

Département de Mathématiques

Master Mathématique et Application au Calcul Scientifique (MACS)

Mémoire de Fin d'Études

En vue de l'obtention du diplôme de **Master Sciences et Techniques**
(MST)

Sous titre :

SÉRIES CHRONOLOGIQUES UNIVARIÉES-APPLICATION SOUS LE LOGICIEL R

Réalisé par :

Lamiae SEDDATI

Dirigé par :

Pr. Ezzaki Fatima **Faculté des Sciences et Techniques Fès** **Encadrante**
Pr. Boutahar Mohamed **Université Aix-Marseille** **Co-encadrant**

Soutenu le 17 Avril 2019 devant le jury composé de :

Pr. Hilali Abdelmajid **Faculté des Sciences et Techniques Fès**
Pr. Ammor Ouafae **Faculté des Sciences et Techniques Fès**
Pr. Rahmouni Hassani Aziza **Faculté des Sciences et Techniques Fès**
Pr. Ezzaki Fatima **Faculté des Sciences et Techniques Fès**
Pr. Boutahar Mohamed **Université Aix-Marseille**

Année Universitaire : 2018-2019

FACULTE DES SCIENCES ET TECHNIQUES FES – SAISS
B.P. 2202 – Route d'Imouzzer – FES

REMERCIEMENT

Louange à Dieu le tout puissant d'avoir mis sur mon chemin plusieurs personnes dont l'intervention a constitué une aide considérable pour l'élaboration de ce projet.

Je suis digne à mes parents qui m'ont toujours encouragée et aidée dans mes études. Ils ont su me donner toutes les chances pour réussir. J'espère qu'ils trouvent, dans la réalisation de ce travail, l'aboutissement de leurs efforts ainsi que l'expression de ma plus affectueuse gratitude.

Je voudrais d'abord exprimer mes profonds remerciements et ma reconnaissance à, Mme *ELZAKI* Fatima, l'enseignante-chercheur de haut niveau qui, par ses conseils valeureux, ses remarques perspicaces, son temps précieux bref, pour son meilleur encadrement m'a soutenue dans les différentes étapes de la réalisation de ce projet.

Mes sincères remerciements s'adressent à M. *BOUCHEB* Mohamed, mon co-encadrant qui par son expertise, ses conseils valeureux, ses remarques pertinentes et ses commentaires judicieux, m'a aidé pour l'élaboration de ce projet.

J'adresse ensuite mes remerciements les plus sincères aux membres du jury respectable : Pr. Hilali Abdelmajid, Pr. Ammor Ouafae, Pr. Rahmouni Hassani Aziza, Pr. Ezzaki Fatima d'avoir bien voulu participer à l'évaluation de ce travail .

Je remercie enfin toutes les personnes qui ont participé de près ou de loin à la réussite de ce projet.

TABLE DES MATIÈRES

0.1	Un peu d'histoire	3
0.2	Définition d'une série chronologique	5
0.3	Objectifs principaux d'étudier les séries chronologiques	5
0.4	Domaines d'application des séries chronologiques	6
1	Description et décomposition d'une série chronologique	9
1.1	Description d'une série chronologique :	9
1.1.1	Correction des données	10
1.1.2	Observation de la série	10
1.1.3	Modélisation	10
1.1.4	Analyse de la série à partir de ses composantes	11
1.1.5	Diagnostic du modèle - ajustement au modèle	11
1.1.6	Prédiction	12
1.2	Modélisation d'une série chronologique	12
1.2.1	Modèle additif	12
1.2.2	Modèle multiplicatif	13
1.2.3	Modèle mixte	14
1.2.4	Choix du modèle	14
2	Étude de la partie déterministe	15
2.1	Analyse de la tendance	15
2.1.1	Ajustement tendanciel linéaire par la méthode des moindres carrés	15
2.1.2	Ajustement tendanciel linéaire par points médians	16
2.1.3	Ajustement tendanciels non linéaires	17
2.1.4	Élimination de la tendance par différenciation	17
2.2	Les moyennes mobiles	17
2.2.1	Définitions des moyennes mobiles	18
2.2.2	Propriétés d'un lissage par moyenne mobile	20
2.3	Prévision par lissage exponentiel	23
2.3.1	Lissage exponentiel simple(LES)	23
2.3.2	Lissage exponentiel double(LED)	24
2.4	L'étude d'une série chronologique	27
2.4.1	Application d'une moyenne mobile d'ordre judicieusement choisi	28

2.4.2	Estimation de la saisonnalité	28
2.4.3	Estimation de la tendance	29
2.4.4	Itération éventuelle de la procédure	29
2.4.5	Prévision des valeurs futures	29
2.4.6	Analyse des résidus	30
2.5	Application sous le logiciel R	31
3	Généralités et Rappels sur les processus stochastiques	35
3.1	Espace L^2	35
3.2	La régression affine	37
3.3	Notions de Convergence	38
3.3.1	Convergence en moyenne quadratique	38
3.3.2	Convergence en loi	39
3.4	processus stochastique	39
3.4.1	Premiers exemples de séries temporelles	40
3.5	Processus stationnaire du second ordre	41
3.5.1	Fonctions d'autocovariance et d'autocorrélation	43
4	Processus aléatoires stationnaires	49
4.1	Théorème de Wold	49
4.1.1	Conditions sur les pondérations ψ_i	50
4.1.2	Prévisions à partir de la décomposition de Wold	51
4.2	Processus auto-régressifs d'ordre p	52
4.2.1	Définition et représentation canonique	53
4.2.2	Prédiction d'un $AR(p)$	58
4.2.3	Intervalle de confiance pour la prévision	59
4.2.4	Exemple sous le logiciel R	59
4.3	Processus moyennes mobiles d'ordre q	63
4.3.1	Définition et représentation canonique	63
4.3.2	Prédiction d'un $MA(q)$ /Intervalle de confiance pour la prévision	66
4.3.3	Exemple sous le logiciel R	66
4.4	Les processus mixtes ARMA	69
4.4.1	Définition d'un $ARMA(p,q)$	70
4.4.2	Écritures $MA(\infty)$ et $AR(\infty)$	71
4.4.3	Exemple sous le logiciel R	73
5	Processus non stationnaires	79
5.1	Les processus TS	80
5.2	Les processus DS	81
5.2.1	Processus ARIMA	84
5.2.2	Processus SARIMA	85
5.3	Construction d'un système de prévision	86
5.3.1	Identification du modèle	86
5.3.2	Estimation du modèle	87
5.3.3	Le critère d'information d'akaike (AIC)	90
5.3.4	Diagnostic du modèle/ Validation du modèle	90
5.4	Simulation sous le logiciel R	93
	Liste des notations	107

TABLE DES FIGURES

1	Exemples des séries chronologiques	6
1.1	Modèle additif. Amplitude constante autour de la tendance	13
1.2	Modèle multiplicatif. L'amplitude proportionnelle à la tendance	13
2.1	Les Modèles de tendance	16
2.2	Lissage exponentiel	25
2.3	Chronogramme des ventes d'un produit	32
2.4	Chronogramme des ventes d'un produit/ Méthode de la bande	32
2.5	la série filtrée et la tendance des ventes d'un produit	33
2.6	Décomposition du chronogramme des ventes d'un produit	34
3.1	Simulation de la convergence en loi.	40
3.2	Illustration du théorème central limite.	40
3.3	Simulation selon un modèle de régression simple avec bruit gaussien.	41
3.4	La fonction d'autocorrélation empirique d'un bruit blanc	47
3.5	La fonction d'autocorrélation empirique partielle d'un bruit blanc	47
4.1	Graphe de trajectoire.	60
4.2	Corrélogramme du processus auto-régressif d'ordre 2.	60
4.3	Corrélogramme partiel du processus auto-régressif d'ordre 2.	60
4.4	Graphe de trajectoire 1 de $AR(1)$	61
4.5	Représentation du corrélogramme 1 de $AR(1)$	61
4.6	Représentation du corrélogramme partielle 1 de $AR(1)$	61
4.7	Graphe de trajectoire 2 de $AR(1)$	62
4.8	Représentation du corrélogramme 2 de $AR(1)$	62
4.9	Représentation du corrélogramme partielle 2 de $AR(1)$	62
4.10	Graphe original de la fonction moyenne mobile.	67
4.11	Graphe de trajectoire.	67
4.12	Corrélogramme du processus moyenne mobile d'ordre 3.	67
4.13	Corrélogramme partiel du processus moyenne mobile d'ordre 3.	67
4.14	Graphe de trajectoire 1 de $MA(1)$	68
4.15	Représentation du corrélogramme 1 de $MA(1)$	68
4.16	Représentation du corrélogramme partielle 1 de $MA(1)$	68

4.17	Graphe de trajectoire 2 de $MA(1)$	69
4.18	Représentation du corrélogramme 2 de $MA(1)$	69
4.19	Représentation du corrélogramme partielle 2 de $MA(1)$	69
4.20	Graphe de trajectoire 1 de $ARMA(1, 1)$	74
4.21	Graphe de trajectoire 2 de $ARMA(1, 1)$	74
4.22	Graphe de trajectoire 3 de $ARMA(1, 1)$	74
4.23	Graphe du variogramme 1 de $ARMA(1, 1)$	75
4.24	Graphe du variogramme 2 de $ARMA(1, 1)$	75
4.25	Graphe du variogramme 3 de $ARMA(1, 1)$	75
4.26	Graphe du corrélogramme 1 de $ARMA(1, 1)$	76
4.27	Graphe du corrélogramme 2 de $ARMA(1, 1)$	76
4.28	Graphe du corrélogramme 3 de $ARMA(1, 1)$	76
4.29	Graphe du corrélogramme partielle 1 de $ARMA(1, 1)$	77
4.30	Graphe du corrélogramme partielle 2 de $ARMA(1, 1)$	77
4.31	Graphe du corrélogramme partielle 3 de $ARMA(1, 1)$	77
5.1	Graphe du marche aléatoire	83
5.2	Graphe du marche aléatoire différencier	84
5.3	Chronogramme de la température d’Oran	95
5.4	Corrélogramme de la température d’Oran	95
5.5	Corrélogramme partielle de la température d’Oran	95
5.6	Chronogramme de la température d’Oran après la différenciation	96
5.7	Corrélogramme de la température d’Oran après la différenciation	96
5.8	Corrélogramme partielle de la température d’Oran après la différenciation	96
5.9	La prévision de la température de l’année 2019 d’Oran	97
5.10	Chronogramme de la série ciment	98
5.11	Corrélogramme de la série ciment	98
5.12	Corrélogramme partielle de la série ciment	99
5.13	Chronogramme de la série ciment après la différenciation	99
5.14	Corrélogramme de la série ciment après la différenciation	100
5.15	Corrélogramme partielle de la série ciment après la différenciation	100
5.16	Chronogramme des indices boursiers	101
5.17	Corrélogramme des indices boursiers	101
5.18	Corrélogramme partielle des indices boursiers	101
5.19	La prévision des indices boursiers	103
5.20	la série CAC40 + La prévision par lissage exponentiel des indices boursiers	103
5.21	La prévision par lissage exponentiel des indices boursiers	104

AVANT-PROPOS

L'étude de l'évolution temporelle d'une ou plusieurs variables, depuis longtemps, a intéressé des spécialistes de domaines très variés. Les probabilistes ont construit une importante théorie mathématique, la théorie des processus stochastiques, qui permet d'analyser avec un très grand degré de finesse les propriétés d'un ensemble de variables aléatoires indicées par le temps. Les théoriciens de la statistique ont proposé diverses approches des problèmes d'estimation, de tests, de prévision et de filtrage pour les processus stochastiques. Des praticiens appartenant à des domaines très variés (physique, automatique, économie,...) ont développé des techniques adaptées à leurs problèmes spécifiques et ont ainsi créé une grande gamme d'outils d'analyse qui n'est pas toujours reliée très clairement à la littérature théorique.

L'analyse d'une série chronologique consiste à mettre en évidence et à classer les facteurs qu'influent sur la grandeur considérée. Ces mouvements peuvent être de plus ou moins long terme. La périodicité d'une série chronologique est la durée qui sépare deux observations. L'étude d'une série chronologique exige que les relevés successifs soient comparables.

Nous avons choisi le logiciel **R** pour étudier les séries chronologiques, à cause de plusieurs raisons, parmi eux, R possède :

- un système efficace de manipulation et de stockage des données.
- différents opérateurs pour le calcul sur tableaux (et spécialement les matrices).
- un grand nombre d'outils pour l'analyse des données et les méthodes statistiques.
- un langage de programmation simple et performant comportant : conditions, boucles, moyens d'entrées sorties, possibilité de définir des fonctions récursives.
- des moyens graphiques pour visualiser les analyses.

Les outils de visualisation de série, nombreux dans **R**, permettent de comprendre la structure d'une série avant toute modélisation, ils sont présentés dès le chapitre 1 et utilisés systématiquement dans l'étude des séries. Les graphiques sont importants à toutes les étapes du traitement. Avant modélisation, ils aident à saisir la structure de la série, après modélisation, ils offrent une vision globale de l'ajustement, vision que ne peut donner un niveau de signification empirique considéré isolément.

Venons à une description plus détaillée du contenu. Ce rapport comprend six

chapitres.

D'un chapitre introductif qui est consacré à une présentation générale des séries chronologiques. Le premier chapitre montre, de façon détaillée, l'analyse des différentes composantes d'une série chronologique et le choix du modèle. Le deuxième chapitre complète l'analyse et l'étude de la partie déterministe d'une série chronologique en présentant l'analyse de la tendance, les moyennes mobiles, les facteurs saisonniers, le lissage exponentiel et se terminant par un exemple d'application sous le logiciel R. Les théorèmes et les notions nécessaires à l'étude des processus aléatoires stationnaires et non stationnaires sont décrites dans le chapitre 3. Le chapitre 4 présente le théorème de Wold, le théorème fondamental de l'analyse des séries chronologiques stationnaires et les processus $AR(p)$, $MA(q)$, $ARMA(p, q)$ qui forment une classe flexible de modèles pour de nombreux phénomènes observés. Le chapitre 5 est consacré d'une part à l'étude des processus non stationnaires TS et DS , et d'autre part à la construction d'un système de prévision.

Finalement, on va faire une étude complète des séries chronologiques basée sur des données réelles.

CHAPITRE INTRODUCTIF

L'étude des séries temporelles, ou séries chronologiques, correspond à l'analyse statistique d'observations régulièrement espacées dans le temps. Elles ont été utilisées en astronomie, en météorologie, en théorie du signal, en biologie, en économie ...etc.

0.1 Un peu d'histoire

Historiquement, c'est en astronomie qu'apparaissent les premières séries chronologiques constituées volontairement à des fins d'analyse. L'observation systématique du ciel remonte à l'antiquité.

- En 1662 **John Graunt** analyse de façon informelle des séries chronologiques relatives à la population et à la mortalité.

- 1760/1780 : **Johann Heinri Lambert** (mathématicien et philosophe suisse, 1728-1777) publie des graphiques d'une grande qualité pour toutes sortes de séries chronologiques, par exemple, son graphique (publié en 1779) de la variation de la température du sol en relation avec la profondeur sous la surface.

- 1772 : Méthode mathématique pour détecter les périodicités (Lagrange, 1772).

Louis de Lagrange(astronome et mathématicien français, 1736-1813) propose une méthode mathématique pour mesurer les périodicités dans les données astronomiques.

- 1807 : Analyse harmonique.

Joseph Fourier(mathématicien français, 1768-1830) propose en 1807 que toute fonction périodique peut s'écrire comme une somme(ou série)de fonctions sinusoïdales et cosinusoidales.

- 1825 : Courbe de Gompertz.

Benjamin Gompertz propose une courbe de tendance non linéaire pour graduer les tables de mortalité.

- 1844 : Courbe logistique (**Verhulst**) est une courbe de tendance non linéaire, elle est très utilisée dans les études de population et en biologie.

- 1847 : Méthode de **Buys-Ballot** pour l'analyse des périodicités. Buys-Ballot(météorologue néerlandais, 1817-1890)propose en 1847 une méthode pour analyser les phénomènes périodiques (saisonniers, notamment) dans les données météorologiques.

- 1876-78 : Analyseur harmonique de **Thomson**.

Thomson propose et construit un instrument permettant de calculer mécaniquement l'intégrale du produit de deux fonctions.

-
- 1879 : Méthode d'analyse harmonique de **Stewart** et **Dodgson**.
Stewart et Dodgson proposent une méthode plus générale que celle de Buys-Ballot pour l'analyse de périodicités dans les séries chronologiques.
 - 1884 : Moyennes mobiles. Corrélation sérielle (**Poynting**).
en 1884, le physicien Poynting propose l'utilisation de moyennes mobiles arithmétiques pour éliminer les variations accidentelles ou périodiques. Cette idée sera ensuite fortement utilisée dans le cadre des méthodes de désaisonnalisation.
 - 1901 : L'utilisation de moyennes mobiles pour éliminer les fluctuations accidentelles est discutée dans l'ouvrage de *Bowley Elements of Statistics* (1901).
 - 1904 : Moyenne mobile de **Spencer**, cette méthode a joué ensuite un rôle important dans le développement de méthodes de désaisonnalisation.
 - 1914 : Critère de **Walker** pour tester si un pic dans le périodogramme est statistiquement significatif.
 - 1914 : Utilisation de la différenciation pour éliminer la tendance d'une série chronologique(**Student, Moore**). - 1915 : Désaisonnalisation (**Copeland**).
Copeland(1915) étudie la séparation des fluctuations saisonnières des autres composantes d'une série chronologique.
 - 1917 : Première étude des propriétés harmoniques d'une série chronologique par le biais des auto-corrélations (**Clayton**).
 - 1919 : Approche standard de décomposition(**Persons**). En 1919, Persons propose une approche globale de décomposition d'une série chronologique en quatre composantes : $X_t = T_t + C_t + S_t + \varepsilon_t$ où est une composante tendancielle, C_t est une composante cyclique, S_t est une composante saisonnière et ε_t est une composante accidentelle (bruit).
 - 1926-27 : Processus de moyenne mobile (**Yule**, 1926 ; **Slutsky**, 1927).
Yule démontre que le fait de différencier un bruit aléatoire introduit des auto-corrélations artificielles. Dans la même veine, indépendamment de Yule, Slutsky montre qu'en calculant une moyenne mobile à partir d'un bruit blanc, on obtient une série dont les observations ne sont pas indépendantes et qui peut présenter des cycles apparents (« *effet Slutsky* »).
 - 1927 : Modèle auto-régressif (**Yule**) en 1927, Yule propose le modèle auto-régressif et trouve qu'un modèle auto-régressif d'ordre 2 représente assez bien le comportement des données de Wolfer sur les taches solaires.
 - 1929 : Test de **Fisher** pour la présence d'un pic dans le périodogramme.
 - 1929 : Propriétés du périodogramme (**Slutsky**). Slutsky étudie les propriétés du périodogramme pour des séries chronologiques auto-corrélées.
 - 1931 : Relation entre périodogramme et auto-corrélations (**Walker**).
Gilbert Walker étudie les propriétés harmoniques d'une série chronologique à partir des auto-corrélations d'une série chronologique. Pour un processus auto-régressif, il se sert des fameuses « *équations de Yule-Walker* »
 - 1938 : Théorème de décomposition de **Wold**.
 - 1939-42 : Théorie de la prévision optimale (**Kolmogorov, Wiener**).
 - 1969 : Test d'auto-corrélation basé sur le périodogramme (**Durbin**).
 - 1940-42 : Premières contributions à l'étude statistique de tests d'auto-corrélation pour des séries gaussiennes.
 - 1949-51 : Tests d'auto-corrélation dans les modèles de régression (**Durbin** et **Watson**).
 - 1954 : première version de la méthode de désaisonnalisation Census.
 - 1960 : Méthode du lissage exponentiel (**Winters, Brown**).
 - 1969 : Notion de causalité de **Granger**.
 - 1970 : Approche de **Box** et **Jenkins** pour l'ajustement des modèles *ARMA* et la prévision.
 - 1971 : Test de causalité (**Sims**).
 - 1981 : Généralisation de l'approche de Box et Jenkins à des processus vectoriels (**Tiao** et

Box).

- 1982 : Méthode généralisée des moments (**Hansen**).

- 1982 : La terminologie de **Nelson** et **Plosser** des processus non stationnaires.

0.2 Définition d'une série chronologique

Une série chronologique ou série temporelle (ou chronique) est une suite finie de données quantitatives indexée par le temps, c'est-à-dire une suite d'observations d'une famille de variables aléatoires réelles notées $(X_t)_{t \in \Gamma}$ ou $(X_t, t \in \Gamma)$ où l'ensemble Γ est appelé espace des temps qui peut être :

- **discret** (nombre de voyageurs SNCF quotidien, température maximale...). Dans ce cas, $\Gamma \subset \mathbb{Z}$, les dates d'observations sont le plus souvent équidistantes : par exemple relevés mensuels, trimestriels... Ces dates équidistantes sont alors indexées par des entiers : $t = 1, 2, \dots, T$ et T est le nombre d'observations. On dispose donc des observations des variables $\{X_1, X_2, \dots, X_T\}$ issues de la famille $(X_t)_{t \in \Gamma}$ où $\Gamma \subset \mathbb{Z}$ (le plus souvent $\Gamma = \mathbb{Z}$).

- **continu** (signal radio, résultat d'un électrocardiogramme...). L'indice de temps est à valeurs dans un intervalle de \mathbb{R} et on dispose (au moins potentiellement) d'une infinité d'observations issues d'un processus $(X_t)_{t \in \Gamma}$ où Γ est un intervalle de \mathbb{R} . Un tel processus est dit à temps continu. Les méthodes présentées dans ce cadre sont différentes de celles pour les séries chronologiques à temps discret.

Exemple : 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).

0.3 Objectifs principaux d'étudier les séries chronologiques

L'étude d'une série chronologique permet d'*analyser*, de *décrire* et d'*expliquer* un phénomène au cours du temps et d'en tirer des conséquences pour des prises de décision (marketing...).

Cette étude permet aussi de faire un *contrôle*, par exemple pour la gestion des stocks, le contrôle d'un processus chimique... 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} où $h = 1, 2, 3, \dots$ 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 $\widehat{X}_T(h)$, 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 $\widehat{X}_T(h) - X_{T+h}$ "en moyenne" avec l'idée que plus h est grand, plus l'erreur est grande. 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*, par exemple, d'un changement de politique (économique). Ces ruptures peuvent être de deux ordres : une rupture de niveau ou une rupture de pente. La prévision de ces dates de rupture est bien évidemment très importante.

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 $Y_t = a_1 X_{t+1} + a_2 X_{t+2}$

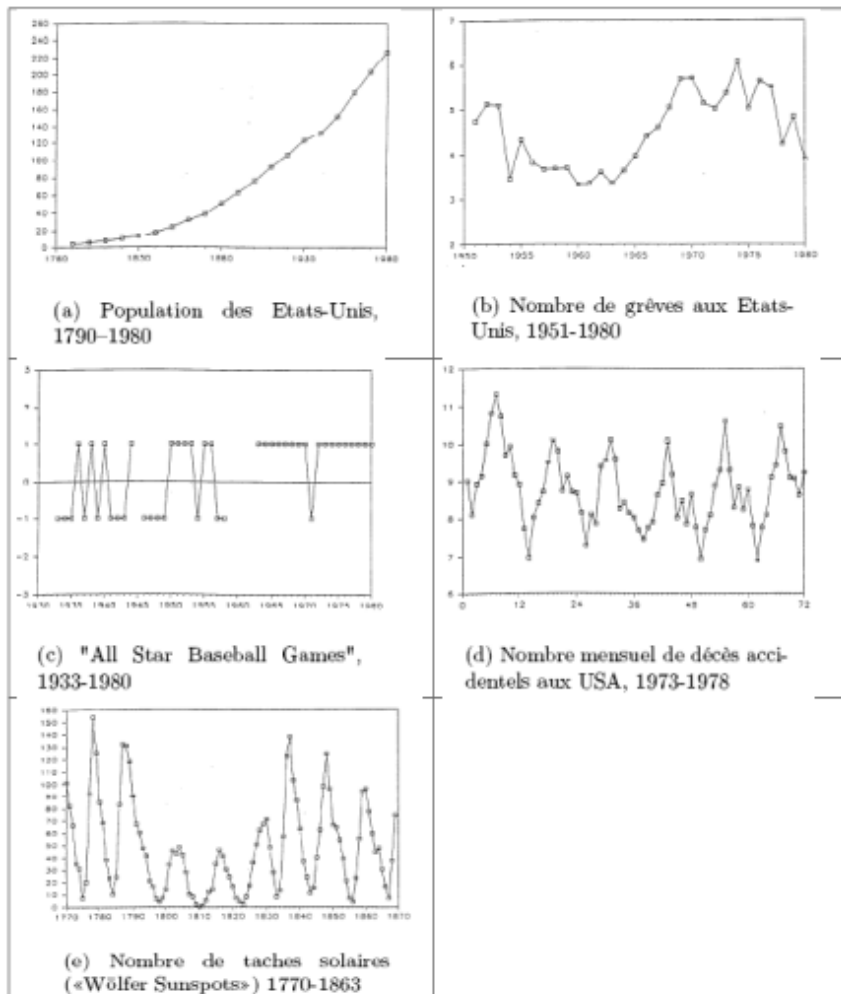


FIGURE 1 – Exemples des séries chronologiques

Ici, deux questions se posent : tout d'abord, la question de *la causalité* c'est-à-dire quelle variable va expliquer l'autre? 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 ?

0.4 Domaines d'application des séries chronologiques

On trouve des exemples de séries chronologiques uni-variées dans de très nombreux domaines. La liste suivante n'est qu'un échantillon :

- 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,...etc
- assurance : analyse des sinistres,..
- médecine / biologie : suivi des évolutions des pathologies, analyse d'électro-encéphalogrammes et d'électrocardiogrammes, ..
- sciences de la Terre et de l'espace : indices de marées, variations des phénomènes physiques (météorologie), évolution des taches solaires, phénomènes d'avalanches, ...
- traitement du signal : signaux de communications, de radars, de sonars, analyse de la parole, ..
- traitement des données : mesures successives de position ou de direction d'un objet mobile, ...

– métrologie : variation de phase ou de fréquence des oscillateurs, dérive et bruit des capteurs inertiels.

CHAPITRE

1

DESCRIPTION ET DÉCOMPOSITION D'UNE SÉRIE CHRONOLOGIQUE

1.1 Description d'une série chronologique :

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

- **la tendance** (ou **trend**) (T_t) représente l'évolution à long terme de la série étudiée. Elle traduit le comportement "moyen" de la série.
- **la composante saisonnière** (ou **saisonnalité**) (S_t) correspond à un phénomène qui se répète à intervalles de temps réguliers (périodiques). En général, c'est un phénomène saisonnier d'où le terme de variations saisonnières.

- **la composante résiduelle** (ou **bruit** ou **résidu**) (ε_t) correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire. On parle aussi d'aléas.

Les modèles présentés généralement tiennent compte de ces trois composantes (tendance, saisonnalité et fluctuations irrégulières). Il faut cependant remarquer que l'on pourrait envisager d'autres composantes.

- Des phénomènes accidentels (grèves, conditions météorologiques exceptionnelles, crash financier) peuvent notamment intervenir.

- Une autre composante parfois étudiée de manière spécifique a trait au phénomène cyclique : c'est souvent le cas en climatologie et en économie (exemple : récession et expansion...). Il s'agit d'un phénomène se répétant mais contrairement à la saisonnalité sur des durées qui ne sont pas fixes et généralement plus longues. Sans informations spécifiques, il est généralement très difficile de dissocier tendance et cycle.

Dans le cadre de ce projet, la composante correspondant aux phénomènes accidentels sera intégrée aux fluctuations irrégulières de la série et la composante tendance regroupera à la fois la tendance et le cycle.

Comme nous avons vu, l'un des objectifs principaux de l'étude d'une série chronologique est la prévision des valeurs futures de cette série. Pour cela, on a besoin de connaître ou au moins de modéliser le mécanisme de production de la série chronologique.

Notons que les variables X_t ne sont le plus souvent ni indépendantes ni identiquement distribuées (dans la plupart des cas, le phénomène évolue, se modifie au cours du temps ce qui entraîne que les variables le décrivant ne sont pas équidistribuées). Cela nécessite des méthodes statistiques de traitement et de modélisation spécifiques puisqu'en particulier dans un cadre standard les méthodes statistiques classiques sont basées sur des hypothèses d'indépendance.

Schématiquement, les principales étapes de traitement d'une série chronologique sont les suivantes :

1. correction des données.
2. observation de la série.
3. modélisation (avec un nombre fini de paramètres).
4. analyse de la série à partir de ses composantes.
5. diagnostic du modèle - ajustement au modèle.
6. prédiction.

1.1.1 Correction des données

Avant de se lancer dans l'étude d'une série chronologique, il est souvent nécessaire de traiter, modifier les données brutes. Par exemple :

- L'évaluation de données manquantes, remplacement de données accidentelles,...
- Découpage en sous-séries ;
- standardisation afin de se ramener à des intervalles de longueur fixe.

1.1.2 Observation de la série

Une règle générale en statistique descriptive consiste à commencer par regarder les données avant d'effectuer le moindre calcul. Ainsi, une fois la série corrigée et pré-traitée, on trace son graphique c'est-à-dire la courbe de coordonnées (t, X_t) .

1.1.3 Modélisation

Un modèle est une image simplifiée de la réalité qui vise à traduire les mécanismes de fonctionnement du phénomène étudié et permet de mieux les comprendre. On distingue principalement deux types de modèles :

Les modèles déterministes :

Ces modèles relèvent de la statistique descriptive. Ils ne font intervenir que de manière sous-jacente, le calcul des probabilités et consistent à supposer que l'observation de la série à la date t est une fonction du temps t et d'une variable ε_t centrée faisant office d'erreur au modèle, représentant la différence entre la réalité et le modèle proposé : $X_t = f(t, \varepsilon_t)$. On suppose de plus que les ε_t sont dé-corrélées. Les deux modèles de ce type les plus utilisés sont les suivants :

a- Modèle additif : C'est le "modèle classique de décomposition" dans le traitement des modèles d'ajustement. La variable X_t s'écrit comme la somme de trois termes : $X_t = T_t + S_t + \varepsilon_t$ où T_t représente la tendance (déterministe), S_t la saisonnalité (déterministe aussi) et ε_t les composantes ("erreurs au modèle") aléatoires.

b- Modèle multiplicatif : La variable X_t s'écrit au terme d'erreur près comme le produit de la tendance et d'une composante de saisonnalité : $X_t = T_t(1 + S_t)(1 + \varepsilon_t)$.

Les modèles stochastiques :

Ils sont du même type que les modèles déterministes à ceci près que les variables de bruit ε_t ne sont pas i.i.d mais possèdent une structure de corrélation non nulle : ε_t est une fonction des valeurs passées et d'un terme d'erreur $\varepsilon_t = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \eta_t)$.

La classe des modèles de ce type la plus fréquemment utilisée est la classe des modèles ARMA, ARIMA, SARIMA,...

1.1.4 Analyse de la série à partir de ses composantes

Une fois l'étape de modélisation effectuée, on étudie les composantes du modèle les unes après les autres.

La série corrigée de la tendance :

La tendance agit comme une forte corrélation entre les variables X_t mais cette corrélation n'exprime aucune liaison à caractère explicatif. Il s'agit donc d'isoler cette tendance puis de l'étudier à part et enfin de l'éliminer de la série pour voir si des liaisons à caractère explicatif existent et étudier seulement ces corrélations sans tendance. On définit la série corrigée de la tendance en supprimant la tendance. La série détendancialisée est :

- pour le modèle additif : $X'_{cvs,t} = S_t + \varepsilon_t$.
- pour le modèle multiplicatif : $X'_{cvs,t} = S_t(1 + \varepsilon_t)$.

La série corrigée des variations saisonnières :

Dans le même ordre d'idée, nous corrigerons les éventuelles variations saisonnières qui résultent d'un comportement périodique dans la série observée. Pour pouvoir réellement comparer les ventes d'un trimestre à l'autre, on doit donc supprimer l'effet de la saisonnalité et on définit la série corrigée des variations saisonnières $(X_{cvs,t})_t$ en supprimant la composante saisonnière $(S_t)_t$ du modèle. La série désaisonnalisée est :

- pour le modèle additif : $X_{cvs,t} = T_t + \varepsilon_t$.
- pour le modèle multiplicatif : $X_{cvs,t} = T_t(1 + \varepsilon_t)$.

La série lissée des prédictions :

On définit la série lissée des prédictions en supprimant les fluctuations irrégulières $(\varepsilon_t)_t$ du modèle. C'est à partir de cette série que nous ferons les prédictions en utilisant les modélisations, estimations de la tendance et de la saisonnalité. Par exemple, après avoir supprimé les fluctuations irrégulières, on obtient :

- pour le modèle additif : $X_{SLP,t} = T_t + S_t$.
- pour le modèle multiplicatif : $X_{SLP,t} = T_t(1 + S_t)$.

1.1.5 Diagnostic du modèle - ajustement au modèle

Une fois le modèle construit et ses paramètres estimés, on vérifie que le modèle proposé est bon c'est-à-dire l'ajustement au modèle :

- en étudiant les résidus.
- en faisant des tests,...

1.1.6 Prédiction

Enfin, une fois ces différentes étapes réalisées, nous sommes en mesure de faire de la prédiction.

1.2 Modélisation d'une série chronologique

Ce qui rend l'étude d'une série chronologique complexe c'est la non indépendance entre les variables aléatoires X_t . Afin de pouvoir traiter une telle série, il n'y a d'autre possibilité que celle d'imposer un modèle de la forme : $X_t = f(t, \varepsilon_t)$ où $t \in \mathbb{Z}$ et ε_t est une variable aléatoire (v.a) centrée qui représente la composante d'erreur d'une série chronologique et $f(\cdot)$ fonction inconnue. Évidemment, ce modèle est encore trop général pour pouvoir être utilisé en pratique. On se restreint souvent à trois types de modèles.

1.2.1 Modèle additif

Définition 1

Soit $p \in \mathbb{N}^*$, Une série chronologique $(X_t, t \in \mathbb{Z})$, est issue d'un modèle additif de période p s'il existe deux fonctions déterministes inconnues $T(\cdot)$ et $S(\cdot, p)$ telles que $X_t = T(t) + S(t, p) + \varepsilon(t)$ où $\varepsilon(t)$ est une variable aléatoire et $S(\cdot, p)$ est une fonction périodique de période p . [11]

La tendance est représentée par la fonction $T(\cdot)$ et la composante saisonnière par la fonction $S(\cdot, p)$. La tendance est souvent supposée polynomiale¹, la fonction $S(\cdot, p)$ étant périodique et elle ne prend que p valeurs différentes voir la figure 1.1.

Définition 2

Les p valeurs prises par la fonction $S(\cdot, p)$ sont appelées les coefficients saisonniers. Ils sont donnés par $S_1^p := S(1, p), \dots, S_p^p := S(p, p)$. En d'autres termes, on a pour tout $t \in \mathbb{Z}$, $S(t, p) = S_{j(t,p)}^p$ avec :

$$j(t, p) := \begin{cases} t - p(\lceil t/p \rceil - 1) & \text{si } t \geq 1 \\ t + p(1 + \lfloor |t|/p \rfloor) & \text{si } t < 1 \end{cases} \quad [11]$$

Enfin, on suppose très souvent que les effets saisonniers se compensent sur une période². Cette hypothèse se justifie intuitivement par le fait que sur une période, les creux sont compensés par les pics.

Un exemple simple de modèle additif est le modèle de Buys-Ballot donné par : $X_t = \beta_0 + \beta_1 t + S_{j(t,p)}^p + \varepsilon(t)$ où $t \in \mathbb{Z}$ et $j(t,p)$ est défini comme ci-dessus et $S_1^p + S_2^p + \dots + S_p^p = 0$.

1. $T(t) = \sum_{i=0}^q a_i t^i$

2. c'est-à-dire que $S_1^p + S_2^p, \dots, S_p^p = 0$.

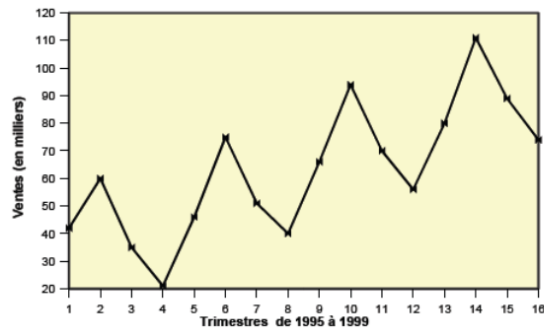


FIGURE 1.1 – Modèle additif. Amplitude constante autour de la tendance

1.2.2 Modèle multiplicatif

Définition 3

Soit $p \in \mathbb{N}^*$, Une série chronologique $(X_t, t \in \mathbb{Z})$, est issue d'un modèle multiplicatif de période p s'il existe deux fonctions déterministes inconnues $T(\cdot)$ et $S(\cdot, p)$ telles que $X_t = T(t) \times S(t, p) \times (1 + \varepsilon(t))$ où ε_t est une variable aléatoire telle que $\mathcal{P}(\varepsilon_t > -1) = 1$ pour tout $t \in \mathbb{Z}$.

Les fonctions $T(\cdot)$ et $S(\cdot, p)$ sont les mêmes que dans le cas additif à la différence que la compensation des effets saisonniers s'écrit $S_1^p \times S_2^p \times \dots \times S_p^p = 1$. [11]

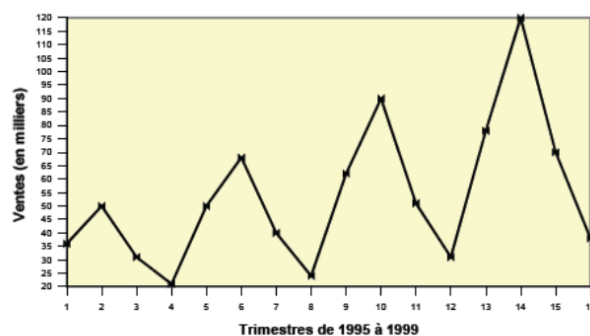


FIGURE 1.2 – Modèle multiplicatif. L'amplitude proportionnelle à la tendance

1.2.3 Modèle mixte

Définition 4

Soit $p \in \mathbb{N}^*$, Une série chronologique $\{X_t, t \in \mathbb{Z}\}$, est issue d'un modèle mixte de période $p \in \mathbb{N}^*$ s'il existe deux fonctions déterministes inconnues $T(\cdot)$ et $S(\cdot, p)$ telles que $X_t = T(t) \times S(t, p) + \varepsilon(t)$ où ε_t est une variable aléatoire.

Les fonctions $T(\cdot)$ et $S(\cdot, p)$ sont toujours les mêmes que dans les deux modèles précédents avec la condition $S_1^p \times S_2^p \times \dots \times S_p^p = 1$ pour la compensation des effets saisonniers. [11]

1.2.4 Choix du modèle

Avant toute modélisation et étude approfondie du modèle, on tente d'abord de déterminer si on est en présence d'une série dans laquelle pour une observation X donnée :

– la variation saisonnière S s'ajoute simplement à la tendance T , c'est le modèle additif.

– la variation saisonnière S est proportionnelle à la tendance T , c'est le modèle multiplicatif.

Afin de faire cette distinction, on peut se baser sur une méthode graphique ou utiliser une méthode analytique.

Méthode du profil :

Pour faire la détermination entre modèle additif et modèle multiplicatif graphiquement, on peut par exemple superposer les saisons représentées par des courbes de profil sur un même graphique. Si ces courbes sont parallèles, le modèle est additif, autrement le modèle est multiplicatif.

Méthode de la bande :

On fait un graphique représentant la série chronologique, puis on trace deux droites une passant par les minima et l'autre par les maxima de chaque saison. Si ces deux droites sont parallèles, nous sommes en présence d'un modèle additif. Dans le cas contraire, c'est un modèle multiplicatif.

Méthode analytique / Méthode du tableau de Buys et Ballot :

On calcule les moyennes et les écarts-types pour chacune des périodes considérées puis la droite des moindres carrés $\sigma_t = a\bar{X}_t + b$, c'est-à-dire, on régresse en fonction de la moyenne \bar{X}_t . Si a est nul, c'est le modèle additif, sinon c'est le modèle multiplicatif³.

3. Il faut bien tester avec les trois méthodes pour décider du modèle.

CHAPITRE

2

ÉTUDE DE LA PARTIE DÉTERMINISTE

2.1 Analyse de la tendance

Dans ce chapitre, nous nous plaçons dans le cadre d'un modèle composé uniquement d'une tendance et des fluctuations irrégulière et donnons différentes méthodes permettant d'estimer la tendance. La fonction tendance peut prendre plusieurs forme :

- Tendance linéaire : $T_t = \alpha_0 + \alpha_1 t$.
- Tendance quadratique : $T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$.
- Tendance polynomiale : $T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_n t^n$.
- Tendance exponentielle : $T_t = c_0 + c_1 \alpha^t$.
- Tendance de Gompertz : $T_t = \exp(c_0 + c_1 \alpha^t)$.
- Tendance Logistique : $\frac{1}{c_0 + c_1 \alpha^t}$.
- ou bien des mélanges de ces types de fonctions.

2.1.1 Ajustement tendanciel linéaire par la méthode des moindres carrés

L'idée est alors d'estimer les paramètres inconnus α_i , pour $i = 1, \dots, n$ par moindres carrés, i.e.

$$(\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_n) = \arg \min_{(\alpha_0, \alpha_1, \dots, \alpha_n) \in \mathbb{R}^{n+1}} \sum_{t=1}^n (X_t - T_t)^2$$

Un cas relativement simple est celui où la tendance est polynomiale ou plus généralement sous forme d'une combinaison linéaire de fonctions entièrement connues, ce qui n'est pas le cas avec les tendances exponentielles, Gompertz et Logistique.

Supposons que l'on observe n valeurs d'une série dont la tendance semble être linéaire. La méthode des moindres carrés consiste à estimer la tendance par une fonction linéaire,

$$\hat{T}_t = \hat{a}t + \hat{b}.$$

Le couple solution (\hat{a}, \hat{b}) est donné par : $\hat{a} = \frac{\text{cov}(t, X_t)}{\text{var}(t)}$ et $\hat{b} = \bar{X} - \hat{a}\bar{t}$.

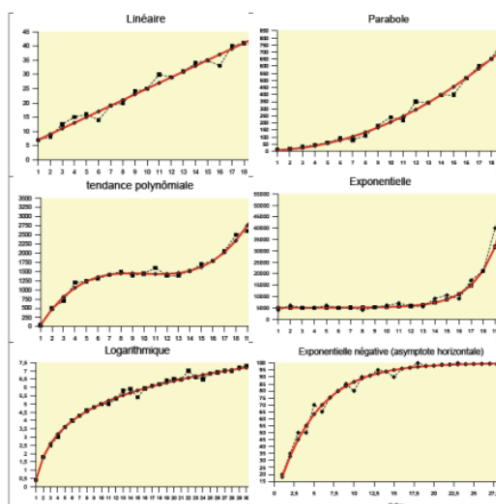


FIGURE 2.1 – Les Modèles de tendance

$$\text{où } \bar{X} = \frac{\sum_{t=1}^n X_t}{n}, \quad \bar{t} = \frac{\sum_{i=1}^n t}{n}, \quad \text{Var}(t) = \frac{\sum_{t=1}^n (t - \bar{t})^2}{n}$$

$$\text{et } \text{Cov}(t, X_t) = \frac{\sum_{t=1}^n (t - \bar{t})(X_t - \bar{X})}{n} = \frac{\sum_{t=1}^n tX_t}{n} - \bar{t}\bar{X}.$$

Afin de confirmer qu'il est raisonnable d'approximer le nuage de points par une droite, on calcule le *coefficient de corrélation*¹ linéaire encore appelé coefficient de **Bravais-Pearson** :

$$r(t, X) = \frac{\text{Cov}(t, X)}{\sigma_t \sigma_X}$$

Le coefficient de corrélation linéaire est symétrique : $r(t, X) = r(X, t)$ et d'après l'inégalité de **Cauchy-Schwarz** on a : $-1 \leq r(t, X) \leq 1$.

En pratique, on considèrera la corrélation linéaire croissante faible lorsque $0 \leq r(t, X) \leq 0.3$, moyenne lorsque $0.3 \leq r(t, X) \leq 0.7$ et forte lorsque $r(t, X) > 0.7$.

2.1.2 Ajustement tendanciel linéaire par points médians

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, X_α) et (t_β, X_β) et on fait passer la droite par ces deux points. Les coefficients (\hat{a}, \hat{b})

1. -Si $r(t, X) = 0$, dans ce cas l'éloignement des points du nuage avec la droite de régression de X en t est maximal. On dira alors que t et X sont *linéairement indépendants*.

-Si $r(t, X) \simeq 1$ (resp $\simeq -1$) les points du nuage sont donc presque alignés, on a donc une *forte corrélation linéaire croissante* (resp décroissante) entre t et X .

-Si $r(t, X) = 1$ (resp -1), les points du nuage sont alors parfaitement alignés, on peut donc parler de corrélation linéaire croissante totale (resp décroissance totale).

-Si $r(t, X) < 0$ (resp > 0), dans ce cas la droite de régression de X en t est décroissante, on parle alors de *corrélation linéaire décroissante* (resp corrélation linéaire croissante) entre X et t .

vérifient :

$$\begin{cases} X_\alpha = \hat{a}t_\alpha + \hat{b} \\ X_\beta = \hat{a}t_\beta + \hat{b} \end{cases} \quad \text{alors} \quad \begin{cases} \hat{a} = \frac{X_\alpha - X_\beta}{t_\alpha - t_\beta} \\ \hat{b} = X_\alpha - \frac{X_\alpha - X_\beta}{t_\alpha - t_\beta}t_\alpha \end{cases}$$

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. On peut également prendre les points moyens ou choisir à la main des points judicieux.

2.1.3 Ajustement tendanciels non linéaires

Dans certaines situations, la modélisation linéaire de la tendance peut être trop simplificatrice. Lorsque la tendance n'est pas linéaire, une technique simple consiste à se ramener à un ajustement linéaire après un changement de variable approprié. Évidemment ce procédé n'est pas toujours possible et on verra plus loin qu'il existe des méthodes d'ajustement non linéaires directes.

Exemple 2.1.1. – Si $T_t = at^2 + b$, en posant $Z_t = t^2$ et on se ramène à $T_t = aZ_t + b$ et on peut faire un ajustement linéaire entre T_t et Z_t .

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

2.1.4 Élimination de la tendance par différenciation

Définition 5

L'opérateur différenciation ∇ (à l'ordre 1) d'une série temporelle est défini comme opérateur qui associe à une série (X_t) une série (\hat{X}_t) déterminée par :

$$\hat{X}_t = \nabla X_t = X_t - X_{t-1} = X_t - BX_t = (I - B)X_t, \quad t \in T$$

Naturellement, la différenciation à l'ordre k de la série X_t est la série \hat{X}_t définie par :

$$\hat{X}_t = \nabla^k X_t = (I - B)^k X_t, \quad t \in T$$

[12]

Remarque 2.1.2. Pour supprimer une saisonnalité de périodicité P , on peut utiliser une méthode de différenciation similaire à celle utilisée pour éliminer la tendance.

2.2 Les moyennes mobiles

Dans la section précédente, nous avons vu comment ajuster la tendance dans un modèle composé simplement d'une tendance et de variations irrégulières. Dans la présente section, nous nous intéressons à un ensemble d'outils, les moyennes mobiles ou filtres linéaires, transformations de séries chronologiques.

Le but sera de lisser une série temporelle, en gardant la tendance et en supprimant la saisonnalité pour ensuite procéder à l'estimation de ces deux composantes. Nous présentons dans

les sous-sections les moyennes mobiles, leurs propriétés et la mise en œuvre de la méthode en pratique, puis on va voir l'estimation des composantes déterministes dans le cadre des modèles déterministes.

2.2.1 Définitions des moyennes mobiles

Ces outils font partie des premières méthodes pour l'analyse des séries chronologiques. Il semble que le physicien **Poynting** soit le premier, en 1884, a utilisé les moyennes mobiles² pour éliminer les variations accidentelles ou périodiques d'une série. En 1904, **Spencer** introduit une moyenne mobile (symétrique d'ordre 15) permettant de conserver les polynômes de degré 3. Puis à partir de 1914, les grands personnages de la statistique tels que **Student**, **Pearson** et **Yule** par exemple, s'intéressent à ce genre de problèmes.

Définition 6

On appelle moyenne mobile, une transformation de X_t s'écrivant comme combinaison linéaire finie des valeurs de la série correspondant à des dates entourant t . La série transformée s'écrit : $X_t^* = \sum_{i=-m_1}^{m_2} \theta_i X_{t+i}$ où $\theta_{-m_1}, \dots, \theta_{m_2}$ sont des réels et $m_1, m_2 \in \mathbb{N}$.

On appelle $m_1 + m_2 + 1$ l'ordre de la moyenne mobile, il représente simplement le nombre de termes considérés dans la somme. [12]

Notation On peut réécrire la moyenne mobile³ en termes d'opérateurs. On définit pour cela l'opérateur B appelé **opérateur retard**, qui à tout processus $(X_t)_{t \in \mathbb{Z}}$ associe le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par :

$$\forall t \in \mathbb{Z} \quad Y_t = BX_t = X_{t-1}$$

- Si on compose B avec lui-même on obtient $B^2 = BoB$ tel que $\forall t \in \mathbb{Z} \quad B^2 X_t = X_{t-2}$
- On peut itérer cette application et définir par récurrence $\forall t \in \mathbb{Z} \quad B^k X_t = X_{t-k}, k \in \mathbb{N}$, Par convention, B^0 est l'opérateur identité I .
- L'opérateur B est linéaire et inversible. Son inverse $B^{-1} = F$ est défini par : $\forall t \in \mathbb{Z} \quad FX_t = X_{t+1}$, l'opérateur est appelé **opérateur avance d'ordre k**.
- On peut réécrire la moyenne mobile en termes d'opérateurs B et F :

$$M_{m_1+m_2+1} = \sum_{i=-m_1}^{m_2} \theta_i B^{-i} = \sum_{i=-m_1}^{m_2} \theta_i F^i$$

Vocabulaire 2.2.1. • si $m_1 = m_2$, on dit que la moyenne mobile est centrée.

- si $\sum_i \theta_i = 1$, on dit que la moyenne mobile est normalisé.
- si $\theta_{-i} = \theta_i$ et $m_1 = m_2$, on dit que la moyenne mobile est symétrique.
- Une moyenne mobile arithmétique est une moyenne mobile centrée et symétrique, d'ordre (impair)

$2m + 1$ et telle que $\theta_i = \frac{1}{2m + 1}, \forall i = -m, \dots, m$.

- Le polynôme P de $M_{m_1+m_2+1}$ est défini par :

$$P(x) = \sum_{i=-m_1}^{m_2} \theta_i x^{m_1+i}$$

est appelé polynôme caractéristique de la moyenne mobile $M_{m_1+m_2+1}$.

2. Cette moyenne est dite mobile parce qu'elle est recalculée de façon continue, en utilisant à chaque calcul un sous-ensemble d'éléments dans lequel un nouvel élément remplace le plus ancien ou s'ajoute au sous-ensemble.

3. Mathématiquement, toute moyenne mobile est un exemple de **convolution**. Physiquement, une moyenne mobile est un **filtre passe-bas** et possède ainsi un lien profond avec le traitement du signal.

Moyenne mobile arithmétique 2

La série des moyennes mobiles arithmétiques d'ordre k , notée $(M_k(t))_t$, 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 :

$$\frac{t_1 + t_2 + \dots + t_k}{k} \quad \text{puis} \quad \frac{t_2 + t_3 + \dots + t_{k+1}}{k} \quad \dots \quad \text{jusqu'à} \quad \frac{t_{n-k+1} + \dots + t_n}{k}$$

et pour la variable d'intérêt

$$\frac{X_1 + X_2 + \dots + X_k}{k} \quad \text{puis} \quad \frac{X_2 + X_3 + \dots + X_{k+1}}{k} \quad \dots \quad \text{jusqu'à} \quad \frac{X_{n-k+1} + \dots + X_n}{k}$$

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 :

$$\frac{1}{k}(0.5X_{t-m} + X_{t-m+1} + \dots + X_{t+m-1} + 0.5X_{t+m}) = 0.5M_k^1 X_t + 0.5M_k^2 X_t$$

combinaison linéaire des moyennes mobiles arithmétiques sur $2m$ valeurs :

$$M_k^1 X_t = \frac{1}{k}(X_{t-m} + X_{t-m+1} + \dots + X_{t+m-1}) = M_{m_1+m_2+1} X_t \quad \text{avec} \quad m_1 = m \quad \text{et} \quad m_2 = m - 1$$

et

$$M_k^2 X_t = \frac{1}{k}(X_{t+m-1} + \dots + X_{t+m}) = M_{m_1+m_2+1} X_t \quad \text{avec} \quad m_1 = m - 1 \quad \text{et} \quad m_2 = m$$

Exemple 2.2.1. Calculons les séries des moyennes mobiles d'ordre 3 de la série initiale X_t suivante :

Temps t	Série X_t	Date $M_3(t)$ de la MM	$M_3(X_t)$
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		

Exemple 2.2.2. Calculons les séries des moyennes mobiles d'ordre 2 et 4 de la série initiale X_t suivante :

Temps t	Série X_t	Date $M_2(t)$ de la MM	$M_2(X_t)$
1	5		
2	3	$(\frac{1}{2} + 2 + \frac{3}{2})/2 = 2$	$(\frac{5}{2} + 3 + \frac{4}{2})/2 = 3,75$
3	4	$(\frac{2}{2} + 3 + \frac{4}{2})/2 = 3$	$(\frac{3}{2} + 4 + \frac{5}{2})/2 = 4$
4	5	$(\frac{3}{2} + 4 + \frac{5}{2})/2 = 4$	$(\frac{4}{2} + 5 + \frac{4}{2})/2 = 4,5$
5	4	$(\frac{4}{2} + 5 + \frac{6}{2})/2 = 5$	$(\frac{5}{2} + 4 + \frac{4}{2})/2 = 4,25$
6	4		

Temps t	Série X_t	Date $M_4(t)$ de la MM	$M_4(X_t)$
1	5		
2	3		
3	4		
4	5	$(\frac{1}{2} + 2 + 3 + 4 + \frac{5}{2})/4 = 3$	$(\frac{5}{2} + 3 + 4 + 5 + \frac{4}{2})/4 = 4,125$
5	4	$(\frac{2}{2} + 3 + 4 + 5 + \frac{6}{2})/4 = 4$	$(\frac{3}{2} + 4 + 5 + 4 + \frac{4}{2})/4 = 4,125$
6	4		

2.2.2 Propriétés d'un lissage par moyenne mobile

la moyenne mobile transfère la série chronologique à une série filtrée, afin d'éviter les pics et les creux.

1- Effet d'une moyenne mobile sur une tendance :

Proposition 1

- 1- Une moyenne mobile conserve les constantes $\Leftrightarrow \theta_{-m_1} + \dots + \theta_{m_2} = 1$.
- 2- Une moyenne mobile symétrique conservant les constantes conserve les polynômes de degré 1. [12]

Proposition 2

La moyenne mobile d'ordre $m_1 + m_2 + 1$ conserve les polynômes de degré inférieur ou égal à p si et seulement si 1 est racine d'ordre $p + 1$ du polynôme $P(x) - x^{m_1}$. [12]

Preuve 2.2.1. *Étudions le polynôme $Q(x) := P(x) - x^{m_1}$, Tout d'abord, nous avons pour tout k*

$$\left\{ \begin{array}{l} P(x) = \sum_{i=-m_1}^{m_2} \theta_i x^{m_1+i} \\ P'(x) = \sum_{i=-m_1}^{m_2} \theta_i (m_1 + i) x^{m_1+i-1} \\ P''(x) = \sum_{i=-m_1}^{m_2} \theta_i (m_1 + i)(m_1 + i - 1) x^{m_1+i-2} \\ \cdot \\ \cdot \\ \cdot \\ P^k(x) = \sum_{i=-m_1}^{m_2} \theta_i (m_1 + i)(m_1 + i - 1) \dots (m_1 + i - k + 1) x^{m_1+i-k} \end{array} \right.$$

et immédiatement

$$\left\{ \begin{array}{l} Q(1) = \sum_{i=-m_1}^{m_2} \theta_i - 1 \\ Q'(1) = \sum_{i=-m_1}^{m_2} \theta_i (m_1 + i) - m_1 \\ Q''(1) = \sum_{i=-m_1}^{m_2} \theta_i (m_1 + i)(m_1 + i - 1) - m_1(m_1 - 1) \\ \cdot \\ \cdot \\ \cdot \\ Q^k(1) = \sum_{i=-m_1}^{m_2} \theta_i (m_1 + i)(m_1 + i - 1) \dots (m_1 + i - k + 1) - m_1(m_1 - 1) \dots (m_1 + 1 - k) \end{array} \right.$$

Par conséquent 1 est racine d'ordre $p + 1$ de $P(x) - x^{m_1}$ si et seulement si

$$\left\{ \begin{array}{l} Q(1) = 0 \\ Q'(1) = 0 \\ Q''(1) = 0 \\ \vdots \\ Q^k(1) = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \sum_{i=-m_1}^{m_2} \theta_i = 1 \\ \sum_{i=-m_1}^{m_2} \theta_i i = 0 \\ \sum_{i=-m_1}^{m_2} \theta_i i^2 = 0 \\ \vdots \\ \sum_{i=-m_1}^{m_2} \theta_i i^k = 0 \end{array} \right.$$

Soit maintenant la série $X_t = t^k$, pour $k \leq p$. Sa transformée par la moyenne mobile centrée d'ordre $2m + 1$ est donnée par :

$$\begin{aligned} X_t^* &= \sum_{i=-m_1}^{m_2} \theta_i (t+i)^k \\ &= \sum_{i=-m_1}^{m_2} \theta_i \sum_{l=0}^k C_k^l t^{k-l} i^l \\ &= \sum_{l=0}^k C_k^l t^{k-l} \sum_{i=-m_1}^{m_2} \theta_i i^l \\ &= C_k^0 t^k \sum_{i=-m_1}^{m_2} \theta_i i^0 + C_k^1 t^{k-1} \sum_{i=-m_1}^{m_2} \theta_i i^1 + \dots + C_k^k t^0 \sum_{i=-m_1}^{m_2} \theta_i i^k \\ &= C_k^0 t^k \\ &= t^k \end{aligned}$$

si on suppose que 1 est racine d'ordre $p + 1$ de $P(x) - x^{m_1}$. Et on en déduit que la moyenne mobile d'ordre $m_1 + m_2 + 1$ conserve bien les polynômes de degré inférieur ou égal à p . On vérifie aisément que la réciproque est vraie \square .

2- Effet d'une moyenne mobile sur la composante saisonnière :

Si la série X_t possède une composante saisonnière de période P alors l'application d'une moyenne mobile d'ordre P supprime cette saisonnalité. La série $(M_P X_t)_t$ ne possède plus de composante saisonnière de période P . Plus précisément, cherchons les séries chronologiques qui sont arrêtées par le filtre moyenne mobile centrée d'ordre $2m + 1$. Ce sont les séries S_t telles que leur transformée S_t^* par la moyenne mobile vérifie $S_t^* = 0$.

On cherche donc à déterminer le noyau d'une moyenne mobile arithmétique d'ordre $2m + 1$ ou encore les vecteurs propres associés à la valeur propre nulle.

Pour trouver les séries arrêtées par une moyenne mobile, on considère le polynôme : $P(x) = \sum_{i=-m_1}^{m_2} \theta_i x^{m_1+i}$ dont on cherche les racines.

Proposition 3

La moyenne mobile d'ordre $m_1 + m_2 + 1$ arrête les fonctions de la forme a^t si et seulement si a est racine du polynôme $P(x)$. [12]

Proposition 4

Une moyenne mobile absorbe les composantes saisonnières de période p si et seulement si son polynôme caractéristique P est divisible par $1 + x + \dots + x^{p-1}$. [12]

3- Effet d'une moyenne mobile sur les fluctuations irrégulières :

Jusqu'à présent nous nous sommes intéressés qu'à l'effet d'une moyenne mobile sur la partie déterministe de la série (tendance et saisonnalité). Nous allons étudier maintenant l'effet d'une moyenne mobile sur le résidu lorsque celui-ci est un bruit blanc[nous verrons plus précisément au chapitre suivant ce processus]. Par construction, une moyenne mobile consiste à faire des moyennes partielles de proche en proche. On obtient donc un lissage de la série. L'effet de la composante irrégulière est d'autant plus atténué que l'ordre de la moyenne mobile est grand.

Proposition 5

1- Les moyennes mobiles arithmétiques d'ordre $2m + 1$ sont les moyennes mobiles minimisant la variance d'un bruit blanc parmi les moyennes mobiles centrées telles

que $\sum_{i=-m}^m \theta_i = 1$.

2- Les moyennes mobiles arithmétiques d'ordre $2m + 1$ transforment un bruit blanc en un processus centré (inchangé) et de variance $\frac{\sigma^2}{2m + 1}$ réduite. Notons cependant que les variables ε_t^* sont à présent corrélées alors que les ε_t ne l'étaient pas. [12]

Preuve 2.2.2. *On a en effet :*

$$\begin{aligned}
 \gamma(h) &= Cov(\varepsilon_t^*, \varepsilon_{t+h}^*) \\
 &= E(\varepsilon_t^* \varepsilon_{t+h}^*) - E(\varepsilon_t^*)E(\varepsilon_{t+h}^*) \\
 &= E(\varepsilon_t^* \varepsilon_{t+h}^*) \\
 &= E\left(\sum_{i=-m}^m \theta_i \varepsilon_{t+i} \sum_{j=-m}^m \theta_j \varepsilon_{t+h+j}\right) \\
 &= \sum_{i=-m}^m \sum_{j=-m}^m \theta_i \theta_j E(\varepsilon_{t+i} \varepsilon_{t+h+j}) \\
 &= \begin{cases} \sigma^2 \sum_{i=h-m}^m \theta_i \theta_{i-h} & \text{si } h \leq 2m \\ 0 & \text{sinon} \end{cases} \\
 &= \begin{cases} \sigma^2 \frac{2m+1-h}{(2m+1)^2} & \text{si } h \leq 2m \\ 0 & \text{sinon} \end{cases}
 \end{aligned}$$

On voit ainsi que la covariance entre ε_t^* et ε_{t+h}^* ne dépend que de σ^2 (variance du bruit blanc) et de h et non de t et qu'elle s'annule dès que $h > 2m$. Nous verrons plus loin la notion qui généralise les processus vérifiant cette propriété d'indépendance en t de la covariance. En divisant $\gamma(h)$ par $Var(\varepsilon_t^*) = \gamma(0)$, on obtient la corrélation linéaire entre ε_t^* et ε_{t+h}^* :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} \frac{2m+1-h}{(2m+1)} & \text{si } h \leq 2m \\ 0 & \text{sinon} \end{cases}$$

La corrélation est donc indépendante de la variance du bruit blanc.

Cas particulier des moyennes mobiles arithmétiques On a vu qu'une moyenne mobile arithmétique d'ordre $2m+1$:

- laisse invariants les polynômes de degré 1.
- arrête les fonctions périodiques de période $2m+1$.
- réduit la variance d'un bruit blanc de manière optimale (c'est-à-dire d'un facteur $\frac{1}{2m+1}$).
- laissent invariantes certaines suites géométriques de la forme a^t (si a est racine de $P(x) - x^m$).
- absorbent certaines suites géométriques de la forme a^t (si a est racine de P).

C'est pourquoi ce sont celles que nous utiliserons le plus fréquemment en pratique.

Remarque 2.2.3. 1- Les moyennes mobiles permettent de lisser directement la série sans hypothèse a priori sur le modèle sous-jacent. La méthode est donc valable quel que soit le modèle de décomposition.

Pour cette raison, on peut classer ce type de lissage dans les méthodes non-paramétriques. C'est un outil simple à mettre en œuvre qui met en évidence l'allure de la tendance en supprimant la composante saisonnière et en atténuant le bruit.

2- Dans tout ce qui précède, on peut remplacer la moyenne par la médiane et on obtient alors un **lissage par médiane mobile**. Ce procédé a alors l'avantage d'être moins sensible aux valeurs aberrantes.

2.3 Prédiction par lissage exponentiel

Introduite par **Holt** en 1958 ainsi que par **Winters** en 1960 et popularisée par le livre de **Brown** en (1963), le lissage exponentiel est une méthode non paramétrique permettant, lorsque l'on dispose d'observations (X_1, \dots, X_T) d'une série chronologique, d'effectuer la prédiction de la valeur à l'instant $T+1$ (horizon $h=1$). Cette méthode est couramment utilisée notamment en gestion de stocks. Il est important de noter dès à présent que la méthode du lissage exponentiel doit être théoriquement utilisée pour des séries ne présentant ni tendance ni variations saisonnières. Cependant cette méthode est parfois utilisée sur des séries quelconques afin de les lisser ou de prédire des éventuelles valeurs manquantes. Nous donnons dans un premier temps la définition du lissage exponentiel et nous expliquons ensuite comment utiliser cette méthode en pratique. Enfin, nous justifierons par un résultat théorique la définition du lissage exponentiel.

2.3.1 Lissage exponentiel simple(LES)

Définition 7

Soit une série $\{X_t ; t \leq T\}$ où $T \in \mathbb{N}$ que l'on supposera bornée, c'est-à-dire $\sup_{t \leq T} |X_t| < \infty$. La prédiction à l'instant $T+1$ par la méthode du lissage exponentiel simple (LES) est donnée par :

$$X_{T+1}^{(LES)} := (1 - \alpha) \sum_{k=0}^{\infty} \alpha^k X_{T-k}$$

où $\alpha \in]0, 1[$ est paramètre de lissage. [11]

Avant de montrer comment cette méthode est utilisée en pratique, voici quelques remarques importantes.

Remarque 2.3.1. 1- Il ressort de la définition ci-dessus, que l'idée du lissage exponentiel simple est de prédire la valeur à l'instant $T + 1$ par une moyenne pondérée, puisque $\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1 - \alpha}$ des valeurs passées. On voit donc bien que cette méthode n'a théoriquement de sens que si la série ne présente ni tendance ni variations saisonnières.
 2- Le paramètre de lissage α a pour but de donner un poids plus faible aux observations les plus anciennes. Plus α est proche de 0, moins on tient compte des valeurs passées.
 3- Pour calculer le lissage exponentiel d'une série, on utilise la plupart du temps la formule de mise à jour ci-dessous : pour tout $t \leq T$, $X_{t+1}^{(LES)} = \alpha X_t^{(LES)} + (1 - \alpha)X_t$.

La définition du lissage exponentiel suppose que l'on dispose d'une infinité d'observations. Ce n'est évidemment jamais le cas en pratique. Pour appliquer la méthode du lissage exponentiel, on procède de la façon suivante :

Supposons que l'on dispose d'observations (X_1, \dots, X_T) .

On initialise la formule de mise à jour par $X_1^{(LES)} = X_0$.

On applique ensuite pour $t = 1, \dots, T$ la formule de mise à jour : $X_{t+1}^{(LES)} = \alpha X_t^{(LES)} + (1 - \alpha)X_t$.

La série $X_t^{(LES)}$ tel que $t = 1, \dots, T + 1$ est appelée la série lissée. Elle peut être utilisée pour travailler sur une série plus propre que la série initiale (moins de fluctuations) ou pour estimer une valeur manquante dans la série initiale.

Il reste à aborder le problème du choix de α . Il est courant de prendre pour α la valeur minimisant l'écart quadratique entre les valeurs observées de la série et les valeurs lissées par la méthode du lissage exponentiel. Autrement dit, on prend pour α la valeur $\hat{\alpha}$ donnée par :

$$\hat{\alpha} = \arg \min_{\alpha \in]0,1[} \sum_{t=1}^T (X_t^{(LES)} - X_t)^2$$

Terminons ce paragraphe sur le lissage exponentiel simple par une justification théorique de la méthode.

Proposition 6

Soit une série $\{X_t ; t \leq T\}$ une série bornée et soit $\alpha \in]0, 1[$. On a :

$$X_{t+1}^{(LES)} = \arg \min_{b \in \mathbb{R}} \sum_{k=0}^{\infty} \alpha^k (X_{t-k} - b)^2$$

[11]

Ainsi, le lissage exponentiel est la meilleure prévision (au sens des moindres carrés pondérés) de la série constante.

2.3.2 Lissage exponentiel double(LED)

Une critique que l'on peut faire de la méthode du lissage exponentiel simple est qu'elle suppose l'absence de composante tendancielle dans la série. Pour remédier à cela, on peut utiliser la méthode du lissage exponentiel double qui est définie de la façon suivante :

Définition 8

Soit une série $\{X_t ; t \leq T\}$ où $T \in \mathbb{N}$ et $\alpha \in]0, 1[$. La prévision de cette série à l'horizon $h \in \mathbb{N}$ par la méthode du lissage exponentiel double (LED) est donnée par :

$$X_{t+h}^{(LED)} = \hat{a}_t h + \hat{b}_t \quad \text{avec} \quad (\hat{a}_t, \hat{b}_t) = \arg \min_{(a,b) \in \mathbb{R}^2} \sum_{k=0}^{\infty} \alpha^k (X_{t-k} + ak - b)^2$$

[11]

L'idée du lissage exponentiel double est donc simplement d'ajuster à la série X_t , $t \leq T$, la série $-a(T-t) + b$, $t \in \mathbb{Z}$, en utilisant la méthode des moindres carrés pondérés⁴. On prend donc en compte la présence d'une éventuelle tendance mais, comme pour le lissage exponentiel simple, il faut que la série ne présente pas de composantes saisonnières.

Les valeurs explicites des paramètres \hat{a}_t et \hat{b}_t sont données dans le résultat ci-dessous.

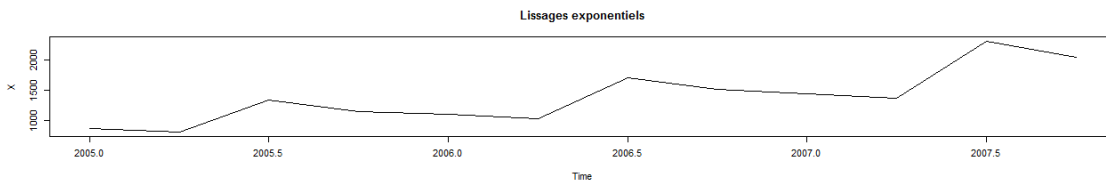


FIGURE 2.2 – Lissage exponentiel

Proposition 7

On pose :

$$L_{T+1}^1 := (1 - \alpha) \sum_{k=0}^{\infty} \alpha^k X_{T-k} \quad \text{et} \quad L_{T+1}^2 := (1 - \alpha) \sum_{k=0}^{\infty} \alpha^k L_{T-k+1}^1 ;$$

Les paramètres du lissage exponentiel double sont donnés par :

$$\hat{a}_t = \frac{1 - \alpha}{\alpha} (L_{T+1}^1 - L_{T+1}^2) \quad \text{et} \quad \hat{b}_t = 2L_{T+1}^1 - L_{T+1}^2$$

[11]

4. Cette méthode est souvent utilisée lorsque les variances des résidus ne sont pas constantes sur l'étendue des valeurs des variables indépendantes

Ce résultat justifie l'appellation de lissage exponentiel double puisque, pour le calcul des paramètres, on effectue deux lissages exponentiel simple consécutifs : le premier sur la série X_t avec $t \leq T$, et le second sur la série lissée L_t^1 avec $t \leq T + 1$. Comme pour le lissage exponentiel simple, il existe des formules de mises à jour permettant de mettre en œuvre le lissage exponentiel double.

Proposition 8

La formule de mise à jour pour les coefficients \hat{a}_t et \hat{b}_t est donnée pour $t \leq T$:

$$\hat{b}_t = (1 - \alpha^2)X_t + \alpha^2(\hat{a}_{t-1} + \hat{b}_{t-1}) \quad \text{et} \quad \hat{a}_t = \frac{2\alpha}{1 + \alpha}\hat{a}_{t-1} + \frac{1 - \alpha}{1 + \alpha}(\hat{b}_t - \hat{b}_{t-1})$$

[11]

Preuve 2.3.1. *En utilisant la formule de mise à jour du lissage exponentiel simple⁵, on a :*

$$L_{T+1}^1 = (1 - \alpha)X_T + \alpha L_T^1 \quad \text{et} \quad L_{T+1}^2 = (1 - \alpha)L_{T+1}^1 + \alpha L_T^2.$$

Ainsi,

$$\begin{aligned} L_{t+1}^2 &= \alpha L_t^2 + (1 - \alpha)((1 - \alpha)X_t + \alpha L_t^1). \\ &= \alpha L_t^2 + (1 - \alpha)^2 X_t + \alpha(1 - \alpha)L_t^1. \end{aligned}$$

On en déduit que :

$$\begin{aligned} \hat{b}_t &= 2L_{t+1}^1 - L_{t+1}^2 \\ &= 2((1 - \alpha)X_t + \alpha L_t^1) - \alpha L_t^2 + (1 - \alpha)^2 X_t + \alpha(1 - \alpha)L_t^1. \\ &= (1 - \alpha^2)X_t + \alpha^2\left(\frac{1 + \alpha}{\alpha}L_t^1 - \frac{1}{\alpha}L_t^2\right) \\ &= (1 - \alpha^2)X_t + \alpha^2(\hat{a}_{t-1} + \hat{b}_{t-1}). \end{aligned}$$

enfin :

$$\hat{a}_t = \frac{1 - \alpha}{\alpha}(\alpha(1 - \alpha)X_t + \alpha^2 L_t^1 - \alpha L_t^2)$$

En remarquant que :

$$\hat{b}_t - \hat{b}_{t-1} = (1 - \alpha^2)X_t + \alpha^2\hat{a}_{t-1} - (1 - \alpha^2)\hat{b}_{t-1}.$$

On a alors :

$$\frac{2\alpha}{1 + \alpha}\hat{a}_{t-1} + \frac{1 - \alpha}{1 + \alpha}(\hat{b}_t - \hat{b}_{t-1}) = (1 - \alpha)^2 X_t + \frac{2\alpha + \alpha^2(1 - \alpha)}{1 + \alpha}\hat{a}_{t-1} - (1 - \alpha^2)\hat{b}_{t-1}.$$

et en utilisant le résultat de la Proposition précédente, on conclut la démonstration.

La méthode de **Holt** est une généralisation de la méthode du lissage exponentiel double.

5. On remarque ainsi que pour tout $t \leq T$, \hat{b}_t est la moyenne pondérée de X_t et $\hat{a}_{t-1} + \hat{b}_{t-1}$ et que \hat{a}_t est la moyenne pondérée de \hat{a}_{t-1} et $\hat{a}_t - \hat{a}_{t-1}$. Cette remarque est le point de départ de la méthode de lissage de **Holt** présentée ci-après.

Définition 9

Soit la série $\{X_t \ t \in \mathbb{N}\}$. La prévision de cette série à l'horizon $h \in \mathbb{N}^*$ par la méthode de lissage de **Holt** est donnée pour $\xi \in]0, 1[$ et $\beta \in]0, 1[$ par

$$X_{t+h}^{Holt} = \hat{c}_t h + \hat{d}_t.$$

où, pour $t \leq T$, les paramètres \hat{c}_t et \hat{d}_t sont définis par la formule de récurrence suivante

$$\hat{d}_t = \xi X_t + (1 - \xi)(\hat{c}_{t-1} + \hat{d}_{t-1}) \quad \text{et} \quad \hat{c}_t = \beta \hat{c}_{t-1} + (1 - \beta)(\hat{d}_t - \hat{d}_{t-1}).$$

[11]

La méthode de lissage de *Holt*⁶ est plus souple que celle du lissage exponentiel double au prix du choix de deux paramètres au lieu d'un seul.

Pour finir, montrons comment en pratique on applique la méthode du lissage de *Holt* (et donc en particulier celle du double lissage exponentiel) lorsque l'on dispose des observations $\{X_1, X_2, \dots, X_T\}$. Il faut commencer par initialiser les paramètres. En partant de l'idée qui soutend le lissage exponentiel double, on souhaite que les points $\{X_i \ i = 1, 2\}$ soient sur la droite d'équation $\hat{d}_2 - \hat{c}_2(2 - i)$. Autrement dit, on souhaite avoir :

$$\begin{cases} \hat{d}_2 - \hat{c}_2 = X_1 \\ \hat{d}_2 = X_2 \end{cases}$$

On initialise donc la formule de mise à jour avec $\hat{c}_2 = X_2 - X_1$ et $\hat{d}_2 = X_2$. Par récurrence, on en déduit facilement les valeurs X_t^{Holt} pour $t = 1, \dots, T$ ainsi que les coefficients \hat{c}_t et \hat{d}_t . On peut ensuite faire la prévision à l'horizon h en utilisant la formule $X_{t+h}^{Holt} = \hat{c}_t h + \hat{d}_t$.

2.4 L'étude d'une série chronologique

Nous disposons maintenant des outils de base permettant la décomposition d'une série chronologique. Il est clair qu'afin de pouvoir estimer la tendance, c'est-à-dire le mouvement du phénomène observé sur un grand intervalle de temps, il faut disposer d'une série statistique sur une longue période. Disposant de ces données, comme nous l'avons dit précédemment, le premier travail consiste à effectuer une représentation graphique adéquate permettant d'avoir une vue globale du phénomène en question. Afin d'éliminer ou d'amortir les mouvements cycliques, saisonniers et accidentels, on utilise donc la technique des moyennes mobiles et on procède ainsi en quelque sorte au lissage de la courbe. D'après ce que nous venons de voir, la méthode des moyennes mobiles arithmétiques peut être utilisée pour tout type de modèle. Cependant, d'après ses propriétés, elle est particulièrement adaptée pour le modèle déterministe additif lorsque :

- la tendance est sensiblement linéaire,
- la composante saisonnière est périodique,
- le bruit est de variance faible.

Dans le cas de polynôme de degré supérieur à 1, on peut aussi utiliser des moyennes mobiles autres qu'arithmétiques laissant invariants les polynômes de degré supérieur à 1.

l'étude de la série statistique (X_t) en vue de la prédiction se font selon les étapes suivantes :

6. En prenant $\xi = 1 - \alpha^2$ et $\beta = \frac{2\alpha}{1 + \alpha}$, on retrouve la méthode du lissage exponentiel double.

1. application d'une moyenne mobile d'ordre judicieusement choisi.
2. estimation de la saisonnalité.
3. estimation de la tendance.
4. itération éventuelle de la procédure.
5. prévision des valeurs futures.
6. analyse des résidus.

2.4.1 Application d'une moyenne mobile d'ordre judicieusement choisi

Tout d'abord, on applique une moyenne mobile arithmétique d'ordre $2m + 1$ dans le cas d'une saisonnalité d'ordre impair $2m + 1$ ou une moyenne mobile arithmétique modifiée d'ordre $2m + 1$ dans le cas d'une saisonnalité de période paire $2m$. Dans chaque cas, on a vu que la saisonnalité est ainsi annulée et que la variance du bruit est diminuée. Si le modèle est bon, la série transformée ne contient plus aucun mouvement saisonnier. Notons que la série ainsi obtenue est de longueur $T - 2m$.

2.4.2 Estimation de la saisonnalité

On calcule la série diminuée de sa tendance \tilde{S}_t en retranchant à la série de départ la série transformée X_t^* . On estime ensuite les P coefficients du saisonnier par la moyenne des valeurs de \tilde{S}_t correspondant à chaque temps de la période.

On peut affiner la méthode en retranchant ensuite à chaque coefficient estimé la moyenne des coefficients afin que la condition de nullité de la moyenne des coefficients sur la période soit respectée.

Plus précisément, supposons que les observations sont périodiques de période $P = 2m$ paire et réalisées sur N périodes. A l'issue du premier filtrage, on dispose de la série

$$\tilde{S}_t = X_t - X_t^*, \quad t = m + 1, \dots, PN - m$$

Cette série est appelée **série corrigée de la tendance**. Disposons ces valeurs dans un tableau

inter-période	période	1	...	m	...	j	...	$P - m + 1$...	P
1		***	***	***						
⋮										
i						$\tilde{S}_{j+P(i-1)}$				
⋮										
N								***	***	***

L'estimation du coefficient saisonnier c_j est donc :

$$\tilde{c}_j = \begin{cases} \frac{1}{N-1} \sum_{i=2}^N \tilde{S}_{j+P(i-1)}, & 1 \leq j \leq m \\ \frac{1}{N-1} \sum_{i=1}^{N-1} \tilde{S}_{j+P(i-1)}, & m+1 \leq j \leq P \end{cases}$$

Cette opération consiste à appliquer à la série $(\tilde{S}_t)_{t=m+1, \dots, PN-m}$ à laquelle on ajoute P observations nulles pour $t = 1, \dots, m$ et $t = PN - m + 1, \dots, PN$, une moyenne mobile d'ordre $P(N-1) + 1$ et de coefficients

$$\frac{1}{N-1} \left(\underbrace{1, 0, \dots, 0}_{P-1}, \underbrace{1, 0, \dots, 0, 1, 0, \dots, 0, 1}_{N-1}, \underbrace{0, \dots, 0, 1}_{P-1} \right)$$

Les estimateurs des P coefficients saisonniers sont ensuite

$$\hat{c}_j = \tilde{c}_j - \frac{1}{P} \sum_{l=1}^P \tilde{c}_l$$

de façon à bien vérifier la contrainte d'une somme nulle sur une période.

2.4.3 Estimation de la tendance

Une fois les coefficients saisonniers estimés à l'étape précédente, on retranche l'estimation du saisonnier à la série X_t . La série ainsi obtenue est appelée **série corrigée des valeurs saisonnières** :

$$X_{CVS,t} = X_t - \hat{S}_t \quad \text{où } t = 1, \dots, T;$$

avec $\hat{S}_t = \hat{c}_j$ et $t \equiv j[P]$

Remarque 2.4.1. notons que si l'on avait soustrait \tilde{S}_t à X_t au lieu de \hat{S}_t , nous aurions simplement obtenu : $X_t - \tilde{S}_t = X_t^*$, $t = m + 1, \dots, T - m$ au lieu de $X_{CVS,t}$. Mais ce résultat aurait été moins précis puisque nous n'aurions eu des valeurs t seulement de $m + 1$ à T .

On procède ensuite à l'estimation du terme représentant la tendance par une méthode de régression comme vu au début de ce Chapitre⁷ : on modélise le plus souvent la tendance par un polynôme⁸ $Q(t)$. Puis on ajuste au sens des moindres carrés un polynôme $\hat{Q}(t)$ à la série corrigée des valeurs saisonnières $X_{CVS,t}$.

2.4.4 Itération éventuelle de la procédure

On procède parfois à une itération de la procédure : on estime à nouveau la tendance à partir de la série $X_{CVS,t}$ en utilisant une moyenne mobile d'ordre différent de celui utilisé à l'étape 1. Soit \hat{T}_t cette estimation. On retranche à la série X_t cette tendance et on revient éventuellement à l'étape 2 pour une seconde estimation du saisonnier.

2.4.5 Prédiction des valeurs futures

Afin de prévoir les valeurs futures de la série, on utilise l'estimation de la tendance et celle de la composante saisonnière. Plus précisément, si on souhaite prévoir une valeur de la série à l'instant $T + h$, où $h \geq 1$, c'est-à-dire à l'horizon h , on utilise les estimations de la tendance et de la saisonnalité et on pose :

$$\hat{X}_T(h) = \hat{Q}_{T+h} + \hat{c}_j \quad \text{où } T + h \equiv j[P].$$

Remarque 2.4.2. Cette méthode de prévision a un inconvénient majeur qui est celui d'une prévision ne tenant pas compte des valeurs les plus récentes de la série : celles-ci ont été "éliminées" par application de moyennes mobiles. Pour cette raison, on utilise plus souvent la méthode comme méthode de désaisonnalisation que comme méthode de prévision.

7. Section 1 : Analyse de la tendance

8. en général un polynôme de degré 1 pour être cohérent avec le choix fait lors de la première étape

2.4.6 Analyse des résidus

Une fois estimées les composantes du modèle, on peut contrôler la pertinence du modèle par une analyse des résidus. Ceux-ci sont définis par :

$$\hat{\varepsilon}_t = X_t - \hat{S}_t - \hat{Q}_t = X_{CVS,t} - \hat{Q}_t$$

Si le modèle est bon, il ne doit rester dans les résidus aucune trace de saisonnalité. Pour le vérifier, on trace le corrélogramme⁹ des résidus, mais parfois le modèle peut être bon même si les résidus sont corrélés, il suffit de les modéliser par un des processus traités au chapitre quatre, ARMA par exemple.

Comme nous le verrons plus loin, le corrélogramme n'est tracé en théorie que dans le cas où la série est stationnaire, ce qui implique en particulier qu'il n'y a dans cette série ni tendance ni saisonnalité. En pratique, on s'en sert (dans le cas de l'analyse des résidus) pour vérifier justement l'absence de saisonnalité dans les résidus.

Important : - Si c'est le cas et si le modèle est bon, le corrélogramme ne doit présenter que des valeurs faibles, indiquant une faible corrélation entre les erreurs.

- Si au contraire, le corrélogramme présente des pics régulièrement espacés, cela indique que le saisonnier n'a pas été complètement éliminé et c'est donc le signe que le modèle proposé a échoué, on peut alors réitérer la procédure ci-dessus ou proposer un autre modèle.

- Dans le cas où le corrélogramme (ou le périodogramme) des résidus n'indique pas la présence d'un mouvement saisonnier, on trace le graphe des résidus $(t, \hat{\varepsilon}_t)$ qui sert à repérer d'éventuelles observations exceptionnelles, un mouvement tendanciel...

- Dans le cas de l'hypothèse d'erreurs gaussiennes, on vérifie celle-ci en traçant l'histogramme des résidus, un qq-plot ou encore en effectuant un test de normalité...

Remarque 2.4.3. : Cas du modèle multiplicatif/Mixte

On utilise parfois la méthode des moyennes mobiles dans le cadre du modèle multiplicatif ou du modèle mixte bien qu'elle ne soit pas particulièrement bien adaptée à ces contextes. Dans ce cas, l'algorithme est identique à celui décrit ci-dessus avec quelques adaptations selon le cas.

Cas du modèle multiplicatif : Soit $X_t = T_t(1 + S_t)(1 + \varepsilon_t)$, les différences sont les suivantes :

- à l'étape 2, on pose :

$$\hat{S}_t = \frac{X_t}{X_t^*} - 1;$$

Puis on estime les coefficients saisonniers et on obtient les estimations \bar{c}_j , $j = 1, \dots, P$ de la même manière que pour le modèle additif. On pose ensuite : $\bar{c}_j = P \frac{\hat{c}_j}{\sum_{j=1}^P \hat{c}_j}$, $j = 1, \dots, P$ de

façon à respecter la contrainte de nullité en moyenne sur la période $\sum_{j=1}^P c_j = 0$

- à l'étape 3, on pose :

$$X_{CVS,t} = \frac{X_t}{1 + \hat{S}_t}, \quad t = m + 1, \dots, T - m$$

9. c'est-à-dire le graphe d'un estimateur de la fonction d'autocorrélation.

Cas du modèle mixte : Soit $X_t = T_t S_t + \varepsilon_t$, l'algorithme est identique à celui décrit ci-dessus avec quelques adaptations selon le cas. Par exemple, pour un modèle multiplicatif du type $X_t = T_t S_t + \varepsilon_t$, les différences sont les suivantes :

- à l'étape 2, on pose :

$$\hat{S}_t = \frac{X_t}{X_t^*};$$

Puis on estime les coefficients saisonniers et on obtient les estimations \bar{c}_j , $j = 1, \dots, P$ de la même manière que pour le modèle additif. On pose ensuite : $\bar{c}_j = P \frac{\hat{c}_j}{\sum_{j=1}^P \hat{c}_j}$, $j = 1, \dots, P$ de

façon à respecter la contrainte sur la période $\sum_{j=1}^P c_j = P$

- à l'étape 3, on pose :

$$X_{CVS,t} = \frac{X_t}{\hat{S}_t}, \quad t = m + 1, \dots, T - m$$

2.5 Application sous le logiciel R

R est un logiciel pour l'analyse statistique des données initié par Robert Gentleman et **Ross Ihaka** (Université d'Auckland, Nouvelle-Zélande), auxquels sont venus s'ajouter un noyau de chercheurs du monde entier en 1997, il constitue aujourd'hui un langage et un environnement de programmation intégré d'analyse statistique. Il fournit les procédures usuelles (t-tests, anova, tests non paramétriques...) et possède des possibilités graphiques pour explorer les données, illustrer quelques concepts fondamentaux de la statistique inférentielle. Pouvant être utilisé aussi bien en mode interactif qu'en mode batch, R est un logiciel libre, dont le code source est disponible et qui peut être recopié et diffusé gratuitement. Des versions compilées de R sont disponibles pour Linux, Windows et Mac OS X. Il répond à tous les calculs et représentations nécessaires dans l'exercice professionnel de l'enseignant de mathématique ou de sciences. Étudions les ventes d'un produit P sur quatre années. Les résultats sont consignés dans le tableau suivant :

Année	Numéro du trimestre	Ventes
2011	1	30
	2	11
	3	12
	4	36
2012	5	32
	6	12
	7	13
	8	37
2013	9	33
	10	13
	11	15
	12	39
2014	13	35
	14	14
	15	17
	16	41

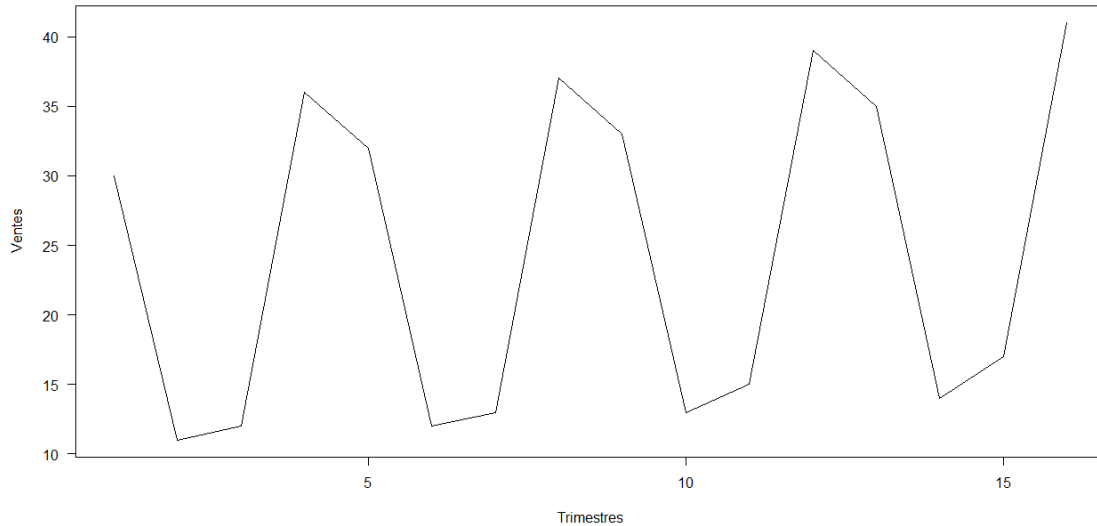


FIGURE 2.3 – Chronogramme des ventes d'un produit

- Commençons par représenter cette série chronologique. Avant de modéliser et étudier la série, on va tester les trois méthodes afin de choisir le bon modèle.

* **Méthode de la bande** : Comme les deux droites sont presque parallèles, alors le modèle

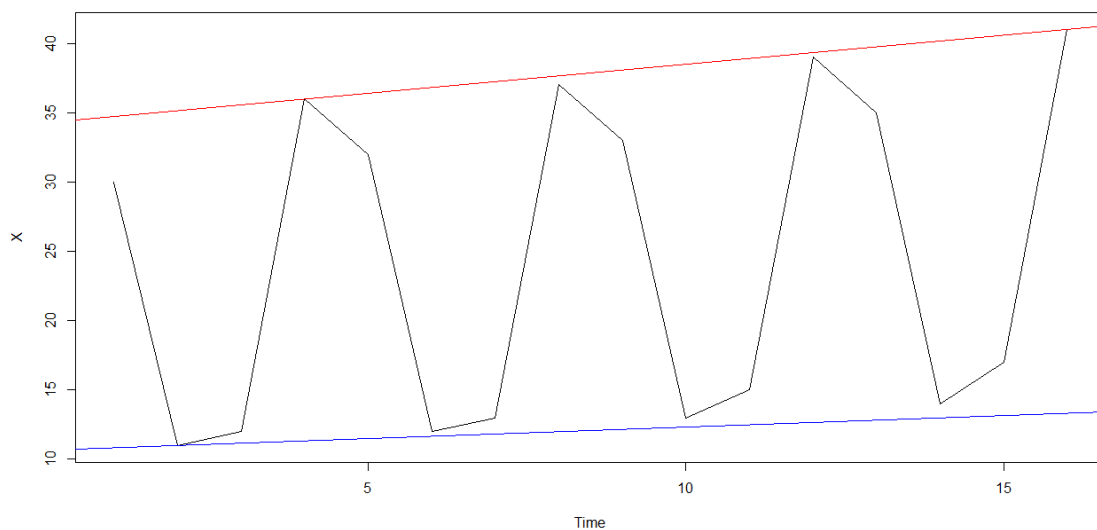


FIGURE 2.4 – Chronogramme des ventes d'un produit/ Méthode de la bande

est additif.

* **Méthode du tableau de Buys-Ballot** :

Dates	T1	T2	T3	T4
A1	30	11	12	36
A2	32	12	13	37
A3	33	13	15	39
A4	35	14	17	41
Moyenne	32,5	12,5	14,25	38,25
Écart-type	8,32	11,93	10,3	14

Donc la régression de σ_t en fonction de la moyenne s'écrit sous la forme : $\sigma_t = a\bar{X}_t + b$ équivalent à dire que $\sigma_t = 0.03166\bar{X}_t + 10.36586$, maintenant il suffit d'appliquer le test de student pour voir est-ce que le coefficient a est significatif ou non, ce test est basé sur les deux hypothèses suivantes :

$$\begin{cases} H_0 : \hat{a} = 0 \\ H_1 : \hat{a} \neq 0 \end{cases}$$

La statistique qui convient pour le test est $t^* = \frac{|\hat{a}|}{\hat{\sigma}_X}$

Si $t^* \in [-t_{\frac{\alpha}{2}, n-2}; t_{\frac{\alpha}{2}, n-2}]$ on accepte H_0 .

On sait que :

$$Var(\bar{X}_t) = \frac{\sum_{i=1}^n (\sigma_t - \hat{\sigma}_t)^2 / (n-2)}{\sum_{i=1}^n (\bar{X}_t - \bar{X})^2}$$

En utilisant le logiciel R, plus précisément la commande `fit=lm(sigma moyenne)` détermine l'équation de la régression de σ_t en fonction de la moyenne, ensuite l'utilisation de la commande `summary(fit)` nous permettons de déterminer les coefficient $a = 0.03166$, $b = 10.36586$ et aussi $t^* = 0.2436$, $t_{\frac{\alpha}{2}, n-2} = 4.3023$. Alors on accepte H_0 , d'où le modèle est additif.

Appliquons la procédure de l'étude d'une série chronologique, que nous avons vu au section précédente, et à l'aide du logiciel R, nous avons obtenue les graphe suivants :

En utilisant la commande " `m=decompose(Ventes)` ", on peut décomposer la série chrono-

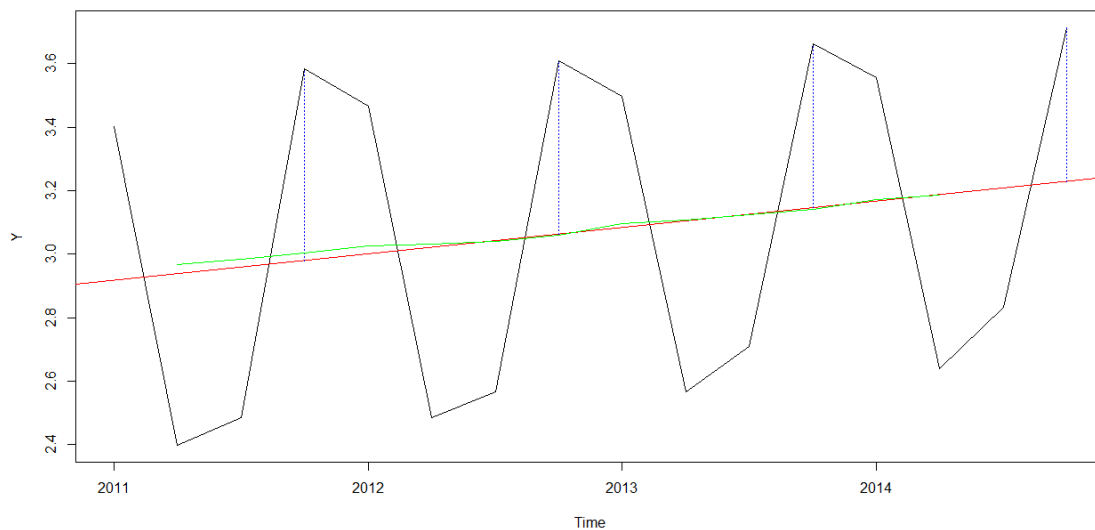


FIGURE 2.5 – la série filtrée et la tendance des ventes d'un produit

gique et on obtient :

Saisonnalité	T1	T2	T3	T4
A1	8.93750	-11.81250	-10.39583	13.27083
A2	8.93750	-11.81250	-10.39583	13.27083
A3	8.93750	-11.81250	-10.39583	13.27083
A4	8.93750	-11.81250	-10.39583	13.27083

La somme des coefficients saisonniers est nulle sur chaque période $A1, A2, A3, A4$.

Trend	T1	T2	T3	T4
A1	NA	NA	22.500	22.875
A2	23.125	23.375	23.625	23.875
A3	24.250	24.750	25.250	25.625
A4	26.000	26.500	NA	NA

Random	T1	T2	T3	T4
A1	NA	NA	-0.1041667	-0.1458333
A2	-0.0625000	0.4375000	-0.2291667	-0.1458333
A3	-0.1875000	0.0625000	0.1458333	0.1041667
A4	0.0625000	-0.6875000	NA	NA

A l'aide de la commande " plot(m) ", on peut décomposer la série chronologique, le graphe est le suivant :

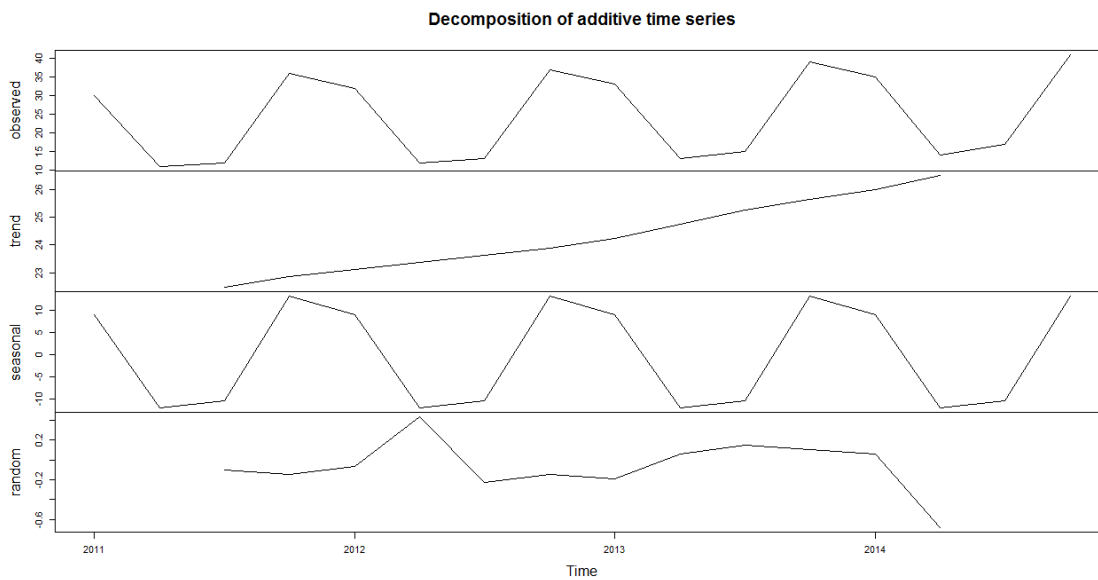


FIGURE 2.6 – Décomposition du chronogramme des ventes d'un produit

CHAPITRE

3

GÉNÉRALITÉS ET RAPPELS SUR LES PROCESSUS STOCHASTIQUES

Avant de définir la modélisation aléatoire des séries temporelles, il convient de faire un certain nombre de rappels succincts. Les rappels proposés dans le cadre de cette section portent avant tout sur les probabilités et les statistiques usuelles.

3.1 Espace L^2

On donne ici brièvement quelques éléments importants sur l'espace L^2 .

Définition 10

On dit qu'une v.a.r. est dans l'espace $L^2(\Omega, A, \mathbb{P})$, si l'on a $E(X^2) < +\infty$.

Cet espace $L^2(\Omega, A, \mathbb{P})$ est un espace de Hilbert muni du produit scalaire $\langle X, Y \rangle = E(XY)$ et de la norme $\|X\|_{L^2} = \sqrt{E(X^2)}$.

On a donc dans cet espace : $X \perp Y \Leftrightarrow E(XY) = 0 \Leftrightarrow Cov(X, Y) = 0$ si les v.a. X et Y sont centrées.

[7]

On rappelle l'inégalité de **Cauchy-Schwarz** pour les v.a. de carré intégrables.

Théorème 1

Si X et Y sont deux v.a. dans $L^2(\Omega, A, \mathbb{P})$ alors la v.a. XY est dans $L^1(\Omega, A, \mathbb{P})$ et on a : $\|XY\|_{L^1} \leq \|X\|_{L^2} \|Y\|_{L^2}$. [7]

Les projections dans l'espace L^2 sont facilitées par la propriété d'Hilbert de cet espace. Rappelons d'abord la définition d'un sous espace vectoriel fermé d'un espace de Hilbert (écrite dans le cas où l'espace de Hilbert est $L^2(\Omega, A, \mathbb{P})$).

Définition 11

Un sous-espace vectoriel \mathcal{H} de l'espace $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est dit fermé s'il contient toutes ses limites, i.e. si :

$$(X_n \in \mathcal{H}, \forall n \in \mathbb{N} \text{ et } \|X_n - X\|_{L^2} \rightarrow 0, \text{ quand } n \rightarrow +\infty) \implies X \in \mathcal{H}.$$

[7]

Le théorème de projection qui suit est très connu et sera utilisé en prévision des séries temporelles.

Théorème 2

Soit X une v.a. de l'espace $L^2(\Omega, \mathcal{A}, \mathbb{P})$ et \mathcal{H} un sous espace vectoriel fermé de cet espace $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Il existe alors une unique v.a. \hat{X} dans \mathcal{H} telle que :

$$\|X - \hat{X}\|_{L^2} = \min_{Y \in \mathcal{H}} \|X - Y\|_{L^2}$$

La v.a. \hat{X} correspond à **la projection orthogonale** de X sur l'espace \mathcal{H} telle que :

$$\hat{X} \in \mathcal{H} \text{ et } X - \hat{X} \perp \mathcal{H}$$

[7]

L'existence dans L^2 d'une notion d'orthogonalité et d'une notion de convergence permet d'introduire la notion de projection orthogonale.

On rappelle également la définition de convergence dans L^2 . Nous l'utiliserons ici pour justifier la convergence de sommes d'un nombre infini de v.a. dans L^2 .

Définition 12

Soient $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ et X une v.a. dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. On dit que $(X_n)_{n \in \mathbb{N}}$ converge dans L^2 vers X si l'on a :

$$\lim_{n \rightarrow +\infty} \|X_n - X\|_{L^2} = 0.$$

[7]

La proposition suivante nous sera très utile dans la suite pour autoriser l'inversion entre les signe somme et le signe intégral (l'espérance mathématique).

Proposition 9

Si la série $\sum_{i=0}^n X_i$ converge dans L^2 (vers $\sum_{i=0}^{+\infty} X_i$), alors on a :

$$E\left(\sum_{i=0}^{+\infty} X_i\right) = \sum_{i=0}^{+\infty} E(X_i)$$

[7]

Proposition 10

Si les séries $\sum_{i=0}^n X_i$ et $\sum_{j=0}^n Y_j$ converge dans L^2 (vers $\sum_{i=0}^{+\infty} X_i$ et $\sum_{j=0}^{+\infty} Y_j$), alors on a :

$$E\left(\sum_{i=0}^{+\infty} X_i \times \sum_{j=0}^{+\infty} Y_j\right) = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} E(X_i Y_j)$$

[7]

L'extension aux séries bidimensionnelles reste vraie ici aussi. Enfin, il découle des deux précédentes proposition, le corollaire suivant.

Corollaire 1

Si les séries $\sum_{i=0}^n X_i$ et $\sum_{j=0}^n Y_j$ converge dans L^2 (vers $\sum_{i=0}^{+\infty} X_i$ et $\sum_{j=0}^{+\infty} Y_j$), alors on a :

$$Cov\left(\sum_{i=0}^{+\infty} X_i, \sum_{j=0}^{+\infty} Y_j\right) = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} Cov(X_i, Y_j)$$

[7]

3.2 La régression affine

Considérons maintenant une famille $(X_i)_{i \in I}$ (I étant fini ou dénombrable) de v.a.r. de carré intégrable. Soit $L_{((X_i)_{i \in I})}$ le plus petit sous-espace vectoriel fermé de L^2 contenant toutes les combinaisons affines des variables X_i (et donc leurs limites au sens L^2).

Définition 13

On appelle régression affine de Y sur $(X_i, i \in I)$ la projection orthogonale de Y sur $L_{((X_i)_{i \in I})}$ notée Y^* .

Etant donnée la variable Y de carré intégrable, la variable Y^* est la meilleure approximation de Y au sens de L^2 par un élément de $L_{((X_i)_{i \in I})}$ puisqu'elle vérifie :

$$E[(Y - Y^*)^2] = \min_{Z \in L_{((X_i)_{i \in I})}} E[(Y - Z)^2]$$

De plus, on a : $Y = Y^* + R = a_0 + \sum_{i \in I} a_i X_i + R$ avec $R \perp X_i, i \in I$. [13]

Proposition 11

Soit $X = (X_1, \dots, X_n)$. Si (X, Y) est un vecteur gaussien, la régression affine de Y sur (X_1, \dots, X_n) coïncide avec la projection orthogonale de Y sur L_X^2 : $Y^* = E(Y|X)$.

[13]

cette propriété montre le lien entre espérance conditionnelle et la régression affine.

3.3 Notions de Convergence

Définition 14

Soit $(X_T)_{T=1, \dots, \infty}$ une séquence de variables aléatoires scalaires. Cette séquence **converge en probabilité** vers c , $\forall c \in \mathbb{C}, \forall \varepsilon > 0, \forall \delta > 0, \exists N, \forall T \geq N : \mathbb{P}[|X_T - c| > \delta] < \varepsilon$.

Alors, on note : $X_T \xrightarrow{p} c \Leftrightarrow \text{plim } X_T = c$ [9]

Exprimée autrement, cette définition signifie que pour un échantillon de taille infinie, la probabilité que la réalisation de la variable X_T diffère de la valeur c de plus ou moins δ (δ étant aussi petit que l'on veut) est inférieure à toute valeur ε aussi petite soit-elle.

Théorème 3

Slutsky

Soit $(X_T)_{T=1, \dots, \infty}$ une suite de $(n, 1)$ vecteurs admettant une limite en probabilité définie par c , et soit $g(\cdot)$ une fonction continue en c , satisfaisant $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, et ne dépendant pas de T , alors :

$$g(X_T) \xrightarrow{T \rightarrow \infty} g(c)$$

[9]

L'idée est la suivante. Si la fonction $g(\cdot)$ est continue, la quantité $g(X_T)$ se situera au voisinage de $g(c)$, dès lors que X_T se situe au voisinage de c . En choisissant une valeur de T suffisamment grande, la probabilité que la réalisation de X_T se situe au voisinage de c peut être définie aussi proche de l'unité que l'on le désire.

3.3.1 Convergence en moyenne quadratique

Une forme de convergence plus restrictive que la convergence en probabilité est la convergence en moyenne quadratique (m.s. pour mean square convergence).

Définition 15

Une suite de suite de v.a.r. $(X_T)_{T=1, \dots, \infty}$ **converge en moyenne quadratique** vers c , $\forall \varepsilon > 0, \exists N, \forall T \geq N : E[(X_T - c)^2] < \varepsilon$.

Alors, on note : $X_T \xrightarrow{m.s.} c$. [9]

La notion de convergence en m.q. nous permet alors d'introduire l'inégalité de **Chebyshev**.

Proposition 12**Inégalité de Chebyshev**

Soit X une v.a.r. telle que la quantité $E(|X|^r)$ existe et soit finie pour $r > 0$; $\forall \delta > 0, \forall c$, on montre que :

$$\mathbb{P}[|X - c| > \delta] \leq \frac{E(|X - c|^r)}{\delta^r}$$

[9]

Le résultat selon lequel la convergence en moyenne quadratique implique la convergence en probabilité peut être démontré à partir de l'inégalité de Chebyshev.

Une autre application possible de la notion de convergence en m.q. consiste à démontrer la loi faible des grands nombres.

3.3.2 Convergence en loi

Le troisième type de convergence que nous utiliserons est la convergence en loi ou convergence en distribution.

Théorème 4**Théorème de Paul Levy**

soit (X_T) une suite de v.a.r. et soit $F_{X_T}(x)$ la fonction de répartition cumulative de X_T . Si (X_T) converge en loi vers une v.a.r. X admettant $F_X(x)$ pour fonction de répartition, alors :

$$\lim_{T \rightarrow +\infty} F_{X_T}(x) = F_X(x) \quad \forall x \in \mathbb{R}$$

[9]

Théorème 5**Théorème Central Limite**

Soit $(X_t)_{t=1, \dots, \infty}$ une suite de v.a.r. i.i.d : telle que $E(X_T) = \mu$ et $Var(X_T) = \sigma^2 \forall t$, alors :

$$\sqrt{T}(\bar{X}_T - \mu) \xrightarrow{T \rightarrow +\infty} N(0, \sigma^2) \text{ converge en loi.}$$

avec $\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$. [9]

Une des hypothèses importantes du T.C.L tient à l'indépendance des v.a.r. X_t .

3.4 processus stochastique

Une série temporelle, quand elle n'est pas déterministe, comme ce sera le cas dans toutes les situations que nous rencontrerons, est vue comme une réalisation d'un processus stochastique.

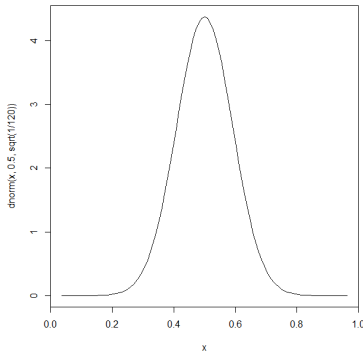


FIGURE 3.1 – Simulation de la convergence en loi.

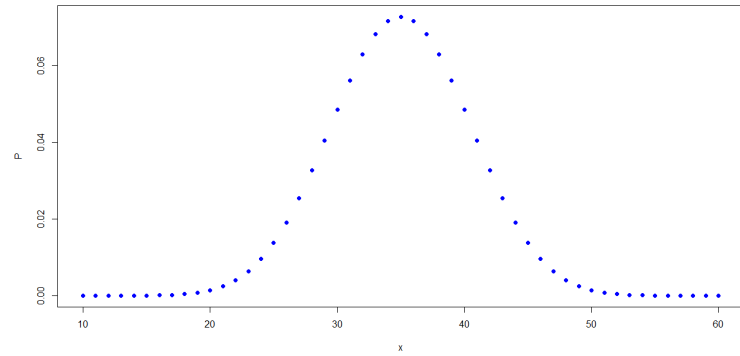


FIGURE 3.2 – Illustration du théorème central limite.

Définition 16

On appelle processus stochastique toute famille de variables aléatoires (v.a.) $(X_t)_{t \in T}$ d'un espace probabilisé (Ω, A, \mathbb{P}) vers un espace probabilisable (E, ε) . L'ensemble T est appelé espace des temps et E espace des états. Chacun de ces espaces peut être discret ou continu. [14]

3.4.1 Premiers exemples de séries temporelles

1- Bruits Blancs

Définition 17

On appelle **bruit blanc faible**, toute suite de v.a.r. $(X_t)_{t \in \mathbb{Z}}$, centrées, non corrélées et de variance σ^2 . On note : $(X_t) \sim bb(0, \sigma^2)$. [14]

Définition 18

Un **bruit blanc fort** est une suite de variables $(X_t)_{t \in \mathbb{Z}}$ centrées, indépendantes et identiquement distribuées (i.i.d.) et de variance σ^2 . On note : $(X_t) \sim bbF(0, \sigma^2)$. [14]

Par définition (18) il n'y a donc aucune dépendance entre les observations. Ainsi, par exemple, la connaissance de X_1, \dots, X_n , n'informe en rien sur la valeur (future) de X_{n+h} . Le processus binaire est un exemple de bruit blanc fort où la loi pour tout t est donnée par : $\mathbb{P}(X_t = 1) = p$ et $\mathbb{P}(X_t = -1) = 1 - p$.

Définition 19

On appelle **bruit blanc gaussien**, tout bruit blanc fort pour lequel la loi commune des v.a.r. $(X_t)_{t \in \mathbb{Z}}$ est une $\mathcal{N}(0, \sigma^2)$. [14]

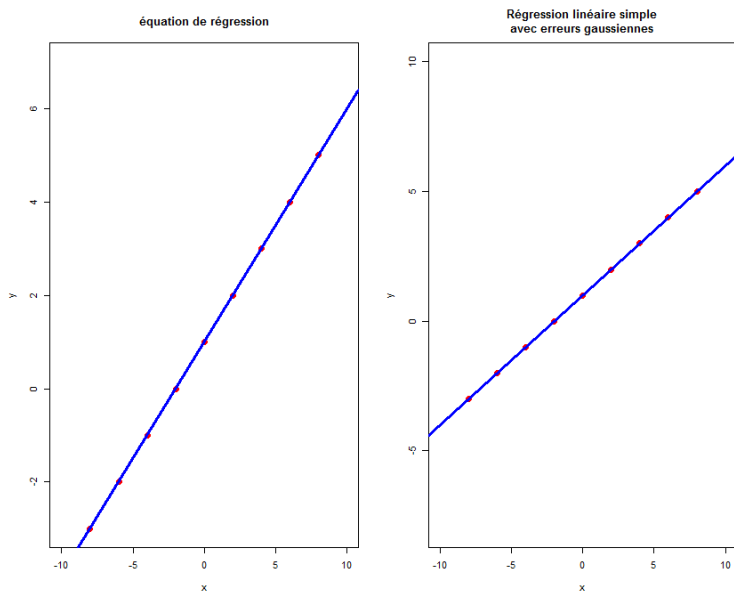


FIGURE 3.3 – Simulation selon un modèle de régression simple avec bruit gaussien.

2- Marche aléatoire

Une marche aléatoire $(S_t)_{t \in \mathbb{N}}$ est obtenue en écrivant $S_t = X_1 + \dots + X_t$ pour tout t dans \mathbb{N} où $(X_t)_{t \in \mathbb{N}}$ est un bruit blanc fort. Si ce dernier est un processus binaire, alors la marche aléatoire $(S_t)_{t \in \mathbb{N}}$ est dite **symétrique simple**.

3- Processus gaussien

Un processus gaussien à temps discret $(X_t)_{t \in \mathbb{Z}}$ est une série temporelle telle que la loi de n'importe quel vecteur extrait est gaussien, i.e.

$$\forall n \in \mathbb{N}^*, \forall (t_1, \dots, t_n) \in \mathbb{Z}^* : (X_{t_1}, \dots, X_{t_n}) \text{ est un vecteur gaussien.}$$

Un bruit blanc gaussien est une série temporelle gaussienne.

3.5 Processus stationnaire du second ordre

Dans toute la suite on considérera $(X_t)_{t \in \mathbb{Z}}$ et on supposera $X_t \in L^2(\Omega, A, \mathbb{P}), \forall t \in \mathbb{Z}$. Dans de très nombreux cas, on ne peut pas renouveler la suite de mesures dans des conditions parfaitement identiques (météo, économie...). Alors pour que le modèle déduit à partir d'une suite d'observations ait un sens, il faut que toute portion de trajectoire observée fournisse des informations sur la loi de X et que des portions différentes mais de même longueur fournissent les mêmes indications. C'est ce qui nous amène à définir la notion de stationnarité.

Définition 20

Stationnarité stricte ou forte

$(X_t)_{t \in \mathbb{Z}}$ est un processus fortement stationnaire si :

$$\forall n \in \mathbb{N}^*, \forall (t_1, \dots, t_n), \quad \forall h \in \mathbb{Z} : (X_{t_1}, \dots, X_{t_n}) \text{ est identique à la loi de } (X_{t_1+h}, \dots, X_{t_n+h}).$$

[7]

Théorème 6

Théorème de Kolmogorov

$(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire au sens strict si et seulement si la loi de $(X_t)_{t \in \mathbb{Z}}$ est identique à la loi de $(Y_t)_{t \in \mathbb{Z}}$ où $Y_t = X_{t+h}$. [7]

Définition 21

Stationnarité faible

$(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire du second ordre (ou un processus faiblement stationnaire) s'il vérifie :

- (i) $\forall t \in \mathbb{Z}, E(X_t) = m$
 - (ii) $\forall t \in \mathbb{Z}, Var(X_t) = \sigma^2 = \gamma(0)$
 - (iii) $\forall t \in \mathbb{Z}, \forall h \in \mathbb{Z} Cov(X_t, X_{t+h}) = \gamma(h)$
- $\gamma(h)$ est l'auto-covariance d'ordre h de X_t . [7]

Dans cette définition, la propriété (i) exprime la stationnarité en moyenne, (ii) assure que la variance de chaque variable est finie et (iii) précise ce que l'on entend par "**invariance de la structure de covariance**". Par cette définition, on peut dire que la fonction $\gamma(h)$ qui est indépendante de t et n'est définie que si le processus est stationnaire du second ordre (et vérifie donc (iii)). Nous reviendrons sur cette fonction et ses propriétés dans la section suivante.

Proposition 13

Si X est un processus stationnaire, $Var(X_t) = \gamma(0) = \sigma^2$ est indépendante de t . On dit que le processus est **homoscédastique**. [7]

Remarque 3.5.1. On peut également définir un concept de stationnarité à partir des lois jointes des variables aléatoires du processus X : Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est strictement stationnaire si et seulement si $\forall k \in \mathbb{N}, \forall t_1, \dots, t_n \in \mathbb{Z}, \forall h \in \mathbb{Z}$, la loi conjointe de $(X_{t_1+h}, \dots, X_{t_k+h})$ ne dépend pas de h , i.e.

$$\mathcal{L}(X_{t+h_1}, \dots, X_{t+h_k}) = \mathcal{L}(X_{h_1}, \dots, X_{h_k}), \forall t, \forall k(h_1, \dots, h_k).$$

Notons que si X est un processus strictement stationnaire, la loi de X_t ne dépend pas de t : les variables X_t , pour $t \in \mathbb{Z}$ sont identiquement distribuées.

Cette dernière définition de stationnarité est plus exigeante que le concept de stationnarité du second ordre puisque tout processus strictement stationnaire est stationnaire du second ordre. La réciproque est fautive. La stationnarité à l'ordre 2 est bien plus facile à étudier et vérifier que la stationnarité stricte.

L'importance pratique de la stationnarité à l'ordre 2 tient surtout aux problèmes de prédiction ou de régression. En effet, on se limite souvent à des critères de moindres carrés pour avoir des estimateurs calculables. Cela signifie alors utiliser des prédicteurs linéaires optimaux dont le calcul ne fait pas intervenir dans sa totalité la structure probabiliste du processus observé X mais seulement la géométrie (angles et longueurs) de la suite X_k considérée comme une suite de vecteurs dans l'espace de Hilbert $L^2(\Omega, \mathbb{P})$. Or cette géométrie ne dépend que des moments d'ordre 2 de X , la notion naturelle de stationnarité est donc l'invariance de ces moments d'ordre

2 par translation dans le temps.

Dans la suite, nous utiliserons le terme "stationnaire" pour un processus stationnaire de second ordre et nous nous cantonnerons simplement à ce type de stationnarité.

3.5.1 Fonctions d'autocovariance et d'autocorrélation

Les principales caractéristiques temporelles d'un processus sont données par l'autocorrélation (simple) et l'autocorrélation partielle.

Définition 22

La fonctions d'autocovariance

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

$$\forall h, t \in \mathbb{Z}, \quad \gamma(h) := Cov(X_t, X_{t+h}).$$

Le graphe de cette fonction est appelé *variogramme*. [13]

Proposition 14

La fonction d'autocovariance d'un processus stationnaire vérifie :

- 1- $\forall h \in \mathbb{Z}, \gamma(-h) = \gamma(h)$: elle est paire.
- 2- $\forall n \in \mathbb{N}, \forall (a_i) \in \mathbb{R}^n, \forall (t_i) \in \mathbb{Z}^n, \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(t_i - t_j) > 0$: elle de type positif.
- 3- $\gamma(0) = Var(X_t)$.
- 4- $|\gamma(h)| \leq \gamma(0), \forall h$. [14]

La fonction d'autocovariance peut être "normalisée" et la nouvelle fonction obtenue est la fonction d'autocorrélation.

Définition 23

La fonction d'autocorrélation

soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire. On appelle fonction d'autocorrélation la fonction ρ définie de \mathbb{Z} dans \mathbb{R} par :

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

Le graphe de cette fonction est appelé *corrélogramme*. [13]

Proposition 15

La fonction d'autocorrélation d'un processus stationnaire vérifie :

- 1- $\forall h \in \mathbb{Z}, \rho(-h) = \rho(h)$: elle est paire.
- 2- $\rho(0) = 1$.
- 3- $|\rho(h)| \leq 1, \forall h$. [14]

La fonction $\rho(h)$ est l'expression du lien linéaire entre X_t et X_{t-h} . Si t est 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 γ entre le présent et le futur d'ordre h); plus $|\rho(h)|$ est proche de 1 et plus ce lien est fort.

les valeurs $\rho(h)$, $h = 1, 2, \dots$ ne sont souvent pas suffisantes pour expliquer comment présent et passé sont reliés car les variables du passé sont elles mêmes reliées entre elles. Autrement dit, X_{t-h} dépend en général des variables $X_{t-h+1}, \dots, X_{t-1}$. On a donc recours à une nouvelle fonction, **la fonction d'autocorrélation partielle** $\rho(h)$ qui exprime le lien entre X_t et X_{t-h} lorsqu'on a retiré leur meilleure explication affine en termes de $X_{t-h+1}, \dots, X_{t-1}$. Cette fonction existe donc pour connaître jusqu'à quel niveau de décalage ou ordre, il existe une relation directe entre X_t et les valeurs précédentes.

Définition 24

La fonction d'autocorrélation partielle soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire. Posons $X_{t,h}^*$ la régression affine de X_t sur $(X_{t-1}, \dots, X_{t-h})$. On a alors :

$$X_t = X_{t,h}^* + R_{t,h} = \lambda_{0,h} + \sum_{s=1}^h \lambda_{s,h} X_{t-s} + R_{t,h}.$$

où $R_{t,h}$ est une variable aléatoire non corrélée avec $(X_{t-1}, \dots, X_{t-h})$. La fonction d'autocorrélation partielle τ est définie par :

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

Le graphe de cette fonction est appelé **corrélogramme partiel**. [13]

Théorème 7

L'autocorrélation partielle $\tau(h)$ d'ordre h représente le coefficient de corrélation linéaire entre :
 - Le résidu $X_t - X_{t,h-1}^*$ de la régression de X_t par $X_{t-h+1}, \dots, X_{t-1}$.
 - Le résidu $X_{t-h} - X_{t-h,h-1}^*$ de la régression de X_{t-h} par $X_{t-h+1}, \dots, X_{t-1}$.

En d'autres termes,

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_{h-1} X_{t-h+1} + U. \\ X_{t-h} &= \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_{h-1} X_{t-h+1} + V. \end{aligned}$$

et

$$\tau(h) = \text{Cor}(U, V).$$

[13]

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

Preuve 3.5.1. En effet, on a :

$$\frac{\text{Cov}(X_t - X_{t,h-1}^*, X_{t-h} - X_{t-h,h-1}^*)}{\sqrt{\text{Var}(X_t - X_{t,h-1}^*) \text{Var}(X_{t-h} - X_{t-h,h-1}^*)}} = \frac{\text{Cov}(R_{t,h-1}, R_{t-h,h-1})}{\sqrt{\text{Var}(R_{t,h-1}) \text{Var}(R_{t-h,h-1})}}. \quad (3.1)$$

De plus, la stationnarité de X entraîne que $Var(R_{t,h-1}) = Var(R_{t-h,h-1})$ et par conséquent le second membre de l'équation précédente ne dépend pas de t . Notons cette expression $\tau(h)$, sa valeur représente le coefficient de régression de $X_t - X_{t,h-1}^*$ sur $X_{t-h} - X_{t-h,h-1}^*$.

Écrivons maintenant la régression affine de X_t sur $(X_{t-1}, \dots, X_{t-h})$, on a

$$X_t = \lambda_{0,h} + \lambda_{1,h}X_{t-1} + \lambda_{h,h}X_{t-h} + R_{t,h}.$$

En projetant chaque membre de l'égalité sur $L_{X_{t-h+1}, \dots, X_{t-1}}$, on obtient :

$$\begin{aligned} X_{t,h-1}^* &= \lambda_{0,h} + \lambda_{1,h}X_{t-1,h-1}^* + \lambda_{h,h}X_{t-h,h-1}^* \\ &= \lambda_{0,h} + \lambda_{1,h}X_{t-1} + \lambda_{h,h}X_{t-h,h-1}^* \end{aligned}$$

Et par conséquent,

$$X_t - X_{t,h-1}^* = \lambda_{h,h}(X_{t-h} - X_{t-h,h-1}^*) + R_{t,h}.$$

où $R_{t,h}$ est orthogonal à $X_{t-h} - X_{t-h,h-1}^*$. De cette expression, on déduit que le coefficient de régression $\tau(h)$ de $X_t - X_{t,h-1}^*$ sur $X_{t-h} - X_{t-h,h-1}^*$ est égal à $\lambda_{h,h} = \rho(h)$. D'où le résultat.

Remarque 3.5.2. Généralement on utilise la statistique de **Durbin-Watson**¹ pour vérifier la présence d'une auto-corrélation dans les erreurs des séries chronologiques, car supposons que votre logiciel favori vous restitue un graphique des résidus et vous remarquez que les points qu'ils représentent tracent une sorte de courbe, ce qui signifie qu'ils contiennent une information "oubliée" dans le modèle, soit vous établissez un test non paramétrique de Durbin-Watson afin de fonder votre opinion de façon quantitative et non sur une visualisation.

Relation entre les fonctions d'autocorrélation et d'autocorrélation partielle :

Nous avons tout d'abord

$$\tau(1) = \lambda_{1,1} = \frac{Cov(X_t, X_{t-1})}{\sigma^2} = \rho(1)$$

Déterminons maintenant une expression explicite de $\tau(2)$. Pour cela, écrivons

$$X_t = \lambda_{0,2} + \lambda_{1,2}X_{t-1} + \lambda_{2,2}X_{t-2} + R_{t,2}.$$

avec $R_{t,2}$ indépendant à la fois de $X_{t-1} - X_{t-2}$. En multipliant les deux membres de cette égalité par X_{t-1} d'une part et X_{t-2} d'autre part, puis en prenant l'espérance, on obtient le système d'équations

$$\begin{cases} \rho(1) = \lambda_{1,2} + \lambda_{2,2}\rho(1). \\ \rho(2) = \lambda_{1,2}\rho(1) + \lambda_{2,2}. \end{cases}$$

On en déduit

$$\tau(2) = \lambda_{2,2} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$$

1. Par définition, la statistique de Durbin-Watson est donnée par : $DW = \frac{\sum_{i=2}^T (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^T \hat{\varepsilon}_i^2}$ La statistique

DW est comprise entre zéro et quatre. L'hypothèse nulle est acceptée lorsque la valeur de cette statistique est proche de deux. [Voir :Savin,N.E.et White, K.J.(1977), The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors, Econometrica,45(8),p.1989-1996.]

En généralisant la démarche précédente, on obtient pour déterminer $\tau(h)$ un système de h équations linéaires

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-2) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \rho(h-1) & \rho(h-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} \lambda_{1,h} \\ \lambda_{2,h} \\ \vdots \\ \lambda_{h,h} \end{pmatrix}$$

D'où l'on déduit que

$$\tau(h) = \lambda_{h,h} = \frac{\begin{vmatrix} 1 & \rho(1) & \dots & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \dots & \rho(h-3) & \rho(2) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho(h-1) & \rho(h-2) & \rho(1) & \rho(h) & \rho(h) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \dots & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-3) & \rho(h-2) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho(h-1) & \rho(h-2) & \rho(1) & \rho(h) & 1 \end{vmatrix}} = \frac{\rho(h) - \sum_{i=1}^{h-1} \lambda_{h-1,i} \rho(h-i)}{1 - \sum_{i=1}^{h-1} \lambda_{h-1,i} \rho(i)}$$

où

$$\lambda_{h,i} = \lambda_{h-1,i} - \lambda_{h,h} \times \lambda_{h-1,h-i} \quad \text{pour } i = 1, \dots, h-1$$

Ceci montre que $\tau(h)$ est fonction de $\rho(1), \dots, \rho(h)$. Inversement, $\rho(h)$ est fonction de $\tau(1), \dots, \tau(h)$ et bien sûr $\tau(1) = \rho(1)$.

La connaissance de $\rho(h)$ est donc équivalente à celle de $\tau(h)$. Selon les cas, il sera plus commode d'utiliser l'une ou l'autre de ces fonctions.

Exemple 3.5.3. Calculons la fonction d'autocorrélation et la fonction d'autocorrélation partielle pour le bruit blanc $(\eta_t)_{t \in \mathbb{Z}} \sim N(0, 50)$:

* **la fonction d'autocorrélation :**

Par définition de la covariance on a :

$$\gamma_h = \text{Cov}(\eta_t, \eta_{t-h}) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}$$

D'où la fonction d'autocorrélation est :

$$\rho(h) = \rho_h = \frac{\gamma_h}{\gamma_0} = \begin{cases} 1 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}$$

Son graphe est le suivant :

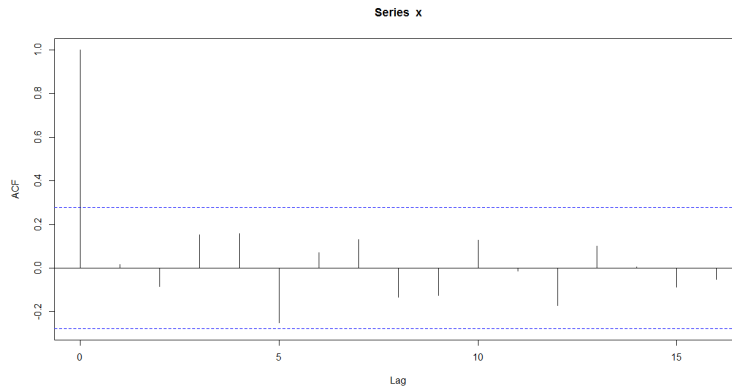


FIGURE 3.4 – La fonction d'autocorrélation empirique d'un bruit blanc

par définition on a $\lambda_{0,0} = 1$ et $\lambda_{1,1} = \rho_1$

$$\lambda_{h,h} = \begin{cases} 1 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}$$

son graphe est le suivant :

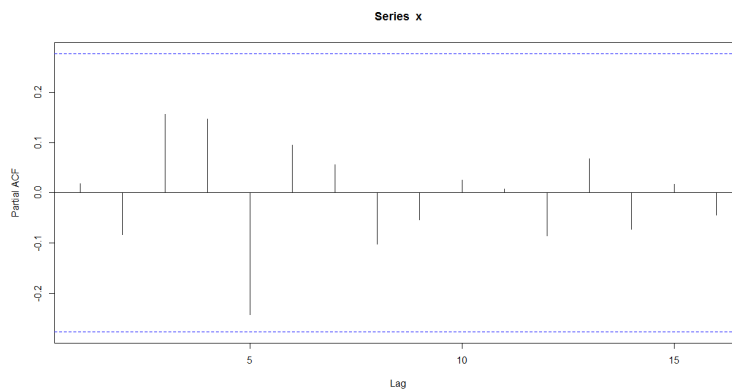


FIGURE 3.5 – La fonction d'autocorrélation empirique partielle d'un bruit blanc

Nous terminons le paragraphe par une dernière définition

Définition 25

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus de carré intégrable et X_t^* la régression affine de X_t sur $(X_s, s < t)$. On appellera **innovation du processus** à la date t la variable aléatoire réelle $\epsilon_t^* := X_t - X_t^*$. [8]

Exemple 3.5.4. Lorsque $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, la suite des innovations $(\epsilon_t^*)_{t \in \mathbb{Z}} := (X_t - X_t^*)_{t \in \mathbb{Z}}$ est un bruit blanc.

CHAPITRE

4

PROCESSUS ALÉATOIRES STATIONNAIRES

Nous avons jusqu'à présent introduit une étude descriptive des séries temporelles afin d'identifier ou supprimer les éventuelles tendance et saisonnalité. Puis nous avons introduit la modélisation probabiliste des séries temporelles en considérant les notions de stationnarité, de fonction d'autocovariance ou d'autocorrélation ainsi que la fonction d'autocorrélation partielle. Avant d'introduire et entamer une étude détaillée des modèles particuliers de processus stationnaires que sont les processus MA, AR, ARMA, Nous verrons que tout processus stationnaire peut s'écrire comme une somme pondérée de bruits blancs (**théorème de Wold**).

Le théorème de Wold(1938) est le théorème fondamental de l'analyse des séries temporelles stationnaires.

4.1 Théorème de Wold

Théorème 8

Tout processus stationnaire d'ordre deux $(X_t)_{t \in \mathbb{Z}}$ peut être représenté sous la forme :

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

où les paramètres ψ_i satisfont $\psi_0 = 1$, $\psi_i \in \mathbb{R}$, $\forall i \in \mathbb{N}^*$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ et où ε_t est un bruit blanc i.i.d. $(0, \sigma^2)$, On dit que la somme des chocs passés correspond à la composante linéaire stochastique de X_t . Le terme η_i désigne la composante linéaire déterministe telle que $Cov(\eta_i, \varepsilon_{t-i}) = 0$, $\forall i \in \mathbb{Z}$. [9]

Ainsi, d'après le théorème de Wold¹, tout processus stationnaire peut s'écrire comme une somme pondérée infinie de chocs passés, ces chocs étant représentés par un bruit blanc de variance finie. L'implication forte de ce théorème est que, si l'on connaît les pondérations ψ_i , $\forall i \in \mathbb{N}$, et si l'on connaît la variance σ_ε^2 du bruit blanc, on est en mesure de proposer une représentation de n'importe quel processus stationnaire. Cette représentation est aussi qualifiée de **représentation moyenne mobile infinie**.

Reste à comprendre ce que peut être cette composante linéaire déterministe η_t . La condition $Cov(\eta_i, \varepsilon_{t-i}) = 0$, implique que ce terme est, par définition (déterministe), indépendant des chocs. Alors le cas le plus simple est celui d'un processus stationnaire $(X_t, t \in \mathbb{Z})$ d'espérance non nulle, tel que $E(X_t) = m \neq 0$. Puisque le bruit blanc est par définition un processus centré, une somme pondérée de ces chocs est elle-même centrée. Par conséquent, la représentation de Wold du processus $(X_t, t \in \mathbb{Z})$ suppose que l'on ajoute à cette somme pondérée des chocs passés, une composante déterministe qui n'est autre que l'espérance du processus, $\eta_t = m$. On a donc :

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

et l'on vérifie que :

$$E(X_t) = E\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}\right) + m = \sum_{i=0}^{\infty} \psi_i E(\varepsilon_{t-i}) + m = m.$$

4.1.1 Conditions sur les pondérations ψ_i

L'énoncé du théorème de Wold fait clairement apparaître trois conditions sur les pondérations ψ_i .

- La première condition $\psi_0 = 1$, impose tout d'abord que le poids du choc présent ε_t dans la définition du processus soit unitaire. Il s'agit ici tout simplement d'une condition de normalisation qui porte sur la détermination de la variance du bruit blanc.
- La seconde condition, $\psi_i \in \mathbb{R}$, $\forall i \in \mathbb{N}^*$, est triviale et signifie simplement que les pondérations des chocs strictement passés peuvent être éventuellement nulles pour certains retards.
- la troisième condition, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ est particulièrement importante.

Définition 26

La condition $\sum_{i=0}^{\infty} \psi_i^2 < \infty$, dite condition de sommabilité des carrés (ou d'intégrabilité des carrés) assure l'existence des moments d'ordre deux du processus $(X_t, t \in \mathbb{Z})$. Sous cette condition, on dit alors que X_t converge en moyenne quadratique. [9]

Pour comprendre en quoi cette condition de sommabilité assure l'existence des moments théoriques d'ordre deux, il faut revenir à la définition de la fonction génératrice d'autocovariance du processus $(X_t, t \in \mathbb{Z})$ et utiliser la représentation de Wold. Si l'on suppose pour simplifier que $E(X_t) = 0$, on obtient alors :

$$\gamma(h) = E(X_{t+h}X_t) = E\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t+h-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right), \quad \forall h \in \mathbb{Z}$$

1. La démonstration de ce théorème est donnée dans Wold(1938), voir Sargent, Macroeconomic Theory, Boston Academic Press, p 286-290 pour l'intuition de la démonstration.

Étant donné la définition d'un bruit blanc on a $E(\varepsilon_t \psi_i \varepsilon_{t-i}) = 0, \forall i \neq 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2$. Dès lors, puisque l'espérance est un opérateur linéaire, la fonction $\gamma(h)$ peut se réécrire sous la forme :

$$\gamma(h) = \begin{cases} \sum_{i=0}^{\infty} \psi_i^2 E(\varepsilon_{t-i}^2) & \text{si } h = 0 \\ \sum_{i=0}^{\infty} \psi_{i+h} \psi_i E(\varepsilon_{t-i}^2) & \text{si } h \neq 0 \end{cases}$$

D'où normalement on tire que :

$$\gamma(h) = \begin{cases} \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2 & \text{si } h = 0 \\ \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_{i+h} \psi_i & \text{si } h \neq 0 \end{cases}$$

On peut montrer que la condition $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ implique $\sum_{i=0}^{\infty} \psi_{i+h} \psi_i < \infty, \forall h \neq 0$. Dès lors cette condition suffit à garantir la convergence, et donc l'existence, de l'ensemble des moments d'ordre deux du processus X_t .

4.1.2 Prévisions à partir de la décomposition de Wold

Comme nous l'avons dit, une des conséquences de ce théorème est que, si l'on connaît les pondérations $\psi_i, \forall i \in \mathbb{N}$, et si l'on connaît la variance σ_ε^2 du bruit blanc, on est mesure de proposer une représentation de n'importe quel processus stationnaire. Partant de cette représentation, on peut alors réaliser une prévision, notée \hat{X}_{t+h} , pour la réalisation du processus X_t à la date $t+h$.

Considérons, pour commencer, une prévision à période ($h=1$). La meilleure prévision possible de la réalisation de X_{t+1} , notée \hat{X}_{t+1} , connaissant les valeurs passées X_t, X_{t-1}, \dots , est donnée par l'espérance conditionnelle suivante $E(X_{t+1}|X_t, X_{t-1}, \dots)$. Or, d'après le théorème de Wold, la connaissance des valeurs passées de X_t est équivalente à la connaissance des valeurs passées des chocs $\varepsilon_t, \varepsilon_{t-1}, \dots$. En effet, on sait que :

$$\begin{aligned} X_t &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \\ X_{t-1} &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-1-i} = \varepsilon_{t-1} + \psi_1 \varepsilon_{t-2} + \psi_2 \varepsilon_{t-3} + \dots \\ X_{t-p} &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i-p} = \varepsilon_{t-p} + \psi_1 \varepsilon_{t-1-p} + \psi_2 \varepsilon_{t-2-p} + \dots \quad \forall p > 0 \end{aligned}$$

Par conséquent, l'espace de conditionnement de notre prévision peut être modifié de la façon suivante :

$$\begin{aligned} \hat{X}_{t+1} &= E(X_{t+1}|X_t, X_{t-1}, \dots) \\ &= E(X_{t+1}|\varepsilon_t, \varepsilon_{t-1}, \dots) \end{aligned}$$

Reste alors à considérer la représentation de Wold du processus X_{t+1} :

$$X_{t+1} = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i+1} = \varepsilon_{t+1} + \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \dots$$

A partir de cette formule, il est facile de déterminer l'espérance conditionnelle :

$$\begin{aligned}\hat{X}_{t+1} &= E(X_{t+1}|\varepsilon_t, \varepsilon_{t-1}, \dots) \\ &= E\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i+1} | \varepsilon_t, \varepsilon_{t-1}, \dots\right) \\ &= \sum_{i=0}^{\infty} \psi_i E(\varepsilon_{t-i+1} | \varepsilon_t, \varepsilon_{t-1}, \dots)\end{aligned}$$

On obtient alors la forme générale de notre prévision :

$$\hat{X}_{t+1} = \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i+1} = \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \dots$$

et par conséquent de l'erreur de prévision :

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

Par définition cette erreur de prévision est d'espérance nulle et de variance égale à σ_ε^2 . Ainsi, pour un processus stationnaire l'erreur de prévision optimale à un horizon d'une période correspond à l'innovation fondamentale (au choc) associé à cette même période. De plus, cette erreur de prévision a de "bonnes" propriétés puisqu'elle est strictement indépendante de l'erreur de prévision que l'on a pu commettre aux dates précédentes puisque $E(\varepsilon_t, \varepsilon_{t-j}) = 0$, ce n'est pas parce qu'hier l'économétrie s'est trompé, qu'il se trompera nécessairement demain.

Bien entendu, on peut généraliser ces résultats pour des prévisions à des ordres h quelconques.

Résultat 1

Soit \hat{X}_{t+h} la meilleure prévision possible de la réalisation du processus X_t à la date $t+h$ compte tenu de l'information disponible à la date t .

$$\hat{X}_{t+h} = \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i+h}$$

L'erreur de prévision associée est alors déterminée par :

$$X_{t+h} - \hat{X}_{t+h} = \sum_{i=0}^{h-1} \psi_i \varepsilon_{t-i+h}$$

[9]

Remarque 4.1.1. *Le théorème de Wold n'est jamais directement applicable, en tant que tel, dans la pratique puisque il suppose que l'on exprime un processus stationnaire comme une somme pondérée infinie de chocs passés. Or, dans la pratique informatique il n'est pas possible de manier des objets mathématiques de dimension infinie. Pour autant, on peut démontrer qu'il existe toujours pour tout processus stationnaire, un ordre fini, tel que ce processus puisse être approximé par une somme pondérée de chocs passés jusqu'à cet ordre (**décomposition de Wold tronqué**).*

Concrètement, cela signifie que n'importe quelle série stationnaire peut être représentée de façon satisfaisante par une somme pondérée finie de chocs passés.

4.2 Processus auto-régressifs d'ordre p

Ces processus forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes, ils sont donc définis implicitement par la relation :

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \delta + \eta_t \quad (4.1)$$

où $\phi_1, \phi_2, \dots, \phi_p, \delta$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . L'observation au temps t est donc la somme d'un choc aléatoire η_t à l'instant t (indépendant de l'historique) et d'une fonction linéaire de son passé $\sum_{i=1}^p \phi_i X_{t-i}$ (qui peut-être vue comme la prédiction de X_t à partir des p dernières valeurs observées. Nous reviendrons en détail sur cette propriété dans la section consacrée à la prédiction).

4.2.1 Définition et représentation canonique

Définition 27

On appelle processus autorégressif d'ordre p tout processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire tel que :

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \delta + \eta_t \quad (4.2)$$

où $\phi_1, \phi_2, \dots, \phi_p, \delta$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 . Un tel processus est dit *AR(p)* (**AutoRegressive of order p**). [7]

Posons :

$$\Phi(B) = I - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

On peut écrire :

$$\Phi(B)X_t = \delta + \eta_t, \quad t \in \mathbb{Z} \quad (4.3)$$

Remarque 4.2.1. Dans ce projet, nous ne considérerons que des processus non centrés, si $\delta = 0$ alors le processus est centré.

Écriture moyenne mobile infinie d'un AR(p)

Dans le cas régulier où le polynôme Φ ne s'annule pas sur le cercle unité, on a le théorème.

Théorème 9

(i) Si Φ n'a pas de racine de module égal à 1, $\Phi(B)$ est inversible et on en déduit que l'équation (1.3) a une solution unique, avec une écriture moyenne mobile infinie

$$\forall t \in \mathbb{Z}, \quad X_t = (\Phi(B))^{-1} \eta_t = \sum_{i \in \mathbb{Z}} \psi_i \eta_{t-i} \text{ avec } \sum_{i=-\infty}^{+\infty} |\psi_i| < \infty$$

Les coefficients ψ_i convergent rapidement vers 0 lorsque $i \rightarrow \infty$, en effet ce sont les coefficients du filtre associé au polynôme $\frac{1}{\Phi}$. [7]

Ainsi le processus X est bien stationnaire, déterminé de manière unique par la relation précédente et la valeur présente de X dépend à la fois du passé, du présent et du futur du bruit blanc.

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

$$\forall t \in \mathbb{Z}, \quad X_t = (\Phi(B))^{-1}\eta_t = \sum_{i=0}^{+\infty} \psi_i \eta_{t-i} \text{ avec } \sum_{i=-\infty}^{+\infty} |\psi_i| < \infty$$

C'est-à-dire que le processus X_t s'exprime en fonction de $\eta_s, s < t$ et d'après la définition, on voit que η_t n'est pas corrélé avec X_{t-1}, \dots . la variable η_t est donc l'innovation du processus à la date t et $\sum_{i=0}^{+\infty} \psi_i \eta_{t-i}$ est la régression affine de X_t sur $(X_s, s \leq t-1)$, il s'agit de la représentation canonique d'un processus $AR(p)$.

Bruit blanc d'innovation d'un AR(p)

Théorème 10

Si X est un $AR(p)$ alors X est purement innovant ^a et il existe un unique polynôme $\tilde{\Phi}$ (dit canonique) tel que le bruit blanc d'innovation de X soit de la forme $\epsilon = \tilde{\Phi}(B)X$. Ce polynôme est de degré p et s'obtient à partir de Φ . Il est donc de la forme

$$\tilde{\Phi}(z) = \prod_{1 \leq k \leq p} (1 - \mu_k z) \text{ avec } \forall k, |\mu_k| < 1.$$

[7]

a. on dit que X est purement innovant si X est non nul et si $H_{-\infty}(X) = \cap_t H_{<t}(X) = 0$. [où $\cap_t H_{<t}(X)$ est appelée *passé lointain* de X ou le passé linéaire de X jusqu'à l'instant t . Les $\cap_t H_{<t}(X)$ forment une chaîne croissante de sous-espaces de $H(X)$ [$H(X)$ est histoire de X c'est l'espace engendré par les $(X_t)_{t \in \mathbb{Z}}$]]

Exemple 4.2.2. Le processus autorégressif d'ordre 1 non centré, $AR(1)$ s'écrit sous la forme

$$\Phi(B)X_t = \delta + \eta_t \Leftrightarrow (1 - \phi_1 B)X_t = \delta + \eta_t \Leftrightarrow X_t = \delta + \eta_t + \phi_1 X_{t-1} \quad (4.4)$$

avec $\delta, \phi_1 \in \mathbb{R}$ et $(\eta_t)_{t \in \mathbb{Z}} \sim bb(0, \sigma^2)$

* **Espérance :**

On a : $(1 - \phi_1 B)X_t = \delta + \eta_t$ donc $X_t = \frac{\delta}{1 - \phi_1} + (1 - \phi_1 B)^{-1}\eta_t$

D'où $E(X_t) = \frac{\delta}{1 - \phi_1} + E[(1 - \phi_1 B)^{-1}\eta_t]$

On a $E[(1 - \phi_1 B)^{-1}\eta_t] = E[(\sum_{i=0}^{\infty} \phi_1^i B^i)\eta_t]$

Comme les racines de $1 - \phi_1 B = 0$ à l'extérieur du cercle unité, c'est-à-dire $|B| > 1$, i.e $B = \frac{1}{\phi_1}$

$|B| > 1 \Rightarrow |\frac{1}{\phi_1}| > 1 \Rightarrow |\phi_1| < 1$ d'où la série est stationnaire.

Donc $E[(\sum_{i=0}^{\infty} \phi_1^i B^i)\eta_t] = \sum_{i=0}^{\infty} \phi_1^i B^i E(\eta_t) = 0 \forall t$

Alors $E(X_t) = \frac{\delta}{1 - \phi_1}$

* **La fonction d'autocorrélation :**

Supposons que $E(X_t) = m$ alors on obtient $m = \frac{\delta}{1 - \phi_1} \Leftrightarrow \delta = m(1 - \phi_1)$

En remplace δ dans (1.6) et on trouve :

$$X_t = \delta + \eta_t + \phi_1 X_{t-1}$$

$$= m(1 - \phi_1) + \eta_t + \phi_1 X_{t-1}$$

$$= m - m \times \phi_1 + \eta_t + \phi_1 X_{t-1}$$

$X_t - m = \phi_1(X_{t-1} - m) + \eta_t$ On multiplie l'équation précédente par $(X_{t-h} - m)$ alors :

$$\begin{aligned} E[(X_{t-h} - m)(X_t - m)] &= \phi_1 E[(X_{t-h} - m)(X_{t-1} - m)] + E[(X_{t-h} - m)\eta_t] \\ \Leftrightarrow \gamma_h &= \phi_1 \gamma_{h-1} + E[(X_{t-h} - m)\eta_t] \end{aligned}$$

pour $h=0$: $\gamma_0 - \phi_1 \gamma_1 = E[(X_t - m)\eta_t]$

Calculons $E[(X_t - m)\eta_t]$:

On a $(X_t - m)\eta_t - \eta_t \phi_1(X_{t-1} - m) = \eta_t \eta_t$

Donc $E[(X_t - m)\eta_t] - \phi_1 E[\eta_t(X_{t-1} - m)] = E[\eta_t \eta_t]$

$$\Leftrightarrow E[(X_t - m)\eta_t] - \phi_1 \times 0 = \sigma^2$$

Donc $\gamma_0 - \phi_1 \gamma_1 = \sigma^2$

D'où

$$\gamma_h = \begin{cases} \sigma^2 + \phi_1 \gamma_1 & \text{si } h = 0 \\ \phi_1 \gamma_{h-1} & \text{sinon} \end{cases}$$

Donc

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \phi_1 \rho_{h-1} = \phi_1^{|h|} \text{ pour } h \neq 0$$

* **La fonction d'autocorrélation partielle :**

Par définition de la fonction d'autocorrélation partielle on a :

$$\begin{aligned} \lambda_{0,0} &= 1, \quad \lambda_{1,1} = \rho_1 = \phi_1 \\ \lambda_{2,2} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\phi_1^2 - \phi_1^2}{1 - \phi_1^2} = 0 \end{aligned}$$

D'où

$$\lambda_{h,h} = \begin{cases} 1 & \text{si } h = 0 \\ \phi_1 & \text{si } h = 1 \\ 0 & \text{si } h > 1 \end{cases}$$

Fonctions caractéristiques d'un AR(p)

Proposition 16

Soit $X_t \rightsquigarrow AR(p)$ tel que $\Phi(B)X_t = \delta + \eta_t$, où $\phi_1, \phi_2, \dots, \phi_p, \delta$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 , alors :

1)-

$$\mathbb{E}(X_t) = \frac{\delta}{1 - (\phi_1 + \phi_2 + \dots + \phi_p)}.$$

2)- l'auto-covariance est :

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \dots + \phi_p\gamma(h-p)$$

3)- la variance est :

$$\text{Var}(X_t) = \phi_1\gamma(1) + \phi_2\gamma(2) + \dots + \phi_p\gamma(p) + \sigma^2, \forall h \geq 0$$

4)- La fonction d'auto-corrélation :

$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2) + \dots + \phi_p\rho(h-p)$$

5)- La fonction d'auto-corrélation partielle :

$$\tau(h) = \begin{cases} \phi_p & \text{pour } |h| = p \\ 0 & \text{pour } h > p \end{cases}$$

Preuve 4.2.1. 1)- On a :

$$\begin{aligned} X_t &= \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \eta_t \\ \mathbb{E}(X_t) &= \delta + \phi_1 \mathbb{E}(X_{t-1}) + \phi_2 \mathbb{E}(X_{t-2}) + \dots + \phi_p \mathbb{E}(X_{t-p}) + \mathbb{E}(\eta_t) \\ m &= \delta + \phi_1 m + \phi_2 m + \dots + \phi_p m \\ m &= \frac{\delta}{1 - (\phi_1 + \phi_2 + \dots + \phi_p)} \end{aligned}$$

D'où

$$\mathbb{E}(X_t) = \frac{\delta}{1 - (\phi_1 + \phi_2 + \dots + \phi_p)}$$

Remarque : Si les processus $AR(p)$ centrés alors $\mathbb{E}(X_t) = 0$.

2/3)- par définition on a :

$$\begin{aligned} \gamma(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \mathbb{E}(X_t X_{t-h}) - \mathbb{E}(X_t)\mathbb{E}(X_{t-h}) \quad \forall h > 0 \\ &= \mathbb{E}(X_t X_{t-h}) \quad \text{car } \mathbb{E}(X_t) = 0 \end{aligned}$$

Or, on sait que :

$$\forall t \in \mathbb{Z}, X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \eta_t$$

Donc

$$X_t^2 = \phi_1 X_t X_{t-1} + \phi_2 X_t X_{t-2} + \dots + \phi_p X_t X_{t-p} + X_t \eta_t$$

Alors

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \dots + \phi_p\gamma(p) + \mathbb{E}(X_t\eta_t)$$

Or

$$\mathbb{E}(X_t\eta_t) = \underbrace{\mathbb{E}((\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p})\eta_t)}_{=0 \text{ car } \eta_t \perp X_{t-i}, \forall i} + \mathbb{E}(\eta_t^2)$$

D'où

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \dots + \phi_p\gamma(p) + \sigma^2$$

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

$$X_t X_{t-h} = \phi_1 X_{t-1} X_{t-h} + \phi_2 X_{t-2} X_{t-h} + \dots + \phi_p X_{t-p} X_{t-h} + \eta_t X_{t-h}$$

Alors :

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \dots + \phi_p\gamma(h-p) + \underbrace{\mathbb{E}(\eta_t X_{t-h})}_{=0 \text{ car } \eta_t \perp X_{t-h}} \quad (4.5)$$

4)- A partir de la relation de récurrence de $\gamma(h)$ on déduit celle sur $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$. Donc :

$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2) + \dots + \phi_p\rho(h-p) \quad (4.6)$$

Remarque :

$$|\phi| = \begin{cases} < 1 & \text{alors le processus est stationnaire} \\ = 1 & \text{alors marche aléatoire} \\ > 1 & \text{alors le processus est explosif} \end{cases}$$

quand on modélise des séries temporelles stationnaires, $|\phi| < 1$ de telle sorte que X a **une représentation causale**² tout processus $AR(p)$ tel que $|\phi| > 1$ peut être ré-exprimé selon un $AR(p)$ avec $|\phi| < 1$ et une nouvelle suite de bruit blanc.

Dans le cas où $|\phi| = 1$, il n'y a pas de solution stationnaire et par conséquent les $AR(p)$ n'existent pas selon la définition des processus AR . Ces dernières équations sont appelées **équations de Yule-Walker**.

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

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 . En particulier elles donneront une estimation préliminaire de $\hat{\phi}_1, \dots, \hat{\phi}_p$ en fonction de $\hat{\rho}_T(1), \dots, \hat{\rho}_T(p)$.

2. Un processus $(X_t)_{t \in \mathbb{Z}}$ admet une représentation causale s'il peut s'écrire comme combinaison linéaire des valeurs passées d'un autre processus, c'est-à-dire qu'il existe une suite $(\phi_i, i \in \mathbb{Z})$ et un processus $(Y_t)_{t \in \mathbb{Z}}$ tels que $\forall t \in \mathbb{Z}, X_t = \sum_{i \in \mathbb{Z}} \phi_i Y_{t-i}$, et un processus $(X_t)_{t \in \mathbb{Z}}$ admet une représentation inversible s'il peut s'écrire comme combinaison linéaire des valeurs d'un autre processus, c'est-à-dire qu'il existe une suite $(\phi_i, i \in \mathbb{Z})$ et un processus $(Y_t)_{t \in \mathbb{Z}}$ tels que $\forall t \in \mathbb{Z}, X_t = \sum_{i \in \mathbb{Z}} \phi_i Y_{t-i}$.

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

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

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

4.2.2 Prédiction d'un $AR(p)$

Les $AR(p)$ sont bien adaptés à la prédiction. Les prévisions optimales $X_t(h)$ pour $h > 0$ sont des combinaisons linéaires de $X_t, X_{t-1}, \dots, X_{t-p+1}$. Le processus $AR(p)$ peut s'écrire de la façon suivante :

$$X_t - m = \sum_{i=1}^p \phi_i(X_{t-i} - m) + \eta_t$$

On veut prévoir les valeurs futures X_{n+h} pour $h > 0$

$$X_n(h) = E[X_{n+h}/X_n, X_{n-1}, \dots] \quad h > 0$$

$$= m + E\left[\sum_{i=1}^p \phi_i(X_{n+h-i} - m) + \eta_{n+h}/X_n, X_{n-1}, \dots\right]$$

$$= m + \sum_{i=1}^p \phi_i E[(X_{n+h-i} - m)/X_n, X_{n-1}, \dots] + E[\eta_{n+h}/X_n, X_{n-1}, \dots]$$

$$= m + \sum_{i=1}^p \phi_i E[(X_{n+h-i}/X_n, X_{n-1}, \dots) - m] + E[\eta_{n+h}/X_n, X_{n-1}, \dots]$$

$$= m + \sum_{i=1}^p \phi_i(X_n(h-i) - m)$$

car

$$E[X_{n+i}/X_n, X_{n-1}, \dots] = \begin{cases} X_{n+i} & i \leq 0 \\ X_n(i) & i > 0 \end{cases} \quad E[\eta_{n+i}/X_n, X_{n-1}, \dots] = \begin{cases} \eta_{n+i} & i \leq 0 \\ 0 & i > 0 \end{cases}$$

Ainsi on a le théorème suivant :

Théorème 11

Soit X un $AR(p)$ de bruit blanc d'innovation η et de polynôme canonique $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$. Les prévisions optimales $X_t(h)$ pour $h > 0$ sont des combinaisons linéaires de $X_t, X_{t-1}, \dots, X_t, X_{t-p+1}$:

$$X_t(h) = m + \sum_{1 \leq i \leq p} \phi_i(X_t(h-i) - m)$$

avec les conditions initiales $X_t(-i) = X_{t-i}$ pour $i \geq 0$. [7]

4.2.3 Intervalle de confiance pour la prévision

L'erreur de prévision à l'horizon $h > 0$ d'une série chronologique $(X_t)_{t \in 1, \dots, n}$:

$$e_n(h) = X_{n+h} - X_n(h) = \sum_{i=0}^{h-1} \psi_i \eta_{n+h-i}$$
$$Var[e_n(h)] = \sigma^2 \sum_{i=0}^{h-1} \psi_i^2$$

avec ψ_i sont les pondérations du théorème de Wold qui égale les autocorrélations du processus à étudié et $e_n(h) \rightsquigarrow \mathcal{N}(0, \sigma^2 \sum_{i=0}^{h-1} \psi_i^2)$

On en déduit que :

$$\frac{e_n(h)}{\sigma \sqrt{\sum_{i=0}^{h-1} \psi_i^2}} \rightsquigarrow \mathcal{N}(0, 1)$$

On en déduit un intervalle de confiance ou de prévision au niveau $1 - \alpha$ pour X_{n+h} :

$$[X_n(h) - z_{\frac{\alpha}{2}} \sqrt{Var(e_n(h))}; X_n(h) + z_{\frac{\alpha}{2}} \sqrt{Var(e_n(h))}]$$

4.2.4 Exemple sous le logiciel R

A titre d'exemple, nous allons étudier le processus $AR(2)$ suivant :

$$X_t - 0,9X_{t-1} + 0,8X_{t-2} = \eta_t$$

On peut utiliser tout simplement la fonction `arima.sim()`.

Pour représenter la fonction d'autocorrélation empirique d'une série observée, il faut utiliser en \mathcal{R} la fonction `acf()` où la série chronologique est placée en argument. Par défaut, le logiciel fournit la fonction d'autocorrélation pour une longueur proportionnelle au logarithme de la série et représente graphiquement la fonction. On peut changer ces paramètres en utilisant la syntaxe suivante : `acf(serie, lag.max = n, plot = F)`.

En précisant à la fonction `acf()` un argument `type`, on peut obtenir les autocorrélations partielles : `acf(serie, type = "partial")`.

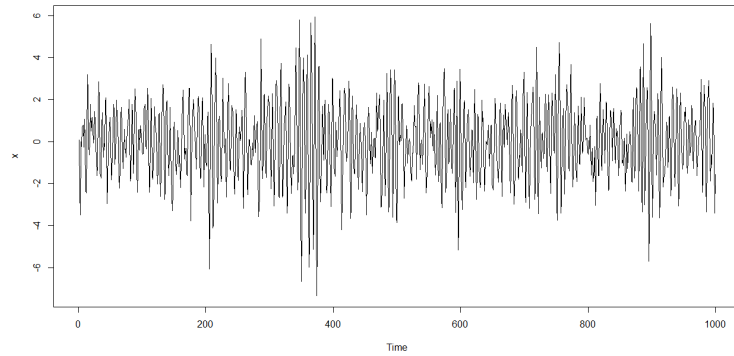


FIGURE 4.1 – Graphe de trajectoire.

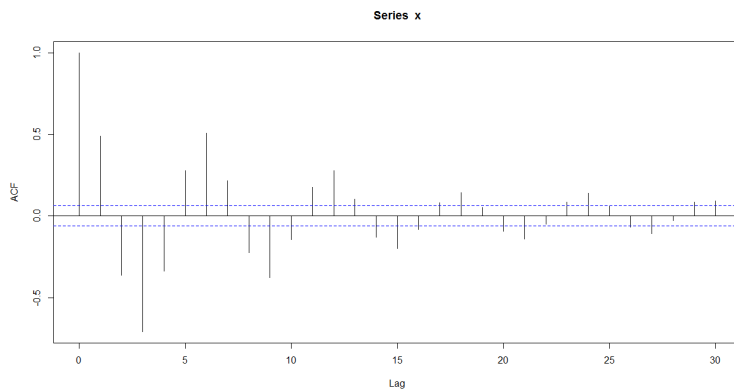


FIGURE 4.2 – Corrélogramme du processus auto-régressif d'ordre 2.

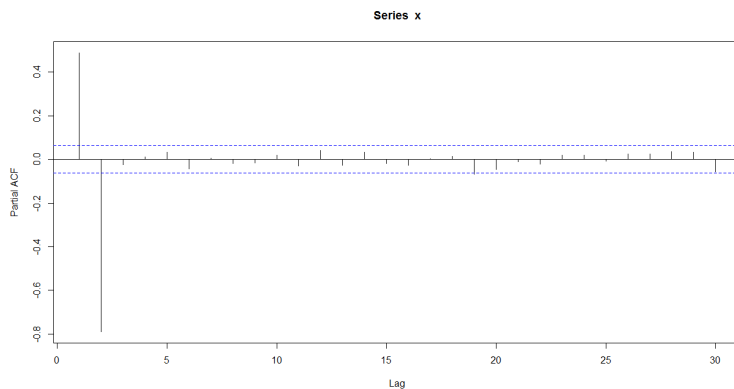


FIGURE 4.3 – Corrélogramme partiel du processus auto-régressif d'ordre 2.

Le polynôme Φ est du second degré et son discriminant est négatif et vaut -2.39 . Les racines sont donc complexes. Ceci explique la forme "sinusoïdale" du corrélogramme[si les racines sont réelles, alors le corrélogramme est de type exponentielle]. D'autre part, les corrélations d'ordre 1 et 4 apparaissent assez élevées en module. Cette propriété devrait réapparaître sur les trajectoires : une forte valeur à une certaine date devrait en général impliquer une faible valeur trois dates après et une forte valeur six dates après (le caractère faible ou fort dépend bien évidemment du signe de ρ). La présence du bruit de variance σ^2 explique évidemment que la fonction ne soit pas strictement périodique. * Prenons un autre exemple $AR(1)$, nous allons

étudier deux cas selon le signe de ϕ_1 .

- Dans le premier cas, on va étudier le processus $AR(1)$ avec $\phi_1 = 0,8$

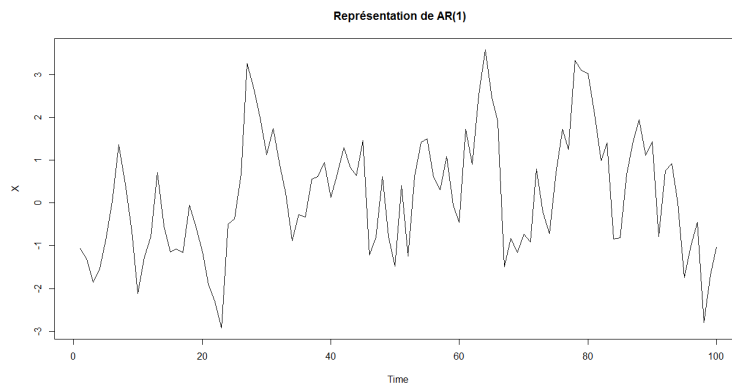


FIGURE 4.4 – Graphe de trajectoire 1 de $AR(1)$

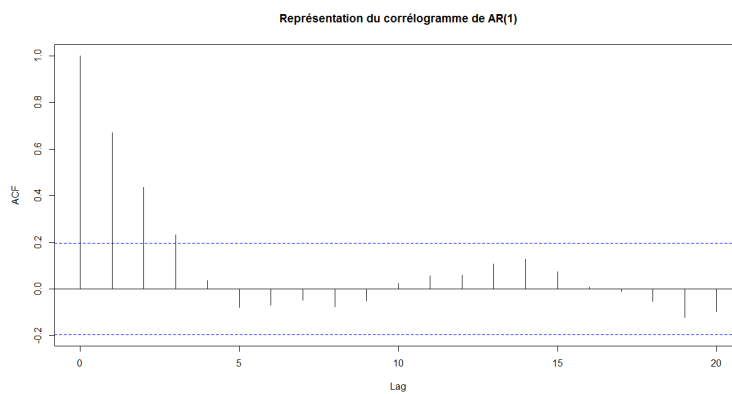


FIGURE 4.5 – Représentation du corrélogramme 1 de $AR(1)$

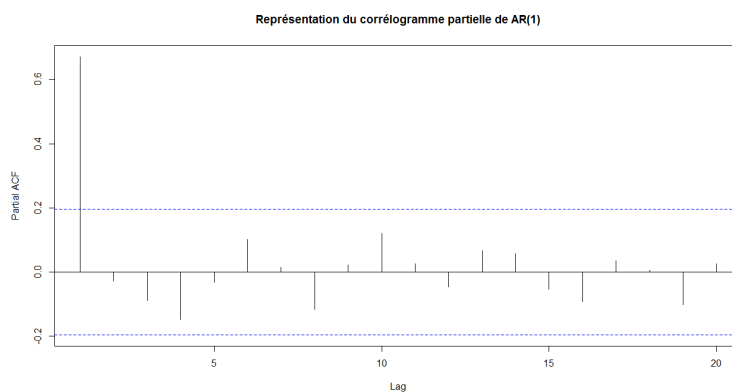


FIGURE 4.6 – Représentation du corrélogramme partielle 1 de $AR(1)$

- Pour le deuxième cas, on va étudier le processus $AR(1)$ avec $\phi_1 = -0,8$

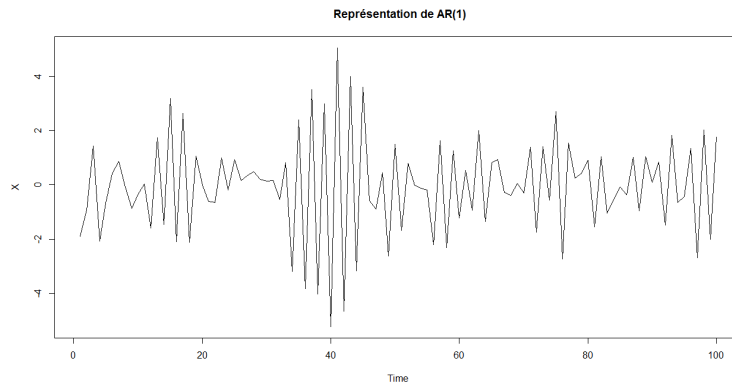


FIGURE 4.7 – Graphe de trajectoire 2 de $AR(1)$

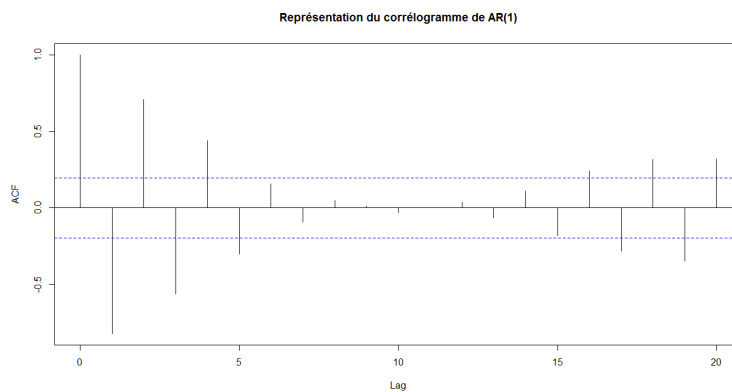


FIGURE 4.8 – Représentation du corrélogramme 2 de $AR(1)$

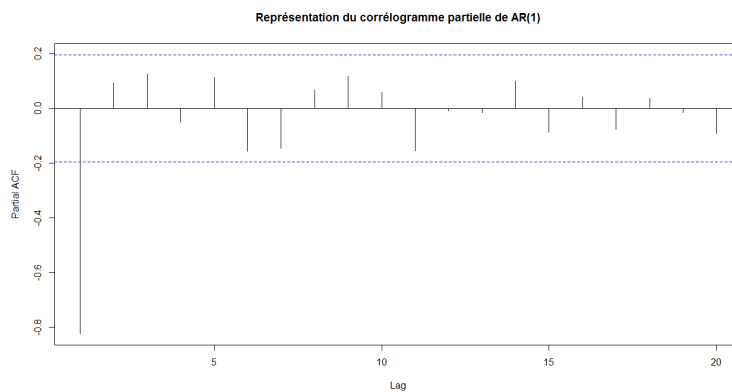


FIGURE 4.9 – Représentation du corrélogramme partielle 2 de $AR(1)$

Cette fonction est donc marquée par une décroissance exponentielle de termes soit presque tous positifs si $\phi_1 > 0$, soit alternant en signe si $\phi_1 < 0$. On notera enfin que la fonction d'autocovariance et la fonction d'autocorrélation obéissent, au bruit blanc près, à la même équation que le processus. En pratique, cela signifie par exemple que si ϕ_1 est positif, alors les autocovariances et autocorrélations seront aussi positives. En revanche, si ϕ_1 est négatif, alors autocovariances et autocorrélations alternent en signe.

4.3 Processus moyennes mobiles d'ordre q

Ces processus forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations d'un bruit blanc, ils sont donc définis par la relation

$$\forall t \in \mathbb{Z} \quad X_t = \delta + \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q}$$

où $\theta_1, \dots, \theta_q, \delta$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 .

L'observation au temps t est donc la somme d'un choc aléatoire η_t à l'instant t et d'une fonction linéaire du passé de ce choc $-\sum_{i=1}^q \theta_i \eta_{t-i}$.

4.3.1 Définition et représentation canonique

Définition 28

On appelle processus moyenne mobile [Moving Average] d'ordre q tout processus $(X_t)_{t \in \mathbb{Z}}$ stationnaire tel que

$$\forall t \in \mathbb{Z} \quad X_t = \delta + \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q}$$

où $\theta_1, \dots, \theta_q, \delta$ sont des réels fixés et $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 .

Un tel processus est dit $MA(q)$ (Moving Average of order q).

Posons $\Theta(B) = I - \theta_1 B - \dots - \theta_q B^q$, on peut écrire :

$$X_t = \delta + \Theta(B)\eta_t, \quad \forall t \in \mathbb{Z}$$

[7]

Remarque 4.3.1. Notons immédiatement que

- un processus à moyenne mobile est défini de manière explicite.
- un processus à moyenne mobile est automatiquement stationnaire.
- un processus à moyenne mobile est par définition purement innovant et causal.

Mais attention son bruit blanc innovant n'est pas nécessairement égal à η . On peut étendre cette définition aux $MA(\infty)$ en faisant croître q. On pourra alors vérifier que X est stationnaire si et seulement si $\sum_{i \in \mathbb{Z}} \theta_i^2 < \infty$.

Forme auto-régressive infinie d'un MA(q)

Dans le cas régulier où le polynôme Θ ne s'annule pas sur le cercle unité, on a le théorème

Théorème 12

(i) Si Θ n'a pas de racine de module égal à 1, alors il existe un choix unique de coefficients π_i pour lesquels on a

$$\eta_t = \sum_{i \in \mathbb{Z}} \pi_i X_{t-i}$$

avec $\sum_{i=-\infty}^{+\infty} |\pi_i| < \infty$. Dans ce cas, on dit que le processus est de représentation inversible.

(ii) Si de plus les racines de Θ sont toutes de module strictement supérieur à 1, alors il existe un choix unique de coefficients π_i pour lesquels on a

$$\eta_t = \sum_{i \in \mathbb{Z}} \pi_i X_{t-i}$$

Les coefficients π_i convergent rapidement vers 0 lorsque $i \rightarrow \infty$, en effet ce sont les coefficients du filtre associé au polynôme $\frac{1}{\Theta}$. Dans ce cas, on dit que le processus est causal et η_t est l'innovation du processus à la date t .

Dans le cas où Θ n'a pas de racine multiple, on a la formule explicite suivante

$$\forall j \geq 0, \quad \pi_j = \sum_{1 \leq k \leq q} m_k \mu_k^j$$

avec

$$m_k = \frac{1}{\prod_{l \neq k} (1 - \frac{\mu_l}{\mu_k})} \quad \text{et} \quad \Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q = \prod_i (1 - \mu_i z)$$

[7]

Par contre, si Θ s'annule sur le cercle unité, on a le théorème suivant :

Théorème 13

Si Θ s'annule en un point du cercle unité, alors il n'existe aucun choix de coefficients π_i pour lesquels $\sum_{i \geq 0} \pi_i Y_{t-i}$ converge et coïncide avec η_t . Cependant, on peut représenter η_t comme limite de telles combinaisons linéaires des X_u , $u \leq t$. [7]

Bruit blanc d'innovation d'un MA(q)

Comme nous l'avons dit précédemment, le bruit blanc d'innovation $(\epsilon_t)_{t \in \mathbb{Z}}$ du processus MA(q) $(X_t)_{t \in \mathbb{Z}}$ n'est pas nécessairement le processus $(\eta_t)_{t \in \mathbb{Z}}$ c'est-à-dire que η_t ne représente pas l'information ajoutée au passé pour obtenir la valeur présente de X_t . Mais nous avons le résultat suivant

Théorème 14

La relation d'un MA(q) avec son bruit blanc d'innovation est aussi du type $X_t = \tilde{\Theta}(B)\epsilon_t$, $t \in \mathbb{Z}$. Le polynôme $\tilde{\Theta}$ de degré q, dit **polynôme canonique**, s'obtient à partir de Θ en remplaçant toutes les racines éventuelles à l'intérieur du cercle unité par leur inversion ($\mu \rightarrow \frac{1}{\mu}$), les déplaçant ainsi à l'extérieur du cercle unité. Notons μ_1, \dots, μ_q les co-racines de $\Theta(z)$:

$$\Theta(z) = \prod_i (1 - \mu_i z) = 1 - \theta_1 z - \dots - \theta_q z^q.$$

Le polynôme $\tilde{\Theta}$ qui relie X à son bruit blanc d'innovation est donné par

$$\tilde{\Theta}(z) = \prod_{|\mu_i| \leq 1} (1 - \mu_i z) = \prod_{|\mu_i| > 1} (1 - \frac{1}{\mu_i} z).$$

[7]

On déduit du théorème précédent que

Proposition 17

La variance du bruit blanc d'innovation ϵ est donnée par

$$Var(\epsilon_t) = \gamma_\epsilon(0) = \gamma_\eta(0) = \prod_{i, |\mu_i| > 1} |\mu_i|^2.$$

A partir de maintenant, nous ne considèrerons plus que des processus MA(q) donnés sous leur représentation canonique i.e. vérifiant l'équation

$$X_t = \Theta(B)\eta_t$$

avec $\Theta(z) = \prod_{i=1}^q (1 - \mu_i z) = 1 - \theta_1 z - \dots - \theta_q z^q$ où $|\mu_i| \leq 1, \forall i$ **puisque nous venons de voir qu'il est toujours possible de se ramener à ce cas quitte changer de bruit blanc.**

Fonctions caractéristiques et propriétés d'un MA(q)

Exemple 4.3.2. Soit X un MA(1). (X_t) est donc généré par un bruit blanc (η_t) sous la forme

$$X_t = \delta + (I - \theta B)\eta_t = \delta + \eta_t - \theta\eta_{t-1}.$$

La fonction de transfert du filtre se réduit à un seul terme.

1- Par simples calculs, on peut montrer que

$$Var(X_t) = (1 + \theta^2)\sigma^2, \forall t.$$

Plus généralement,

$$\gamma(h) = \begin{cases} -\theta\sigma^2, & si |h| = 1 \\ 0 & si |h| \geq 2 \end{cases}$$

2- De même, on peut montrer que

$$\rho(h) = \begin{cases} \frac{-\theta}{1+\theta^2}, & \text{si } |h| = 1 \\ 0 & \text{si } |h| \geq 2 \end{cases}$$

3- Enfin à partir de l'expression générale de la fonction d'autocorrélation partielle, on peut montrer que

$$\tau(h) = -\theta^h \frac{1-\theta^2}{1-\theta^{2h+2}}, \quad h \in \mathbb{N}^*.$$

On voit que cette fonction décroît exponentiellement vers 0 quand h augmente.

Proposition 18

1- La variance de X_t est donnée par

$$\text{Var}(X_t) = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2, \quad \forall t.$$

Ayant de plus $\delta = 0$, on en déduit que tout processus moyenne mobile est un processus stationnaire.

2- Plus généralement, la fonction d'autocovariance est donnée par

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

3- La fonction d'autocorrélation est donnée par

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

4- L'expression de la fonction d'autocorrélation partielle est compliquée, mais il suffit d'utiliser l'expression de la fonction d'autocorrélation partielle étudiée au chapitre 3.

4.3.2 Prédiction d'un MA(q)/Intervalle de confiance pour la prévision

Soit X un MA(q) de polynôme Θ et de bruit blanc d'innovation η (on suppose que η est causal). La relation qui le lie à son bruit blanc d'innovation η :

$$X_t = \delta + \eta_t - \theta_1\eta_{t-1} - \dots - \theta_q\eta_{t-q}$$

Comme nous l'avons fait dans la section précédente, c'est-à-dire, calculons $X_n(h)$ à l'horizon $h > 0$ on utilisant l'espérance conditionnel, et la même expression de la variance, il faut juste remplacer les pondérations par les autocorrélation du MA(q).

L'intervalle de confiance pour la prévision du processus MA(q) est égale l'intervalle de confiance du processus AR(p), prenons on considérations l'expression de la variance et des pondérations.

4.3.3 Exemple sous le logiciel R

A titre d'exemple, nous allons étudier le processus MA(3) suivant : $X_t = \eta_t - 0,8\eta_{t-1} - 0,4\eta_{t-2} - 0,9\eta_{t-3}$

On peut utiliser tout simplement la fonction `arima.sim()`.

Pour représenter la fonction d'autocorrélation empirique d'une série observée, il faut utiliser

en \mathcal{R} la fonction $acf()$ où la série chronologique est placée en argument. Par défaut, le logiciel fournit la fonction d'autocorrélation pour une longueur proportionnelle au logarithme de la série et représente graphiquement la fonction. On peut changer ces paramètres en utilisant la syntaxe suivante : $acf(serie, lag.max = n, plot = F)$.

En précisant à la fonction $acf()$ un argument `type`, on peut obtenir les autocorrélations partielles : $acf(serie, type = "partial")$.

* Prenons un autre exemple $MA(1)$, nous allons étudier deux cas selon le signe de θ_1

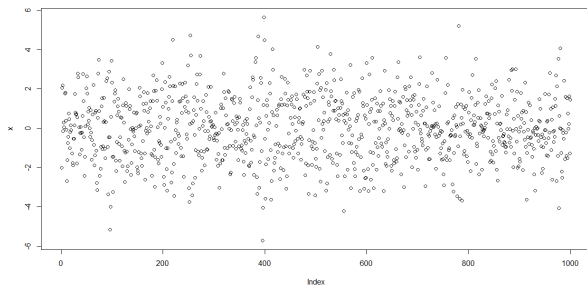


FIGURE 4.10 – Graphe original de la fonction moyenne mobile.

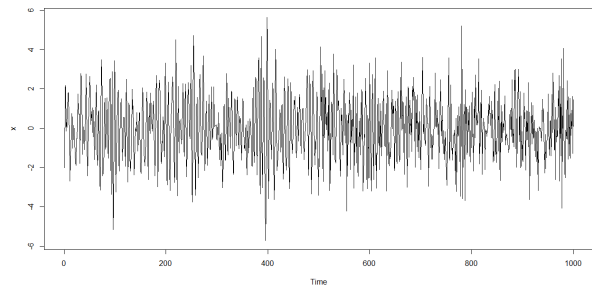


FIGURE 4.11 – Graphe de trajectoire.

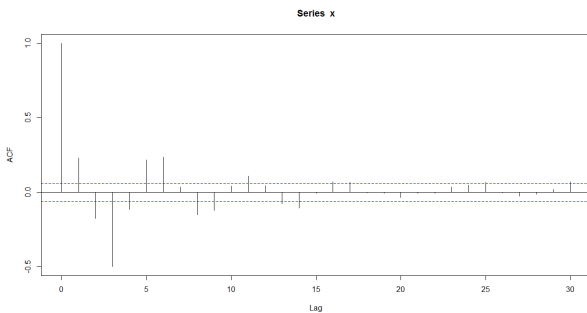


FIGURE 4.12 – Corrélogramme du processus moyenne mobile d'ordre 3.

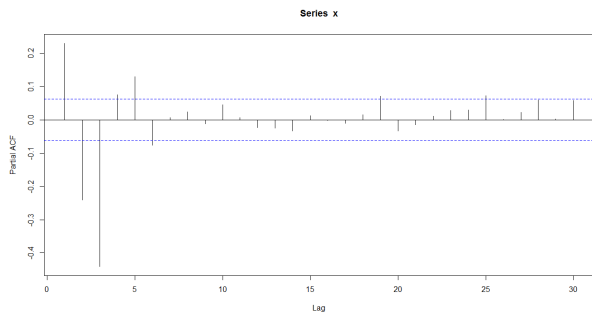


FIGURE 4.13 – Corrélogramme partiel du processus moyenne mobile d'ordre 3.

- Dans le premier cas, on va étudier le processus $MA(1)$ avec $\phi_1 = 0,8$

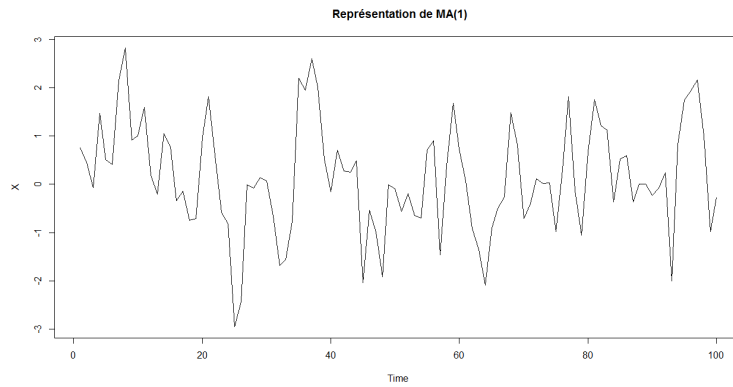


FIGURE 4.14 – Graphe de trajectoire 1 de $MA(1)$

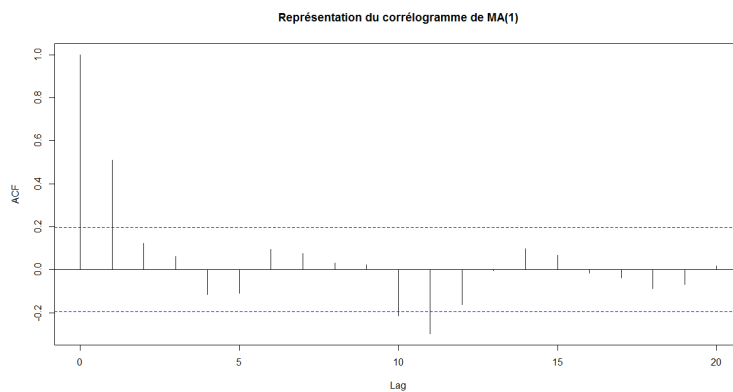


FIGURE 4.15 – Représentation du corrélogramme 1 de $MA(1)$

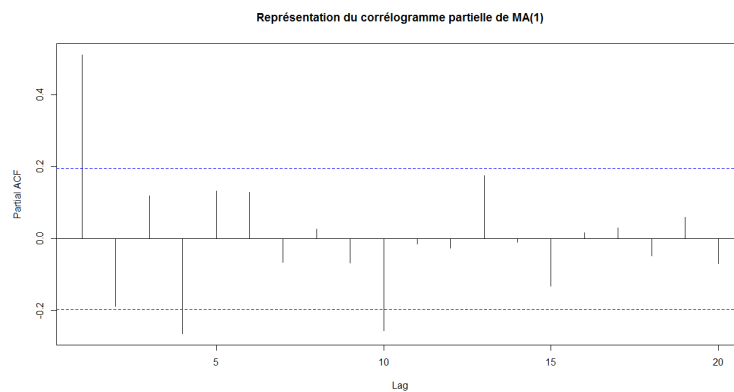


FIGURE 4.16 – Représentation du corrélogramme partielle 1 de $MA(1)$

- Pour le deuxième cas, on va étudier le processus $MA(1)$ avec $\theta_1 = -0,8$

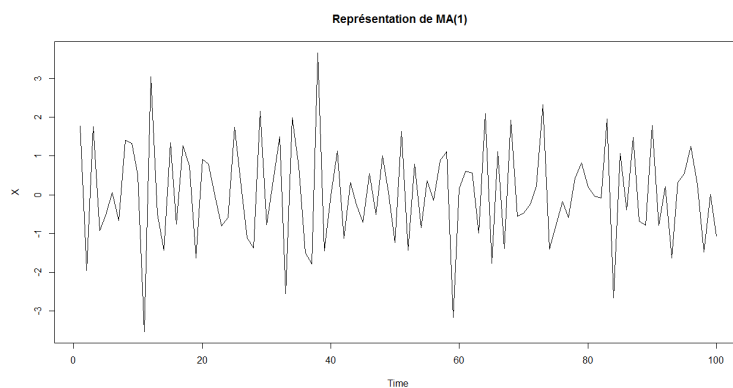


FIGURE 4.17 – Graphe de trajectoire 2 de $MA(1)$

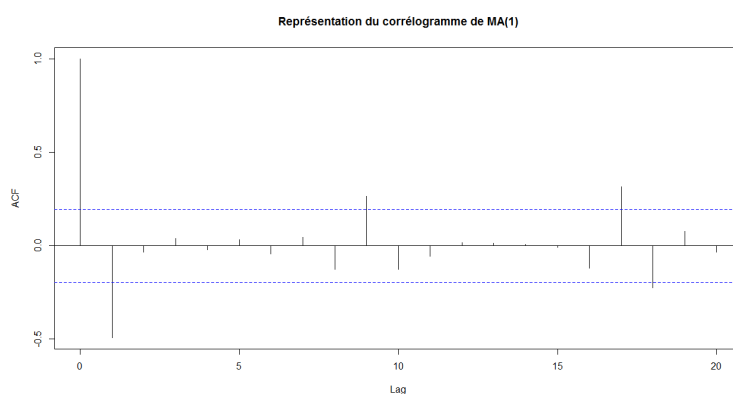


FIGURE 4.18 – Représentation du corrélogramme 2 de $MA(1)$

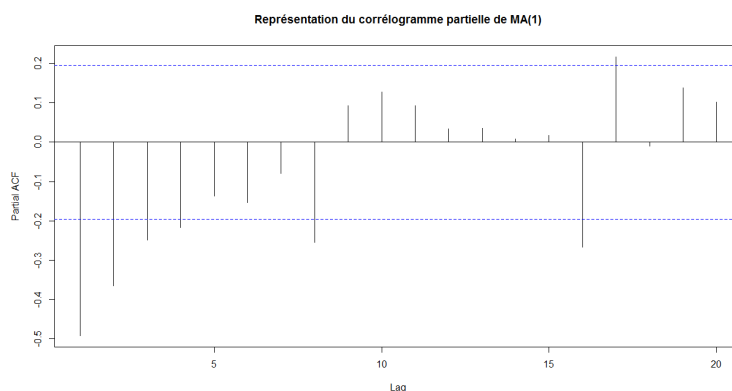


FIGURE 4.19 – Représentation du corrélogramme partielle 2 de $MA(1)$

4.4 Les processus mixtes ARMA

Les processus AR et MA ont des caractéristiques qui se révèlent grâce à leurs fonctions d'autocorrélations et leurs fonctions d'autocorrélations partielles. Pour un processus AR, nous avons vu que la fonction d'autocorrélation partielle possède un point de rupture après un certain nombre d'écarts, ce dernier détermine l'ordre du polynôme AR. Pour un processus MA, nous avons vu que c'est la fonction d'autocorrélation qui possède un point de rupture après un

certain nombre d'écarts, ce dernier détermine l'ordre du polynôme MA.

Cependant pour certains processus, ni la fonction d'autocorrélation, ni la fonction d'autocorrélation partielle ne possèdent de point de rupture. Dans de tels cas, il faut construire un modèle mixte.

Nous définissons dans cette section les séries ARMA qui sont des généralisations directes des deux processus précédents, la combinaison des processus autorégressifs et moyennes mobiles. Cette classe de processus ARMA est encore un cas particulier de processus linéaires et jouera un rôle prépondérant dans la modélisation concrète des processus stationnaires. Elle présente l'avantage d'être plus souple à l'utilisation et de fournir généralement de bonnes approximations des séries réelles avec moins de paramètres que les modèles purs, nous aborderons finalement le problème de la prédiction sur ce modèle.

4.4.1 Définition d'un ARMA(p,q)

Définition 29

On appelle processus autorégressif moyenne mobile d'ordre (p,q) tout processus $(X_t)_t \in \mathbb{Z}$ stationnaire tel que

$$\forall t \in \mathbb{Z}, X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \delta + \eta_t - \theta_1 \eta_{t-1} - \dots - \theta_q \eta_{t-q} \quad (4.7)$$

où $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \delta$ sont des réels fixés et $(\eta_t)_t \in \mathbb{Z}$ est un bruit blanc de variance σ^2 . Un tel processus est dit ARMA(p, q) (AutoRegressive Moving Average of order (p,q)). [7]

Posons comme précédemment $\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p$ et $\Theta(B) = I - \theta_1 B - \dots - \theta_q B^q$. On peut alors écrire

$$\forall t \in \mathbb{Z}, \Phi(B)X_t = \delta + \Theta(B)\eta_t.$$

Remarque 4.4.1. *Il est immédiat qu'un ARMA(p,0) est un AR pur et qu'un ARMA(0,q) est un MA pur. Les seuls processus admettant simultanément une représentation MA pure et une représentation AR pure correspondent au cas ARMA(0,0), c'est-à-dire aux bruits blancs.*

Définition 30

La représentation d'un processus ARMA(p, q)

$$\forall t \in \mathbb{Z}, \Phi(B)X_t = \delta + \Theta(B)\eta_t.$$

est dite :

- **minimale** si les polynômes $\Phi(z)$ et $\Theta(z)$ n'ont pas de racine commune.
- **causale** si le polynôme $\Phi(z)$ a toutes ses racines à l'extérieur du disque unité.
- **inversible** si le polynôme $\Theta(z)$ a toutes ses racines à l'extérieur du disque unité.
- **canonique** si elle est causale et inversible.

[7]

4.4.2 Écritures $MA(\infty)$ et $AR(\infty)$

Proposition 19

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus ARMA(p,q) de représentation minimale et causale. Il admet alors la représentation $MA(\infty)$

$$\forall t \in \mathbb{Z}, X_t = \Phi(B)^{-1} \Theta(B) \eta_t = \delta + \eta_t + \sum_{i=1}^{\infty} \psi_i \eta_{t-i}.$$

où les coefficients $(\psi_i)_{i \in \mathbb{N}}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire, pour $\delta = 0$:

$$\psi_i - \sum_{j=1}^p \phi_j \psi_{i-j} = \theta_i \quad \forall i \in \mathbb{N}$$

avec $\psi_i = 0$, pour $i < 0$, $\psi_0 = 1$ et $\theta_0 = 1$ et $\theta_i = 0$ pour $i > q$. [7]

Proposition 20

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus ARMA(p,q) de représentation minimale, inversible et centré. Il admet alors la représentation $AR(\infty)$

$$\forall t \in \mathbb{Z}, \eta_t = \Theta(B)^{-1} \Phi(B) X_t = X_t + \sum_{i=1}^{\infty} \pi_i X_{t-i}.$$

où les coefficients $(\pi_i)_{i \in \mathbb{N}}$ forment une famille absolument sommable et vérifient l'équation de récurrence linéaire :

$$\pi_i + \sum_{j=1}^q \theta_j \pi_{i-j} = -\phi_i \quad \forall i \in \mathbb{N}$$

avec $\pi_i = 0$, pour $i < 0$, $\phi_0 = -1$ et $\phi_i = 0$ pour $i > p$. [7]

Exemple 4.4.2. Soit $(X_t)_{t \in \mathbb{Z}}$ un ARMA(1,1) non centré, $(X_t)_{t \in \mathbb{Z}}$ est donc généré par un bruit blanc $(\eta_t) \sim N(0, \sigma^2)$, sous la forme :

$$\Phi_1(B) X_t = \delta + \Theta_1(B) \eta_t$$

$$(1 - \phi_1 B) X_t = \delta + (1 - \theta_1 B) \eta_t \quad (4.8)$$

$$X_t = \delta + \phi_1 X_{t-1} + \eta_t - \theta_1 \eta_{t-1} \quad (4.9)$$

★ Condition de stationnarité : $|\phi_1| < 1$

★ Condition d'inversibilité : $|\theta_1| < 1$

★ Degeneracy condition : $\phi_1 \neq \theta_1$ car sinon ARMA(1,1) devient ARMA(0,0)

* **Espérance de ARMA(1,1) :**

D'après l'équation(4.8) on trouve :

$$X_t = \frac{\delta}{1 - \phi_1} + \frac{1 - \theta_1 B}{1 - \phi_1 B} \eta_t$$

Donc

$$E[X_t] = \frac{\delta}{1 - \phi_1} + \frac{1 - \theta_1 B}{1 - \phi_1 B} E[\eta_t] = \frac{\delta}{1 - \phi_1}$$

* **La fonction d'autocovariance de ARMA(1, 1) :**

Comme $E[X_t] = m = \frac{\delta}{1 - \phi_1}$ alors $\delta = m(1 - \phi_1)$

Donc l'équation (4.9) devient :

$$X_t = m(1 - \phi_1) + \phi_1 X_{t-1} + \eta_t - \theta_1 \eta_{t-1} \Leftrightarrow X_t - m = \phi_1 (X_{t-1} - m) + \eta_t - \theta_1 \eta_{t-1}$$

multiplions $(X_t - m)$ par $(X_{t-h} - m)$ on obtient :

$$(X_t - m)(X_{t-h} - m) = \phi_1 (X_{t-1} - m)(X_{t-h} - m) + \eta_t (X_{t-h} - m) - \theta_1 \eta_{t-1} (X_{t-h} - m)$$

Donc

$$E[(X_t - m)(X_{t-h} - m)] = \phi_1 E[(X_{t-1} - m)(X_{t-h} - m)] + E[\eta_t (X_{t-h} - m)] - \theta_1 E[\eta_{t-1} (X_{t-h} - m)]$$

équivalent à :

$$\gamma_h = \phi_1 \gamma_{h-1} + E[\eta_t (X_{t-h} - m)] - \theta_1 E[\eta_{t-1} (X_{t-h} - m)]$$

pour $h = 0$ on a : $\gamma_0 = \phi_1 \gamma_1 + E[\eta_t (X_t - m)] - \theta_1 E[\eta_{t-1} (X_t - m)]$

Cherchons $E[\eta_t (X_t - m)]$ et $E[\eta_{t-1} (X_t - m)]$

On a $X_t - m = \phi_1 (X_{t-1} - m) + \eta_t - \theta_1 \eta_{t-1}$

alors

$$\begin{aligned} E[(X_t - m)\eta_t] &= \phi_1 E[(X_{t-1} - m)\eta_t] + E[\eta_t \eta_t] - \theta_1 E[\eta_{t-1} \eta_t] \\ &= \phi_1 \times 0 + \sigma^2 - \theta_1 \times 0 \\ &= \sigma^2 \end{aligned}$$

De même façon on trouve $E[(X_t - m)\eta_{t-1}]$

$$\begin{aligned} E[(X_t - m)\eta_{t-1}] &= \phi_1 E[(X_{t-1} - m)\eta_{t-1}] + E[\eta_t \eta_{t-1}] - \theta_1 E[\eta_{t-1} \eta_{t-1}] \\ &= \phi_1 \sigma^2 + 0 - \theta_1 \sigma^2 \\ &= \sigma^2 (\phi_1 - \theta_1) \end{aligned}$$

Alors $\gamma_0 = \phi_1 \gamma_1 + \sigma^2 (1 - \theta_1 (\phi_1 - \theta_1))$

pour $h = 1$ on a :

$$\begin{aligned} \gamma_1 &= \phi_1 \gamma_0 + E[\eta_t (X_{t-1} - m)] - \theta_1 E[\eta_{t-1} (X_{t-1} - m)] \\ &= 0 - \theta_1 \sigma^2 \\ &= -\theta_1 \sigma^2 \end{aligned}$$

pour $h = 2$ on a :

$$\begin{aligned} \gamma_2 &= \phi_1 \gamma_1 + E[\eta_t (X_{t-2} - m)] - \theta_1 E[\eta_{t-1} (X_{t-2} - m)] \\ &= 0 \end{aligned}$$

D'où :

$$\begin{aligned} \gamma_0 &= \frac{1 + \theta^2 - 2\phi_1 \theta_1}{1 - \phi_1^2} \sigma^2 \\ \gamma_1 &= \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2} \sigma^2 \end{aligned}$$

* **La fonction d'autocorrélation de ARMA(1, 1) :**

D'après ce qui précède on peut déduire :

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \begin{cases} 1 & \text{si } h = 0 \\ \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta^2 - 2\phi_1\theta_1} & \text{si } h = 1 \\ \phi_1\rho_{h-1} & \text{si } h \geq 2 \end{cases}$$

* **La fonction d'autocorrélation partielle de ARMA(1, 1) :** Par définition de la fonction d'autocorrélation partielle, on a :

$$\lambda_{h,h} = \begin{cases} 1 & \text{si } h = 0 \\ \rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta^2 - 2\phi_1\theta_1} & \text{si } h = 1 \\ \frac{\rho_2 - \lambda_{1,1}\rho_1}{1 - \lambda_{1,1}\rho_1} & \text{si } h = 2 \\ \frac{\rho_3 - \lambda_{2,1}\rho_2 - \lambda_{2,2}\rho_1}{1 - \lambda_{2,1}\rho_1 - \lambda_{2,2}\rho_2} & \text{si } h = 3 \\ \vdots & \end{cases} \quad \text{où } \lambda_{2,1} = \lambda_{1,1} - \lambda_{2,2}\lambda_{1,1}$$

Proposition 21

La fonction d'autocovariance d'un processus $ARMA(p, q)$ vérifie :

pour $h \geq \max(p, q + 1)$: $\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h - i)$

pour $h \leq \max(p, q + 1)$: $\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h - i) + \sigma^2(\psi_0 - \psi_1\theta_{h+1} - \dots - \psi_{q-h}\theta_q)$

[13]

Proposition 22

La somme d'un $ARMA(p_1, q_1)$ et d'un $ARMA(p_2, q_2)$ (non corrélés) est un $ARMA(\leq \max(p_1, p_2), \leq \max(q_1, q_2))$. [13]

4.4.3 Exemple sous le logiciel R

A titre d'exemple, nous allons étudier le processus $ARMA(1, 1)$ suivant :

$$X_t = \phi_1 X_{t-1} + \eta_t - \theta_1 \eta_{t-1} \quad \text{avec } \eta_t \sim bb(0, 1)$$

On va traiter les trois cas selon les signes de ϕ_1 et θ_1

Cas 1 : $X_t = 0,9X_{t-1} + \eta_t + 0,5\eta_{t-1}$

Cas 2 : $X_t = 0,9X_{t-1} + \eta_t - 0,5\eta_{t-1}$

Cas 3 : $X_t = -0,9X_{t-1} + \eta_t - 0,5\eta_{t-1}$

Commençons par la représentation du processus $ARMA(1, 1)$ dans les trois cas.

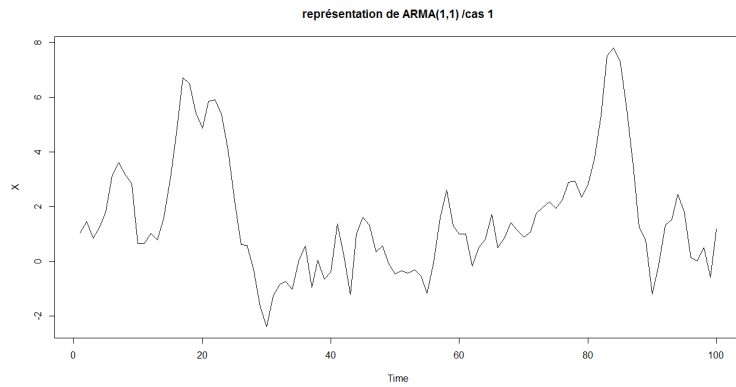


FIGURE 4.20 – Graphe de trajectoire 1 de $ARMA(1, 1)$

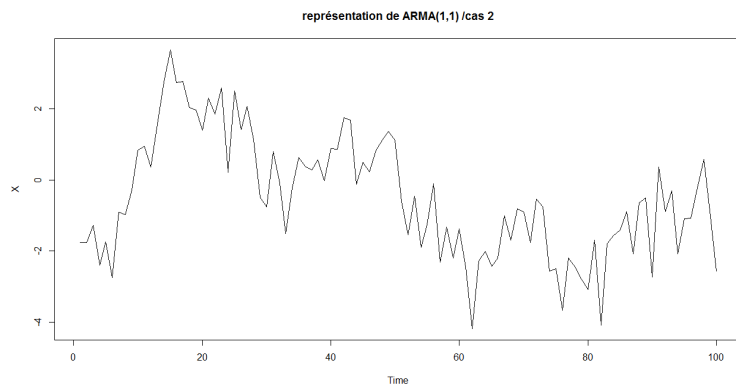


FIGURE 4.21 – Graphe de trajectoire 2 de $ARMA(1, 1)$

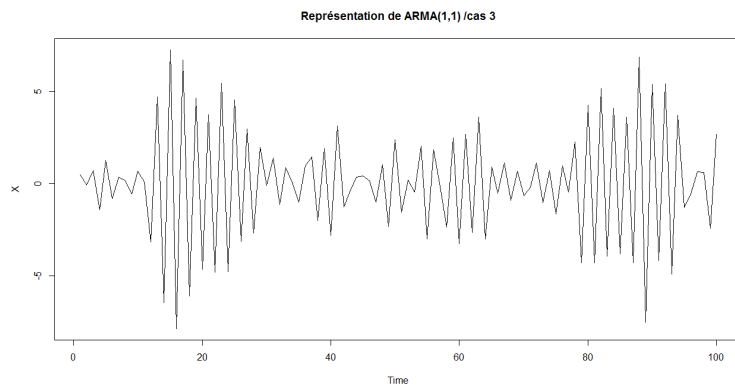


FIGURE 4.22 – Graphe de trajectoire 3 de $ARMA(1, 1)$

Passons maintenant à la représentation du variogramme du processus pour les trois cas.

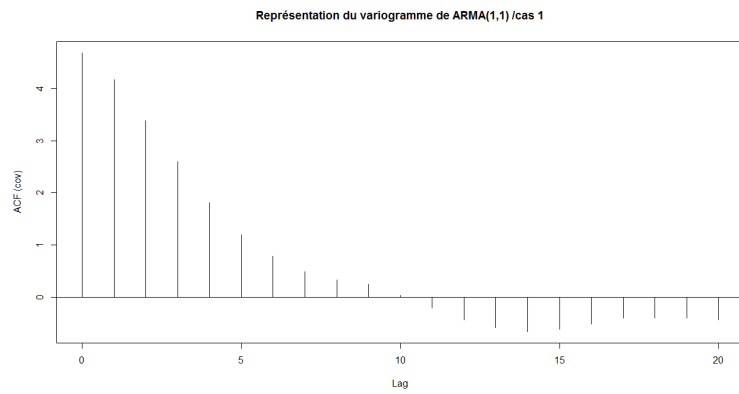


FIGURE 4.23 – Graphe du variogramme 1 de $ARMA(1,1)$

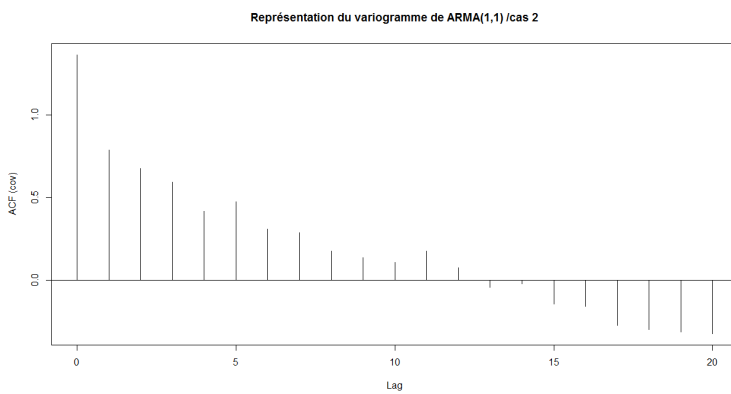


FIGURE 4.24 – Graphe du variogramme 2 de $ARMA(1,1)$

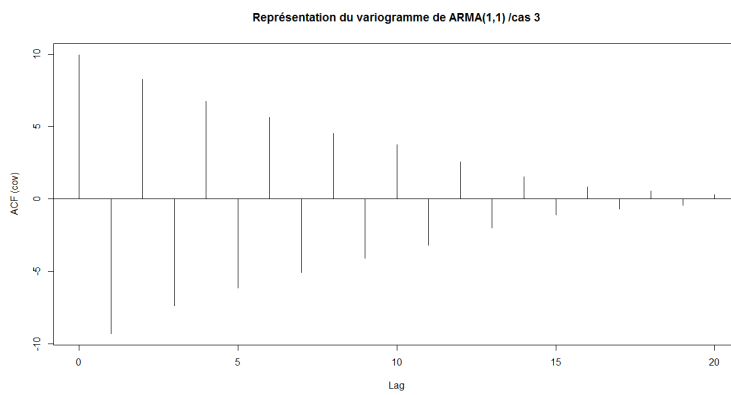


FIGURE 4.25 – Graphe du variogramme 3 de $ARMA(1,1)$

Maintenant on va représenter le corrélogramme du processus $ARMA(1,1)$ pour les trois cas.

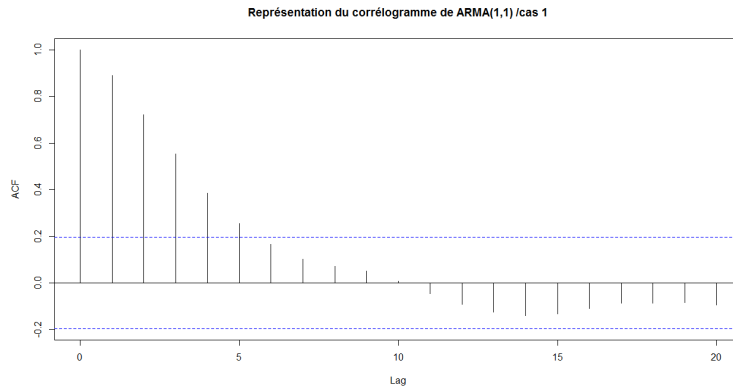


FIGURE 4.26 – Graphe du corrélogramme 1 de $ARMA(1,1)$

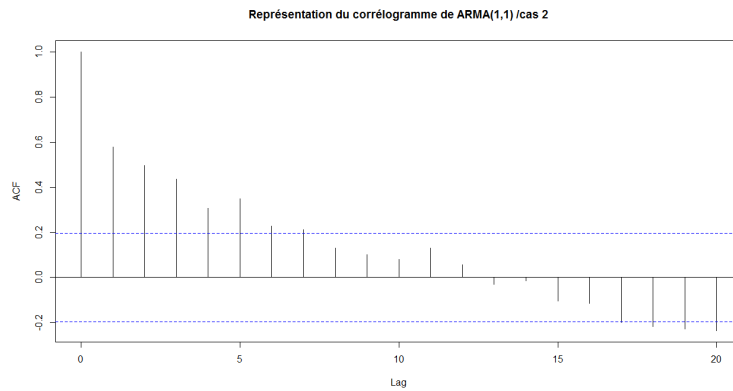


FIGURE 4.27 – Graphe du corrélogramme 2 de $ARMA(1,1)$

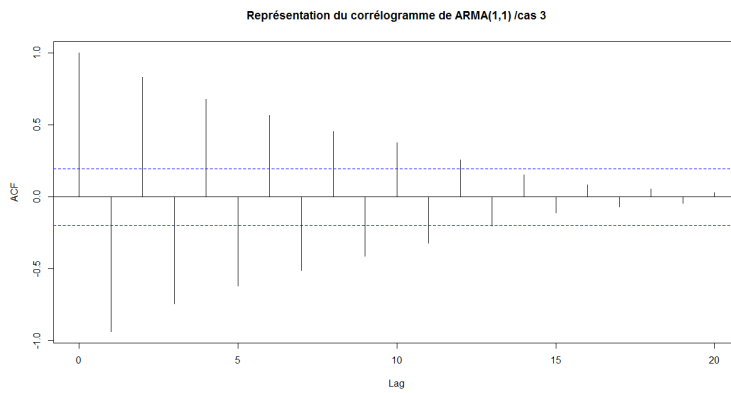


FIGURE 4.28 – Graphe du corrélogramme 3 de $ARMA(1,1)$

La représentation graphique du corrélogramme partielle du processus $ARMA(1,1)$ pour les trois cas.

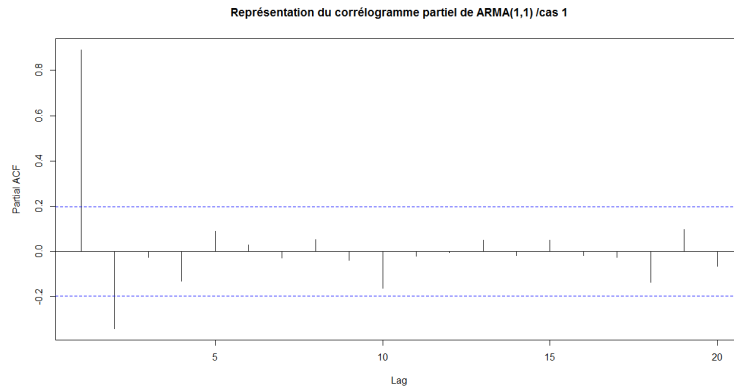


FIGURE 4.29 – Graphe du corrélogramme partielle 1 de $ARMA(1,1)$

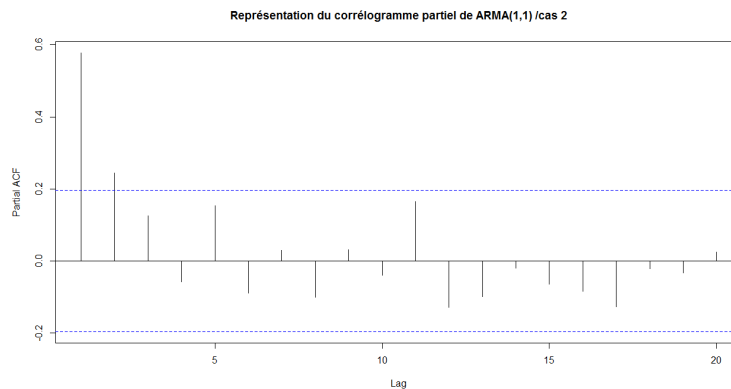


FIGURE 4.30 – Graphe du corrélogramme partielle 2 de $ARMA(1,1)$

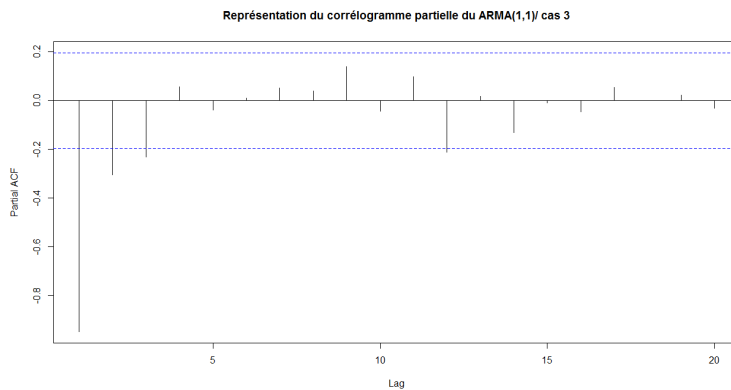


FIGURE 4.31 – Graphe du corrélogramme partielle 3 de $ARMA(1,1)$

On remarque que les autocorrélations du processus $ARMA(1,1)$ soit décroît exponentiellement dans un seul sens, soit alternant en signe, exactement comme le processus $AR(1)$ sauf que la décroissance commence dès ρ_1 . Pour les autocorrélations partielles du processus $ARMA(1,1)$ soit décroît aussi exponentiellement dans un seul sens, soit alternant en signe, exactement comme le processus $MA(1)$ sauf que la décroissance commence après la valeur initiale de $\lambda_{1,1} = \rho_1$.

En résumé, on peut présenter les caractéristiques majeures des deux fonctions d'autocorrélation et d'autocorrélation partielle des différents processus stationnaires considérés jusqu'ici dans un tableau récapitulatif :

Processus	Fonction d'autocorrélation	Fonction d'autocorrélation partielle
$MA(q)$	Annulation après q premiers coefficients	Décroissance exponentielle ou alternance en signe
$AR(p)$	Décroissance exponentielle ou alternance en signe	Annulation après p premiers coefficients

CHAPITRE

5

PROCESSUS NON STATIONNAIRES

Dans les chapitres précédents, nous avons vu qu'une des première étape de la démarche de modélisation d'une série chronologique consiste à vérifier la stationnarité du processus. Généralement, on se limite à vérifier la stationnarité faible ou stationnarité du second ordre. Nous allons à présent étudier de façon plus précise ce qu'est un processus non stationnaire.

Il existe en effet deux sorte de la non stationnarité : la non stationnarité déterministe et la non stationnarité stochastique. Nous verrons que suivant l'origine de la non stationnarité, il convient d'adopter une méthode de stationnarisation particulière.

Dans le troisième chapitre, nous avons introduit la notion de stationnarité du second ordre ou stationnarité faible. D'après cette définition, un processus est stationnaire au second ordre si l'ensemble de ses moments d'ordre un et d'ordre deux sont indépendants du temps. Par opposition, un processus non stationnaire est un processus qui ne satisfait pas l'une ou l'autre de ces deux conditions. Ainsi, l'origine de la non stationnarité peut provenir d'une dépendance du moment d'ordre un (l'espérance) par rapport au temps et/ou d'une dépendance de la variance ou des autocovariances par rapport au temps.

Le fait qu'un processus soit stationnaire ou non conditionne le choix de la modélisation que l'on doit adopter. En règle générale, si la série étudiée est issue d'un processus stationnaire, on cherche alors le meilleur modèle parmi la classe des processus stationnaire pour la représenter, puis on estime ce modèle. En revanche, si la série est issue d'un processus non stationnaire, on doit avant toutes choses, chercher à la "stationnariser", c'est à dire trouver une transformation stationnaire de ce processus. Puis, on modélise et l'on estime les paramètres associés à la composante stationnaire.

La difficulté réside dans le fait qu'il existe différentes sources de non stationnarité et qu'à chaque origine de la non stationnarité est associée une méthode propre de stationnarisation. Nous allons donc commencer dans cette section par présenter deux classes de processus non stationnaires, selon la terminologie de Nelson et Plosser : les processus **TS** (Trend Stationary) et les processus **DS** (Differency Stationary).

5.1 Les processus TS

Commençons par définir ce qu'est un processus **TS** pour Trend Stationary, selon la terminologie proposée par Nelson et Plosser.

Définition 31

$(X_t, t \in \mathbb{Z})$ est un processus TS s'il peut s'écrire sous la forme :

$$X_t = f(t) + z_t$$

où $f(t)$ est une fonction du temps et z_t est un processus stochastique stationnaire. [9]

L'exemple le plus simple d'un processus TS est celui d'une tendance polynomiale perturbée par un bruit blanc.

Exemple 5.1.1. *Soit*

$$X_t = b_0 + b_1 t + \eta_t \tag{5.1}$$

où $(b_0, b_1) \in \mathbb{R}^2$ et $\eta_t \sim bb(0, \sigma^2)$
Alors

$$E(X_t) = b_0 + b_1 t$$

Le processus X_t est non stationnaire puisque l'espérance dépend du temps, ce qui viole la première condition de la définition d'un processus stationnaire.

le processus Y_t défini par l'écart entre X_t et la composante déterministe $f(t) = b_0 + b_1 t$, est quand à lui stationnaire : $Y_t = X_t - f(t) = \eta_t$ est un bruit blanc, par définition stationnaire.

Une des propriétés importantes de ce type de processus réside dans l'influence des innovations stochastiques η_t : En effet, lorsque un processus TS est affecté par un choc stochastique, l'effet de ce choc tend à disparaître au fur et à mesure que le temps passe : c'est la propriété de non persistance des chocs. De façon plus formelle, cette propriété est la suivante :

Proposition 23

L'influence d'un choc η_t à une date T sur un processus X_t défini par

$$X_t = f(t) + z_t$$

avec z_t stationnaire et $E(z_t) = 0$, est transitoire. La tendance du modèle étant déterministe, après le choc η_T , la séquence des X_t converge ainsi vers sa valeur de long terme définie par $f(t)$. Il n'y a pas de persistance des chocs. [9]

Cela signifie que lorsque l'on a un processus TS, en cas de choc positif ou négatif à une date donnée, toutes choses égales par ailleurs, l'influence de ce choc a tendance à s'estomper au cours du temps. La variable considérée rejoint alors sa dynamique de long terme déterminée par $f(t)$. Dans le cas où $f(t)$ est une fonction affine du temps, la variable rejoint la tendance linéaire de long terme. Cette propriété traduit l'existence d'une tendance non stochastique, et qui donc ne présente pas de rupture dès lors que la fonction $f(t)$ est continue.

Afin d'illustrer cette propriété considérons l'exemple suivant où l'on a introduit une structure autorégressive dans la perturbation z_t :

$$\begin{aligned} X_t &= b_0 + b_1 t + z_t \\ z_t &= \phi_1 z_{t-1} + \eta_t \end{aligned}$$

où $(b_0, b_1) \in \mathbb{R}^2$, $|\phi_1| < 1$ et $\eta_t \sim \text{bb}(0, \sigma^2)$

Le processus z_t est un $AR(1)$ stationnaire, puisque la racine associée à son polynôme autorégressif, égale à $1/\phi_1$, est supérieure à l'unité en module. Admettons que $E(z_t) = 0$. Étudions à présent l'influence du choc η_t à une date T quelconque sur la séquence des $(X_t, t \geq T)$. Pour ce faire, appliquons la décomposition de Wold au processus stationnaire z_t , il vient :

$$z_t = \sum_{i=0}^{\infty} \phi_1^i \eta_{t-i}$$

On peut alors réécrire le processus X_t sous la forme suivante :

$$X_t = b_0 + b_1 t + \sum_{i=0}^{\infty} \phi_1^i \eta_{t-i}$$

Supposons qu'à la date T , on a ait une réalisation du choc η_T positive ($\eta_T > 0$) et qu'ensuite les chocs η_t pour $t \geq T$ soient nuls. A la date T , on a donc :

$$X_T = b_0 + b_1 T + \sum_{i=0}^{\infty} \phi_1^i \eta_{T-i}$$

A la date $T + 1$, les chocs η_{T+1} étant nul, on obtient :

$$X_{T+1} = b_0 + b_1(T + 1) + \sum_{i=0}^{\infty} \phi_1^i \eta_{T+1-i}$$

De façon générale, à toute date $T + k, k \geq 0$, le processus X_{T+k} est défini par :

$$X_{T+k} = b_0 + b_1(T + k) + \sum_{i=0}^{\infty} \phi_1^i \eta_{T+k-i}$$

On peut alors montrer que plus le temps "passe", c'est à dire plus k est élevé, plus l'influence du choc η_T de la date T s'estompe. L'influence de ce choc est donc transitoire. En effet, la v.a.r. X_{T+k} va converger vers sa valeur de long terme définie par la tendance linéaire $f()$, Pour le montrer, il suffit de considérer l'écart entre X_{T+k} et la valeur correspondante de la tendance $f(T + k) = b_0 + b_1(T + k)$ et de montrer que cet écart converge vers 0 quand k tend vers l'infini. Ce résultat illustre l'absence de persistance des chocs, ou l'absence d'hystérésis pour les processus TS.

5.2 Les processus DS

Comme nous l'avons précédemment mentionné, il existe une autre forme de non stationnarité, provenant non pas de la présence d'une composante déterministe tendancielle, mais d'une source stochastique. C'est pourquoi nous allons à présent introduire la définition des processus DS pour Differency Stationnary.

Définition 32

Un processus non stationnaire $(X_t, t \in \mathbb{Z})$ est un processus DS d'ordre d , où d désigne l'ordre d'intégration, si le processus filtré défini par $(1 - B)^d X_t$ est stationnaire. On dit aussi que $(X_t, t \in \mathbb{Z})$ est un processus intégré d'ordre d , noté $I(d)$. [9]

Ainsi, on peut définir une classe de processus stochastiques qui ne satisfont pas les conditions de la stationnarité, mais dont la différence à l'ordre d satisfait elle les propriétés de la stationnarité. Par exemple, si un processus z_t n'est pas stationnaire, on dit que ce processus est DS, intégré d'ordre un, noté $I(1)$, si le processus défini par la différence première $\nabla z_t = z_t - z_{t-1}$ est quant à lui stationnaire. De la même façon, le processus z_t est $I(2)$ si le processus défini par la différence seconde $(1 - B)^2 z_t = (1 - B) \nabla z_t = z_t - 2z_{t-1} + z_{t-2}$ est stationnaire.

Proposition 24

Un processus non stationnaire $(X_t, t \in \mathbb{Z})$ est un processus DS intégré d'ordre d , noté $I(d)$, si le polynôme $\Phi(B)$ défini en l'opérateur retard B , associé à sa composante autorégressive admet d racines unitaires :

$$\Phi(B)X_t = z_t \quad \text{avec} \quad \Phi(B) = (1 - L)^d \bar{\Phi}(B) \quad (5.2)$$

où z_t est un processus stationnaire, et si les racines du polynôme $\bar{\Phi}(B)$ sont toutes supérieures strictement à l'unité en module. [9]

En effet, on pose $\Phi(B)X_t = z_t$ avec $\Phi(B) = (1 - B)^d \bar{\Phi}(B)$, Si l'on admet que les racines du polynôme $\bar{\Phi}(B)$ sont inférieures à l'unité en module, ce polynôme est inversible. On peut alors écrire la différence $d^{\text{ème}}$ de X_t sous la forme d'une somme de valeurs retardées de z_t :

$$(1 - B)^d \bar{\Phi}(B)X_t = z_t \Leftrightarrow (1 - B)^d X_t = \Psi(B)z_t \quad (5.3)$$

où $\Psi(B) = (\bar{\Phi}(B))^{-1}$, Si z_t est un processus stationnaire, la somme pondérée de ses valeurs passées $\Psi(L)z_t$ est elle aussi stationnaire. Donc, en résumé, dès lors que le polynôme autorégressif $\Phi(B)$ de X_t admet d racines unitaires, la quantité $(1 - B)^d X_t$ est stationnaire, donc le processus X_t est $I(d)$. L'exemple le plus simple d'un processus DS est celui d'une tendance polynomiale perturbée par un bruit blanc.

Exemple 5.2.1.

$$X_t = b_0 + b_1 t + b_2 t^2 + \eta_t \quad (5.4)$$

où $(b_0, b_1, b_2) \in \mathbb{R}^3$ et $\eta_t \sim bb(0, \sigma^2)$

Alors

$$E(X_t) = b_0 + b_1 t + b_2 t^2$$

Le processus X_t est non stationnaire puisque l'espérance dépend du temps, ce qui viole la première condition de la définition d'un processus stationnaire. il faut introduire un opérateur de différenciation d'ordre 2 $[\nabla^2]$, afin de stabiliser la série, comme suit :

$$\begin{aligned} Y_t &= \nabla^2 X_t \\ &= (X_t - X_{t-1})^2 \\ &= (1 - B^2)X_t \\ &= (1 - 2B + B^2)(b_0 + b_1 t + b_2 t^2 + \eta_t) \\ &= b_0 - 2b_0 + b_0 + b_1 t - 2b_1(t - 1) + b_1(t - 2) + b_2 t^2 - 2b_2(t - 1)^2 + b_2(t - 2)^2 + \eta_t - 2\eta_{t-1} + \eta_{t-2} \\ &= 2b_2 + \eta_t - 2\eta_{t-1} + \eta_{t-2} \\ &= b' + \epsilon_t \end{aligned}$$

où $\epsilon_t \sim bb(0, \sigma'^2)$

Donc $E(Y_t) = b'$ ne dépend pas de t .

Dans la classe générale des processus DS, un type de processus apparaît de façon régulière, si bien que l'on lui a attribué un nom particulier : la marche aléatoire.

Exemple 5.2.2.

$$X_t = X_{t-1} + \eta_t \quad (5.5)$$

où $\eta_t \sim bb(0, \sigma^2)$

Alors

$$X_t = \eta_1 + \eta_2 + \dots + \eta_t$$

On a alors :

$E(X_t) = 0$ ne dépend pas de t

$$Var(X_t) = Var\left(\sum_{i=1}^t \eta_i\right) = \sum_{i=1}^t Var(\eta_i) = t\sigma^2 \text{ dépend de } t$$

Le processus X_t est non stationnaire puisque la variance dépend du temps, ce qui viole la seconde condition de la définition d'un processus stationnaire. On va visualiser la marche aléatoire sous R, prenons $t=10$ et initialiser $\eta_0 = 0$, on obtient :

$$\eta_1 = 0.00000000; \eta_2 = -0.50219235; \eta_3 = 0.13153117; \eta_4 = -0.07891709; \eta_5 = 0.88678481; \\ \eta_6 = 0.11697127; \eta_7 = 0.31863009; \eta_8 = -0.58179068; \eta_9 = 0.71453271; \eta_{10} = -0.82525943$$

Alors :

$$X_1 = 0.00000000; X_2 = -0.5021924; X_3 = -0.3706612; X_4 = -0.4495783; X_5 = 0.4372065; \\ X_6 = 0.5541778; X_7 = 0.8728079; X_8 = 0.2910172; X_9 = 1.0055499; X_{10} = 0.1802905$$

Le graphe est le suivant :

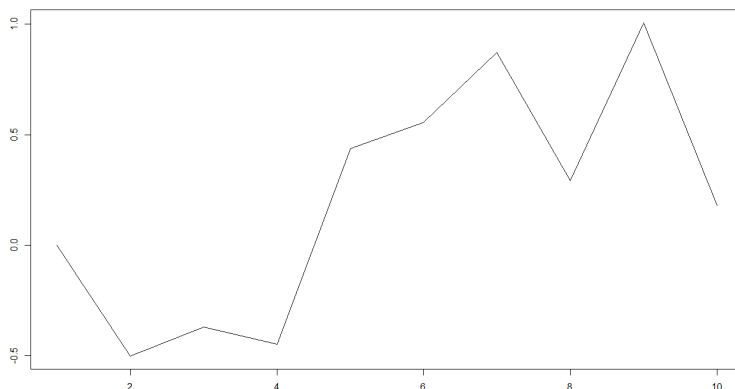


FIGURE 5.1 – Graphe du marche aléatoire

Il est clair que le graphe est non stationnaire. il faut donc introduire un nouveau processus Y_t afin de stabiliser la série, comme suit :

$$Y_t = \nabla X_t = X_t - X_{t-1} = \eta_t$$

Donc $E(Y_t) = 0$ et $Var(X_t) = \sigma^2$ ne dépend pas de t . Revenons à logiciel R, à l'aide de la commande "diff.nom de la série à stationnariser.ts", la série rend stationnaire, et on obtient le graphe suivant :

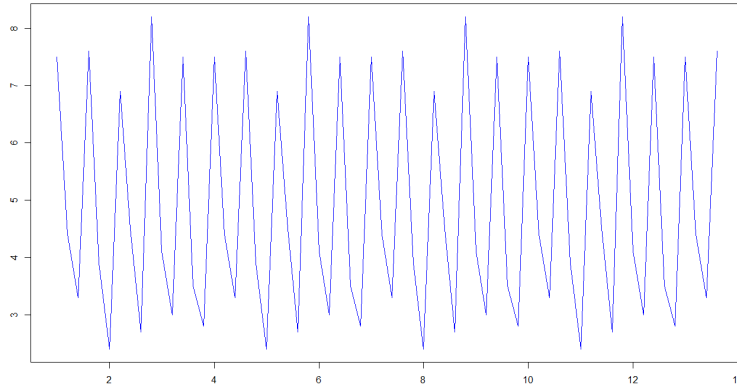


FIGURE 5.2 – Graphe du marche aléatoire différencier

Généralement, si $Var(X_t) = cf(m_t)$ avec $c > 0$ et $f(.)$ fonction connue à valeur positive et m_t moyenne dépend du temps, il faut utiliser la transformation suivante :

$$Y_t = T(X_t) = \frac{X_t^k - 1}{k} \quad \text{où } k \in \mathbb{R}$$

Le tableau suivant donne les valeurs du k les plus utilisées dans la transformation :

k	-0,1	-0,5	0,0	0,5	1,0
Y_t	$\frac{1}{X_t}$	$\frac{1}{\sqrt{X_t}}$	$\ln(X_t)$	$\sqrt{X_t}$	X_t

5.2.1 Processus ARIMA

L'idée générale, essentiellement conçue pour les processus non stationnaires à tendance polynomiale, est de différencier suffisamment le processus initial afin d'obtenir un processus sans tendance et sur la partie différenciée appliquer un modèle *ARMA*. Cette classe de modèle est connue sous le nom de processus *ARIMA* (AutoRegressive Integrated Moving Average).

Définition 33

On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ admet une représentation *ARIMA*(p, d, q) s'il vérifie l'équation :

$$\Phi_p(B) \nabla^d X_t = m + \Theta_q(B)\eta_t \Leftrightarrow \Phi_p(B)(1 - B)^d X_t = m + \Theta_q(B)\eta_t \quad (5.6)$$

où $\eta_t \sim bb(0, \sigma^2)$ [7]

On peut montrer que, dans le cas d'un processus *ARIMA*(p, d, q), le processus $\nabla^d X_t$ est asymptotiquement, au sens quand t tend vers $+\infty$, un processus *ARMA*(p, q).

Exemple 5.2.3. Soit le processus *ARIMA*(1, 1, 0) = *ARI*(1, 1) défini par :

$$\begin{aligned} \Phi_1(B)(1 - B)X_t &= m + \eta_t \\ \Leftrightarrow (1 - \Phi_1 B)(1 - B)X_t &= m + \eta_t \\ \Leftrightarrow [1 - (\Phi_1 + 1)B + \Phi_1 B^2]X_t &= m + \eta_t \end{aligned}$$

Donc $X_t = (\Phi_1 + 1)X_{t-1} - \Phi_1 X_{t-2} + m + \eta_t$ où $|\Phi_1| < 1$

5.2.2 Processus SARIMA

Si l'on veut en même temps traiter les saisonnalités de période s , on est amenés à définir les processus *SARIMA*.

Définition 34

Soient $p, d, q, s \geq 0$ et $P, Q, D \geq 0$. Un processus (X_t) est un processus *SARIMA* $(p, d, q)(P, D, Q)_s$ [Autoregressif moyenne mobile et saisonnalité intégrées] si

$$Y_t = \nabla^d \nabla_s^D X_t = (1 - B)^d (1 - B^s)^D X_t \quad (5.7)$$

est un processus *SARMA* $(p, q)(P, Q)_s$ stationnaire de la forme :

$$\Phi_p(B) \Psi_P(B^s) Y_t = \Theta_q(B) \Xi_Q(B^s) \eta_t \quad (5.8)$$

où Φ (resp. Θ) est le polynôme générateur d'un *AR* (p) (resp. d'un *MA* (q)) [14]

Exemple 5.2.4. Soit le processus *SARMA* $(0, 1)(1, 1)_{12}$ défini par :

$$\Phi_0(B) \Psi_1(B^{12}) Y_t = \Theta_1(B) \Xi_1(B^{12}) \eta_t$$

équivalent à dire :

$$(1 - \psi_1 B^{12}) Y_t = (1 - \theta_1 B) (1 - \xi_1 B^{12}) \eta_t$$

$$Y_t - \psi_1 Y_{t-12} = \eta_t - \theta_1 \eta_{t-1} - \xi_1 \eta_{t-12} + \xi_1 \theta_1 \eta_{t-13}$$

$$Y_t = \psi_1 Y_{t-12} + \eta_t - \theta_1 \eta_{t-1} - \xi_1 \eta_{t-12} + \xi_1 \theta_1 \eta_{t-13} \quad (5.9)$$

On multiplie l'équation (1.9) par Y_t on trouve :

$$E(Y_t Y_t) = \psi_1 E(Y_t Y_{t-12}) + E(Y_t \eta_t) - \theta_1 E(Y_t \eta_{t-1}) - \xi_1 E(Y_t \eta_{t-12}) + \xi_1 \theta_1 E(Y_t \eta_{t-13})$$

$$\Leftrightarrow \gamma_0 = \psi_1 \gamma_{12} + \sigma^2 + \theta_1^2 \sigma^2 - \xi_1 (\psi_1 - \xi_1) + \xi_1 \theta_1 (-\psi_1 \theta_1 + \theta_1 \xi_1)$$

$$\Leftrightarrow \gamma_0 = \psi_1 \gamma_{12} + \sigma^2 (1 + \theta_1^2) (1 + \xi_1 (\psi_1 - \xi_1))$$

On multiplie l'équation (1.9) par Y_{t-12} on obtient :

$$\gamma_{12} = \psi_1 \gamma_0 - \xi_1 \sigma^2 + \xi_1 \theta_1 (-\theta_1) \sigma^2$$

$$\Leftrightarrow \gamma_{12} = \psi_1 \gamma_0 - \xi_1 \sigma^2 (1 + \theta_1^2)$$

Donc :

$$\gamma_0 = \sigma^2 (1 + \theta_1^2) \frac{1 + \xi_1^2 - 2\psi_1 \xi_1}{1 - \psi_1^2}$$

$$\gamma_{12} = \sigma^2 (1 + \theta_1^2) \left[\psi_1 - \xi_1 + \frac{\psi_1 (\xi_1 - \psi_1)^2}{1 - \psi_1^2} \right]$$

de la même façon on trouve : $\gamma_2 = \gamma_3 = \dots = \gamma_{10} = 0$ et $\gamma_{13} = \gamma_{11}$ et $\gamma_k = \psi_1 \gamma_{k-12}$ $k > 13$

Donc :

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1 & k = 0 \\ \frac{-\theta_1}{1 + \theta_1^2} & k = 1 \\ 0 & k = 2, \dots, 10 \\ \frac{\theta_1 (\xi_1 - \psi_1) (1 - \xi_1 \psi_1)}{1 + \theta_1^2 (1 + \xi_1^2 - 2\psi_1 \xi_1)} & k = 11 \\ \frac{-(\xi_1 - \psi_1) (1 - \xi_1 \psi_1)}{1 + \xi_1^2 - 2\psi_1 \xi_1} & k = 12 \\ \rho_{11} & k = 13 \\ \psi_1 \rho_{k-12} & k > 13 \end{cases}$$

5.3 Construction d'un système de prévision

La prévision comporte essentiellement les trois étapes suivantes :

- 1- Identification du modèle.
- 2- Estimation des paramètres du modèle.
- 3- Diagnostic du modèle.

5.3.1 Identification du modèle

Cette étape consiste à comprendre les observations, les données (l'histoire), afin de bien choisir le processus adéquat au problème donnée. Pour atteindre ce but, il faut suivre les étapes suivantes :

1- Transformation

Une des premiers étape de la démarche de modélisation d'une série chronologique consiste à vérifier la stationarité du processus, comme nous avons montrer précédemment.

2- Choix d'ordre d'intégration ou de différenciation d

Si la série est non stationnaire de type DS, premièrement il faut déterminer d l'ordre d'intégration qui rend la série stationnaire, ensuite tracer le graphes de la série, le ACF (graphe des autocorrélations) et la PACF (graphe des autocorrélations partielles) :

- Si le graphe de la série chronologique montre un changement de niveau, alors la série est non stationnaire.
- Si le corrélogramme est lentement décroît, alors la série est non stationnaire.
- Si le corrélogramme partielle donne une seule valeur proche de 1 (ignore le signe) et les autres valeurs proche de 0, alors la série est non stationnaire.

Donc il faut choisir de nouveau l'ordre d'intégration d [enlever d] et répéter l'étude des graphes, généralement d=0 ou d=1 ou d=2.

Remarque 5.3.1. *Comment tracer les graphes d'autocorrélation et d'autocorrélation partielle sans connaitre les paramètres ? la réponse est simple, on va estimer les fonctions d'autocorrélations par $\hat{\rho}_h = r_h$ et les fonctions d'autocorrélations partielles par $\hat{\lambda}_{h,h} = r_{h,h}$ où :*

$$\hat{\rho}_h = r_h = \frac{\sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{n-h} (X_t - \bar{X})^2}$$

$$\hat{\lambda}_{h,h} = r_{h,h} = \frac{r_h - \sum_{i=1}^{h-1} r_{h-1,i} r_{h-i}}{1 - \sum_{i=1}^{h-1} r_{h-1,i} r_i}$$

Où $h \in \mathbb{Z}$ et $r_{h,i} = r_{h-1,i} - r_{h,h} r_{h-1,h-i} \quad \forall i \in 1, \dots, h-1$ et $r_{0,0} = 1; \quad r_{1,1} = r_1 = \hat{\rho}_1$

* **Test sur les autocorrélations** $\begin{cases} H_0 : \rho_h = 0 \\ H_1 : \rho_h \neq 0 \end{cases}$ si $\sqrt{n}|r_h| > z_{\frac{\alpha}{2}}$ on accepte H_1

* **Test sur les autocorrélations partielles** $\begin{cases} H_0 : \lambda_{h,h} = 0 \\ H_1 : \lambda_{h,h} \neq 0 \end{cases}$ si $\sqrt{n}|r_{h,h}| > z_{\frac{\alpha}{2}}$ on accepte H_1

3- Détermination de l'ordre d'autorégressif p et l'ordre de moyenne mobile q

Si l'étape précédente est bien passée, on cherche alors l'ordre d'autorégressif p et l'ordre de moyenne mobile q, on utilisons le graphe des ACF et PACF, le tableau ci-dessus présente les caractères graphiques des processus.

Processus	corrélogramme	corrélogramme partielle
AR(1)	décroissance exponentielle	$\lambda_{h,h} = 0 \quad h > 1$
AR(p)	décroissance exponentielle et/ou périodique	$\lambda_{h,h} = 0 \quad h > p$
MA(1)	$\rho_h = 0 \quad h > 1$	décroissance exponentielle
MA(q)	$\rho_h = 0 \quad h > q$	décroissance exponentielle

4- L'ajout de δ

Pour préciser est ce que la série est centrée ou non, on doit comparer la moyenne de la série Y_t [$Y_t = (1 - B)^d X_t$] avec l'erreur quadratique $se(\bar{Y}_t) \cong \sqrt{\frac{c_0}{n}(1 + 2r_1 + \dots + 2r_K)}$ où $c_0 = \hat{\gamma}_0$ et $r_i = \hat{\rho}_i$ de degré K.

Le test suivant détermine l'existence du paramètre δ .

$$\begin{cases} H_0 : \delta = 0 \\ H_1 : \delta \neq 0 \end{cases}$$

si $|\frac{\bar{Y}}{se(\bar{Y})}| > z_{\frac{\alpha}{2}}$ on accepte H_1

5.3.2 Estimation du modèle

Après la détermination du modèle, il faut déterminer les paramètres $\delta, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma$, existent plusieurs méthodes pour préciser les paramètres, mais on va traiter seulement deux méthodes principales : méthode des moments et méthode des moindres carrés conditionnels (CLSM).

1- Méthode des moments

Présentons la méthode pour le processus $AR(p)$, soit (X_t) une série chronologique.

Considérons $\hat{m}, \hat{\phi}_1, \dots, \hat{\phi}_p$ les estimations de m, ϕ_1, \dots, ϕ_p où $\hat{m} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, d'après les

équations de Yule-Walker on a :

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \dots \\ \dots \\ \dots \\ \rho_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \rho_2 & \rho_1 & \dots & \rho_{p-3} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_p \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \dots \\ \dots \\ \dots \\ \phi_p \end{pmatrix}$$

Donc :

$$\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ \dots \\ \dots \\ \dots \\ r_p \end{pmatrix} = \begin{pmatrix} 1 & r_1 & \dots & r_{p-1} \\ r_1 & 1 & \dots & r_{p-2} \\ r_2 & r_1 & \dots & r_{p-3} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & \dots & 1 \end{pmatrix} \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \\ \dots \\ \dots \\ \dots \\ \hat{\phi}_p \end{pmatrix} \quad \text{D'où} \quad \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \\ \dots \\ \dots \\ \dots \\ \hat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & r_1 & \dots & r_{p-1} \\ r_1 & 1 & \dots & r_{p-2} \\ r_2 & r_1 & \dots & r_{p-3} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ \dots \\ \dots \\ \dots \\ r_p \end{pmatrix}$$

σ^2 est estimé par $\hat{\sigma}^2 = \hat{\gamma}_0(1 - \hat{\phi}_1 r_1 - \dots - \hat{\phi}_p r_p)$ où $\hat{\gamma}_0 = \frac{\sum_{t=1}^n (X_t - \bar{X})^2}{n}$.

Exemple 5.3.2. 1)- Soit $(X_t)_{t \in \mathbb{Z}}$ un $AR(2)$ non centré, $(X_t)_{t \in \mathbb{Z}}$ est donc généré par un bruit blanc $\eta_t \sim \mathcal{N}(0, \sigma^2)$, sous la forme : $\Phi_2(B)X_t = \delta + \eta_t \Leftrightarrow X_t - m = \phi_1(X_{t-1} - m) + \phi_2(X_{t-2} - m) + \eta_t$

D'après les équations de Yule-Walker on obtient :
$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} = \begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \Leftrightarrow \begin{cases} \hat{\phi}_1 = \frac{r_1 - r_1 r_2}{1 - r_1^2} \\ \hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \end{cases}$$

2)- Soit $(X_t)_{t \in \mathbb{Z}}$ un ARMA(1,1) non centré, $(X_t)_{t \in \mathbb{Z}}$ est donc généré par un bruit blanc $\eta_t \sim \mathcal{N}(0, \sigma^2)$, sous la forme : $\Phi_1(B)X_t = \delta + \Theta_1(B)\eta_t \Leftrightarrow X_t - m = \phi_1(X_{t-1} - m) + \eta_t - \theta_1\eta_{t-1}$
On sait que :

$$\rho_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\theta_1\phi_1} \text{ et } \rho_2 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\theta_1\phi_1}\phi_1$$

Donc :

$$r_1 = \frac{(1 - \hat{\theta}_1\hat{\phi}_1)(\hat{\phi}_1 - \hat{\theta})}{1 + \hat{\theta}_1^2 - 2\hat{\theta}_1\hat{\phi}_1} \text{ et } r_2 = \frac{(1 - \hat{\theta}_1\hat{\phi}_1)(\hat{\phi}_1 - \hat{\theta}_1)}{1 + \hat{\theta}_1^2 - 2\hat{\theta}_1\hat{\phi}_1}\hat{\phi}_1$$

On déduit

$$\frac{r_2}{r_1} = \hat{\phi}_1 \quad (5.10)$$

On remplace (5.10) dans r_1 et on trouve :

$$r_1 = \frac{(1 - \frac{r_2}{r_1}\hat{\theta}_1)(\frac{r_2}{r_1} - \hat{\theta})}{1 + \hat{\theta}_1^2 - 2\frac{r_2}{r_1}\hat{\theta}_1} = \frac{(r_1 - r_2\hat{\theta}_1)(r_2 - \hat{\theta}r_1)}{r_1 + r_1\hat{\theta}_1^2 - 2r_2\hat{\theta}_1}$$

On résout l'équation et on choisit $\hat{\theta}_1$ tel que $|\hat{\theta}_1| < 1$

2- Méthode des moindres carrés conditionnels (CLSM)

Soit $(X_t)_{t \in \mathbb{Z}}$ un ARMA(p,q) non centré, $(X_t)_{t \in \mathbb{Z}}$ est donc généré par un bruit blanc $\eta_t \sim \mathcal{N}(0, \sigma^2)$, sous la forme : $\Phi_p(B)X_t = \delta + \Theta_q(B)\eta_t \Leftrightarrow \Phi_p(B)(X_t - m) = \Theta_q(B)\eta_t$

Alors $\eta_t = \frac{\Phi_p(B)}{\Theta_q(B)}(X_t - m)$

Donc $\eta_t(\Phi, \Theta, m) = \frac{\Phi_p(B)}{\Theta_q(B)}(X_t - m)$ où $\phi = (\phi_1, \dots, \phi_p)$ et $\theta = (\theta_1, \dots, \theta_q)$

La méthode des moindres carrés conditionnels consiste à minimiser la fonction :

$$\min_{\phi, \theta, m} S_c(\phi, \theta, m) = \sum_{t=p+1}^n \eta_t^2(\phi, \theta, m/X)$$

et résoudre les équations résultantes, on posant $\phi = \hat{\phi}, \theta = \hat{\theta}, m = \hat{m}$

$$\begin{aligned} \frac{\partial S_c}{\partial \phi}(\phi, \theta, m) &= \frac{\partial}{\partial \phi} \sum_{t=p+1}^n \eta_t^2(\phi, \theta, m/X) = 0 \\ \frac{\partial S_c}{\partial \theta}(\phi, \theta, m) &= \frac{\partial}{\partial \theta} \sum_{t=p+1}^n \eta_t^2(\phi, \theta, m/X) = 0 \\ \frac{\partial S_c}{\partial m}(\phi, \theta, m) &= \frac{\partial}{\partial m} \sum_{t=p+1}^n \eta_t^2(\phi, \theta, m/X) = 0 \end{aligned}$$

l'estimation de σ^2 est donnée par : $\hat{\sigma}^2 = \frac{S_c(\hat{\phi}, \hat{\theta}, \hat{m})}{n - (p + q + 1)}$

Exemple 5.3.3. 1)- Soit $(X_t)_{t \in \mathbb{Z}}$ un AR(1) non centré, $(X_t)_{t \in \mathbb{Z}}$ est donc généré par un bruit blanc $\eta_t \sim \mathcal{N}(0, \sigma^2)$, sous la forme : $\Phi_1(B)X_t = \delta + \eta_t \Leftrightarrow X_t - m = \phi_1(X_{t-1} - m) + \eta_t$
On remplace m par \bar{X} :

$$\eta_t = (X_t - \bar{X}) - \phi_1(X_{t-1} - \bar{X}) \Leftrightarrow \eta_t(\phi_1) = (X_t - \bar{X}) - \phi_1(X_{t-1} - \bar{X}) \quad t \in 2, \dots, n$$

D'où

$$S_c(\phi_1) = \sum_{t=2}^n \eta_t^2(\phi_1) = \sum_{t=2}^n [(X_t - \bar{X}) - \phi_1(X_{t-1} - \bar{X})]^2$$

Donc

$$\frac{\partial S_c}{\partial \phi_1}(\phi_1) = -2 \sum_{t=2}^n (X_{t-1} - \bar{X}) [(X_t - \bar{X}) - \hat{\phi}_1(X_{t-1} - \bar{X})] = 0$$

On déduit alors :

$$\sum_{t=2}^n (X_{t-1} - \bar{X})(X_t - \bar{X}) = \hat{\phi}_1 \sum_{t=2}^n (X_{t-1} - \bar{X})^2$$

D'où

$$\hat{\phi}_1 = \frac{\sum_{t=2}^n (X_{t-1} - \bar{X})(X_t - \bar{X})}{\sum_{t=2}^n (X_{t-1} - \bar{X})^2}$$

2)- Soit $(X_t)_{t \in \mathbb{Z}}$ un MA(1) non centré, $(X_t)_{t \in \mathbb{Z}}$ est donc généré par un bruit blanc $\eta_t \sim \mathcal{N}(0, \sigma^2)$, sous la forme : $X_t - m = \eta_t - \theta_1 \eta_{t-1}$

On remplace m par \bar{X} et $Y_t = X_t - m$:

$$\eta_t = Y_t + \theta_1 \eta_{t-1} \tag{5.11}$$

Donc,

pour $t=1$: $\eta_1 = Y_1$

pour $t=2$: $\eta_2 = Y_2 + \theta_1 \eta_1$

pour $t=3$: $\eta_3 = Y_3 + \theta_1 \eta_2$

.....

pour $t=n$: $\eta_n = Y_n + \theta_1 \eta_{n-1}$

l'équation (5.11) est non linéaire, alors on utilise les méthodes numériques [Gauss-Newton, Sécante, Dichotomie,...] pour trouver θ_1 qui minimise $S_c(\theta_1)$.

Utilisons la méthode de Gauss-Newton, qui consiste à approximer $\eta_t = \eta_t(\theta_1)$ par une fonction affine et une valeur initiale θ^* , c'est-à-dire :

$$\eta_t(\theta_1) \simeq \eta_t(\theta^*) + (\theta_1 - \theta^*) \frac{d\eta_t(\theta^*)}{d\theta_1} \tag{5.12}$$

où on peut calculer $\frac{d\eta_t(\theta^*)}{d\theta_1}$ on dérivons l'équation $\eta_t = Y_t + \theta_1 \eta_{t-1}$ par rapport au θ_1 , on obtient :

$$\frac{d\eta_t(\theta_1)}{d\theta_1} = \frac{\eta_t d\eta_{t-1}(\theta_1)}{\theta_1} + \eta_{t-1}(\theta_1) \tag{5.13}$$

avec une valeur initiale $\frac{d\eta_0(\theta_1)}{d\theta_1} = 0$, l'équation (5.12) affine par rapport à θ_1 et par suite on peut minimiser $S_c(\theta_1)$.

5.3.3 Le critère d'information d'akaike (AIC)

Il existe plusieurs critères d'information dont le but est de faciliter la sélection du processus représentant le mieux possible la réalité. Les critères d'information de Akaike (1974), Hannan et Quinn (1979), Hurvich et Tsai (1989) et Schwarz (1978) sont basés sur l'échantillon observé et peuvent servir à sélectionner le meilleur modèle parmi un ensemble de modèles possibles.

Quelques critères importants Le principe général d'information utilisé pour la construction de critères d'information est basé sur la divergence entre deux fonctions de densité f et g

$$I(f, g) = \int f(x) \ln \frac{f(x)}{g(x)}$$

On peut interpréter cette quantité comme étant une mesure du risque impliqué par la possibilité de choisir une fausse densité g au lieu de la bonne densité f . Alternativement, elle peut également représenter la surprise de découvrir que g est la vraie densité lorsque f avait été considérée comme la vraie. Si les deux densités font partie d'une classe paramétrique et que g contient k paramètres, il est possible d'obtenir une approximation de $I(f, g)$ en utilisant la méthode du maximum de vraisemblance dans un échantillon de taille T :

$$-2 \sum \ln g(\cdot, k) + 2k$$

Il s'agit de l'équation utilisée par Akaike (1974) pour obtenir le critère d'information d'akaike (AIC). Dans le contexte des séries chronologiques, ce critère traduit de la façon suivante :

$$AIC = \ln \sigma(k)^2 + \frac{2k}{T}$$

où k représente l'ordre d'un processus AR ou la somme des ordres AR et MA dans un modèle ARMA.

- Le critère d'information bayésien (BIC) le plus utilisé est :

$$BIC = -2 \sum \ln g(\cdot, k) + k \ln T$$

Il existe autre méthodes pour estimer un modèle, à titre d'exemple :

- Méthode de maximum vraisemblance.
- Méthode des moindres carrés inconditionnel.
- Méthodes des estimations non linéaire.

5.3.4 Diagnostic du modèle/ Validation du modèle

Après avoir estimé les différents paramètres du modèle proposé, il convient à présent de valider ce modèle, en servant d'une analyse sur les résidus estimés $e_t = X_t - \hat{X}_t = \hat{\eta}_t$ $t \in 1, \dots, n$, c'est-à-dire pour que le modèle proposé soit valide, il doit vérifier que les résidus estimés e_t suivent bien un bruit blanc η_t . Le test de bruit blanc repose sur l'analyse des résidus et la vérification des hypothèses suivantes du modèle :

- 1- $E(\eta_t) = 0$.
- 2- η_t aléatoire.
- 3- Test nature des résidus.

Pour cela, en servant des tests sur les résidus pour valider le modèle, s'il vérifie les conditions précédentes, sinon (si un des tests est rejetée) le modèle est mal posé et il faut proposer un autre modèle.

1- Test moyen

ce test est basé sur les deux hypothèses suivantes :

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

La statistique qui convient pour le test est $u = \frac{\bar{e}}{se(\bar{e})} \sim \mathcal{N}(0, \sigma^2)$

Si $|u| < z_{\frac{\alpha}{2}}$ on accepte H_0 .

2- Test aléatoire

Ce test est basé sur le nombre de séquences croissantes et décroissantes. Soit R le nombre de séquences dans l'échantillon de n nombres tel que :

$$E(R) = \frac{2n - 1}{3} \quad \text{et} \quad Var(R) = \frac{16n - 29}{90}$$

Alors on les hypothèses suivantes :

$$\begin{cases} H_0 : \text{l'échantillon est aléatoire} \\ H_1 : \text{l'échantillon n'est pas vraiment aléatoire} \end{cases}$$

La statistique qui convient pour le test est $u = \frac{R - E(R)}{\sigma(R)} \sim \mathcal{N}(0, Var(R))$

Si $|u| < z_{\frac{\alpha}{2}}$ on accepte H_0 .

Exemple 5.3.4. Prenons $n=15$ nombres obtenus par un générateur aléatoire produisant des nombres entiers entre 0 et 99 :

n_i	12	14	65	18	33	77	89	72	76	43	70	81	94	98	3
La croissance		+		-		+		-		-		+		+	-

$R = 8$ séquence : 4 croissantes et 4 décroissantes, où $E(R) = 9,66$ et $Var(R) = 2,34$

Comme $u = \frac{R - E(R)}{\sqrt{Var(R)}} = -1,089$ et $z_{\frac{\alpha}{2}} = 1,96$ au niveau $\alpha = 0,05$

Donc $|u| < z_{\frac{\alpha}{2}}$, on déduit alors que le test est réussi.

3- Test nature des résidus

Il existe plusieurs tests de normalité, pour déterminer si les résidus distribuent normalement ou non, les tests les plus connus sont :

- Goodness of Fit Test (test qualité de l'ajustement) : on trouve le test de Pearson-chi deux, le test de Kolmogorov-Smirnov.
- Normal Probability plot.
- Q-Q plot.

Existe aussi plusieurs tests d'adéquation sur les résidus, pour tester la bruité, les tests les plus connus sont :

- Approche de Box-Jenkins.
- Test de corrélation de Box-Pierce.

1- Test d'ajustement

a)- Test de Kolmogorov-Smirnov Le test de Kolmogorov-Smirnov est un test d'ajustement à une loi continue, qui prend en compte l'ensemble des quantiles. Le modèle est encore un échantillon (X_1, \dots, X_n) d'une loi inconnue P . L'hypothèse nulle est : \mathcal{H}_0 : la loi P a pour fonction de répartition F_0 , où F_0 est la fonction de répartition d'une loi continue donnée. L'idée est la suivante : si l'hypothèse \mathcal{H}_0 est correcte, alors la fonction de répartition empirique \hat{F} de l'échantillon doit être proche de F_0 . La fonction de répartition empirique est la fonction de \mathbb{R} dans $[0,1]$, qui vaut :

$$\hat{F}(x) = \begin{cases} 0 & \text{pour } x < X_1 \\ \dots & \\ \frac{i}{n} & \text{pour } X_i \leq x < X_{i+1} \\ \dots & \\ 1 & \text{pour } x \geq X_n \end{cases}$$

où les X_i sont les statistiques d'ordre de l'échantillon (valeurs de l'échantillon rangées par ordre croissant). En d'autres termes, $\hat{F}(x)$ est la proportion d'éléments de l'échantillon qui sont inférieurs ou égaux à x .

On mesure l'adéquation de la fonction de répartition empirique à la fonction F_0 par la distance de Kolmogorov-Smirnov, qui est la distance de la norme uniforme entre fonctions de répartitions. Pour la calculer, il suffit d'évaluer la différence entre \hat{F} et F_0 aux points X_i .

$$D_{KS}(F_0, \hat{F}) = \max_{i=1, \dots, n} (|F_0(X_{(i)}) - \frac{i}{n}|, |F_0(X_{(i)}) - \frac{i-1}{n}|)$$

Sous l'hypothèse \mathcal{H}_0 , la loi de la statistique $D_{KS}(F_0, \hat{F})$ ne dépend pas de F_0 , car les images de X_i par F_0 sont des variables aléatoires de loi $\mathcal{U}(0,1)$. Mais la fonction de répartition de $D_{KS}(F_0, \hat{F})$ n'a pas d'expression explicite simple et doit être calculée numériquement. Pour des échantillons de taille suffisante, on utilise le résultat asymptotique suivant :

Proposition 25

Sous l'hypothèse \mathcal{H}_0 , on a, pour tout $t \geq 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{H}_0}[\sqrt{n}D_{KS}(F_0, \hat{F}) \leq t] = 1 - 2 \sum_{k=1}^{+\infty} (-1)^{k+1} \exp(-2k^2t^2)$$

La série converge très rapidement. [1]

En pratique, pour $t > 0.56$, la somme des trois premiers termes donne déjà une approximation avec une erreur inférieure à 10^{-4} .

Si l'hypothèse \mathcal{H}_0 est fautive, $\sqrt{n}D_{KS}(F_0, \hat{F})$ tend vers $+\infty$ avec n . Le test est donc nécessairement unilatéral à droite (rejet des valeurs trop grandes). Supposons que la distance $D_{KS}(F_0, \hat{F})$ ait pris la valeur 0.047 pour un échantillon de taille $n = 1000$. La statistique $\sqrt{n}D_{KS}(F_0, \hat{F})$ vaut $t = 1.486$. La p-valeur correspondante est : $p(t) \simeq 2 \sum_{k=1}^3 (-1)^{k+1} \exp(-2k^2t^2) = 0.0241$.

Le test de Kolmogorov-Smirnov s'étend à la comparaison de deux fonctions de répartition empiriques, et permet alors de tester l'hypothèse que deux échantillons sont issus de la même loi. Bien d'autres tests d'ajustement peuvent être utilisés, comme ceux de Stephens, Anderson-Darling et Cramer-von Mises.

b)- Test du khi-deux de Pearson Le test du chi-deux concerne uniquement les lois discrètes, mais on peut l'utiliser aussi pour des échantillons continus regroupés en classes. Le modèle de base est toujours un échantillon (X_1, \dots, X_n) d'une loi inconnue. Les classes, notées c_1, \dots, c_r sont une partition de l'ensemble des valeurs possibles. L'hypothèse à tester porte sur les probabilités des classes, pour lesquelles on se donne des valeurs théoriques $P_0(c_1) \dots, P_0(c_r)$. $\mathcal{H}_0 : \mathbb{P}[X_i \in c_k] = P_0(c_k), \forall k = 1, \dots, r$.

Sous l'hypothèse \mathcal{H}_0 , la distribution empirique de l'échantillon sur les classes doit être proche de la distribution théorique. La distribution empirique est celle des fréquences de l'échantillon dans les classes. $\hat{P}(c_k) = \frac{1}{n} \sum_{i=1}^n 1_{c_k}(X_i)$. On mesure l'adéquation de la distribution empirique à la distribution théorique par la distance du chi-deux.

Définition 35

On appelle distance du chi-deux de \hat{P} par rapport à P_0 , et on note $D_{\chi^2}(P_0, \hat{P})$, la quantité :
$$D_{\chi^2}(P_0, \hat{P}) = \sum_{h=1}^r \frac{(P_0(c_h) - \hat{P}(c_h))^2}{P_0(c_h)}$$
.

La "distance" du chi-deux est donc une moyenne pondérée d'écart quadratiques entre les valeurs de P_0 et \hat{P} . Ce n'est pas une distance au sens usuel du terme, puisqu'elle n'est même pas symétrique.

La loi de probabilité de $D_{\chi^2}(P_0, \hat{P})$ n'a pas d'expression explicite en général. On utilise le résultat suivant. [16]

Théorème 15

Sous l'hypothèse \mathcal{H}_0 , la loi de la variable aléatoire $nD_{\chi^2}(P_0, \hat{P})$ converge quand n tend vers l'infini vers la loi du chi-deux de paramètre $r-1$. Si l'hypothèse \mathcal{H}_0 est fautive, alors la variable $nD_{\chi^2}(P_0, \hat{P})$ tend vers l'infini. C'est donc un test unilatéral à droite que l'on appliquera (rejet des trop grandes valeurs). [16]

3- Test d'adéquation des résidus

a)- L'approche de Box-Jenkins La méthodologie de BOX-JENKINS s'agit de choisir dans la vaste classe des modèles *ARIMA* le modèle reproduisant au mieux la série étudiée. La panoplie statistique habituelle peut s'appliquer : estimation des paramètres du modèle, tests d'hypothèse, analyse des résidus, identification d'observations atypiques, prévision. Généralement c'est le procédé étudié à cette section.

b)- Le test de Box-Pierce Le Test de Box-Pierce est un test statistique qui teste l'auto-corrélation d'ordre supérieur à 1. Il s'agit d'un test asymptotique qui n'a donc qu'une puissance très faible dans le cadre de petits échantillons. Les études ont montré que le test Box-Jenkins était meilleur que ce test pour toutes les tailles d'échantillons même les très petites.

5.4 Simulation sous le logiciel R

Pour la simulation on a utilisé le langage R et plus particulièrement les packages "tseries", "forecast" et "urca".

Exemple 5.4.1. La série température : représente les données de la température d'Oran de 2010 à 2015.

Année	Numéro du mois	Température	Année	Numéro du mois	Température
2010	1	24	2011	1	28
	2	33		2	21
	3	27		3	24
	4	32		4	29
	5	30		5	30
	6	34		6	35
	7	37		7	37
	8	37		8	37
	9	34		9	35
	10	33		10	30
	11	25		11	28
	12	28		12	22
2012	1	21	2013	1	29
	2	18		2	23
	3	22		3	25
	4	26		4	27
	5	31		5	29
	6	36		6	33
	7	35		7	34
	8	39		8	37
	9	36		9	32
	10	31		10	34
	11	29		11	29
	12	20		12	24
2014	13	24			
	2	28			
	3	28			
	4	30			
	5	32			
	6	36			
	7	38			
	8	36			
	9	36			
	10	36			
	11	30			
	12	23			

La série n'est pas stationnaire car le corrélogramme montre l'existence d'une saisonnalité. A l'aide de \mathcal{R} , plus précisément, en utilisant la commande "dd=diff(d, difference=1)" pour rendre la série stationnaire, d'après le test de Dickey-Fuller sous R avec la commande "adf.test(dd)" on obtient "p-value = 0.03387" inférieure à 0.05 donc la série est stationnaire, on passe à l'étape suivante, c'est-à-dire, la détermination des paramètres du modèle, toujours à l'aide de la commande "auto.arima(dd)" dans \mathcal{R} , on obtient :

$ARIMA(1, 0, 0)(1, 0, 0)[12]$ with zero mean.

	AR(1)	SAR(1)
Coefficients	-0.4960	0.6362
s.e	0.1489	0.1102

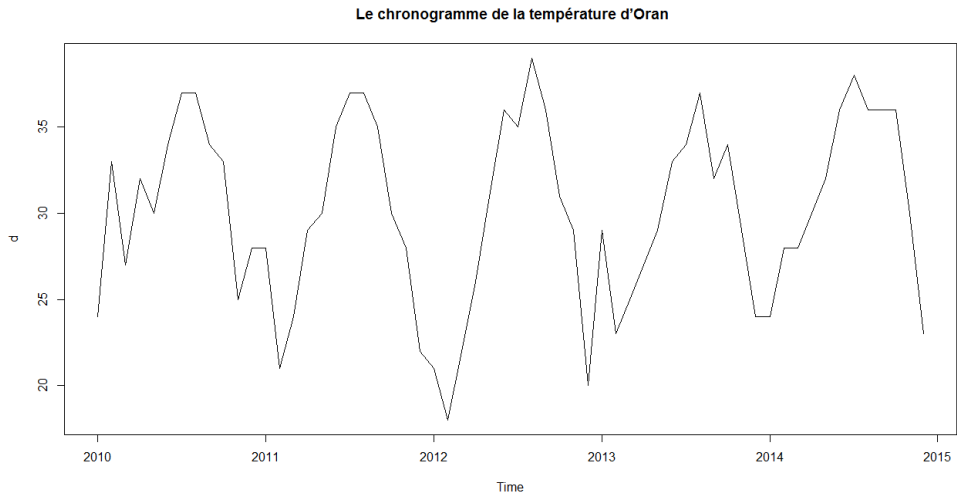


FIGURE 5.3 – Chronogramme de la température d'Oran

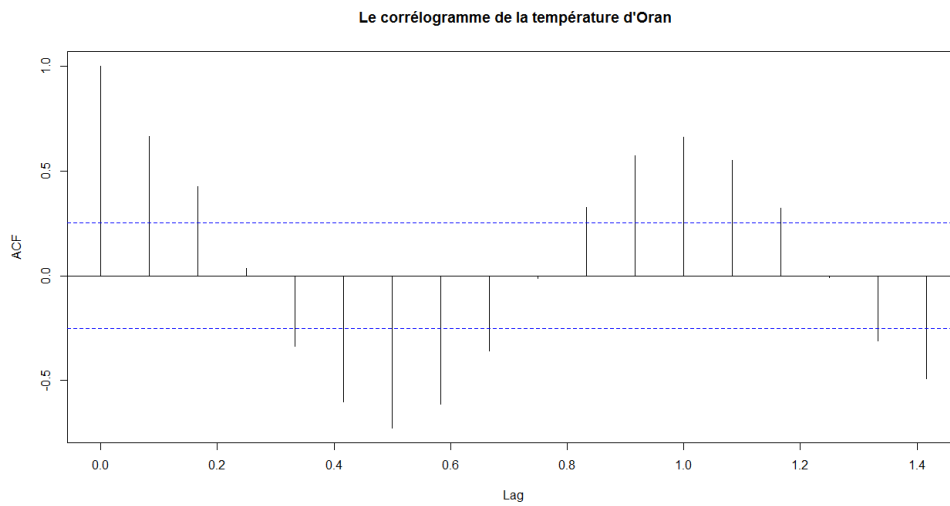


FIGURE 5.4 – Corrélogramme de la température d'Oran

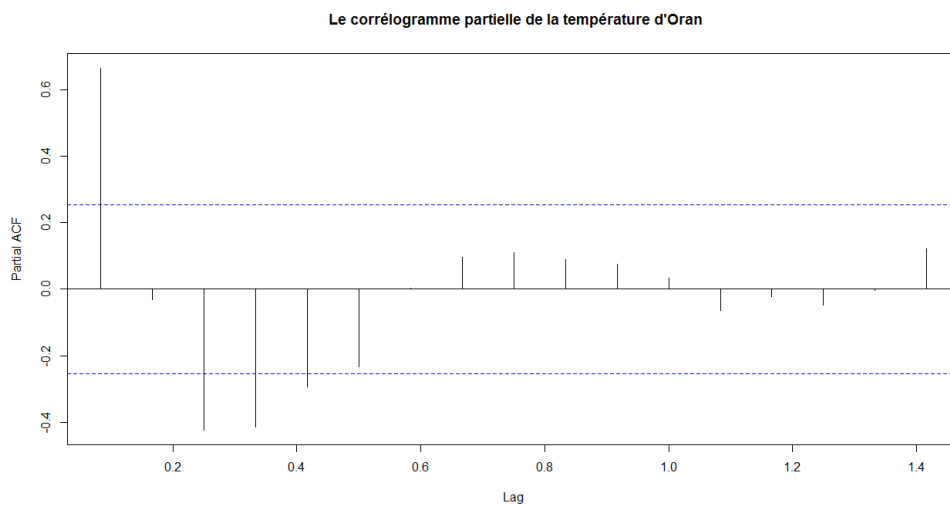


FIGURE 5.5 – Corrélogramme partielle de la température d'Oran

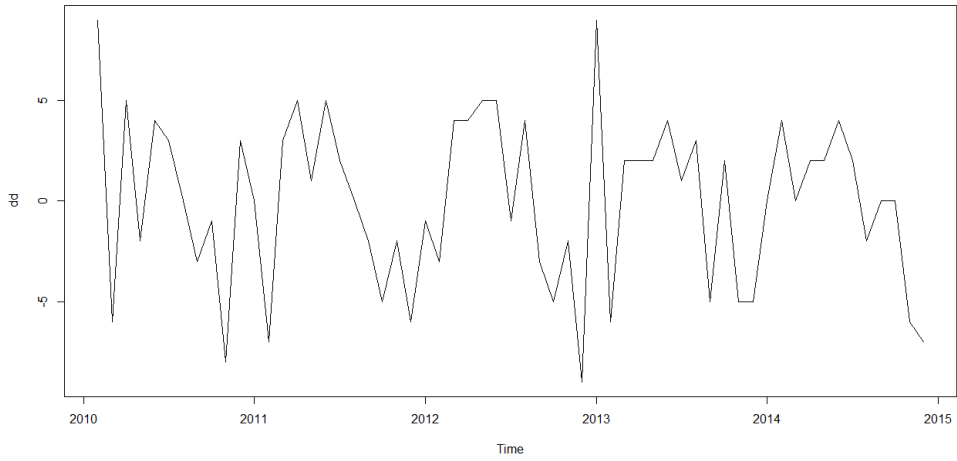


FIGURE 5.6 – Chronogramme de la température d’Oran après la différentiation

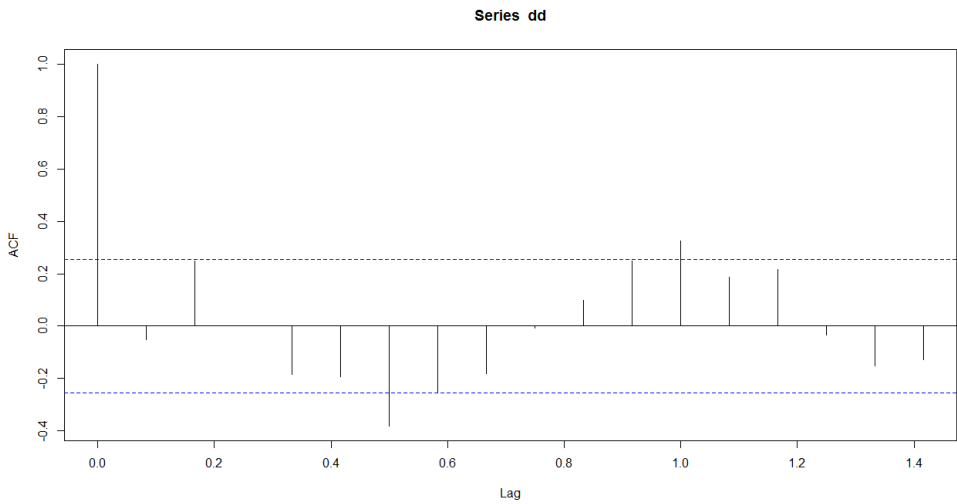


FIGURE 5.7 – Corrélogramme de la température d’Oran après la différentiation

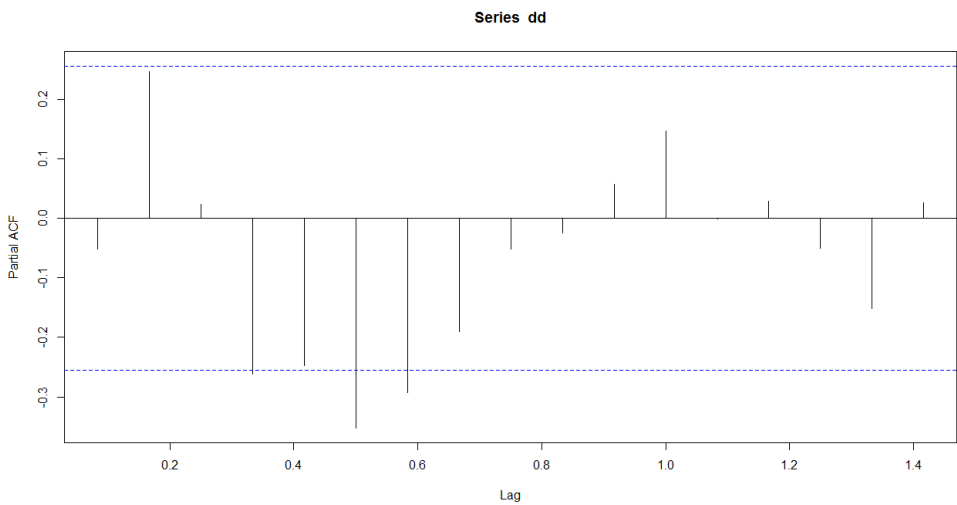


FIGURE 5.8 – Corrélogramme partielle de la température d’Oran après la différentiation

$AIC = 325.36$	$AIC_c = 325.8$	$BIC = 331.6$
----------------	-----------------	---------------

et $\hat{\sigma}^2 = 12.17$

On sait qu'un processus *SARIMA* s'écrit sous la forme :

$$\Phi_p(B)\Psi_P(B^s)\nabla^d\nabla_s^D X_t = \Theta_q(B)\Xi_Q(B^s)\eta_t$$

Donc le modèle retenu est :

$$\begin{aligned} \Phi_1(B)\Psi_1(B^{12})\nabla^0\nabla_{12}^0 X_t &= \Theta_0(B)\Xi_0(B^{12})\eta_t \\ \Leftrightarrow (1 - \phi_1 B)(1 - \psi_1 B^{12})X_t &= \eta_t \\ \Leftrightarrow (1 + 0.4960B)(1 - 0.6362B^{12})X_t &= \eta_t \\ \Leftrightarrow X_t &= -0.4960X_{t-1} + 0.6362X_{t-12} + 0.315552X_{t-13} + \eta_t \end{aligned}$$

Passons maintenant aux tests d'adéquation sur les résidus c'est-à-dire tester la bruité, en utilisons la commande "Box.test(dd)", on obtient alors "p-value = 0.6927" supérieure à 0.05, donc la partie résiduelle est bien un bruit blanc.

La commande "forecast(dd,h=60)" qui nous permet de prévoir et de connaître les valeurs de X_t jusqu'au l'horizon $h = 60$ c'est-à-dire jusqu'au l'année 2019, on obtient alors :

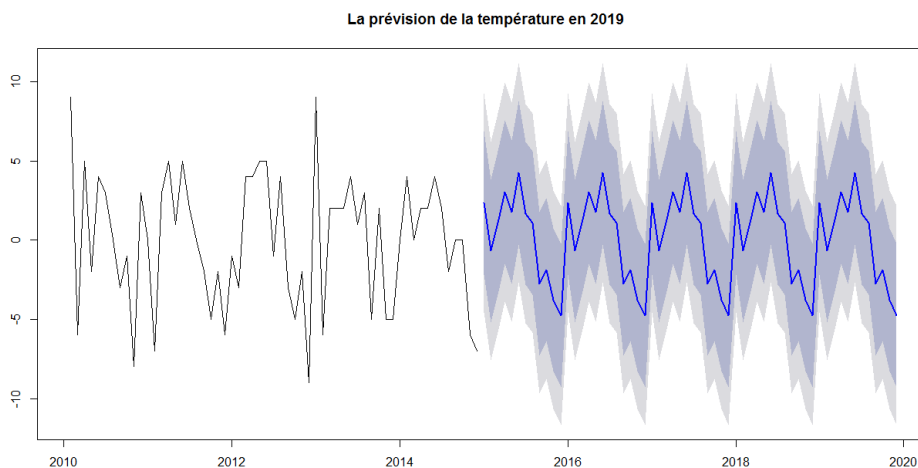


FIGURE 5.9 – La prévision de la température de l'année 2019 d'Oran

Exemple 5.4.2. *La série ciment : représente des données mensuelles des ventes du ciment au sein S.C.I.S de 2015 au premier trimestre du 2017.*

<i>Année</i>	<i>Numéro du mois</i>	<i>Ciment</i>	<i>Année</i>	<i>Numéro du mois</i>	<i>Ciment</i>
<i>2015</i>	<i>1</i>	<i>26040</i>	<i>2016</i>	<i>1</i>	<i>19460</i>
	<i>2</i>	<i>25260</i>		<i>2</i>	<i>10660</i>
	<i>3</i>	<i>33350</i>		<i>3</i>	<i>28950</i>
	<i>4</i>	<i>30750</i>		<i>4</i>	<i>22250</i>
	<i>5</i>	<i>11220</i>		<i>5</i>	<i>25110</i>
	<i>6</i>	<i>21750</i>		<i>6</i>	<i>16300</i>
	<i>7</i>	<i>27540</i>		<i>7</i>	<i>22130</i>
	<i>8</i>	<i>30520</i>		<i>8</i>	<i>31420</i>
	<i>9</i>	<i>28310</i>		<i>9</i>	<i>29960</i>
	<i>10</i>	<i>30580</i>		<i>10</i>	<i>32640</i>
	<i>11</i>	<i>22910</i>		<i>11</i>	<i>25030</i>
	<i>12</i>	<i>24890</i>		<i>12</i>	<i>25080</i>
<i>2017</i>	<i>1</i>	<i>20260</i>			
	<i>2</i>	<i>25100</i>			
	<i>3</i>	<i>25200</i>			

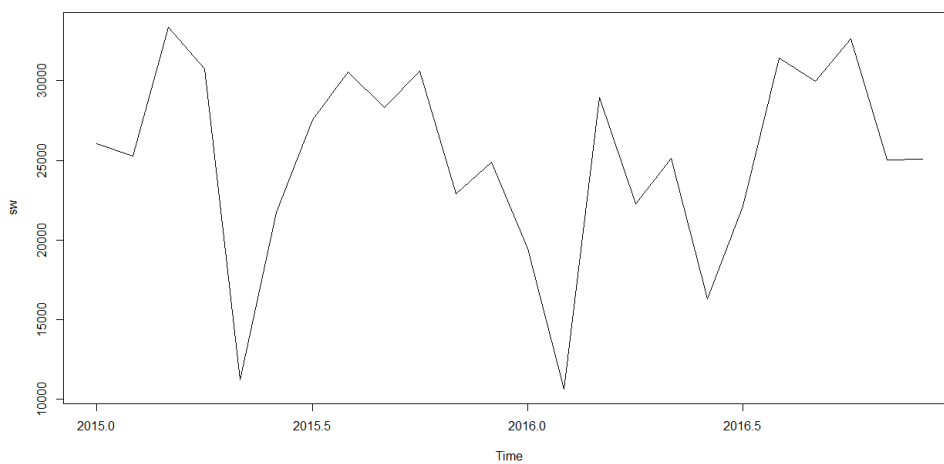


FIGURE 5.10 – Chronogramme de la série ciment

Le corrélogramme de la série ciment

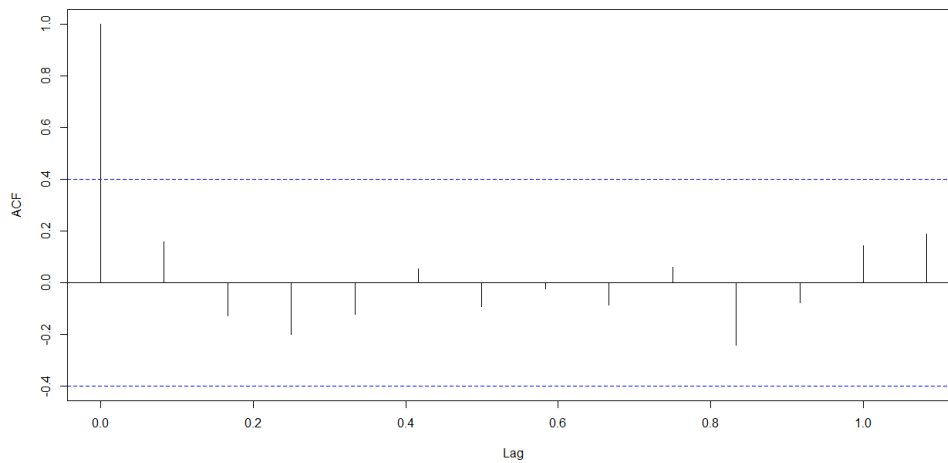


FIGURE 5.11 – Corrélrogramme de la série ciment

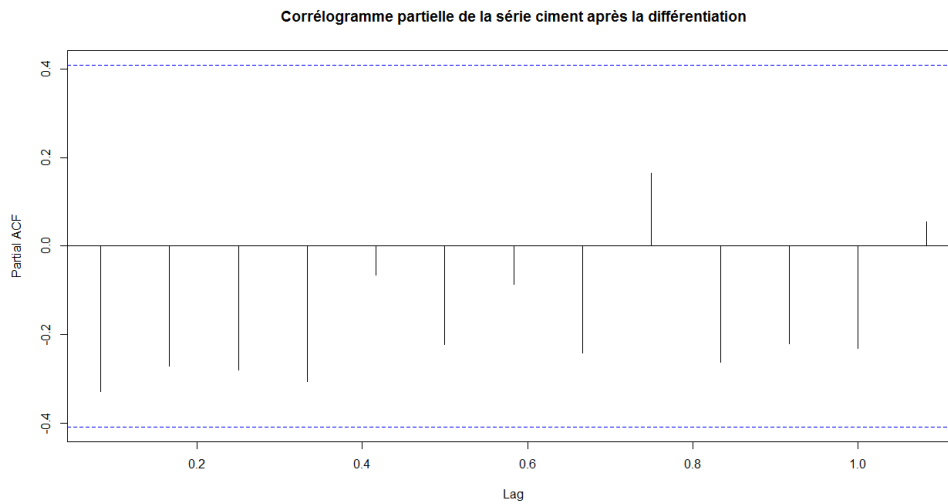


FIGURE 5.12 – Corrélogramme partielle de la série ciment

La série n'est pas stationnaire car le corrélogramme montre l'existence d'une saisonnalité, et d'après le test de Dickey-Fuller sous R avec la commande "adf.test(sw)", on trouve "p-value = 0.1899" supérieure à 0.05 donc la série est non stationnaire, donc on doit rendre la série stationnaire avec la commande "dsw=diff(sw, difference=1)", on obtient les graphes suivants :

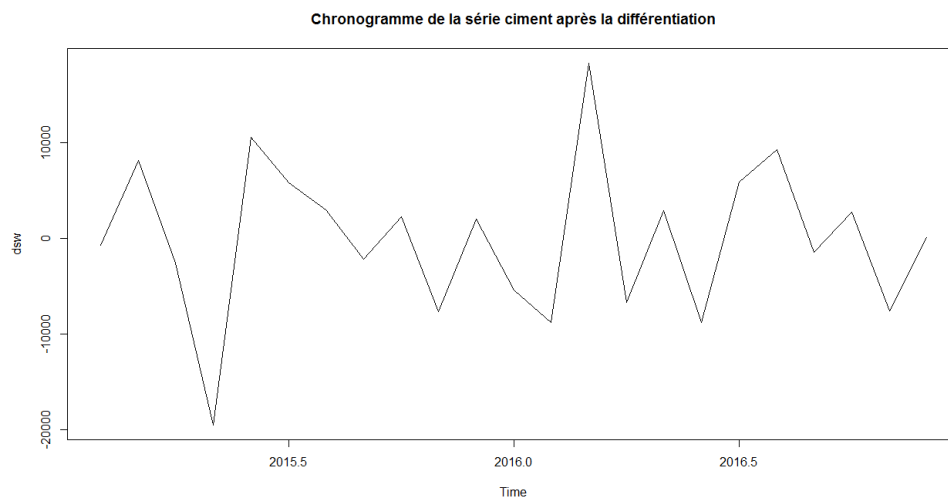


FIGURE 5.13 – Chronogramme de la série ciment après la différenciation

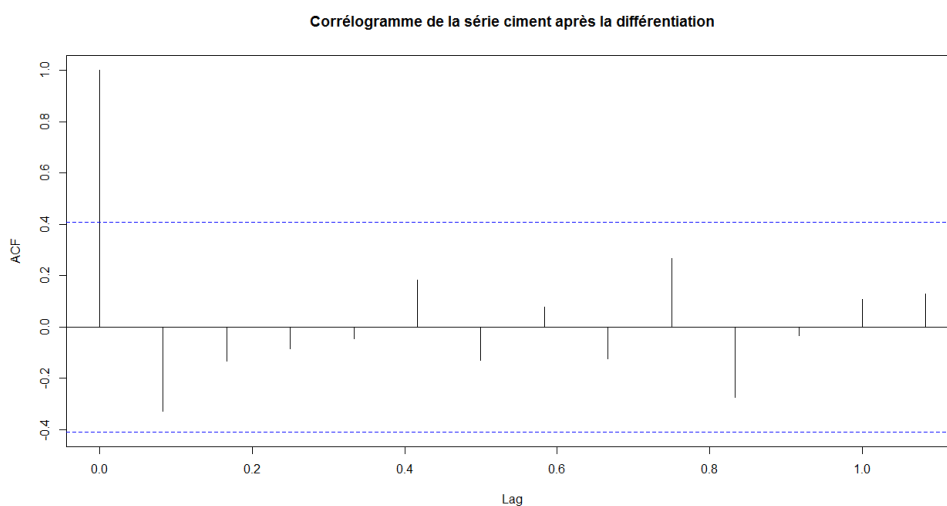


FIGURE 5.14 – Corrélogramme de la série ciment après la différenciation

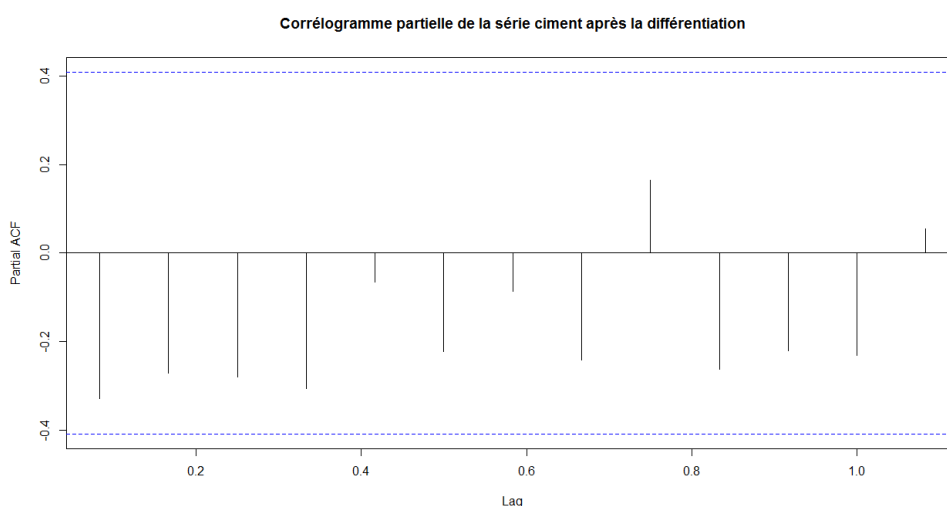


FIGURE 5.15 – Corrélogramme partielle de la série ciment après la différenciation

Exemple 5.4.3. *La série CAC40 : La modélisation des indices boursiers (bourse de Paris). On considère les prix de clôture de l'indice boursier CAC40 de 01/01/2019 jusqu'au 06/02/2019 (26 observations).*

<i>La date</i>	<i>Les prix de clôture</i>	<i>La date</i>	<i>Les prix de clôture</i>
02/01/2019	4689.39	21/01/2019	4867.78
03/01/2019	4611.48	22/01/2019	4847.53
04/01/2019	4737.12	23/01/2019	4840.38
07/01/2019	4719.17	24/01/2019	4871.96
08/01/2019	4773.27	25/01/2019	4925.82
09/01/2019	4813.58	28/01/2019	4888.58
10/01/2019	4805.66	29/01/2019	4928.18
11/01/2019	4781.34	30/01/2019	4974.76
14/01/2019	4762.75	31/01/2019	4992.72
15/01/2019	4786.17	01/02/2019	5019.26
16/01/2019	4810.74	04/02/2019	5000.19
17/01/2019	4794.37	05/02/2019	5083.34
18/01/2019	4875.93	06/02/2019	5079.05

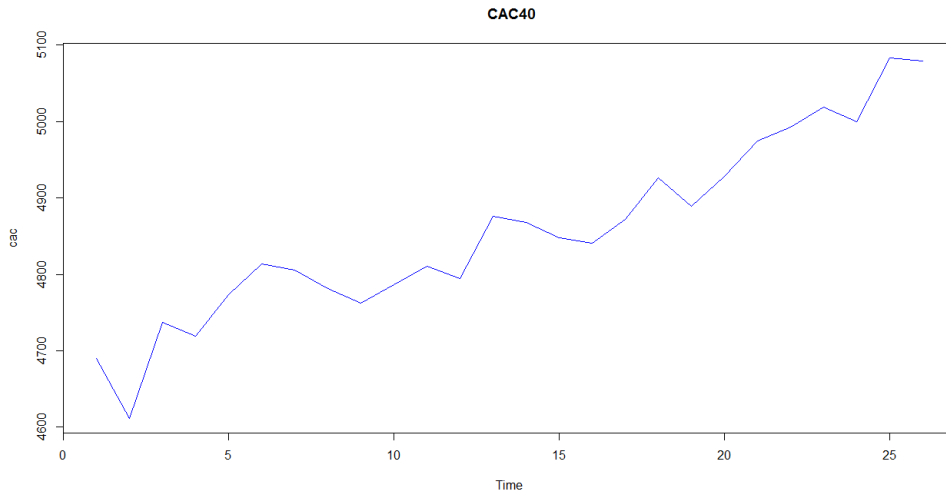


FIGURE 5.16 – Chronogramme des indices boursiers

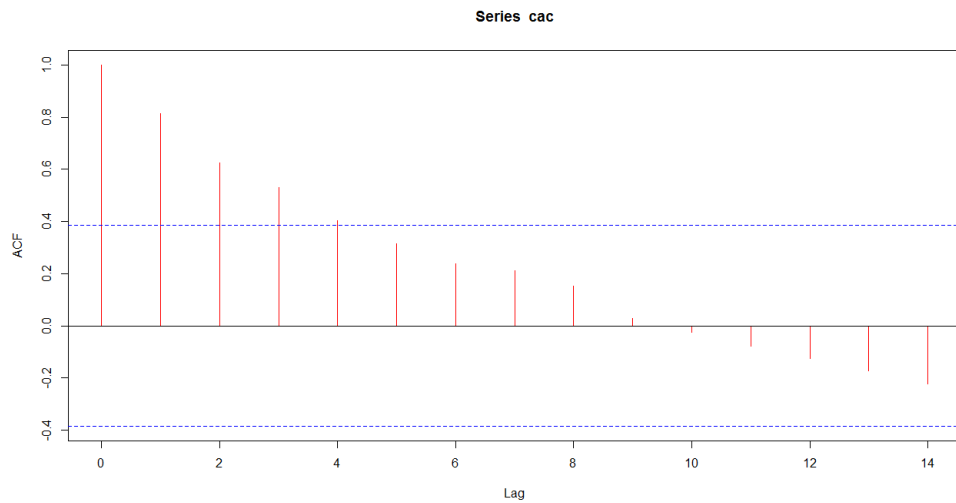


FIGURE 5.17 – Corrélogramme des indices boursiers

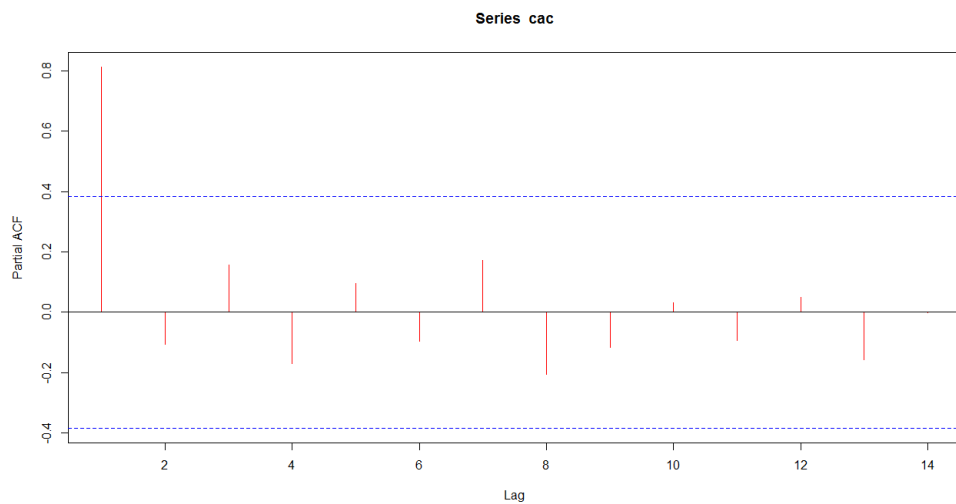


FIGURE 5.18 – Corrélogramme partielle des indices boursiers

Graphiquement, le corrélogramme montre que la série *CAC40* n'est pas stationnaire. Analytiquement, on va faire des tests de stationnarité, à l'aide de **R**, on obtient pour :

Nom du test	Dickey-Fuller	PP	KPSS
Commande dans R	adf.test(cac)	pp.test(cac)	kpss.test(cac)
p-value	0.674	0.02884	0.01
Commentaire	non stationnaire	non stationnaire	non stationnaire

Tous les tests confirment l'existence d'une racine unitaire, c'est-à-dire que la série des indices boursiers est non stationnaire.

Maintenant, on doit rendre la série *CAC40* stationnaire avec la commande "d=diff(cac)", répétant le test de stationnarité, par exemple le test KPSS donne p-value= 0.1, donc on rejette l'hypothèse nulle de stationnarité. L'étape suivante est la détermination des paramètres du modèle avec la commande "auto.arima(d)", on obtient :

ARIMA(0, 0, 1) with non-zero mean.

	<i>MA</i> (1)	mean
Coefficients	-0.6216	16.039
s.e	0.2012	3.080

<i>AIC</i> = 258.08	<i>AICc</i> = 259.22	<i>BIC</i> = 261.74
---------------------	----------------------	---------------------

et $\hat{\sigma}^2 = 1494$

On sait qu'un processus *ARMA* s'écrit sous la forme :

$$\Phi_p(B)X_t = \Theta_q(B)\eta_t$$

Donc le modèle retenu est :

$$\begin{aligned} \Phi_0(B)X_t &= \Theta_1(B)\eta_t \\ \Leftrightarrow X_t &= (1 - \theta_1 B)\eta_t \\ \Leftrightarrow X_t &= (1 + 0.6216B)\eta_t \\ \Leftrightarrow X_t &= \eta_t + 0.6216\eta_{t-1} \end{aligned}$$

Passons maintenant aux tests d'adéquation sur les résidus, en utilisons le test de Box-Pierce, avec la commande "Box.test(d)" on obtient p-value = 0.02489 supérieure à 0.01, donc la partie résiduelle est bien un bruit blanc, c'est-à-dire, avec 0.01 de risque on ne rejette pas l'hypothèse que les accroissements du *CAC40* est un bruit blanc.

Il reste à savoir la normalité de la série, on va faire des tests de normalité, à l'aide de **R**, et plus précisément "library(nortest)", on obtient pour :

Nom du test	Pearson chi-square	Kolmogorov-Smirnov
Commande dans R	pearson.test(d)	lillie.test(d)
p-value	0.2455	0.1412
Commentaire	loi normale	loi normale

Donc tous les tests confirment que les données suivent une loi de Gauss.

Passons au dernière étape, qui consiste à prédire les valeurs future de la série, en utilisant la commande "forecast()".

L'horizon h	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
27	5083.535	5032.858	5134.212	5006.032	5161.038
28	5099.492	5046.180	5152.804	5017.958	5181.026
29	5115.449	5059.624	5171.274	5030.072	5200.827
30	5131.406	5073.175	5189.638	5042.349	5220.463
31	5147.363	5086.820	5207.906	5054.770	5239.956
32	5163.320	5100.549	5226.092	5067.319	5259.321
33	5179.277	5114.353	5244.202	5079.984	5278.571
34	5195.234	5128.224	5262.244	5092.752	5297.717
35	5211.191	5142.158	5280.224	5105.614	5316.768
36	5227.148	5156.148	5298.148	5118.563	5335.733

Ce tableau représente les valeurs de la série *CAC40*, leurs prévisions et les bornes d'intervalles de confiance, on obtient le graphe suivant : En utilisant la prévision par lissage exponen-

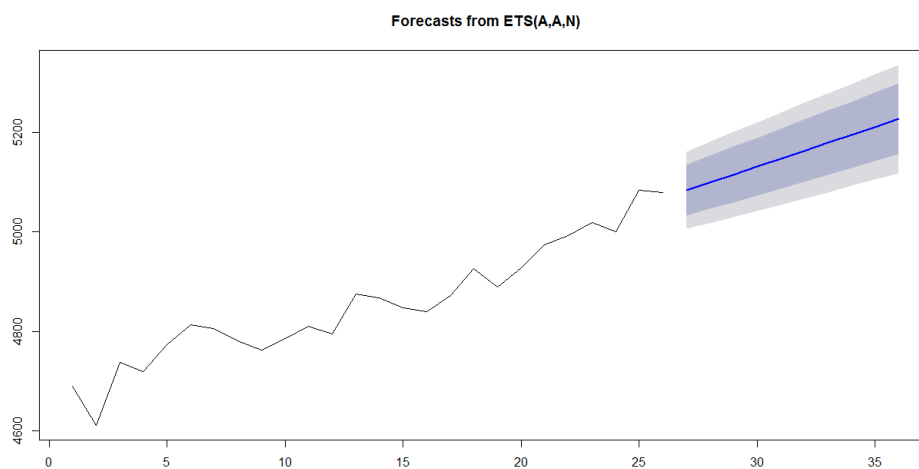


FIGURE 5.19 – La prévision des indices boursiers

tiel, on trace d'abord la série *CAC40* et prévision par lissage exponentiel avec la commande "p=HoltWinters(cac, gamma=FALSE)", on obtient alors $\alpha = 0.8655365$ et $\beta = 0.3834376$ et le graphe suivant : La figure montre que les prévisions par lissage sont assez bonnes.

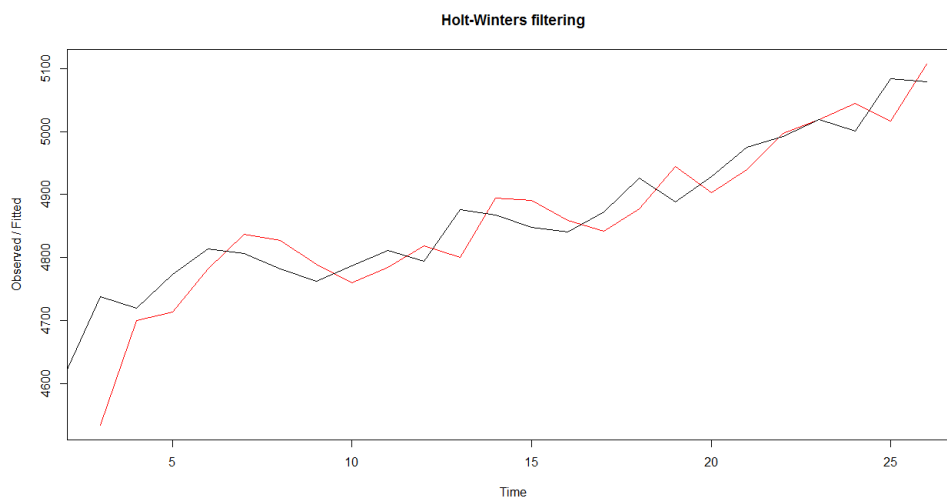


FIGURE 5.20 – la série *CAC40* + La prévision par lissage exponentiel des indices boursiers

Donc la commande "forecast(c=stats : :p, h=10)", on obtient :

L'horizon h	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
27	5106.299	5034.436	5178.161	4996.394	5216.203
28	5129.759	5017.648	5241.869	4958.300	5301.217
29	5153.218	4996.226	5310.211	4913.119	5393.318
30	5176.678	4970.437	5382.920	4861.259	5492.098
31	5200.138	4940.575	5459.702	4803.171	5597.106
32	5223.598	4906.902	5540.295	4739.253	5707.944
33	5247.058	4869.640	5624.476	4669.847	5824.269
34	5270.518	4828.983	5712.054	4595.249	5945.788
35	5293.978	4785.096	5802.861	4515.710	6072.247
36	5317.438	4738.123	5896.754	4431.452	6203.425

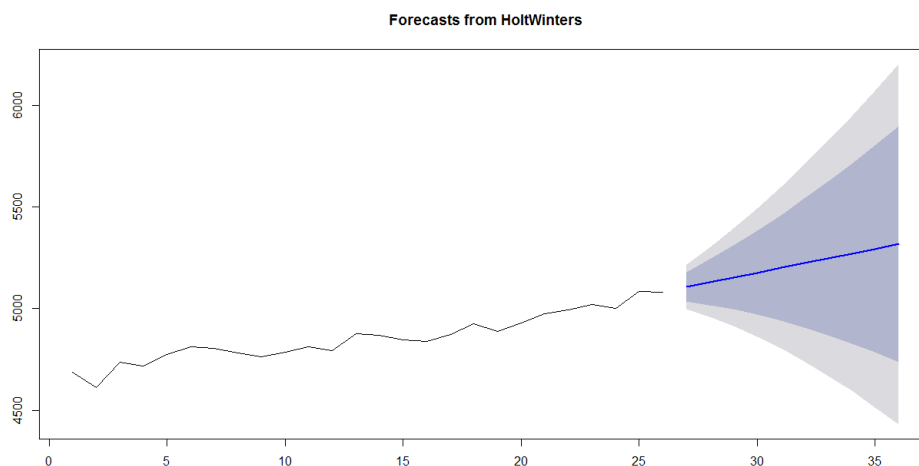


FIGURE 5.21 – La prévision par lissage exponentiel des indices boursiers

Maintenant nous allons comparer nos résultats avec ceux rapportés par CAC40 :

L'horizon h	les prix de clôture
27	5079,05
28	4985,56
29	4961,64

On peut conclure, d'une part que la méthode de Box-Jenkins donne des prévisions meilleures que le lissage exponentiel, d'autre part nous avons trouvé des prévisions assez proches de celles annoncées par CAC40.

CONCLUSION

Le domaine des séries chronologiques est en pleine expansion et les notions présentées dans ce mémoire, quoique largement utilisées, ne constituent qu'une petite partie des connaissances actuelles sur le sujet. Car on peut distinguer :

- Temps discret et temps continu.
- Stationnarité et non stationnarité.
- Scalaire et vectoriel.
- Linéaire et non linéaire.
- Domaine temporel et domaine des fréquences.
- Paramétrique et non paramétrique.
- Uni-variées et multivariées.
- La recherche de zones stationnaires : ce point a une grande importance pratique, car pour une réalisation particulière, une zone stationnaire est un sous-ensemble de la série à laquelle on peut appliquer un modèle unique, afin de pouvoir ensuite y appliquer des opérations particulières.
- La causalité : l'existence d'une influence entre les séries. En pratique, la plupart des séries chronologiques sont causales, car elles rendent compte de phénomènes qui sont causaux.

Comme nous avons vu dans l'introduction, un des objectifs, parmi d'autres, de l'étude des séries chronologiques est la prédiction, qui consiste à donner la meilleure valeur estimée du processus pour des instants futurs, compte tenu des valeurs observées précédemment.

Les différentes entreprises aussi bien dans le domaine de finance et autre, ont besoin des prévisions pour anticiper un suivi plus au moins raisonnable des résultats futures, Il est prouvé dans la littérature que l'étude des séries chronologiques est un moyen efficace pour répondre aux besoins de ces derniers.

LISTE DES NOTATIONS

- (Ω, A, \mathbb{P}) Espace probabilisé, page 35
 $(X'_{cvs,t})_t$ La série corrigée de la tendance, page 11
 $(X_{cvs,t})_t$ La série corrigée des variations saisonnières, page 11
 $(X_{SLP,t})_t$ La série lissée des prédictions, page 11
 ∇^k Opérateur de différenciation d'ordre k , page 17
 ϵ_t^* Innovation du processus à la date t , page 47
 $\gamma(h)$ La fonction d'auto covariance d'ordre h de X_t , page 56
 ϕ_i Les paramètres du processus $AR(p)$, page 53
 ψ_i Les pondérations du théorème du Wold, page 49
 $\rho(h)$ La fonction d'autocorrélation d'ordre h de X_t , page 56
 σ_t L'écart-type, page 14
 $\tau(h)$ La fonction d'autocorrélation partielle d'ordre h de X_t , page 44
 θ_i Les paramètres du processus $MA(q)$, page 63
 ϵ_t La composante résiduelle ou le bruit blanc, page 10
 $AR(p)$ Autoregressive Models, page 107
 $ARIMA$ Autoregressive Integrated Moving Average Models, page 84
 $ARIMA(p, d, q)$, page 84
 $ARMA(p, q)$ Autoregressive-Moving Average Models, page 70
 B^k Opérateur retard d'ordre k , page 18
 $Cov(t, X_t)$ La covariance entre les variables t et X_t , page 16
 DS Differency Stationary, page 81
 $e_n(h)$ L'erreur de prévision à l'horizon h d'une série chronologique $(X_t)_{t=1, \dots, n}$, page 59
 F^k Opérateur avance d'ordre k , page 18

-
- $F_{X_T}(x)$ La fonction de distribution cumulative de X_t , page 39
- $H(X)$ L'histoire de la série X, page 54
- $M_{m_1+m_2+1}$ La moyenne mobile d'ordre $m_1 + m_2 + 1$, page 18
- $MA(q)$ Moving Average Models, page 107
- S_t La composante saisonnière, page 4
- $SARIMA$ Seasonal Autoregressive Integrated Moving Average Models, page 85
- T_t La tendance ou trend, page 4
- TS Time Stationary, page 80
- $Var(t)$ La variance de la variable t, page 16
- $X_{T+1}^{(LES)}$ La prévision de la série à l'instant $T + 1$ par la méthode du lissage exponentiel simple, page 23
- $X_{t+h}^{(LED)}$ La prévision de la série à l'horizon h par la méthode du lissage exponentiel double, page 25
- X_{t+h}^{Holt} La prévision de la série à l'horizon h par la méthode du lissage exponentiel de Holt, page 27
- X_t Série chronologique, page 5
- X_t^* La série transformée, page 18

BIBLIOGRAPHIE

- [1] Adnane Majid Abd Arhman, *forecasting methods* , Chapitre 1-4, Edition, King Saud University, 2004.
- [2] Florin Avram, *Séries Temporelles avec R*, Edition, 2017.
- [3] Jérémie Bigot, *Séries chronologiques*, Edition, Université de Bordeaux, 2017.
- [4] M. Boutahar, *Séries temporelles : Estimation paramétrique et non paramétrique avec le logiciel R*, Edition,2007.
- [5] Arthur Charpentier, *Modèles de prévision : Séries temporelles*,Edition,2014.
- [6] Béchir Dola, *Problèmes économétriques d'analyse des séries temporelles à mémoire longue*, Edition, Université Sorbonne, 2013.
- [7] Jean-Yves Dauxois, *Introduction à l'étude des Séries Temporelles*, Edition, 2017.
- [8] M.Doiz, *Séries Temporelles linéaires*, Edition, ENSAE, 2005.
- [9] Dauphine-Eurisko, *Séries Temporelles*, Edition, 2017.
- [10] Christian GourierouxAlain Monfort, *Time Series and Dynamic Models*.
- [11] Laurent Gardes, *Séries chronologiques*, Edition, Université Paris-Sud, 2010.
- [12] Agnès Lagnoux, *Séries Chronologiques*, Edition, Université de Toulouse, 2015.
- [13] Agnès Lagnoux, *Renforcement Statistique Séries chronologiques*, Edition, Université de Toulouse, 2015.
- [14] Michel Prenat, *Séries chronologiques*, Edition, Université Paris-Sud, 2010.
- [15] Sylvain Rubenthaler, *Séries Chronologiques avec R*, Edition, Université Nice Sophia Antipolis, 2018.
- [16] R.S.Stoica, *Séries chronologiques*, Edition, Université Lille 1, 2009.