

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira de Béjaïa
Faculté des Sciences Exactes
Département de Mathématiques



Mémoire

En vue de l'Obtention du Diplôme de

Master en Mathématiques

Option : Probabilités Statistiques Applications

Thème

*Modélisation et prévision
par processus court et long mémoire*

Préparé par :

M^{elle} AOURIR Wardia

Soutenu devant le jury composé de :

Présidente	M ^{me} H. TABTI	M.C.B	Université de Béjaïa
Rapportrice	M ^{me} K.TIMERIDJINE	M.C.A	Université de Béjaïa
Examinatrice	M ^{elle} N. SAADI	M.C.B	Université de Béjaïa

Juin 2018

Remerciements

Au terme de notre modeste travail, je tiens à remercier :

Le bon Dieu, tous puissant de nous avoir accordé la volonté, force, santé, détermination et courage afin d'accomplir notre mémoire ;

Ma promotrice Mme TIMERIDJINE pour la manière dont elle nous a encadré ainsi que sa compréhension vis à vis de notre disponibilité ;

Je suis également reconnaissante pour la confiance qu'elle nous a accordée, ses remarques, sa compréhension, ses conseils et sa grande disponibilité qui nous a donné plus de volonté pour réaliser ce mémoire ;

Les membres du jury : Madame H. TABTI qui a honoré de sa présence en présidant ce jury et mademoiselle N. SAIDI d'avoir accepté d'examiner notre travail ;

Tout particulièrement je souhaite remercier vivement les premières personnes qui ont cru en moi ; Monsieur REHAL, Monsieur HABET le Directeur de l'Environnement , Melle REDOUANI ma Chef de Service ainsi que tous mes collègues ;

A tous les enseignants du département de Mathématiques pour leur disponibilité et leurs savoir faire durant toute cette année de formation à leurs tête Mme BOURAINE et Mme LAKADIR

A tous ceux et celles qui ont contribué de manière directe ou indirecte pour la réalisation de ce travail.

Dédicaces

- * Toutes les lettres ne sauraient trouver les mots qu'il faut ...
 - * Tous les mots ne sauraient exprimer la gratitude
 - * L'amour, le respect, la reconnaissance ...
 - * Aussi, c'est tout simplement que

*J*e dédie ce modeste travail

- * *A* ma très chère maman et à la mémoire de mon père ;
 - * *A* mes chères grand père et grand mère ;
 - * *A* mes chères tantes ;
 - * *A* mes frères ;
 - * *A* ma petite soeur ;

A toutes les personnes qui ont participé a l'élaboration de ce travail à
tous ceux que j'ai omis de citer.

Table des matières

Liste des figures	v
Liste des tableaux	1
Introduction	2
1 Généralités sur les processus stochastiques et les séries temporelles	5
1.1 Processus stochastiques stationnaires	6
1.2 Fonctions d'auto-covariance	8
1.3 Fonctions d'auto-corrélation simple et partielle	9
1.3.1 La fonction d'auto-corrélation (FAC)	9
1.3.2 La fonction d'auto-corrélation partielle (FAP)	9
1.4 Tests de stationnarité(ou tests de racine unitaire)	10
1.4.1 Test de Dickey-Fuller simple	10
1.4.2 Test de Dickey-Fuller Augmenté	11
1.5 Processus non stationnaire	12
1.5.1 Le processus TS	12
1.5.2 Le processus DS	13
1.6 Séries temporelles (chronologiques)	14

1.6.1	Les trois composantes d'une série chronologique	14
1.6.2	Les modèles de composition des trois composantes d'une série	15
1.6.3	Choix du modèle	15
1.7	Prévisions	16
2	Les processus stationnaires à court mémoire	18
2.1	Typologie des modèles AR , MA et $ARMA$	19
2.1.1	Modèle AR (Autorégressif)	19
2.1.2	Modèle MA (Moving Average : Moyenne Mobile)	20
2.1.3	Modèle $ARMA$ (Combinaison de processus AR et MA)	21
2.2	Processus $ARIMA$ et $SARIMA$	22
2.3	Modélisation par la méthode de Box et Jenkins	23
2.3.1	Les étapes de la méthode de Box et Jenkins :	24
3	Les processus à longue mémoire "ARFIMA"	31
3.1	Les processus à mémoire longue	32
3.2	Processus $ARFIMA(p, d, q)$	33
3.3	Processus long mémoire pure ($FI(d)$ ou $ARFIMA(0, d, 0)$)	34
3.3.1	Causalité	34
3.3.2	Inversibilité	35
3.3.3	Fonction d'auto-covariance	36
3.3.4	Fonction d'auto-corrélation	36
3.3.5	Propriétés	36
3.4	Estimation du paramètre d	37
3.5	Tests de mémoire longue	40

4	Application : modélisation et prévision	42
4.1	Première approche "Méthode Box Jenkins"	43
4.1.1	Étape 1 : Transformation des données afin de stabiliser la variance (<i>log, sqrt, ...</i>) et différenciations des données pour les stationnariser	44
4.1.2	Étape 2 : Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés.	46
4.1.3	Étape 3 : Estimation des paramètres du (des) modèle(s) sélectionné(s).	46
4.1.4	Étape 4 : Diagnostic et tests adéquation du modèle.	47
4.1.5	Étape 5 : Prévision	48
4.2	Deuxième approche "Méthode de Geweke Porter-Hudak (GPH)"	49
4.2.1	Étape 1 : Transformation des données afin de stabiliser la variance (<i>log, sqrt, ...</i>) et différenciations des données pour les stationnariser	50
4.2.2	Étape 2 : Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés.	52
4.2.3	Étape 3 : Estimation des paramètres du (des) modèle(s) sélectionné(s).	52
4.2.4	Étape 4 : Diagnostic et tests adéquation du modèle.	53
4.2.5	Étape 5 : Prévision	54
	Conclusion	58
	Annexe	60
	A Annexes	60

Table des figures

1.1	Exemple de la normalité.	15
1.2	Exemple d'un modèle multiplicatif.	16
2.1	Exemple de la normalité.	28
4.1	Évolution de X_t depuis 1998 à 2018, transformée en logarithme.	44
4.2	Corrélogrammes de la série X_t	45
4.3	La série X_t stationnaire.	45
4.4	Corrélogrammes de la série X_t	46
4.5	Identification des paramètres.	47
4.6	Teste de normalité des résidus $ARIMA(2, 1, 0)$	48
4.7	Prévision des exportations Algériennes.	48
4.8	Prévision de " X_t " sur un intervalle de confiance.	49
4.9	Évolution de la série " cp " depuis 2006 jusqu'à 2018.	50
4.10	Corrélogrammes de la série cp	51
4.11	La série cp stationnaire.	51
4.12	Corrélogramme de la série cp stationnaire.	52
4.13	Teste de normalité des résidus $ARFIMA(10, 0.255857, 9)$	53

4.14	Prévision de "cp" à long terme.	55
4.15	Prévision des "cp" sur un intervalle de confiance.	55
A.1	Stratégie simplifiée des tests de racine unitaire.	61
A.2	La modélisation de Box-Jenkins.	62

Liste des tableaux

2.1	Résumé des propriétés des modèles <i>ARMA</i>	22
4.1	Modèle retenu.	47
4.2	Comparaison des prévisions avec les valeurs réelles.	49
4.3	Identification des paramètres.	54
4.4	Comparaison des prévisions avec les valeurs réelles.	56

Introduction

Nous pouvons considérer qu'un modèle consiste en une présentation formalisée d'un phénomène sous forme d'équations mathématiques dont les variables sont des grandeurs relatives aux phénomènes à étudier dans le but de les expliquer. La construction d'un modèle comporte un certain nombre d'étapes qui sont toutes importantes tel que : référence à une théorie, formalisation, sélection des variables . . . etc. depuis les informations disponibles à l'instant "t" où elle a été effectuée les chercheurs tendent d'estimer des modèles et des paramètres qui constituent un ensemble de probabilités associées à un ensemble d'événements futurs d'où des prévisions sont construites.

La prévision représente un outil essentiel pour gérer des diverses situations, car prévoir et ensuite préparer un plan de travail aide à éviter beaucoup de problèmes et de gérer la situation, étant armé de certaines informations cela, nous permet de planifier, identifier, gérer et réagir rapidement.

Les séries chronologiques ont connu un certain dynamisme, car ces applications sont très larges : commerce extérieur, Astronomie, environnement .. etc. chaque domaine était une cause, d'une manière ou d'une autre, d'apparition d'un modèle mathématique ou d'un développement ou un prolongement. Mais l'innovation est marquée par le livre de Box et Jenkins (1976) [8], où les principales propriétés des processus stationnaires autorégressifs moyenne mobile (*ARMA*) ont été décrites avec les méthodes d'identification, d'estimation et de validation.

A cet effet, si on est devant un phénomène économique, en particulier, l'analyse des exportations ou bien l'indice boursier, constitue à lui seul une activité de recherche importante

très liée à celle effectuée en séries chronologiques. Cela sera intéressant de déduire le quel de ses processus sus-cités est adéquat.

L'objet de notre mémoire porte sur l'inference statistique dans une classe de modèles à court et longue mémoire, plus précisément, les modèles *ARIMA* (autorégressifs, moyenne mobile, intégrés) ensuite fractionnaires *ARFIMA*(p, d, q) définis par l'équation suivante :

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t.$$

Avec B l'opérateur retard et ε_t un processus de bruit blanc centré et de variance inconnue.

La question concernant la persistance des chocs occupe une place centrale particulièrement en économie, même si l'utilisation de la mémoire longue est nettement moins nombreuse que celle de la mémoire courte.

Dans ce contexte, Diebold et Rudebusch (1989) [12] ont appliqué la procédure de Geweke et Porter-Hudak (1983) [16] sur la série américaine du PNB trimestriel d'après guerre. Ils mettent en avant la présence d'une structure de dépendance de long terme dans cette série .

Un tel résultat est confirmé par Sowell (1992) [26] qui, utilisant la méthode du maximum de vraisemblance exact, met en avant la supériorité des processus *ARFIMA* sur les processus de type *ARIMA* et *ARMA* usuels.

L'objectif de notre mémoire est d'établir une étude statistique sur "les exportations Algérienne" d'une part et "l'indice de clôture quotidien du marché boursier koweïtien" d'autre part, vu l'importance capitale que porte ce domaine comme un outil d'aide à la décision dans différents domaines.

Ce thème nous procure une motivation suffisante pour l'étudier. Dans ce cadre, on présenteras les deux processus à savoir à court mémoire (*AR, MA, ARMA, ARIMA, . . . , etc*) d'une part et les processus à long terme (*ARFIMA*) d'autre part. Dans le même contexte on doit se poser certaines questions de développement, à savoir :

- Quels sont les processus " court terme " et " long terme " ?
- Quels sont les conditions d'utilisation de chaque processus ?
- Quelle est la décomposition de chaque processus ainsi que ses caractéristiques ?
- Quelles sont les prévisions à court termes et a long terme ?

A cet effet, nous allons présenter dans notre mémoire une modeste recherche relative à l'objet, structurée comme suit :

Le chapitre 1 consacré aux notions fondamentales relatives aux processus stochastiques et séries temporelles, ensuite nous allons entamer le chapitre 2 qui présentera les processus stationnaires à courte mémoire tel que les différentes classes des modèles (AR , MA , $ARMA$, \dots etc) et leurs propriétés, suivie de la démarche d'analyse la plus fréquente "la méthode Box et Jenkins", le chapitre 3 sera présenté d'une manière équivalente au chapitre 2 sur les processus à longue mémoire précisément " $ARFIMA$ " (Autorégressifs, Moyenne Mobile, Fractionnairement Intégrés) et ces propriétés suivi aussies de la démarche d'analyse la plus utilisée "la méthode de Geweke Porter-Hudak", le chapitre 4 sera réservé à l'application des deux méthodes présentées dans le chapitre 2 et 3.

1

Généralités sur les processus stochastiques et les séries temporelles

Introduction

L'analyse des séries chronologiques se focalise essentiellement sur deux questions, l'identification et l'estimation statistique et prévision.

D'importantes corrélations pour de faibles retards peuvent être parfois détectées en général par les processus à courtes mémoires tel que le modèle *ARIMA*, cela était avant 1951, car au delà de cette année, l'étude de longue mémoire est apparue grâce à son fondateur Hurst [20] sur les crues du Nil, ces travaux qui ont par la suite été développés par plusieurs chercheurs tel que Mandelbrot [23], Lardic et mignon [25], et ce, dans divers

domaines.

Comme on va le constater dans notre travail, ces études soit à courte ou à longue mémoire sont intégrales, car les processus à longue mémoire tel que le modèle *ARFIMA* noté aussi *FARIMA* "Autoregressive, Moyenne mobile, fractionnairement intégrés" ne sont qu'un prolongement des processus *ARIMA* définis comme processus à courte mémoire.

Un processus aléatoire généralise la notion de variable aléatoire utilisée en statistiques élémentaires. On le définit comme une famille de variables aléatoires $(X_t, t \in T)$ (souvent, on aura $T = \mathbb{R}, \mathbb{R}_+, \mathbb{N} \dots$), l'ensemble des observations disponibles (X_t) constitue une réalisation du processus.

On distingue généralement les processus en temps discret et en temps continu de \mathbb{R} , dans ce mémoire nous nous intéresserons au cas discret qui porte sur une durée déterminée [7].

A cette effet, ce chapitre introductif afin d'entamer les trois chapitres ou on va présenter les processus objet de notre mémoire.

Définition 1.1. : Un processus stochastique est une famille de variables aléatoires indicées par le temps t noté $(X_t, t \in T)$. Dans ce qui suit, l'espace des indices T est le temps, t est alors l'instant d'observation de la variable aléatoire X sur l'individu ω , $\omega \in \Omega$ avec Ω l'espace fondamental permettant de définir la variable aléatoire.

Si T est un intervalle de \mathbb{R} , le processus est dit continu, sinon si $T \subseteq \mathbb{Z}$, le processus est dit discret.

Définition 1.2. : Opérateur retard L'opérateur retard noté B , qui décale le processus d'une unité de temps vers le passé : $B(X_t) = X_{t-1}$.

Si on applique h fois cet opérateur, on décale le processus de h unité de temps :

$$\underbrace{B(B(\dots B(X_t) \dots))}_{h \text{ fois}} = B^h(X_t) = X_{t-h}. \quad (1.1)$$

$h \in T$.

1.1 Processus stochastiques stationnaires

La stationnarité est une caractéristique d'un processus stochastique qui implique que le comportement de la série ne dépend pas du temps. Un processus stochastique ne comporte

pas de tendance saisonnière, ni de tendance à la hausse ou à la baisse.

Définition 1.3. $(X_t, t \in T)$ est **strictement stationnaire** si :

$$E(X_t) = m, \forall t \in T \text{ (constante ne dépend pas du temps);}$$

$$\text{Var}(X_t) = \sigma^2 < \infty, \forall t \in T \text{ (constante ne dépend pas du temps);}$$

$$\text{Cov}(X_t, X_{t+h}) = E[(X_{t+h} - m)(X_t - m)] = \gamma_h, \forall t \in T \text{ (constante ne dépend pas du temps).}$$

Définition 1.4. : Le processus $(X_t, t \in T)$ est dit **stationnaire au second ordre**, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :

$$E(X_t) = m, \forall t \in T \text{ (constante ne dépend pas du temps);}$$

$$E(X_t^2) < \infty, \forall t \in T;$$

$$\text{Cov}(X_t, X_{t+h}) = \gamma_h, \forall t \in T \text{ (constante ne dépend pas du temps).}$$

Remarque 1.1. :

- *La plupart des processus stochastiques ne sont pas stationnaires, mais on peut se ramener à ce cas par des transformations (logarithmes).*
- *Pour un processus du second ordre, Une suite X de v.a. indépendantes de même moyenne et même variance est toujours stationnaire à l'ordre 2; mais si les $(X_n, n \in \mathbb{N})$ n'ont pas tous la même loi, X n'est pas stationnaire au sens strict[7].*

Définition 1.5. Un processus $(\varepsilon_t; t \in \mathbb{Z})$ est dit bruit blanc si $(\varepsilon_t; t \in \mathbb{Z})$ est une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance nulle et de variance constante.

$$E(\varepsilon_t) = 0, \forall t;$$

$$\text{Var}(\varepsilon_t) = \sigma^2, \forall t;$$

$$\text{Cov}(\varepsilon_t, \varepsilon_{\dot{t}}) = 0, \forall t \neq \dot{t};$$

Un tel processus n'a ni tendance ni mémoire : la connaissance de la valeur du processus à une date donnée n'apporte aucune information pour la prédiction de sa valeur à une date ultérieure.

Définition 1.6. :Un processus $(X_t, t \in T)$ est **inversible** s'il peut s'écrire comme combinaison linéaire des valeurs d'un autre processus, c'est à dire qu'il existe une suite $(\psi_i, i \in \mathbb{Z})$ tels que :

$$\varepsilon_t = \sum_{i \geq 0} \psi_i X_{t-i}. \quad (1.2)$$

Où $\sum_{i \in \mathbb{Z}} |\psi_i| < \infty$.

Définition 1.7. :Un processus stochastique $(X_t, t \in \mathbb{Z})$ est **causal** si les innovations s'écrivent en fonction des erreurs passées :

$$X_t = \sum_{j \geq 0} \psi_j \varepsilon_{t-j}. \quad (1.3)$$

1.2 Fonctions d'auto-covariance

Dans un processus $(X_t, t \in T)$ la variable aléatoire au temps t dépend généralement, d'une certaine façon, des observations précédentes (X_i) . Une manière de comprendre les liens d'indépendance entre les termes d'une telle série consiste à considérer les fonctions d'auto-covariance et d'auto-corrélation, ces dernières mesurent respectivement la covariance et la corrélation entre les termes d'une série chronologique et caractérisent complètement le processus[11]. La fonction d'auto-covariance mesure la covariance entre une variable et cette même variable à des dates différentes, pour un délai $h \in \mathbb{Z}$ et $\hat{t} = t + h$:

$$cov(X_t, X_{\hat{t}}) = E[(X_t - E(X_t))(X_{\hat{t}} - E(X_{\hat{t}}))]. \quad (1.4)$$

ainsi si $h = 0$:

$$V(X_t) = E[(X_t - E(X_t))^2] = \sigma_t^2.$$

1.3 Fonctions d'auto-corrélation simple et partielle

1.3.1 La fonction d'auto-corrélation (FAC)

La fonction d'auto-corrélation est la fonction notée ρ_h qui mesure la corrélation de la série avec elle-même décalée de h périodes.

On étudie la "mémoire" d'un processus en calculant son auto-corrélation de retard h et $\hat{t} = t + h$ noté :

$$\rho_h = \text{Corr}(X_t, X_{\hat{t}}) = \frac{\text{Cov}(X_t, X_{\hat{t}})}{\sqrt{V(X_t)V(X_{\hat{t}})}}. \quad (1.5)$$

qui mesure le lien entre les valeurs du processus à deux dates distantes de h . ρ_h Pour un processus stationnaire, prend une forme plus simple :

$$\rho_h = \frac{\text{Cov}(X_t, X_{t+h})}{V(X_t)} = \frac{\gamma_k}{\gamma_0}. \quad (1.6)$$

On peut tracer la courbe $\rho_h = f(h)$ qui est appelée auto-corrélogramme [11].

Propriétés 1.1. *La fonction d'auto-corrélation d'un processus stationnaire vérifie :*

$$\begin{aligned} \forall h \in Z, \rho(-h) &= \rho(h); \\ \rho(0) &= 1; \\ |\rho(h)| &\leq 1. \end{aligned}$$

1.3.2 La fonction d'auto-corrélation partielle (FAP)

Le coefficient de corrélation partielle mesure la liaison entre deux variables lorsque l'influence d'une ou des autres variables explicatives est retirée. Le coefficient de corrélation partielle comme étant le calcul de l'influence de X_1 sur X_2 en éliminant les influences des autres variables X_3, X_4, \dots, X_h .

On définit l'auto-corrélation partielle de retard h comme la corrélation entre $(X_t - X_t^*)$ et $(X_{t-h} - X_{t-h}^*)$ ou X_t^* désigne la régression de X_t sur les $(h-1)$ valeurs $(X_{t-1}, X_{t-2}, \dots, X_{t-h+1})$:

$$\tau_h = \text{Corr}(X_t - X_t^*, X_{t-h} - X_{t-h}^*) = \frac{\text{Cov}(X_t - X_t^*, X_{t-h} - X_{t-h}^*)}{\sqrt{V(X_t - X_t^*)V(X_{t-h} - X_{t-h}^*)}}. \quad (1.7)$$

avec :

$$X_t^* = \sum_{k=1}^{h-1} \alpha_k X_{t-k};$$

$$X_{t-h}^* = \sum_{k=1}^{h-1} \beta_k X_{t-k}. \text{Où } \alpha_k, \beta_k \text{ sont les coefficients des régressions.}$$

Cette quantité rend compte de l'intensité de la liaison entre X_t et X_{t-h} en supprimant les liaisons induites par des variables intermédiaires ($X_{t-1}, X_{t-2} \cdots X_{t-h+1}$) [11].

1.4 Tests de stationnarité(ou tests de racine unitaire)

Il existe plusieurs tests de racine unitaire : test de Dickey-Fuller simple et Dickey-Fuller augmenté, test de phillips et perron, test de Kwiatkowski, philips .. etc. Nous ne présenterons dans notre travail que les deux premiers[17].

1.4.1 Test de Dickey-Fuller simple

Les tests de Dickey-Fuller (DF) permettent de mettre en évidence le caractère stationnaire ou non d'une chronique par la détermination d'une tendance déterministe ou stochastique, ainsi de déterminer la bonne manière de stationnariser la série.

si l'hypothèse $H_0 : \Phi_1 = 1$ est retenue dans l'un de ces trois modèles, le processus est

alors non stationnaire. Les hypothèses du test sont les suivantes :

$$\begin{cases} \mathbf{H0} : & \Phi_1 = 1; \\ \mathbf{H1} : & |\Phi_1| < 1. \end{cases}$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

Sous H_0 vrai, la statistique de test pour l'estimateur de Φ_1 est donné par :

$$t_{\hat{\Phi}_1} = \frac{\hat{\Phi}_1 - 1}{\hat{\sigma}_{\hat{\Phi}_1}}. \quad (1.8)$$

Un processus non stationnaire correspond à l'un des modèles suivants :

Modèle 1 : $X_t = \Phi_1 X_{t-1} + \varepsilon_t$, Modèle autorégressif d'ordre 1 ;

Modèle 2 : $X_t = \Phi_1 X_{t-1} + c + \varepsilon_t$, Modèle autorégressif avec constante ;

Modèle 3 : $X_t = \Phi_1 X_{t-1} + b_t + c + \varepsilon_t$, Modèle autorégressif avec tendance.

Dans le **modèle 3**, si on accepte $H1 : |\Phi_1| < 1$ et si le coefficient b est significativement différent de 0, alors le processus est un processus TS , on peut le rendre stationnaire en calculant les résidus par rapport a la tendance estimée par les moindres carrés ordinaires. On estime par les moindres carrés ordinaires le paramètre Φ_1 noté $\hat{\Phi}_1$ pour les trois modèles. L'estimation des coefficients et des écarts types du modèle par les moindres carrés ordinaires fournit $t_{\hat{\Phi}_1}$ qui est analogue a la statistique de Student (rapport du coefficient sur son écart type). Si $t_{\hat{\Phi}_1} \geq t_{\text{tabulée}}$, alors on accepte l'hypothèse $H0$; il existe une racine unitaire, le processus n'est donc pas stationnaire.

1.4.2 Test de Dickey-Fuller Augmenté

Dans le test de Dickey-Fuller que nous venons de présenter, le processus ε_t est par hypothèse un bruit blanc, or il n'y a aucune raison pour qu'à priori, l'erreur soit non corrélée.

On à : $X_t = \Phi_1 X_{t-1} + \varepsilon_t$, s'écrit aussi :

$$X_t - X_{t-1} = \Phi_1 X_{t-1} - X_{t-1} + \varepsilon_t;$$

$$\Delta X_t = (\Phi_1 - 1)X_{t-1} + \varepsilon_t.$$

Le test de Dickey-Fuller augmenté se définit de la façon suivante :

$$\begin{cases} \mathbf{H0} : & \Phi_1 = 1; \\ \mathbf{H1} : & |\Phi_1| < 1. \end{cases}$$

$\varepsilon_t \sim N(0, \sigma^2)$ Un processus non stationnaire correspond à l'un des modèles suivants :

$$\text{Modèle 1 : } \Delta X_t = \rho X_{t-1} - \sum_{k=2}^p \gamma_k \Delta X_{t-k+1} + \varepsilon_t;$$

$$\text{Modèle 2 : } \Delta X_t = \rho X_{t-1} - \sum_{k=2}^p \gamma_k \Delta X_{t-k+1} + c + \varepsilon_t;$$

$$\text{Modèle 3 : } \Delta X_t = \rho X_{t-1} - \sum_{k=2}^p \gamma_k \Delta X_{t-k+1} + b_t + c + \varepsilon_t.$$

Le test se déroule de manière similaire aux tests DF simples, seule la distribution statistiques de différent. La valeur de p peut être déterminée selon les critères de Akaike ou de Schwarz, ou encore, en partant d'une valeur suffisamment importante de p , on estime un modèle à $p-1$ retards, puis à $p-2$ retards, jusqu'à ce que le coefficient du $p^{\text{ème}}$ retard soit significatif.

1.5 Processus non stationnaire

Les séries non stationnaires peuvent être soit un processus TS ou DS.

1.5.1 Le processus TS

Le processus TS (Trend Stationary) s'écrit sous la forme suivante :

$$X_t = \alpha + \beta_t + \varepsilon_t. \quad (1.9)$$

Où ε_t représente l'erreur du modèle à la date t .

Il présente une non stationnarité de nature déterministe. Le processus TS est non stationnaire car $E(X_t) = \alpha + \beta_t$ dépend du temps.

Le processus X_t peut être stationnarisé en retranchant à X_t la valeur estimée $\hat{\alpha} + \hat{\beta}_t$ par la méthode des moindres carrés ordinaires[17].

1.5.2 Le processus DS

Le processus DS (Differency stationary) s'écrit sous la forme suivante :

$$X_t = X_{t-1} + \beta_t + \varepsilon_t. \quad (1.10)$$

- Si ($\beta_t \neq 0$) le processus DS est dit **avec dérive**. Il est appelé aussi marche au hasard (ou marche aléatoire). Il présente une non stationnarité de nature stochastique.

Par récurrence, on obtient (dans le cas avec dérive) :

$$\begin{aligned} X_1 &= X_0 + \beta + \varepsilon_1; \\ X_2 &= X_1 + \beta + \varepsilon_2 = X_0 + 2\beta + \varepsilon_1 + \varepsilon_2; \\ &\vdots \\ X_t &= X_0 + \beta_t + \sum_{i=1}^t \varepsilon_i. \end{aligned}$$

avec ε_i indépendant et identiquement distribuées. Le processus DS avec dérive est non stationnaire car on a $E(X_t) = X_0 + \beta_t$ qui dépend du temps t .

- Si ($\beta_t = 0$) le processus DS est dit **sans dérive**. Par récurrence, on obtient (dans le cas sans dérive) :

$$\begin{aligned} X_1 &= X_0 + \varepsilon_1; \\ X_2 &= X_1 + \varepsilon_2; \\ &\vdots \\ X_t &= X_0 + \sum_{i=1}^t \varepsilon_i. \end{aligned}$$

avec ε_i indépendant et identiquement distribuées. Le processus DS sans dérive est non stationnaire car on a :

$$Var(X_t) = Var\left(\sum_{i=1}^t \varepsilon_i\right) = t\sigma^2.$$

On constate que la variance du processus DS sans dérive dépend du temps.

Pour stationnariser le processus DS (avec ou sans dérive) , il suffit de le passer en

différence première[17] :

$$X_t - X_{t-1} = \beta + \varepsilon_t (\text{cas avec dérive});$$

$$X_t - X_{t-1} = \varepsilon_t (\text{cas sans dérive}).$$

1.6 Séries temporelles (chronologiques)

Une série chronologique est un ensemble de données mesurant un phénomène est repéré dans l'ordre croissant du temps son analyse repose sur :

- L'observation des valeurs dans le passé à des dates fixées et équidistantes ;
- La suggestion d'hypothèses de travail permettent de justifier l'emploi de certaines méthodes de prévisions.

Le but de la décomposition d'une série chronologique est de distinguer dans l'évolution de la série, une tendance " générale ", des variations saisonnières qui se répètent chaque année, et des variations accidentelles imprévisibles.

L'intérêt de ceci est d'une part de mieux comprendre, de mieux décrire l'évolution de la série, et d'autre part de prévoir son évolution (à partir de la tendance et des variations saisonnières).

1.6.1 Les trois composantes d'une série chronologique

- i) La tendance : C_t la tendance correspond à l'évolution à long terme de la série, l'évolution fondamentale de la série ;
- ii) Les variations saisonnières : S_t les variations saisonnières sont des fluctuations périodiques à l'intérieur d'une année, et qui se reproduisent d'une façon plus ou moins permanente d'une année sur l'autre.
Ces variations sont dues au rythme des saisons : matières premières, congés, ..., etc ;
- iii) Les variations accidentelles ou résiduelles : ε_t les variations accidentelles sont des fluctuations irrégulières et imprévisibles. Elles sont en général de faibles amplitudes. Elles proviennent de circonstances non prévisibles : catastrophes naturelles, crises boursières, grèves, .., etc.

1.6.2 Les modèles de composition des trois composantes d'une série

On va se limiter sur les deux modèles les plus fréquents.

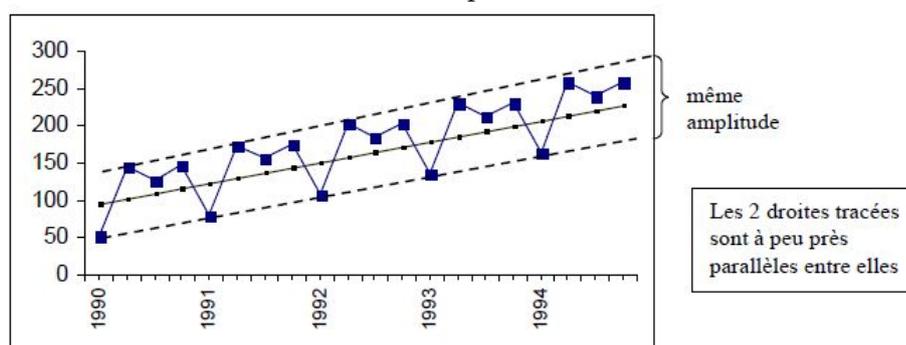
- i) **Le modèle additif** : Dans un modèle additif, on suppose que les trois composantes : tendance, variations saisonnières et variations accidentelles sont indépendantes les unes des autres.

On considère que la série X_t s'écrit comme la somme de ces trois composantes :

$$X_t = C_t + S_t + \varepsilon_t.$$

Graphiquement, l'amplitude des variations est constante autour de la tendance

FIGURE 1.1 – Exemple de la normalité.



Source : [25]

- ii) **Le modèle multiplicatif** : Dans ce modèle, l'amplitude de la composante saisonnière et du bruit n'est plus constante au cours du temps : elle varie au cours du temps proportionnellement à la tendance C_t :

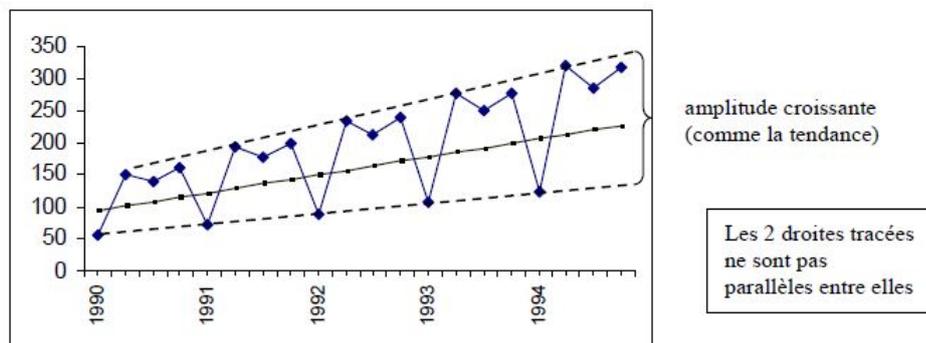
$$X_t = C_t \times S_t \times \varepsilon_t. \quad (1.11)$$

Graphiquement, l'amplitude des variations (saisonniers) varie.

1.6.3 Choix du modèle

- i) **Méthode de la bande** : On utilise le graphe de la série et la droite passant par les minima et celle passant par les maxima.

FIGURE 1.2 – Exemple d'un modèle multiplicatif.



Source :[25]

- Si ces 2 droites sont à peu près parallèles : le modèle est additif(voir figure :1.1) ;
- Si ces 2 droites ne sont pas parallèles : le modèle est multiplicatif(voir figure :1.2).

ii) **Méthode du profil** : On utilise le graphique des courbes superposées.

- Si les différentes courbes sont à peu près parallèles : le modèle est additif(voir figure :1.1) ;
- Sinon (les pics et les creux s'accroissent) : le modèle est multiplicatif(voir figure :1.2).

A cet effet, Pour décomposer une série chronologique on doit commencer par :

1. Tracer son graphique ;
2. Choisir un modèle de composition (additif ou multiplicatif) ;
3. Estimer la tendance C_t ;
4. Estimer les variations saisonnières.

1.7 Prévisions

La prévision des processus stochastiques représente l'un des intérêts majeurs de nos jours (prévision de température, des prix, des indices boursiers, . . . etc) du fait qu'elle aide à l'établissement de futures stratégies et à la prise de décisions pour les buts de planification . . . etc. Dans le but de réaliser des prévisions, plusieurs méthodes ont été envisagées, tel qu'on va le constater dans les chapitres 2 et 3. Les formules et les calculs

associés à l'analyse statistique et à la prévision peuvent être très complexes, à cet effet, des techniques prévisionnelles ont été mises en œuvre telle que la modélisation Box-Jenkins, le lissage exponentiel \dots etc. Pour se faire, on cherche tout d'abord un modèle adéquat pour la série Chronologique étudiée et on procède à une analyse systématique de celle-ci dans le but de déterminer ses caractéristiques et de voir si celles-ci correspondent bien à celles d'un des modèles mentionnés. En outre, on procède à l'estimation des paramètres du modèle considéré et on teste sa compatibilité avec l'un des modèles autorégressifs, le modèle obtenu est alors utilisé pour le calcul des prévisions des processus stochastiques considérés.

Conclusion

Dans ce chapitre on a essayé de présenter des généralités sur les processus aléatoires et les séries temporelles nécessaires et indispensables, à fin d'entamer les trois chapitres objet de notre mémoire, vu que ces notions sont indispensables.

2

Les processus stationnaires à court mémoire

Introduction

Les techniques traditionnelles de prévision des séries temporelles se révèlent dans la plupart des cas insuffisantes pour prévoir les phénomènes économiques, c'est ce qui explique l'engouement des chercheurs à investir ce domaine dans les années 1970 suite aux travaux de Box et Jenkins qui proposent une nouvelle " philosophie " du traitement des séries temporelles.

Cependant, nous nous intéresserons ici uniquement à une catégorie particulière de processus à mémoire courte (*ARIMA* autorégressifs, moyenne mobile, intégrés). Dans un premier temps, nous nous attacherons à la représentation de la mémoire courte des modèles *AR*, *MA* et *ARMA* (autorégressifs, moyenne mobile et mixte) ensuite à l'élargissements de ces derniers nommés (*ARIMA* et *SARIMA*), au travers d'un enchaînement de façon

équivalente, de la définition des modèles avant de détailler de la manière la plus exhaustive possible les diverses étapes de modélisation et prévision, choisissant la méthode la plus utilisée, à savoir, Box et Jenkins.

2.1 Typologie des modèles AR, MA et $ARMA$

Dans l'étude d'une série chronologique, il est naturel de penser que la valeur de la série à la date t peut dépendre des valeurs prises aux dates précédentes :

$$X_t = f(X_{t-1}, X_{t-2}, \dots).$$

Il n'est généralement pas nécessaire de prendre en compte tout le passé de la série et on peut le plus souvent se limiter à p valeurs :

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-p}) + \varepsilon_t. \quad (2.1)$$

Où ε_t est un bruit blanc[7].

2.1.1 Modèle AR (Autorégressif)

i) **Formulation** : De façon générale, un processus $AR(p)$ est un processus qui dépend linéairement des p valeurs antérieures :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t. \quad (2.2)$$

Où ε_t est un bruit blanc.

On peut, sans restriction de généralité, supprimer le terme constant et obtenir un modèle de la forme :

$$X_t = \sum_{1 \leq h \leq p} \phi_h X_{t-h} + \varepsilon_t \Leftrightarrow \Phi(B)X_t = \varepsilon_t \quad (2.3)$$

Où Φ est le polynôme de degré p dont les coefficients sont : $1, \phi_1, \dots, \phi_p$.

ii) **Auto-covariance** :

$$\gamma_h = \sum_{i=1}^p \phi_i \gamma_{h-i}. \quad (2.4)$$

iii) **Auto-corrélation** : On montre que les auto-corrélations sont des solutions de l'équation Yule et Walker suivante :

$$\rho_h = \sum_{i=1}^p \phi_i \rho_{h-i}. \quad (2.5)$$

iv) **Auto-corrélation partielle** : Dans un tel processus, X_t et X_{t-p+1} sont indépendants conditionnellement aux valeurs intermédiaires $(X_{t-1}, X_{t-2}, \dots, X_{t-p})$ et donc :

$$\tau(h) = \begin{cases} 0, & \text{si } |h| > p; \\ \phi_p, & \text{si } |h| = p. \end{cases}$$

La valeur à la date t dépend des p dates précédentes et pas des autres. Cette propriété sert à l'identification des modèles et à déterminer l'ordre p d'un processus $AR(p)$ au vu du corrélogramme partiel[7].

2.1.2 Modèle MA (Moving Average : Moyenne Mobile)

i) **Formulation** : Le terme ε_t est souvent comme un "choc" (une innovation pour les économètres) : il rend compte d'un élément nouveau dans l'évolution du processus. On peut envisager que ces chocs (non-corrélés et d'espérances nulles) aient des effets sur l'évolution du processus non seulement à la date à laquelle ils se produisent mais aussi à des dates ultérieures.

Un processus $(X_t, t \in T)$ est dit Moyenne Mobile s'il vérifie l'équation stochastique suivante :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} = \Theta(B)\varepsilon_t. \quad (2.6)$$

Où Θ est le polynôme de degré q dont les coefficients sont $1, \theta_1, \dots, \theta_q$, dans un tel modèle, on suppose que l'influence des chocs passés se manifeste au travers d'une fonction linéaire.

Un tel modèle est appelé moyenne mobile d'ordre q car X_t est une moyenne mobile (Moving Average) appliquée aux variables aléatoires $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$. Le terme moyenne est à prendre dans un sens très large dans la mesure où la somme des coefficients n'est pas nécessairement égale à 1.[7]

ii) **Auto-covariance :**

$$\gamma(h) = \begin{cases} (\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h})\sigma^2, & \text{si } 0 < |h| \leq q; \\ 0, & \text{si } |h| > q. \end{cases}$$

iii) **Auto-corrélation :**

$$\rho(h) = \begin{cases} \frac{-\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h}}{1 + \theta_1^2 + \dots + \theta_q^2}, & \text{si } 0 < |h| \leq q; \\ 0, & \text{si } |h| > q. \end{cases}$$

iv) **Auto-corrélation partielle :** L'expression de la fonction d'auto-corrélation partielle est compliquée. Notons simplement qu'elle ne s'annule pas à partir d'un certain rang mais qu'il existe un nombre $r \in]0, 1[$ tel que :

$$\tau(h) \leq r^h, h \geq 2.$$

2.1.3 Modèle *ARMA* (Combinaison de processus *AR* et *MA*)

i) **Formulation :** On peut bien évidemment envisager de combiner les deux modèles précédents en introduisant :

- une dépendance du processus vis-à-vis de son passé : modèle *AR*(p) ;
- un effet retardé des chocs : modèle *MA*(q).

Un tel modèle, appelé autoregressif - moyenne mobile (*ARMA*), est caractérisé par le paramètre p de la partie autoregressive et le paramètre q de la partie moyenne mobile.

Un processus (X_t) est dit *ARMA*(p, q) s'il vérifie l'équation stochastique suivante :

$$(1 - \phi_1 B^1 - \dots - \phi_p B^p)X_t = (1 - \theta_1 B^1 - \dots - \theta_q B^q)\varepsilon_t. \quad (2.7)$$

Soit :

$$\Phi(B)X_t = \Theta(B)\varepsilon_t. \quad (2.8)$$

Le traitement d'un tel processus est plus complexe que celui des deux précédents. On peut cependant démontrer que ces auto-corrélations et ces auto-corrélations partielles sont des fonctions amorties tendant vers 0 en valeur absolue à vitesses exponentielles.[7] On peut ainsi dresser un tableau comparatif des corrélogrammes et corrélogrammes partiels des processus :

TABLE 2.1 – Résumé des propriétés des modèles *ARMA*

Processus	Auto-corrélation ρ_h	Auto-corrélation partielle τ_h
AR(p)	amortie	nulle pour $h > q$
MA(q)	nulle pour $h > q$	amortie
ARMA(p,q)	amortie	amortie

L'analyse des corrélogrammes constitue un des outils privilégiés dans l'identification du modèle[7].

2.2 Processus *ARIMA* et *SARIMA*

La condition de stationnarité des modèles présentés ici n'est évidemment pas toujours convenable. On peut intégrer certains types de non stationnarités en élargissant le modèle *ARMA*, Ces élargissements consistent en une série d'opérations préalables visant à éliminer la tendance ou la saisonnalité.

1. **Modèle *ARIMA*** : On a vu que si un processus X_t admet une tendance polynomiale de degré d , le processus différencié d fois est stationnaire :

$$(1 - \phi_1 B^1 - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B^q - \dots - \theta_q B^q) \varepsilon_t. \quad (2.9)$$

$$\Phi(B) \Delta^d X_t = \Theta(B) \varepsilon_t. \quad (2.10)$$

Le modèle *ARIMA* revient à appliquer un modèle *ARMA* sur le processus différencié :

$$Y_t = ARMA(p, q) \Leftrightarrow X_t = ARIMA(p, d, q)$$

L'équation d'un modèle *ARIMA*(p, d, q) est donc donnée par :

$$\Phi(B) Y_t = \Theta(B) \varepsilon_t \Leftrightarrow \Phi(B) \Delta^d X_t = \Theta(B) \varepsilon_t. \quad (2.11)$$

Où Φ et Θ sont deux polynômes de degrés respectifs p et q . Le '*I*' de *ARIMA* signifie 'integrated' comme réciproque de la différenciation. Évidemment, le degré

d n'est généralement pas connu. Pour le déterminer on peut agir par tâtonnements ou avoir recours à des tests de stationnarité, puisqu'un processus $ARMA(p, q)$ est stationnaire, on cherche d tel qu'on puisse accepter l'hypothèse de stationnarité pour le processus $Y_T = \Delta^d X_t$.

De façon générale, on se réfère à un principe de parcimonie et on cherche la valeur satisfaisante minimale de d . Sa sera menée plus en détail dans la section 2 concernant le choix de modèles.

2. **Modèle SARIMA** : La saisonnalité est un autre facteur de non-stationnarité. On a vu qu'une façon simple d'éliminer une saisonnalité de période s consiste à appliquer l'opérateur Δ_ω tel que :

$$Z_t = \Delta_s X_t = (I - B^s)X_t = X_t - X_{t-s}.$$

De façon générale, on peut supposer que l'influence des chocs se transmet entre des dates distantes d'un nombre entier de périodes selon un processus $ARIMA(P, D, Q)$:

$$\Phi_s(B^s)\Delta_s^D X_t = \Theta_s(B^s)\mu_t. \quad (2.12)$$

et que ces chocs eux-mêmes suivent un modèle $ARIMA(p, d, q)$:

$$\Phi(B)\Delta^d X_t = \Theta(B)\varepsilon_t. \quad (2.13)$$

Où ε_t est un bruit blanc.

Un tel modèle est noté $SARIMA(p, d, q) \times (P, D, Q)$ et son équation générale est[?] :

$$\Phi(B)\Phi_s(B^s)\Delta^d \Delta_s^D X_t = \Theta(B)\Theta_s(B)\varepsilon_t. \quad (2.14)$$

Où Φ , Φ_s , Θ et Θ_s sont des polynômes de degrés respectifs p, P, q, Q .

2.3 Modélisation par la méthode de Box et Jenkins

Parmi les méthodes de modélisation des séries chronologiques les plus utilisées est l'approche Box-Jenkins.

George Box et Gwilym Jenkins sont les deux statisticiens qui ont contribué en 1970 à populariser la théorie des séries temporelles univariées. Les procédures de modélisation sont présentées dans leur célèbre ouvrage intitulé " Time Series Analysis : Forecasting and control ". Ils ont proposé une démarche générale de prévision pour les séries chronologiques [18].

En général, la méthode de Box et Jenkins est constituée de cinq étapes qui seront détaillées dans cette section, sera appuyées par un exemple d'application dans le chapitre 3. La méthode de Box et Jenkins est un outil systématique qui permet :

1. De déterminer le meilleur modèle de type *ARMA* décrivant le processus stochastiques d'une série observée ou d'une transformation stationnaire de ce dernier ;
2. D'estimer les paramètres du modèle ;
3. Prévision.

2.3.1 Les étapes de la méthode de Box et Jenkins :

Étape 1 : Transformation des données afin de stabiliser la variance (*log, sqrt, ...*) et différencier les données pour les stationnariser ;

Étape 2 : Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés ;

Étape 3 : Estimation des paramètres du (des) modèle(s) sélectionné(s) ;

Étape 4 : Diagnostic et tests d'adéquation du modèle ;

Étape 5 : Prévision, la dernière étape consiste à la prévision des valeurs futures à travers le modèle retenu.

À travers la représentation graphique de la série, on peut décider si la série en question est stationnaire ou non.

Dans la plus part des cas, une première différenciation rend la série stationnaire.

Il arrive parfois d'appliquer une transformation (log ou racine carré) à la série pour stabiliser la variance.

Étape 1 : différencier les données afin de stabiliser la variance (*log, sqrt, ...*) et différenciations des données pour les stationnariser

a) Stationnarité

Comme on l'a déjà défini dans la section précédente, une série X_t est considérée comme étant stationnaire lorsque ses propriétés statistiques (moyenne, variance) sont constantes dans le temps. Il existe plusieurs tests de stationnarité. Mentionnons entre autres le test de la racine unitaire. Il est possible de vérifier si la série est stationnaire ou pas en analysant le graphique des valeurs de la série en fonction du temps ou en étudiant le graphique de sa fonction d'auto-corrélation lui étant associé.

b) Transformations

Dans le cas où la série n'est pas stationnaire, il existe des moyens pour la rendre stationnaire. Elle se résume comme suit :

$$Y_t = \ln X_t.$$

Sur le même principe, on peut extraire la racine carrée des données initiales de la série :

$$Y_t = \sqrt{X_t}.$$

Ensuite, il est possible de stabiliser la moyenne lorsque la série présente une tendance à l'aide d'une ou de plusieurs différenciations avec l'opérateur Δ . Il est important d'appliquer la transformation logarithmique avant cette dernière si elle est nécessaire car le processus de différenciation implique des valeurs négatives qui ne sont pas compatibles avec le modèle logarithmique.

Soit Y_t le processus : $Y_t = \Delta X_t = X_t - X_{t-1}$.

Remarque 2.1. *En cas d'une série non stationnaire il existe deux cas principaux :*

- *Pour un processus TS, la bonne méthode de stationnarisation, est celle des moindres carrés ordinaires (MCO) ;*
- *Pour un processus DS, la bonne méthode de stationnarisation est celle de filtre aux différenciations premières [7].*

Étape 2 : Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés.

La méthode d'identification est essentiellement fondée sur l'analyse conjointe des

auto-corrélations et des auto-corrélations partielles. Nous supposons que le processus est stationnaire. La méthode s'appuie sur les résultats suivants :

* Pour un processus AR(p) minimal :

- les auto-corrélations sont à l'intérieur d'une enveloppe à décroissance géométrique ;
- les auto-corrélations partielles sont identiquement nulles au delà de l'ordre p .

* Pour un processus MA(q) minimal :

- les auto-corrélations sont identiquement nulles au delà de l'ordre q ;
- les auto-corrélations partielles sont à l'intérieur d'une enveloppe à décroissance géométrique.

La fonction d'auto-corrélation empirique vérifie les deux propriétés suivantes :

$$E(\hat{\rho}_h) = \frac{-1}{n}; \quad (2.15)$$

$$Var(\hat{\rho}_h) = \frac{1}{n}. \quad (2.16)$$

Où n est la taille de la série en question. $\hat{\rho}_h$ est asymptotiquement normalement distribuées. D'où pour tester :

$$\begin{cases} H0, & \hat{\rho}_h = 0; \\ H1, & \hat{\rho}_h \neq 0. \end{cases}$$

On compare la statistique $t_{\hat{\rho}_h} = \frac{\hat{\rho}_h}{\hat{\sigma}(\hat{\rho}_h)}$ à la valeur tabulée de la loi Normale.

Si $|t_{\hat{\rho}_h}| < Z_{1-\alpha/2} \Rightarrow$ on accepte H0.

C'est à dire $\hat{\rho}_h = 0$ ou $Z_{1-\alpha/2}$ est le quantile d'ordre $(1 - \alpha/2)$ de la loi Normale centrée réduite. Graphiquement les limites de l'intervalle de confiance des auto-corrélations seront représentées par $-1/n \pm Z_{1-\alpha/2}/\sqrt{n}$

De même pour la fonction d'auto-corrélation partielle empirique et asymptotiquement et normalement distribuée de variance $1/n$. D'où pour tester la nullité des auto-corrélations partielle, on compare la statistique $t_{\hat{\phi}_h} = \hat{\phi}_h \times \sqrt{n}$ à la valeur tabulée de la loi Normale.

Si $|t_{\hat{\phi}_h}| < Z_{1-\alpha/2}$ alors on accepte H0, c'est à dire : $\hat{\phi}_h = 0$.

Étape 3 : Estimation des paramètres du (des) modèle(s) sélectionné(s).

Il existe plusieurs méthodes d'estimation des paramètres (MCO, Yule-Walker,...),

la méthode la plus répandue est celle du Maximum de Vraisemblance (MV). Cette méthode repose sur l'hypothèse de Normalité des résidus $(\varepsilon) \sim N(0, \sigma^2)$. Le logarithme de vraisemblance d'un processus $ARMA(p, q)$ est donnée par ;

$$\ln B_n = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln(\det(\dot{Z}Z) - \frac{S(\phi, \theta)}{2\sigma_\varepsilon^2}). \quad (2.17)$$

Z est une matrice de taille $(p + q + n, p + q)$ qui dépend des paramètres : $\phi_i (i = 1, \dots, p)$ et $\theta_j (j = 1, \dots, q)$.

$$S(\phi, \theta) = \sum_{t=0}^n (E(\varepsilon_t))^2$$

En maximisant cette fonction, on déduit les estimateurs : $\hat{\phi}_i, \hat{\theta}_j$ et $\hat{\sigma}_\varepsilon^2$.

Remarque 2.2. Pour tester si on inclut le terme constant lors de l'estimation ou non.

$$\begin{cases} H0, & c = 0 \\ H1, & c \neq 0 \end{cases}$$

On procède comme suit :

Si $|t| = \sqrt{n} \left| \frac{\bar{X}}{\hat{\sigma}_\varepsilon} \right| < Z_{1-\alpha/2}$, on accepte $H0$.

Étape 4 : Diagnostic et tests adéquation du modèle (validation).

Souvent, il n'est pas facile de déterminer un modèle unique qui représente le processus générateur de données, et il est souvent estimer plusieurs modèles à l'étape initiale.

Le modèle qui est finalement choisi est celui considéré comme le meilleur basé sur un ensemble de critères de contrôle et de diagnostique.

Ces critères comprennent :

1. **Les tests de significativité :** Ce test, nous permet d'effectuer le test de Student sur chacun des paramètres de processus $ARMA$ en divisant le paramètre par son écart type. Il peut arriver qu'un ou plusieurs paramètres ne soit pas significativement différent de 0 : le modèle est alors rejeté et on retourne à l'étape d'estimation en éliminant la variable dont le coefficient n'est pas significatif. Il s'agit de tester :

$$\begin{cases} H0, & \hat{p} = p - 1 \text{ et } \hat{q} = q; \\ H1, & \hat{p} = p \text{ et } \hat{q} = q - 1. \end{cases} \quad \text{En d'autres termes, tester :}$$

$H0 : ARMA(p - 1, q)$ vs $H1 : ARMA(p, q - 1)$

Il s'agit d'un test de significativité sur les coefficients : ϕ_p et θ_q .

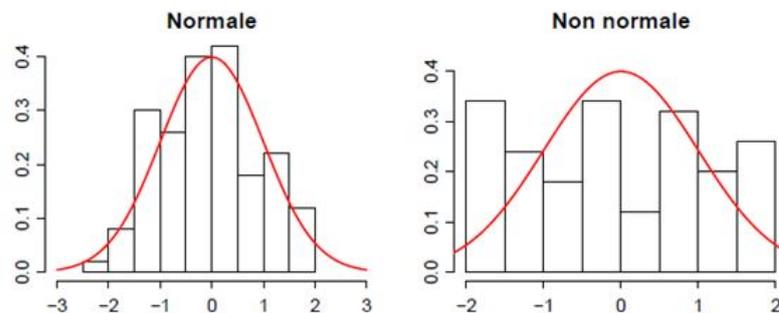
On compare les statistiques : $t_{\hat{\phi}_p} = \hat{\phi}_p / \hat{\sigma}(\hat{\phi}_p)$ et $t_{\hat{\theta}_q} = \hat{\theta}_q / \hat{\sigma}(\hat{\theta}_q)$ à la valeur tabulée $t_{1-\alpha/2}$.

Si la valeur de $|t_{\hat{\phi}_p}| < t_{1-\alpha/2}$ on accepte $H0$ et on retient $ARMA(p, q)$.

2. **Analyse des résidus** : Les résidus issus d'une estimation doivent vérifier quelques propriétés statistiques :

- **La normalité** : La normalité peut être testée graphiquement, soit en représentant l'histogramme des résidus, soit par le graphe quantile-quantile (qq-plot).

FIGURE 2.1 – Exemple de la normalité.



Source : [25]

- **tests d'auto-corrélation** : Il existe plusieurs tests d'absence d'auto-corrélation. Ces tests se regroupent en deux les tests paramétriques et non paramétrique.
 - Les tests paramétriques les plus connus dans la littérature étant ceux de Box et Pierce (1970) et de Ljung et Box (1978) ;
 - Les tests non paramétriques les plus répandus sont : le test de retournement et le test de monotonie. Mais dans ce travail on va s'intéresser aux tests paramétriques.

- **Les tests paramétriques : Le test de Box et Piere** : ce test est connu sous le nom du test de porte manteau, il s'agit de tester :

$$H_0 : \hat{\rho}_1(\hat{\varepsilon}_t) = \hat{\rho}_2(\hat{\varepsilon}_t) = \dots = \hat{\rho}_h(\hat{\varepsilon}_t) = 0$$

La statistique utilisée est :

$$BP(h) = n \sum_{h=1}^h \hat{\rho}_h^2(\hat{\varepsilon}_t)$$

Sous l'hypothèse nulle, $BP(h) \sim \chi^2(h - q - p)$, si $BP(k) < \chi_{1-\alpha/2}^2(h - q - p)$, on accepte H_0 .

- 3. Les critères d'informations (choix du modèle)** : Les critères les plus fréquemment utilisés sont :

- * Le critère d'informations d'Akaike (1969) : $AIC = \ln(\hat{\sigma}_\varepsilon^2) + \frac{2(p+q)}{n}$;

- * Le critère d'informations de Schwarz (1978) : $SIC = \ln(\hat{\sigma}_\varepsilon^2) + (p+q) \frac{\ln n}{n}$;

- * Le critère d'informations de Hannan-Quinn (1979) :

$$SIC = \ln(\hat{\sigma}_\varepsilon^2) + \alpha(p+q) \frac{2(p+q)}{n}$$

ou $\alpha > 2$.

Un modèle choisi par critère d'informations est un modèle qui minimise l'un des critères d'Akaike, Schwarz ou Hannan-Quinn .

Étape 5 : Prédiction : La dernière étape consiste en la prédiction des valeurs futures à travers le modèle retenu.

Considérons un processus $ARMA(p, q)$ et on note \hat{X}_{t+h} la prédiction de X . Faite en t à l'horizon h .

$\hat{X}_{t+h} = E[X_{t+h}|T_t]$, $|_t$ est l'information disponible à l'instant t .

Formule déduite de la forme $ARMA$:

$$\hat{X}_{t+h} = \sum_{j=1}^{p+q} \phi_j X_{t+h-j} + \sum_{j=1}^q \hat{\varepsilon}_{t+h-j} \cdot [18] \quad (2.18)$$

Remarque 2.3. Le schéma (voir : A.2) nous résume toutes ces étapes, objet de cette section.

Conclusion

Après avoir présenté l'aspect théorique des processus à courte mémoire ainsi que la méthode de modélisation et de prévision classique de Box-Jenkins, on va entamer les processus les plus complexes qui peuvent être considérés comme un élargissement aux modèles à court terme.

3

Les processus à longue mémoire

”*ARFIMA*”

Introduction

Nous nous intéresserons ici uniquement à une catégorie particulière de processus à mémoire longue, précisément *ARFIMA* (Autorégressifs, Moyenne Mobile, Fractionnairement Intégrés). Par rapport aux autres modèles à mémoire longue, les processus *ARFIMA* présentent l’avantage d’être relativement facilement applicables en économie dans la mesure où ils constituent une extension directe des processus *ARIMA* usuels.

Dans un premier temps, nous nous attacherons à la représentation de la mémoire longue (*ARFIMA*), à savoir, sa définition et ses caractéristiques avant de détailler de la manière la plus exhaustive possible les diverses procédures de modélisation et prévision.

La famille des modèles $ARIMA(p, d, q)$ est largement utilisée dans tous les domaines d'application de l'analyse des séries chronologiques, économie, etc et peuvent modéliser une classe assez large des séries.

Les processus ARMA étudiés jusqu'à présent sont des processus à mémoire courte dans le sens où l'effet d'un choc à un instant donné n'est pas durable et n'affecte donc pas l'évolution future de la chronique.

Les processus à mémoire infinie comme les processus DS ont un comportement opposé, l'effet d'un choc est permanent et se répercute sur l'ensemble des valeurs futures de la série temporelle. Cette dichotomie s'avère insuffisante pour rendre compte de phénomènes à long terme[7].

Cependant, les modèles à mémoire courte sont avérés insuffisants dans le domaine de la modélisation, ce qui a fait apparaître une classe de processus plus large, sont les processus à mémoire longue. Dans ce mémoire on se limitera à l'étude des modèles "ARFIMA".

3.1 Les processus à mémoire longue

Définition 3.1. Les processus à mémoire longue peuvent être définis de façon équivalente dans le domaine spectral et le domaine temporel.

1. Dans le domaine fréquentiel(spectral) :

Les processus à mémoire longue sont caractérisés par une densité spectrale qui croît sans limite quand la fréquence tend vers zéro.

Un processus stationnaire $\{X_t\}_{t \in \mathbb{Z}}$ est un Processus à mémoire longue s'il existe un nombre réel $\beta, 0 < \beta < 1$ et une constante $c > 0$ vérifiant :

$$\lim_{\lambda \rightarrow 0} \frac{f(\lambda)}{c|\lambda|^{-\beta}} = 1. \quad (3.1)$$

$f(\lambda)$: la densité spectrale du processus $(X_t, t \in \mathbb{Z})$ à la fréquence λ On en déduit immédiatement que :

$f(\lambda) \sim c|\lambda|^{-\beta}$, quand $\lambda \rightarrow 0$. Ainsi la densité spectrale exhibe un pôle à la fréquence zéro, contrairement à la densité spectrale des processus à mémoire courte qui est finie et positive aux basses fréquences.

2. Dans le domaine temporel :

Les processus à mémoire longue sont caractérisés par une fonction d'auto-corrélation décroissante hyperboliquement au fur et à mesure que le retard s'accroît, à l'encontre des processus à mémoire courte où elle décroît exponentiellement.

3.2 Processus $ARFIMA(p, d, q)$

Les processus $ARFIMA(p, d, q)$ s'écrivent :

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B)\varepsilon_t. \quad (3.2)$$

Où (ε_t) est indépendamment et identiquement distribuées, $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$.

Où encore :

$$(1 - B)^d X_t = \mu_t. \quad (3.3)$$

Où μ_t le processus $ARMA(p, q)$:

$$\phi_p(B)\mu_t = \theta_q(B)\varepsilon_t$$

Il est stationnaire si :

* $d < 1/2$, les racines $\phi_p(B)$ sont à l'extérieur du cercle unite du plan complexe,

Il est inversible si :

* $d > -1/2$, les racines $\theta_p(B)$ sont à l'extérieur du cercle unite du plan complexe,

La fonction d'auto-corrélation des $ARFIMA$ possède le comportement de décroissance hyperbolique mentionné pour les processus $FI(d)$.

On peut vérifier qu'au-delà d'un certain retard l'influence de la mémoire courte (due à la partie $ARMA$) s'estompe pour ne laisser subsister que l'influence de la partie intégrée fractionnaire.

Le spectre du processus $ARFIMA(p, d, q)$ est : $f_x(\omega) = [2 \sin(\frac{\omega}{2})]^{-2d} f_\mu(\omega)$ avec $\omega \in [0, \pi]$ ou $f_\mu(\omega)$ est la densité spectrale d'une composante μ_t de mémoire courte comme un processus ARMA.

3.3 Processus long mémoire pure ($FI(d)$ ou $ARFIMA(0, d, 0)$)

$FI(d)$ est le processus qui contient les composantes de long terme, la partie $ARIMA$ rassemblant les composants de court terme (voir section 2)[3]. Le processus $FARIMA$ le plus simple est le bruit fractionnaire pure, ou $ARFIMA(0, d, 0)$:

$$(1 - B)^d X_t = \varepsilon_t \quad (3.4)$$

Où ε_t est une suite de variables aléatoires indépendantes et identiquement distribuées ;

B opérateur retard, $B^j X_t = X_{t-j}$;

d est un paramètre inconnu (d n'est pas forcément entier). Considérons l'opérateur $A(B) = (1 - B)^d$, et en utilisant la fonction $\Gamma(\cdot)$ pour $d \in R$, on obtient :

$$\pi_j = \frac{d(d-1)\dots(d-j-1)}{j!} = \frac{\Gamma(j-d)}{\Gamma(-d)j!}$$

La représentation autoregressive du modèle donné par l'équation 3.4, s'écrit :

$$\varepsilon_t = \sum_{j \geq 0} \pi_j X_t$$

Considérons maintenant l'inverse de l'équation de l'opérateur $A(B)$, pour exprimer (X_t) en fonction de ε_t , la représentation moyenne mobile du modèle est donnée par :

$$X_t = \sum_{j \geq 0} \pi_j \varepsilon_{t-j}, \quad \pi_j = \frac{\Gamma(j+d)}{\Gamma(d)j!}$$

3.3.1 Causalité

Pour caractériser la causalité du modèle considéré, on exprime les innovations en terme des valeurs présentes et passées du processus.

On définit l'opérateur fractionnaire $A(B) = (1 - B)^d$. Explicitant l'opérateur fractionnaire $A(B)$ comme suit :

$$(1 - B)^d = \sum_{j=0}^{\infty} \frac{d(d-1)\dots(d-j-1)}{j!} (-1)^j B^j. \quad (3.5)$$

En utilisant la fonction $\Gamma(\cdot)$ pour $d \in R$, on obtient :

$$\frac{d(d-1)\dots(d-j-1)}{j!}(-1)^j = \frac{(d-j-1)\dots(1-d)(-d)}{j!}(-1)^j B^j = \frac{\Gamma(j-d)}{\Gamma(-d)j!}. \quad (3.6)$$

On pose :

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(-d)j!}. \quad (3.7)$$

Les coefficients de l'opérateur $A(B)$.

La présentation moyenne mobile du modèle s'écrit sous la forme suivante :

$$\varepsilon_t = \sum_{j \geq 0} \pi_j X_j$$

Proposition 3.1. *Le processus stochastique vérifiant l'équation :*

$$(1-B)^d X_t = \varepsilon_t \quad (3.8)$$

existe et est causale si seulement si : ($d < 1/2$).

3.3.2 Inversibilité

Le concept de l'inversibilité est fondamentale, il permet d'une part d'estimer les paramètres d'un modèle de séries chronologiques et d'autre part de faire des prévisions.

Le processus $(X_t, t \in Z)$ s'exprime en fonction des valeurs passées et présentes des innovations $(\varepsilon_t, t \in Z)$. Explicitant maintenant l'inverse de l'opérateur $A(B)$:

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

On note par :

$$\Psi_j = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)j!}.$$

Les coefficients de l'opérateur inverse.

La représentation autoregressive du modèle donné par ?? s'écrit sous la forme :

$$X_t = \sum_{j \geq 0} \Psi_j \varepsilon_{t-j}.$$

Proposition 3.2. *Le modèle 3.8 est inversible si et seulement si $d > 1/2$.*

Proposition 3.3. *Le modèle 3.8 est causale et inversible si et seulement si $-1/2 < d < 1/2$.*

3.3.3 Fonction d'auto-covariance

Soit le modèle donnée par l'équation stochastique 3.4 et en utilisant la formule de Stirling pour $h \rightarrow \infty$, on obtient :

$$\gamma_X(h) \simeq \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} (h)^{2d-1}. \quad (3.9)$$

3.3.4 Fonction d'auto-corrélation

La fonction d'auto-corrélation du processus (X_t) vérifiant l'équation stochastique 3.9 est donnée par :

$$\rho_h = \frac{\Gamma(-d+1)\Gamma(d)}{\Gamma(d)\Gamma(h-d+1)}.$$

Si $h = 0, \pm 1, 2, \dots$

On a l'approximation pour h grand :

$$\rho_h = \frac{\Gamma(-d+1)}{\Gamma(d)} h^{2d-1}.$$

Si $h \rightarrow \infty$, et $h^{2d-1}\rho_h = \frac{\Gamma(-d+1)}{\Gamma(d)} = \frac{\pi}{\sin(\pi d)}$.

Remarque 3.1. *Les corrélations partielles du processus $(X_t, t \in Z)$ vérifiant l'équation 3.9 sont donnée par :*

$$\phi_h = \frac{d}{h-d}, h = 1, 2, \dots$$

3.3.5 Propriétés

Les principales propriétés du processus le plus simple ($ARFIMA(0, d, 0)$) sont données par :

– Lorsque $d < 1/2$

Le processus est stationnaire et possède une représentation moyenne mobile infinie :

$$x_t = (1 - B)^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B) \varepsilon_t.$$

Avec $\psi_j = \frac{\Gamma(d+j)}{\Gamma(d)\Gamma(j+1)}$ ou la fonction $\Gamma(h)$ est telle que :

$$\Gamma(h) = (h - 1)!$$

– Lorsque $d > -1/2$

le processus est inversible et possède la représentation auto-régressive infinie :

$$(1 - B)^d x_t = \pi(B) x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = \varepsilon_t. \text{ Avec } \pi_j = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}.$$

– Lorsque $d \geq 1/2$

Le processus est stationnaire. Les valeurs asymptotiques des coefficients ψ_j et π_j : $\lim_{j \rightarrow \infty} \psi_j \approx \frac{j^{d-1}}{\Gamma(d)}$ et $\lim_{j \rightarrow \infty} \pi_j \approx \frac{j^{-d-1}}{\pi(-d)}$ décroissent à une vitesse hyperbolique qui est plus faible que la vitesse exponentielle des processus ARMA. La FAC possède ce même type de comportement ce qui permet de caractériser les processus $FI(d)$.

En définitive si :

- * $0 < d < 1/2$, le processus $FI(d)$ est un processus à mémoire longue ;
- * $d < 0$, le processus $FI(d)$ est un processus anti-persistant ;
- * $-1/2 < d < 0$, le processus $FI(d)$ est anti-persistant.

Le dernier cas ou $-1/2 < d < 0$ n'a pas le comportement des ARMA. Ce cas intermédiaire appelé anti-persistence par Mandelbrot correspond à des alternances de hausses et de baisses dans le processus.

Ce comportement est aussi appelé « effet Joseph » par référence à la bible.

3.4 Estimation du paramètre d

Il existe diverses techniques d'estimation des paramètres d'un processus *ARFIMA*. Les méthodes d'estimation peuvent être regroupées en deux catégories selon que l'on

estime ou non simultanément tous les paramètres de la représentation $ARFIMA(p, d, q)$. On distingue ainsi les méthodes en deux étapes, qui sont les plus anciennes, et les méthodes en une étape.

a) Les méthodes en deux étapes

Les méthodes en deux étapes peuvent être réunies en trois techniques principales dont la caractéristique commune est d'estimer dans une première étape le paramètre d'intégration fractionnaire d puis, dans une seconde étape, d'estimer par les méthodes usuelles de séries temporelles les paramètres autorégressifs et moyenne mobile de la représentation $ARMA(p, q)$ de la série transformée.

Les trois techniques (Méthode basée sur l'exposant de Hurst, les méthodes spectrales, la méthode de Janacek) se distinguent uniquement par la méthode de calcul du paramètre de différenciation fractionnaire d [?]

b) Les méthodes en une étape

Parmi les méthodes d'estimation en une étape, on peut distinguer celles basées sur une approximation de la méthode du maximum de vraisemblance et celles fondées sur le maximum de vraisemblance exacte. Une maximisation directe de la log-vraisemblance a été utilisée par McLeod et Hipel (1978) pour estimer les paramètres d'un bruit gaussien fractionnaire et peut également être employée pour des processus $ARFIMA(p, d, q)$.

Hosking (1984) souligne cependant que cette méthode est très coûteuse du point de vue, du temps de calcul. Dans ces circonstances, il semble dans un premier temps naturel de chercher à remplacer la vraisemblance par une approximation qui pourra être plus rapidement évaluée et de chercher à maximiser cette vraisemblance approximée.

Concernant les méthodes du maximum de vraisemblance approché, nous distinguerons la procédure de Li et Me Leod (1986), où l'approximation provient du fait qu'ils tronquent la somme infinie définissant le développement binomial et les procédures de Brockwell et Davis (1991) ou Fox et Taqqu (1986) utilisant une approximation de la fonction de log-vraisemblance fournie par la fonction de

Whittle. La méthode du maximum de vraisemblance exacte repose quant à elle sur la procédure élaborée par Sowell (1992)[?].

Remarque 3.2. *Nous présentons ici la méthode la plus utilisée "la méthode de Geweke Porter-Hudak".*

c) Estimation par la méthode de Geweke Porter-Hudak (1983)

Cette méthode est l'une des méthodes d'estimation semi paramétrique du paramètre d'intégration fractionnaire d'un processus $ARFIMA(p, d, q)$.

D'une manière générale, on considère un processus scalaire, (X_t) , faiblement stationnaire, dont la densité spectrale est de la forme suivante sur l'intervalle $[0, 2\pi[$:

$$f(\lambda) = |2 \sin(\frac{\lambda}{2})|^{-2d} f^*(\lambda)$$

ou d est le paramètre de mémoire compris dans l'intervalle $[-1/2, 1/2[$ et f^* est une fonction continue bornée sur tout l'intervalle $[0, 2\pi[$. le paramètre d contrôle le comportement de la densité spectrale dans un voisinage de zéro alors que f^* contrôle le comportement de courte mémoire.

Le périodogramme associé à un échantillon de taille T est défini par :

$$I_T(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t \exp^{-i\lambda t} \right|^2. \quad (3.10)$$

On définit les fréquences de Fourier par : $\lambda_j = 2\pi j/T$ pour $j = 1, \dots, \frac{T-1}{2}$.

A partir de la densité spectrale et du périodogramme pris aux fréquences de Fourier, on obtient la régression suivante :

$$\log I_T(\lambda_j) = \alpha + \beta Z_j + \varepsilon_j.$$

Où ξ est la constante d'Euler.

$$\begin{cases} \alpha = \log f^*(0) - \xi; \\ \beta = d; \\ Z_j = -2 \log |2 \sin(\lambda_j/2)|; \\ \varepsilon_j = I_T(\lambda_j) - \log f^*(\lambda_j + \xi). \end{cases}$$

Pour $0 \leq s < t < T$, on utilise l'estimateur :

$$\hat{d}_R(s, t) = \frac{\sum_{j=s+1}^t (Z_j - \bar{Z}) - \log I_T(\lambda_j)}{\sum_{j=s+1}^t (Z_j - \bar{Z})^2}. \quad (3.11)$$

Où Z est la moyenne des Z_j .

Dans cette section on a essayé, de façon équivalente à la première section, d'étudier les grands axes relatifs aux processus à mémoire longue de modèle "ARFIMA", suivi de la méthode la plus fréquente à cette fin. Cependant, nous pensons avoir touché les aspects principaux dans le but de pouvoir l'appliquer dans le prochain chapitre.

3.5 Tests de mémoire longue

L'exposant de Hurst et processus ARFIMA lors de son analyse sur les crues du Nil, Hurst (1951) a introduit une statistique permettant de détecter la présence de phénomènes de mémoire longue.

Cette statistique, appelée analyse R/S , est particulièrement intéressante dans la mesure où elle donne lieu à un coefficient, appelé exposant de Hurst, permettant de classer les séries temporelles en fonction de la nature de leur mémoire[7]

a) Analyse R/S et exposant de Hurst : La statistique R/S se définit comme l'étendue (R) des sommes partielles des écarts d'une série temporelle à sa moyenne divisée par son écart type (s_T).

Ainsi, soit une série temporelle $X_t, t = 1, \dots, T$, de moyenne \bar{X}_T , la statistique R/S , notée ici Q_T , s'écrit :

$$Q_t = R/s_T = \frac{[\max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}_T)]}{[\frac{1}{T} \sum_{j=1}^T (X_j - \bar{X}_T)^2]^{1/2}} \quad (3.12)$$

Cette statistique est asymptotiquement proportionnelle à T^H (voir Hurst, 1951), où la constante $H, 0 < H < 1$ est appelée exposant de Hurst. L'exposant de Hurst est ainsi donné par :

$$H \cong \frac{\log Q_T}{\log T} \quad (3.13)$$

L'exposant de Hurst est particulièrement intéressant dans la mesure où sa valeur permet de classer les séries temporelles en fonction de leur structure de dépendance. Le paragraphe suivant propose d'établir une telle classification en mettant en avant la relation existante entre exposant de Hurst et paramètre d'intégration fractionnaire des processus *ARFIMA*.

Conclusion

A travers ce chapitre nous avons présenté les différents aspects théoriques du processus objet de notre mémoire ainsi que la méthode ou technique de modélisation la plus fréquente destinée à mesurer les grandeurs statistiques, et ce, sans s'étaler dans les détails qui ont fait objet de divers ouvrages qui peuvent être consultés en cas de besoin.

A cet effet, et afin de mettre en œuvre cet aspect, nous consacrons le prochain et dernier chapitre à l'illustration et l'appui de l'aspect théorique par application.

4

Application : modélisation et prévision

Introduction

Dans ce chapitre, nous allons établir deux applications sur des données réelles, la première sur "les exportations Algériennes", la deuxième sur "l'indice de clôture quotidien du marché boursier koweïtien", en utilisant les deux méthodes de modélisation présentées dans les chapitres 2 et 3.

Nous allons modéliser cette série en utilisant de manière compétitive deux approches différentes, La première approche est basée sur une modélisation à mémoire courte de type "ARIMA", et la seconde approche est basée sur une modélisation à mémoire longue de type "ARFIMA", de plus, pour chacune de ces deux approches, nous étudierons l'impact en prévision des différents modèles estimés.

4.1 Première approche "Méthode Box Jenkins"

Dans cette première approche, nous utilisons la méthode classique de modélisation et de prévision de Box et Jenkins (1970). Les séries économiques sont rarement stationnaires. Elles peuvent présenter une tendance et des variations saisonnières. L'amplitude des variations de la série peut aussi dépendre de son niveau. Il est cependant souvent possible de stationnariser la série à l'aide d'une transformation permettant d'éliminer la tendance et la saisonnalité, et de stabiliser la variance.

L'une des conditions qui semble être nécessaire pour avoir de meilleurs résultats est de travailler sur une série chronologique assez longue ($n \geq 50$) dans l'objectif de Prévoir la valeur des exportations (à horizon de 03 mois).

Présentation des données :

Dans cette section, nous allons étudier la série des "exportations Algérienne" : La série relevée a partir de Janvier 1998 à Novembre 2017, soit 227 **Observations**, ces quantités sont présentées en Dinar Algérien. L'application est à l'aide du logiciel "EViews 9.1". On note par X la série des exportations Algérienne transformée comme suit :

$$Y_t = \log X_t$$

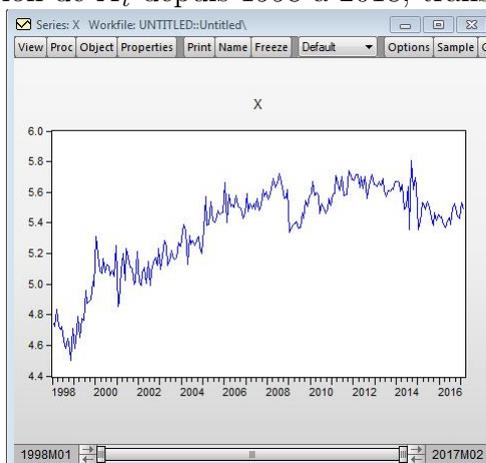
Nous avons transformé la variable en logarithme afin de pouvoir couvrir l'effet de la non-stationnarité. Ainsi, nous utiliserons les séries logarithmiques qui offrent les avantages suivant :

- * Minimisation de l'influence de la variable du temps sur les séries;
- * Minimiser les étapes dans le processus de stationnarisation.

4.1.1 Étape 1 : Transformation des données afin de stabiliser la variance ($\log, \sqrt{}, \dots$) et différenciations des données pour les stationnariser

A partir des données sur les exportations, on a pu établir le graphe (voir figure :4.1), nous utiliserons les séries logarithmiques nous allons étudier la série X comme suit :

FIGURE 4.1 – Évolution de X_t depuis 1998 à 2018, transformée en logarithme.



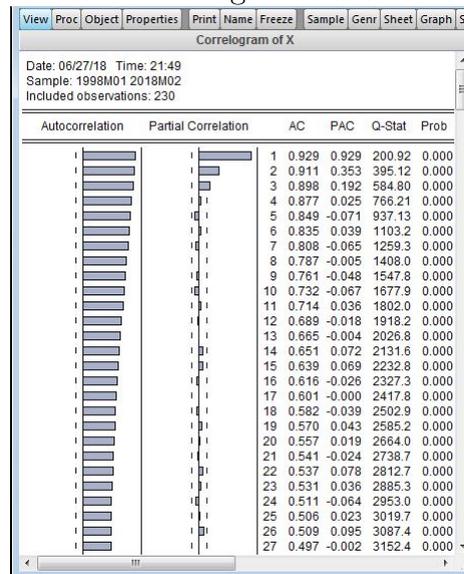
Le graphe de la série laisse apparaître une tendance qui peut être déterministe ou stochastique sur toute la durée.

A partir du corrélogramme d'auto-corrélation (voir figure :4.4) sur 36 retards, on peut dire que le corrélogramme d'auto-corrélation simple diminue lentement (converge) jusqu'à presque dire une stabilité, on remarque aussi trois pics au niveau du corrélogramme d'auto-corrélation partielle ($h = 1, h = 2, h = 3$ ou h : représente la périodicité des données mensuelles) d'où on soupçonne que notre série est générée par un modèle AR.

Ce qui nous amène à constater que la série X des exportations est non stationnaire, qui est vérifié par des tests statistiques (les tests ADF plus tard).

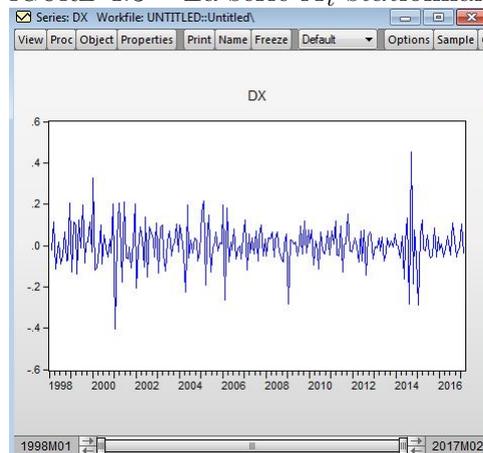
A l'aide des trois modèles (voir : chapitre 2) qui vont nous permettre de situer notre processus s'il est un TS ou DS, dans le but de choisir la bonne méthode pour stationnariser la série.

On confirme l'absence de la tendance et l'existence de la racine unitaire, donc le processus

FIGURE 4.2 – Corrélogrammes de la série X_t .

est non stationnaire de type DS.

Pour stationnariser la série, nous allons utiliser la méthode de stationnarisation des différences premières.

FIGURE 4.3 – La série X_t stationnaire.

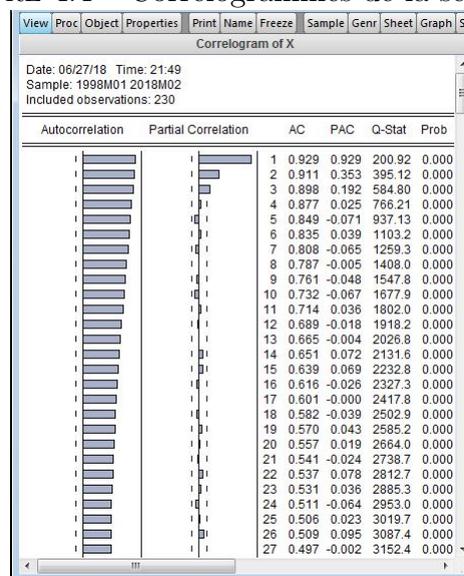
A partir du graphe (voir figure : 4.3) on remarque que la série est stationnaire autour de sa moyenne à partir de la première différenciation, ainsi d'après le correlogramme la fonction d'auto-correlation converge lentement vers 0, mais pour mieux confirmer la stationnarité de la série on a appliqué les tests de Dickey fuller .

A cet effet, notre série est devenue stationnaire.

4.1.2 Étape 2 : Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés.

On cherchera maintenant le processus qui génère la série.

FIGURE 4.4 – Corrélogrammes de la série X_t .



On utilise le corrélogramme (voir figure : 4.5).

4.1.3 Étape 3 : Estimation des paramètres du (des) modèle(s) sélectionné(s).

On va estimer les modèles (voir figure : 4.5) un par un et on doit choisir le modèle qui minimise les critères d'Akaike et schwarz.

Pour valider un modèle il faut que toutes les probabilités critique soient inférieur à 5% (voir tableau 4.1), et les résidus forment un bruit blanc et distribuer avec une loi normale. D'après l'analyse des critères Akaike et schwarz, nous avons choisi les valeurs qui minimisent ces dernières, d'où le modèle $ARIMA(2, 1, 0)$. D'après le tableau 4.1 notre

FIGURE 4.5 – Identification des paramètres.

Model Selection Criteria Table				
Dependent Variable: DX				
Date: 06/10/18 Time: 18:32				
Sample: 1998M01 2018M02				
Included observations: 229				
Model	LogL	AIC*	BIC	HQ
(0,1)(0,0)	233.476024	-1.904761	-1.861509	-1.887337
(2,3)(0,0)	237.436769	-1.904436	-1.803516	-1.863782
(0,2)(0,0)	233.718599	-1.898501	-1.840832	-1.875270
(1,1)(0,0)	233.684910	-1.898222	-1.840554	-1.874991
(3,0)(0,0)	234.089477	-1.893301	-1.821216	-1.864263
(2,1)(0,0)	233.985160	-1.892439	-1.820354	-1.863401
(2,0)(0,0)	232.856510	-1.891376	-1.833708	-1.868145
(0,3)(0,0)	233.847130	-1.891299	-1.819213	-1.862260
(1,2)(0,0)	233.760295	-1.890581	-1.818495	-1.861542
(3,1)(0,0)	234.599229	-1.889250	-1.802747	-1.854403
(4,0)(0,0)	234.165190	-1.885663	-1.799160	-1.850816
(0,4)(0,0)	234.056092	-1.884761	-1.798259	-1.849915
(2,2)(0,0)	233.985164	-1.884175	-1.797672	-1.849329
(1,3)(0,0)	233.880573	-1.883311	-1.796808	-1.848464
(3,2)(0,0)	234.613905	-1.881107	-1.780187	-1.840453
(4,1)(0,0)	234.605731	-1.881039	-1.780119	-1.840385
(3,3)(0,0)	235.584977	-1.880868	-1.765531	-1.834406
(4,2)(0,0)	235.457894	-1.879817	-1.764481	-1.833356
(2,4)(0,0)	235.445057	-1.879711	-1.764374	-1.833249
(1,4)(0,0)	234.059620	-1.876526	-1.775506	-1.835872
(4,3)(0,0)	235.629175	-1.872968	-1.743215	-1.820699
(3,4)(0,0)	235.600325	-1.872730	-1.742976	-1.820460
(4,4)(0,0)	236.122256	-1.868779	-1.724608	-1.810702
(1,0)(0,0)	224.835494	-1.833351	-1.790100	-1.815928
(0,0)(0,0)	199.053932	-1.628545	-1.599711	-1.616929

TABLE 4.1 – Modèle retenu.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.561607	0.050598	-11.09937	0.0000
AR(2)	-0.257154	0.066593	-3.861593	0.0001

modèle s'écrit comme suit :

$$\Phi(B)\nabla^d X_t = \Theta(B)\varepsilon_t$$

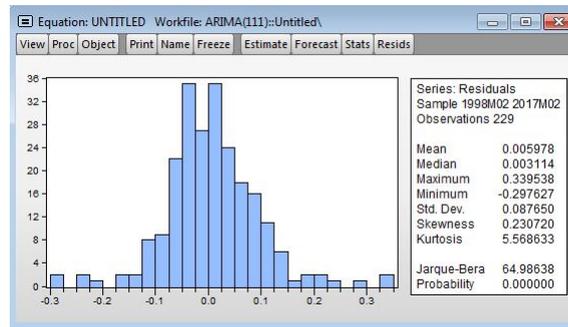
$$X_t = 0.561606685184X_{t-1} + 0.257153590091X_{t-2} + \varepsilon_t \quad (4.1)$$

4.1.4 Étape 4 : Diagnostic et tests adéquatation du modèle.

Après avoir estimé le modèle, on passe à l'étape de validation, l'auto-corrélogramme d'auto-corrélation des résidus forme un bruit blanc.

En analysant la représentation graphique, nous constatons que le graphe de la série estimée est presque semblable à celui de la série initiale, à quelque pics près.

A cet effet, les résidus forment un bruit blanc non gaussien.

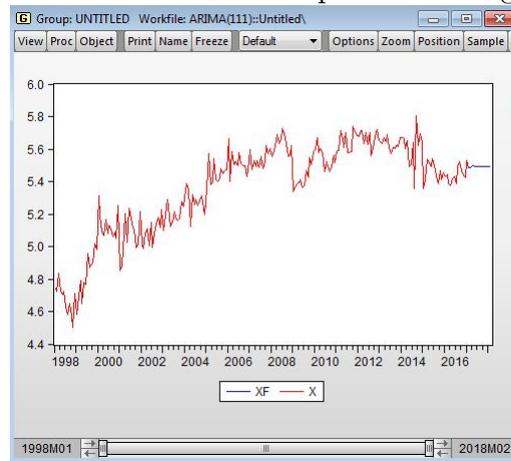
FIGURE 4.6 – Teste de normalité des résidus $ARIMA(2, 1, 0)$.

4.1.5 Étape 5 : Prédiction

Le modèle de la série sera utilisé pour calculer des prévisions à l'horizon $h=03$ mois qui suit à partir du mois de mars 2017.

On comparant nos résultats de prévision avec ceux de l'année précédente, on constate

FIGURE 4.7 – Prédiction des exportations Algériennes.



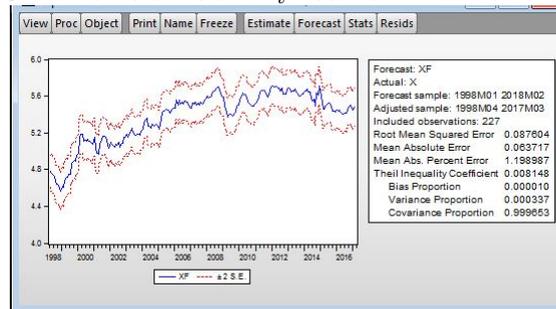
une légère progression dans les valeurs de prévision pour tous les mois, ce qui peut être interprété par la situation économique assez stable et stagné plus que de dire conforme au développement et à l'investissement.

A cet effet, d'après le tableau 4.2 nous remarquons que les prévisions sont bonnes.

Ce graphe nous prouve que ses prévisions sont fiables vu que toutes les valeurs appartiennent à l'intervalle de confiance et l'erreur quadratique moyenne de prévision $RMSE = 0.087604$ est petite.

TABLE 4.2 – Comparaison des prévisions avec les valeurs réelles.

Mois	Réel en DA	ARIMA en DA
12/2017	5,575203	5,493243
01/2018	5,6079214	5,493259
02/2018	5,64064	5,493243

FIGURE 4.8 – Prédiction de ” X_t ” sur un intervalle de confiance.

l’objectif de limiter notre traitement des données jusqu’à Septembre 2017, c’est de comparer les deux valeurs réelles et prédites. Où on constate que les valeurs sont assez proches des valeurs réelles, se qui prouve que notre série est solide et donne de meilleurs résultats à court terme.

D’après le tableau 4.2, on a obtenue de bonnes prévisions.

4.2 Deuxième approche”Méthode de Geweke Porter-Hudak (GPH)”

Comme on la constater dans la section 1 du chapitre 4, la modélisation de la série en question, ” les exportations Algériennes ”, avec les moyens des modèles à courts mémoire ”ARMA”, dans cette section on va procéder à la modélisation par le modèle ”ARFIMA” relatif aux processus à long mémoire.

Comme la série des exportations est anti persistante, ce qui nous prouve qu’elle n’est pas assez solide pour élaborer des prévisions à long terme. Nous avons opté pour une autre série plus adéquate qui nous permet d’étudier les processus ARFIMA.

Présentation des données :

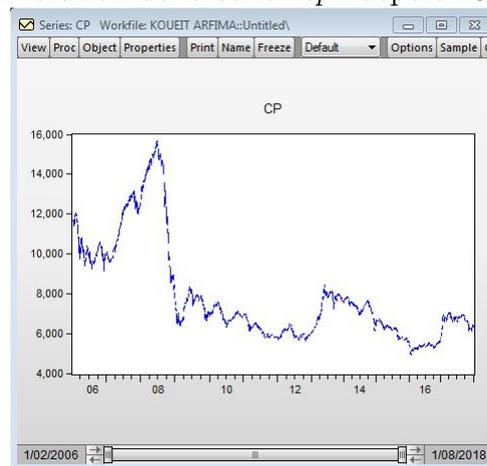
Dans cette section, nous allons étudier la série de "l'indice de clôture quotidien du marché boursier koweïtien" : la série irrégulière a partir de 2006 jusqu'à 2018, ce qui nous génère 3038 **Observations**, ces valeurs sont présentées en Dinar koweïtien.

On note par CP la série de "l'indice de clôture quotidien du marché boursier koweïtien".

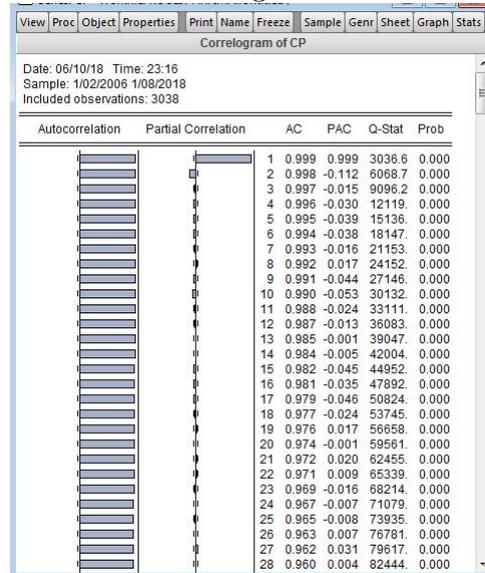
4.2.1 Étape 1 : Transformation des données afin de stabiliser la variance ($\log, \sqrt{}, \dots$) et différenciations des données pour les stationnariser

A partir des données sur l'indice boursier, on a pu établir le graphe 4.9, nous allons étudier la série cp comme suit :

FIGURE 4.9 – Évolution de la série " cp " depuis 2006 jusqu'à 2018.

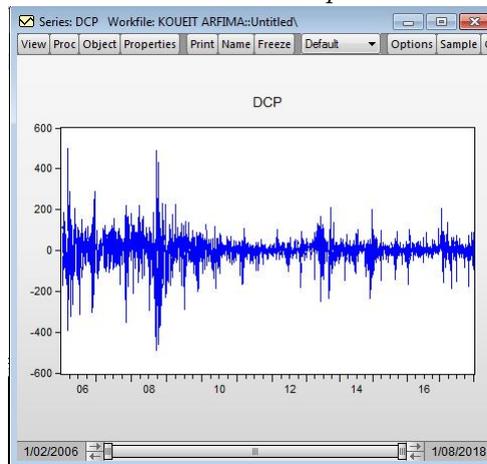


Le graphe de la série " cp " laisse apparaître une tendance qui peut être déterministe ou stochastique sur toute la durée ainsi que des fluctuations, des pics et des creux qui traduisent la saisonnalité.

FIGURE 4.10 – Corrélogrammes de la série cp .

Dans un premier temps, on suppose que les ordres p et q des polynômes autorégressif et moyenne mobile sont nuls.

On estime alors le paramètre de mémoire longue par la méthode GPH présentée précédemment, le processus $ARFIMA(0, d, 0)$ on obtiens $d = 0.5$, ce qui signifie que le processus $ARFIMA$ est un processus non stationnaire. Les auto-corrélation sont positives et constante lorsque le retard augmente.

FIGURE 4.11 – La série cp stationnaire.

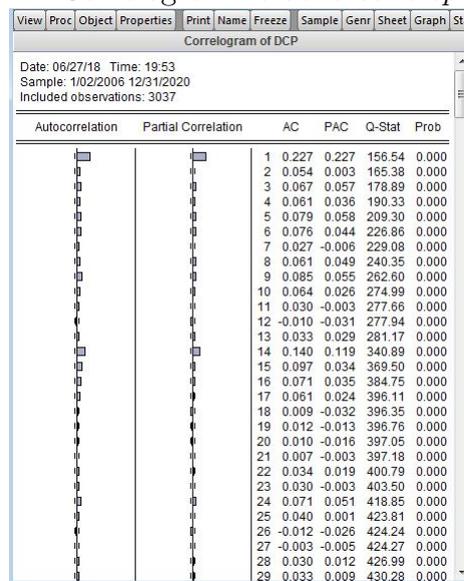
A partir du graphe (voir figure : 4.11) on remarque que la série est stationnaire autour de sa moyenne à partir de la première différenciation, ainsi d'après le corrélogramme la fonction d'auto-correlation converge lentement vers 0, mais pour mieux confirmer la stationnarité de la série on a appliqué les tests de Dickey fuller .

A cet effet, notre série est devenue stationnaire.

4.2.2 Étape 2 : Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés.

On cherchera maintenant le processus qui générer la série, tel que on utilise le corrélogramme pour estimer le modèle adéquat.

FIGURE 4.12 – Corrélogramme de la série cp stationnaire.



4.2.3 Étape 3 : Estimation des paramètres du (des) modèle(s) sélectionné(s).

On a estimé les modèles un par un et on a choisi le modèle qui minimise les critères d'Akaike et schwarz.

Pour valider un modèle il faut que toutes les probabilités soient inférieure à 5%, et les résidus forment un bruit blanc et distribuer avec une loi normale.

D’après l’analyse des critères Akaike et schwarz, nous avons choisi les valeurs qui minimisent ces dernières, d’où le modèle retenu est le $ARFIMA(10, 0.255857, 9)$, et $d \in]0, 1/2[$

Notre modèle s’écrit comme suit :

$$\Phi(B)\Delta^d X_t = \Theta(B)\varepsilon_t$$

Les coefficients sont identifiés comme on le constate sur le tableau(voir figure :4.3).

4.2.4 Étape 4 : Diagnostic et tests adéquation du modèle.

Après avoir estimé le modèle, on passe à l’étape de validation, l’auto-corrélogramme d’auto-corrélation des résidus forme un bruit blanc.

En analysant la représentation graphique, nous constatons que le graphe de la série estimée est presque semblable à celui de la série initiale, à quelque pics près.

A cet effet, les résidus forment un bruit blanc non gaussien.

FIGURE 4.13 – Teste de normalité des résidus $ARFIMA(10, 0.255857, 9)$.

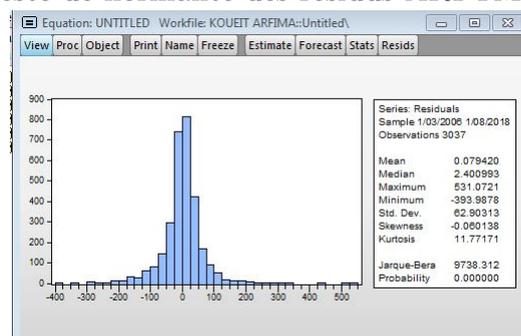


TABLE 4.3 – Identification des paramètres.

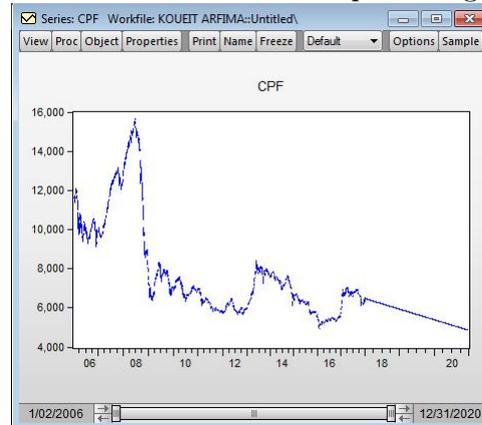
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1,491627	6,354691	-0,234729	0,8144
D	0,255857	0,051140	5,003086	0,0000
AR(1)	-0,511532	0,265461	-1,926957	0,0541
AR(2)	-0,022988	0,145207	-0,158313	0,8742
AR(3)	-0,686896	0,170313	-4,033152	0,0001
AR(4)	-0,738720	0,253744	-2,911281	0,0036
AR(5)	0,064344	0,225279	0,285618	0,7752
AR(6)	0,158595	0,153650	1,032187	0,3021
AR(7)	-0,468182	0,132205	-3,541343	0,0004
AR(8)	0,174385	0,200233	0,870912	0,3839
AR(9)	0,452880	0,134217	3,374239	0,0007
AR(10)	0,033894	0,030935	1,095645	0,2733
MA(1)	0,465263	0,282484	1,647044	0,0997
MA(2)	-0,119888	0,160400	-0,747430	0,4549
MA(3)	0,588120	0,152849	3,847730	0,0001
MA(4)	0,657207	0,247396	2,656498	0,0079
MA(5)	-0,202877	0,215368	-0,942002	0,3463
MA(6)	-0,260331	0,128991	-2,018214	0,0437
MA(7)	0,406294	0,104213	3,898672	0,0001
MA(8)	-0,208208	0,181080	-1,149813	0,2503
MA(9)	-0,440626	0,132744	-3,319360	0,0009

4.2.5 Étape 5 : Prédiction

Le modèle de la série sera utilisé pour calculer des prévisions à l'horizon $h \simeq 03$ ans qui suit à partir du mois de mars 2017.

Les résultats nous révèle une diminution des valeurs . Ses prévisions (voir figure :

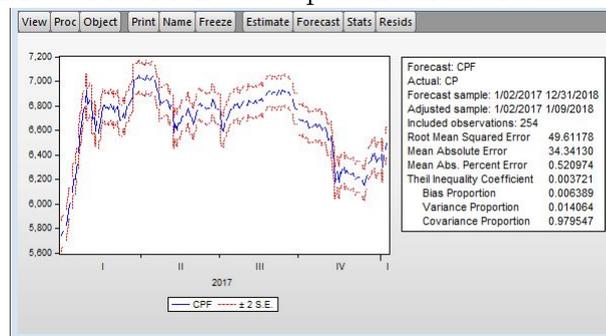
FIGURE 4.14 – Prédiction de "cp" à long terme.



4.14) sont dites en dehors des valeurs de la série, mais pour s'assurer de l'efficacité de ses prévisions on aura ce graphe, on a établie des prévisions de telle sorte que si ses prévisions sont comprises dans l'intervalle de confiance alors on confirme leurs efficacité :

A cet effet, d'après le tableau 4.4 nous remarquons que les prévisions sont bonnes.

FIGURE 4.15 – Prédiction des "cp" sur un intervalle de confiance.



Ce graphe (voir figure :4.15) nous prouve que ses prévisions sont fiables vue que toutes les valeurs appartiennent à l'intervalle de confiance et l'erreur quadratique moyenne de prévision $RMSE = 49.61178$ est petite.

Dans cette section, on a appliqué "la Méthode de Geweke Porter-Hudak (GPH)" avec toutes les étapes et les tests nécessaires , à l'aide du logiciel EViews, sur la série de "l'indice de clôture quotidien du marché boursier koweïtien", on a aboutis à un modèle adéquat "ARFIMA(10,0.255857,9)" qui représente le mieux notre série qui nous a

TABLE 4.4 – Comparaison des prévisions avec les valeurs réelles.

Jours	Réel en DA	<i>ARFIMA</i> en DA
10/12/2017	6162.35	6187.8117
11/12/2017	6155.6	6160.7081
12/12/2017	6197.58	6151.9035
13/12/2017	6237.87	6195.3024
14/12/2017	6331.72	6244.6623
17/12/2017	6355.9	6326.7777
18/12/2017	6370.24	6339.101
19/12/2017	6420.64	6356.9288
20/12/2017	6384.48	6403.6995
21/12/2017	6413.72	6376.7809
24/12/2017	6342.66	6413.8791
25/12/2017	6344.77	6337.8117
26/12/2017	6369.43	6356.7851
27/12/2017	6371.29	6373.8616
28/12/2017	6408.01	6368.6488
02/01/2018	6312.52	6404.4099
03/01/2018	6401.08	6300.7377
04/01/2018	6454.16	6403.5355
07/01/2018	6504.97	6447.8234
08/01/2018	6498.49	6494.4513
09/01/2018	/	6492.5872

permet l'élaboration d'une prévision à un horizon d'environ trois ans.

Cette partie nous a permis une modélisation et prévision sur un processus estimé à longue mémoire.

Conclusion

D'après les résultats obtenus dans ce chapitre, pour modéliser la première série "exportations Algériennes", il fallait choisir "la méthode de Box et Jenkins" la plus adéquate qui génère notre série, vu qu'on n'a pas pu appliquer la deuxième méthode.

Cependant, les hydrocarbures fortement vulnérables aux divers facteurs telle que les guerres, crises, monopoles des pays OPEP,.. , ainsi que les exportations appartenant a un marché non stable, nous explique pourquoi la série est anti persistante et ne contient pas une mémoire aussi longue pour nous permettre d'élaborer des prévisions.

D'autre part, la modélisation de la deuxième série "l'indice de clôture quotidien du marché boursier koweïtien", nous a orienté vers " la Méthode de Geweke Porter-Hudak (GPH)" la plus adéquate qui génère cette série, où on à eu une série persistante et fortement liée nous a permit d'effectuer des prévisions à longue mémoire.

Conclusion

A la lumière des résultats reportés dans ce modeste travail, on a présenté les deux méthodes de modélisation et prévision objet de notre mémoire, ensuite on a procédé à l'étude du comportement de "la série des exportations Algérienne" ensuite "l'indice de clôture quotidien du marché boursier koweïtien", objet de notre étude, à l'aide des processus à court mémoire comme à longue mémoire. Puisque les deux processus prennent la même forme et la seule différence est le domaine ou appartient le coefficient d'intégration " d ".

Après avoir stationnariser la série, nous avons procédé à l'estimation des deux processus respectivement à court puis à longue mémoire, il fallut s'assurer de la convergence ainsi que la signification de " d ", aussi il a été nécessaire d'avoir des résidus constituant un bruit blanc.

Nous avons validé le modèle *ARIMA*, après avoir appliqué les tests nécessaires, qui représente au mieux notre série, cela nous a permis d'établir des prévisions. En contrepartie l'application de la deuxième approche " *ARFIMA* " n'a été possible vu l'absence de la mémoire longue dans notre série des exportations. Cela peut être interprété par la situation économique Algérienne et la structure des exportations, ou on a constaté que les exportations Algériennes contiennent à 98% des hydrocarbures ce qui le relie fortement avec le marché non stable.

A cet effet, et d'après notre modeste étude on peut dire que l'utilisation des processus *ARFIMA* qui élargit la classe des processus *ARIMA* est nécessaire dans certaines séries.

Les statistiques calculées témoignent la présence du marche aléatoire (*ARMA*) et

rejette la présence de mémoire longue. Mais si le marché Algérien été stable, autrement dit un marché varié avec d'autres ressources telle que le tourisme, industrie, agriculture,..etc,on pouvait avoir une série persistante c'est à dire un processus à mémoire longue comme cité sur le *PNB* Américain.

L'apport principal de notre analyse est que l'on est capable de battre sur certain horizon la simple marche aléatoire. Ne nous pourrons pas pour autant affirmer que toutes les séries, et ce, nous amène vers d'autres horizons.

perspectives

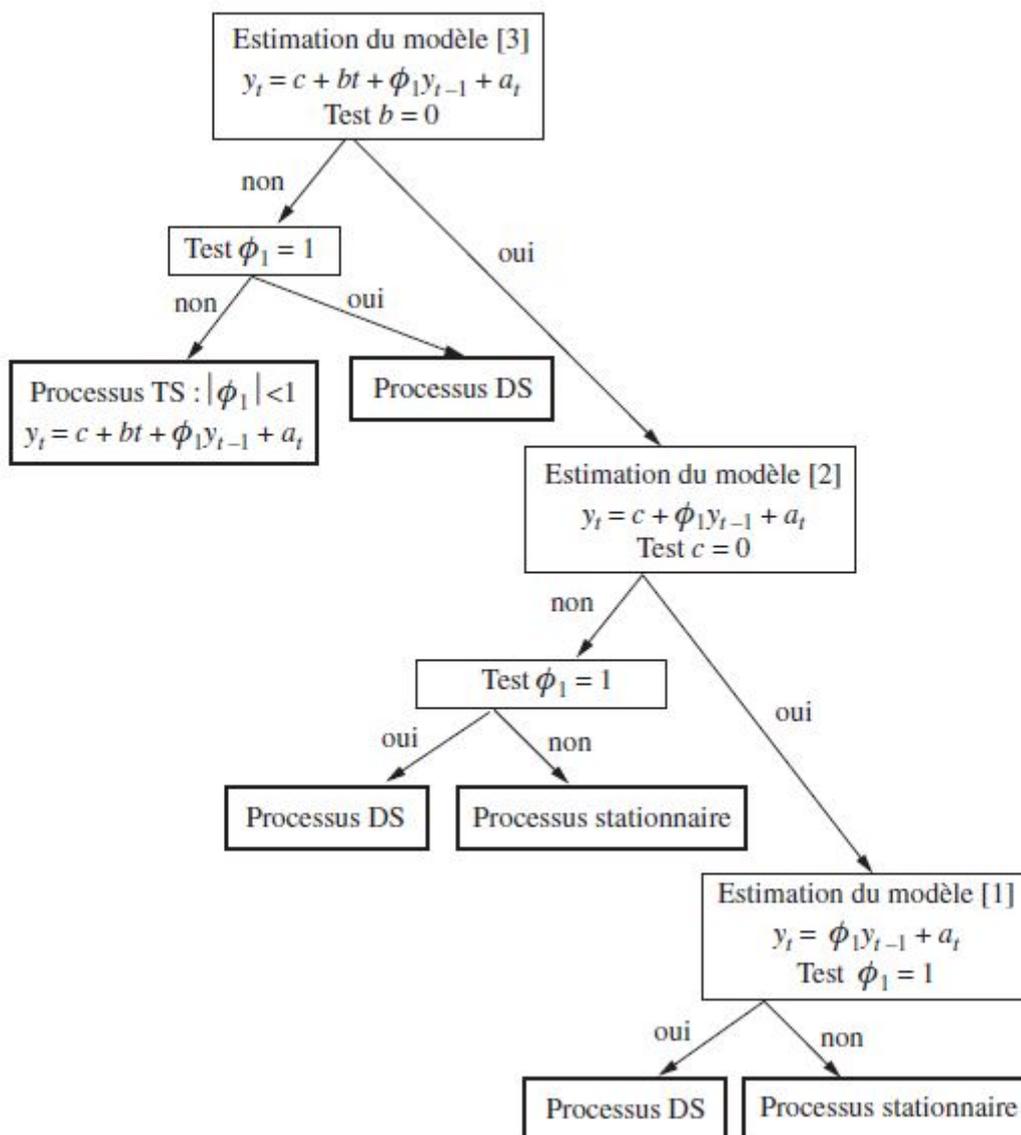
Aussi en matière de perspectives de recherches, nous préconisons l'application des modèles *FARIMA* pour la modélisation des chroniques se rapportant aux phénomènes relatifs à l'Environnement, vu la disponibilité des données et l'importance de ce domaine, la pollution fera un thème d'actualité, à savoir :

- La pollution de l'air ;
- La pollution des déchets ménagers.

A

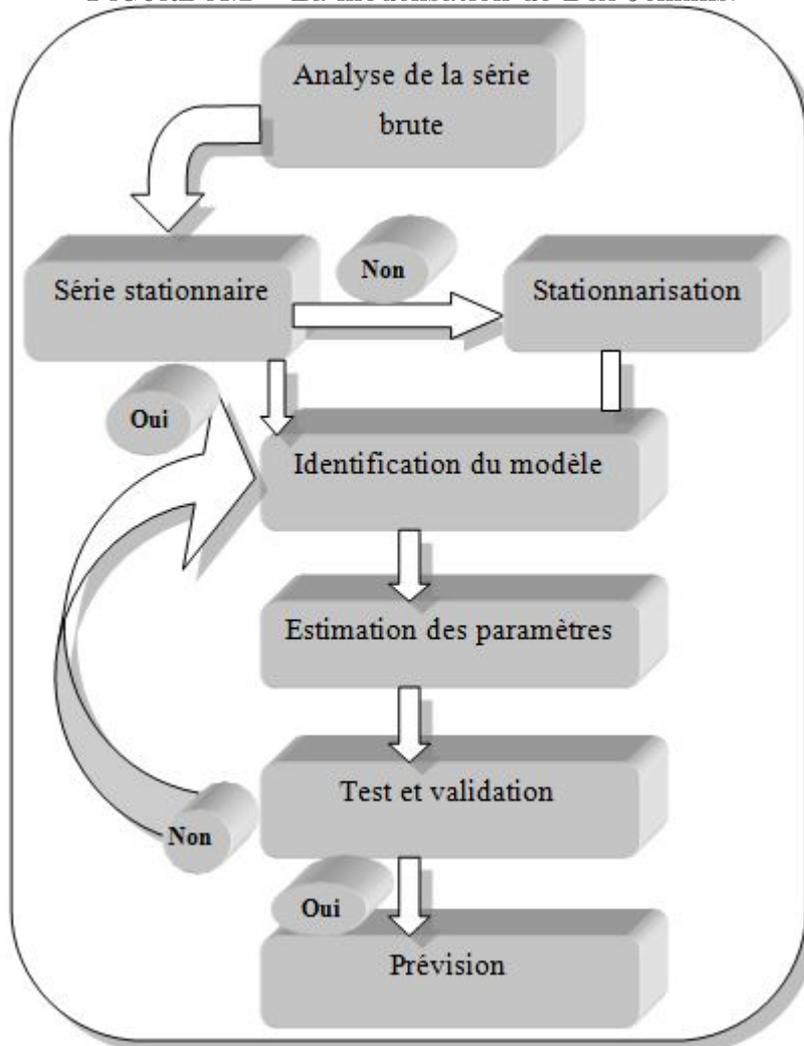
Annexes

FIGURE A.1 – Stratégie simplifiée des tests de racine unitaire.



Source : [18]

FIGURE A.2 – La modélisation de Box-Jenkins.



Source :[18]

Bibliographie

- [1] Alexandre, BROUSTE. 2014. "Contributions à la statistique des processus fractionnaires". *allintitle : "introduction-séries-chronologiques chapitre- Google Scholar"*. s. d. Consulté le 25 mai 2018. <https://scholar.google.com/scholar>.
- [2] Arthur, CHARPENTIER. 2011. s. d. "cours de séries temporelles théorie et applications". Vol. 2.
- [3] Arthur, CHARPENTIER. 2011. s. d. "Modèles de prévision Séries temporelles", 196.
- [4] Bernard, RAPACCHI. 1993. s. d. "Dur, dur!! Les series chronologiques", 42.
- [5] BOX G, E, P., JENKINS G, M. 1976. "Time series analysis : forecasting and control. Holdenday."
- [6] Bourbonnais, REGIS. 2010. *Économétrie : cours et exercices corrigés. Paris : Dunod.* s. d. "Analyse des séries temporelles", 346.
- [7] Bourbonnais, REGIS. 2015. *Économétrie : cours et exercices corrigés. Paris : Dunod.* s. d. "Analyse des séries temporelles", 394.
- [8] Butler, J. G. Jayson, et A. J. Swallow. 1975. "The Reaction between the Superoxide Anion Radical and Cytochrome C". *Biochimica Et Biophysica Acta* 408 (3) : 215-22.
- [9] COLLETAZ, Gilbert. 2018. s. d. "LES PROCESSUS VAR notes du cours de séries temporelles", 80.
- [10] Corinne, P. 2005. *Séries chronologiques : Quelques éléments du cours. Paris.*

- [11] Daudin, J J, C Duby, S Robin, et P Trecourt. 1996. s. d. "Analyse de Séries Chronologiques", 47.
- [12] DIEBOLD, F, X. RUDEBUSCH, G, D. 1989. "Long Memory and Persistence in, Aggregate Output", *Journal of Monetary Economies*, Vol. 24, n° 2, pp. 189-209.
- [13] Faÿ, Gilles. 2008. s. d. "Contributions à l'analyse statistique des processus à longue mémoire et du fond cosmologique", 64.
- [14] Florence, NICOLAU. 2005. "Chapitre 1 : Généralités sur les séries chronologiques"
- [15] Godet, Fanny. 2008. s. d. "Prévision linéaire des processus à longue mémoire", 136.
- [16] GEWEKE J, PORTER-HUDAK S. 1983. "The Estimation and Application of Long Memory time series models". *Journal of time series analysis*, Vol.4, N°4, pp. 221-238.
- [17] Hamisultane, Hélène. s. d. "Econométrie des séries temporelles", 24.
- [18] Hamrita, Mohamed Essaied. 2013. s. d. "Séries temporelles - Chapitre 3 : Méthodologie de Box et Jenkins", 47.
- [19] HURLIN, Christophe. s. d. "Econométrie Appliquée Séries Temporelles", 80.
- [20] HURST ,H.E. 1951. "Long-Term Storage Capacity of Réservoirs", *Transactions of the American Society of Civil Engineers*. Vol, 116.
- [21] Kouamo, Olaf. 2011. s. d. "Analyse des séries chronologiques à mémoire longue dans le domaine des ondelettes", 167.
- [22] Kröger, H., I. Donner, et G. Skiello. 1975. "Influence of a New Virostatic Compound on the Induction of Enzymes in Rat Liver". *Arzneimittel-Forschung* 25 (9) : 1426-29.
- [23] MANDELBROT B.B, TAQQU M. 1979. "Robust R/S Analysis of Long Run Sériai Corrélation", *Bulletin of the International Statistical Institute*, 48, n° 2.
- [24] Mignon, Sandrine., Lardic Valérie. 1999. s. d. " La mémoire longue en économie : une revue de la littérature", 45.
- [25] Ramdane, DERGUAOUI. "2015. "Mémoire de Master. Processus de longue mémoire". *Université de Bejaia*.
- [26] SOWELL, F. 1992. "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated time series models. *Journal of econometrics*. Vol, 53.

-
- [27] Walid , SACI. 2014. "*Mémoire de Master. Etude du modèle autoregréssif fractionnaire d'ordre 1*". Université de Bejaia.
- [28] Youcef ,BENYAMMI. 2015. "*Etude du pouvoir estimatif de la méthode Geweke Porter-Hudak sur les modèles ARFIMA : Application sur la température de l'aire de la ville d'Alger*". ENSSEA, no 60 : 24.

Résumé

Devant la complexité des séries économiques, divers modèles statistiques ont été développés. Mais l'innovation est marquée par la parution du livre de Box et Jenkins (1970) sur les processus court mémoire "ARMA", ces derniers s'avère insuffisants pour expliquer tous les phénomènes.

L'objectif de ce mémoire est de modéliser le comportement de certaines séries.

Nous nous sommes intéressés aux études théoriques des processus long mémoire "ARFIMA(p,d,q)" qui représente une extension des processus "ARIMA(p,d,q)" à court mémoire avant de dresser un panorama d'application par les méthodes, et ce, afin d'appuyer notre théorie.

Mots-Clés : ARMA, Mémoire longue, ARFIMA(p,d,q).

Abstract

In the front of the complexity of economic series, different statistic patterns are developed.

But the innovation has been marked by the publication of the book : "Box and Jenkins" (1970), on short memory process "ARMA". These later are not enough to explain all the phenomena the goal of this end studies thesis is to modelize the behavior of some series.

We were interested in the theoretical studies of long memory process "FARIMA" which represent an extension of "ARIMA" short memory process before setting up an application panorama in order to support our theory.

Keywords : ARMA, Long memory, ARFIMA(p,d,q).