3 nous généralisons les résultats d'Ing et Wei (2003) (voir [24]) pour trouver une expression asymptotique de l'erreur moyenne quadratique pour une série chronologique à longue mémoire. Dans la dernière section, nous prouvons un théorème central limite. Plus précisément, nous prouvons la convergence dans la distribution de la différence normalisée entre notre prédicteur et le prédicteur de Wiener-Kolmogorov, qui est le prédicteur des moindres carrés sachant tout le passé.

Définition du prédicteur

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de moyenne nulle et de variance finie. Nous supposons que la fonction d'autocovariance du processus est connue. Notre but est de prévoir la valeur de X_{n+1} , en utilisant les k précédentes données observées. Le prédicteur linéaire optimal est défini comme étant projection sur le sous-espace fermé de l'espace de Hilbert engendré par $\{X_n, \dots, X_{n-k+1}\}$ muni du produit scalaire $\langle X, Y \rangle = E(X'Y)$ où X' est la transposée du vecteur X . C'est le prédicteur de moindres carrés sachant (X_{n-k+1}, \dots, X_n) . On note ce prédicteur par $\hat{X}_{n+1}(k)$ et par $-a_{j,k}$ les coefficients théoriques de prédiction c'est-à-dire

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

Ils sont donnés par (voir Brockwell et Davis (1988) [13] Section 5.1)

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

où $\Sigma(k)$ est la matrice de covariance du vecteur (X_1, \dots, X_k) .

Estimation des coefficients de prédiction

L'estimation est construite sur la base des n dernières observations (X_n, \dots, X_1) et notre prédicteur est la projection sur les k dernières observations ($k \leq n$). La matrice de covariance est estimée par

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

où

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

et K_n est la dimension maximum du sous-espace sur lequel nous projetons c'est-à-dire : nous étudions la famille des prédicteurs $(\hat{X}_{n+1}(k))_{1 \leq k \leq K_n}$. K_n est une suite croissante de nombres entiers qui peut être bornée ou peut croître vers l'infini.

Les coefficients de prédiction $a_{i,k}$ sont estimés à partir de (X_1, \dots, X_n) par

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

Le prédicteur d'un pas est :

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

Dans ce travail nous utilisons C pour noter les constantes positives génériques qui sont indépendantes de la dimension de l'échantillon n .

Les hypothèses suivantes sur le processus $(X_t)_{t \in \mathbb{Z}}$ sont essentielles pour les résultats présentés ici.

Il existe $d \in]0, 1/2[$ tel que :

- H.1 Le processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est gaussien et admet une représentation moyenne mobile infinie et une représentation autorégressive infinie comme suit :

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

avec $a_0 = b_0 = 1$, pour tout $j \geq 1$ et pour tout $\delta > 0$, $|a_j| \leq C j^{-d-1+\delta}$ et $|b_j| \leq C j^{d-1+\delta}$ et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un processus de bruit blanc. Ces hypothèses sur les coefficients sont vérifiées pour les deux processus à longue mémoire et à courte mémoire ;

- H.2 La covariance $\gamma(k)$ est équivalente à $L(k)k^{2d-1}$ quand k tend vers l'infini, où L est une fonction à variation lente. Dans cette hypothèse les autocovariances ne sont pas absolument sommables et le processus est donc un processus à longue mémoire ;
- H.3 La densité spectrale du processus $(X_t)_{t \in \mathbb{Z}}$ existe et a une borne inférieure strictement positive ;
- H.4 Les coefficients $(a_j)_{j \in \mathbb{N}}$ vérifient :

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

où L est une fonction à variation lente.

5.2 Moment borné

Dans cette section, nous établissons des moments bornés pour la matrice inverse de covariance d'échantillon et nous appliquons ces résultats pour obtenir la vitesse de convergence de $\hat{\Sigma}_n(k)$ vers $\Sigma(k)$.

On note par $\lambda_{\min}(Y)$ et $\lambda_{\max}(Y)$ respectivement la plus petite et la plus grande valeurs propres de la matrice Y . On munit l'ensemble des matrices de la norme

$$\|Y\|^2 = \lambda_{\max}(Y'Y). \quad (5.7)$$

Pour une matrice symétrique, cette norme est égale au rayon spectral et pour un vecteur (X_1, \dots, X_n) , elle est égale à $\sqrt{\sum_{i=1}^n X_i^2}$. Cette norme est une norme de matrice qui vérifie pour toutes les matrices A et B :

$$\|AB\| \leq \|A\| \|B\|. \quad (5.8)$$

Lemme 4 *Soit $(K_n)_{n \in \mathbb{N}}$ une suite croissante de nombres entiers positifs satisfaisant $K_n = o(\sqrt{n})$. Supposons (H.1). Puis, pour tout $q > 0$, pour tout $\theta > 0$ et pour tout $1 \leq k \leq K_n$,*

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

où $\hat{\Sigma}_n(k)$ est définie en (5.3).

Preuve. (voir Godet (2008) [21]). •

Pour n suffisamment grand le Lemme 4 garantit que $\hat{\Sigma}_n^{-1}(k)$ existe presque sûrement car la valeur propre minimum de $\hat{\Sigma}_n(k)$ est presque sûrement positive. Nous obtenons également une borne supérieure pour la moyenne de la valeur propre maximum de $\hat{\Sigma}_n^{-1}(k)$. Mais cette borne supérieure n'est pas uniforme car $k \rightarrow +\infty$ et donc ne fournit pas un équivalent asymptotique de l'erreur de prédiction. Néanmoins la borne donnée dans le Lemme 4 sert de base au Théorème suivant.

Théorème 7 *Supposons que le processus $(X_t)_{t \in \mathbb{Z}}$ vérifie les hypothèses H.1-H.3*

- Si $d \in]0, 1/4[$ et s'il existe $\delta > 0$ tel que $K_n^{2+\delta} = O(n)$ alors $\forall q > 0$ et $\forall 1 \leq k \leq K_n$, on a

$$E\|\hat{\Sigma}_n^{-1}(k)\|^q = O(1) \quad (5.9)$$

et

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \leq C \left(\frac{K_n^2}{(n - K_n + 1)} \right)^{q/4} \quad (5.10)$$

pour n suffisamment grand;

- Si $d \in]1/4, 1/2[$ et s'ils existent $\delta > 0$ et $\delta' > 0$ tels que $K_n^{2+\delta} = O(n^{2-4d-\delta'})$ alors pour tout $q > 0$ et pour tout $1 \leq k \leq K_n$, on a

$$E\|\hat{\Sigma}_n^{-1}(k)\|^q = O(1) \quad (5.11)$$

et

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \leq C \left(\frac{K_n^2 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right)^{q/4} \quad (5.12)$$

pour n suffisamment grand;

- Si $d = 1/4$ et s'ils existent $\delta > 0$ et $\delta' > 0$ tels que $K_n^{2+\delta} = O(n^{1-\delta'})$ alors pour tout $q > 0$ et pour tout $1 \leq k \leq K_n$, on a

$$E\|\hat{\Sigma}_n^{-1}(k)\|^q = O(1) \quad (5.13)$$

et

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

pour n suffisamment grand.

Pour démontrer le Théorème précédent, nous avons besoin du Lemme suivant.

Lemme 5 *Si le processus $(X_t)_{t \in \mathbb{Z}}$ vérifie (H.2), si $1 \leq k \leq K_n$ et :*

– *Si $d \in]0, 1/4[$, alors pour tout $q > 0$, on a*

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C \left(\frac{K_n^2}{(n - K_n + 1)} \right)^{q/2}; \quad (5.15)$$

– *Si $d \in]1/4, 1/2[$, alors pour tout $q > 0$, on a*

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

– *Si $d = 1/4$, alors pour tout $q > 0$, on a*

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C \left(\frac{K_n^2 L^2 (n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)} \right)^{q/2}. \quad (5.17)$$

Preuve. (voir Godet (2008) [21]). •

Preuve du Théorème 7. Puisque $\|\cdot\|$ est une norme de matrice (voir (5.8)),

$$\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq \|\hat{\Sigma}_n^{-1}(k)\|^q \|\hat{\Sigma}_n(k) - \Sigma(k)\|^q \|\Sigma^{-1}(k)\|^q.$$

En outre, par l'hypothèse H.3, la densité spectrale du processus $(X_t)_{t \in \mathbb{Z}}$ a une borne inférieure strictement positive. Ainsi, de Grenander et Szegö (1958) (voir [22]), il existe une constante C telle que pour tout $n \geq 0$:

$$\|\Sigma^{-1}(k)\|^q \leq C. \quad (5.18)$$

En utilisant l'inégalité de Hölder avec $1/p' + 1/q' = 1$, nous obtenons :

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C \left(E\|\hat{\Sigma}_n^{-1}(k)\|^{qq'} \right)^{1/q'} \left(E\|\hat{\Sigma}_n(k) - \Sigma(k)\|^{qp'} \right)^{1/p'}.$$

Par le Lemma 4, nous avons pour tout $\theta > 0$ et pour n grand

$$\left(E\|\hat{\Sigma}_n^{-1}(k)\|^{qq'} \right)^{1/q'} \leq C(k^{2+\theta})^q$$

alors

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C(k^{2+\theta})^q \left(E\|\hat{\Sigma}_n(k) - \Sigma(k)\|^{qp'} \right)^{1/p'}. \quad (5.19)$$

Nous appliquons maintenant le Lemme 5. Afin de traiter en même temps les trois situations $d \in]0, 1/4[$, $d \in]1/4, 1/2[$ et $d = 1/4$, nous définissons $h(n)$ par :

$$h(n) = \begin{cases} \frac{K_n^2}{n-K_n+1} & \text{si } d \in]0, 1/4[\\ \frac{K_n^2 L^2 (n-K_n+1)}{(n-K_n+1)^{2-4d}} & \text{si } d \in]1/4, 1/2[\\ \frac{K_n^2 L^2 (n-K_n+1) \log(n-K_n+1)}{(n-K_n+1)} & \text{si } d = 1/4. \end{cases}$$

pour n grand, nous obtenons :

$$\left(E\|\hat{\Sigma}_n(k) - \Sigma(k)\|^{qp'} \right)^{1/p'} \leq C(h(n))^{q/2}. \quad (5.20)$$

D'après l'inégalité (5.19) et de la borne (5.20), il existe $\theta > 0$ tel que pour n suffisamment grand :

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^q \leq C(k^{4+\theta} h(n))^{q/2}. \quad (5.21)$$

Des inégalités (5.21) et (5.18), nous avons :

$$E\|\hat{\Sigma}_n^{-1}(k)\|^q \leq C \left(1 + (k^{4+\theta} h(n))^{q/2} \right). \quad (5.22)$$

Cette inégalité n'est pas suffisante pour obtenir (5.9) et (5.11) car, en vertu des hypothèses du Théorème 7, $(k^{4+\theta} h(n))^{q/2}$ n'est pas nécessairement bornée. Nous devons améliorer l'inégalité intermédiaire (5.22).

L'inégalité de Cauchy-Schwarz et (5.18) donne :

$$E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \leq C \left(E\|\hat{\Sigma}_n^{-1}(k)\|^q \right)^{1/2} \left(E\|\hat{\Sigma}_n(k) - \Sigma(k)\|^q \right)^{1/2}. \quad (5.23)$$

Il existe $C > 0$ indépendant de q tel que :

$$E\|\hat{\Sigma}_n^{-1}(k)\|^{q/2} \leq C \left(E\|\Sigma^{-1}(k)\|^{q/2} + E\|\hat{\Sigma}_n^{-1}(k) - \Sigma^{-1}(k)\|^{q/2} \right). \quad (5.24)$$

Les inégalités (5.23), (5.18), (5.22) et le Lemme 5 impliquent que :

$$E\|\hat{\Sigma}_n^{-1}(k)\|^{q/2} \leq C \left(1 + (k^{4+\theta}h(n)^2)^{q/4}\right). \quad (5.25)$$

En répétant $(s - 1)$ fois cet argument (c'est-à-dire : en utilisant les inégalités (5.24), (5.23) (5.18) (5.25) et Lemme 5, on aura pour n grand

$$E\|\hat{\Sigma}_n^{-1}(k)\|^{q/2-s} \leq C \left(1 + (k^{4+\theta}h(n)^{1+s})^{q/2-(s+1)}\right). \quad (5.26)$$

Par hypothèse, il existe $\delta > 0$ tel que $h(n)k^\delta$ converge vers 0 quand n tend vers l'infini, donc il existe s , tel que $E\|\hat{\Sigma}_n^{-1}(k)\|^{q/2-s}$ est borné. Puisque q de (5.26) est arbitraire, (5.9) et (5.11) sont prouvés. Les inégalités (5.10) et (5.12) suivies de (5.23) et du Lemme 5. •

Dans la section suivante, nous établissons une expression asymptotique de l'erreur moyenne quadratique de prédiction du prédicteur de moindres carrés en utilisant la borne supérieure stricte pour $E\|\hat{\Sigma}_n^{-1}(k)\|^q$ donnée dans le Théorème 7.

5.3 L'erreur moyenne quadratique de prédiction du prédicteur de moindres carrés

Dans cette section, notre but est de donner une expression asymptotique de l'erreur moyenne quadratique de prédiction du prédicteur défini en (5.5). D'abord nous décomposons $X_{n+1} - \tilde{X}_{n+1}(k)$ comme suit :

$$X_{n+1} - \tilde{X}_{n+1}(k) = \varepsilon_{n+1} + f(k) + S_n(k) \quad (5.27)$$

où ε_{n+1} est le bruit blanc d'innovation au temps $n + 1$, $S_n(k)$ est l'erreur due à la projection sur le sous-espace fermé engendré par (X_n, \dots, X_{n-k+1}) , et $f(k)$ est l'erreur due à l'estimation des coefficients de prédiction. Plus précisément, si nous posons $a_{i,k} = 0$ pour $i > k$ (pour $j \leq k$, les coefficients $a_{j,k}$ sont définis en (5.2)), nous avons

$$S_j(k) = - \sum_{i=1}^{+\infty} (a_i - a_{i,k}) X_{j+1-i} \quad (5.28)$$

et $\varepsilon_{j+1,k}$ est égal à l'erreur de prédiction de X_{j+1} due à la projection sur (X_j, \dots, X_{j-k+1}) c'est-à-dire :

$$\varepsilon_{j+1,k} = X_{j+1} - P_{[X_{j-k}, \dots, X_j]}(X_{j+1}) = X_{j+1} + \sum_{l=1}^k a_{l,k} X_{j+1-l}.$$

Nous avons

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

où $\mathbf{X}'_n(k)$ est défini en (5.4).

Compte tenu de (5.27), on obtient la décomposition de l'erreur moyenne quadratique de prédiction comme étant la somme de la variance σ_ε^2 du bruit blanc et l'erreur due à la méthode de prédiction $E(f(k) + S_n(k))^2$:

$$E\left(X_{n+1} - \tilde{X}_{n+1}(k)\right)^2 = \sigma_\varepsilon^2 + E(f(k) + S_n(k))^2.$$

Théorème 8 *Supposons que les hypothèses H.1-H.3 sont vérifiées et considérons la suite $(K_n)_{n \in \mathbb{N}}$ telle que pour tout $\delta > 0$*

$$K_n^4 = o(n^{1-2d-\delta}), \quad (5.29)$$

alors, on a

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \left| \frac{E\left(X_{n+1} - \tilde{X}_{n+1}(k)\right)^2 - \sigma_\varepsilon^2}{L_n(k)} - 1 \right| = 0,$$

où

$$L_n(k) = E(S_n(k))^2 + \frac{k}{n - K_n + 1} \sigma_\varepsilon^2. \quad (5.30)$$

$S_n(k)$ étant défini en (5.28).

Preuve. D'après (5.27), nous avons :

$$\left| \frac{E\left(X_{n+1} - \tilde{X}_{n+1}(k)\right)^2 - \sigma_\varepsilon^2}{L_n(k)} - 1 \right| = \left| \frac{E(f(k) + S_n(k))^2}{L_n(k)} - 1 \right|.$$

Notre preuve est partagée en trois étapes :

1. Nous fournissons une approximation de $E(f(k))^2$ qui est plus facile à estimer. Cette approximation notée $E(f_1(k))^2$ sera définie en (5.31).
2. Nous montrons que l'équivalent asymptotique de $E(f_1(k))^2$ est $\frac{k}{n-K_n+1} \sigma_\varepsilon^2$.

3. Nous montrons que le terme $E(f(k)S_n(k))$ est négligeable par rapport à $L_n(k)$.

Première étape Nous introduisons

$$f_1(k) = -\mathbf{X}_n^{*'}(k)\Sigma^{-1}(k)\frac{1}{n-K_n+1}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{X}_j(k)\varepsilon_{j+1} \quad (5.31)$$

avec

$$\mathbf{X}_n^{*'}(k) = \left(\sum_{j=0}^{\sqrt{n}/2-K_n} b_j\varepsilon_{n-j}, \dots, \sum_{j=0}^{\sqrt{n}/2-K_n} b_j\varepsilon_{n-k+1-j} \right).$$

Lemme 6 *Si les hypothèses du Théorème 8 sont vérifiées, alors on a*

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} E \left(\sqrt{\frac{1}{L_n(k)}} (f(k) - f_1(k)) \right)^2 = 0. \quad (5.32)$$

Preuve. (voir Godet (2008) [21]). •

Deuxième étape Nous prouvons que

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \left| E \left(\frac{n-K_n+1}{k\sigma_\varepsilon^2} f_1^2(k) \right) - 1 \right| = 0. \quad (5.33)$$

Observons d'abord que

$$\begin{aligned} & E \left(\frac{n-K_n+1}{k\sigma_\varepsilon^2} f_1^2(k) \right) \\ &= \frac{n-K_n+1}{k\sigma_\varepsilon^2} E \left(\mathbf{X}_n^{*'}(k)\Sigma^{-1}(k)\frac{1}{n-K_n+1}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{X}_j(k)\varepsilon_{j+1} \right)^2 \\ &= \frac{n-K_n+1}{k\sigma_\varepsilon^2} E \left[\text{trace} \left(\mathbf{X}_n^{*'}(k)\Sigma^{-1}(k)\frac{1}{n-K_n+1}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{X}_j(k)\varepsilon_{j+1} \right)^2 \right] \end{aligned}$$

$$= \frac{n - K_n + 1}{k\sigma_\varepsilon^2} \operatorname{trace} \left[E \left(\Sigma^{-1}(k) \frac{1}{(n - K_n + 1)^2} \sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(k) \varepsilon_{j+1} \sum_{l=K_n}^{n-\sqrt{n}-1} \mathbf{X}'_l(k) \varepsilon_{l+1} \Sigma^{-1}(k) \mathbf{X}_n^* \mathbf{X}_n^{*'}(k) \right) \right].$$

Puisque les vecteurs $\mathbf{X}_n^{*'}$ et $\sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(k) \varepsilon_{j+1}$ sont non-corrélés parce que $k \leq K_n$:

$$\begin{aligned} & E \left(\frac{n - K_n + 1}{k\sigma_\varepsilon^2} f_1^2(k) \right) \\ &= \frac{n - K_n + 1}{k\sigma_\varepsilon^2 (n - K_n + 1)^2} \operatorname{trace} \left(\Sigma^{-1}(k) (n - K_n + 1 - \sqrt{n}) \sigma_\varepsilon^2 \Sigma(k) \Sigma^{-1}(k) \Sigma^*(k) \right) \\ &= \operatorname{trace} \left(\Sigma^{-1}(k) \Sigma^*(k) k^{-1} \right) (n - K_n + 1 - \sqrt{n}) (n - K_n + 1)^{-1}, \end{aligned}$$

où $\Sigma^*(k)$ est la matrice de covariance du vecteur $X_n^*(k)$. Nous notons que :

$$(n - K_n + 1 - \sqrt{n}) (n - K_n + 1)^{-1} \rightarrow 1 \quad \text{quand } n \rightarrow +\infty.$$

Donc, nous n'avons qu'à étudier la trace de $(\Sigma^{-1}(k) \Sigma^*(k) k^{-1})$. Nous allons utiliser l'inégalité suivante : pour toutes $(k \times k)$ -matrices A et B

$$\begin{aligned} |\operatorname{trace}(AB)| &\leq \sqrt{\operatorname{trace}(AA')} \sqrt{\operatorname{trace}(BB')} \\ &\leq k \|A\| \|B\|. \end{aligned}$$

Nous obtenons

$$\begin{aligned} & \max_{1 \leq k \leq K_n} |\operatorname{trace}(\Sigma^{-1}(k) \Sigma^*(k) k^{-1}) - 1| \\ &= \max_{1 \leq k \leq K_n} |\operatorname{trace}(\Sigma^{-1}(k) (\Sigma^*(k) - \Sigma(k)) k^{-1})| \\ &\leq \max_{1 \leq k \leq K_n} \|\Sigma^{-1}(k)\| \|\Sigma^*(k) - \Sigma(k)\| \\ &\leq \max_{1 \leq k \leq K_n} \|\Sigma^{-1}(k)\| \max_{1 \leq k \leq K_n} \|\Sigma^*(k) - \Sigma(k)\|, \end{aligned}$$

$\Sigma(k) - \Sigma^*(k)$ est symétrique parce que $\Sigma(k)$ et $\Sigma^*(k)$ sont deux matrices symétriques, et leurs normes sont inférieures à toutes les autres normes matricielles. Nous utilisons la norme subordonnée définie pour toute matrice $Y = (y_{i,j})_{1 \leq i,j \leq k}$ par :

$$\|Y\|_1 = \max_j \sum_{i=1}^k |y_{i,j}|.$$

Pour n grand, nous obtenons

$$\begin{aligned} & \max_{1 \leq k \leq K_n} \|\Sigma^{-1}(k)\| \max_{1 \leq k \leq K_n} \|\Sigma^*(k) - \Sigma(k)\| \\ \leq & \max_{1 \leq k \leq K_n} \|\Sigma^{-1}(k)\| \max_{1 \leq k \leq K_n} \|\Sigma(k) - \Sigma^*(k)\|_1 \\ \leq & \max_{1 \leq k \leq K_n} \|\Sigma^{-1}(k)\| \max_{1 \leq k \leq K_n} k \max_{0 \leq j \leq k-1} \sum_{l=\sqrt{n}/2-K_n+1}^{+\infty} |b_l b_{l+j}| \\ = & O\left(\frac{K_n}{(\sqrt{n})^{1-2d-\delta}}\right), \end{aligned}$$

pour tout $\delta > 0$. Alors

$$\max_{1 \leq k \leq K_n} \|\Sigma^{-1}(k)\| \max_{1 \leq k \leq K_n} \|\Sigma^*(k) - \Sigma(k)\| = o(1)$$

résulte de la condition (5.29).

Troisième étape Nous considérons le terme $E(f(k)S_n(k)L_n^{-1}(k))$ et montrons qu'il est négligeable. Ing et Wei (2003) (voir [24]) ont démontré que :

$$|E(f(k)S_n(k)L_n^{-1}(k))| = |E((f(k) - f_1(k))S_n(k)L_n^{-1}(k))|.$$

Par l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} & \max_{1 \leq k \leq K_n} |E((f(k) - f_1(k))S_n(k)L_n^{-1}(k))| \\ \leq & \left[\max_{1 \leq k \leq K_n} E((f(k) - f_1(k))^2 L_n^{-1}(k)) \max_{1 \leq k \leq K_n} E(S_n^2(k)L_n^{-1}(k)) \right]^{1/2}. \end{aligned}$$

Par (5.32), nous obtenons :

$$\max_{1 \leq k \leq K_n} E((f(k) - f_1(k))^2 L_n^{-1}(k)) = o(1)$$

et en utilisant la définition en (5.30) de $L_n(k)$, nous avons :

$$\max_{1 \leq k \leq K_n} E(S_n^2(k)L_n^{-1}(k)) = O(1).$$

Enfin, nous avons :

$$\max_{1 \leq k \leq K_n} |E((f(k) - f_1(k))S_n(k)L_n^{-1}(k))| = o(1).$$

•

Dans ce Théorème, nous avons obtenu une expression asymptotique de l'erreur moyenne quadratique de prédiction de $\tilde{X}_{n+1}(k)$, qui est vérifiée uniformément pour tout $1 \leq k \leq K_n$. Dans le cas de courte mémoire c'est-à-dire : dans S opposant que le processus $(X_t)_{t \in \mathbb{Z}}$ est gaussien, admet les représentations autorégressive et moyenne mobile infinies définies en (5.6), que les coefficients $(a_j)_{j \in \mathbb{N}}$ vérifient $\sum_{j=1}^{+\infty} \sqrt{j}|a_j| < \infty$, et que les coefficients $(b_j)_{j \in \mathbb{N}}$ sont absolument sommables, Ing et Wei (2003) ont prouvés que si $K_n^{2+\delta} = O(n)$ pour tout $\delta > 0$:

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq K_n} \left| \frac{E\left(X_{n+1} - \tilde{X}_{n+1}(k)\right)^2 - \sigma_\varepsilon^2}{L_n(k)} - 1 \right| = 0$$

$L_n(k)$ étant défini en (5.30).

Le terme $L_n(k)$ a la même expression dans le cas de courte mémoire comme dans le cas de longue mémoire.

Dans la section suivante, nous allons utiliser la démonstration du Théorème 8 pour obtenir un théorème central limite pour notre prédicteur.

5.4 Théorème central limite

Nous recherchons un facteur de normalisation pour obtenir une convergence en distribution de la différence entre notre prédicteur $\tilde{X}_{n+1}(K_n)$ et le prédicteur Wiener-Kolmogorov $\hat{X}_{n+1} = -\sum_{j=1}^{+\infty} a_j X_{n+1-j}$, ce qui est le prédicteur linéaire de moindres carrés basé sur tout le passé.

Théorème 9 *Supposons que les hypothèses H.1-H.3 sont vérifiées et considérons la suite $(K_n)_{n \in \mathbb{Z}}$ telle que :*

$$K_n^4 = O(n) \quad \text{et} \quad K_n^{1+2d} = o(n^{1-2d}) \quad (5.34)$$

alors, on a

$$\frac{1}{\sqrt{E[S_n^2(K_n)]}} \left(\hat{X}_{n+1} - \tilde{X}_{n+1}(K_n) \right) \xrightarrow{L} N(0, 1).$$

Preuve. La différence entre notre prédicteur $\tilde{X}_{n+1}(K_n)$ et le prédicteur de Wiener-Kolmogorov \hat{X}_{n+1} est égale à :

$$\hat{X}_{n+1} - \tilde{X}_{n+1}(K_n) = f(K_n) + S_n(K_n). \quad (5.35)$$

Puisque $(X_t)_{t \in \mathbb{Z}}$ est un processus gaussien de moyenne nulle, $\left(\sum_{i=1}^l (a_i - a_{i,K_n}) X_{t+1-i}\right)_{t \in \mathbb{Z}}$ est une variable aléatoire gaussienne de moyenne nulle pour tout entier l . Mais $\left(\sum_{i=1}^l (a_i - a_{i,K_n}) X_{t+1-i}\right)_{t \in \mathbb{Z}}$ converge en moyenne quadratique et donc en distribution quand l tend vers l'infini. Alors $S_n(K_n)$ est une variable aléatoire gaussienne de moyenne nulle. Par conséquent, il suffit de prouver que

$$\frac{1}{\sqrt{E[S_n^2(K_n)]}} f(K_n) \xrightarrow{P} 0 \quad \text{quand } n \rightarrow +\infty.$$

Etudions d'abord la limite de $1/\sqrt{E[S_n^2(K_n)]}$. Pour tout entier l ,

$$E \left(\sum_{i=1}^l (a_i - a_{i,K_n}) X_{n+1-i} \right)^2 \geq 2\pi \underline{f} \sum_{i=1}^l (a_i - a_{i,K_n})^2$$

parce que la densité spectrale f est minorée par une constante positive \underline{f} (voir Grenander et Szegö (1958) [22]). en prenant la limite pour $l \rightarrow +\infty$, on obtient :

$$\begin{aligned} E \left(\sum_{i=1}^{+\infty} (a_i - a_{i,K_n}) X_{n+1-i} \right)^2 &\geq 2\pi \underline{f} \sum_{i=1}^{+\infty} (a_i - a_{i,K_n})^2 \\ &\geq 2\pi \underline{f} \sum_{i=K_n+1}^{+\infty} a_i^2 \end{aligned}$$

puisque $a_{i,K_n} = 0$ quand $i > K_n$. Sous l'hypothèse H.4,

$$\sum_{i=K_n+1}^{+\infty} a_i^2 \sim \frac{1}{1+2d} K_n^{-2d-1} L^2(K_n) \quad \text{quand } n \rightarrow +\infty$$

(voir la Proposition 1.5.10 de Bingham *et al.* (1987) [10]). Alors pour tout $\delta > 0$, il existe $C > 0$ tel que :

$$\frac{1}{E[S_n^2(K_n)]} \leq C K_n^{2d+1+\delta}. \quad (5.36)$$

En introduisant f_1 défini dans la démonstration du Théorème 8, nous décomposons la preuve de la convergence en moyenne quadratique en deux parties. Nous allons d'abord démontrer que :

$$\frac{1}{\sqrt{E[S_n^2(K_n)]}} [f(K_n) - f_1(K_n)] \longrightarrow 0 \quad \text{quand } n \rightarrow +\infty \quad (5.37)$$

et ensuite

$$\frac{1}{\sqrt{E[S_n^2(K_n)]}} f_1(K_n) \longrightarrow 0 \quad \text{quand } n \rightarrow +\infty. \quad (5.38)$$

Plus précisément, nous allons prouver la convergence en moyenne quadratique (5.37), en utilisant la décomposition de $f(k) - f_1(k)$ en quatre termes suivante :

$$\begin{aligned} & f(k) - f_1(k) \\ = & \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{n - K_n + 1} \left(\sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(k) \varepsilon_{j+1} - \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right) \end{aligned} \quad (5.39)$$

$$+ \mathbf{X}_n^{*'}(k) \left(\Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \quad (5.40)$$

$$+ \left(\mathbf{X}_n^{*'} - \mathbf{X}_n'(k) \right) \hat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \quad (5.41)$$

$$+ \mathbf{X}_n'(k) \hat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) (\varepsilon_{j+1} - \varepsilon_{j+1,k}). \quad (5.42)$$

Étude de l'expression donnée en (5.39)

$$\begin{aligned} & \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{n - K_n + 1} \left[\sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(k) \varepsilon_{j+1} - \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right] \\ & = \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{n - K_n + 1} \left(- \sum_{j=n-\sqrt{n}}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right). \end{aligned}$$

L'inégalité de Hölder appliquée deux fois avec $(1/p) + (1/q) = 1$ et $(1/p') + (1/q') = 1$ donne

$$\begin{aligned}
& E \left(\mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=n-\sqrt{n}}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\
& \leq \left(E \left\| \mathbf{X}_n^{*'}(k) \Sigma^{-1}(k) \right\|^{2q} \right)^{1/q} \left(E \left\| \frac{1}{n - K_n + 1} \sum_{j=n-\sqrt{n}}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \\
& \leq \left(E \left\| \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} \left(\left\| \Sigma^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \\
& \quad \left(E \left\| \frac{1}{n - K_n + 1} \sum_{j=n-\sqrt{n}}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p}.
\end{aligned}$$

Sous l'hypothèse H.3 pour tout p' et q' :

$$\left(\left\| \Sigma^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} = \left\| \Sigma^{-1}(k) \right\|^2 = O(1) \quad (5.43)$$

puisque la densité spectrale du processus $(X_t)_{t \in \mathbb{Z}}$ admet une borne inférieure positive et alors la plus grande valeur propre de $\Sigma^{-1}(k)$ est bornée. En outre, par la convexité de la fonction $x \mapsto x^{2q'q}$ et la stationnarité du processus $(\varepsilon_t)_{t \in \mathbb{Z}}$,

$$\begin{aligned}
\left(E \left\| \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} &= \left(E \left\| \frac{\sqrt{k}}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} \\
&= k \left(E \left\| \frac{1}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} \\
&\leq k \left(E \left[\sum_{j=0}^{\sqrt{n}/2 - K_n} b_j \varepsilon_{n-j} \right]^{2q'q} \right)^{1/(q'q)}
\end{aligned}$$

et par le Lemme 2 de Wei (1987) (voir [37]) :

$$\begin{aligned}
\left(E \left\| \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} &\leq C' k \left(\sum_{j=0}^{\sqrt{n}/2 - K_n} b_j^2 \right) \\
&\leq Ck \\
&\leq CK_n
\end{aligned} \quad (5.44)$$

parceque la suite $(b_j^2)_{j \in N}$ est absolument sommable.

Lemme 7 *Supposons que les hypothèses du Théorème 8 sont vérifiées. Si $q > 1$, alors pour tout $1 \leq k \leq K_n$, on a*

$$E \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^q \leq C k^{q/2}$$

avec C indépendante de n et k .

Preuve. (voir Godet (2008) [21]). •

Enfin par le Lemme précédent, nous avons :

$$\begin{aligned} & \left(E \left\| \frac{1}{n - K_n + 1} \sum_{j=n-\sqrt{n}}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \\ & \leq \frac{(\sqrt{n} + 1)^{1/2}}{n - K_n + 1} \left(E \left\| \frac{1}{(\sqrt{n} + 1)^{1/2}} \sum_{j=n-\sqrt{n}}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \\ & \leq C \left(\frac{k}{n^{1/4}} \right). \end{aligned} \quad (5.45)$$

Par les inégalités (5.43), (5.44), (5.45) et (5.36), le terme (5.39) vérifi pour tout $\delta > 0$:

$$\begin{aligned} & E \left(\frac{1}{\sqrt{E[S_n^2(K_n)]}} \mathbf{X}_n^{*'}(K_n) \Sigma^{-1}(K_n) \frac{1}{n - K_n + 1} \left[\sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{X}_j(K_n) \varepsilon_{j+1} - \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) \varepsilon_{j+1}^2 \right]^2 \right) \\ & = O \left(\frac{K_n^{3+2d+\delta}}{n^{1/4}} \right). \end{aligned} \quad (5.46)$$

Sous l'hypothèse (5.34), la moyenne (5.46) converge vers 0.

Étude de l'expression donnée en(5.40)

Montrons que :

$$\lim_{n \rightarrow +\infty} E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \mathbf{X}_n^{*'}(k) \left[\Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 = 0. \quad (5.47)$$

En appliquant deux fois l'inégalité de Hölder, nous avons :

$$\begin{aligned} & E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \mathbf{X}_n^{*'}(k) \left[\Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq \left(E \left\| \frac{1}{\sqrt{k}} \mathbf{X}_n^{*'}(k) \right\|^{2q'q} \right)^{1/(q'q)} \left(E \left\| \Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \\ & \quad \left(E \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p}. \end{aligned}$$

En appliquant le Lemme 7, on obtient

$$\left(E \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p} \leq Ck \leq CK_n. \quad (5.48)$$

Maintenant nous montrons la convergence en moyenne-quadratique vers zéro quand $d \in]0, 1/2[$.

Pour $d \in]0, 1/4[$, nous appliquons le Théorème 7 et nous obtenons :

$$\left(E \left\| \Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \leq C \left(\frac{K_n^2}{n - K_n + 1} \right). \quad (5.49)$$

D'après (5.44), (5.48) et (5.49), on a

$$\begin{aligned} & E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \mathbf{X}_n^{*'}(k) \left[\Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq C \left(\frac{K_n^3}{n - K_n + 1} \right) \end{aligned}$$

qui converge vers zéro si la condition (5.29) est vérifiée.

Si $d \in]1/4, 1/2[$, nous obtenons par le Théorème 7

$$\left(E \left\| \Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \leq C \left(\frac{K_n^2 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right). \quad (5.50)$$

Les inégalités (5.44), (5.48) et (5.50) nous permettent de conclure que :

$$\begin{aligned} & E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \mathbf{X}_n^{*'}(k) \left[\Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq C \left(\frac{K_n^3 L^2 (n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right) \end{aligned}$$

qui converge vers zéro sous l'hypothèse (5.29).

Si $d = 1/4$, nous obtenons par le Théorème 7

$$\begin{aligned} & \left(E \left\| \Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \\ & \leq C \left(\frac{K_n^2 L^2 (n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right). \quad (5.51) \end{aligned}$$

Les inégalités (5.44), (5.48) et (5.51) nous permettent de conclure que :

$$\begin{aligned} & E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \mathbf{X}_n^{*'}(k) \left[\Sigma^{-1}(k) - \hat{\Sigma}_n^{-1}(k) \right] \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq C \left(\frac{K_n^3 L^2 (n - K_n + 1) \log(n - K_n + 1)}{(n - K_n + 1)^{2-4d}} \right) \end{aligned}$$

qui converge vers zéro sous l'hypothèse (5.29).

De même pour le terme (5.40) en utilisant (5.47) et (5.36) nous obtenons pour tout $\delta > 0$:

$$\begin{aligned} & E \left(\frac{1}{\sqrt{E[S_n^2(K_n)]}} \mathbf{X}_n^{*'}(K_n) \left[\Sigma^{-1}(K_n) - \hat{\Sigma}_n^{-1}(K_n) \right] \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) \varepsilon_{j+1} \right)^2 \\ & = O \left(\frac{K_n^{5+2d+\delta}}{(n - K_n + 1)^2} \right) \end{aligned}$$

qui converge vers zéro sous l'hypothèse (5.34) pour δ suffisamment petit.

Étude de l'expression donnée en (5.41)

Montrons que :

$$\lim_{n \rightarrow +\infty} E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \left[\mathbf{X}_n^{*'}(k) - \mathbf{X}_n'(k) \right] \hat{\Sigma}_n^{-1}(k) \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 = 0. \quad (5.52)$$

En utilisant l'inégalité de Hölder deux fois, nous avons :

$$\begin{aligned} & E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \left[\mathbf{X}_n^{*'}(k) - \mathbf{X}_n'(k) \right] \hat{\Sigma}_n^{-1}(k) \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq \left(E \left\| \frac{1}{\sqrt{k}} \left[\mathbf{X}_n^{*'}(k) - \mathbf{X}_n'(k) \right] \right\|^{2q'q} \right)^{1/(q'q)} \left(E \left\| \hat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \\ & \quad \left(E \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right\|^{2p} \right)^{1/p}. \end{aligned}$$

Par la convexité de la fonction $x \mapsto x^{q'q}$ et la stationarité du processus $(\varepsilon_t)_{t \in \mathbb{Z}}$, nous avons :

$$\begin{aligned} & \max_{1 \leq k \leq K_n} \left(E \left\| \frac{1}{\sqrt{k}} \left[\mathbf{X}_n^{*'}(k) - \mathbf{X}_n'(k) \right] \right\|^{2q'q} \right)^{1/(q'q)} \\ & \leq \left(E \left(\sum_{j=\sqrt{n}/2 - K_n + 1}^{+\infty} b_j \varepsilon_{n-j-1} \right)^{2q'q} \right)^{1/(q'q)}. \end{aligned}$$

et par le Lemme 2 de Wei (1987) (voir [37]), nous obtenons :

$$\begin{aligned} \max_{1 \leq k \leq K_n} \left(E \left\| \frac{1}{\sqrt{k}} \left[\mathbf{X}_n^{*'}(k) - \mathbf{X}_n'(k) \right] \right\|^{2q'q} \right)^{1/(q'q)} & \leq C \left(\sum_{j=\sqrt{n}/2 - K_n + 1}^{+\infty} b_j^2 \right) \\ & \leq C n^{\frac{2d-1}{2}}. \end{aligned} \quad (5.53)$$

Puis par le Théorème 7

$$\left(E \left\| \hat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \leq C. \quad (5.54)$$

Par les inégalités (5.48), (5.53) et (5.54), nous obtenons alors

$$\begin{aligned} & E \left(\frac{1}{\sqrt{k(n - K_n + 1)}} \left[\mathbf{X}_n^{*'}(k) - \mathbf{X}_n'(k) \right] \hat{\Sigma}_n^{-1}(k) \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) \varepsilon_{j+1} \right)^2 \\ & \leq C \left(K_n n^{\frac{2d-1}{2}} \right) \end{aligned}$$

qui converge vers zéro sous l'hypothèse (5.29).

Pour le troisième terme (5.41), par (5.52) et (5.36) on obtient :

$$\begin{aligned} & E \left(\frac{1}{\sqrt{E[S_n^2(K_n)]}} \left[\mathbf{X}_n^{*'}(K_n) - \mathbf{X}_n'(K_n) \right] \hat{\Sigma}_n^{-1}(K_n) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) \varepsilon_{j+1} \right)^2 \\ & = O \left(\frac{K_n^{3+2d+\delta}}{(n - K_n + 1)^{\frac{3-2d}{2}}} \right) \end{aligned}$$

qui converge vers zéro sous l'hypothèse (5.34).

Étude de l'expression donnée en (5.42)

On veut montrer que :

$$\lim_{n \rightarrow +\infty} E \left(\sqrt{\frac{1}{E(S_n(k))^2}} \mathbf{X}_n'(k) \hat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 = 0. \quad (5.55)$$

En utilisant l'inégalité de Hölder deux fois, nous avons :

$$\begin{aligned} & E \left(\sqrt{\frac{1}{E(S_n(k))^2}} \mathbf{X}_n'(k) \hat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 \\ & \leq \frac{1}{(n - K_n + 1) E(S_n(k))^2} \left(E \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right\|^{2q'q} \right)^{1/(q'q)} \\ & \quad \left(E \left\| \hat{\Sigma}_n^{-1}(k) \right\|^{2p'q} \right)^{1/(p'q)} \left(E \left\| \mathbf{X}_n'(k) \right\|^{2p} \right)^{1/p}. \end{aligned}$$

En appliquant le Lemme suivant,

Lemme 8 *Sous l'hypothèse H.2. Si $q \geq 1$, alors pour tout $\delta > 0$, il existe une constante C telle que pour tout $1 \leq k \leq K_n$:*

$$\begin{aligned} & E \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) (\varepsilon_{j+1,k} - \varepsilon_{j+1}) \right\|^q \\ & \leq C (k(n - K_n + 1)^{2d+\delta} E(S_n(k))^2)^{q/2} \end{aligned} \quad (5.56)$$

où la norme $\|\cdot\|$ est définie en (5.7).

Preuve. (voir Godet (2008) [21]). •

Nous obtenons pour tout $\delta > 0$:

$$\begin{aligned} & \left(E \left\| \frac{1}{\sqrt{n - K_n + 1}} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right\|^{2q'} \right)^{1/(q'q)} \\ & \leq C (k(n - K_n + 1)^{2d+\delta} E(S_n(k))^2). \end{aligned} \quad (5.57)$$

Enfin nous choisissons $p = 2$ et nous avons :

$$\begin{aligned} (E \|\mathbf{X}'_n(k)\|^4)^{1/2} &= \sqrt{E \left(\sum_{j=1}^k X_j^2 \right)^2} \\ &= \sqrt{\left(\sum_{j=1}^k \gamma(j) \right)^2 + 2 \sum_{j=1}^k \sum_{l=1}^k \gamma(j-l)^2} \end{aligned}$$

puisque le processus $(X_t)_{t \in \mathbb{Z}}$ est gaussien. En utilisant l'hypothèse H.2 sur les covariances, nous vérifions que pour tout $\delta > 0$:

$$(E \|\mathbf{X}'_n(k)\|^4)^{1/2} \leq C \sqrt{k^{4d+\delta}} \leq C \sqrt{k}, \quad (5.58)$$

si $d \in]0, 1/4[$. Par les inégalités (5.54), (5.57), (5.58) et $1 \leq k \leq K_n$, nous concluons que $\forall \delta > 0$:

$$\begin{aligned}
& E \left(\sqrt{\frac{1}{E(S_n(k))^2}} \mathbf{X}'_n(k) \hat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 \\
& \leq C \frac{1}{(n - K_n + 1) E(S_n(k))^2} \left(\sqrt{k} k (n - K_n + 1)^{2d+\delta} E(S_n(k))^2 \right) \\
& \leq C \frac{K_n^{3/2}}{(n - K_n + 1)^{1-2d-\delta}}
\end{aligned}$$

qui converge vers zéro sous l'hypothèse (5.29).

D'autre part, si $d \in [1/4, 1/2[$, l'inégalité (5.58) devient

$$\forall \delta > 0, \left(E \|\mathbf{X}'_n(k)\|^4 \right)^{1/2} \leq C k^{4d+\delta}. \quad (5.59)$$

En utilisant les inégalités (5.54), (5.57) et (5.59), nous avons pour tout $\delta > 0$

$$\begin{aligned}
& E \left(\sqrt{\frac{1}{E(S_n(k))^2}} \mathbf{X}'_n(k) \hat{\Sigma}_n^{-1}(k) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(k) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 \\
& \leq C \left(\frac{K_n^{1+2d+\delta}}{(n - K_n + 1)^{1-2d-\delta}} \right)
\end{aligned}$$

qui converge vers 0 sous l'hypothèse (5.29).

Enfin, l'estimation du quatrième terme (5.42) est directement donnée par (5.55) :

$$\begin{aligned}
& E \left(\sqrt{\frac{1}{E(S_n(K_n))^2}} \mathbf{X}'_n(K_n) \hat{\Sigma}_n^{-1}(K_n) \frac{1}{n - K_n + 1} \sum_{j=K_n}^{n-1} \mathbf{X}_j(K_n) [\varepsilon_{j+1,k} - \varepsilon_{j+1}] \right)^2 \\
& = O \left(\frac{K_n^{1+2d+\delta}}{(n - K_n + 1)^{1-2d-\delta}} \right)
\end{aligned}$$

qui converge vers zéro sous l'hypothèse (5.34).

Nous avons prouvé que, pour tout $d \in]0, 1/2[$:

$$\frac{1}{\sqrt{E[S_n^2(K_n)]}} (f(K_n) - f_1(K_n)) \xrightarrow{L^2} 0 \quad \text{quand } n \rightarrow +\infty.$$

Maintenant, nous allons prouver la convergence en moyenne-quadratique (5.38).

D'après (5.33), on a

$$E \left(\frac{n - K_n + 1}{K_n \sigma_\varepsilon^2} f_1^2(K_n) \right) \sim \frac{K_n \sigma_\varepsilon^2}{n - K_n + 1} \quad \text{quand } n \rightarrow +\infty.$$

Sous la condition (5.34), l'inégalité (5.36) implique que :

$$\frac{1}{E(S_n^2(K_n))} \frac{K_n \sigma_\varepsilon^2}{n - K_n + 1} \rightarrow 0 \quad \text{quand } n \rightarrow +\infty.$$

Alors nous avons

$$\lim_{n \rightarrow +\infty} \frac{1}{E(S_n^2(K_n))} E(f_1^2(K_n)) = 0.$$

•

Chapitre 6

Comparaison des prédicteurs sur des séries réelles et simulées

Nous allons sur des simulations numériques vérifier l'efficacité des deux prédicteurs linéaires que nous avons étudiés dans ce mémoire : le prédicteur de Wiener-Kolmogorov tronqué et le prédicteur projeté sur un passé fini. Notre étude porte sur 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) [14] chapitre 13.2) ce qui donne accès à certaines méthodes de simulations 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 . Pour un processus fractionnaire de paramètre d , ils sont égaux à

$$\begin{aligned}
a_j &= \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \\
&= \prod_{m=1}^j \frac{m-1-d}{m}.
\end{aligned}$$

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

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

6.1.2 Méthodes de simulation

Il existe de nombreuses méthodes permettant de simuler les processus fractionnaires gaussiens. Les trajectoires simulées par ces modèles sont acceptées par les tests effectués sur la densité spectrale ou sur l'estimation du paramètre de longue mémoire par l'estimateur de Whittle. Ainsi ces modèles sont plus compatibles pour des longues trajectoires.

Nous avons étudiés trois modèles FARIMA(0,d,0) pour $d = 0.1$, $d = 0.25$ et $d = 0.4$. Pour calculer le prédicteur de Wiener-Kolmogorov (tronqué), il faut connaître la représentation autorégressive du processus. On peut calculer le développement autorégressif du processus fractionnaire intégré et donc calculer le prédicteur de Wiener-Kolmogorov tronqué. On a tronqué le développement à l'ordre $k = 100$, $k = 1000$ et puis à l'ordre $k = 8000$. On a calculé également le prédicteur projeté sur $k = 100$, $k = 1000$ et $k = 8000$ observations des trois séries précédentes.

On a calculé tous les prédicteurs à l'horizon 1 sur ces séries pour tous les points dont on connaît le passé sur k observations, les résultats sont donnés dans les figures (6.1), (3.2) et (6.3).

6.1.3 Comparaison des prédicteurs de Wiener-Kolmogorov tronqué et projeté

Dans le chapitre 2, pour un processus fractionnaire, nous avons montré que l'écart entre les erreurs commises par les deux prédicteurs croît avec le

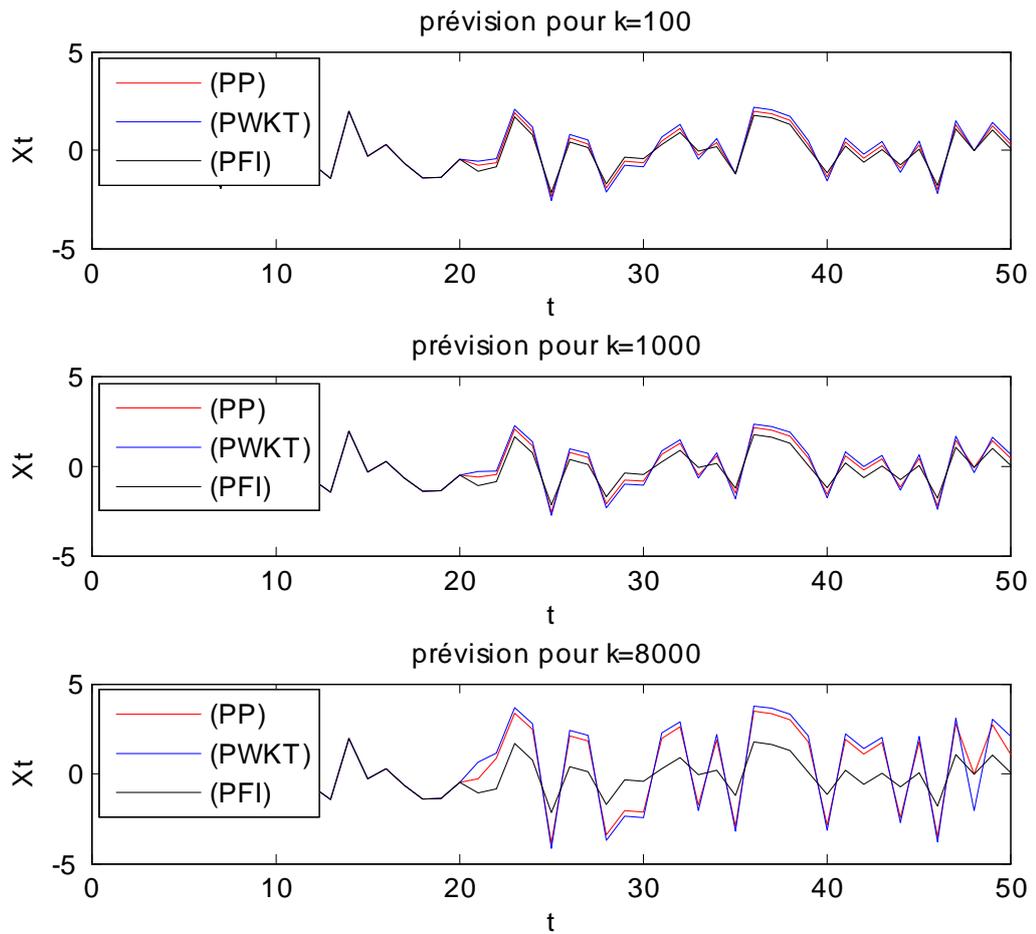


FIG. 6.1 – Prédiction à l’horizon 1 d’un processus FARIMA(0,0.1,0) (PFI) par le prédicteur de Wiener-Kolmogorov tronqué (PWKT) et par le prédicteur projeté (PP) pour $k = 100$, $k = 1000$ et $k = 8000$.

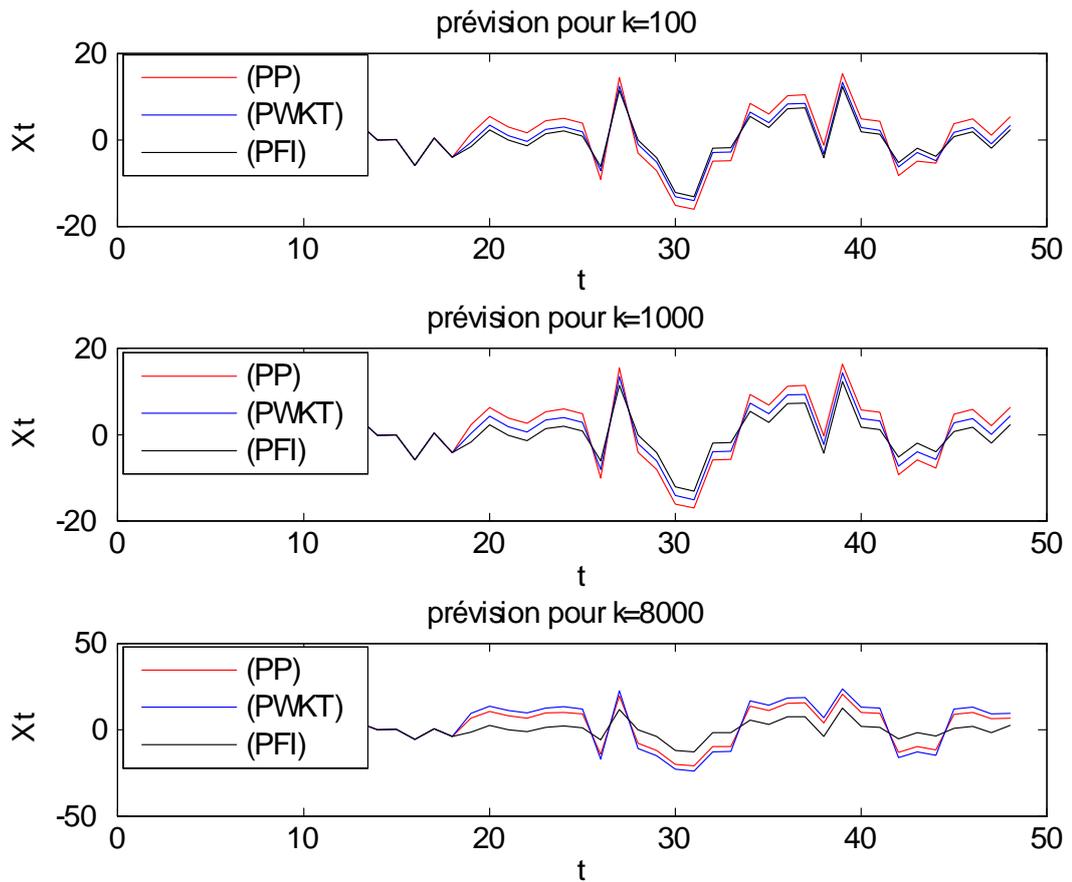


FIG. 6.2 – Prévision à l’horizon 1 d’un processus FARIMA(0,0.25,0) (PFI) par le prédicteur de Wiener-Kolmogorov tronqué (PWKT) et par le prédicteur projeté (PP) pour $k = 100$, $k = 1000$ et $k = 8000$.

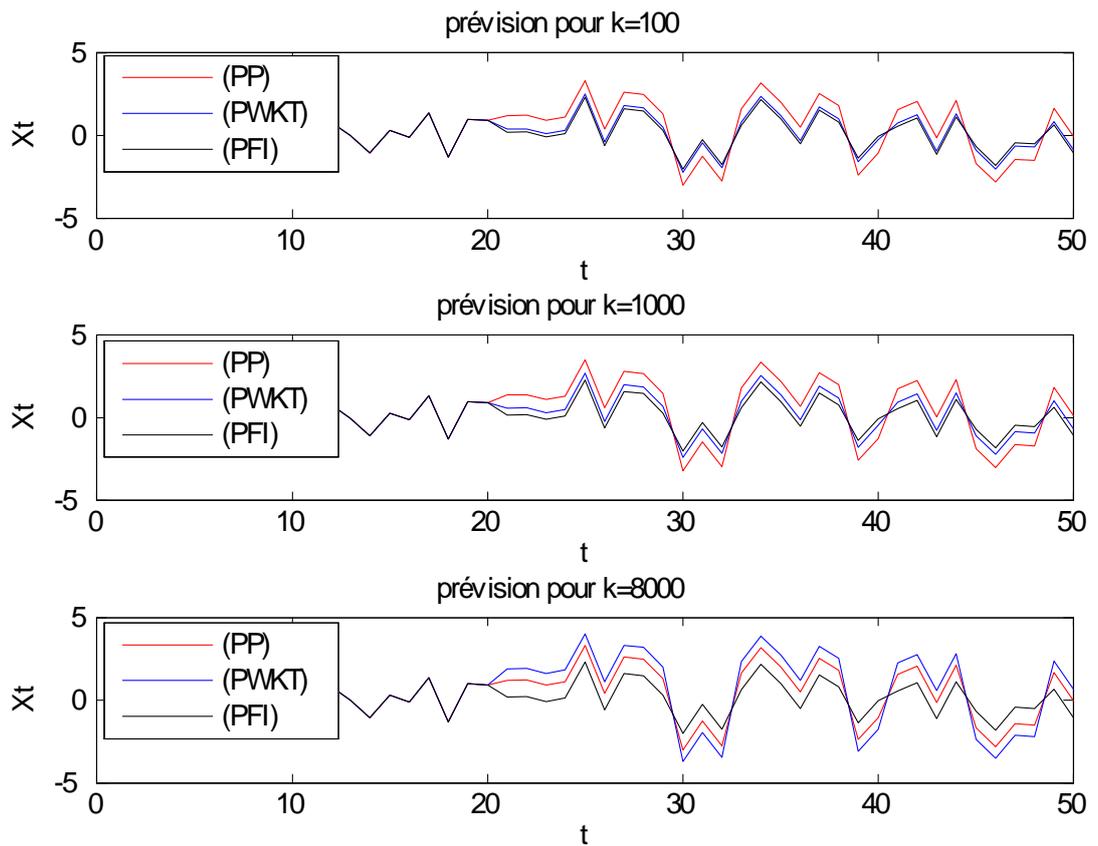


FIG. 6.3 – Prédiction à l’horizon 1 d’un processus FARIMA(0,0.4,0) (PFI) par le prédicteur de Wiener-Kolmogorov tronqué (PWKT) et par le prédicteur projeté (PP) pour $k = 100$, $k = 1000$ et $k = 8000$.

paramètre de longue mémoire d . Il atteint par exemple 60% pour $d = 0.4$ (voir la figure (2.1)) et ce qui est clair dans les figures (6.1), (6.2) et (6.3) . Si on projette ou on tronque à un ordre plus élevé plus on aura la meilleur prédiction mais dans notre cas on observe que plus l'ordre k augmente plus l'écart entre les deux prédicteurs et les vraies valeurs du processus augmente. En effet, à chaque fois qu'on projette ou on tronque à un ordre plus élevé on aura plus de coefficient à estimer et les erreurs commises à chaque estimation peuvent se cumuler. nous avons montrés, aussi au chapitre 2, 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. 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 pour $k = 8000$. 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) [38]).

6.2 Etude de la rivière Nile

Nous avons une série de données concernant la rivière Nile. Cette série montre le niveau annuel minimum de l'eau à la jauge de Roda pour les années 622-1281 (voir la figure (6.4)). Nous allons commencer par modéliser cette série hydrologique puis prévoir les valeurs futures.

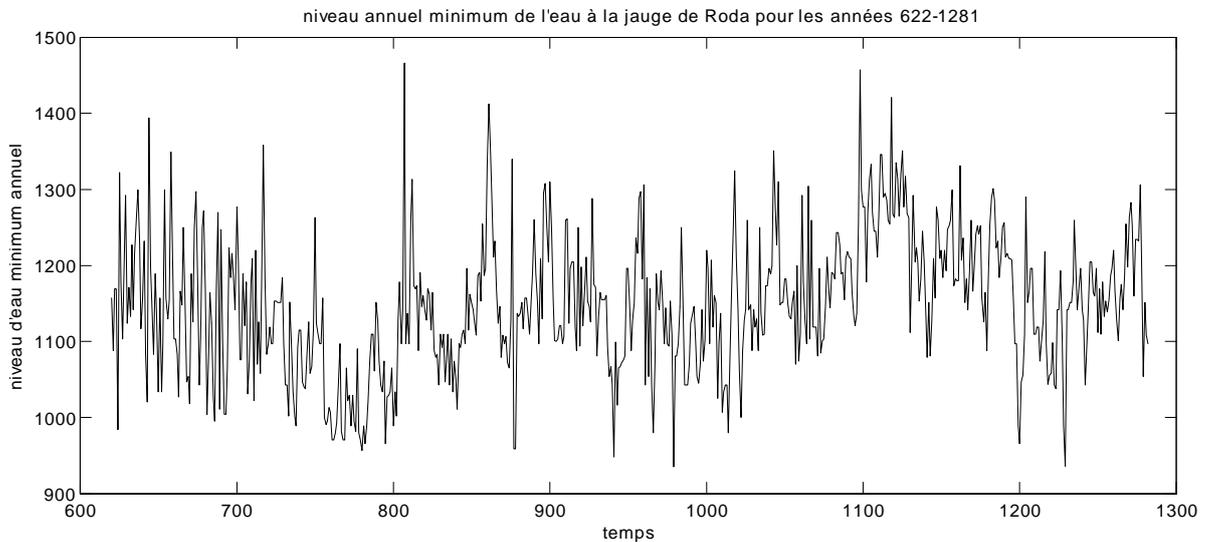


FIG. 6.4 – Niveau annuel minimum de l’eau à la jauge de Roda (la rivière Nile) pour les années 622-1281.

6.2.1 Modélisation

Nous allons modéliser la série que nous avons en utilisant le logiciel R par le modèle correspondant à la valeur minimum de l’AIC. Nous avons le modèle FARIMA(1,d,1) est approprié à la valeur AIC= 7571.18 qui est la plus petite valeur. Alors nous proposons d’associer le modèle FARIMA(1,0.3,1) à la série de données qui vérifie l’équation suivante

$$(1 - 0.9998B)(1 - B)^{0.3}X_t = (1 + 0.6797B)\varepsilon_t,$$

ainsi le programme utilisé nous donne une variance estimée par la méthode du maximum de vraisemblance égale à 5237, et les erreurs dues à l’estimation des coefficients des polynômes autorégressif et moyen-mobile sont égaux respectivement à 0.0003 et 0.0429.

6.2.2 Prévision par le prédicteur projeté

Nous allons utiliser le prédicteur projeté sur un passé fini de longueur k du chapitre 5 donnée par

$$\tilde{X}_{T+1}(k) = \sum_{i=1}^k \hat{a}_{i,k} X_{T+1-i}$$

les coefficients $\hat{a}_{i,k}$ vérifient les équations de Yule-Walker, ils sont estimés en utilisant $T = 633$ observations dont nous avons utilisé également pour prédire. Nous avons choisi de projeter sur les k dernières valeurs observées pour $k = 98$ puis pour $k = 633$, par le logiciel Matlab (version 7.1) on a calculé les coefficients $\hat{a}_{i,k}$ par la méthode de Yule-Walker puis tous les prédicteurs à l'horizon 1 projetés sur les k dernières observations et on a tracé les prévisions (voir la figure (6.5)). On a calculé et tracé également la fonction d'autocovariance (voir la figure (6.6)) pour montrer la propriété de longue mémoire pour le modèle FARIMA par lequel on a modélisé la série de données.

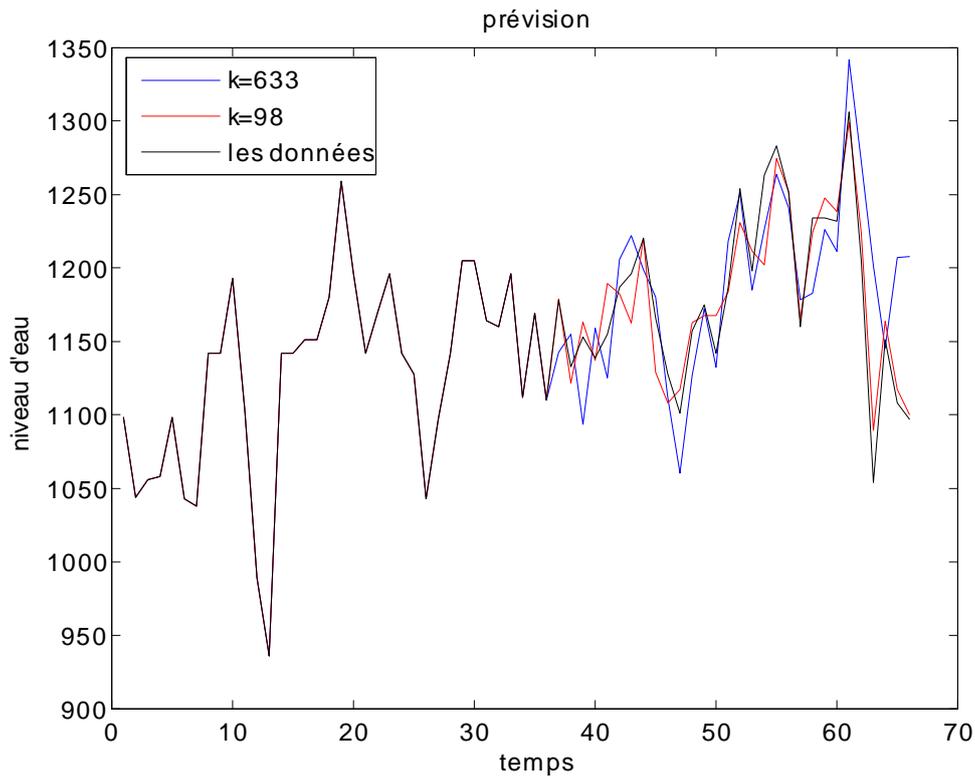


FIG. 6.5 – Prédiction à l'horizon 1 par le prédicteur projeté sur les k dernières valeurs.

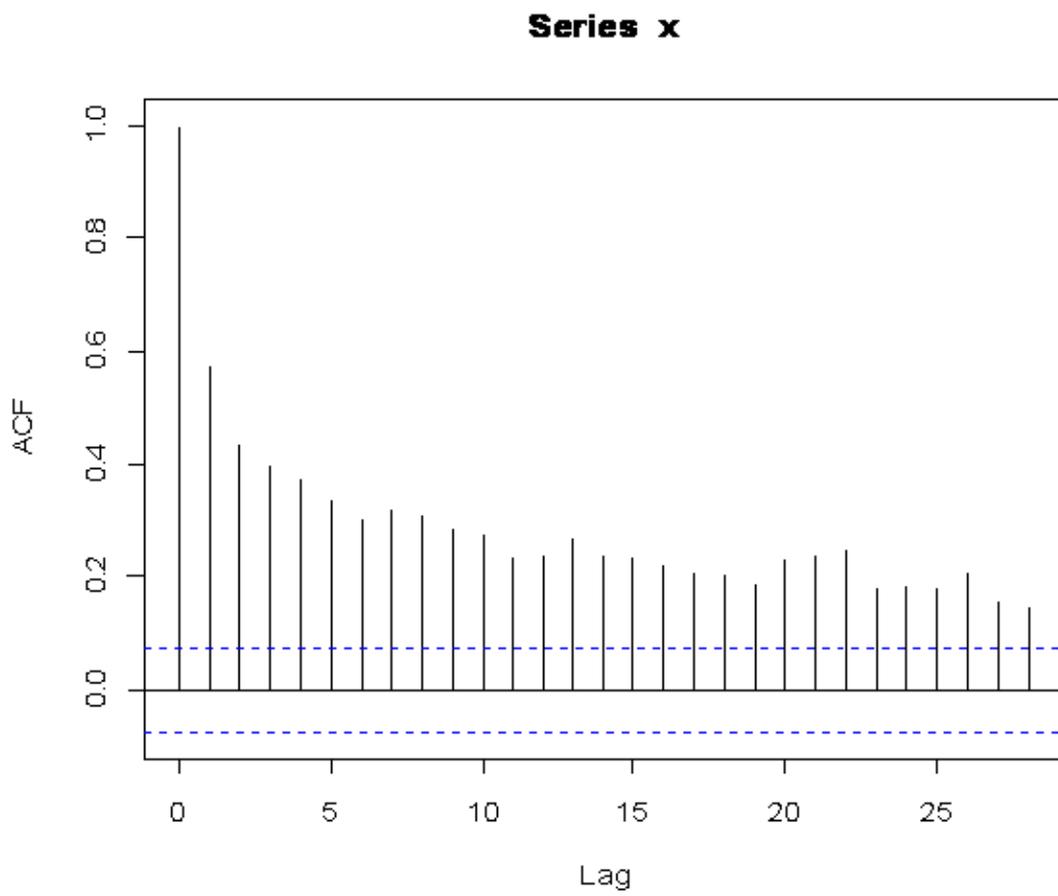


FIG. 6.6 – Fomction d'autocovariace.

6.3 Exemples

6.3.1 Etude de l'erreur quadratique due à la prévision d'un processus FARIMA(0,d,0)

Dans cette partie nous allons étudier l'erreur quadratique de prévision des prédicteurs que nous avons étudié dans notre travail 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 :

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

où $\hat{\theta}_T$ est l'estimateur de Whittle défini en (3.5). On estime

$$E \left[\left(X_{k_T+1} - \hat{X}_{T,k_T}(1) \right)^2 \right].$$

On simule $n = 10000$ itérations indépendantes de la variable aléatoire $X_{k_T+1} - \hat{X}_{T,k_T}(1)$. Puis on estime sa variance par la variance empirique S^2 pour laquelle l'intervalle de confiance de coefficient de confiance $1 - \alpha$ 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 α de la loi du χ^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 d le paramètre de longue mémoire.

On calcule deux jeux 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).

T	100			1000		
d	0.1	0.25	0.4	0.1	0.25	0.4
$k = 50$	$1.02_{\pm 0.03}$	$1.06_{\pm 0.03}$	$1.18_{\pm 0.03}$	$1.03_{\pm 0.03}$	$1.01_{\pm 0.03}$	$1.02_{\pm 0.03}$
$k = 100$	$1.03_{\pm 0.03}$	$1.06_{\pm 0.03}$	$1.16_{\pm 0.03}$	$0.98_{\pm 0.03}$	$1.04_{\pm 0.03}$	$1.01_{\pm 0.03}$

T	10000		
d	0.1	0.25	0.4
$k = 50$	$1.01_{\pm 0.03}$	$1.04_{\pm 0.03}$	$1.00_{\pm 0.03}$
$k = 100$	$1.01_{\pm 0.03}$	$1.01_{\pm 0.03}$	$1.00_{\pm 0.03}$

Tab.6.1 - Erreurs quadratiques empiriques pour le prédicteur de Wiener-Kolmogorov tronqué dont les coefficients sont estimés sur une série indépendante par l'estimateur de Whittle et leurs intervalles de confiance avec un coefficient de confiance à 0,95.

T	100			1000		
d	0.1	0.25	0.4	0.1	0.25	0.4
$k = 50$	$1.03_{\pm 0.03}$	$1.06_{\pm 0.03}$	$1.15_{\pm 0.03}$	$1.01_{\pm 0.03}$	$1.06_{\pm 0.03}$	$1.03_{\pm 0.03}$
$k = 100$	$1.04_{\pm 0.03}$	$1.06_{\pm 0.03}$	$1.15_{\pm 0.03}$	$1.01_{\pm 0.03}$	$1.02_{\pm 0.03}$	$1.04_{\pm 0.03}$

T	10000		
d	0.1	0.25	0.4
$k = 50$	$1.00_{\pm 0.03}$	$1.01_{\pm 0.03}$	$1.01_{\pm 0.03}$
$k = 100$	$1.03_{\pm 0.03}$	$0.99_{\pm 0.03}$	$0.98_{\pm 0.03}$

Tab.6.2 - Erreurs quadratiques empiriques pour le prédicteur de Wiener-Kolmogorov tronqué dont les coefficients sont estimés par l'estimateur de Whittle calculé sur la même série et leurs intervalles de confiance avec un coefficient de confiance à 0,95.

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 prédiction 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 montre 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 $d = 0.4$.

Le prédicteur projeté

Pour estimer l'erreur quadratique de prédiction pour le prédicteur projeté à l'horizon 1 étudié au chapitre 5 :

$$\tilde{X}_{T+1}(k) = \sum_{i=1}^k \hat{a}_{i,k} X_{T+1-i}$$

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 procédures 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 procédures de simulation dont les résultats sont résumés dans les tableaux 6.3 et 6.4. 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) [38]). Cette dégradation de l'estimation des coefficients et de la prédiction se voit bien lorsque T est égal à 100 (voir les tableaux 6.3 et 6.4). Lorsque le nombre k 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 s'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
$k = 50$	$1.48_{\pm 0.04}$	$1.58_{\pm 0.04}$	$1.75_{\pm 0.04}$	$1.05_{\pm 0.03}$	$1.05_{\pm 0.03}$	$1.07_{\pm 0.03}$
$k = 100$				$1.08_{\pm 1.03}$	$1.10_{\pm 0.03}$	$1.11_{\pm 0.03}$

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.6.3 - 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 avec un coefficient de confiance à 0,95.

T	100			1000		
d	0.1	0.25	0.4	0.1	0.25	0.4
$k = 50$	$1.20_{\pm 0.03}$	$1.32_{\pm 0.04}$	$1.54_{\pm 0.04}$	$1.04_{\pm 0.03}$	$1.05_{\pm 0.03}$	$1.07_{\pm 0.03}$
$k = 100$				$1.06_{\pm 1.03}$	$1.09_{\pm 0.03}$	$1.10_{\pm 0.03}$

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.6.4 - 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 avec un coefficient de confiance à 0,95.

Si on compare les deux méthodes de simulations selon que l'on estime les coefficients du prédicteur sur la même série ou sur une série indépendante, alors la comparaison est dans F faveur de l'estimation sur la même série dans tous les cas que nous avons testés. Dans nos résultats, rien ne laisse envisager ce résultat empirique.

6.3.2 Etude du niveau du lac majeur

Les données

Nous allons nous intéresser à la série correspondant à la quantité d'eau arrivée quotidiennement dans le lac Majeur (qui se trouve entre l'Italie et la

Suisse) entre le 1er janvier 1943 et le 30 avril 1994.

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.

Prévision par le prédicteur tronqué

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. L'estimateur de Whittle nous donne le modèle FARIMA suivant pour le résidu :

$$(1 - 0.11B)(1 - B)^{0.38}X_t = (1 + 0.23B)\varepsilon_t. \quad (6.1)$$

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. Si on calcule tous les prédicteurs à l'horizon 1 sur cette série pour tous 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 dans Y remplaçant la densité spectrale f par le périodogramme. On obtient $\hat{\sigma}_\varepsilon^2 = 2960$. L'autre valeur de référence 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.

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 p rojetant 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 (5.34) du Théorème 9. On applique alors le Théorème 9 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 nos k dernières observations, on a tracé en pointillé l'intervalle asymptotique de confiance à 95% (voir la figure (6.7)). Sur l'ensemble de la série, on trouve que 90% des valeurs prises par la série sont dans l'intervalle de confiance.

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 l'exemple précédent. 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.

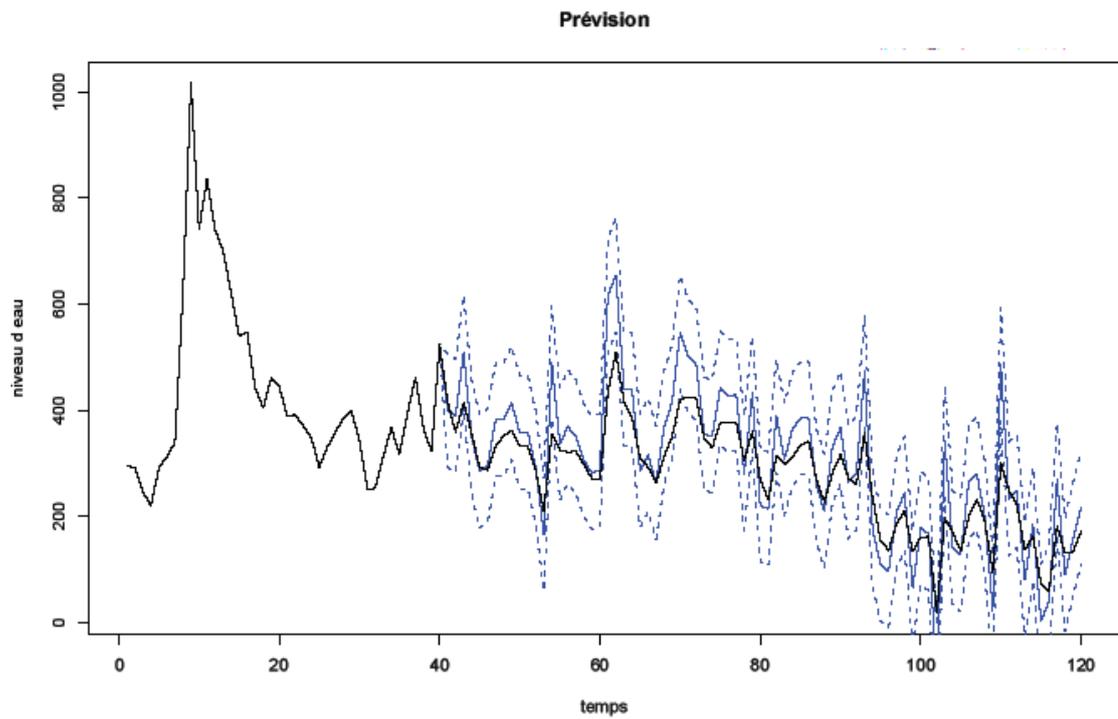


FIG. 6.7 – Prvisions à l’horizon 1 par le prédicteur projeté sur les $k = 80$ dernières valeurs observées et intervalles de confiance asymptotiques de coefficient de confiance 0,95.

Bibliographie

- [1] Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Stat.*, 2 :1095–1107.
- [2] Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Stat. Math.*, 22 :203– 217.
- [3] 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. Birkhauser Boston, Boston, MA.
- [4] Bhansali, R. (1974). Asymptotic properties of the Wiener-Kolmogorov predictor. I. *J. Roy. Statist. Soc. Ser. B*, 36 :61–73.
- [5] Bhansali, R. (1977). Asymptotic properties of the Wiener-Kolmogorov predictor. II. *J. Roy. Statist. Soc. Ser. B*, 39(1) :66–72.
- [6] Bhansali, R. (1978). Linear prediction by autoregressive model fitting in the time domain. *Ann. Stat.*, 6 :224–231. 129
- [7] 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.
- [8] Bhansali, R. et Kokoszka, P. (2001). Prediction of long-memory time series : An overview. *Estadística* 53, No.160-161, 41-96.
- [9] 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.
- [10] Bingham, N., Goldie, C., et Teugels, J. (1987). *Regular variation*. Encyclopedia of Mathematics and its applications, Vol. 27. Cambridge etc. : Cambridge University Press.
- [11] Boettcher, A. et Virtanen, J. (2006). Norms of Toeplitz Matrices with Fisher-Hartwig Symbols. Technical report, arXiv :math.FA/0606016.
- [12] Box, G. E. P. et Jenkins, G. M. (1976). *Time Series analysis : Forecasting and Control*, 3 rd ed. San Francisco : Holden-Day.

- [13] Brockwell, P. et Davis, R. (1988). Simple consistent estimation of the coefficients of a linear filter. *Stochastic Processes and their Application*.
- [14] Brockwell, P. et Davis, R. (1991). *Time Series : Theory and Methods*. Springer Series in Statistics.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] Fuller, W. (1976). *Introduction to statistical time series*. Wiley Series in Probability and Mathematical Statistics. New York.
- [19] Godet, F. (2007a). Linear prediction of long-memory processes : Asymptotic results on mean-squared errors. <http://arxiv.org/abs/0705.1927v1>.
- [20] Godet, F. (2007b). Linear Prediction of Long-Range Dependent Time Series. *ESAIM Probab. Statist.*
- [21] Godet, F. (2008). Prediction of long memory processes on same-realisation. <http://arxiv.org/abs/0712.1922>.
- [22] Grenander, U. et Szegő, G. (1958). *Toeplitz forms and their applications*. California Monographs in Mathematical Sciences. University of California Press, Berkeley.
- [23] Hidalgo, J. et Yajima, Y. (2002). Prediction and signal extraction of strongly dependent processes in the frequency domain. *Econometric Theory*, 18(3) :584–624.
- [24] 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.
- [25] Inoue, A. (1997). Regularly varying correlation functions and KMO-Langevin equations. *Hokkaido Math. J.*, 26(2) :457–482.
- [26] Inoue, A. (2000). Asymptotics for the partial autocorrelation function of a stationary process. *J. Anal. Math.*, 81 :65–109.
- [27] Inoue, A. et Kasahara, Y. (2006). Explicit representation of finite predictor coefficients and its applications. *Ann. Stat.*, 34(2) :973–993.

- [28] Kolmogorov, A. (1941). Stationary sequences in Hilbert's space. *Bull. Mosk. Gos. Univ. Mat.*, 2(6) :1–40.
- [29] Lewis, R. et Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.*, 16 :393–411.
- [30] Pourahmadi, M. (1989). On the convergence of finite linear predictors of stationary processes. *J. Multivariate Anal.*, 30(2) :167–180.
- [31] 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.
- [32] Ray, B. (1993a). Model selection and forecasting for long-range dependent processes. *International Journal of Forecasting*, 9(2) :255–269.
- [33] Ray, B. (1993b). Modeling long-memory processes for optimal long-range prediction. *J. Time Ser. Anal.*, 14(5) :511–525.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] Wei, C.-Z. (1987). Adaptative prediction by least-squares predictors in stochastic regression models with applications to time series. *Ann Stat.*, 15(4) :1667–1682.
- [38] 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.
- [39] 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 une seconde étape nous ne supposons plus la loi du processus connue, il nous faut alors estimer les fonctions de prévision obtenues dans la deuxième 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.

Summary : We study methods for forecasting 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 : Long-memory process, linear model, mean-squared forecast error, asymptotic normality.

خلاصة

في هذه المذكرة ندرس طرق التنبؤ لسلاسل المتغيرات العشوائية ذات الذاكرة الطويلة. بفرض أنها تحقق الاستقرار من الدرجة الثانية، خطية و انعكاسية. نفرض أولاً أننا نعرف قانون سلسلة المتغيرات العشوائية و المتقطع و $Kolmogorov-Wiener$ لكن لدينا عدد منته من الملاحظات للتنبؤ. نقتراح إذن متنبئين خطيين: متنبئ متنبئ مركب بالإسقاط على الملاحظات الماضية المنتهية. درسنا سلوكهما عندما يؤول عدد الملاحظات المتوفرة إلى ما لا نهاية. في خطوة ثانية لا نفرض بأننا نعرف قانون السلسلة، إذن يجب أن نقوم بتقدير المتقطع، نستعمل طريقة بار $Kolmogorov-Wiener$ توابع المتنبئ المحصل عليها في الجزء الأول. بالنسبة لمتنبئ المحسوب انطلاقاً من سلسلة مستقلة عن السلسلة التي بها نقوم $Whittle$ مثرية لتقدير معاملات المتنبئ بمقدر المعاملات التجريبية المحسوبة انطلاقاً من سلسلة التنبؤ أو سلسلة مستقلة عنها. بالنسبة لكلا المتنبئين، نقوم بتقدير الخطأ المربع الناتج عن تقدير المعاملات و نبين تغاربهما العادي

كلمات مفتاحية

سلاسل المتغيرات العشوائية ذات الذاكرة الطويلة، نموذج خطي، الخطأ المربع للتنبؤ، التغارب العادي