

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire

وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Centre Universitaire Salhi Ahmed- Naama  
Institut des sciences et technologies  
Département de Mathématiques et Informatique



## *Mémoire de fin d'études*

En vue de l'obtention du diplôme de Master  
En : Mathématiques

Spécialité : Analyse fonctionnelle et EDP

## **Intitulé**

---

**Comparaison entre la méthode de Box-Jenkins et la  
méthode de lissage exponentiel dans les prévisions à court  
terme**

---

**Présenté par :**  
SAHLI Wahiba

**Soutenu :** Juillet 2021

**Devant le jury composé de :**

Dr.KENOUZA Jamel	MCB	C-Univ Naâma	Président
Dr.LAALA Zeyneb	MCB	C-Univ Naâma	Examinatrice
Dr.BELGUERNA Abderrahmane	MCA	C-Univ Naâma	Encadreur

**Année universitaire 2020/2021**

# Remerciement

Je tiens à remercier en premier lieu et avant tout ALLAH le tout puissant, qui nous a donné la force et la patience d'accomplir notre travail dans les meilleures conditions.

Je tiens également à remercier du coeur mon encadreur Dr.BELGUERNA Abderrahmane, son aide précieuse et ses conseils judicieuses.

Je voudrai également remercier les membres de jury Dr.Jamel KENOUZA et Dr.Zeyneb LAALA maîtres de conférences au centre universitaire Salhi Ahmed de Naâma, qui m'ont fait l'honneur de juger ce modeste travail.

# Table des matières

<b>Introduction</b>	<b>7</b>
<b>1 Concepts de base et processus</b>	<b>8</b>
1.1 Série chronologique . . . . .	8
1.2 Représentation graphique . . . . .	10
1.2.1 La composante tendancielle (trend) $T_t$ . . . . .	10
1.2.2 La composante saisonnière $S_t$ . . . . .	10
1.2.3 La composante résiduelle $R_t$ . . . . .	10
1.2.4 La composante cyclique $C_t$ . . . . .	11
1.3 Le schéma de décomposition . . . . .	11
1.3.1 Le modèle multiplicatif . . . . .	11
1.3.2 Le modèle additif . . . . .	11
1.4 Calcule des statistiques de base . . . . .	11
1.5 Stationnarité . . . . .	12
1.5.1 Stationnarité Stricte . . . . .	12
1.5.2 Stationnarité au second ordre . . . . .	12
1.6 Prévission . . . . .	12
1.6.1 Prévission économique . . . . .	13
1.6.2 Prévission des ventes . . . . .	13
1.6.3 Prévission d'une série chronologique . . . . .	13
1.7 Processus . . . . .	13
1.8 Bruit blanc . . . . .	14
1.9 Corrélacion, Autocorrélacion, Corrélogramme . . . . .	14
1.9.1 Fonction d'autocovariance . . . . .	14
1.9.2 Fonction d'autocorrélacion . . . . .	15
1.9.3 Fonction d'autocorrélacion partielle . . . . .	15
1.10 Les opérateurs linéaires . . . . .	16
1.10.1 L'opérateur retard $\mathbf{B}$ . . . . .	16
1.10.2 L'opérateur différence $\Delta$ . . . . .	16
1.10.3 L'opérateur De déssaisonalisation $\Delta_s$ . . . . .	17

1.11	Processus Stationnaires . . . . .	17
1.11.1	Processus auto-régressif <b>AR</b> . . . . .	17
1.11.2	Processus moyenne mobiles <b>MA</b> . . . . .	17
1.11.3	Processus moyenne mobiles <b>ARMA</b> . . . . .	18
1.12	Processus non stationnaire . . . . .	18
1.12.1	Processus <b>TS</b> . . . . .	18
1.12.2	Processus <b>DS</b> . . . . .	18
1.12.3	Processus <b>ARIMA</b> . . . . .	19
1.12.4	Processus <b>SARIMA</b> . . . . .	19
<b>2</b>	<b>La méthodologie de Box-Jenkins</b>	<b>21</b>
2.1	Présentation de cette méthodologie . . . . .	21
2.2	Tests des composantes saisonnières et tendancielle . . . . .	23
2.2.1	Test de Fisher . . . . .	23
2.2.2	Test de Dickey-Fuller . . . . .	24
2.3	Les étapes de Box-Jenkins . . . . .	24
2.3.1	L'identification . . . . .	24
2.3.2	Estimation . . . . .	25
2.3.3	Tests et validation . . . . .	25
2.3.3.1	Test sur les paramètres . . . . .	25
2.3.3.2	Critères d'informations . . . . .	25
2.3.3.3	Principe de parcimonie . . . . .	26
2.3.3.4	Tests sur les résidus . . . . .	26
2.3.4	Prévision . . . . .	31
<b>3</b>	<b>Lissage exponentiel</b>	<b>33</b>
3.1	Lissage exponentiel simple . . . . .	33
3.1.1	Définition . . . . .	33
3.1.2	Formules récursives de mise à jour . . . . .	34
3.1.3	Interprétation géométrique . . . . .	34
3.1.4	Choix de la constante de lissage . . . . .	35
3.2	Lissage exponentiel amélioré, ou double . . . . .	36
3.2.1	Définition . . . . .	36
3.2.2	Calcul des coefficients . . . . .	37
3.2.3	Formules récursives de mise à jour . . . . .	38
3.3	Méthode de Holt-Winters . . . . .	40
3.3.1	Introduction . . . . .	40
3.3.2	Ajout d'une composante saisonnière. Décomposition additive . . . . .	41
3.4	TP : Lissage Exponentiel . . . . .	44
3.4.1	Lissage et prévision de données simulées . . . . .	44

3.4.2	Lissage et prévision de la concentration en <i>co2</i> . . . . .	45
3.4.3	Lissage et prévision du <i>CAC40</i> . . . . .	45
<b>4</b>	<b>Comparaison des 2 méthodes</b>	<b>46</b>
4.1	Application de la méthode BOX-JENKINS . . . . .	47
4.2	Application de la méthode de lissage exponentiel . . . . .	48
4.3	Comparison Entre les Résultats . . . . .	49
	<b>Bibliographie</b>	<b>50</b>

# Table des figures

2.1	Différentes étapes de Box-Jenkins . . . . .	22
3.1	Lissage et prévision par la méthode de lissage exponentiel simple d'un bruit blanc [ gauche ] et d'une tendance linéaire $X(t) = 0.5t + \epsilon_t, \epsilon_t \sim N(0, 1)$ [ droite ]( $\alpha = .1, .5, .9$ ) . . . . .	36
3.2	Lissage et prévision par la méthode de lissage exponentiel double d'un bruit blanc [ gauche ] et d'une tendance linéaire $X(t) = 0.5t + \epsilon_t, \epsilon_t \sim N(0, 1)$ [ droite ] ( $\alpha = .1, .5, .9$ ) . . . . .	39
3.3	Lissage et prévision par la méthode de lissage exponentiel double sur la série $X(t) = 0.5t + \epsilon_t + 2 \cos(t\pi/6), \epsilon_t \sim N(0, 1) \cdot (\alpha = .1, .5, .9)$ . . . . .	39
3.4	Lissage et prévision par la méthode de Holt-Winters sur la série $x(t) = 0.5t + \epsilon_t + 2\cos(t\pi/6), \epsilon_t \sim N(0, 1)$ . . . . .	43
3.5	Prévision par la méthode de Holt-Winters sur les données de vente de champagne . . . . .	44
3.6	Valeurs de cloture journalières du CAC40 de 1991 à 1998 . . . . .	45
4.1	Taux de change de 1\$ contre le Dinar Algérien . . . . .	46
4.2	Taux de change de 1\$ contre le Dinar Algérien . . . . .	47
4.3	Les valeurs prévisionnelles selon ARIMA(2,0,2) . . . . .	48
4.4	Les valeurs prévisionnelles selon ARIMA(2,0,2) . . . . .	49

# Liste des tableaux

4.1	Résumé de notre série . . . . .	47
-----	---------------------------------	----

# Introduction

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, pour n'en citer qu'une petite partie. On s'intéresse à l'évolution au cours du temps d'un phénomène, dans le but de décrire, expliquer puis prévoir ce phénomène dans le futur. On dispose ainsi d'observations à des dates différentes, c'est à dire d'une suite de valeurs numériques indicées par le temps.

Ce travail se décompose en quatre chapitres. Dans le premier chapitre nous donnons les notions de bases sur les séries temporelles et l'étude de quelques propriétés nécessaires pour achever ce travail en arrivant à la comparaison entre les méthodes de prévisions des valeurs futures 1. Dans le deuxième chapitre on entame la fameuse méthode dite de Box-Jenkins, les étapes et les procédures d'estimation des paramètres et l'identification du modèle approprié pour faire des prévisions à court terme 2. Le troisième chapitre sera consacré à la méthode de lissage exponentiel 3 et ses applications. Dans le dernier chapitre en comparant les deux méthodes statistique étudiées dans le 2<sup>ème</sup> et 3<sup>ème</sup> chapitre à travers un exemple numérique concret 4. Et en fin on termine avec une conclusion 4.3.

# Chapitre 1

## Concepts de base et processus

L'analyse des séries temporelles est fondée sur l'exploitation de données historiques recueillies sur un phénomène donné, durant une certaine période, la recherche d'un modèle pouvant constituer un cadre mathématique reflétant l'évolution de ces données et partant de l'hypothèse que la passée pourrait garantir de l'avenir c'est-à-dire : effectuer des prévisions. Après avoir présenté la partie théorique de notre travail, nous présentons dans ce cours ce que nous aurons besoin au cours de notre application.

### 1.1 Série chronologique

Le terme "série temporelle" désigne à la fois les séries chronologiques réelles et une suite théorique des variables aléatoires indexées par le temps ( $t \in \mathbb{T}$ ), qui va servir à modéliser ces premières.

#### **Définition 1**

*La théorie des séries chronologiques (ou temporelles) abordée dans ce cours est appliquée de nos jours dans des domaines aussi variés que l'économétrie, la médecine ou la démographie, pour n'en citer qu'une petite partie. On s'intéresse à l'évolution au cours du temps d'un phénomène, dans le but de décrire, expliquer puis prévoir ce phénomène dans le futur. On dispose ainsi d'observations à des dates différentes, c'est à dire d'une suite de valeurs numériques indicées par le temps.*

#### **Exemple : 1**

*: On peut songer par exemple à l'évolution du nombre de voyageurs utilisant le train, à l'accroissement relatif mensuel de l'indice des prix ou encore à l'occurrence d'un phénomène naturel (comme le nombre de taches solaires). Cette suite d'observations d'une famille de variables aléatoires réelles notées  $(X_t)_{t \in \theta}$  est appelée série chronologique (ou temporelle). Dans la suite de ce cours, nous la noterons*

$$(X_t)_{t \in \theta}$$

**Objectifs principaux** L'étude d'une série chronologique permet **d'analyser**, **d'écrire** et **d'expliquer** un phénomène au cours

du temps et d'en tirer des conséquences pour des prises de décision. Cette étude permet aussi de faire **un contrôle**, par exemple pour la gestion des stocks, le contrôle d'un processus chimique... Plus généralement, nous pouvons déjà poser quelques problèmes lorsqu'on étudie une série chronologique.

Mais l'un des objectifs principaux de l'étude d'une série chronologique est **la prévision** qui consiste à prévoir les valeurs futures  $X_{T+h}$  ( $h = 1, 2, 3, \dots$ ) de la série chronologique à partir de ses valeurs observées jusqu'au temps  $T$  :  $X_1, X_2, \dots, X_T$ . La prédiction de la série chronologique au temps  $t + h$  est notée  $\hat{X}(h)$  et, en général, est différente de la valeur réelle  $X_{T+h}$  que prend la série au temps  $T + h$ . Pour mesurer cette différence, on définira **l'erreur de prédiction** par la différence  $\hat{X}_T(h) - X_{T+h}$  "en moyenne" avec l'idée que plus  $h$  est grand, plus grande est l'erreur. L'intervalle de précision, défini par les valeurs  $\hat{X}_T^{(1)}$  et  $\hat{X}_T^{(2)}(h)$ , est susceptible de contenir la valeur inconnue  $X_{T+h}$ . La qualité de la prédiction pourra être mesurée en se basant sur 80% des observations, puis en simulant une prédiction sur les 20% d'observations restantes. Cette technique est aussi utile pour :

- les séries qui contiennent des "trous"
- mesurer l'effet d'un phénomène accidentel (erreur,...)

Un autre problème intéressant est **la détection de ruptures résultantes en gras**, par exemple, d'un changement de politique (économique). Ces ruptures peuvent être de deux ordres :

une rupture de niveau ou une rupture de pente. Il existe encore bien d'autres objectifs immédiats relatifs à l'étude des **séries chronologiques**. Par exemple, si deux séries sont observées, on peut se demander quelle influence elles exercent l'une sur l'autre. En notant  $X_t$  et  $Y_t$  les deux séries en question, on examine s'il existe par exemple des relations du type

$$Y_t = a_1 X_{t+1} + a_3 X_{t+3}$$

Ici, deux questions se posent : tout d'abord, la question de **la causalité** i.e. quelle variable (ici ( $X_t$ )) va expliquer l'autre (ici ( $Y_t$ )), ce qui amène la deuxième question, celle du **décalage temporel** : si une influence de ( $X_t$ ) sur ( $Y_t$ ) existe, avec quel délai et pendant combien de temps la variable explicative ( $X_t$ ) influence-t-elle la variable expliquée ( $Y_t$ ). Un dernier problème important de la macro-économétrie est de déterminer les relations persistances (de long terme) des autres relations (de court terme).

## 1.2 Représentation graphique

On considère qu'une série chronologique  $(X_t)$  est la résultante de différentes composantes fondamentales :

Lorsque le prévisionniste dispose d'un historique, le premier réflexe qu'il doit avoir est de tracer le graphique de la série chronologique (série d'origine) dans un repère ortho-normé en portant en ordonnée, les valeurs des observations et en abscisse, les dates correspondantes (jour, mois, années ...). La visualisation de l'évolution de la chronique via le graphique, nous permettra de mettre en évidence les composantes constitutives d'une série chronologique qui sont très importantes pour le choix du modèle à adopter. Dans l'examen de l'allure générale de la courbe représentative de la série, on distingue quatre composantes fondamentales :

### 1.2.1 La composante tendancielle (trend) $T_t$

La composante fondamentale ou tendance (trend) traduit l'évolution à moyen terme du phénomène. On parle aussi de mouvement conjoncturel ou mouvement extra-saisonnier. La chronique correspondante, notée  $f_t = 1, \dots, T$ , est une fonction 'a variation lente supposée déterministe dans cette approche. Elle sera estimée sous forme paramétrique (polynôme, exponentielle,...)ou comme le résultat d'un opération de lissage.

### 1.2.2 La composante saisonnière $S_t$

La composante saisonnière ou mouvement saisonnier représente des effets périodiques de période connue  $p$  qui se reproduisent de façon plus ou moins identique d'une période sur l'autre. La chronique correspondante, également déterministe, est notée  $S_T = 1, \dots, T$ . Elle est généralement supposée rigoureusement périodique :  $S_{t+p} = S_t$  et les valeurs  $S_j = S_{i+j}, j = 1, \dots, p$  d'une période sont appelées coefficients saisonniers. Le bilan de l'effet saisonnier sur une période doit être nul car il est pris en compte dans la tendance. La composante saisonnière permet simplement de distinguer à l'intérieur d'une même période une répartition stable dans le temps d'effets positifs ou négatifs qui se compensent sur l'ensemble de la période.

### 1.2.3 La composante résiduelle $R_t$

La composante résiduelle ou variations accidentelles est la partie non structurée du phénomène. Elle est modélisée par une suite de variables aléatoires "t,  $t = 1, \dots, T$ , centrées, non corrélées et de même variance, on parle de bruit blanc.

### 1.2.4 La composante cyclique $C_t$

Certains phénomènes économiques étudiés à très long terme présentent une composante cyclique (cycles d'activité) dont la période, de plusieurs années, est souvent mal définie. Cette composante est prise en compte dans la tendance sur les séries de taille moyenne et ne sera pas étudiée en tant que telle ici

## 1.3 Le schéma de décomposition

Pour pouvoir séparer les deux 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.

### 1.3.1 Le modèle multiplicatif

Schéma additif :

$$Y_t = f_t + S_t + \varepsilon_t$$

### 1.3.2 Le modèle additif

Schéma multiplicatif :

$$Y_t = f_t \times S_t \times (1 + \varepsilon_t)$$

En utilisant  $(1 + \varepsilon_t)$  dans le cas multiplicatif, on conserve la même signification et les mêmes propriétés à chacune des trois composantes  $f_t$ ,  $S_t$  et  $\varepsilon_t$  dans les deux schémas de composition. Cependant il est nécessaire de supposer que  $(1 + \varepsilon_t)$  reste positif dans le modèle multiplicatif car la composante résiduelle ne peut être responsable du signe de la grandeur observée. Dans les schémas multiplicatif et mixte les oscillations dues à l'effet saisonnier ont une amplitude proportionnelle à la valeur de la tendance. C'est précisément l'argument utilisé pour faire le choix entre le schéma additif et le schémas multiplicatif au vu de la représentation graphique de la chronique.

## 1.4 Calcule des statistiques de base

**la Moyenne** La la formule générale de la moyenne d'une série temporelle de termes générale :

$$\bar{x} = \frac{1}{T} \sum_{t=1}^T X_t$$

**La variance**

La variance d'une série temporelle permet d'évaluer la dispersion autour de la moyenne

.La formule de la variance (Var) dans le cas d'une population connue totalement est :

$$Var(x) = \sum_{t=1}^T (x_t + \bar{x})^2 / T$$

## 1.5 Stationnarité

. La connaissance d'un processus équivaut à connaître la loi de tout vecteur  $1 ( \dots, ) (X_{t+1}, \dots, X_{t+h})$  entier. La notion de stationnarité au sens strict, analogue à celle de régime permanent en physique, est donnée par la

### 1.5.1 Stationnarité Stricte

$X_t$  est stationnaire au sens strict si et seulement si la loi de  $(X_{t+1}, \dots, X_{t+h})$  dépend seulement de h.

### 1.5.2 Stationnarité au second ordre

stationnaire, si et seulement si pour tout entier h,  $EX_{t+h}$  et  $cov(X_t, X_{t+h})$  ne dépendent que de h.

Notons que ces définitions sont équivalentes lorsque  $( ) t X$  est un processus gaussien (c'est-à-dire lorsque la loi de tout vecteur  $(X_{t+1}, X_{t+h})$  est gaussienne). Pour un processus stationnaire, les moments d'ordre 1 et 2,  $EX_t$  et  $var(X_t, X_{t+h})$ , sont donc constants au cours du temps. Toutefois cette propriété n'est pas suffisante. Quitte à retrancher  $\mu = EX_t$ , on peut toujours se ramener à un processus centre. Dans la suite tous les processus sont centrés, et la notion de stationnarité doit être comprise au sens faible.

## 1.6 Prévision

Le mot " prévision ", peut à la fois représenter un résultat pour l'entreprise et recouvrir un ensemble de méthodes très diverses qui ont en commun de chercher à réduire l'incertitude liée à la non connaissance du futur. La décision pour une entreprise est difficile car elle risque d'être coûteuse si les prévisions sur lesquelles elle est fondée sont infirmées. On peut distinguer trois types de prévision selon la période

- **Prévision à court terme** : prévision effectuée sur une période allant de 6 à 18 mois selon les domaines.
- **Prévision à long terme** : la limite inférieure d'une prévision à long terme se situe de 5 à 10 ans, cette limite varie d'un secteur à un autre. La prévision à long terme détermine une tendance générale qui considère les perturbations inconnues futur comme des événements aléatoire.

- **Prévision à moyen terme** : le moyen terme se situe entre le court terme et le long terme. Il apparaît donc comme une sorte de "vision floue " entre un horizon plus net, le long terme, et un aspect plus accessible, le court terme. En conséquence, les frontières du moyen terme avec le long terme et le court terme sont souvent imprécises.

### 1.6.1 Prévision économique

C'est une ou plusieurs hypothèses concernant l'évolution d'un phénomène bien connue. On peut représenter une prévision économique comme une projection d'un phénomène étudié sur l'avenir.

### 1.6.2 Prévision des ventes

Demande effective adressée à l'entreprise pour une période donnée (mois, année). Elle est issue de la prévision du marché et de la fixation d'un objectif de part de marché par l'entreprise.

### 1.6.3 Prévision d'une série chronologique

Le but de l'étude des séries chronologiques est la prévision, qui consiste à évaluer les valeurs futures  $X_{t+h}$ ; ( $h \geq 1$ ) d'une variable à partir de l'observation de ses valeurs passées  $X_1, X_2, \dots, X_t$ . Or la valeur que la variable prendra à l'instant  $t + h$  sera plus au moins égale à la valeur prédite; pour cette raison, notre objectif est de proposer un intervalle de prévision susceptible de contenir la valeur inconnue. La qualité de la prévision dépend d'une part de l'évolution de la série, d'autre part de l'horizon  $h$ , en d'autres termes, la qualité de la prévision est meilleure lorsque la série évolue d'une façon régulière dans le temps et ou lorsque l'horizon est petit.

## 1.7 Processus

D'un point de vue sémantique, un processus est un ensemble de phénomènes conçus comme une chaîne causale progressive à travers le temps, du point de vue mathématique, il est solution d'une équation composée de variables indicées par le temps; si l'évolution temporelle d'une ou plusieurs variables n'est pas totalement déterministe (présence d'une ou plusieurs variables aléatoires) on parle alors de processus stochastique.

## 1.8 Bruit blanc

### Définition 2

On dit de la suite de variable aléatoire  $\varepsilon_t$  qu'elle constitue un bruit faible si elle possède les propriétés suivantes :

$$\begin{aligned} \mathbf{E}[\varepsilon_t] &= 0 \\ \mathbf{E}[\varepsilon_t^2] &= \sigma_\varepsilon^2 \neq 0 \text{ et constante} \\ \text{Cov}(\varepsilon_s, \varepsilon_t) &= 0 \text{ si } t \neq s \end{aligned} \tag{1.1}$$

En d'autres termes, les variables aléatoires  $\varepsilon_t$  sont de moyenne nulle, de variance constante et non corrélées. On dit que  $\varepsilon_t$  est un bruit blanc fort s'il est un bruit blanc faible et que les variables  $\varepsilon_t$  sont i.i.d (indépendantes et identiquement distribuées).

### Notation 1

1. Si  $\varepsilon_t$  bruit blanc faible, alors  $\varepsilon_t \sim \mathbf{WN}(0, \sigma_\varepsilon^2)$
2. Si  $\varepsilon_t$  bruit blanc fort alors  $\varepsilon_t \sim \mathbf{I.I.D}(0, \sigma_\varepsilon^2)$

## 1.9 Corrélation, Autocorrélation, Corrélogramme

La corrélation est un moyen utilisé pour comparer l'évolution de deux ou plusieurs phénomènes. L'autocorrélation est un concept lié à celui de la corrélation, il s'agit d'une corrélation entre la série et elle-même à différents décalages dans le temps

### 1.9.1 Fonction d'autocovariance

Soit  $(X_t)_{t \in Z}$ , un processus stationnaire. On appelle fonction d'autocovariance la fonction  $\gamma(h)$  définie de  $Z$  dans  $\mathbb{R}$  par :

$$\forall h, t \in Z, \gamma(h) = \text{Cov}(X_t, X_{t+h})$$

. Le graphe de cette fonction est appelé corrélogramme, la fonction d'auto-covariance d'un processus stationnaire vérifie

- $\forall h, t \in Z, \gamma(-h) = \gamma(h)$  : elle est paire
- $\forall n \in \mathbb{N}, \forall (a_1) \in \mathbb{R}^n, \forall (t_1) \in \mathbb{Z}^n \sum_{i=1}^n \sum_{j=1}^n a_i a_j (t_i - t_j) > 0$
- $\gamma(0) = \text{Var}(X_t)$
- $|\gamma(h)| \leq \gamma(0), \forall h$

### 1.9.2 Fonction d'autocorrélation

Soit  $(\mathbf{X}_t)_{t \in Z}$  un processus stationnaire. On appelle fonction d'autocorrélation la fonction  $f$  définie de  $Z$  dans  $\mathbb{R}$  par :

$$\forall h, \in Z, \rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

Le graphe de cette fonction est appelé corrélogramme La fonction d'autocorrélation d'un processus stationnaire vérifie :

- $\forall h, \in Z, \rho(-h) = \rho(h)$  : elle est paire
- $\forall \rho(0) = 1$
- $|\rho(h)| \leq 1, \forall h$

La fonction  $\rho(h)$  est l'expression du lien linéaire entre  $(\mathbf{X}_t)_{t \in Z}$  et  $(\mathbf{X}_{t-h})_{t \in Z}$  à l'instant présent et  $h > 0$ ,

$\rho(h)$  est l'expression du lien linéaire entre le présent et le passé d'ordre  $h$  (mais aussi par parité de la fonction  $\gamma$  entre le présent et le futur d'ordre  $h$ ) ; plus  $|\rho(h)|$  est proche de 1 et plus ce lien est fort Soit  $(\mathbf{X}_t)_{t \in Z}$  un processus stationnaire. Définissons tout d'abord la régression affine de  $(\mathbf{X}_t)$  sur  $(\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-h})$  notée  $\mathbf{X}_{t,h}^*$ . On a

$$\mathbf{X}_t = \mathbf{X}_{t,h}^* + \mathbf{R}_{t,h} = \lambda_{0,h} \sum_{s=1}^h \lambda_{s,h} \mathbf{X}_{t-s} + \mathbf{R}_{t,h}$$

### 1.9.3 Fonction d'autocorrélation partielle

Où est une variable aléatoire non corrélée avec  $\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-h}$ . La fonction d'autocorrélation partielle  $\tau$  est définie par :

$$h \in \mathbb{Z}, \tau(h) = \lambda_{h,h}$$

Le graphe de cette fonction est appelée corrélogramme partiel

L'autocorrélation partielle d'ordre  $h$   $\tau(h)$  représente le coefficient de corrélation linéaire

- \* le résidu  $\mathbf{X}_t - \mathbf{X}_{t,h-1}^*$  de la régression de  $\mathbf{X}_t$  par  $\mathbf{X}_{t-h+1}, \dots, \mathbf{X}_{t-1}$
- \* le résidu  $\mathbf{X}_{t-h} - \mathbf{X}_{t-h,h-1}^*$  de la régression de  $\mathbf{X}_{t-h}$  par  $\mathbf{X}_{t-h+1}, \dots, \mathbf{X}_{t-1}$

En d'autres termes,

$$\begin{aligned} \mathbf{X}_t &= \alpha_1 \mathbf{X}_{t-1} + \alpha_2 \mathbf{X}_{t-2} + \dots + \alpha_{h-1} \mathbf{X}_{t-h+1} + V \\ \mathbf{X}_{t-h} &= \beta_1 \mathbf{X}_{t-1} + \beta_2 \mathbf{X}_{t-2} + \dots + \beta_{h-1} \mathbf{X}_{t-h+1} + V \end{aligned} \tag{1.2}$$

et

$$\tau(h) = Cor(U, V)$$

Il faut bien comprendre que l'autocorrélation partielle est la corrélation entre  $X_{t-h}$  et  $X_{t-h}$  une fois que l'on a expliqué ceux-ci par les valeurs entre eux deux  $X_{t-h+1}, \dots, X_{t-1}$ .

## 1.10 Les opérateurs linéaires

### 1.10.1 L'opérateur retard B

l'opérateur retard B (Backward noté aussi pour L log) 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

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

L'opérateur retard possède les propriétés suivantes

- $B^j X_{t-j}$ , en particulier,  $B^0 X_t = X_t$
- Si  $X_t = c \in \mathbb{R}$  pour tout  $t \in \mathbb{Z}$ , alors  $B^j X_t = jc = c$  pour tout  $j \in \mathbb{Z}$
- $B^j (B^k X_t) = B^{j+k} X_t = X_{t-j-k}$
- $B^{-j} X_t = X_{t+j}$
- $(B^j + B^k) X_{t+j}$
- $(B^j + B^k) X_t = X_{t-j} + X_{t-k}$

Si l'on applique le polynôme retard  $\phi(B)$  défini comme suit

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

à une série  $X_t$ , on a :

$$\phi(B)X_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p}$$

Où  $\phi_i$ , des nombres réels, ..., p

### 1.10.2 L'opérateur différence $\Delta$

L'opérateur  $\Delta$  fait la différence entre le processus et sa version décalée d'une unité de temps. L'opérateur  $\Delta$  est défini par :

$$\Delta X_t = X_t - X_{t-1}$$

En effet, on peut écrire :

$$\Delta X_t = X_t - \mathbf{B}X_t = (1 - \mathbf{B})X_t$$

On peut écrire  $\Delta$  sous la forme d'un polynôme en  $\mathbf{B}$  avec :  $\Delta = 1 - \mathbf{B}$

### 1.10.3 L'opérateur De désaisonnalisation $\Delta_s$

L'opérateur  $\Delta_s$  est défini par :

$$\Delta_s X_t = X_t - X_{t-s}$$

s : la période de la saisonnalité en d'autre termes :

$$\Delta_s = (1 - \mathbf{B}^s)$$

## 1.11 Processus Stationnaires

Les processus  $\text{ARMA}(p, q)$  généralise les modèles autorégressifs et moyenne mobiles. Ces modèles sont très utilisés en pratique pour modéliser des séries réelles en nécessitant moins de paramètres que les modèles  $\text{AR}$  ou  $\text{MA}$  simples.

### 1.11.1 Processus auto-régressif $\text{AR}$

#### Définition 3

on appelle processus autorégressif d'ordre  $p$ , usuellement noté  $\text{AR}(p)$ , un processus stationnaire  $(X_t)_{t \in \mathbb{Z}}$  vérifiant une relation du type :

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

avec  $\phi_i \in \mathbf{R}$  et  $\varepsilon_i$  un bruit blanc de variance  $\sigma^2$ . Cette relation s'écrit également où  $\Phi(L) + 1 + \phi_1 L + \dots + \phi_p L^p$

### 1.11.2 Processus moyenne mobiles $\text{MA}$

#### Définition 4

on appelle processus moyenne mobile d'ordre  $q$ , usuellement noté  $\text{MA}(q)$ ; un processus  $X_t$  défini par :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}; \forall t \in \mathbb{Z}$$

où les  $\theta_i$  sont des réels et  $\varepsilon_t$  est un bruit blanc de variance  $\sigma^2$ . De même que pour les processus autorégressif cette relation s'écrit :

$$X_t = (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t = \Theta(L) \varepsilon_t$$

### 1.11.3 Processus moyenne mobiles ARMA

un processus stationnaire  $X_t$  admet une représentation ARMA( $p, q$ ) minimale s'il satisfait

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

soit, avec les notations suivantes :

1.  $\phi_p \neq 0$  et  $\theta_q \neq 0$
2.  $\Phi$  et  $\Theta$  n'est pas racines communes et leurs racines sont de modules  $> 1$
3.  $\varepsilon_t$  est un BB de variance  $\sigma^2$

## 1.12 Processus non stationnaire

A la section précédentes, quelque processus aléatoire stationnaires ont été présentés et étudiés. Toutefois dans ces conditions, il arrive souvent que des séries observées montrent un comportement non stationnaire. Dans ces conditions, il convient alors d'identifier le type de non stationnarité auquel la série est contrainte. cela permet d'utiliser un modèle approprié de façon à rendre stationnaire la série en question. Dans la suite, notamment les TS et DS :

### 1.12.1 Processus TS

Un processus  $(X_t)_{t \in \mathbb{Z}}$  fait partie de la famille des modèles à tendance saisonnière (TS) s'il peut s'écrire sous la forme

$$(X_t) = f(t) + \varepsilon_t$$

Pour stationnariser un processus de type TS on estime  $f(t)$  par les moindres carrés ordinaires.

### 1.12.2 Processus DS

Un processus non stationnaire  $(X_t)_{t \in \mathbb{Z}}$  fait partie de la famille des modèles à différence stationnaire (DS) d'ordre  $d$  si le processus filtré  $F_t = (1 - \beta)^d$ , comme cas particuliers

des modèles dans la suite

### 1.12.3 Processus ARIMA

L'objectif essentiel des modèles ARIMA est de permettre une prédiction de l'évolution future d'un phénomène. Son développement dans le domaine de l'économétrie est basé sur ce principe. Un autre intérêt, peut-être plus essentiel en ce qui concerne la recherche scientifique, est de comprendre la signification de ces différents processus. Il est clair cependant que cette interprétation dépend de la nature du phénomène étudié; et des modèles dont le chercheur dispose pour en rendre compte.

Un modèle ARIMA est noté  $(p, d, q)$ , correspondant au nombre de termes autorégressifs, au nombre de différenciation, et à un nombre de termes de moyenne mobile. De nombreuses combinaisons sont envisageables

Le processus stationnaire  $X_t)_{t \geq 0}$  satisfait une représentation ARIMA  $(p, d, q)$  si et seulement :

$$\phi(\beta^s)(1 - \beta^s)^d X_t = c + \theta(\beta^s)\varepsilon_t$$

Où  $c \in \mathbb{R}$

$$\phi(\beta) = \sum_{j=0}^p \phi_j \beta^j \text{ et } \theta(\beta) = \sum_{j=0}^q \theta_j \beta^j$$

on suppose que toutes les racines des polynômes  $\theta(\beta)$  et  $\phi(\beta)$  sont situées à l'extérieur du cercle unité.

### 1.12.4 Processus SARIMA

Le processus SARIMA est un modèle ARIMA dans lequel une composante saisonnière est ajoutée. Ainsi  $(X_t)_{t \in \mathbb{Z}}$  est un processus SARIMA  $(p, d, q)$  de périodes  $s$  si et seulement si :

$$\phi(\beta^s)(1 - \beta^s)^d X_t = c + \theta(\beta^s)\varepsilon_t$$

où  $c \in \mathbb{R}$

$$\phi(\beta) = \sum_{j=0}^p \phi_j \beta^j \text{ et } \theta(\beta) = \sum_{j=0}^q \theta_j \beta^j$$

La période de saisonnalité  $s$  est associée à la composante AR, alors que la période de saisonnalité  $s'$  est associée à la composante MA, on suppose que toutes les racines des

polynômes  $\theta(\beta)$  et  $\phi(\beta)$  sont situées à l'extérieur du cercle unité.

# Chapitre 2

## La méthodologie de Box-Jenkins

Les travaux de G. BOX et G. JENKINS (dans les années soixante dix) ont beaucoup contribué dans la théorie et la pratique des modèles des séries temporelles. L'objectif auquel ils se proposent de répondre dans leur ouvrage, "Time Series Analysis, Forecasting and Control", est de construire un modèle aléatoire de type ARMA permettant de reproduire au mieux les réalisations d'une série temporelle. Trois étapes sont nécessaires à la construction des modèles : une phase d'identification, une phase d'estimation, une phase de validation. A l'issue de ces trois phases, une fois déterminé le meilleur modèle ARMA, on utilise ce modèle à des fins de prévision

### 2.1 Présentation de cette méthodologie

Box – Jenkins (1970) ont élaboré une méthodologie pour identifier un modèle adéquat pour une série chronologique. Leur méthode est fondée sur les modèles..... Pour les méthodes d'inférence présentées dans la suite, on supposera que  $T$  réalisations d'une série chronologique uni-variée, notées  $y_1, y_2, \dots, y_T$  ont été observés. En décrivent leur méthode de prédiction George Box et Gwilym Jenkins Cette démarche divise le problème de prévision en 3 étapes pour pouvoir trouver le modèle susceptible de représenter la série temporelle, et la dernière étape c'est l'étape de prédiction :

1. **Identification.**
2. **Estimation.**
3. **Validation**
4. **Prévision**

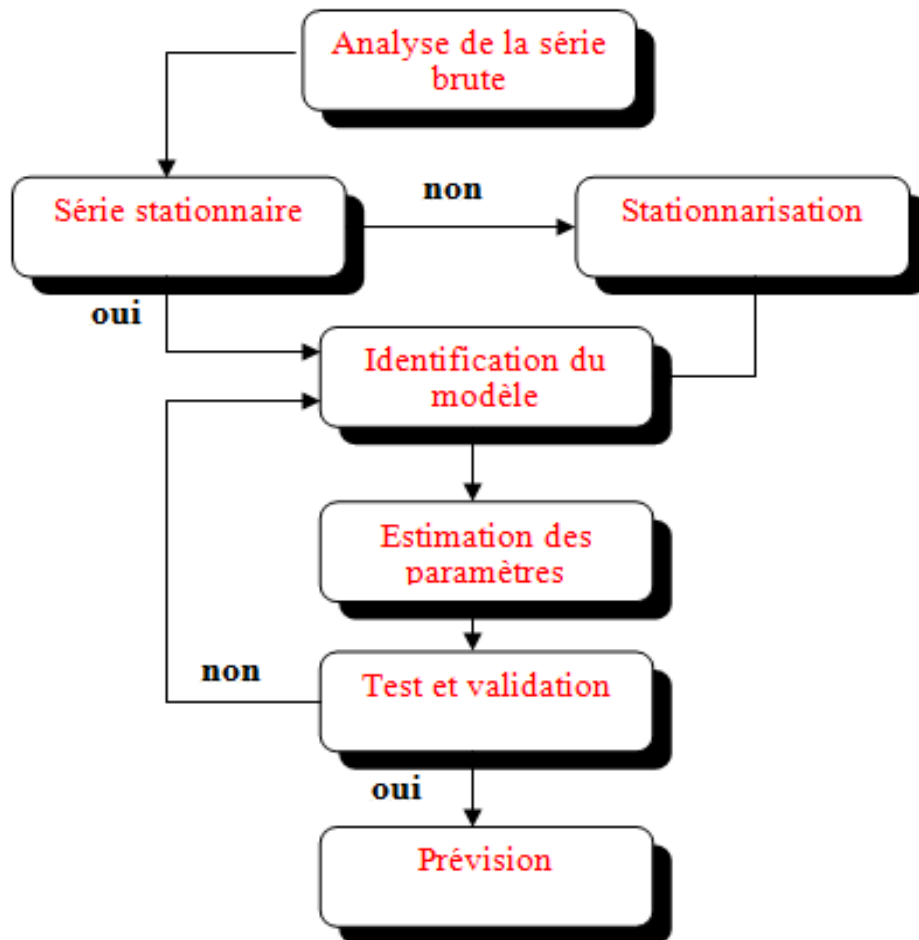


FIGURE 2.1 – Différentes étapes de Box-Jenkins

## 2.2 Tests des composantes saisonnières et tendanciennes

### 2.2.1 Test de Fisher

Avant toute étude d'une chronique, il convient d'élaborer un test permettant de détecter l'existence d'une saisonnalité. Le test le plus communément employé est celui de Fisher par analyse de la variance du facteur période (mensuel, trimestriel.).

On considère  $n$  : Le nombre d'années.  $p$  : Le nombre d'observations dans l'année

$X_{ij}$  : La valeur de la série pour la  $i^{me}$  année et la  $j^{me}$  période.

La moyenne générale  $\bar{X}_{..}$ , la moyenne de l'année  $i$   $\bar{X}_{i.}$ , la moyenne de la période  $j$   $\bar{X}_{.j}$

La variance année et la variance période sont définies respectivement par :

$$var_A = \frac{P \sum_{i=1}^n (\bar{X}_{i.} - \bar{X}_{..})^2}{n-1}, var_P = \frac{n \sum_{j=1}^p (\bar{X}_{.j} - \bar{X}_{..})^2}{n-1}$$

$$\text{La variance résiduelle : } var_R = \frac{P \sum_{i=1}^n n \sum_{j=1}^p (\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2}{(n-1)(n-1)}$$

L'équation de la variance totale

$$var_T = var_A + var_P + var_R = \frac{P \sum_{i=1}^n n \sum_{j=1}^p (\bar{X}_{ij} - \bar{X}_{..})^2}{(n-1)(n-1)}$$

L'hypothèse est :  $H_0$  : " pas de saisonnalité " contre  $H_1$  : "**il existe une saisonnalité**" La valeur calculée  $F_0 \frac{var_P}{var_R}$  que l'on compare à la valeur tabulée  $F_\alpha^{V_1 V_2}$  avec  $v_1 = (p-1)$ ,  $v_2 = (n-1)(p-1)$  degré de liberté Si  $F_0 > F_\alpha^{V_1 V_2}$  on rejette  $H_0$ , la série est saisonnière.

Soient les hypothèses :  $H_0$  : "**La série n'est pas affectée d'une tendance**"

Contre  $H_1$  : "**La série est affectée d'une tendance**"

On calcule  $F_1 \frac{var_P}{var_R}$  que l'on compare avec  $F_\alpha^{V_3 V_2}$

$v_3 = (n-1)$ ,  $v_2 = (n-1)(p-1)$  degré de liberté  $F_\alpha^{V_2 V_3}$

Si  $F_1 > F_\alpha^{V_2 V_3}$  on rejette l'hypothèse, la série est affectée d'une tendance. Concernant l'existence de la tendance, **le test de Fisher** s'avère faible, il convient d'effectuer un autre test

## 2.2.2 Test de Dickey-Fuller

IL permet de détecter non seulement l'existence d'une tendance mais aussi de quel type elle est. Pour la construction de ce test, Dickey et Fuller ont proposé trois modèles de base :

$$(1 - \phi_1 \mathbf{B})X_t = \varepsilon_t : \text{modèle autorégressif d'ordre 1} \dots \text{Modèle(1)}$$

$$(1 - \phi_1 \mathbf{B})(X_t - c) = \varepsilon_t : \text{modèle autorégressif d'ordre 1 avec constante} \dots \text{Modèle(2)}$$

$$(1 - \phi_1 \mathbf{B})(X_t - c - b_t) = \varepsilon_t : \text{modèle autorégressif d'ordre 1 avec tendance} \dots \text{Modèle(3)}$$

Les hypothèses à tester sont  $H_0 : \phi_1 = 1$  contre  $H_1 : |\phi_1| < 1$

C'est un test itératif qui se déroule comme suit :

- Si Dans l'un des modèles la statistique relative  $\phi_1$  à et supérieure aux valeurs tabulées, on déduit qu'il existe une racine unitaire. On estime en premier le modèle (3); on commence par tester la significativité de la tendance :

- Si la tendance n'est pas significative alors on teste la significativité de la constante
  1. recommence.
- Sinon, le processus est déterministe; pour le stationnariser on effectue une régression sur le temps et on recommence le test.

Sinon

- Si dans le modèle (3), le coefficient de la tendance est significativement différent de 0 alors le processus est TS
- sinon la série est stationnaire

## 2.3 Les étapes de Box-Jenkins

### 2.3.1 L'identification

Cette étape est la plus difficile dans le cycle de trois étapes définie par Box – Jenkins. Alors, plusieurs modèles candidats seront retenus et éliminés successivement dans les étapes deux et trois du cycle jusqu'à en retenir le bon. Pour cela plusieurs critères sont nécessaires pour identifier les modèles. Pour commencer, il n'y a pas de règles exactes qui nous permettent de connaître les degrés maximales des polynômes du modèle  $\text{ARMA}(p, q)$

de la série stationnaire. D'après Box et Jenkins (1976) La méthode d'identification est fondée sur la comparaison des moments empiriques de la série considérée aux moments théoriques associés aux différentes représentations potentielles. On se concentre ici sur les moments d'ordre deux résumés par la fonction d'autocorrélation (FAC) et la fonction d'autocorrélation partielle (FACP). D'une manière générale, on peut déduire que :

- ◀ : Pour un modèle  $AR(p)$  : la FAC de la série stationnaire ou devenue stationnaire décroît exponentiellement ou sinusoidalement et la FACP s'annule après un retard  $k = p$ .
- ◀ : Pour un modèle  $MA(q)$  : la FACP de la série stationnaire ou devenue stationnaire décroît exponentiellement ou sinusoidalement et la FAC s'annule après un retard  $k = q$ .

### 2.3.2 Estimation

On utilise pour cette phase la méthode d'estimation du maximum de vraisemblance ou bien celle des moindres carrés ordinaires. Elle consiste à identifier les coefficients autorégressifs et moyenne mobiles saisonniers ou non

### 2.3.3 Tests et validation

Après la phase précédente, si on dispose de plusieurs modèles candidats, il est nécessaire d'effectuer des tests sur les paramètres et sur les résidus afin de les départager

#### 2.3.3.1 Test sur les paramètres

C'est un test classique de Student qui permet de tester la significativité de chaque paramètre du processus en utilisant les hypothèses :  $H_0$  : Le coefficient  $\hat{B} = 0$  contre  $H_1$  : " $\hat{B}_p$  est significativement différent de zéro " La statistique de Student représente le ratio du coefficient à estimer sur son écart type  $\frac{\hat{B}_p}{\hat{\sigma}_{\hat{B}_p}}$

si  $|\frac{\hat{B}_p}{\hat{\sigma}_{\hat{B}_p}}| > 1.96(\alpha = 5)$  on accepte  $H_1$  dans le cas contraire on rejette le modèle et on envisage une nouvelle spécification ( retour à l'étape identification). Après avoir validé plusieurs modèles, il convient de choisir le modèle optimal en se basant sur une comparaison de leur qualité en utilisant les critères d'informations et /ou le principe de Parcimonie

#### 2.3.3.2 Critères d'informations

Ce sont des statistiques aidant à décider de l'ordre d'un modèle dans le cas où il en existerait plusieurs. Entre autres ; on peut citer :

- **Critère d'information d'Akaike (AIC)**

Il tient compte de la qualité d'ajustement du modèle à la série observée et du nombre de paramètres utilisés dans l'ajustement

$$AIC(p, q) = n \log \hat{\sigma}_\varepsilon^2 + 2(p + q)$$

— **Critère d'information Bayésien (BIC)**

Il a les mêmes caractéristiques que le AIC , néanmoins ; jugé plus intéressant dans la mesure où il pénalise les paramètres en nombre excessif plus fortement que le AIC

$$BIC(p, q) = \log \hat{\sigma}_\varepsilon^2 - (n - p - q) \text{Log} \left[ 1 - \frac{p+q}{n} \right] + (p + q) \text{Log} \left[ (p + q)^{-1} \left[ \frac{\sigma_x^2}{\hat{\sigma}_\varepsilon^2 - 1} \right] \right]$$

— **Critère de Schwartz 1978**

$$SC(p, q) = n \text{Log} \hat{\sigma}_\varepsilon^2 + (p + q) \text{Log} n$$

— **Critère de Hannan-Quin 1979**

$$HQ(p, q) = \text{Log} \hat{\sigma}_\varepsilon^2 + (p + q) c \text{Log} \left[ \frac{\text{Log} n}{n} \right] \text{ où } c \text{ est une constante à spécifier.}$$

Les critères les plus utilisés sont AIC, BIC, SC. Le modèle choisi (optimal) est celui qui minimise ces critères. Ils existent d'autres

critères appelés "critères de pouvoir prédictif " utilisés pour vérifier les performances prévisionnelles.

### 2.3.3.3 Principe de parcimonie

Dans le cas où les critères d'informations AIC et BIC de deux ou plusieurs modèles retenus seraient très proches ou contradictoires, on fait intervenir ce principe puisque le but de la modélisation d'une chronique est de minimiser le nombre de paramètres à estimer

### 2.3.3.4 Tests sur les résidus

Lorsque le modèle est bien estimé, les résidus entre les valeurs observées et les valeurs estimées doivent se comporter comme un bruit blanc (gaussien ou pas). Les résidus (ou erreurs de prévision) sont notés  $\varepsilon_t; t \in Z$ .

#### 1. Test des points de retournement (turning point)

Un point de retournement est une observation telle que

$$(\varepsilon_{i-1} < \varepsilon_i \text{ et } \varepsilon_i > \varepsilon_{i+1}) \text{ OU } (\varepsilon_{i-1} > \varepsilon_i \text{ et } \varepsilon_i < \varepsilon_{i+1})$$

La variable aléatoire  $X_i$  suit la loi de Bernoulli de paramètre  $\frac{2}{3}$

On appelle P le nombre total de ces points de retournement on note  $P = \sum_{i=1}^{n-2} X_i$

Pour n observation on a :  $E(P) = E \sum_{i=1}^{n-2} X_i = \frac{2}{3}(n - 2)$

$$E(P) = E[(\sum_{i=1}^{n-2} X_i)^2] = \frac{40n^2 - 144n + 131}{90} \text{ et } \text{var}(P) = \frac{16n - 29}{90}$$

Sous l'hypothèse  $H_0$  "les forment un échantillon aléatoire"

La statistique :  $E = \left| \frac{P - E5 + (P)}{\sqrt{\text{Var}(P)}} \right| \rightarrow (0, 1)$

On accepte l'hypothèse de bruit blanc au seuil de  $\alpha = 0.05$  si  $U < 1.96$

## 2. Test de Box-Ljung

Appelé aussi test de "porte manteau", il permet de vérifier l'hypothèse de bruit blanc des résidus, ce qui signifie qu'il n'existe aucune autocorrélation significativement non nulle, il se base sur la fonction d'autocorrélation en utilisant deux hypothèses :

$$H_0 : \rho_1 = 0, \rho_2 = 0, \rho_k = 0 \text{ contre}$$

$H_1$  Il existe au moins un  $\rho_i$  significativement différent de zéro

Sa statistique est :  $Q = n(n + 1) \sum_{h=1}^k \frac{r_h^2(\varepsilon_t)}{n-2}$  avec :

K = nombre de retards choisis et n = nombre d'observations.

En l'absence d'autocorrélation la statistique q obéit à une  $\chi^2(K - p - q - P - Q)$  degrés de liberté.

L'hypothèse  $H_0$  est rejetée au seuil de 5 si Q est supérieure au quantile 0.95 de la loi de  $\chi^2$

## 3. Test de nullité de la moyenne des résidus

$$H_0 : "m = 0" \text{ contre } H_1 : "m \neq 0" . \text{ La statistique utilisée est } t = \frac{\bar{\varepsilon}_t}{\sigma_\varepsilon \sqrt{n-1}}$$

Le test est basé sur la statistique de Student pour tester

#### 4. Test de normalité

celui de Jarque et Bera (1984), ce dernier est fondé sur la notion de Skewness (asymétrie) et de Kurtosis (Le test le plus fréquent qui permet de vérifier la normalité d'une distribution statistique est queue de distribution), par ailleurs il existe un autre test celui de Kolmogorov-Smirnov.

Ces deux tests permettent par la suite de calculer les intervalles de prévisions

- **Test de Jarque et Bera**

$$\text{soit } \mu_k = \frac{1}{n} \sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})^k$$

le moment centré d'ordre  $k$  du processus  $\varepsilon_t, t \in \mathbb{Z}$

Le coefficient de **Skewness** est défini par :  $B_1^{1/2} = \frac{\mu_3}{\mu_2^{3/2}}$

Le coefficient de **Kurtosis** est défini par :  $B_2 = \frac{\mu_4}{\mu_2^2}$

Si la distribution est normale et le nombre d'observation grand ( $n > 30$ ) :

$$B_1^{1/2} \rightarrow N(0, \sqrt{\frac{6}{n}}) \text{ et } B_2 \rightarrow N(3, \sqrt{\frac{24}{n}})$$

On construit alors les statistiques centrées réduites correspondantes à  $B_1^{1/2}$  et  $B_2$

$$v_1 = \frac{|B_1^{1/2} - 0|}{\sqrt{\frac{6}{n}}} \text{ et } v_2 = \frac{|B_2 - 3|}{\sqrt{\frac{24}{n}}}$$

Que l'on compare à 1.96 (valeur de la loi normale au seuil de 5)

$H_0 : v_1 = 0$  (symétrie) et  $v_2 = 0$  (aplatissement normal)

On acceptera l'hypothèse  $H_0$  si  $v_1 \leq 1.96$  et  $v_2 \leq 1.96$ ; dans le cas contraire, l'hypothèse de normalité est rejetée

Le test de **Jarque et Bera** synthétise les résultats précédents

Si  $B_1^{1/2}$  et  $B_2$  obéissent à des lois normales alors la statistique :

$JB = \frac{n}{6}B_1 + \frac{n}{24}(B_2 - 3)^2$  suit une  $\chi^2$  à deux degrés de liberté

Si  $JB > \chi^2_{1-\alpha}(2)$  on rejette l'hypothèse de normalité des résidus au seuil  $\alpha$ .

• **Test de Kolmogorov-Smirnov**

C'est un test d'adéquation qui est aussi important que le précédent. Il est fondé sur les statistiques fonctions de  $F_n$  et  $F$ , tel que fonction de répartition empirique qui est un estimateur sans de  $F$ .

On teste :  $H_0 : \mathcal{F} = F_0$  contre  $H_1 : \mathcal{F} \neq F_0$

Les statistiques définies sont :

$$D_n^+ = \max\left(\frac{i-1}{n} - F_0(X_i)\right) \quad D_n^- = \max\left(F_0(X_i) - \frac{i}{n}\right)$$

$$D_n = \max\{|D_n^+|, |D_n^-|\}$$

avec :  $X_i$  est la valeur de la coordonnée  $n^\circ i$  de la statistique d'ordre croissant.

Si  $D_n < c$  on accepte  $H_0$ ,  $c$  étant la valeur critique lue sur la table pour  $n \geq 35$ ,  $c = \frac{1.36}{\sqrt{n}}$  puisque  $\alpha = 5$

5. **Test de Von-Neuman (indépendance)**

C'est un test valide dans le cas où les résidus sont gaussiens, sous les hypothèses :

$H_0$  : " les résidus sont indépendants et identiquement distribués " contre

$H_1$  : " au moins deux observations successives tendent à être corrélées "

Il est fondé sur les deux estimateurs suivants :

$$D^2 = \frac{1}{n-1} \sum_{t=1}^{n-1} (\varepsilon_{t+1} - \varepsilon_t)^2; \quad S^2 = \frac{1}{n-1} \sum_{t=1}^n (\varepsilon_{t+1} - \bar{\varepsilon}_t)^2$$

La statistique utilisée est :  $VN = \frac{(D^2)^{-1}}{\sqrt{\frac{n-2}{n^2-1}}}$

Si  $|VN| < VN_\alpha = 1.96$  avec  $\alpha = 0.05$  on accepte  $H_0$ .

6. **Test de Durbin et Watson**

Les modèles ajustés à des séries chronologiques manifestent parfois un certain degré de corrélation entre les valeurs successives des erreurs. En

terme probabiliste, cela signifie que les erreurs sont autocorrélées, ou encore qu'une erreur produite en  $t-1$  a une influence sur l'erreur en  $t$ .

Le test de Durbin et Watson (1951) permet de détecter l'autocorrélation des résidus pour un ordre un (corrélation entre  $\varepsilon_t$

et  $\varepsilon_{t-1}$  ) sous la forme :

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \text{ où } v_t \rightarrow N(0, \sigma_v^2) \text{ on teste}$$

$H_0 : \rho = 0$  (absence d'autocorrélation à l'ordre 1 des résidus).

Contre  $H_1 : \rho \neq 0$  (présence d'autocorrélation à l'ordre 1 des résidus).

La statistique de **Durbin et Watson**, notée DW, est donnée par :

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} \text{ avec } \varepsilon_t : \text{ sont les}$$

résidus de l'estimation du modèle.

De par sa construction, cette statistique varie entre 0 et 4 et nous avons  $DW = 2$  lorsque  $\hat{\rho}$

( $\hat{\rho}$  est l'estimateur de  $\rho$  ) on a :

- ◇  $DW = 0$ , il existe une autocorrélation positive
- ◇  $DW = 4$  il existe une autocorrélation négative
- ◇  $DW \approx 2$  indique l'absence d'autocorrélation

### 7. Test d'hétéroscédasticité : (l'existence d'effet ARCH)

Pour ce faire, on utilise les corrélogrammes des carrés des résidus, si un ou plusieurs termes sont significativement différents de zéro, on

déduit qu'il y a effet ARCH qui est détecté par la statistique de Box et Ljung, ceci est confirmé par la statistique du Multiplicateur de Lagrange

$LM = nR^2$  avec  $n$ = le nombre d'observations servant au calcul de la régression .

$R^2$ = le coefficient de détermination.

Soit une spécification de type ARCH pour les erreurs  $\varepsilon_t$  tel que :

$$\varepsilon_t = u_t \cdot h_t \text{ avec } u_t \rightarrow N(0, 1) \text{ et } h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 \text{ et Soit l'hypothèse } H_0 : \alpha_1 = \alpha_2 = \dots \alpha_p = 0 \text{ contre } H_1 : \alpha_i \text{ non tous nul}$$

Si  $LM < \chi^2(1)$  on accepte  $H_0$ , la variance de l'erreur est constante  $h_t^2 = \alpha_0$

Dans le cas contraire  $LM > \chi^2(p)$  à  $p$  degrés de libertés compris entre 1 et 3, on rejette  $H_0$  et le processus est justifiable d'un modèle ARCH( $p$ ).

Si  $p > 3$  le modèle sera justifié d'un modèle de type GARCH.

### 2.3.4 Préviation

C'est l'ultime étape de cette méthode, voire de la quasi-totalité des analyses statistiques. C'est une extrapolation des observations d'une série en se basant sur ses observations et ses erreurs passées. Lors de cette étape, il est indispensable de prendre en considération les transformations effectuées sur la chronique afin d'aboutir à des valeurs prévisionnelles appropriées. Par définition :

$$X_{t+1} = E(X_{t+1}/I_t) \text{ ou est l'information disponible à } t.$$

$$I_t = (X_1, X_2, \dots, X_t, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_t)$$

Considérons un ARMA(1, 1) et un horizon de prévision 1

$$X_t = \phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$X_{t+1} = \phi_1 X_t + \varepsilon_{t+1} - \theta_1 \varepsilon_t$$

$$\hat{X}_{t+1} = E(X_{t+1}/I_t) = \phi_1 X_t + \theta_1 \varepsilon_t$$

$$\hat{X}_{t+2} = E(X_{t+2}/I_t) = \phi_1 E(X_{t+1}/I_t)$$

$$\hat{X}_{t+2} = \phi_1 E((\phi_1 X_t + \varepsilon_{t+1} - \theta_1 \varepsilon_t)/I_t) = \phi_1 E((\phi_1 X_t - \theta_1 \varepsilon_t)/I_t)$$

$$\hat{X}_{t+2} = \phi_1 (\phi_1 X_t - \theta_1 \varepsilon_t) = \phi_1 \hat{X}_{t+1}$$

En généralisant pour un horizon  $h$  :

$$\phi(B)X_t = \theta(B)\varepsilon_t \Leftrightarrow X_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t = \Psi(B)\varepsilon_t$$

$$X_t = \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i} \text{ et } \hat{X}_{t+h} = E(X_{t+h}/I_t) = E(\sum_{i=1}^{\infty} \Psi_i \varepsilon_{t+h-i}/I_t)$$

avec

$$E(\varepsilon_{t+h-i}/I_t) = \begin{cases} \varepsilon_{t+h-i} & \text{si } h - i \leq 0 \Rightarrow i \geq h, \\ 0 & \text{si non.} \end{cases}$$

$$\text{d'où } \hat{X}_{t+h} = \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t+h-i}$$

Comme la prévision n'est pas identique à la valeur réelle, il existe un écart entre ces deux valeurs appelé erreur de prévision  $\hat{\varepsilon}_{t+1}$

$$\hat{\varepsilon}_{t+1} = X_{t+1} - \hat{X}_{t+1} = \varepsilon_{t+1}$$

$$\hat{\varepsilon}_{t+2} = X_{t+2} - \hat{X}_{t+2} = \varepsilon_{t+2} + \Psi \varepsilon_{t+1}$$

$$\hat{\varepsilon}_{t+h} = X_{t+h} - \hat{X}_{t+h} = \sum_{i=0}^{h-1} \Psi_i \varepsilon_{t+h-i} \text{ avec } \Psi_0 = 1$$

Sous réserve que les erreurs forment un bruit blanc gaussien, on pourra construire un intervalle de prévision tel que

$$v(\hat{\varepsilon}_{t+h}) = v[\sum_{i=0}^{h-1} \Psi_i \varepsilon_{t+h-i}] = E(\sum_{i=0}^{h-1} \Psi_i \varepsilon_{t+h-i})^2 - 0 = \sigma_{\varepsilon}^2 \sum_{i=0}^{h-1} \Psi_i^2$$

$$\text{D'où l'intervalle de prévision : } \hat{X}_{t+h} \pm 1.96 \sigma_3 (\sum_{i=0}^{h-1} \Psi_i^2)^{\frac{1}{2}}$$

Nous pouvons dire pour cette méthode, qu'en dépit de sa simplicité dans son principe de base, elle semble complexe dans sa mise en œuvre, d'autant plus qu'elle ne prend pas en compte les informations externes. Néanmoins, elle permet non seulement de traiter et d'analyser les processus stationnaires mais également les processus qui ne le sont pas.

# Chapitre 3

## Lissage exponentiel

Les méthodes de lissages exponentielles, dont il existe plusieurs variantes que nous présentons ici, sont des méthodes empiriques de prévision de série temporelle. Elles présentent l'intérêt d'être facilement compréhensibles et leur implementation récursive en font un outil efficace pour le traitement de gros volumes de données ou dans des systèmes embarqués disposant de peu de mémoire. Il faut noter que, bien que largement utilisés, ces méthodes souffrent d'un manque de bases théoriques solides comme celles des *ARMA*, *ARIMA* et *SARIMA*.

toutes ces méthodes consistent à ajuster à une chronique de série temporelle une estimation locale de ce que va être sa valeur future. Selon les variantes :

1. une constante pour le lissage exponentiel simple
2. une droite pour le lissage exponentiel double ou de Holt
3. des fonctions polynomiales ou périodiques pour les lissages plus généraux

### 3.1 Lissage exponentiel simple

#### 3.1.1 Définition

La prévision  $\hat{x}_{n,h}$  est construite :

- en prenant en compte toute l'histoire de la chronique
- de sorte que, plus on s'éloigne de la base  $n$  de la prévision, moins l'influence des observations correspondantes est importante. Cette décroissance de l'influence est de type exponentiel. De là vient le nom de la technique.

On se donne  $\alpha$ , appelé constante de lissage, avec  $0 < \alpha < 1$ , et on définit (la prévision  $\hat{x}_{n,h}$  ne dépend pas de  $h$ ) :

$$\hat{x}_{n,h} = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j} \quad (3.1)$$

Remarquons que plus  $\alpha$  est voisin de 1, plus l'influence des observations éloignées dans le temps de la base  $n$  est grande.

### 3.1.2 Formules récursives de mise à jour

la formule (3.1) vérifie les formules de récurrences suivantes

$$\hat{x}_{n,h} = (1 - \alpha)x_n + \alpha\hat{x}_{n-1,h} \quad (3.2)$$

$$\hat{x}_{n,h} = \hat{x}_{n-1,h} + (1 - \alpha)(x_n - \hat{x}_{n-1,h}) \quad (3.3)$$

Par conséquent, la connaissance de la prévision à l'horizon sur la base  $n - 1$ , soit  $\hat{x}_{n-1,h}$ , et l'observation  $x_n$  suffisent pour calculer immédiatement la prévision suivante  $\hat{x}_{n,h}$ .

Cependant, l'utilisation des formules récursives nécessite d'initialiser la récurrence en prenant par exemple  $\hat{x}_{1,h} = x_1$ , on remarque que pour  $n$  assez grand la valeur initiale a peu d'influence.

Remarquons aussi que la formule (3.2) permet d'interpréter  $\hat{x}_{n,h}$  comme le barycentre de  $x_n$  et de  $\hat{x}_{n-1,h}$  affectés respectivement des masses  $1 - \alpha$  et  $\alpha$ . L'influence de la dernière observation est d'autant plus grande que  $\alpha$  est proche de 1.

### 3.1.3 Interprétation géométrique

Asymptotiquement  $\hat{x}_{n,h}$  est la meilleure approximation au sens des moindres carrés (pour des pondérations exponentielles) de  $x$  par le vecteur constant  $a\delta$  où  $\delta = (1, \dots, 1) \in \mathbb{R}^n$ . En effet la valeur de  $a$  qui minimise :

$$\sum_{j=0}^{n-1} \alpha^j (x_{n-j} - a)^2 \quad (3.4)$$

est fournie par :

$$\hat{a}(n) = \frac{1 - \alpha}{1 - \alpha^n} \sum_{j=0}^{n-1} \alpha^j x_{n-j}.$$

Si l'on transforme légèrement le modèle en considérant que

$$x = (x_1, \dots, x_n) = (\dots, 0, \dots, 0, x_1, \dots, x_n)$$

c'est à dire que les coordonnées du nouveau vecteur sont nulles pour tous les indices

négatifs ou nul. On cherche donc la valeur de  $a$  qui minimise

$$\sum_{j=0}^{\infty} \alpha^j (x_{n-j} - a)^2$$

la solution est fournie par :

$$\hat{x}_{n,h} = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j}$$

Il est facile de montrer que les deux estimateurs sont asymptotiquement équivalents, on a :

$$\lim_{n \rightarrow \infty} \hat{\alpha}(n) = \lim_{n \rightarrow \infty} \hat{x}_{n,h}$$

Dans la minimisation de (3.4) là aussi, l'influence des observations décroît exponentiellement lorsqu'on s'éloigne de la base. On peut dire que l'ajustement par la constante  $a$  est local et se fait au voisinage de  $n$ . La taille réelle de la plage des observations influentes dépend de  $\alpha$ . Plus  $\alpha$  est grand plus la taille est grande.

### 3.1.4 Choix de la constante de lissage

La méthode n'est pas bonne en présence de tendance, de composante saisonnière ou de fluctuations de hautes fréquences. En pratique le lissage exponentiel simple est assez peu utilisé.

Des valeurs de  $\alpha$  comprises entre 0.7 et 0.99 produisent des prévisions qui tiennent compte du passé lointain. Elles sont rigides et peu sensibles aux variations conjoncturelles. Il n'en n'est plus de même si  $\alpha$  est compris entre 0.01 et 0.3, la prévision est souple c'est à dire influencée par les observations récentes.

On peut aussi déterminer un  $\alpha$  adapté à la prévision d'une série donnée. Pour un horizon  $h$  on cherche la valeur de  $\alpha$  qui minimise  $\sum_{t=1}^{n-h} (x_{t+h} - (1 - \alpha) \sum_{j=0}^{t-1} \alpha^j x_{t-j})^2$ , c'est à dire la valeur qui en moyenne réalise la meilleure prévision à l'horizon  $h$  sur les bases  $1, 2, \dots, n - h$  en utilisant le critère des moindres carrés. On choisit une grille de valeurs de  $\alpha$  pour lesquelles on calcule le critère et on retient la valeur qui le minimise. Cette technique de choix de constante peut être reprise dans les méthodes qui vont suivre. Les calculs sont souvent plus longs car ce n'est, en fait, plus une constante de lissage que nous sommes amenés à déterminer mais souvent deux ou trois.

## 3.2 Lissage exponentiel amélioré, ou double

### 3.2.1 Définition

On ajuste au voisinage de  $n$  une droite d'équation

$y_t = a_1 + a_2(t - n)$ . Le prédicteur sera :

$$\hat{x}_{n,h} = \hat{a}_1 + \hat{a}_2 h \tag{3.5}$$

Les coefficients  $\hat{a}_1, \hat{a}_2$  sont solution de :

$$\sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (\hat{a}_1 - \hat{a}_2 j))^2 = \inf_{a_1 \in R, a_2 \in R} \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2 \tag{3.6}$$

En remplaçant, comme dans le lissage simple, la sommation finie par une somme de 0 à  $+\infty$ , les équations précédentes deviennent

$$\sum_{j=0}^{\infty} \alpha^j (x_{n-j} - (\hat{a}_1 - \hat{a}_2 j))^2 = \inf_{a_1 \in R, a_2 \in R} \sum_{j=0}^{\infty} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2$$

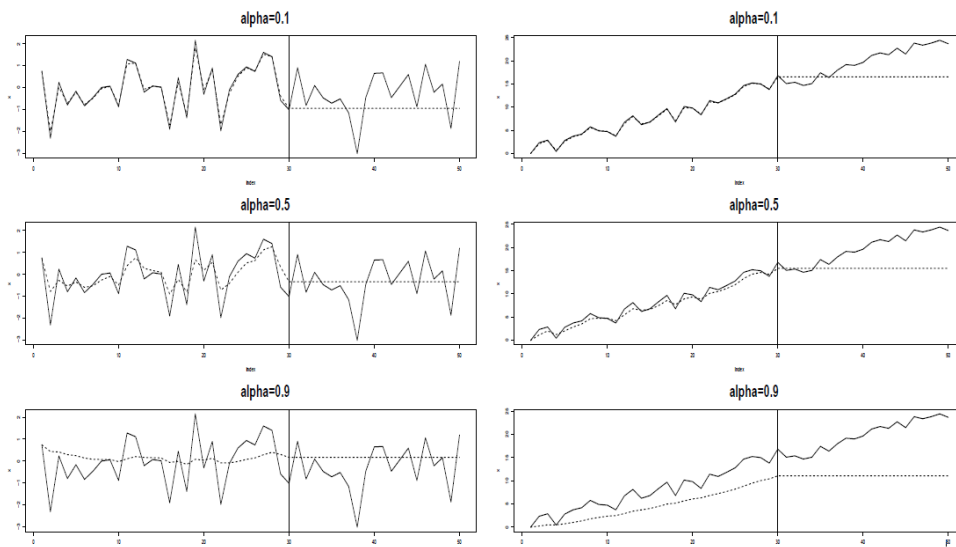


FIGURE 3.1 – Lissage et prévision par la méthode de lissage exponentiel simple d'un bruit blanc [ gauche ] et d'une tendance linéaire  $X(t) = 0.5t + \epsilon_t, \epsilon_t \sim N(0, 1)$  [ droite ] ( $\alpha = .1, .5, .9$ )

### 3.2.2 Calcul des coefficients

On annule les dérivées partielles par rapport à  $a_1$  et  $a_2$  du critère (3.6) et on vérifie à l'aide des dérivées secondes que l'on a bien un minimum. En notant :

$$C = C(a_1, a_2) = \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j))^2$$

$$\frac{\delta C}{\delta a_1} = -2 \sum_{j=0}^{n-1} \alpha^j (x_{n-j} - (a_1 - a_2 j))$$

$$\frac{\delta C}{\delta a_2} = 2 \sum_{j=0}^{n-1} j \alpha^j (x_{n-j} - (a_1 - a_2 j))$$

on a :

$$\sum_{j=0}^{n-1} \alpha^j x_{n-j} - a_1 \sum_{j=0}^{n-1} \alpha^j + a_2 \sum_{j=0}^{n-1} j \alpha^j = 0 \quad (3.7)$$

$$\sum_{j=0}^{n-1} j \alpha^j x_{n-j} - a_1 \sum_{j=0}^{n-1} j \alpha^j + a_2 \sum_{j=0}^{n-1} j^2 \alpha^j = 0$$

Puisque  $\sum_0^\infty \alpha^j = \frac{1}{1-\alpha}$ ,  $\sum_0^\infty j \alpha^j = \frac{\alpha}{(1-\alpha)^2}$ ,  $\sum_0^\infty j^2 \alpha^j = \frac{\alpha(1+\alpha)}{(1-\alpha)^3}$ , les relations (3.7) deviennent :

$$(1-\alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j} - a_1 + a_2 \frac{\alpha}{1-\alpha} = 0$$

$$(1-\alpha)^2 \sum_{j=0}^{n-1} j \alpha^j x_{n-j} - a_1 \alpha + a_2 \frac{\alpha(1+\alpha)}{1-\alpha} = 0$$

Introduisons deux lissages exponentiels simples successifs, celui des  $x$  produit  $L_1$ , celui de  $L_1$  produit  $L_2$  :

$$L_1(n) = (1-\alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j}$$

$$L_2(n) = (1-\alpha) \sum_{j=0}^{n-1} \alpha^j L_1(n-j)$$

On a donc

$$L_2(n) - (1-\alpha)L_1(n) = (1-\alpha)^2 \sum_{j=1}^{n-1} j \alpha^j x_{n-j}$$

et  $L_1$  et  $L_2$  sont liés à  $a_1$  et  $a_2$  par les relations

$$\begin{aligned} L_1(n) - a_1 + a_2 \frac{\alpha}{1 - \alpha} &= 0 \\ L_2(n) - (1 - \alpha)L_1(n) - a_1\alpha + a_2 \frac{\alpha(1 + \alpha)}{1 - \alpha} &= 0 \end{aligned}$$

Les solutions cherchées vérifient donc,

$$\begin{aligned} \hat{a}_1(n) &= 2L_1(n) - L_2(n) \\ \hat{a}_2(n) &= \frac{1 - \alpha}{\alpha} [L_1(n) - L_2(n)] \end{aligned} \quad (3.8)$$

Soit aussi :

$$\begin{aligned} L_1(n) &= \hat{a}_1(n) - \frac{\alpha}{1 - \alpha} \hat{a}_2(n) \\ L_2(n) &= \hat{a}_1(n) - \frac{2\alpha}{1 - \alpha} \hat{a}_2(n) \end{aligned} \quad (3.9)$$

### 3.2.3 Formules récursives de mise à jour

Commençons par  $L_1(n)$  et  $L_2(n)$ .

$$L_1(n) = (1 - \alpha) \sum_{j=0}^{n-1} \alpha^j x_{n-j} = (1 - \alpha)x_n + \alpha L_1(n - 1) \quad (3.10)$$

$$\begin{aligned} L_2(n) &= (1 - \alpha)L_1(n - 1) + \alpha L_2(n - 1) \\ &= (1 - \alpha)^2 x_n + \alpha(1 - \alpha)L_1(n - 1) + \alpha L_2(n - 1) \end{aligned} \quad (3.11)$$

récrivons (3.8) en utilisant (3.10) et (3.11).

$$\begin{aligned} \hat{a}_1(n) &= 2[(1 - \alpha)x_n + \alpha L_1(n - 1)] - [(1 - \alpha)^2 x_n + \alpha(1 - \alpha)L_1(n - 1) + \alpha L_2(n - 1)] \\ &= (1 - \alpha^2) x_n + \alpha(1 + \alpha)L_1(n - 1) - \alpha L_2(n - 1) \end{aligned}$$

en tenant compte de (3.9)

$$\begin{aligned} \hat{a}_1(n) &= (1 - \alpha^2) x_n + \alpha(1 + \alpha) \left[ \hat{a}_1(n - 1) - \frac{\alpha}{1 - \alpha} \hat{a}_2(n - 1) \right] \\ &\quad - \alpha \left[ \hat{a}_1(n - 1) - \frac{2\alpha}{1 - \alpha} \hat{a}_2(n - 1) \right] \\ &= (1 - \alpha^2) x_n + \alpha^2 [\hat{a}_1(n - 1) + \hat{a}_2(n - 1)] \end{aligned}$$

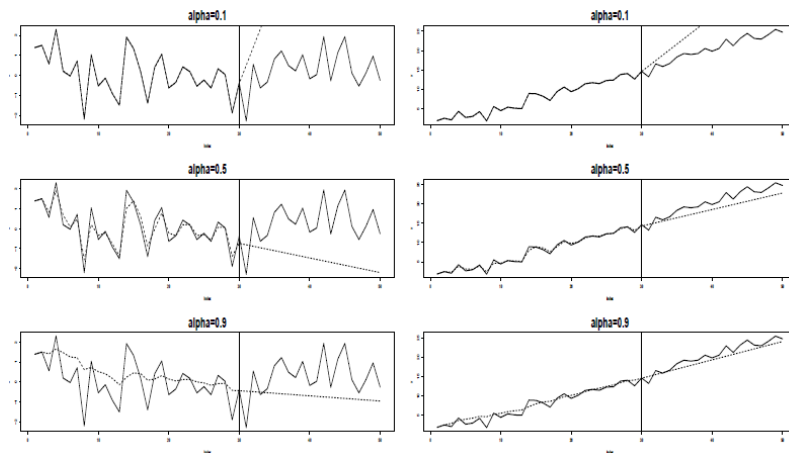


FIGURE 3.2 – Lissage et prévision par la méthode de lissage exponentiel double d’un bruit blanc [ gauche ] et d’une tendance linéaire  $X(t) = 0.5t + \epsilon_t$ ,  $\epsilon_t \sim N(0, 1)$ . [ droite ] ( $\alpha = .1, .5, .9$ )

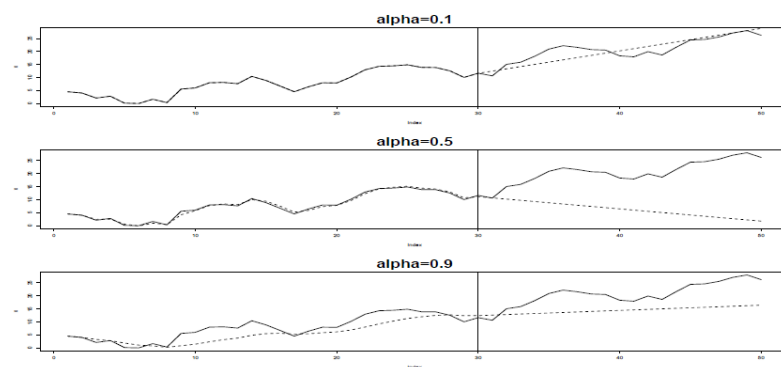


FIGURE 3.3 – Lissage et prévision par la méthode de lissage exponentiel double sur la série  $X(t) = 0.5t + \epsilon_t + 2 \cos(t\pi/6)$ ,  $\epsilon_t \sim N(0, 1)$ . ( $\alpha = .1, .5, .9$ )

Les formules analogues pour  $\hat{a}_2(n)$  sont :

$$\begin{aligned}\hat{a}_2(n) &= \frac{1-\alpha}{\alpha} [\alpha(1-\alpha)x_n + \alpha^2 L_1(n-1) - \alpha L_2(n-1)] \\ &= (1-\alpha)^2 x_n + \alpha(1-\alpha)L_1(n-1) - (1-\alpha)L_2(n-1) \\ &= (1-\alpha)^2 x_n - (1-\alpha)^2 \hat{a}_1(n-1) + (2\alpha - \alpha^2) \hat{a}_2(n-1)\end{aligned}$$

En utilisant (3.5) avec  $h = 1$  et  $n - 1$  comme instant de base, on obtient :

$$\hat{a}_1(n) = \hat{a}_1(n-1) + \hat{a}_2(n-1) + (1-\alpha^2) [x_n - \hat{x}_{n-1,1}] \quad (3.12)$$

$$\hat{a}_2(n) = \hat{a}_2(n-1) + (1-\alpha)^2 [x_n - \hat{x}_{n-1,1}] \quad (3.13)$$

On peut prendre comme valeurs initiales  $\hat{a}_1(0) = x_1, \hat{a}_2(0) = x_2 - x_1$ .

## 3.3 Méthode de Holt-Winters

### 3.3.1 Introduction

Cette approche a pour but d'améliorer et de généraliser le L.E.S. Nous étudions plusieurs cas particuliers de cette méthode :

- ajustement d'une droite affine (sans saisonnalité)
- ajustement d'une droite affine + une composante saisonnière
- ajustement d'une constante + une composante saisonnière

le dernier cas étant déduit du second. On ajuste au voisinage de l'instant  $n$  une droite

$$x_t = a_1 + (t - n)a_2$$

La variante proposée par Holt-Winters porte sur les formules de mise à jour dans l'estimation de  $a_1$  et  $a_2$  par les formules (3.12) et (3.13). La formule (3.14) s'interprète comme le barycentre de l'observation  $x_n$  et de la prévision à l'horizon 1 à la date  $n - 1$ . La formule (3.15) s'interprète comme le barycentre de la pente estimée à l'instant  $n - 1$  et la différence  $[\hat{a}_1(n) - \hat{a}_1(n-1)]$  entre les ordonnées à l'origine aux dates  $n$  et  $n - 1$ .

Soit  $0 < \beta < 1$  et  $0 < \gamma < 1$  deux constantes fixées et les formules de mise à jour

$$\hat{a}_1(n) = \beta x_n + (1-\beta) [\hat{a}_1(n-1) + \hat{a}_2(n-1)] \quad (3.14)$$

$$\hat{a}_2(n) = \gamma [\hat{a}_1(n) - \hat{a}_1(n-1)] + (1-\gamma) \hat{a}_2(n-1) \quad (3.15)$$

La prévision prend la forme :

$$\hat{x}_{n,h} = \hat{a}_1(n) + h\hat{a}_2(n) \quad (3.16)$$

Cette méthode est plus souple que le lissage exponentiel amélioré, dans la mesure où elle fait intervenir deux constantes  $\beta$  et  $\gamma$  au lieu d'une  $\alpha$ . Les équations (3.12) et (3.13), qui sont un cas particulier des formules (3.14) et (3.15) peuvent s'écrire :

$$\begin{aligned}\hat{a}_1(n) &= \alpha^2 [\hat{a}_1(n-1) + \hat{a}_2(n-1)] \\ &\quad + (1 - \alpha^2) [x_n - \hat{x}_{n-1,1} + \hat{a}_1(n-1) + \hat{a}_2(n-1)] \\ &= \alpha^2 [\hat{a}_1(n-1) + \hat{a}_2(n-1)] + (1 - \alpha^2) x_n \\ \hat{a}_2(n) &= \hat{a}_2(n-1) + \frac{(1 - \alpha)^2}{1 - \alpha^2} [\hat{a}_1(n) - \alpha^2 [\hat{a}_1(n-1) + \hat{a}_2(n-1)] \\ &\quad - (1 - \alpha^2) [\hat{a}_1(n-1) + \hat{a}_2(n-1)]] \\ &= \hat{a}_2(n-1) + \frac{(1 - \alpha)^2}{1 - \alpha^2} [\hat{a}_1(n) - \hat{a}_1(n-1) - \hat{a}_2(n-1)]\end{aligned}$$

qui sont identiques à (3.14) et (3.15) si  $\beta = 1 - \alpha^2$  et  $\gamma = \frac{1-\alpha}{1+\alpha}$ .

On remarque que si  $\beta$  et  $\gamma$  sont petits, le lissage est fort puisqu'alors  $\alpha$  est grand et que l'on tient compte du passé lointain.

### 3.3.2 Ajout d'une composante saisonnière. Décomposition additive

On ajuste au voisinage de  $n$ , la chronique

$$x_t = a_1 + (t - n)a_2 + s_t$$

où  $s_t$  est une composante périodique de période  $T$ . Les formules récursives de mise à jour sont : soit  $0 < \beta < 1$ ,

$$\begin{aligned}\hat{a}_1(n) &= \beta (x_n - \hat{s}_{n-T}) + (1 - \beta) [\hat{a}_1(n-1) + \hat{a}_2(n-1)] \\ \hat{a}_2(n) &= \gamma [\hat{a}_1(n) - \hat{a}_1(n-1)] + (1 - \gamma) \hat{a}_2(n-1) \\ \hat{s}_n &= \delta [x_n - \hat{a}_1(n)] + (1 - \delta) \hat{s}_{n-T}\end{aligned}$$

La prévision prend la forme :

$$\begin{aligned}\hat{x}_{n,h} &= \hat{a}_1(n) + h\hat{a}_2(n) + \hat{s}_{n+h-T} \quad 1 \leq h \leq T \\ \hat{x}_{n,h} &= \hat{a}_1(n) + h\hat{a}_2(n) + \hat{s}_{n+h-2T} \quad T + 1 \leq h \leq 2T\end{aligned}$$

et ainsi de suite pour  $2T < h$ . Dans cette procédure la difficulté de choisir les constantes de lissage est encore plus grande que pour les autres situations.

**Initialisation** : Les valeurs initiales suivantes permettent (mais on pourrait en proposer

d'autres) de commencer le calcul récursif des prévisions à partir de l'instant  $T + 2$  :

$$\begin{aligned}\hat{a}_1(T + 1) &= x_{T+1} \\ \hat{a}_2(T + 1) &= \frac{x_{T+1} - x_1}{T} \\ \hat{s}_j &= x_j - (x_1 + (T - 1)\hat{a}_2(T + 1)) \quad \text{pour } j \in \{1, \dots, T\}\end{aligned}$$

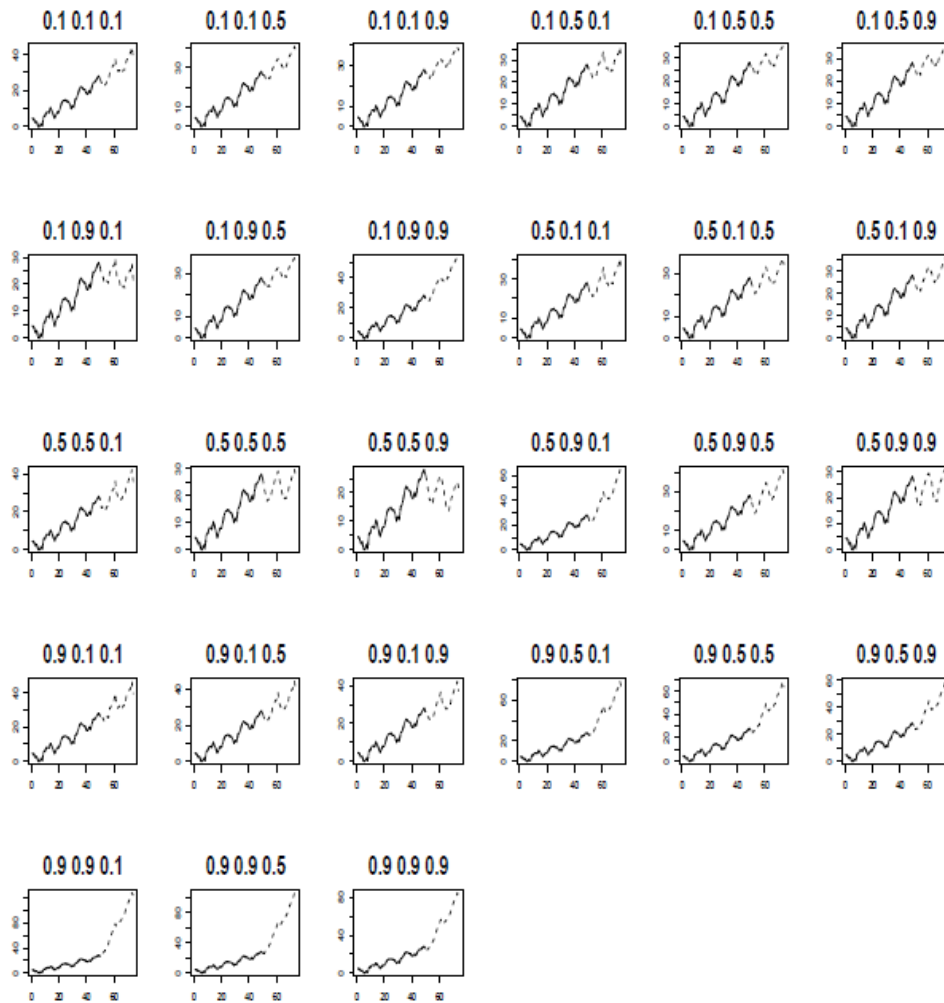


FIGURE 3.4 – Lissage et prévision par la méthode de Holt-Winters sur la série  $x(t) = 0.5t + \epsilon_t + 2\cos(t\pi/6)$ ,  $\epsilon_t \sim N(0, 1)$

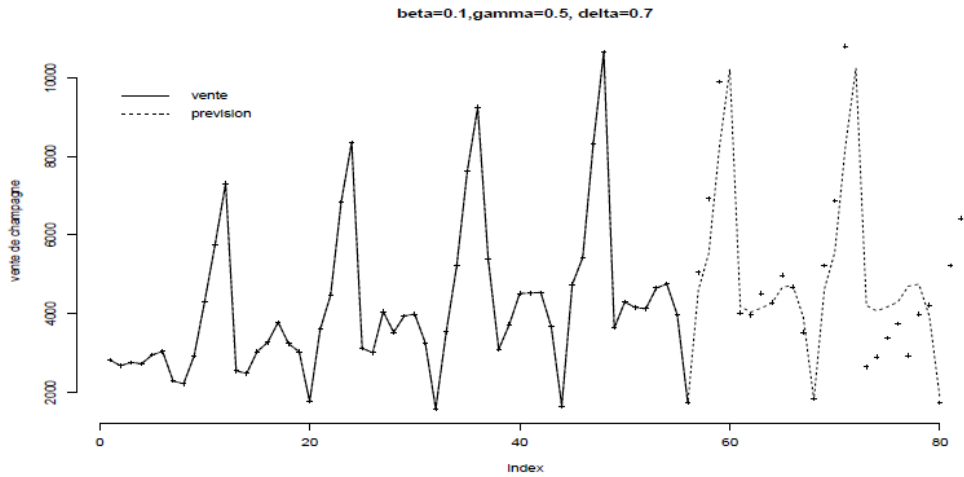


FIGURE 3.5 – Prédiction par la méthode de Holt-Winters sur les données de vente de champagne

### 3.4 TP : Lissage Exponentiel

#### 3.4.1 Lissage et prévision de données simulées

Simuler les trois séries temporelles suivantes (cf. TP1), de taille 100 :

- $X_1(t) = \epsilon_t$
- $X_2(t) = 0.5t + 2\epsilon_t$
- $X_3(t) = 0.5t + \epsilon_t + 3 \cos\left(t\frac{\pi}{6}\right)$

où  $\epsilon_t \sim \mathcal{N}(0, 1)$ . Les 70 premières données de chaque série vont être utilisées pour prédire les 30 dernières.

1. Pour chaque série, effectuer la prévision par lissage exponentiel simple et double. Tester différentes valeurs du paramètre de lissage  $\alpha$  (4 à 5 valeurs), et représenter graphiquement la série ainsi que la prévision. Commenter chaque résultat, et essayer de déterminer graphiquement le lissage le plus adaptée pour chaque série.
2. Calculer pour chaque prévision effectuée la somme des carrés des erreurs, et sélectionner le meilleur modèle à l'aide de cette quantité .
3. Tester maintenant le lissage exponentiel de Holt-Winters avec composante saisonnière additive puis multiplicative.
4. Les prédictions obtenues sont-elles meilleures ?

Rq : la commande par (`mfrOW = c(n, p)`) permet d'afficher  $n \times p$  graphique sur la même page.

### 3.4.2 Lissage et prévision de la concentration en $CO_2$

Le fichier de données  $CO_2$  contenu dans  $R$  contient les concentrations en  $CO_2$  à proximité du volcan Mauna Loa (Hawaï) de 1959 à 1997.

Après avoir représenté graphiquement ces données, quel modèle de lissage exponentiel vous semble le mieux approprié ?

Afin de valider ce modèle, tester la prédiction des données de 1990 à 1997 en utilisant celles de 1959 à 1989. Si cela vous semble graphiquement correct, utilisez cette méthode pour prédire les concentrations en  $CO_2$  de 1997 à 2007. Sinon, tester d'autres méthodes de lissage exponentiel.

### 3.4.3 Lissage et prévision du $CAC40$

Récupérer le fichier contenant les valeurs de clôture journalière du  $CAC40$  de 1991 à 1998 (données R `EuStockMarkets`). Essayer de prédire par lissage exponentiel les valeurs de clôture de 1998 en utilisant les valeurs de clôture de 1991 à 1997.

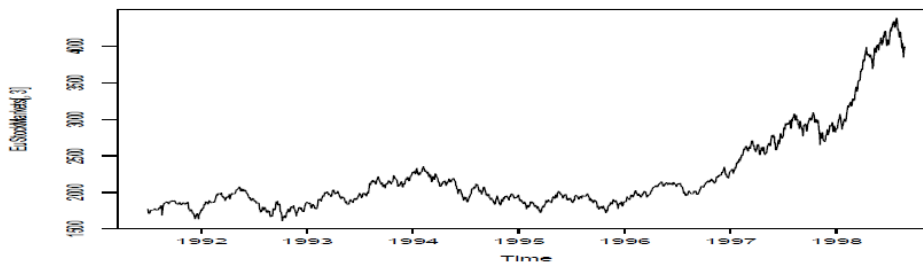


FIGURE 3.6 – Valeurs de clôture journalières du  $CAC40$  de 1991 à 1998

## Chapitre 4

# Comparaison entre la méthode de Box-Jenkins et la méthode de Lissage Exponentiel

L'objectif principale dans ce chapitre est la comparaison des deux méthodes de prévisions étudiés dans les deux chapitres précédents 2 et 3 et l'illustration à travers un exemple concret.

Pour cela nous allons pris en considération la série chronologique journalier du taux de change monétaire d'un \$ Américain contre le Dinar Algérien. Nous avons 61 observations du 01 Mai 2021 au 30 Juin 2021.

Comme logiciel, on utilise R. R est un langage de programmation et logiciel, gratuit et open source, spéciale et très puissant dans les études statistiques.

Utilisant 50 observations pour l'estimation et on garde 11 observations pour les comparer avec les valeurs que nous allons prédire.

La visualisation de notre série avec la commande plot donne le graphe suivant : La com-

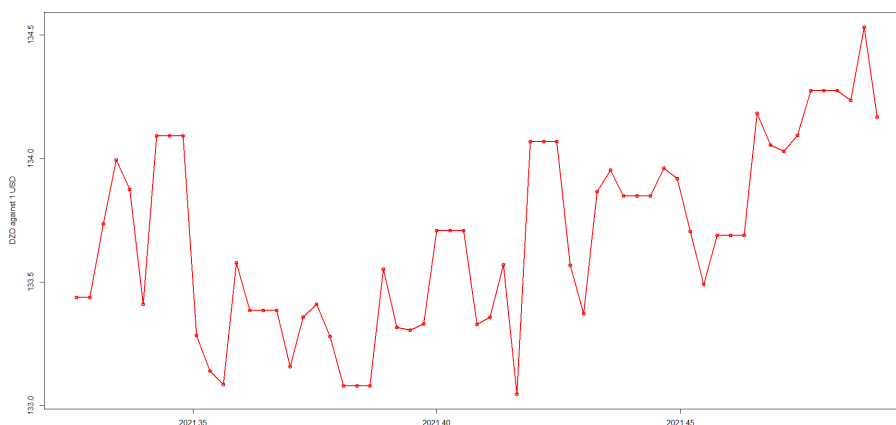


FIGURE 4.1 – Taux de change de 1\$ contre le Dinar Algérien

mande summary dans R donne un résumé sur notre série Pour mieux comprendre on

Min	$Q_1$	Median	Mean	$Q_3$	Max
133.0	133.3	133.6	133.6	133.8	134.1

TABLE 4.1 – Résumé de notre série

exécute la commande boxplot qui donne la figure suivante

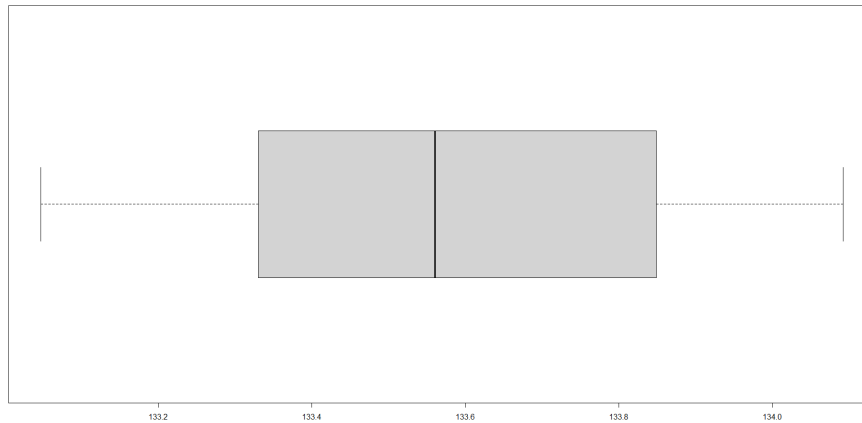


FIGURE 4.2 – Taux de change de 1\$ contre le Dinar Algérien

## 4.1 Application de la méthode BOX-JENKINS

Les travaux de *G.BOX* et *G.JENKINS* (dans les années soixante dix) ont beaucoup contribué dans la théorie et la pratique des modèles des série temporelles, L'objectif auquel ils se proposent de répondre dans leur ouvrage. "Time Series Analysis, Forecasting and Control", est de construire un modèle aléatoire de type *ARMA* permettant de reproduire au mieux les réalisations d'une série temporelle. Trois étapes sont nécessaires à la construction des modèles : une phase d'identification, une phase d'estimation, une phase de validation. A l'issue de ces trois phases, une fois déterminé le meilleur modèle *ARMA*, on utilise ce modèle à des fins de prévision.

La commande `auto.arima` dans R nous donne un *ARIMA*(2, 0, 2) comme modèle auto-régressif approprié.

```
arima(x , order = c(2, 0, 2))
```

Coefficients :

```
ar1 -0.1256 ar2 -0.3385 ma1 0.7121 ma2 0.8664 intercept 133.5801
```

$\sigma^2$  estimated as 0.06053

log likelihood = -1.86, aic = 15.72

Training set error measures :

ME -0.00188616 RMSE 0.2460309 MAE 0.1918204 MPE -0.001761065  
 MAPE 0.1435942 MASE 0.9609312

Le graphe de la série incluant les valeurs prévisionnelles selon le modèle ainsi construit

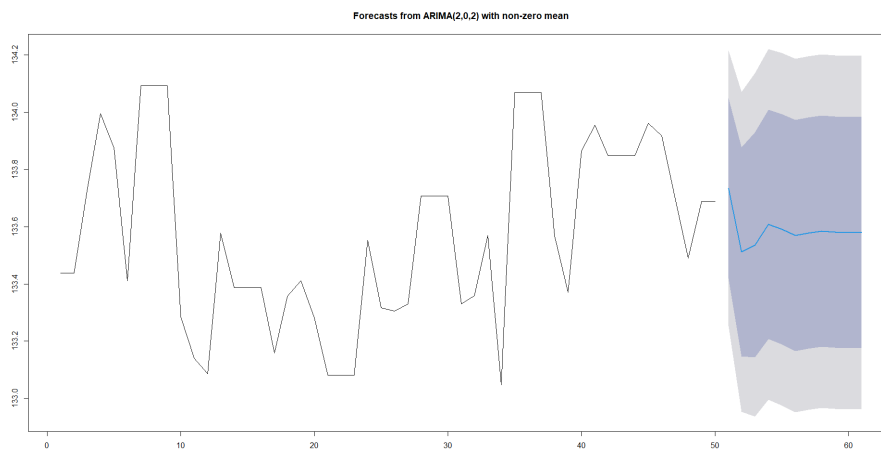


FIGURE 4.3 – Les valeurs prévisionnelles selon ARIMA(2,0,2)

## 4.2 Application de la méthode de lissage exponentiel

Les méthodes de lissages exponentiels constituent un outil permettant de réaliser des prévisions à partir de l'observation d'une série temporelle. Ces méthodes étant relativement basiques et simples de mise en oeuvre, elles sont souvent utilisées dans l'industrie, notamment lorsque le nombre de prévisions à réaliser est important (par exemple, prévisions des ventes de centaines de produits dans une grande surface).

Nous présentons trois types de lissage exponentiel :

- le lissage exponentiel simple qui consiste à ajuster localement à la série temporelle une constante.
- le lissage exponentiel double qui ajuste quant à lui une droite.

La commande ets dans R nous donne le modèle de lissage exponentiel approprié suivant.  
 ets(y , model = "ANN")

Smoothing parameters :

alpha = 0.3581

Initial states :

l = 133.5905

sigma : 0.2914

AIC 76.25517    AICc 76.77691    BIC 81.99124

Training set error measures :

ME 0.005877789    RMSE 0.2855153    MAE 0.2218967    MPE 0.004023178  
MAPE 0.1660999    MASE 1.111599

Le graphe de la série incluant les valeurs prévisionnelles selon le modèle ainsi construit

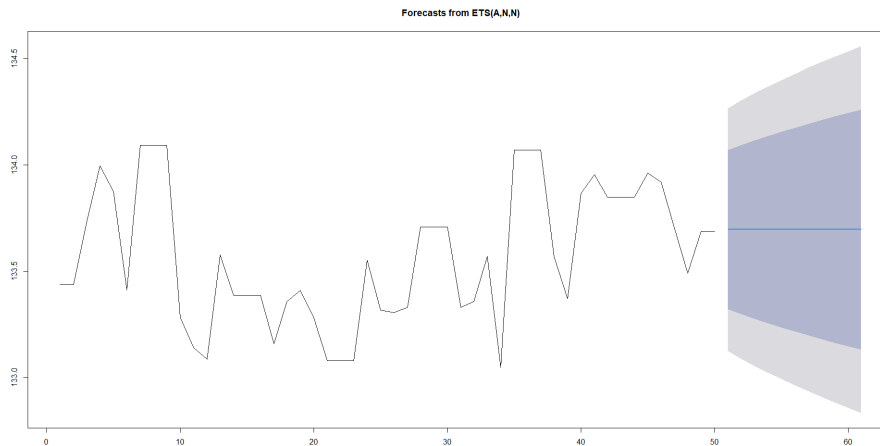


FIGURE 4.4 – Les valeurs prévisionnelles selon ARIMA(2,0,2)

### 4.3 Comparaison Entre les Résultats

On peut comparer les deux modèles selon le critère MSE (Mean square error), nous voyons que le MSE du modèle ARIMA est plus petit que celle du modèle de lissage exponentiel, ce qui nous laisse penser que le modèle ARIMA est meilleur ici.

# Conclusion

Les modèles étudiés dans ce mémoire sont très puissants et possèdent de très bonnes propriétés. Il est clair que les modèles auto-régressifs sont appropriés pour les prévisions à court terme bien qu'il existe d'autres familles de modèles telle que les modèles NN (réseau neurone) et les modèles des algorithmes génétiques.

# Bibliographie

- [1] Séries Chronologiques Agnès Lagnoux lagnoux@univ-tlse2.fr
- [2] Gardner E.S Jr. Exponential Smoothing : The State of the Art. Journal of Forecasting, 4(1) :1-28, 1985.
- [3] cours de s'eries chronologique Serge D'egerine 4 janvier 2005 univ PERRE MEN-DÉS FRANCE
- [4] Modèles de prévision Séries temporelles UQAM, ACT6420, Hiver 2011
- [5] Brockwell P.J. et Davis R.A. Introduction to Time Series and Forecasting, Springer, 2001.
- [6] Modèles de prévision Séries temporelles Arthur Charpentier 1 UQAM, ACT6420, Hiver 2011
- [7] introduction aux séries chronologiques O. ROUSTANT Novembre 2008
- [8] Les Prévisions tirées des processus ARMA(p,q) Gilbert Colletaz 4 décembre 2018
- [9] Renforcement Statistique Séries chronologiques Agnès Lagnoux lagnoux@univ-tlse2.fr
- [10] Gourieroux C.- Monfort A. Cours de series temporelles. Economica, 1983.
- [11] Bernard Rapacchi, ( Dur, les série chronologique ) Center Interuniversitaire de calcul de Grenoble, 18 aout 1993
- [12] Bensaber A. et Bleuse-Trillon B. Pratique des chroniques et de la prévision à court terme, Masson, Paris, 1989.
- [13] Aragon, Y. Séries temporelles avec R. Méthodes et cas., Springer, Paris, 2011.