

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



UNIVERSITE KASDI MERBAH
OUARGLA

Faculté des Mathématiques et des Sciences
de la Matière

N° d'ordre :
N° de série :

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire Présenté En Vue De L'obtention Du
DIPLÔME DE MASTER

Spécialité : Mathématiques

Option : Probabilité et Statistique

Présenté par :

GOHMES Khadidja

Prévision de la Série Chronologique
Nombre des Etudiants de l'Université de
Ouargla

Soutenu publiquement le : mai 2016

Devant le jury composé de :

Mr. AMARA Abdelkader. Université de KASDI Merbah - Ouargla	Président
M ^{elle} . SAIDANE Hadda. Université de KASDI Merbah - Ouargla	Examineur
M ^{elle} . ARBIA Hanane. Université de KASDI Merbah - Ouargla	Rapporteur

Table des matières

Dédicace	i
Remerciements	ii
Table des figures	iii
Notations et abreviations	iv
Introduction	1
1 Généralités sur les séries chronologique	2
1.1 Définition d'une série chronologique	2
1.2 Objectifs de l'analyse d'une chronique	3
1.3 Domaines d'application	4
1.4 Les composantes d'une série chronologique	4
1.5 Les schémas de composition	5
1.5.1 Choix du modèle	6
1.6 Autocorrélations simple et partielle	6
1.6.1 La fonction d'autocovariance et d'autocorrélation	7
1.6.2 La fonction d'autocorrélation partielle	7
1.7 Bruit blanc	9
1.8 stationnarité	9
1.8.1 Relation entre stationnarité faible et stricte	10
1.8.2 Test de Dickey Fuller Augmenté	10
1.9 Théorème de Wold	11
1.10 Opérateurs définis sur une série chronologique	13
1.10.1 Opérateur de retard	13
1.10.2 Opérateur de différence d'ordre d	13
1.11 Filtres linéaires et moyennes mobiles	14
1.11.1 Filtres linéaires	14
1.11.2 Moyennes mobiles	14

1.11.3	Moyenne mobile et composante saisonnière	14
1.12	Analyse de la tendance	17
1.12.1	Rappels sur la régression linéaire	17
1.12.2	Ajustement tendanciel linéaire par moindres carrés	19
1.12.3	Ajustement tendanciel linéaire par points médians	19
1.12.4	Ajustements tendanciels non linéaires	20
1.12.5	Estimation non paramétrique	20
2	Modélisation <i>ARIMA</i> des séries chronologique	21
2.1	Les modèles autorégressifs, (<i>Autoregressive</i>)	21
2.1.1	Représentation stationnaire	22
2.1.2	Représentation inversible	22
2.1.3	Représentation causale	22
2.1.4	Auto-covariance, auto-corrélations et équivalence de Yule-Walker	24
2.1.5	Prédiction dans les modèles autorégressifs	27
2.2	Les modèles moyennes mobiles, <i>MA</i> (<i>Moving Average</i>)	28
2.2.1	Représentation stationnaire	28
2.2.2	Représentation causale	28
2.2.3	Représentation inversible	28
2.2.4	Structure d'autocorrélation	29
2.2.5	Prévisions à l'aide d'un modèle <i>MA</i> (<i>q</i>)	30
2.2.6	Dualité des processus <i>AR</i> (<i>p</i>) et <i>MA</i> (<i>q</i>)	31
2.3	Les modèles <i>ARMA</i> (<i>p, q</i>) (<i>Autorégressive Moving Average</i>)	31
2.3.1	Inversibilité et stationnarité en terme de valeurs passées	31
2.3.2	Auto-corrélation	32
2.3.3	Auto-corrélations partielles	33
2.3.4	Prévisions à l'aide d'un modèle <i>ARMA</i> (<i>p, q</i>)	33
2.4	Les modèles <i>ARIMA</i> (<i>p, d, q</i>)	34
2.4.1	Prévisions dans le cas d'un processus <i>ARIMA</i> (<i>p, d, q</i>)	34
2.5	Méthode de Box & Jenkins	35
2.5.1	Identification du modèle	35
2.5.2	Estimation des paramètres	36
2.5.3	Adéquation et validation du modèle	37
2.5.4	Prévision	39
3	Application	40
3.1	Identification de la série	40
3.1.1	Les tests informels	40
3.1.2	Tests formels : Test de racine unitaire (<i>ADF</i>)	42

3.2	Stationnarisation de la série	45
3.3	Identification du modèle optimal provisoire	48
3.4	Estimation	49
3.5	Validation du modèle	50
3.5.1	Test de Student des paramètres (significativité statistique)	50
3.5.2	Test de bruit blanc	50
3.5.3	Test de normalité des résidus	51
3.6	La prévision	52
	Conclusion générale	53
	Bibliographie	54

Dédicace

Je dédie ce modeste travail

A ma très cher Mère et mon très cher Père

A ceux qui ont veillé pour mon bien être

A ceux qui m'ont toujours encouragé pour que je réussisse dans mes études

A tout ce qui est m'encouragé dans la réalisation de ce modeste travail

A mes chers soeurs

A tous mes frères

A toutes la familles : Gohmes, Terriche

N'oublier pas de dédier à mon très cher amies intime et ma copine de chambre

Tous mes camarades surtout la promotion 2015/2016

Option Sciences Proba-Stat

Tous qui m'on aider de proche ou de loin.

Remerciements

Tout d'abord, je remercie ALLAH Le Tout Puissant qui m'a accordée la volonté et le courage pour réaliser ce mémoire.

Mes plus vifs remerciements et ma profonde gratitude vont à Pr.ARBIA HANANE notre promoteur de mémoire pour sa grande patience, pour sa disponibilité, pour ses nombreux conseils, pour ses corrections et son appréciation au cours de l'élaboration de ce travail, notamment quand à la portée de la problématique. Elle a toujours été présente dans les moments où j'en avais besoin ; elle je faudrait des pour le remercier.

Je remercie sincèrement, les membres de jury d'avoir bien voulu accepter de faire partie de la commission d'examinateur.

Je remercie aussi l'ensemble de mes camarades de master, mes amis, mes proches et mes familles (au sens large) qui m'ont soutenue durant ce travail, et spécialement à celles ou ceux, elles ou ils reconnaîtront, qui j'ont encouragée à finir ce travail et qui j'ont accompagnée dans tous les moments de joie et de tristesse.

Enfin, je ne saurais terminer cette partie sans exprimer notre gratitude à mes parents, mes grands parents, mes frères et mes soeurs qui j'ont toujours soutenue, encouragé et stimulée pendant mes études.

Merci à tous et à toutes.

Table des figures

1.1	Population algérienne de 1960 à 2015	3
1.2	Les composantes d'une série chronologique	5
3.1	Graphique de l'évolution du nombre des étudiants de l'université de ouargla 1987- 2015.	41
3.2	Corrélogramme de la série.	42
3.3	Test ADF sur premier série	44
3.4	Estimation de la tendance	45
3.5	Test ADF sur la série d'après enlève la tendance	46
3.6	La série stationnaire $d(jj)$	47
3.7	corrélogramme de la séries $d(jj)$	48
3.8	Estimation de modèle $AR(1)$	49
3.9	corrélogramme des résidus du modèle $AR(1)$	51
3.10	Normalité des résidus du modèle $AR(1)$	52

Notations et abriviations

ADF	Dickey-Fuller Augmenté
AIC (p, q)	Critère d'Akaike
AR	Autorégressifs, (Autoregressive)
ARMA(p, q)	Autoregressive Moving Average
ARIMA(p, d, q)	Autoregressive Integrated Moving Average
B	Opérateur retard
BB	Bruit blanc
BIC (p, q)	Critère de Schwarz
DS	Difference Stationary
F	Opérateur d'avance
JB	Test de Jarque & Bira
MA	Moyennes Mobiles (Moving Average)
MCO	Moindres carrés ordinaires
MM	Moyenne mobile
TS	Trend Stationary
VCM	Valeur Critique de MacKinnon
$\xrightarrow{\mathcal{L}}$	Convergence en loi

Introduction

Une série chronologique est une suite formée d'observations au cours du temps. L'analyse des séries chronologiques est un outil couramment utilisé de nos jours pour la prédiction de données futures. Ce domaine possède beaucoup d'applications en finance, en médecine, en économétrie et en météorologie et dans bien d'autres domaines. L'idée est de prendre un échantillon de données et de construire le meilleur modèle qui ajuste ces données. Ce modèle nous permet de tirer certaines conclusions sur la série.

Notre travail est structuré de la façon suivante ; dans le premier chapitre, on donne les définitions nécessaires à la compréhension du corps de ce travail. Ensuite, le deuxième chapitre explique les trois principaux modèles : AR, MA, ARIMA, et la méthode de Box & Jenkins. Finalement, le troisième chapitre présente l'application de la méthode de Box & Jenkins avec le programme Eviews sur le nombre des étudiants de l'université Ouargla (1987 – 2015).

Chapitre 1

Généralités sur les séries chronologiques

La théorie des séries chronologiques (ou temporelles) est appliquée de nos jours dans des domaines aussi variés que l'économétrie, la médecine ou la démographie. Dans ce chapitre, plusieurs concepts importants liés à l'analyse des séries chronologiques seront abordés. Parmi ceux-ci, on retrouve les notions de bruit blanc, d'autocorrélation, de stationnarité et opérateurs définis sur une série chronologique.

1.1 Définition d'une série chronologique

Une série statistique est désignée comme étant chronologique quand les données qui la constituent sont les valeurs d'une variable enregistrée en fonction de la date pendant une certaine période [8], [9].

Définition 1.1.1 *On appelle série chronologique une suite d'observations numériques d'une grandeur effectuées à intervalles réguliers au cours du temps.*

Définition 1.1.2 *Une série chronologique, ou chronique ou série temporelle, est une suite finie de données indexée par le temps.*

Si t_1, t_2, \dots, t_n sont les n instants successifs d'observation et si X_{t_i} est la valeur mesurée à l'instant t_i , on notera la série chronologique $\{X_t\}_{t \in T}$ où T est l'ensemble ordonné $T = \{t_1, t_2, \dots, t_n\}$.

Exemple 1.1.1 (*Sur les séries chronologique*)

1. Nombre mensuel de vente de voitures neuves en France.
2. Ecologie (pollution).
3. Taille de la population algérienne de 1960 à 2015.

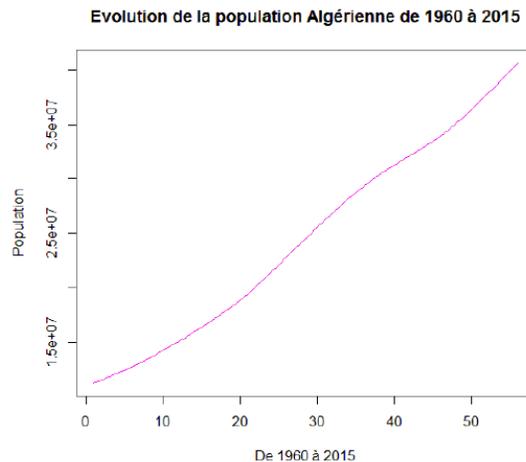


FIG. 1.1 – Population algérienne de 1960 à 2015

1.2 Objectifs de l'analyse d'une chronique

Les objectifs de l'analyse d'une chronique sont :

1. Modélisation

Elle consiste à :

- Développer des modèles permettant de décrire le comportement d'une ou plusieurs séries chronologiques.
- Mettre au point une méthodologie pour spécifier, estimer, valider (juger) un modèle approprié pour des données particulières.

2. Prévision

Étant données des observations X_1, \dots, X_T , la prévision consiste à évaluer une valeur non observée, X_{T+h} . La prévision peut être ponctuelle, ou prendre la forme d'un intervalle de prévision.

1.3 Domaines d'application

On trouve des exemples de séries chronologiques dans de très nombreux domaines. La liste suivante n'est qu'un échantillon [13] :

1. Finance et économétrie : évolution des indices boursiers, des prix, des données économiques des entreprises, des ventes et achats de biens, des productions agricoles ou industrielles.
2. Médecine : suivi des évolutions des pathologies.
3. Sciences de la terre et de l'espace : indices de marées, variations des phénomènes physiques (météorologie), évolution des taches solaires.
4. Traitement du signal : signaux de communications, de radars, de sonars, analyse de la parole.
5. Traitement des données : mesures successives de position ou de direction d'un objet mobile (trajectographie).
6. Traitement d'images : analyse d'images satellites.
7. Démographie : analyse de l'évolution d'une population.
8. Météorologie : analyse de données climatiques.
9. Energie : prévision de la consommation d'électricité.

1.4 Les composantes d'une série chronologique

La première étape dans l'étude des séries chronologiques consiste à représenter graphiquement l'événement qui permet d'entrevoir les quatre composantes fondamentales d'une chronique [9].

1. **La tendance** Z_t : Représente l'évolution à long terme de la grandeur étudiée, et traduit l'aspect général de la série.
2. **Les variations saisonnières** S_t : Sont liées au rythme imposé par les saisons météorologiques (production agricole, consommation de gaz,...), ou encore par des activités économiques et sociales (fêtes, vacances, solde, etc). Elles sont de nature périodique, c'est-à-dire qu'il existe un entier p , appelé période, tel que $S_t = S_{t+p}$ pour tout $t \geq 1$.
3. **Cycle** C_t : Regroupe les variations autour de la tendance avec des alternances de phases d'expansion et de recession. Ces phases durent généralement plusieurs années, mais n'ont pas de durée fixe.
4. **Les variations accidentelles ou résiduelles** ε_t : les variations accidentelles sont des fluctuations irrégulières et imprévisibles. Elles sont supposées en général de faible amplitude. Elles proviennent de circonstances non prévisibles : catastrophes naturelles, crise boursière, grèves,...

Le graphique suivant présente l'ensemble des composantes précitées

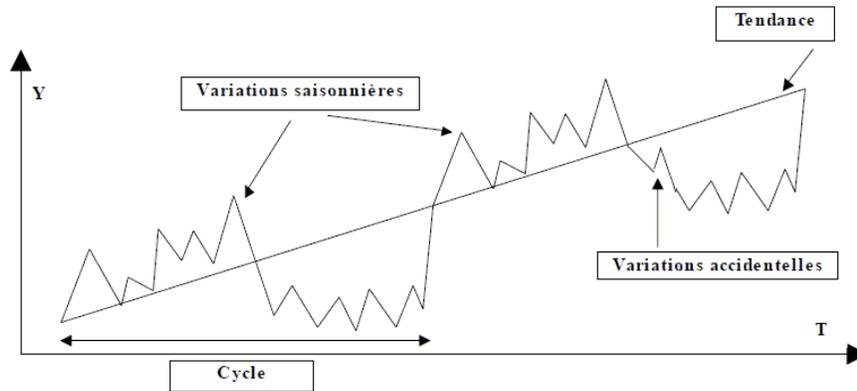


FIG. 1.2 – Les composantes d'une série chronologique

1.5 Les schémas de composition

Pour pouvoir séparer les composantes servant à décrire la série observée, il est nécessaire de préciser leur mode d'interaction. La plupart des séries chronologiques entrent dans l'un des schémas suivants [8], [14] :

1. Schéma additif

Dans ce modèle, la série chronologique, s'écrit de la façon suivante :

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

La tendance, les variations saisonnières, les variations accidentelles et cyclique, ont un effet additif dans ce modèle de série chronologique.

2. Schéma multiplicatif

Dans ce modèle, la série chronologique, s'écrit de la façon suivante :

$$X_t = Z_t \times S_t \times C_t \times \varepsilon_t.$$

Dans ce modèle, La tendance, les variations saisonnières, les variations accidentelles et cyclique, ont un effet multiplicatif sur la taille de la série chronologique.

3. Schéma mixte

Il s'agit là des modèles où addition et multiplication sont utilisées. On peut supposer, par exemple, que la tendance, la composante saisonnière, et cyclique ont la forme multiplicative, alors que les fluctuations irrégulières sont additives :

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

Remarque 1.5.1 La tendance (Z_t) prend différentes forme :

- a. *Linéaire* : $Z_t = a + bt$.
- b. *Quadratique* : $Z_t = a + bt + ct^2$.
- c. *Exponentielle* : $Z_t = ae^{bt}$.

1.5.1 Choix du modèle

Graphiquement : Constatation graphique du parallélisme ou non des droites des minima et des maxima :

- Droites parallèles : modèle additif.
- Droites non parallèles : modèle multiplicatif.

Algébriquement : On recherche le coefficient directeur de ces deux droites par une méthode analytique (deux points).

- Coefficients directeurs très proches : le modèle choisi sera additif.
- Coefficients directeurs très différents : le modèle choisi sera multiplicatif.

1.6 Autocorrélations simple et partielle

Les principales caractéristiques temporelles sont données par l'autocorrélation (simple) et l'autocorrélation partielle [7].

1.6.1 La fonction d'autocovariance et d'autocorrélation

La fonction d'autocovariance $\{\gamma(h)\}_{h \in \mathbb{Z}}$ mesure la covariance entre une variable et cette même variable à des dates différentes, pour un délai h :

$$\gamma(h) = \text{cov}(X_t, X_{t-h}) = E[(X_t - E(X_t))(X_{t-h} - E(X_{t-h}))].$$

Ainsi

$$\gamma(0) = \text{Var}(X_t) = E[(X_t - E(X_t))^2] = \sigma_X^2.$$

Elle fournit une information sur la variabilité de la série et sur les liaisons temporelles qui existent entre les diverses composantes de la série X_t .

La fonction d'autocorrélation est définie par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad h \in \mathbb{Z}.$$

Avec $\rho(0) = 1$ et $|\rho(h)| < 1$.

L'équivalent empirique de la fonction d'autocorrélation, noté $\hat{\rho}(h)$, est obtenu à partir de l'estimateur suivant pour l'autocovariance $\hat{\gamma}(h)$ à l'ordre h :

$$\hat{\gamma}(h) = \frac{1}{T-h-1} \sum_{t=h+1}^T (X_t - \bar{X})(X_{t-h} - \bar{X}).$$

On définit la matrice de corrélation (de dimension m) de la manière suivante :

$$R(m) = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(m-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(m-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(m-1) & \rho(m-2) & \dots & \rho(1) & 1 \end{pmatrix}$$

1.6.2 La fonction d'autocorrélation partielle

Elle mesure la liaison (linéaire) entre X_t et X_{t-h} une fois retirés les liens transitant par les variables intermédiaires $X_{t-1}, \dots, X_{t-h+1}$.

Le coefficient d'autocorrélation partielle d'ordre h , noté $r(h)$, est définie par :

$$r(h) = \text{corr}(X_t, X_{t-h} / X_{t-1}, \dots, X_{t-h+1}).$$

Le coefficient d'autocorrélation partielle d'ordre h d'un processus stationnaire est se calcule de la manière suivante :

$$r(h) = \frac{|R(h)^*|}{|R(h)|},$$

avec

$$R(h) = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-2) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \rho(h-1) & \rho(h-2) & \dots & 1 \end{pmatrix}$$

et $R(h)^*$ la matrice $R(h)$ dans laquelle on a remplacé la colonne h par $\begin{pmatrix} \rho(1) \\ \rho(2) \\ \cdot \\ \cdot \\ \rho(h) \end{pmatrix}$,

soit :

$$R(h)^* = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \rho(h-1) & \rho(h-2) & \dots & \rho(h) \end{pmatrix}$$

Ainsi,

$$r(1) = \rho(1), r(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} \dots$$

De manière empirique, les autocorrélations partielles s'estiment par l'estimation des autocorrélations simples et en calculant $\hat{r}(h)$ à partir de la formule ci-dessus.

1.7 Bruit blanc

Définition 1.7.1 On dit que la suite de variables aléatoires $\{\varepsilon_t\}$ qu'elle constitue un bruit blanc faible si elle possède les propriétés suivantes [10] :

$$\begin{cases} E(\varepsilon_t) = 0 \text{ pour tout } t \in \mathbb{Z}, \\ E(\varepsilon_t^2) = \sigma_\varepsilon^2 \neq 0, \\ \text{cov}(\varepsilon_s, \varepsilon_t) = 0, \text{ si } t \neq s. \end{cases}$$

En d'autres termes, les variables aléatoires ε_t sont de moyenne nulle, de variance constante et non corrélées.

Remarque 1.7.1 On dit que $\{\varepsilon_t\}$ est un bruit blanc fort s'il est un bruit blanc faible et que les variables aléatoires ε_t sont i.i.d. (indépendantes et identiquement distribuées).

1.8 stationnarité

La stationnarité est une caractéristique d'une série chronologique qui implique que le comportement de la série ne dépend pas du temps. En particulier, on dit qu'une série X_t est stable si elle ne comporte pas de tendance à la hausse ou à la baisse. Plus formellement, on distingue deux types de stationnarité, à savoir forte et faible [12], [17].

Définition 1.8.1 (Stationnarité faible) Un processus $(X_t)_{t \in \mathbb{Z}}$ est dit **stationnaire** ou **faiblement stationnaire** ou **stationnaire au second ordre** si

- i $E(X_t^2) < \infty$,
- ii $E(X_t) = \mu, \forall t \in \mathbb{Z}$ (donc ne dépend pas du temps),
- iii $\text{Cov}(X_t, X_{t-h}) = E(X_t X_{t-h}) - E(X_t)E(X_{t-h}) = \gamma(h) \quad \forall t \in \mathbb{Z}, \quad \forall h \in \mathbb{Z}$ (ne dépend pas du temps).

En résumé, un processus X_t est dit stationnaire du second ordre si sa moyenne, sa variance et sa covariance sont indépendantes du temps et si sa variance est finie.

Définition 1.8.2 (Stationnarité stricte) Un processus $(X_t)_{t \in \mathbb{Z}}$ est dit **strictement stationnaire** ou **stationnaire au sens strict** si les lois jointes de $(X_{t_1}, \dots, X_{t_k})$ et de $(X_{t_1+h}, \dots, X_{t_k+h})$ sont identiques pour tout entier positif k et pour tous $t_1, \dots, t_k, h \in \mathbb{Z}$.

Intuitivement, une série chronologique strictement stationnaire doit avoir le même comportement statistique sur des intervalles de temps égaux.

1.8.1 Relation entre stationnarité faible et stricte

Proposition 1.8.1 *Un processus strictement stationnaire du second ordre est faiblement stationnaire. La réciproque n'est pas vraie en général.*

Contre-exemple : Soit (X_t) une suite de variables aléatoires indépendantes telle que

- $X_t \sim \exp(1)$ lorsque t est pair,
- $X_t \sim N(1, 1)$ lorsque t est impair,

alors (X_t) est stationnaire avec $\gamma_X(0) = 1$ et $\gamma_X(h) = 0$ lorsque $h \neq 0$. Cependant X_1 et X_2 n'ont pas la même loi donc (X_t) n'est pas strictement stationnaire.

Difference-Stationary

Un processus X_t est dit difference-stationary (DS) ou stationnaire en différence s'il peut s'écrire sous l'une des formes suivante :

- (1) $X_t = X_{t-1} + \varepsilon_t$.
- (2) $X_t = X_{t-1} + \beta + \varepsilon_t$ (avec $\beta \neq 0$).

Trend-Stationary

Il sera dit Trend-stationary (TS) ou stationnaire en tendance s'il peut se mettre sous la forme :

$$X_t = g(t) + \varepsilon_t,$$

$g(t)$ est une fonction polynômiale du temps.

La bonne manière de stationnariser une série TS consiste à estimer, en général par les moindres carrés ordinaires (MCO), l'expression de la tendance et à la retirer. Tandis que la stationnarisation des séries DS se fait par passage aux différences.

1.8.2 Test de Dickey Fuller Augmenté

Les tests de stationnarité permettent de vérifier si une série est stationnaire ou non. Le test de Dickey-Fuller, le test augmenté de Dickey-Fuller (ADF), ou encore le test de Phillips-Perron (PP) sont des tests de racine unitaire, pour lesquels l'hypothèse nulle est que la série a été générée par un processus présentant une

racine unitaire, et donc, qu'elle n'est pas stationnaire. On donne ici le test de *ADF*.

Les hypothèses du test de *ADF* se définissent de la façon suivante ; Un processus non stationnaire, il correspond à une de ces formes de non stationnarité :

$$\Delta X_t = \phi X_{t-1} + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \varepsilon_t \quad \text{avec } \phi = \rho - 1 \quad \text{"Modèle[1]"},$$

$$\Delta X_t = \alpha + \phi X_{t-1} + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \varepsilon_t \quad \text{"Modèle[2]"},$$

$$\Delta X_t = \alpha + \beta t + \phi X_{t-1} + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \varepsilon_t \quad \text{"Modèle[3]}."$$

1.9 Théorème de Wold

Le théorème de Wold est fondamental pour l'analyse de séries temporelles stationnaires [10].

Théorème 1.9.1 *Tout processus $(X_t)_{t \in \mathbb{Z}}$ faiblement stationnaire peut s'écrire sous la forme :*

$$X_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j} + k_t,$$

où les paramètres ψ_1, ψ_2, \dots sont des réels tels que $\psi_0 = 1$ et $\sum_{j=0}^{\infty} \psi_j^2 < +\infty$, $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc et k_t est une composante linéaire telle que

$$\text{Cov}(k_t, \varepsilon_{t-j}) = 0 \quad \text{pour tout } j \in \mathbb{Z}.$$

Selon le théorème de Wold, tout processus stationnaire d'ordre 2 peut être représenté comme une somme pondérée infinie de chocs passés, caractérisés par un bruit blanc.

Dans ce qui suit, deux exemples de séries chronologiques stationnaires sont présentés. On obtiendra, dans chaque cas, leur représentation selon la formule du théorème de Wold.

Exemple 1.9.2 *Soit le processus*

$$X_t = \sum_{j=0}^{+\infty} \left(\frac{1}{2}\right)^j \nu_{t-j+1},$$

où $(\nu_t)_{t \in \mathbb{Z}}$ est un bruit blanc gaussien de variance unitaire. Afin que la condition $\psi_0 = 1$ du théorème de Wold soit satisfaite, il s'agit de poser

$$\varepsilon_t = \frac{1}{2} \nu_t.$$

on a alors

$$X_t = \sum_{j=0}^{+\infty} \left(\frac{1}{2}\right)^{j-1} \varepsilon_{t-j+1} = \varepsilon_t + \sum_{j=0}^{+\infty} \left(\frac{1}{2}\right)^j \varepsilon_{t-j},$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance $Var(\varepsilon_t) = Var(\nu_t/2) = \frac{1}{4}$. Ainsi, $\psi_0 = 1$ et $\psi_j = (1/2)^j$. On note que la condition $\sum_{j=0}^{+\infty} \psi_j^2 < +\infty$ est également satisfaite car

$$\sum_{j=0}^{+\infty} \psi_j^2 = \sum_{j=0}^{+\infty} \left(\frac{1}{2}\right)^{2j} = \sum_{j=0}^{+\infty} \left(\frac{1}{4}\right)^j = \frac{4}{3},$$

où le résultat

$$\sum_{j=0}^{+\infty} \alpha^j = \frac{1}{1-\alpha}, \quad |\alpha| < 1,$$

été utilisé.

Exemple 1.9.3 *Soit le processus*

$$X_t = \rho X_{t-1} + \varepsilon_t,$$

où $|\rho| < 1$ et $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance $\sigma_\varepsilon^2 = 1$.

Alors X_t a une représentation sous la forme

$$X_t = \sum_{j=0}^{+\infty} \rho^j \varepsilon_{t-j},$$

qui est précisément un cas particulier de la formule $X_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j} + k_t$ avec $\psi_j = \rho^j$. En utilisant le résultat sur la somme géométrique infinie, on note que

$$\sum_{j=0}^{+\infty} \psi_j^2 = \sum_{j=0}^{+\infty} (\rho^j)^2 = \sum_{j=0}^{+\infty} (\rho^2)^j = \frac{1}{1-\rho^2} < +\infty.$$

1.10 Opérateurs définis sur une série chronologique

Si la série X_t n'est pas stationnaires l'opérateur suivant rend cette série stationnaire [6], [12].

1.10.1 Opérateur de retard

Définition 1.10.1 *L'opérateur de retard B se définit de la manière suivante :*

$$B(X_t) = X_{t-1}.$$

Remarque 1.10.1 *L'opérateur B est linéaire et inversible. Son inverse $B^{-1} = F$ est défini par*

$$\forall t \in \mathbb{Z}, FX_t = X_{t+1}.$$

L'opérateur F est appelé opérateur d'avance.

Si on compose B avec lui-même on obtient $B^2 = B \circ B$ tel que

$$\forall t \in \mathbb{Z}, B^2 X_t = X_{t-2}.$$

On peut itérer cette application et définir par récurrence

$$B^k X_t = X_{t-k}, k \in \mathbb{N}.$$

Par convention, B^0 est l'opérateur identité I .

1.10.2 Opérateur de différence d'ordre d

Définition 1.10.2 *On définit l'opérateur Δ^d de différence d'ordre d comme l'opérateur linéaire tel que*

$$\Delta^d X_t = (1 - B)^d X_t.$$

On peut aussi prendre l'opérateur d'ordre 1 et l'appliquer plusieurs fois :

$$\Delta^2(X_t) = \Delta(\Delta(X_t)) = \Delta(X_t - X_{t-1}) = (1 - B)(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}.$$

1.11 Filtres linéaires et moyennes mobiles

Nous donnons ici deux définitions importants, pour l'étude des séries chronologiques [21].

1.11.1 Filtres linéaires

Un filtre linéaire d'ordre $m = p_1 + p_2 + 1$ est défini par

$$\begin{aligned} FL &= \sum_{j=-p_1}^{p_2} \omega_j B^{-j} \\ &= \omega_{-p_1} B^{p_1} + \omega_{-p_1+1} B^{p_1-1} + \dots + \omega_{-1} B + \omega_0 I + \omega_1 F + \dots + \omega_{p_2-1} F^{p_2-1} + \omega_{p_2} F^{p_2} \end{aligned}$$

où $p_1, p_2 \in \mathbb{N}$ et $\omega_j \in \mathbb{R}$.

1.11.2 Moyennes mobiles

Une moyenne mobile d'ordre $m = p_1 + p_2 + 1$ est un filtre linéaire tel que

$$\sum_{j=-p_1}^{p_2} \omega_j = 1, \text{ pour tout } j = -p_1, \dots, p_2.$$

Beaucoup de moyennes mobiles ont des poids ω_j positifs, mais pas toutes.

Une moyenne mobile est symétrique si $p_1 = p_2 = p$ et $\omega_j = \omega_{-j}$, pour tout $j = 1, \dots, p$.

Une moyenne mobile symétrique est dite non-pondérée si $\omega_j = cst$ pour tout $j = -p_1, \dots, p$.

1.11.3 Moyenne mobile et composante saisonnière

Une moyenne mobile est un outil intéressant pour lisser une série temporelle et donc pour enlever une composante saisonnière. On utilise de préférence des moyennes mobiles non-pondérées d'ordre égal à la période, par exemple d'ordre 7 pour des données journalières, d'ordre 12 pour des données mensuelles. Par exemple, pour enlever la composante saisonnière due au jour de la semaine, on peut appliquer une moyenne mobile non-pondérée d'ordre 7.

$$MM_7 = \frac{1}{7} (B^3 + B^2 + B + I + F + F^2 + F^3).$$

Cette moyenne mobile accorde le même poids à chaque jour de la semaine. En effet,

$$MM_7(X_t) = \frac{1}{7} (X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3}).$$

Remarque 1.11.1 *Pour les composantes saisonnières d'une période paire, il n'existe pas de moyennes mobiles centrées non-pondérées.*

Si la période est paire et égale à m , on utilise une moyenne mobile accordant un demi-poids aux deux extrémités.

Par exemple, pour des données trimestrielles ($m = 4$), la moyenne mobile est définie par

$$MM_4 = \frac{1}{4} \left(\frac{1}{2}B^2 + B + I + F + \frac{1}{2}F^2 \right).$$

En effet,

$$MM_4(X_t) = \frac{1}{4} \left(\frac{1}{2}X_{t-2} + X_{t-1} + X_t + X_{t+1} + \frac{1}{2}X_{t+2} \right).$$

Cas particulier : Moyenne mobile arithmétique

La série des moyennes mobiles arithmétiques d'ordre k (k est impair), est la série des moyennes de k observations consécutives et elle prend ses valeurs aux dates moyennes correspondantes. Plus précisément, on calcule les moyennes de k termes consécutifs pour les dates

$$\frac{t_1 + \dots + t_k}{k} \quad \text{puis} \quad \frac{t_2 + \dots + t_{k+1}}{k} \quad \text{jusqu'à} \quad \frac{t_{T-k+1} + \dots + t_T}{k},$$

et pour la variable d'intérêt

$$\frac{X_1 + \dots + X_k}{k} \quad \text{puis} \quad \frac{X_2 + \dots + X_{k+1}}{k} \quad \text{jusqu'à} \quad \frac{X_{T-k+1} + \dots + X_T}{k}.$$

Exemple 1.11.1 *Calcul d'une moyenne mobile arithmétique d'ordre 3*

Date t	Série y_t	Date $M_3(t)$ de la MM	$M_3(y_t)$
1	5		
2	3	$(1 + 2 + 3) / 3 = 2$	$(5 + 3 + 4) / 3 = 4$
3	4	$(2 + 3 + 4) / 3 = 3$	$(3 + 4 + 5) / 3 = 4$
4	5	$(3 + 4 + 5) / 3 = 4$	$(4 + 5 + 4) / 3 = 4.33$
5	4	$(4 + 5 + 6) / 3 = 5$	$(5 + 4 + 4) / 3 = 4.33$
6	4		

TAB. 1.1 – Moyenne Mobile d'ordre 3

Lorsque k est impair, $k = 2m + 1$, la série moyenne mobile est calculée aux mêmes instants que les observations initiales.

En revanche, lorsque k est pair, $k = 2m$, la moyenne mobile est calculée entre les dates d'observations. Si l'on veut comparer la série transformée à la série initiale, on a besoin d'avoir les valeurs pour les mêmes dates d'observations. Pour pallier cet inconvénient, on prendra plutôt comme transformation

$$\frac{1}{k} \left(\frac{1}{2} X_{t-m} + X_{t-m+1} + \dots + X_{t+m-1} + \frac{1}{2} X_{t+m} \right).$$

Exemple 1.11.2 *Calcul des moyennes mobiles arithmétiques d'ordre 2 et 4*

Date t	Série y_t	Date $M_2(t)$ de la MM	$M_2(y_t)$
1	5		
2	3	$(1/2 + 2 + 3/2) / 2 = 2$	$(5/2 + 3 + 4/2) / 2 = 3.75$
3	4	$(2/2 + 3 + 4/2) / 2 = 3$	$(3/2 + 4 + 5/2) / 2 = 4$
4	5	$(3/2 + 4 + 5/2) / 2 = 4$	$(4/2 + 5 + 4/2) / 2 = 4.5$
5	4	$(4/2 + 5 + 6/2) / 2 = 5$	$(5/2 + 4 + 4/2) / 2 = 4.25$
6	4		

TAB. 1.2 – Moyenne Mobile d'ordre 2

Date t	Série y_t	Date $M_4(t)$ de la MM	$M_4(y_t)$
1	5		
2	3		
3	4	$(1/2 + 3 + 4 + 5/2) / 4 = 3$	$(5/2 + 3 + 4 + 5 + 4/2) / 4 = 4.125$
4	5	$(2/2 + 3 + 4 + 5 + 6/2) / 4 = 4$	$(3/2 + 4 + 5 + 4 + 4/2) / 4 = 4.125$
5	4		
6	4		

TAB. 1.3 – Moyenne Mobile d'ordre 4

1.12 Analyse de la tendance

Dans ce section, nous nous plaçons dans le cadre d'un modèle composé uniquement d'une tendance et de fluctuations irrégulières et donnons différentes méthodes permettant d'estimer la tendance.

1.12.1 Rappels sur la régression linéaire

Lorsqu'une liaison linéaire forte entre deux variables X et Y semble raisonnable au vu du nuage de points, on a alors une relation du type :

$$Y \simeq aX + b.$$

où les coefficients a et b sont inconnus.

Le problème est que les points du nuage sont rarement (parfaitement) alignés : ils sont proches d'une droite.

Nous cherchons maintenant la droite qui passe au plus près des points du nuage. Pour cela, il faut donc mesurer l'éloignement des points du nuage par rapport à une droite D d'équation $y = ax + b$ puis minimiser un critère d'erreur donné. On peut envisager de minimiser

- la somme des erreurs en valeur absolue : $\min_{a,b} \sum_{i=1}^n |y_i - ax_i - b|$.
- la somme des erreurs au carré : $\min_{a,b} \sum_{i=1}^n (y_i - ax_i - b)^2$.

La méthode des moindres carrés minimisant le second critère est la plus usité.

La méthode des moindres carrés

On démontre en minimisant la fonction de deux variables

$$g(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2,$$

que le couple solution (\hat{a}, \hat{b}) est donné par

$$\hat{a} = \frac{\frac{1}{n} \sum_{i=1}^n y_i x_i - \left(\frac{1}{n} \sum_{i=1}^n y_i\right) \left(\frac{1}{n} \sum_{i=1}^n x_i\right)}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2} = \frac{Cov(X, Y)}{Var(X)} \quad \text{et} \quad \hat{b} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{a} \frac{1}{n} \sum_{i=1}^n x_i = \bar{Y} - \hat{a} \bar{X}.$$

La droite d'équation $y = \hat{a}x + \hat{b}$ est appelée droite de régression de Y en X et est notée : $\Delta_{Y/X}$.

1. Cette droite passe par le point moyen $M(\bar{X}; \bar{Y})$.
2. Le coefficient directeur a de $\Delta_{Y/X}$, $Cov(X, Y)$ et $r(X, Y)$ sont de même signe :

- Lorsqu'ils sont positifs, on parle de corrélation positive (y augmente quand x augmente).

- Lorsqu'ils sont négatifs, on parle de corrélation négative (y diminue quand x augmente).

Propriétés et interprétation du coefficient de corrélation linéaire

Afin de confirmer qu'il est raisonnable d'approximer le nuage de points par une droite, on calcule le coefficient de corrélation linéaire :

$$r(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

- a. Le coefficient de corrélation linéaire est symétrique :

$$r(X, Y) = r(Y, X).$$

- b. L'inégalité de **Cauchy-Schwarz** donne :

$$-1 \leq r(X, Y) \leq 1.$$

En pratique, il faut commencer par tracer le nuage de points puis calculer $r(X, Y)$ et ce n'est que si la corrélation linéaire est assez forte que l'on cherchera la droite de régression de Y en X .

1.12.2 Ajustement tendanciel linéaire par moindres carrés

Supposons que l'on observe T valeurs d'une série dont la tendance semble être linéaire. La méthode des moindres carrés décrite au paragraphe précédent consiste à estimer la tendance par une fonction linéaire

$$\widehat{Z}_t = \widehat{a}t + \widehat{b}.$$

Le couple solution $(\widehat{a}, \widehat{b})$ est donné par

$$\widehat{a} = \frac{\text{Cov}(t, Z_t)}{\text{Var}(t)} \text{ et } \widehat{b} = \bar{Z} - \widehat{a}\bar{t},$$

en posant $\bar{t} = \frac{1}{T} \sum_{t=1}^T t$, $\bar{Z} = \frac{1}{T} \sum_{t=1}^T Z_t$ et

$$\begin{cases} \text{Cov}(t, Z_t) = \frac{1}{T} \sum_{t=1}^T (t - \bar{t})(Z_t - \bar{Z}), \\ \text{Var}(t) = \frac{1}{T} \sum_{t=1}^T (t - \bar{t})^2. \end{cases}$$

Remarque 1.12.1 *Le coefficient de corrélation linéaire est défini par*

$$r = \frac{\text{Cov}(t, Z_t)}{\sqrt{\text{Var}(t)\text{Var}(Z_t)}}.$$

La corrélation linéaire entre la date t et la variable Z_t est d'autant plus importante que $|r|$ est proche de 1.

1.12.3 Ajustement tendanciel linéaire par points médians

On suppose ici aussi que la tendance est linéaire. Cette méthode est empirique et ne repose sur aucun critère d'erreur à minimiser. Elle peut cependant s'avérer efficace en présence de valeurs aberrantes. On choisit deux points de coordonnées (t_α, Z_α) et (t_β, Z_β) et on fait passer la droite par ces deux points.

Les coefficients $(\widetilde{a}, \widetilde{b})$ vérifient

$$\begin{cases} Z_\alpha = \widetilde{a}t_\alpha + \widetilde{b} \\ Z_\beta = \widetilde{a}t_\beta + \widetilde{b} \end{cases} \text{ soit } \begin{cases} \widetilde{a} = \frac{Z_\alpha - Z_\beta}{t_\alpha - t_\beta} \\ \widetilde{b} = Z_\alpha - \frac{Z_\alpha - Z_\beta}{t_\alpha - t_\beta} t_\alpha \end{cases}$$

Pour choisir les deux points, on constitue deux sous-séries d'observations en général d'effectifs égaux. Puis on prend les points médians de chaque sous-série.

1.12.4 Ajustements tendanciels non linéaires

Lorsque la tendance n'est pas linéaire, une technique simple consiste à se ramener à un ajustement linéaire après un changement de variable approprié.

Exemple 1.12.1 Si $Z_t = at^2 + b$, en posant $Y_t = t^2$, on se ramène à $Z_t = aY_t + b$ et on peut faire un ajustement linéaire entre Y_t et Z_t .

Si $Z_t = b \exp(at)$, en posant $Y_t = \ln(Z_t)$, on se ramène à $Y_t = at + \ln(b)$ et on peut faire un ajustement linéaire entre Y_t et t .

1.12.5 Estimation non paramétrique

Dans certaines situations, il n'est pas facile de trouver le degré du polynôme d'ajustement pour Z_t ou de changement de variable adéquat. On pourrait utiliser un polynôme avec un degré élevé mais le nombre de paramètres à estimer serait important et rendrait les calculs fastidieux. Par ailleurs, on ne sait pas non plus déterminer l'allure de cette fonction.

Dans cette situation, on a recours à la théorie non paramétrique de l'estimation de la tendance qui ne suppose rien sur celle-ci a priori et on approxime la tendance par la moyenne mobile arithmétique d'ordre k .

$$Z_t^* = M_k(Z_t).$$

Alors la tendance à la date t peut être estimée par la moyenne mobile (centrée) d'ordre k à la date t .

Chapitre 2

Modélisation *ARIMA* des séries chronologiques

Les modèles *ARIMA* permettent de combiner trois types de modèle temporels : les modèles autorégressifs (*AR*), les modèles moyenne mobile (*MA*) et les modèles intégrés (*I*). Nous exposons ces modèles dans ce chapitre.

2.1 Les modèles autorégressifs, (Autoregressive)

De tels processus seront notés *AR* [7],[12],[15].

Définition 2.1.1 On dit que la série X_t suit un processus autorégressif d'ordre 1 (*AR*(1)) si on peut écrire :

$$X_t = \phi X_{t-1} + \varepsilon_t,$$

$$X_t - \phi X_{t-1} = \varepsilon_t,$$

$$(1 - \phi B) X_t = \varepsilon_t.$$

où ε_t est un bruit blanc.

On peut remarquer qu'on fait une régression de la série décalée de 1 sur la série elle-même et les résidus forment un bruit blanc.

Définition 2.1.2 On appelle processus autorégressif d'ordre p , noté *AR*(p), un processus $\{X_t\}_{t \in \mathbb{Z}}$ vérifiant une relation du type :

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t, \quad \forall t \in \mathbb{Z}.$$

où les ϕ_i sont des réels, $\phi_p \neq 0$ et $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ_ε^2 .

cette relation équivalent à

$$(1 - \phi_1 B - \dots - \phi_p B^p) X_t = \varepsilon_t,$$

$$\iff \Phi(B) X_t = \varepsilon_t.$$

2.1.1 Représentation stationnaire

Ce processus est pour l'instant défini sous forme implicite et en particulier il n'est pas certain que cette dernière équation admette toujours une solution stationnaire.

Si le polynôme Φ a toutes ses racines de module différent de 1, on peut inverser l'opérateur $\Phi(B)$. On en déduit que l'équation admet une solution unique, avec l'écriture :

$$X_t = \Phi(B)^{-1} \varepsilon_t = \sum_{i=-\infty}^{+\infty} h_i \varepsilon_{t-i}.$$

On peut alors montrer que l'on a $\sum_{i=-\infty}^{+\infty} |h_i| < +\infty$ et donc que la représentation est stationnaire.

2.1.2 Représentation inversible

Définition 2.1.3 Si $(\xi_t)_t$ peut être exprimé en fonction de $\{X_s, s \leq t\}$ seulement, i.e. [7]

$$\xi_t = \sum_{j=0}^{+\infty} \varphi_j X_{t-j},$$

alors $(X_t)_t$ est inversible.

La représentation $AR(p)$ est inversible par définition.

2.1.3 Représentation causale

Définition 2.1.4 Si $(X_t)_t$ peut être exprimé en fonction de $\{\xi_s, s \leq t\}$ seulement, i.e. [7]

$$X_t = \sum_{i=0}^{+\infty} \phi_i \xi_{t-i},$$

alors $(X_t)_t$ est causal.

Si le polynôme Φ a toutes ses racines de module strictement supérieur à 1, l'opérateur inverse $\Phi(B)^{-1}$ admet un développement ne faisant intervenir que les puissances positives de B . On a alors :

$$X_t = \sum_{i=0}^{+\infty} h_i \varepsilon_{t-i}.$$

Dans ce cas, on montre que l'on a

$$\sum_{i=0}^{+\infty} |h_i| < +\infty, \quad h_0 = 1$$

Dans ce cas, X_{t-1}, X_{t-2}, \dots sont fonctions linéaires de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$

Exemple 2.1.1

$$X_t = \phi X_{t-1} + \varepsilon_t \quad \text{avec} \quad \varepsilon_t \sim BB(0, \sigma_\varepsilon^2) \quad \text{"modèle AR(1)"}$$

La racine de $\Phi(Z) = 1 - \phi Z$ est $Z = 1/\phi$.

• Si $|Z| = 1$ et donc $|\phi| = 1$, par exemple $\phi = 1$ alors :

$$X_t = X_{t-1} + \varepsilon_t = \dots = X_{t-n} + \varepsilon_{t-n+1} + \dots + \varepsilon_{t-1} + \varepsilon_t,$$

et la variance dépend de t donc il n'existe pas de solution stationnaire.

• Si $|Z| > 1$ et donc $|\phi| < 1$ alors :

$$X_t = (1 - \phi B)^{-1} \varepsilon_t = \sum_{i=0}^{+\infty} \phi^i B^i \varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty} \phi^i \varepsilon_{t-i}.$$

La représentation est alors causale, en plus d'être stationnaire et inversible, elle est donc canonique, ε_t est le processus d'innovation de X_t , puisque le passé de X_t dépend du passé de ε_t .

• Si $|Z| < 1$ et donc $|\phi| > 1$ alors :

$$X_t = (1 - \phi B)^{-1} \varepsilon_t = - \sum_{i=1}^{+\infty} \phi^{-i} B^{-i} \varepsilon_t = - \left(\frac{\varepsilon_{t+1}}{\phi} + \frac{\varepsilon_{t+2}}{\phi^2} + \dots \right).$$

Cette forme n'est pas tournée vers le passé et la représentation n'est donc pas canonique.

Conclusion 2.1.2 La représentation $\Phi(B) X_t = \varepsilon_t$ est canonique si les racines de $\Phi(B)$ sont supérieures à 1 en module.

2.1.4 Auto-covariance, auto-corrélations et équivalence de Yule-Walker

On considère le cas $\Phi(B)X_t = \varepsilon_t$ avec $E(X_t) = 0$ [21].

L'auto-covariance

$$\gamma(h) = \text{Cov}(X_t, X_{t-h}) = E(X_t X_{t-h}) \text{ pour } h \geq 0.$$

On a

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t.$$

Donc

$$X_t^2 = \phi_1 X_t X_{t-1} + \dots + \phi_p X_t X_{t-p} + X_t \varepsilon_t.$$

Alors

$$\gamma(0) = \phi_1 \gamma(1) + \dots + \phi_p \gamma(p) + E(X_t \varepsilon_t).$$

Or

$$E(X_t \varepsilon_t) = E[(\phi_1 X_{t-1} + \dots + \phi_p X_{t-p}) \varepsilon_t] + E(\varepsilon_t^2).$$

D'où

$$\gamma(0) = \phi_1 \gamma(1) + \dots + \phi_p \gamma(p) + \sigma_\varepsilon^2.$$

Si $h > 0$, on procède de la même façon :

$$X_t X_{t-h} = \phi_1 X_{t-1} X_{t-h} + \dots + \phi_p X_{t-p} X_{t-h} + \varepsilon_t X_{t-h}.$$

Donc

$$\gamma(h) = \phi_1 \gamma(h-1) + \dots + \phi_p \gamma(h-p) + \underbrace{E(\varepsilon_t X_{t-h})}_{=0 \text{ car } \varepsilon_t \perp X_{t-h}}.$$

Les auto-corrélations

A partir de la relation de récurrence de $\gamma(h)$ on déduit celle sur

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

$$\rho(h) = \phi_1 \rho(h-1) + \dots + \phi_p \rho(h-p), \quad \forall h \geq 0.$$

Ces dernières équations sont appelées **équations de Yule-Walker**.

Pour $h > 0$, les $\gamma(h)$ et les $\rho(h)$ vérifient une relation de récurrence d'ordre p et

$$1 = \phi_1 \rho(1) + \dots + \phi_p \rho(p) + \frac{\sigma_\varepsilon^2}{\gamma(0)},$$

donc

$$\gamma(0) = \sigma_\varepsilon^2 \frac{1}{1 - (\phi_1 \rho(1) + \dots + \phi_p \rho(p))}.$$

Les équations de Yule-Walker pour $h = 1, \dots, p$ peuvent s'écrire :

$$\begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ \rho(p-1) & \dots & \dots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \vdots \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \vdots \\ \vdots \\ \vdots \\ \rho(p) \end{pmatrix}$$

Les solutions de l'équation de récurrence sont complètement déterminées par la donnée de conditions initiales $\rho(1), \dots, \rho(p)$: elles permettent d'obtenir ϕ_1, \dots, ϕ_p .

$$\Leftrightarrow \begin{cases} \rho(1) = \phi_1 + \phi_2 \rho(1) + \dots + \phi_p \rho(p-1) \\ \vdots \\ \rho(p) = \phi_1 \rho(p-1) + \dots + \phi_{p-1} \rho(1) + \phi_p \\ \phi_1 = (1 - \phi_2) \rho(1) - \dots - \phi_p \rho(p-1) \\ \vdots \\ \phi_p = \rho(p) - \phi_1 \rho(p-1) - \dots - \phi_{p-1} \rho(1) \end{cases}$$

On peut donc aussi obtenir $\rho(1), \dots, \rho(p)$ en fonction de ϕ_1, \dots, ϕ_p .

Auto-corrélations partielle

Dans un processus auto-régressif d'ordre p ; il est possible de montrer que

$$\Gamma(h) = \begin{cases} 1 & \text{si } h = 0, \\ \rho(1) & \text{si } h = 1, \\ \phi_p & \text{si } h = p, \\ 0 & \text{si } h > p. \end{cases}$$

$\hat{\rho} = (\hat{\rho}(i))_{i=1, \dots, p}$ et $\hat{\phi}$ le vecteur des paramètres

$$\hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \cdot \\ \cdot \\ \hat{\phi}_p \end{pmatrix}.$$

2.1.5 Prédiction dans les modèles autorégressifs

Dans ce paragraphe, on suppose que $\{X_t\}$ est un processus stationnaire qui suit un modèle $AR(p)$, c'est à dire vérifie [6] :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

avec ε_t un bruit blanc.

Objectif : on cherche à prédire la valeur prise par le processus aux instants $T+1, T+2, \dots$ à partir de la connaissance des valeurs prises par ce processus jusqu'à l'instant T , c'est à dire de x_1, \dots, x_T .

Le modèle se réécrit sous la forme

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

La prévision à la date $T+1$ est

$$\hat{X}_{T+1} = E(X_{T+1}/X_T, X_{T-1}, \dots, X_1).$$

Alors

$$\hat{X}_{T+1} = \phi_1 X_T + \dots + \phi_p X_{T-p+1}.$$

Pour prédire X_{T+2} à partir de X_1, \dots, X_T on a

$$\hat{X}_{T+2} = \phi_1 \hat{X}_{T+1} + \dots + \phi_p X_{T-p+2}.$$

Et de façon générale

$$\hat{X}_{T+h} = \phi_1 \hat{X}_{T+1} + \dots + \phi_p X_{T+h-p}.$$

Remarque 2.1.1 Dans le cas des modèles d'ordre 1, on a $\hat{X}_{T+1} = \phi_1 X_T$ et $\hat{X}_{T+2} = \phi_1 \hat{X}_{T+1} = \phi_1^2 X_T, \dots$. On vérifie aisément par récurrence que

$$\hat{X}_{T+h} = \phi_1^h X_T,$$

donc en particulier que $\hat{X}_{T+h} \rightarrow 0$ quand h tend vers l'infini.

2.2 Les modèles moyennes mobiles, MA (Moving Average)

Définition 2.2.1 On appelle processus moyenne mobile d'ordre q , noté $MA(q)$ pour Moving Average, un processus $\{X_t\}_{t \in \mathbb{Z}}$ défini par [7],[12] :

$$X_t = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \forall t \in \mathbb{Z}.$$

où les θ_i sont des réels, $\theta_q \neq 0$ et $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ est un processus bruit blanc de variance σ_ε^2 .

cette relation équivalent à

$$\begin{aligned} X_t &= (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \\ \iff X_t &= \Theta(B) \varepsilon_t. \end{aligned}$$

2.2.1 Représentation stationnaire

La définition d'un $MA(q)$ est explicite et ne pose donc pas de problème : le processus X_t est parfaitement défini et est automatiquement stationnaire.

2.2.2 Représentation causale

La représentation est causale par définition.

2.2.3 Représentation inversible

$\Theta(B)$ est un polynôme en B de degré q , que l'on peut factoriser en ayant calculé ses racines $Z_i = 1/\lambda_i$, $i = 1, \dots, q$:

$$\begin{aligned} \Theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q \\ &= (1 - \lambda_1 B) (1 - \lambda_2 B) \dots (1 - \lambda_q B). \end{aligned}$$

Si $\Theta(Z)$ n'a pas de racine de module égal à 1, on peut calculer l'inverse de $\Theta(B)$, qui est alors donnée par :

$$\begin{aligned} \Theta(B)^{-1} &= (1 - \lambda_1 B)^{-1} (1 - \lambda_2 B)^{-1} \dots (1 - \lambda_q B)^{-1} \\ &= \frac{k_1}{1 - \lambda_1 B} + \dots + \frac{k_q}{1 - \lambda_q B}. \end{aligned}$$

si toutes les racines de Θ (c'est-à-dire $1/\lambda_i$, $i = 1, \dots, q$) sont distinctes et où k_1, \dots, k_q sont des paramètres qui dépendent de $\lambda_1, \dots, \lambda_q$.

On obtient alors l'expression suivante :

$$\Theta(B)^{-1} X_t = \sum_{i=-\infty}^{+\infty} \pi_i X_{t-i} = \varepsilon_t$$

où les π_i sont fonctions des paramètres θ_j , et on peut montrer que $\sum_{i=-\infty}^{+\infty} |\pi_i| < +\infty$.

Si les racines de $\Theta(Z) = 0$ sont toutes de module supérieur à 1, on peut montrer que $\pi_i = 0 \forall i < 0$. On dit que le processus est inversible.

2.2.4 Structure d'autocorrélation

Comme $MA(q)$ est stationnaire, on peut étudier sa structure d'autocorrélation. En effet, les autocorrélations existent toujours puisque le modèle est stationnaire [21],

$$\gamma(h) = E[X_t X_{t+h}],$$

ce qui est plus adéquat ici, car il y a non-corrélation des ε_t avec le futur (on suppose par simplicité que $\mu = 0$) :

$$\begin{aligned} X_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \\ X_{t+h} &= \varepsilon_{t+h} - \theta_1 \varepsilon_{t+h-1} - \theta_2 \varepsilon_{t+h-2} - \dots - \theta_q \varepsilon_{t+h-q}. \end{aligned}$$

Comme les ε_t sont un bruit blanc, les ε_t sont des variables aléatoires non-corrélées, autrement dit

$$\begin{cases} E[\varepsilon_{t+i} \varepsilon_{t+i}] = \sigma_\varepsilon^2, \\ E[\varepsilon_{t+i} \varepsilon_{t+j}] = 0, \quad i \neq j. \end{cases}$$

On obtient

$$\gamma(h) = \begin{cases} \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) & h = 0, \\ \sigma_\varepsilon^2 (-\theta_h + \theta_1 \theta_{h+1} + \dots + \theta_q \theta_{q+h}) & 1 \leq h \leq q, \\ 0 & h > q. \end{cases}$$

La fonction d'autocorrélation vaut donc

$$\rho(h) = \begin{cases} 1 & h = 0, \\ \frac{-\theta_h + \theta_1\theta_{h+1} + \dots + \theta_{q+h}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & 1 \leq h \leq q, \\ 0 & h > q. \end{cases}$$

2.2.5 Prévisions à l'aide d'un modèle MA(q)

On supposera là aussi que l'on s'est ramené à un processus centré (X_t) , satisfaisant [5] :

$$\begin{aligned} X_t &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q} \\ &= \Theta(B)\varepsilon_t. \end{aligned}$$

La prévision optimale à la date $T + 1$, faite à la date T est

$$\widehat{X}_{T+1} = E(X_{T+1}/X_T, X_{T-1}, \dots, X_1).$$

Donc

$$\widehat{X}_{T+1} = 0 + \theta_1\varepsilon_T + \dots + \theta_q\varepsilon_{T+1-q}.$$

De façon analogue, X_{T+h} est estimé par

$$\widehat{X}_{T+h} = E(X_{T+h}/X_T, X_{T-1}, \dots, X_1),$$

et donc

$$\widehat{X}_{T+h} = \begin{cases} \theta_h\widehat{X}_{T+1} + \dots + \theta_q X_{T+h-q} & \text{pour } h \leq q, \\ 0 & \text{pour } h > q. \end{cases}$$

Toutefois, cette méthode présente le désavantage d'estimer X_{T+h} à partir des résidus passés, a priori non observables, et non pas du passé de la variable.

2.2.6 Dualité des processus $AR(p)$ et $MA(q)$

Processus	$AR(p)$	$MA(q)$
Modèle	$\Phi(B) X_t = \varepsilon_t$	$X_t = \Theta(B) \varepsilon_t$
Présentation en terme de bruits blancs	$X_t = \Phi^{-1}(B) \varepsilon_t$	$X_t = \Theta(B) \varepsilon_t$
Présentation en terme de valeurs passées	$\Phi(B) X_t = \varepsilon_t$	$\varepsilon_t = \Theta^{-1}(B) X_t$
Nombre de coeff. ψ_j en terme de bruit blanc	∞	q
Nombre de coeff. π_i en terme de valeurs passées	p	∞
Condition d'inversibilité	toujours	$\ \Theta(B)\ > 1$
Condition de la stationnarité	$\ \Phi(B)\ > 1$	toujours
Nombre de $\rho(h)$ non-nuls	∞	$\rho(h) = 0, h > q$
Nombre de $\Gamma(h)$ non-nuls	$\Gamma(h) = 0, h > p$	∞

TAB. 2.1 – Table Dualité des AR(p) et MA(q)

2.3 Les modèles $ARMA(p, q)$ (Autorégressive Moving Average)

Les modèles $ARMA$ sont représentatifs de processus générés par une combinaison des valeurs passées et des erreurs passées [15].

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad t \in \mathbb{Z}.$$

On peut aussi écrire le modèle $ARMA(p, q)$ sous la forme :

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

où B est l'opérateur de décalage et (ε_t) est le processus de bruit blanc.

2.3.1 Inversibilité et stationnarité en terme de valeurs passées

Pour que le processus puisse s'écrire [21] :

$$\Pi(B) X_t = \varepsilon_t,$$

où

$$\Pi(B) = \Theta_q^{-1}(B) \Phi_p(B),$$

il faut que toutes les racines de $\Theta_q(z)$ soient à l'extérieur du cercle unité. Le processus est alors inversible.

Pour que le processus puisse s'écrire

$$X_t = \Psi(B)\varepsilon_t,$$

où

$$\Psi(B) = \Theta_q(B)\Phi_p^{-1}(B),$$

il faut que toutes les racines de $\Phi_p(z)$ soient à l'extérieur du cercle unité. Le processus est alors stationnaire.

On suppose donc que $\Phi_p(z)$ et $\Theta_q(z)$ ont leurs racines à l'extérieur du cercle unité de manière à avoir la stationnarité et l'inversibilité.

2.3.2 Auto-corrélation

La covariance vaut

$$\begin{aligned}\gamma(h) &= E[X_t X_{t-h}] \\ &= E\left[\left(\sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}\right) X_{t-h}\right].\end{aligned}$$

Si on note

$$\gamma_{X_\varepsilon}(h) = E[\varepsilon_t X_{t-h}].$$

la covariance croisée entre X_{t-h} et ε_t . On a

$$\begin{aligned}\gamma(h) &= \sum_{j=1}^p \phi_j E[X_{t-j} X_{t-h}] - \sum_{i=1}^q \theta_i E[\varepsilon_{t-i} X_{t-h}] + \gamma_{X_\varepsilon}(h) \\ &= \sum_{j=1}^p \phi_j \gamma(h-j) - \sum_{i=1}^q \theta_i \gamma_{X_\varepsilon}(h-i) + \gamma_{X_\varepsilon}(h), h \geq 0,\end{aligned}$$

or

$$E[X_t \varepsilon_t] = E\left[\left(\sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}\right) \varepsilon_t\right] = E[\varepsilon_t \varepsilon_t] = \sigma_\varepsilon^2,$$

et ε_t n'est pas corrélé avec le passé.

$$\begin{cases} \gamma_{X_\varepsilon}(h) = 0 & h > 0, \\ \gamma_{X_\varepsilon}(0) = \sigma_\varepsilon^2, \\ \gamma_{X_\varepsilon}(h) \neq 0 & h < 0. \end{cases}$$

La fonction $\gamma_{X_\varepsilon}(h)$ n'est pas paire.

Si $h > q$, tous les $\gamma_{X_\varepsilon}(h) = 0$, et on a

$$\begin{cases} \gamma(h) = \sum_{j=1}^p \phi_j \gamma(h-j), \\ \rho(h) = \sum_{j=1}^p \phi_j \rho(h-j). \end{cases}$$

Ce qui est un système d'équation de type Yule-Walker. En effet, la partie MA ne joue que jusqu'au délai q , après elle ne joue plus. Il est donc normal de retrouver des équations ressemblant à celles trouvées dans les $AR(p)$ à partir de $h > q$.

Pour un $ARMA(p, q)$ la structure d'auto-corrélation ne suit pas un schéma connu jusqu'au délai q mais ensuite le comportement est le même que celui d'un $AR(p)$.

Il y a donc une infinité de $\gamma(h)$ non-nuls, mais à partir de $h > q$; la fonction d'auto-corrélation a le même comportement que pour $AR(p)$.

2.3.3 Auto-corrélations partielles

Ici la partie, AR ne se fait sentir que jusqu'à l'ordre p . Pour $h \leq p$, il n'y a pas de schéma bien précis. Pour $h > p$, on retrouve le même comportement que pour un $MA(q)$.

2.3.4 Prévisions à l'aide d'un modèle ARMA(p,q)

On supposera là aussi que l'on s'est ramené à un processus centré (X_t) , satisfaisant [6] :

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

Sous cette forme $ARMA$, alors

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

et donc

$$X_{t+h} = \sum_{j=1}^p \phi_j X_{t+h-j} + \varepsilon_{t+h} + \sum_{i=1}^q \theta_i \varepsilon_{t+h-i}.$$

On a alors

$$\hat{X}_{T+h} = E(X_{T+h} / X_T, X_{T-1}, \dots, X_1).$$

On peut noter que pour $h > q$

$$\widehat{X}_{T+h} = \begin{cases} \phi_1 \widehat{X}_{T+h-1} + \dots + \phi_{h-1} \widehat{X}_{T+1} + \dots + \phi_h X_T + \dots + \phi_p X_{T+h-p} & h \leq p, \\ \phi_1 \widehat{X}_{T+h-1} + \dots + \phi_p \widehat{X}_{T+h-p} & h > p. \end{cases}$$

2.4 Les modèles $ARIMA(p, d, q)$

Une série (X_t) suit un processus $ARIMA$ (Auto Regressive Integrated Moving Average) d'ordre (p, d, q) si elle écrit sous la forme [15] :

$$\Psi(B) X_t = \Theta(B) \varepsilon_t.$$

On modélise alors la série sous la forme :

$$\begin{aligned} \Phi(B) (1 - B)^d X_t &= \Theta(B) \varepsilon_t, \\ \Phi(B) \Delta^d X_t &= \Theta(B) \varepsilon_t. \end{aligned}$$

où le polynôme $\Phi(B)$ est de degré p et le polynôme $\Theta(B)$ est de degré q . On écrit que la série X_t suit un processus $ARIMA(p, d, q)$.

2.4.1 Prévisions dans le cas d'un processus $ARIMA(p, d, q)$

On considérons ici (X_t) satisfaisant une équation de la forme [5] :

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

Posons alors $\Psi(B) = \Phi(B) (1 - B)^d$. La forme $ARIMA(p, d, q)$ peut s'écrire

$$X_t = \sum_{j=1}^{p+d} \psi_j X_{t-j} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

et donc

$$X_{t+h} = \sum_{j=1}^{p+q} \psi_j X_{t+h-j} + \varepsilon_{t+h} + \sum_{i=1}^q \theta_i \varepsilon_{t+h-i}.$$

Notons \widehat{X}_{T+h} la prévision faite à la date $T + h$

$$\widehat{X}_{T+h} = E(X_{T+h} / X_T, X_{T-1}, \dots, X_1).$$

Alors

$$\widehat{X}_{T+h} = \sum_{j=1}^{p+d} \psi_j \widehat{X}_{T+h-j} + 0 + \sum_{i=1}^q \theta_i \widehat{\varepsilon}_{T+h-i},$$

où

$$\widehat{X}_{T+h-j} = X_{T+h-j} \text{ pour } j \geq h,$$

et

$$\widehat{\varepsilon}_{T+h-i} = \begin{cases} 0 & \text{pour } i < h, \\ \varepsilon_{T+h-i} & \text{pour } i \geq h. \end{cases}$$

En particulier, pour $h \geq q$, on obtient une relation de récurrence de la forme

$$\widehat{X}_{T+h} = \sum_{j=1}^{p+d} \psi_j \widehat{X}_{T+h-j}.$$

2.5 Méthode de Box & Jenkins

Box & Jenkins en 1970 ont élaboré une méthodologie pour identifier un modèle adéquat pour une série chronologique. Leur méthode est fondée sur les modèles *ARIMA*. Dans ce section, les principales étapes de cette technique sont présentées. Pour les méthodes d'inférence présentées dans la suite, on supposera que T réalisations d'une série chronologique univariée, notées X_1, \dots, X_T ont été observées [10].

2.5.1 Identification du modèle

De façon générale, l'étape d'identification du modèle consiste à identifier le modèle qui représente au mieux la série étudiée. En d'autres mots, il s'agit de trouver un modèle stationnaire qui tient compte de la variabilité dans le temps et pour lequel il y a absence d'autocorrélation des résidus. Plus particulièrement, cette étape implique les méthodes d'estimation du paramètre d'intégration d , l'estimation des ordres p et q , les tests de non stationnarité ou de racine unitaire.

Estimation du paramètre d'intégration

Lorsqu'un processus possède une non stationnarité de type *DS*, on parle alors d'un modèle intégré. Dans ce cas, il convient de déterminer l'ordre d'intégration d du processus filtré $(1 - \beta)^d X_t$ pour lequel le processus est stationnaire, d'où le processus $I(d)$. L'ordre d'intégration " d " est le nombre de différentiation pour que la série initiale sera stationnaire.

Remarque 2.5.1 *L'objectif visé par les différenciations est de stabiliser la moyenne d'une série.*

Estimation des ordres p et q

Identification d'un processus MA Soit un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ satisfaisant une représentation $MA(q)$. Pour déterminer la valeur de q , on se base sur la fonction d'autocorrélation du processus MA . En fait, q correspondra au plus grand délai tel que l'autocorrélation n'est pas statistiquement égale à 0.

Identification d'un processus AR Soit un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ satisfaisant une représentation $AR(p)$. Globalement, l'identification d'un processus $AR(p)$ s'effectue de la même façon que celle d'un processus $MA(q)$. La seule différence réside dans le fait que c'est l'autocorrélation partielle, plutôt que l'autocorrélation, qui est utilisée.

Identification d'un processus $ARMA$ Il existe plusieurs méthodes pour déterminer les ordres p et q d'un processus $ARMA$. On peut également se baser sur les autocorrélations et les autocorrélations partielles. Pour obtenir l'ordre de la composante MA , il faut identifier l'autocorrélation significative dont l'ordre est le plus élevé; pour la composante AR , il faut identifier l'autocorrélation partielle significative dont l'ordre est le plus élevé.

Tests de non stationnarité

Il existe plusieurs tests de non stationnarité ou de racine unitaire, ces tests ayant pour objet de déterminer la présence de non stationnarité. Parmi les tests de non stationnarité les plus populaires, on a ceux de Dickey & Fuller, Dickey Fuller augmenté et de Phillips & Perron.

2.5.2 Estimation des paramètres**Maximum de vraisemblance**

Une méthode populaire pour estimer les paramètres d'un modèle est le maximum de vraisemblance. La fonction de vraisemblance associée à un échantillon X_1, \dots, X_T i.i.d. d'une loi dont la densité est $f(x|\theta)$, avec $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$, est définie par

$$L(\theta) = \prod_{t=1}^T f(X_t | \theta).$$

L'estimateur du maximum de vraisemblance (EMV) est la valeur $\hat{\theta}_{EMV}$ qui maximise $L(\theta)$. Parfois, il est possible de déduire cet estimateur en dérivant $L(\theta)$

par rapport à chacun des paramètres $(\theta_1, \dots, \theta_k)$ et de résoudre le système à k équations

$$\frac{\partial L(\theta)}{\partial \theta_j} = 0, \text{ où } j = 1, \dots, k.$$

2.5.3 Adéquation et validation du modèle

Cette étape consiste à s'assurer que le modèle retenu décrit bien la série d'observations. Il s'agit de vérifier les conditions d'un processus bruit blanc à l'aide de tests appropriés et que les paramètres estimés sont statistiquement différents de 0.

Test de significativité des coefficients

Les coefficients du modèle doivent être significativement différent de zéro, pour ce faire on utilise le test classique de student.

Test sur la moyenne des résidus

Le test sur la moyenne des résidus fait référence à la première condition d'un processus bruit blanc, à savoir une moyenne égale à 0. Ainsi, pour vérifier que les résidus d'un certain modèle sont de moyenne nulle, comme il se doit, il s'agit de confronter les hypothèses

$$\begin{cases} H_0 : E(\varepsilon_t) = 0, \\ H_1 : E(\varepsilon_t) \neq 0. \end{cases}$$

Une statistique de test est définie par

$$\xi = \frac{\bar{\hat{\varepsilon}}}{S_\varepsilon / \sqrt{T}},$$

où

$$\bar{\hat{\varepsilon}} = \frac{1}{T} \sum_{t=1}^T \varepsilon_t \text{ et } S_\varepsilon = \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2.$$

La statistique ξ est approximativement distribuée selon la loi normale centrée réduite. Ainsi, on rejette H_0 si $|\xi| > z_{\alpha/2}$, où $z_{\alpha/2}$ est la fractile d'ordre $\alpha/2$ de la distribution normale centrée réduite.

Test sur la fonction d'autocorrélation

Le test de Box Pierce permet d'identifier les processus de bruit blanc. Ce test est

$$\begin{cases} H_0 : \rho(1) = \rho(2) = \dots = \rho(h) = 0, \\ H_1 : \text{il existe au moins un } \rho(i) \text{ significativement différent de } 0. \end{cases}$$

Pour effectuer ce test, on recourt à la statistique Q qui est donnée par

$$Q = T \sum_{k=1}^h \hat{\rho}^2(k).$$

h : nombre de retards, T : nombre d'observations, $\hat{\rho}(k)$: autocorrélation empirique d'ordre k .

La statistique Q est distribuée de manière asymptotique comme un χ^2 à h degrés de liberté. Nous rejetons donc l'hypothèse de bruit blanc, au seuil α , si la statistique Q est supérieure au χ^2 lu dans la table au seuil $(1 - \alpha)$ et h degrés de liberté.

Test de normalité des résidus

Plusieurs des modèles de séries chronologiques supposent que les résidus sont indépendants et distribués selon la loi Normale. Un des tests permettant de vérifier la normalité des résidus est celui de Jarque & Bera. Les hypothèses à confronter sont

$$\begin{cases} H_0 \sim N(0, 1), \\ H_1 \approx N(0, 1). \end{cases}$$

Avant de décrire le test pour les résidus d'un modèle de séries chronologiques, prenons le cas de n observations X_1, \dots, X_n indépendantes et de même loi. Dans ce cas, la statistique du test de Jarque & Bera, de loi asymptotique khi-carré à deux degrés de liberté, est définie par

$$JB = \frac{n}{6} \beta_1^2 + \frac{n}{24} (\beta_2 - 3)^2,$$

où

$$\beta_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S} \right)^3 \quad \text{et} \quad \beta_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S} \right)^4,$$

sont respectivement les coefficients d'asymétrie et d'aplatissement. Ici, \bar{X} et S sont respectivement la moyenne et l'écart type empiriques. Sous l'hypothèse de normalité, on peut montrer que

$$\frac{\sqrt{n}\beta_1}{\sqrt{6}} \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{et} \quad \frac{\sqrt{n}(\beta_2 - 3)}{\sqrt{24}} \xrightarrow{\mathcal{L}} N(0, 1),$$

et que

$$\frac{n\beta_1^2}{6} \xrightarrow{\mathcal{L}} \chi_1^2 \quad \text{et} \quad \frac{n(\beta_2 - 3)^2}{24} \xrightarrow{\mathcal{L}} \chi_1^2,$$

où χ_ν^2 représente la loi khi-carré à ν degrés de liberté. Ainsi, la loi asymptotique de JB est la distribution khi-carré à deux degrés de liberté. On rejette donc l'hypothèse de normalité, si $JB > \chi_{1-\alpha;2}^2$.

2.5.4 Prévision

Une fois qu'on a trouvé un modèle acceptable pour la série chronologique étudiée, il est possible de calculer les prévisions. On note \hat{X}_{T+h} la prévision de X_{T+h} au temps $T+h$ où h l'horizon de la prévision.

La prévision est calculée par la formule suivante :

$$\hat{X}_{T+h} = E(X_{T+h} / X_T, X_{T-1}, \dots, X_1).$$

Chapitre 3

Application

Le tableau suivant présente l'évolution du nombre des étudiants de l'université de ouargla depuis 1987 jusqu'à 2015.

Année	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
N des étudiants	139	232	255	624	1211	1772	2001	2354	3014	4016
Année	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
N des étudiants	5703	6018	6562	7409	10048	11523	13450	15430	17125	18287
Année	2007	2008	2009	2010	2011	2012	2013	2014	2015	
N des étudiants	20370	21199	21539	21794	21357	22015	22298	23354	25129	

TAB. 3.1 – L'évolution du nombre des étudiants de l'université de ouargla depuis 1987 jusqu'à 2015

3.1 Identification de la série

Pour identifier le type de chronique, nous allons procéder à des tests formels et informels afin de connaître la tendance et la stationnarité.

3.1.1 Les tests informels

L'analyse graphique de la série

L'analyse graphique nous permet de visualiser l'évolution temporelle de la série du nombre d'étudiant dans l'université de ouargla (1987-2015), Ainsi, à partir des données de la série, nous obtenons le graphique suivant :

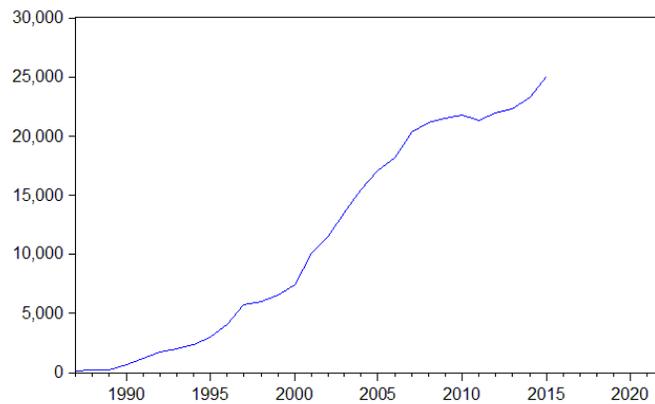


FIG. 3.1 – Graphique de l'évolution du nombre des étudiants de l'université de ouargla 1987- 2015.

De ce graphique, nous observons que la série évolue de manière croissante dans le temps. Ce qui nous permet de tirer quelques conclusions intéressantes au vu de l'allure de la courbe ci-dessus :

- La présomption d'une tendance de la série.
- La série n'est pas stationnaire.

Corrélogramme de la série initiale :

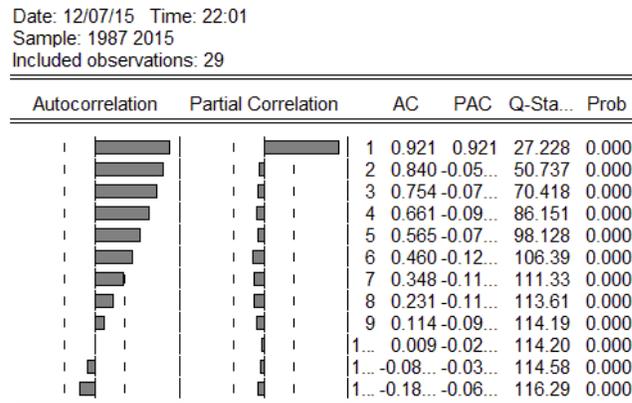


FIG. 3.2 – Corrélogramme de la série.

Les coefficients d'autocorrélation simples décroissent lentement, de manière linéaire, ce qui appuie notre affirmation sur la non stationnarité de la série.

Nous observons aussi que le premier coefficient d'autocorrélation partielle est significatif.

Seulement ces indicateurs sont nécessaires mais non-suffisants.

Par conséquent, nous utilisons le test de Dickey-Fuller Augmenté (ADF) qui est plus performant dans l'étude de la stationnarité des séries.

3.1.2 Tests formels : Test de racine unitaire (ADF)

Ce test a un double objectif :

- Il permet de vérifier la stationnarité d'une série.
- Il donne une idée sur la structure de la série.

Nous avançons les hypothèses suivantes pour notre test :

$$\left\{ \begin{array}{l} H_0 : \rho = 1, \text{ présence de la racine unitaire c\`ad la s\`erie est non stationnaire;} \\ H_1 : \rho \neq 1, \text{ absence de la racine unitaire c\`ad la s\`erie est stationnaire.} \end{array} \right.$$

Ainsi, nous rencontrons les cas de figures ci-apr\`es :

-Si $|t_ADF| < |VCM|$, on accepte l'hypoth\`ese nulle, la s\`erie est non stationnaire.

-Si $|t_ADF| > |VCM|$, on rejette l'hypoth\`ese nulle, la s\`erie est stationnaire. (Avec VCM =Valeur Critique de MacKinnon au seuil de 5%).

• Significativité du Trend

$$\left\{ \begin{array}{l} H_0 : b = 0, \text{ le trend est non significatif;} \\ H_1 : b \neq 0, \text{ le trend est significatif.} \end{array} \right.$$

Ainsi, nous rencontrons les cas de figures ci-apr\`es :

-Si $|t - stat| < 2$, on accepte l'hypoth\`ese nulle ;

-Si $|t - stat| > 2$, on rejette l'hypoth\`ese nulle.

• Significativité de l'intercept

$$\left\{ \begin{array}{l} H_0 : c = 0, \text{ la constante est non significatif; mod\`ele sans d\`erive;} \\ H_1 : c \neq 0, \text{ la constante est significatif; mod\`ele avec d\`erive.} \end{array} \right.$$

Ainsi, nous rencontrons les cas de figures ci-apr\`es :

-Si $|t - stat| < 2$, on accepte l'hypoth\`ese nulle.

-Si $|t - stat| > 2$, on rejette l'hypoth\`ese nulle.

Null Hypothesis: J has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 2 (Automatic - based on SIC, maxlag=6)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.495733	0.3272
Test critical values: 1% level	-4.356068	
5% level	-3.595026	
10% level	-3.233456	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(J)
 Method: Least Squares
 Date: 12/08/15 Time: 18:14
 Sample (adjusted): 1990 2015
 Included observations: 26 after adjustments

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
J(-1)	-0.194544	0.077950	-2.495733	0.0210
D(J(-1))	0.403462	0.192805	2.092586	0.0487
D(J(-2))	0.270695	0.197621	1.369765	0.1852
C	-692.9816	529.8801	-1.307808	0.2051
@TREND("1987")	213.5708	85.46850	2.498825	0.0208
R-squared	0.439470	Mean dependent var		956.6923
Adjusted R-squared	0.332702	S.D. dependent var		741.5487
S.E. of regression	605.7586	Akaike info criterion		15.82188
Sum squared resid	7705813.	Schwarz criterion		16.06382
Log likelihood	-200.6845	Hannan-Quinn criter.		15.89155
F-statistic	4.116130	Durbin-Watson stat		2.001568
Prob(F-statistic)	0.012871			

FIG. 3.3 – Test ADF sur premier série

Il ressort de cet output du test d'ADF, que la série est non stationnaire ($|t_ADF| = 2.495733 < |VCM| = 3.595026$, le trend est significatif ($|t - stat| = 2.498825 > 2$), et l'intercept est non significatif ($|t - stat| = 1.307808 < 2$).

Ce qui nous amène à mettre fin à l'étude de la stationnarité de notre série et à rejoindre la conclusion des tests préliminaires : notre série présente donc une non stationnarité de type Déterministe (TS : Trend stationnarity).

3.2 Stationnarisation de la série

Etant donné que la série présente donc une non stationnarité de type Déterministe (TS : Trend stationnarity).

Nous allons la stationnariser par l'extraction du trend grâce à la méthode d'écart à la tendance.

L'élimination de la tendance peut être observée par l'estimation suivante :

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
@TREND	862.6701	27.58358	31.27477	0.0000
R-squared	0.926480	Mean dependent var	11249.24	
Adjusted R-squared	0.926480	S.D. dependent var	8934.838	
S.E. of regression	2422.649	Akaike info criterion	18.45698	
Sum squared resid	1.64E+08	Schwarz criterion	18.50413	
Log likelihood	-266.6263	Hannan-Quinn criter.	18.47175	
Durbin-Watson stat	0.092946			

FIG. 3.4 – Estimation de la tendance

Nous constatons que le coefficient de tendance est significatif. Alors notre variable j n'est pas encore stationnaire.

Test de Stationnarité sur la série extrait du trend

Null Hypothesis: JJ has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 2 (Automatic - based on SIC, maxlag=6)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.495733	0.3272
Test critical values: 1% level	-4.356068	
5% level	-3.595026	
10% level	-3.233456	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(JJ)
 Method: Least Squares
 Date: 12/08/15 Time: 21:05
 Sample (adjusted): 1990 2015
 Included observations: 26 after adjustments

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
JJ(-1)	-0.194544	0.077950	-2.495733	0.0210
D(JJ(-1))	0.403462	0.192805	2.092586	0.0487
D(JJ(-2))	0.270695	0.197621	1.369765	0.1852
C	-806.2500	431.5727	-1.868167	0.0758
@TREND("1987")	45.74393	23.18848	1.972700	0.0618

R-squared	0.439470	Mean dependent var	94.02223
Adjusted R-squared	0.332702	S.D. dependent var	741.5487
S.E. of regression	605.7586	Akaike info criterion	15.82188
Sum squared resid	7705813.	Schwarz criterion	16.06382
Log likelihood	-200.6845	Hannan-Quinn criter.	15.89155
F-statistic	4.116130	Durbin-Watson stat	2.001568
Prob(F-statistic)	0.012871		

FIG. 3.5 – Test ADF sur la série d'après enlève la tendance

Ainsi, nous allons stationnariser cette série au moyen de la méthode de différence première. Ce qui conduit donc à la série $d(jj)$.

La représentation graphique de la série stationnarisée est donnée sur la figure ci-dessous.

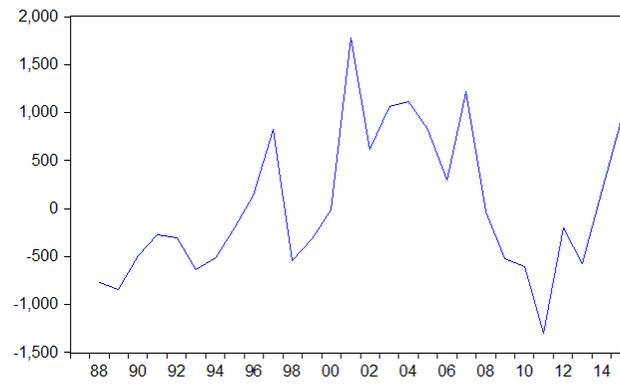


FIG. 3.6 – La série stationnaire $d(jj)$.

Identification de la série stationnaire

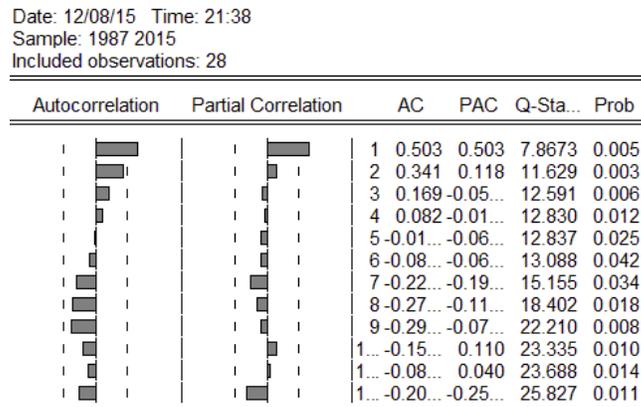


FIG. 3.7 – corrélogramme de la séries $d(jj)$.

3.3 Identification du modèle optimal provisoire

Cette phase consiste à déterminer les modèles adéquats dans la famille *ARIMA*. Elle est fondée sur l'étude des corrélogrammes simples et partiels de la série étudiée (éventuellement stationnarisée).

Nous constatons que les corrélogrammes et les corrélogrammes partiels ont leurs premier termes différent de 0. Nous pouvons donc anticiper des modèle de type *AR*(1), *MA*(1) où *ARIMA*(1, 1, 1).

Et pour pouvoir choisir un bon modèle parmi ceux présente, on construit un tableau de comparaison de leurs critères d'information et nous retenons le modèle qui minimise ces critères.

Tableau de comparaison :

	<i>AKAIKE</i>	<i>SCHWARZ</i>
<i>AR</i> (1)	15.87492	15.97007
<i>MA</i> (1)	15.98562	16.08077
<i>ARIMA</i> (1, 1, 1)	15.93376	16.07650

TAB. 3.2 – Comparaison entre les critères des modèles *AR*(1), *MA*(1), *ARIMA*(1,1,1).

D'après le tableau nous pouvons retenir le modèle *AR*(1) parce que c'est le modèle qui minimise les critères d'information de *AKAIKE* et *SCHWARZ* par rapport aux autres modèles. Alors notre modèle est

$$\widehat{W}_t = \theta_1 AR(1) + \varepsilon_t \quad \text{ou} \quad \widehat{W}_t = \theta_1 W_{t-1} + \varepsilon_t.$$

3.4 Estimation

Après avoir identifié le modèle, il convient de passer à la phase d'estimation des paramètres de ce modèle, nous avons :

Dependent Variable: DJJ
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/08/15 Time: 21:52
Sample: 1988 2015
Included observations: 28
Convergence achieved after 19 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
AR(1)	0.534039	0.209729	2.546332	0.0172
SIGMASQ	393246.0	94034.42	4.181937	0.0003
R-squared	0.277962	Mean dependent var		29.82992
Adjusted R-squared	0.250191	S.D. dependent var		751.5349
S.E. of regression	650.7655	Akaike info criterion		15.87492
Sum squared resid	11010889	Schwarz criterion		15.97007
Log likelihood	-220.2488	Hannan-Quinn criter.		15.90401
Durbin-Watson stat	2.034996			
Inverted AR Roots	.53			

FIG. 3.8 – Estimation de modèle *AR*(1).

Le modèle $AR(1)$ estimé est :

$$\widehat{W}_t = \theta_1 AR(1) + \varepsilon_t \quad \text{ou} \quad \widehat{W}_t = \theta_1 W_{t-1} + \varepsilon_t \quad \text{avec } \varepsilon_t \sim iid(0, \sigma^2)$$

$$\widehat{W}_t = 0.534039 * W_{t-1} + \varepsilon_t.$$

3.5 Validation du modèle

Le modèle ne peut être mis à défaut. Nous allons diagnostiquer notre modèle à partir des tests suivants :

3.5.1 Test de Student des paramètres (significativité statistique)

Ce test consiste à vérifier que les paramètres du modèle qui ont été estimés sont statistiquement différents de 0.

Les hypothèses du test sont :

$$\left\{ \begin{array}{l} H_0 : \theta_1 = 0, \text{ le coefficient est non significatif;} \\ H_1 : \theta_1 \neq 0, \text{ le coefficient est significatif.} \end{array} \right.$$

Ainsi, au risque de 5%, le paramètre du modèle est statistiquement différent de Zéro car $|t - stat| = 2.546332 > 2$.

3.5.2 Test de bruit blanc

Une des premières validations du modèle consiste à vérifier que le résidu issu de l'estimation est un bruit blanc. Cela équivaut à faire le Test de bruit blanc normal.

Les hypothèses de ce test sont :

$$\left\{ \begin{array}{l} H_0 : \varphi_1 = \varphi_2 = \dots = \varphi_p = 0, \text{ la série des résidus est un bruit blanc;} \\ H_1 : \text{Au moins un } \varphi_i \neq 0, \text{ la série des résidus n'est pas un bruit blanc.} \end{array} \right.$$

Il est basé sur l'observation des corrélogrammes et sur le test de BOX-PIERCE & LJUNG-BOX. Ainsi en visualisant les corrélogrammes du résidu, nous avons :

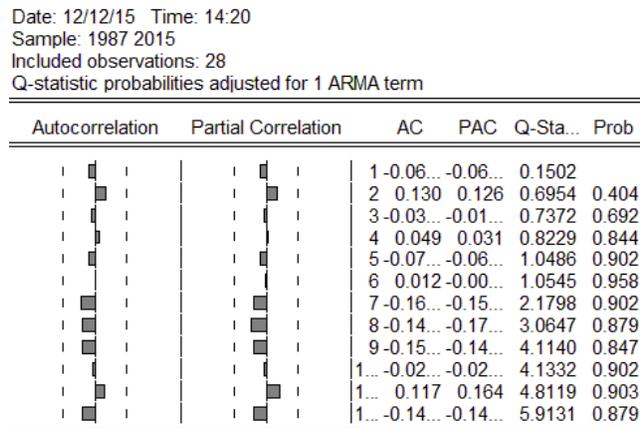


FIG. 3.9 – corrélogramme des résidus du modèle AR(1).

Les hypothèses du test sont :

$$\left\{ \begin{array}{l} H_0 : \text{les résidus sont des bruits blancs (prob pour le retards } 12 > 5\%); \\ H_1 : \text{les résidus ne sont pas des bruits blancs (prob pour le retards } 12 < 5\%). \end{array} \right.$$

Les Prob (Q-stat) n'étant pas inférieures à 5%, nous acceptons l'hypothèse nulle. Les résidus du modèle estimé sont des bruits blancs. Par conséquent, le modèle $AR(1)$ estimé est celui qui a généré notre variable. Il est donc opportun de choisir un modèle AR dont le décalage est égal à 1.

3.5.3 Test de normalité des résidus

Le test que nous utilisons est celui de Jarque et Bera. Il stipule que les erreurs suivent une distribution normale de moyenne nulle et de variance constante.

Ce test se formule avec les hypothèses suivantes :

$$\left\{ \begin{array}{l} H_0 : \text{les résidus sont normalement distribués;} \\ H_1 : \text{les résidus ne sont pas normalement distribués.} \end{array} \right.$$

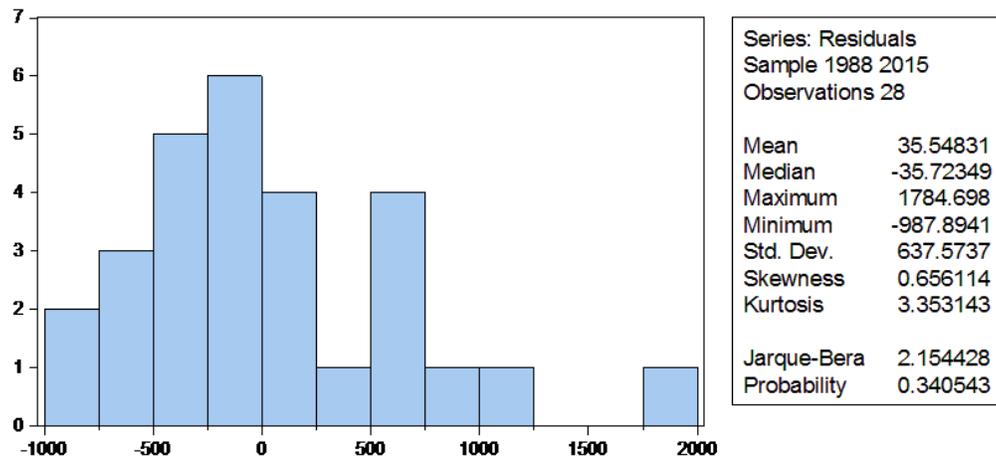


FIG. 3.10 – Normalité des résidus du modèle AR(1)

Les résidus de ce modèle sont normalement distribués. Ils suivent un processus bruit blanc gaussien.

3.6 La prévision

D'après le programme Eviews, les prévisions de cinq années sont dans le tableau suivant :

année	N-étudiant	IC (95%)
2016	26179	[23951, 29556]
2017	28186	[25931, 30536]
2018	29348	[28673, 31041]
2019	30823	[29850, 33000]
2020	32728	[31023, 34052]

TAB. 3.3 – Prévision du nombre des étudiants.

Conclusion

Depuis les années soixante-dix les techniques de modélisation des séries chronologiques ont beaucoup évolué et se sont perfectionnées. Parmi les plus connues et utilisées s'impose la méthodologie de Box-Jenkins. Elle s'avère être un outil puissant et performant pour faire de la prévision.

Le but de notre étude est de trouver le meilleur modèle pour prédire le nombre des étudiants à l'Université de Ouargla (1987-2015), les résultats de l'analyse de notre série montre que le modèle approprié pour cette étude est AR(1). Finalement, nous obtenons les prévisions du nombre des étudiants pour cinq années.

Bibliographie

- [1] Bourbonnais, R. Terraza, M. (2004). Analyse des séries Temporelles. Dunod, Paris.
- [2] Bourbonnais, R. (2005). Econométrie, 6^{ème} Edition. Dunod, Paris.
- [3] Dormont, B. (1999). Introduction à l'économétrie. Montchretien.
- [4] Johnston, J. et Dinardo, J. (1999). Méthodes de économétriques, 4^{ème} Edition. Economica .
- [5] Charpentier, A. (2012). Modèles de prévision Séries temporelles.
- [6] Charpentier, A. (2004). Cours des séries temporelles. Théorie et Applications. Université de Paris.
- [7] Corinne, P. (2005). Séries chronologiques : Quelques éléments du cours. Paris.
- [8] Dégerine, S. (2005). Cours de série chronologique. Université Joseph Fourier.
- [9] Florenece, N. (2006). Généralités sur les séries chronologiques.
- [10] Girard, Y. (2011). Série chronologiques à une et plusieurs variables : synthèse des méthodes classiques et modèles à base de copules. Université du Québec.
- [11] Gouriéroux, C. et Monfort, A. (1990). Séries Temporelles et Modèles Dynamiques. Economica.
- [12] Lévy, C.L. (2013). Introduction à l'étude des séries temporelles.
- [13] Mechgoug, R. (2013). La Prédiction des Séries Temporelles utilisant les Paradigmes de Soft Computing. Algérie.
- [14] Marine, E. (2015). Centre de Formation aux Métier de la Montagne.
- [15] Monbet, V. (2013). Modélisation de séries temporelles.
- [16] Roesbeck, J. D., Fichet, B. et Tassi, Ph. (1989) . Séries Chronologique : Théorie et Pratique des Modèles ARIMA. Economica.
- [17] Serigne, M. (2007). Analyse et prévision des série temporelles par la méthode de Box & Jenkins.
- [18] Shumway, R.H. and Stoffer, D.S. (2005). Time Series Analysis and Its Applications. SPRINGER, New York.

- [19] Tenenhaus, M. (2007). Statistique ; Méthodes pour décrire, expliquer et prévoir. Dunod, Paris.
- [20] Tenenhaus M. (1994). Méthodes statistiques en gestion. Dunod, paris.
- [21] Yves, T. (2004). Résumé du cours de séries temporelles.

Résumé

L'analyse des séries temporelles et plus particulièrement la prévision à court et moyen terme, a connu un développement important. La prévision est fondamentale dans la mesure où elle est à la base de la prise de décision optimale. Prévoir le comportement futur d'une série chronologique, nécessite l'utilisation de plusieurs méthodes de prévision. Notre objectif de cette note à l'analyse des séries chronologique par la méthode de Box & Jenkins.

Mots clés: Série chronologique, prévision, méthode de Box & Jenkins.

Abstract

The analysis of time series and especially the short and medium term forecasts, was a significant development. The forecast is fundamental because it is the basis of the optimal decision-making. Predict the future behavior of a time series, requires the use of several methods of forecasting. Our purpose of this note to the analysis of time series by the Box & Jenkins.

Keywords: Time series, prediction, method of Box & Jenkins.

المخلص

شهد تحليل السلاسل الزمنية وخاصة التوقعات على المدى القصير والمتوسط تطورا مهما. كما يعتبر التنبؤ بالسلوك المستقبلي للسلاسل الزمنية أمر أساسي لأنه أساس صنع القرار حيث يتطلب استخدام عدة أساليب. نهدف من خلال هذه المذكرة إلى تحليل السلاسل الزمنية باستخدام منهجية بوكس و جنكينز.

الكلمات المفتاحية: السلاسل الزمنية, التنبؤ, منهجية بوكس و جنكينز .