
Année: 2016

N° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

La méthodologie de Box Jenkins et applications

Mémoire présenté en vue de l'obtention du diplôme de

Master

Universitaire de Saida

Discipline : MATHÉMATIQUES

Spécialité : Analyse Stochastique Statistique des Processus et Applications

par

Mohamed Tahar El Amine Bouazza¹

Sous la direction de

Encadreur : M^{me} F.Mokhtari

Soutenue le 21/juin/2017 devant le jury composé de

T. Guendouzi	Maître de conférence Univ. Saida	Président
F. Mokhtari	Maître de conférence Univ. Saida	Directrice de mémoire
F. Benziadi	Maître assistant Univ. saida	Examineur
F. Maaref	Maître de conférence Univ. Saida	Examineur

1. e-mail : bouazzat@yahoo.fr

Dédicaces

À mes chers parents, pour leur soutien, leurs sacrifices et leurs prières tout au long de mes études.

À toute ma famille.

À tous mes compagnons de promotion et à tous ce que ma réussite leur tient à cœur.

Que ce travail soit le témoignage de ma gratitude et de mon profond respect.

Remerciements

J'adresse tout d'abord mes remerciements les plus sincères au Dr Mme F Mokhtari qui a accepté de diriger ce travail. Sa grande disponibilité et ses encouragements ont joué un rôle important dans la réalisation de ce mémoire.

Mes remerciements s'adressent également au président et aux membres du jury qui ont accepté de juger ce travail.

Qu'ils trouvent ici, le gage de mon infinie gratitude.

Résumé

On introduit dans ce travail la méthodologie de Box Jenkins. Cette méthode vise à formuler un modèle permettant de représenter une chronique avec comme finalité de prévoir des valeurs futures. De ce fait l'objet de cette méthode est de modéliser une série temporelle en fonction de ses valeurs passées en utilisant des modèles ARMA et ARIMA. On termine ce travail par quelques applications sur des données réelles (UKgas, ciment, température d'Oran, skirts).

Title : Introduction to the methodology of Box Jenkins and applications

Abstract

We introduce in this work the methodology of Box Jenkins. This method aims at formulating a model allowing to represent a chronicle with the aim of foreseeing future values. Therefore, the object of this method is to model a temporal series according to its past values using ARMA and ARIMA models. We finish this work with some applications on real data (UKgas, ciment, temperature of Oran, skirts).

Mots clés : Séries chronologiques, Stationnarité, processus (ARMA,ARIMA), tests, estimation, prévision.

Key words : Time series, Stationarity, process (ARMA,ARIMA), tests, estimation, prediction.

Table des matières

Introduction	4
1 Concepts des series temporelles	10
1.1 Vocabulaire et exemples	10
1.1.1 Objectifs principaux	12
1.2 Description d'une série chronologique	12
1.2.1 La tendance (ou trend)	12
1.2.2 La composante saisonnière (ou saisonnalité)	13
1.2.3 La composante résiduelle (bruit ou variable accidentelle)	13
1.3 Processus stationnaire du second ordre	14
1.3.1 Polynomes retard, avance et différence	18
1.4 Processus ARMA et ARIMA	21
1.4.1 Processus auto-régressifs d'ordre p (AR(p))	21
1.4.2 Processus moyenne mobile d'ordre q (MA(q))	23
1.4.3 Processus ARMA(p,q)	26
1.4.4 Processus ARIMA(p,d,q)	26
2 Méthodologie de Box-Jenkins	28
2.1 Transformation	29
2.2 Identification	30
2.2.1 Choix de p pour un processus AR(p)	30
2.2.2 Choix de q pour un processus MA(q)	32
2.2.3 Choix de p et q pour un processus ARMA(p,q)	33
2.3 Estimation des paramètres d'un modèle ARMA	35

2.3.1	Estimation de la partie autorégressive	36
2.3.2	Estimation de la partie moyenne mobile	36
2.4	Validation	37
2.4.1	Tests sur les résidus	37
2.4.2	Tests sur les paramètres	40
2.5	Choix d'un modèle parmi plusieurs	40
2.5.1	Critères d'information	41
2.5.2	Critères d'erreur	41
2.6	Prévision dans les <i>ARMA</i> et les <i>ARIMA</i>	42
2.7	Prévision dans un <i>AR(p)</i>	42
2.8	Prévision dans un <i>MA(q)</i>	44
2.9	Prévision pour un <i>ARMA(p, q)</i>	45
2.9.1	Prévision optimale	45
2.9.2	Prévision approchée	46
2.9.3	Equation de récurrence	46
2.10	Prévision pour un <i>ARIMA(p, d, q)</i>	46
2.10.1	Prévision optimale	47
2.10.2	Prévision approchée	47
2.10.3	Equation de recurrence	48
3	Simulation	49
3.1	Mise en oeuvre de la méthode de Box-Jenkins avec R	49
3.1.1	Simuler une série chronologique	49
3.1.2	Transformation	50
3.1.3	Identification	52
3.1.4	Estimation des paramètres	53
3.1.5	Tests d'adéquation et choix du modèle	54
3.1.6	Prévision	54
3.2	Prévision par lissage exponentiel	54
3.3	Exemples et comparaisons	57
3.3.1	Application sur des données simulées	57
3.3.2	Application sur des données réelles	62

Bibliographie

78

Introduction

L'analyse et la prédiction de série temporelle sont des défis scientifiques majeurs qui trouvent leurs applications dans des domaines aussi variés que la finance, la biologie, l'économie, la météorologie, etc.

L'étude de ces séries temporelles est utile lorsqu'on cherche à analyser, comprendre ou encore prévoir un phénomène évoluant dans le temps. Le tout est donc de tirer des conclusions à partir des séries observées.

L'étude d'un processus aléatoire à partir d'une série chronologique peut avoir généralement les objectifs suivants :

- Comprendre le passé : expliquer les variations observées
- Prédire les valeurs futures (proches)
- Etudier le lien avec d'autres séries.

Le premier chapitre a pour but de regrouper les outils indispensables pour l'étude de la méthodologie de Box-Jenkins et d'introduire la notion de série temporelle tel le processus "AR-MA-ARMA", cette présentation suppose qu'on définisse au préalable des notions essentielles à l'analyse de ces séries, et en particulier la notion de "stationnarité". Comme dans l'économie et dans plusieurs domaines les séries sont le plus souvent non stationnaires, on donne alors la définition des processus ARIMA (processus non stationnaires).

Le deuxième chapitre est particulièrement consacré à la méthodologie de Box-Jenkins. Cette méthode suggère une procédure à cinq étapes (transformation identification validation estimation prévision) qui seront bien détaillées.

Enfin dans le troisième chapitre on présente une étude de la méthodologie sur des données simulées et sur des données réelles en utilisant le langage R. cette mé-

thode est ensuite comparée avec une autre méthode non paramétrique appelée lissage exponentiel. La comparaison est faite en calculant l'erreur quadratique moyenne.

Chapitre 1

Concepts des series temporelles

La théorie des séries chronologiques est appliquée de nos jours dans de vastes domaines tels que l'économétrie, la médecine ou la démographie. On s'intéresse à l'évolution au cours du temps d'un phénomène, dans le but d'expliquer puis prévoir ce phénomène dans le futur.

1.1 Vocabulaire et exemples

Définition 1.1.1. Une suite d'observations d'une famille de variables aléatoires réelles notées $(X_t)_{t \in \Theta}$ indexée par le temps est appelée série chronologique (ou temporelle). Nous la noterons

$$(X_t)_{t \in \Theta} \text{ où } (X_t, t \in \Theta).$$

où l'ensemble Θ est appelé espace de temps qui peut être :

- **discret** (nombre de voyageurs SNCF quotidien, température maximale,...). Dans ce cas, $\Theta \subset \mathbb{Z}$.

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 $(X_t)_{t \in \Theta}$ où $\Theta \subset \mathbb{Z}$ (le plus souvent $\Theta = \mathbb{Z}$). Ainsi si h est l'intervalle de temps séparant deux observations et t_0 l'instant de la première observation, on a le schéma suivant

$$\begin{array}{c}
 t_0 \quad t_0 + h \quad \dots \quad t_0 + (T-1)h \\
 X_{t_0} \quad X_{t_0+h} \quad \dots \quad X_{t_0+(T-1)h} \\
 X_1 \quad X_2 \quad \dots \quad X_T
 \end{array}$$

- **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 $(X_t)_{t \in \Theta}$ où Θ est un intervalle de \mathbb{R} . Un tel processus est dit à temps continu.

Exemples 1.1.1. On peut songer par exemple à l'évolution du nombre de voyageurs utilisant le train, à l'accroissement relatif mensuel de l'indice des prix ou encore à l'occurrence d'un phénomène naturel, prenant les exemples :

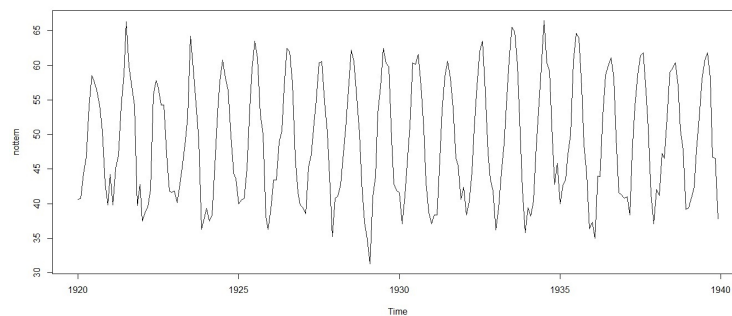


Figure 1.1 : La température mensuelle moyenne à Nottingham, 1920-1939

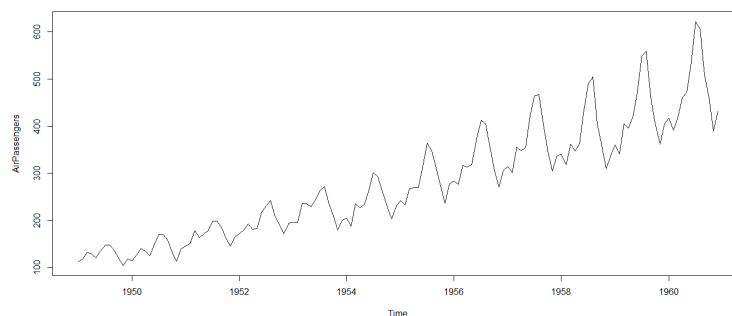


Figure 1.2 : le nombre de passagers (en milliers) dans les transports aériens

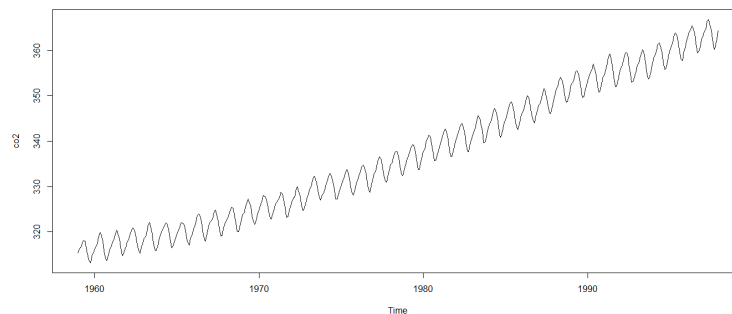


Figure 1.3 : la concentration atmosphérique de CO₂ à Maunaloa

1.1.1 Objectifs principaux

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} ($h = 1, 2, 3, \dots$) de la série chronologique à partir de ses valeurs observées jusqu'au temps T : X_1, X_2, \dots, X_T .

Il existe encore bien d'autres objectifs immédiats relatifs à l'étude des séries chronologiques. Par exemple, la détection des ruptures qui peuvent être résultantes (un changement de politique), la détermination de la causalité,...

1.2 Description d'une série chronologique

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

1.2.1 La tendance (ou trend)

Représente l'évolution à long terme de la série étudiée. Elle traduit le comportement "moyen" de la série.

Par exemple la figure 1.4 représente une tendance qui augmente d'une façon linéaire

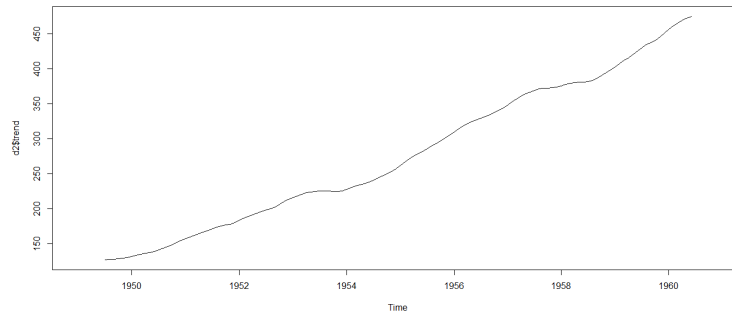


Figure 1.4

1.2.2 La composante saisonnière (ou saisonnalité)

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.

Par exemple la figure 1.5 représente des cycles réguliers au cours du temps et de même amplitude

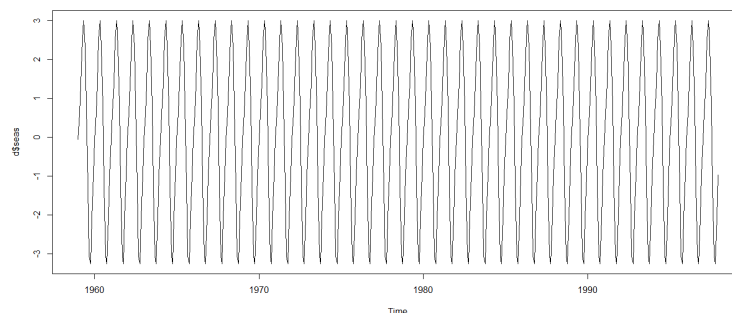


Figure 1.5

1.2.3 La composante résiduelle (bruit ou variable accidentelle)

Correspond à des fluctuations irrégulières, en général de faible intensité mais de nature aléatoire. Elle correspond aussi à des phénomènes accidentels (grèves, conditions

météorologique exceptionnelles, crash financier) peuvent nottament intervenir. Par exemple la figure 1.6 représente des fluctuation irrégulières

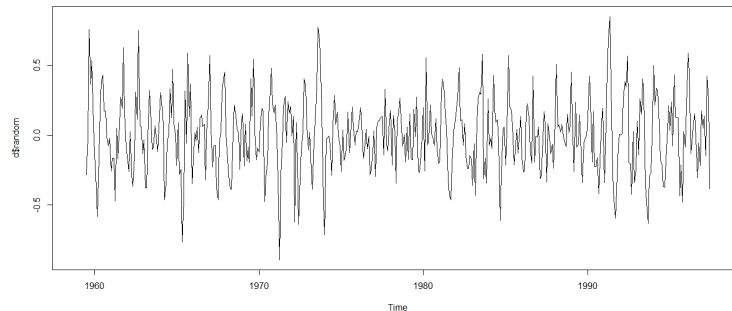


Figure 1.6

1.3 Processus stationnaire du second ordre

Une des grandes questions dans l'étude de séries temporelles (ou chronologiques) est de savoir si celles-ci suivent un processus stationnaire. On entend par là le fait que la structure du processus sous-jacent supposé évolue ou non avec le temps. Si la structure reste la même, le processus est dit alors stationnaire.

Dans toute la suite on considerera $(X_t)_{t \in \mathbb{Z}}$ et on supposera $X_t \in L^2(\Omega, \mathbf{A}, \mathbb{P}), \forall t \in \mathbb{Z}$.

Définition 1.3.1. (Stationnarité faible) $(X_t)_{t \in \mathbb{Z}}$ est un **processus stationnaire du second ordre** (ou un processus faiblement stationnaire) s'il vérifie :

(i) $\forall t \in \mathbb{Z}, \mathbb{E}(X_t) = m$

(ii) $\forall t \in \mathbb{Z}, \mathbb{V}(X_t) = \sigma^2 = \gamma(0)$

(iii) $\forall t \in \mathbb{Z}, \forall h \in \mathbb{Z}, \text{Cov}(X_t, X_{t+h})$ (ne dépend que de h)

$\gamma(h)$ est l'**auto-covariance** d'ordre h de X_t

Remarque 1.3.1. (1) Dans la suite, les processus stationnaires désignent les processus de la définition précédente.

(2) (iii) \Rightarrow (ii) : $h = 0$ et $\gamma(0) = \sigma^2$.

Exemple 1.3.1. (Processus stationnaire)

(1) **Bruit blanc faible** : $(\varepsilon_t)_{t \in \mathbb{Z}}$, si et seulement si :

$$\begin{aligned}\mathbb{E}(\varepsilon_t) &= 0, \forall t \in \mathbb{Z} \\ \mathbb{V}(\varepsilon_t) &= \sigma^2, \forall t \in \mathbb{Z} \\ \text{Cov}(\varepsilon_t, \varepsilon_s) &= 0, \text{ si } t \neq s\end{aligned}$$

On notera $\varepsilon \sim \mathcal{BB}(0, \sigma^2)$

(2) ε_t est un **bruit blanc fort** si et seulement si les ε_t sont i.i.d., $\mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{V}(\varepsilon_t) = \sigma^2$.

(3) Processus **moyenne mobile** d'ordre 1 noté $MA(1)$.

Soit $(X_t)_{t \in \mathbb{Z}}$ défini par : $\forall t \in \mathbb{Z}$, $X_t = \varepsilon_t - \theta \varepsilon_{t-1}$ où $\theta \in \mathbb{R}^*$ et $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$.

Alors $(X_t)_{t \in \mathbb{Z}}$ est un **processus stationnaire**.

Exemple 1.3.2. (Processus non stationnaire) (1) Marche aléatoire :

Soit $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$, $(X_t)_{t \in \mathbb{Z}}$ est une marche aléatoire sans dérive si et seulement si

(i) $X_t = X_{t-1} + \varepsilon_t$

(ii) $\text{Cov}(\varepsilon_t, X_{t-k}) = 0, \forall 0 < k \leq t$.

Même si on a la propriété $\mathbb{E}(X_t) = \mathbb{E}(X_{t-1}) \Rightarrow \mathbb{E}(X_t) = m, \forall t \in \mathbb{Z}$, $(X_t)_t$ n'est pas stationnaire :

$$\left. \begin{aligned} X_t &= X_{t-1} + \varepsilon_t \\ X_{t-1} &= X_{t-2} + \varepsilon_{t-1} \\ &\vdots \\ X_1 &= X_0 + \varepsilon_1 \end{aligned} \right\} \Rightarrow X_t = X_0 + \sum_{k=1}^t \varepsilon_k$$

D'où

$$\begin{aligned}\mathbb{V}(X_t) &= \mathbb{V}(X_0) + 2 \sum_{k=1}^t \text{Cov}(\varepsilon_k, X_0) + \mathbb{V}\left(\sum_{k=1}^t \varepsilon_k\right) \\ &= \mathbb{V}(X_0) + t\sigma^2\end{aligned}$$

Le processus n'est pas stationnaire en variance.

(2) (*Processus avec tendance ou saisonnalité*)

- $X_t = at + b + \varepsilon_t$ où $a, b \in \mathbb{R}$ et $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$. C'est un processus non stationnaire car $\mathbb{E}(X_t) = at + b$ dépend de t .
- $X_t = S_t + \varepsilon_t$ où S_t est une fonction périodique de période p . X_t n'est pas stationnaire car $\mathbb{E}(X_t) = S_t$

Définition 1.3.2. (*Fonction d'auto-covariance*) L'auto-covariance d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est définie par :

$$\begin{aligned} \gamma &: \mathbb{Z} \rightarrow \mathbb{R} \\ h &\mapsto \gamma(h) = \text{Cov}(X_t, X_{t-h}) \end{aligned}$$

Proposition 1.3.1. (i) γ est une fonction paire :

$$\gamma(-h) = \gamma(h) \quad \forall h$$

(ii) γ est de type positif : $\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n), \forall (a_1, \dots, a_n) \in \mathbb{R}^n$

$$\sum_{1 \leq i, j \leq n} a_i a_j \gamma(t_i - t_j) > 0$$

Preuve

(i) Parité :

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

(ii) Positivité :

$$\begin{aligned} \mathbb{V} \left(\sum_i a_i X_{t_i} \right) &= \text{Cov} \left(\sum_i a_i X_{t_i}, \sum_j a_j X_{t_j} \right) \\ &= \sum_{i,j} a_i a_j \text{Cov}(X_{t_i}, X_{t_j}) \\ &= \sum_{i,j} a_i a_j \gamma(t_i - t_j) \geq 0 \end{aligned}$$

On suppose toujours qu'il n'y a pas de relations linéaires entre les X_t . En effet, si on avait $\mathbb{V}\left(\sum_i a_i X_{t_i}\right) = 0$ alors $\sum_i a_i X_{t_i} = \text{constante}$ presque sûrement.

Définition 1.3.3. (Fonction d'auto-corrélation)

La fonction d'**auto-corrélation** d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est définie par :

$$\forall h \in \mathbb{Z}, \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \text{Corr}(X_t, X_{t+h})$$

Définition 1.3.4. (Fonction d'auto-corrélation partielle)

On appelle **auto-corrélation partielle** d'ordre p

$$r(p) = \text{Corr}(X_t - EL(X_t|X_{t-1}, \dots, X_{t-p+1}), X_{t-p} - EL(X_{t-p}|X_{t-1}, \dots, X_{t-p+1}))$$

$$= \frac{\text{Cov}(X_t - EL(X_t|X_{t-1}, \dots, X_{t-p+1}), X_{t-p} - EL(X_{t-p}|X_{t-1}, \dots, X_{t-p+1}))}{[\text{Var}(X_t - EL(X_t|X_{t-1}, \dots, X_{t-p+1}))\text{Var}(X_{t-p} - EL(X_{t-p}|X_{t-1}, \dots, X_{t-p+1}))]^{1/2}}$$

Définition 1.3.5. (Densité spectrale)

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de la forme :

$$X_t = m + \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \text{ où } (\varepsilon_t)_{t \in \mathbb{Z}} \sim \mathcal{BB} \text{ et } \sum_{j=0}^{+\infty} |a_j| < +\infty$$

Alors

- (i) $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$
- (ii) $\forall \omega \in [-\pi, \pi], \mathfrak{f}_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{i\omega h}$

\mathfrak{f}_X est la densité spectrale de $(X_t)_{t \in \mathbb{Z}}$

Définition 1.3.6. (Fonction d'auto-covariance inverse)

On appelle **auto-covariance inverse** d'ordre h de $(X_t)_{t \in \mathbb{Z}}$

$$\gamma_X^i(h) = \int_{[-\pi, \pi]} \frac{1}{\mathfrak{f}_X(\omega)} e^{i\omega h} d\omega$$

Définition 1.3.7. (Fonction d'auto-corrélation inverse) L'**auto-corrélation inverse** d'ordre h est alors définie comme

$$\rho_X^i(h) = \frac{\gamma_X^i(h)}{\gamma_X^i(0)}$$

Définition 1.3.8. (Estimateurs associés et lois limites) On considère un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ tel que pour tout $t \in \mathbb{Z}$, $\mathbb{E}(X_t) = m$. On cherche à estimer les grandeurs associées $\gamma_X(h)$, $\rho_x(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$, $r_X(h)$, $\mathfrak{f}_X(w)$, $\gamma_X^i(h)$ et $\rho_x^i(h)$ et ceci sachant qu'on observe X_1, \dots, X_T .

On prend comme estimateurs :

$$- \hat{m} = \frac{1}{T} \sum_{t=1}^T X_t = \bar{X}_t, \text{ moyenne empirique.}$$

$$- \hat{\gamma}_X(h) = \frac{1}{T-h} \sum_{t=h+1}^T (X_t - \bar{X}_t)(X_{t-h} - \bar{X}_t), \text{ auto-covariance empirique d'ordre } h.$$

$$- \hat{\rho}_x(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}.$$

- $\hat{r}(h)$ = la régression empirique (m.c.o.) de x_t sur 1, x_{t-1}, \dots, x_{t-h} .

$$- \hat{\mathfrak{f}}_X(w) = \frac{1}{2\pi} \sum_{h=-H}^H \hat{\gamma}_X(h) e^{iwh}, \text{ le problème ici étant que l'on voudrait un } H \text{ suffisamment}$$

grand mais prendre un H trop grand est risqué pour l'estimation de $\hat{\gamma}_X(h)$. On prend alors un estimateur corrigé :

$$\hat{\mathfrak{f}}_X(w) = \frac{1}{2\pi} \sum_{h=-H}^H \left(1 - \frac{|h|}{H+1}\right) \hat{\gamma}_X(h) e^{iwh}$$

On donne moins de poids aux $\hat{\gamma}_X(h) e^{iwh}$.

Si X_t est un processus stationnaire alors tous les estimateurs présentés ci-dessus sont convergents.

1.3.1 Polynomes retard, avance et différence

Définition 1.3.9. - (i) l'opérateur retard L (Lag) ou B (Backward) est défini sur la classe des processus stationnaires comme étant :

$L : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}}$ tel que $Y_t = X_{t-1}$

On note : $LX_t = X_{t-1}$.

– (ii) De la même façon, l' **opérateur avance** F (forward) correspond à

$$F : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}} \text{ tel que } Y_t = X_{t+1}$$

On note : $FX_t = X_{t+1}$

– (iii) l'**opérateur de différence** $\Delta = (1-L)$

l'**opérateur de différence d'ordre k** $\Delta_k = (1-L^k)$

Proposition 1.3.2. (i) $L^k = \underbrace{L \circ \dots \circ L}_{k \text{ fois}}$ vérifie $L^k X_t = X_{t-k}$

(ii) $F^k = \underbrace{F \circ \dots \circ F}_{k \text{ fois}}$ vérifie $F^k X_t = X_{t+k}$

Définition 1.3.10. Soit P un polynôme $P(z) = \sum_{k=0}^p a_k z^k$, $a_k \in \mathbb{R}$, on lui associe le

polynôme retard $P(L)$ défini comme suit :

$$P(L) = \sum_{k=0}^p a_k L^k$$

Et

$$P(L)X_t = \left(\sum_{k=0}^p a_k L^k \right) X_t = \sum_{k=0}^p a_k X_{t-k}$$

De façon similaire on obtient le **polynôme avance** $P(F)$:

$$P(F)X_t = \left(\sum_{k=0}^p a_k F^k \right) X_t = \sum_{k=0}^p a_k X_{t+k}$$

Définition 1.3.11. (**Séries en L (ou en F)**) (polynômes de degré infini)

Soit $A(z) = \sum_{k=0}^{+\infty} a_k z^k$ et $Y_t = A(L)X_t = \sum_{k=0}^{+\infty} a_k X_{t-k}$ alors $(Y_t)_{t \in \mathbb{Z}}$ est bien un processus

stationnaire car : $\sum_{k=0}^{+\infty} |a_k| < +\infty$

Proposition 1.3.3. *On suppose $\sum_{k=0}^{+\infty} |a_k| < +\infty$ et $\sum_{k=0}^{+\infty} |b_k| < +\infty$ et*

$$A(L) = \sum_{k=0}^{+\infty} a_k L^k \quad B(L) = \sum_{k=0}^{+\infty} b_k L^k$$

Alors

$$(i) \forall \alpha \in \mathbb{R}, \alpha A(L) = (\alpha A)(L) = \alpha \left(\sum_{k=0}^{+\infty} a_k L^k \right) = \sum_{k=0}^{+\infty} (\alpha a_k) L^k$$

$$(ii) A(L) + B(L) = (A + B)(L) = \left(\sum_{k=0}^{+\infty} a_k L^k \right) + \left(\sum_{k=0}^{+\infty} b_k L^k \right) = \sum_{k=0}^{+\infty} (a_k + b_k) L^k$$

$$(iii) A(L) \circ B(L) = (AB)(L) = B(L) \circ A(L) \text{ avec } (AB)(L) = (BA)(L) = C(L) = \sum_{k=0}^{+\infty} c_k L^k \text{ et } c_k = \sum_{j=0}^{+\infty} a_j b_{k-j}$$

Proposition 1.3.4. *Si X_t est un processus admettant une tendance m_t et une saisonnalité de période T alors :*

$$\Delta_T X_t = (m_t - m_{t-T}) + (\varepsilon_t - \varepsilon_{t-T})$$

(1) X_t est un processus désaisonnalisé. De plus, si la tendance du processus est linéaire, elle est également supprimée.

(2) Si X_t est un processus admettant une tendance polynomiale d'ordre k ie. :

$$X_t = \underbrace{\sum_{j=0}^k a_j t^j}_{m_t} + \varepsilon_t$$

alors le processus ΔX_t admet une tendance polynomiale d'ordre $k - 1$.

1.4 Processus ARMA et ARIMA

1.4.1 Processus auto-régressifs d'ordre p (AR(p))

Définition 1.4.1. $(X_t)_{t \in \mathbb{Z}}$ est un processus AR(p) si

- (X_t) est stationnaire
- (X_t) vérifie une équation $X_t = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$ avec $\varphi_p \neq 0$ et $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$

On note $\phi(L)X_t = \mu + \varepsilon_t$ où $\phi(L) = 1 - \varphi_1 L + \dots + \varphi_p L^p$

Exemple 1.4.1. $X_t \sim AR(1)$ i.e. $(1 - \rho L)X_t = \mu + \varepsilon_t$ où $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ et $|\rho| < 1$.

Propriétés des processus AR(p)

On suppose que $\phi(L)X_t = \mu + \varepsilon_t$ où

- les racines de ϕ sont de module strictement supérieur à 1
- ε_t suit un bruit blanc.

On peut se ramener ensuite à $\mu=0$ par centrage car

$$\phi(L)(X_t - m) = \varepsilon_t \text{ où } m = \frac{\mu}{\phi(1)} = \mathbb{E}(X_t).$$

On considère donc le cas où $\phi(L)X_t = \varepsilon_t$ (et $\mathbb{E}(X_t) = 0$).

Auto-covariance, auto-corrélations et équations de Yule-Walker

- **L'auto-covariance :**

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

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$$

Donc

$$X_t^2 = \varphi_1 X_t X_{t-1} + \dots + \varphi_p X_t X_{t-p} + X_t \varepsilon_t$$

et

$$\gamma(0) = \varphi_1 \gamma(1) + \dots + \varphi_p \gamma(p) + \mathbb{E}(X_t \varepsilon_t)$$

Or

$$\mathbb{E}(X_t \varepsilon_t) = \underbrace{\mathbb{E}[(\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}) \varepsilon_t]}_{=0 \text{ car } \varepsilon_t \perp \varepsilon_{t-i} \forall i > 0} + \mathbb{E}(\varepsilon_t^2)$$

D'où

$$\gamma(0) = \varphi_1 \gamma(1) + \dots + \varphi_p \gamma(p) + \sigma_\varepsilon^2$$

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

$$\begin{aligned} X_t X_{t-h} &= \varphi_1 X_{t-1} X_{t-h} + \dots + \varphi_p X_{t-p} X_{t-h} + \varepsilon_t X_{t-h} \\ \gamma(h) &= \varphi_1 \gamma(h-1) + \dots + \varphi_p \gamma(h-p) + \mathbb{E}(\varepsilon_t X_{t-h}) \end{aligned}$$

- **Les auto-corrélations :**

À partir de la relation de récurrence de $\gamma(h)$ on déduit celle sur $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$.

$$\rho(h) = \varphi_1 \rho(h-1) + \dots + \varphi_p \rho(h-p), \forall h \geq 0$$

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

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

$$\begin{aligned} 1 &= \varphi_1 \rho(1) + \dots + \varphi_p \rho(p) + \frac{\sigma_\varepsilon^2}{\gamma(0)} \\ \Rightarrow \gamma(0) &= \sigma_\varepsilon^2 \frac{1}{1 - (\varphi_1 \rho(1) + \dots + \varphi_p \rho(p))} \end{aligned}$$

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

$$\begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \rho(p-2) \\ \vdots & & \ddots & \vdots \\ \rho(p-1) & \dots & & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix}$$

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 $\varphi_1, \dots, \varphi_p$. En particulier elles donneront une estimation préliminaire de $\hat{\varphi}_1, \dots, \hat{\varphi}_p$ en fonction de $\hat{\rho}_T(1), \dots, \hat{\rho}_T(p)$.

$$\begin{aligned}
\rho(1) &= \varphi_1 + \varphi_2\rho(1) + \cdots + \varphi_p\rho(p-1) \\
\rho(p) &= \varphi_1\rho(p-1) + \cdots + \varphi_{p-1}\rho(1) + \varphi_p \\
\varphi_1 &= (1 - \varphi_2)\rho(1) - \cdots - \varphi_p\rho(p-1) \\
&\quad \dots \\
\varphi_1 &= \rho(p) - \varphi_1\rho(p-1) - \cdots + \varphi_{p-1}\rho(1)
\end{aligned}$$

On peut donc aussi obtenir $\rho(1), \dots, \rho(p)$ en fonction de $\varphi_1, \dots, \varphi_p$.

Proposition 1.4.1. *Si $X_t \sim AR(p)$ alors les $|\rho(h)|$ et les $\gamma(h)$ décroissent vers 0 exponentiellement avec h .*

Proposition 1.4.2. *Si $(X_t)_{t \in \mathbb{Z}} \sim AR(p)$ et si $\phi(L)X_t = \mu + \varepsilon_t$ est sa représentation canonique, alors :*

$$r(h) = \begin{cases} 0 & \text{si } h > p \\ \neq 0 & \text{sinon} \end{cases}$$

1.4.2 Processus moyenne mobile d'ordre q (MA(q))

Définition 1.4.2. $(X_t)_{t \in \mathbb{Z}} \sim MA(q)$ s'il existe $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ et $\theta_1, \dots, \theta_q$ tel que

$$\begin{aligned}
X_t &= m + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} \\
\text{On note } X_t &= m + \theta(L)\varepsilon_t \text{ où } \theta(L) = 1 - \theta_1L - \cdots - \theta_qL^q
\end{aligned}$$

Ecriture $AR(\infty)$ quand les racines de θ sont de module > 1

Sous ces hypothèses $\theta(L)$ est inversible et $\theta(L)^{-1} = \sum_{k=0}^{\infty} a_k L^k$ avec

$a_0=1$ et $\sum |a_k| < +\infty$. Il s'en suit que

$$X_t - m = \theta(L)\varepsilon_t \Leftrightarrow \theta(L)^{-1}(X_t - m) = \varepsilon_t \Leftrightarrow \theta(L)^{-1}X_t - \frac{m}{\theta(1)} = \varepsilon_t$$

Soit encore

$$\sum_{k=0}^{\infty} a_k X_{t-k} - \mu = \varepsilon_t \text{ où } \mu = \frac{m}{\theta(1)}$$

D'où la représentation canonique $\text{AR}(\infty)$

$$X_t = - \sum_1^{\infty} a_k X_{t-k} + \frac{m}{\theta(1)} + \varepsilon_t$$

Propriétés des processus $\text{MA}(q)$

On suppose que la représentation étudiée est la représentation canonique

$$\left\{ \begin{array}{l} X_t = m + \theta(L)\varepsilon_t \\ \text{toutes les racines de } \theta \text{ sont de module } > 1 \\ \theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q \\ \varepsilon_t \sim \mathcal{BB}(0, \sigma^2) \end{array} \right.$$

Proposition 1.4.3. *Auto-covariance* :

Sous les hypothèses précédentes

$$\gamma(h) = \left\{ \begin{array}{l} 0 \text{ si } |h| > q \\ -\theta_q \sigma_\varepsilon^2 \neq 0 \text{ si } |h| = q \\ \sigma_\varepsilon^2 (-\theta_h + \sum_{i=h+1}^q \theta_i \theta_{i-h}) \text{ si } 1 \leq |h| < q \\ \sigma_\varepsilon^2 (1 + \sum_{i=1}^q \theta_i^2) \text{ si } h = 0 \end{array} \right.$$

On en déduit $\rho(h)=0$ si $|h| > q$ et $\rho(q) \neq 0$.

Preuve

$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ après centrage.

- Si $h = 0$

$$\gamma(0) = \text{Var}(X_t) = \sigma_\varepsilon^2 (1 + \theta_1^2 + \dots + \theta_q^2) \neq 0$$

Car $\text{Cov}(\varepsilon_{t-j}, \varepsilon_{t-k}) = 0$ si $j \neq k$

- Si $h > q$

$$\begin{aligned}
\gamma(h) &= \text{Cov}(X_t, X_{t-h}) \\
&= \text{Cov}\left[\underbrace{(\varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q})}_{\varepsilon_{t-j}, j \in [0, q]}, \underbrace{(\varepsilon_{t-h} - \theta_1\varepsilon_{t-h-1} - \dots - \theta_q\varepsilon_{t-h-q})}_{\varepsilon_{t-k}, k \in [h, q+h]}\right] \\
&\Rightarrow \gamma(h) = 0 \text{ car } t - j \neq t - k \\
&\text{- Si } |h|=q \\
\gamma(q) &= \text{Cov}[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}), (\varepsilon_{t-q} - \theta_1\varepsilon_{t-q-1} - \dots - \theta_q\varepsilon_{t-2q})] \\
&= -\theta_q\sigma_\varepsilon^2 \\
&\text{- Si } 1 \leq |h| < q \\
\gamma(h) &= \text{Cov}[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}), (\varepsilon_{t-h} - \theta_1\varepsilon_{t-h-1} - \dots - \theta_q\varepsilon_{t-h-q})] \\
&= -\sum_{i=1}^q \theta_i \text{Cov}[(\varepsilon_{t-i}, \varepsilon_{t-h} - \sum_{k=1}^q \theta_k \varepsilon_{t-h-k}] \text{ car } \text{Cov}(\varepsilon_t, \varepsilon_{t-i})=0 \forall i>0 \\
&= -\theta_h\sigma_\varepsilon^2 + \sum_{i=1}^q \sum_{k=1}^q \theta_i\theta_k \underbrace{\text{Cov}(\varepsilon_{t-i}, \varepsilon_{t-h-k})}_{0 \text{ si } i \neq h+k} \\
&= -\theta_h\sigma_\varepsilon^2 + \left(\sum_{i=h+1}^q \theta_i\theta_{i-h} \right) \sigma_\varepsilon^2
\end{aligned}$$

Remarque 1.4.1. *On n'a pas de résultat particulier pour les auto-corrélations partielles.*

Proposition 1.4.4. $\rho^i(h)$ décroît exponentiellement avec h .

Théorème 1.4.1. *Soit ε_t un bruit blanc (au sens fort) tel que $\mathbb{E}[\varepsilon_t^4] = cte < +\infty$. Considérons $\phi(L)$ un filtre linéaire, μ une constante et X_t le processus stationnaire défini par $X_t = \mu + \phi(L)\varepsilon_t$. Alors :*

- Si X_t est un MA(q), $\sqrt{N}\hat{\rho}(k)$ converge en loi vers une gaussienne centrée de variance $1+2\sum_{u=1}^q \rho^2(u)$ pour tout $k > q$.
- Si X_t est un AR(p), $\sqrt{N}\hat{r}(k)$ converge en loi vers une gaussienne centrée de variance 1 pour tout $k > p$.

On déduit de ces comportements asymptotiques différents tests sur les paramètres du processus X_t .

1.4.3 Processus ARMA(p,q)

Définition 1.4.3. *Un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une représentation ARMA(p,q) canonique minimale s'il vérifie une équation :*

$$\phi(L)X_t = \mu + \theta(L)\varepsilon_t$$

où

- (i) $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$
- (ii) $\phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$, avec $\varphi_p \neq 0$
- (iii) $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$, avec $\theta_q \neq 0$
- (iv) $\phi(L)$ et $\theta(L)$ ont toutes leurs racines de module strictement supérieur à 1 (représentation canonique).
- (v) $\phi(L)$ et $\theta(L)$ n'ont pas de racines communes (représentation minimale).

Auto-covariance et auto-corrélation

- (i) Pour $h > q$, les $\gamma(h)$ et les $\rho(h)$ vérifient les équations de récurrence d'ordre p :

$$\begin{aligned} \gamma(h) - \varphi_1 \gamma(h-1) - \dots - \varphi_p \gamma(h-p) &= 0 \\ \rho(h) - \varphi_1 \rho(h-1) - \dots - \varphi_p \rho(h-p) &= 0 \end{aligned}$$

- (ii) Elles décroissent donc vers 0 exponentiellement avec h, pour $h > q$.

Un processus ARMA(p,q) vérifie alors (équations de Yule-Walker) :

$$\begin{pmatrix} \rho(p) & \cdots & \rho(q+p-1) \\ \vdots & \ddots & \vdots \\ \rho(q+p-1) & \cdots & \rho(q) \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} \rho(p+1) \\ \vdots \\ \rho(p+q) \end{pmatrix}$$

Quand ρ est connu ou estimé, on peut alors calculer les ϕ_j ou inversement, quand les φ_j sont connus, on calcule $\rho(q+1), \dots, \rho(q+p)$.

1.4.4 Processus ARIMA(p,d,q)

Ces processus sont non stationnaires dès que $d \leq 1$. Les séries économiques sont souvent non stationnaires.

Définition 1.4.4. $(X_t)_{t \in \mathbb{Z}}$ est un processus ARIMA(p,d,q) en représentation canonique minimale s'il vérifie une équation du type :

$$\forall t \geq 0, (1-L)^d \phi(L)X_t = \mu + \theta(L)\varepsilon_t$$

Et ceci avec :

(i) $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$

(ii) $\phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$ où $\varphi_p \neq 0$

$\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ où $\theta_q \neq 0$

(iii) $\phi(L)$ et $\theta(L)$ ont leurs racines de module > 1 et n'ont pas de racines communes

(iv) conditions initiales

$$Z = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q})$$

telles que $\text{Cov}(\varepsilon_t, Z) = 0$

Approximation auto-régressive d'un ARIMA(p,d,q)

Proposition 1.4.5. Avec les notations précédentes, $\exists A_t(L) = \sum_{j=0}^t a_j^t L^j$ et $a_t^0 = 1$,

$\exists \mu_0 \in \mathbb{R}$, $\exists h(t) \in \mathbb{R}^{p+d+q}$ et $\lim_{t \rightarrow +\infty} h(t) = 0$ tels que

$$A_t(L)X_t = \mu_0 + \varepsilon_t + h(t)'Z \Leftrightarrow X_t = \sum_{j=1}^t X_{t-j} + \varepsilon_t + h(t)'Z$$

Chapitre 2

Méthodologie de Box-Jenkins

Box and Jenkins (1976) ont promu une méthodologie consistant à modéliser les séries temporelles univariées au moyen des processus ARMA. Ces processus sont parcimonieux et constituent une bonne approximation de processus plus généraux pourvu que l'on se restreigne au cadre linéaire. Les modèles ARMA donnent souvent de bons résultats en prévision et ont bénéficié de la vague de scepticisme quant à l'intérêt des gros modèles économétriques.

Les étapes de la méthodologie

- La méthodologie de Box et Jenkins comporte essentiellement les étapes suivantes :
- Transformation des données afin de stabiliser la variance (log, sqrt,...) et différenciation des données pour les stationariser.
 - Visualisation des ACF et des PACF empiriques pour identifier les paramètres p et q appropriés.
 - Estimation des paramètres du(des) modèle(s) sélectionné(s).
 - Diagnostique et tests d'adéquation du modèle.
 - Prévision : La dernière étape consiste à la prévision des valeurs futures à travers le modèle retenu.

2.1 Transformation

Les modèles ARMA sont assez restrictifs dans la mesure où ils supposent que les observations sont stationnaires, que les paramètres sont constants et que les erreurs sont de variance constante. Pour se ramener à un tel cas, il faut souvent transformer les données. La transformation log permet de stabiliser la variance d'une série si la variance de la série originale croît avec les valeurs de celle-ci. Cette transformation permet aussi parfois de se rapprocher de la normalité.

Une série non stationnaire est une série qui en général croît avec le temps. On peut donc être tenté de retirer par régression un trend temporel de cette série. L'autre solution consiste à utiliser un filtre du type $(1 - L)$. On va alors différencier la série. Si l'on a tout d'abord pris le logarithme, la série transformée

$$(1 - L) \log y_t = \log y_t - \log y_{t-1}$$

représentera les accroissements en pourcentage de la série. Par exemple, au lieu de considérer un indice de prix, on modélisera un taux d'inflation. Un indice de prix croît en général toujours, ce qui fait qu'il n'est pas stationnaire. Un taux d'inflation aura lui beaucoup plus de chances d'être stationnaire. d représente le nombre de différenciation pour obtenir une série stationnaire, si une série représente une forte composante saisonnière, on peut éliminer celle-ci en différenciant la série non plus à l'ordre un, mais à l'ordre de la saisonnalité. Ainsi pour des séries trimestrielles, on pourra utiliser $(1 - L^4)$. Donc on peut définir un taux d'inflation annuel glissant par

$$(1 - L^4) \log p_t = \log p_t - \log p_{t-4}$$

Remarques 2.1.1. (1) *Il faut toutefois être prudent dans l'utilisation de ces transformations, car elles peuvent avoir des conséquences sur le processus étudié et introduire artificiellement des caractéristiques qui ne s'y trouvaient pas.*

(2) *Si les autocorrélations empiriques restent proche de 1 ou décroissent linéairement avec h alors le processus est sans doute non stationnaire.*

(3) *On utilise l'autocorrélogramme inverse pour étudier une sur-différenciation éventuelle.*

(4) *On peut aussi utiliser l'approche par des tests de racine unité.*

2.2 Identification

La méthode d'identification est essentiellement fondée sur l'analyse conjointe des auto-corrélations et des auto-corrélations partielles. Nous supposons le processus stationnaire ARMA(p,q). La méthode s'appuie sur les résultats suivants :

2.2.1 Choix de p pour un processus AR(p)

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

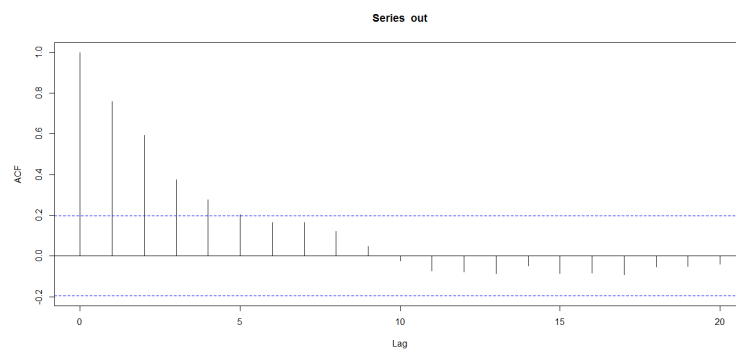


Figure 2.1 : Les auto-corrélations du processus $X_t = -0.8X_{t-1} + \varepsilon_t$

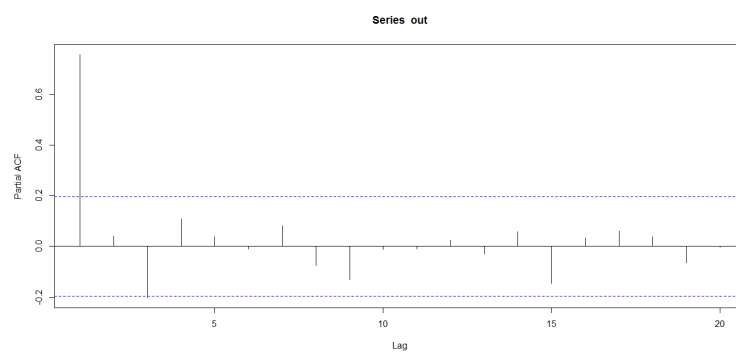


Figure 2.2 : Les auto-corrélations partielles du processus $X_t = -0.8X_{t-1} + \varepsilon_t$

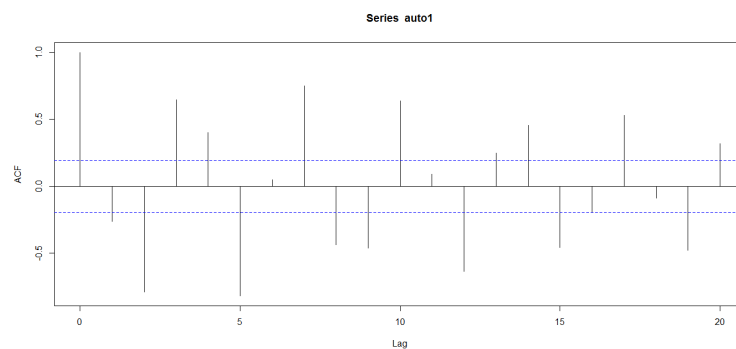


Figure 2.3 : Les auto-corrélations du processus $X_t = -0.5X_{t-1} - 0.9X_{t-2} + \varepsilon_t$

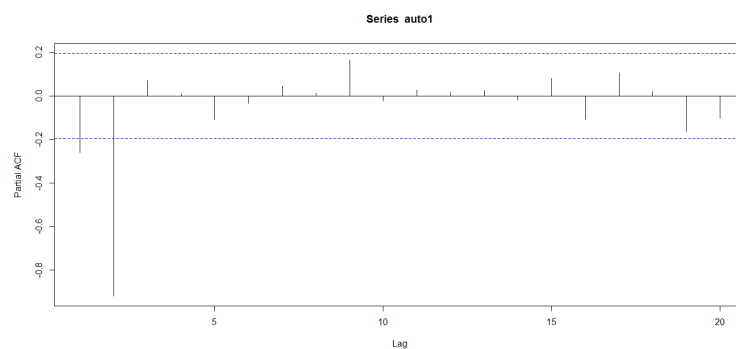


Figure 2.4 : Les auto-corrélations partielles du processus

$$X_t = -0.5X_{t-1} - 0.9X_{t-2} + \varepsilon_t$$

Test sur le paramètre p : Soit $X = (X_t, t \in \mathbb{Z})$ un processus $AR(p)$. D'après le théorème (1.4.1.) pour tout $r > p$, l'autocorrélation partielle empirique $r_X^{(N)}(r)$ est approximation une gaussienne centrée de variance $1/N$. On en déduit comme précédemment un test de $H_0 : X$ est un $AR(p)$, pour (presque) tout $r > p$ et $r < N/4$, on a

$$|r_X^{(N)}(r)| < 1,96\sqrt{\frac{1}{N}},$$

alors on accepte H_0 .

2.2.2 Choix de q pour un processus MA(q)

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

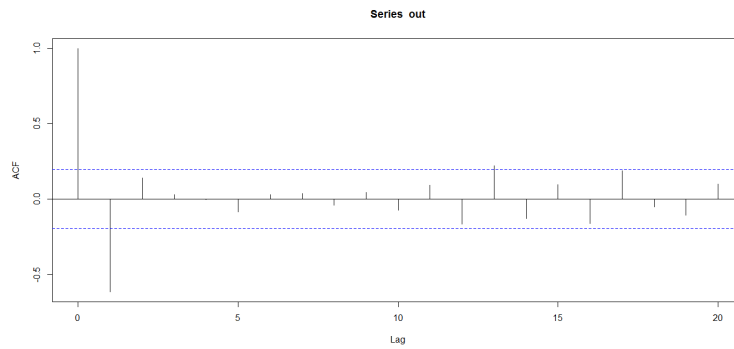


Figure 2.5 : Les auto-corrélations du processus $X_t = \varepsilon_t - 0.8\varepsilon_{t-1}$

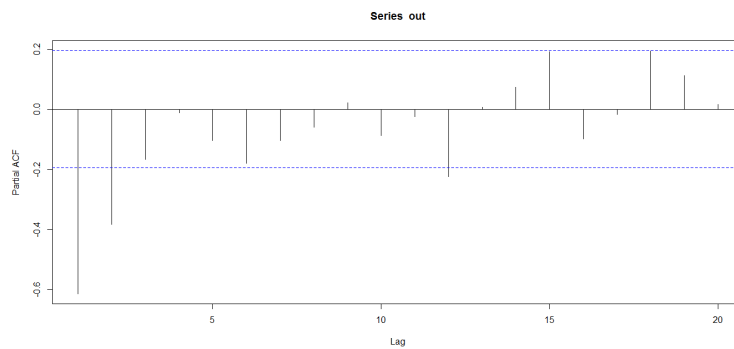


Figure 2.6 : Les auto-corrélations partielles du processus $X_t = \varepsilon_t - 0.8\varepsilon_{t-1}$

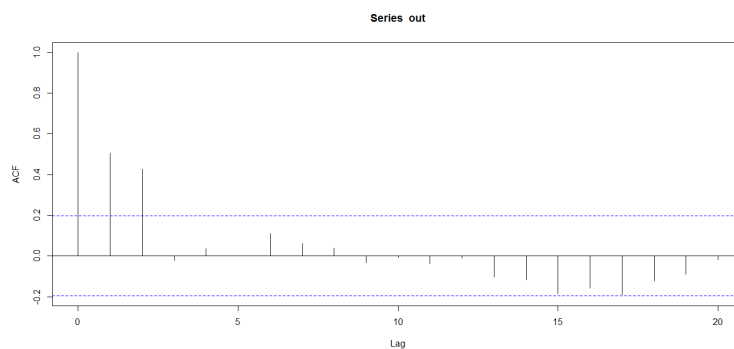


Figure 2.7 : Les auto-corrélations du processus $X_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.8\varepsilon_{t-2}$

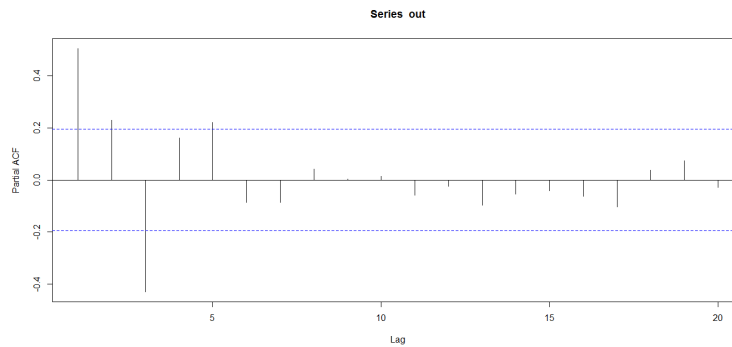


Figure 2.8 : Les auto-corrélations partielles du processus $X_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.8\varepsilon_{t-2}$

Test sur le paramètre q : Si $X = (X_t, t \in \mathbb{Z})$ un processus $MA(q)$. D'après le théorème (1.4.1.) pour tout $r > q$, alors l'autocorrélation empirique $\rho_X^{(N)}(r)$ est approximativement une gaussienne centrée de variance

$$\frac{1}{N} \left(1 + 2 \sum_{u=1}^q \rho_X(u)^2 \right) = \frac{1}{N} \left(1 + 2 \sum_{u=1}^r \rho_X(u)^2 \right)$$

En pratique, on fait donc l'approximation

$$\forall r > q, \sqrt{N} \frac{\rho_X^{(N)}(r)}{\sqrt{1 + 2 \sum_{u=1}^r \rho_X^{(N)}(u)^2}} \rightsquigarrow \mathcal{N}(0, 1).$$

On en déduit un test (pratique mais peu régoureux) de $H_0 : X$ est un $MA(q)$, si pour (presque) tout $r > q$ et $r < N/4$, on a

$$|\rho_X^{(N)}(r)| < 1,96 \sqrt{\frac{1 + 2 \sum_{u=1}^r \rho_X^{(N)}(u)^2}{N}},$$

alors on accepte H_0 . Si l'on se limite à $r < N/4$, c'est que l'estimation de $\rho_X^{(N)}(r)$ devient de moins en moins précise quand r croît vers N .

2.2.3 Choix de p et q pour un processus ARMA(p, q)

Les auto-corrélations partielles sont identiquement nulles au delà de l'ordre p et les auto-corrélations sont identiquement nulles au delà de l'ordre q .

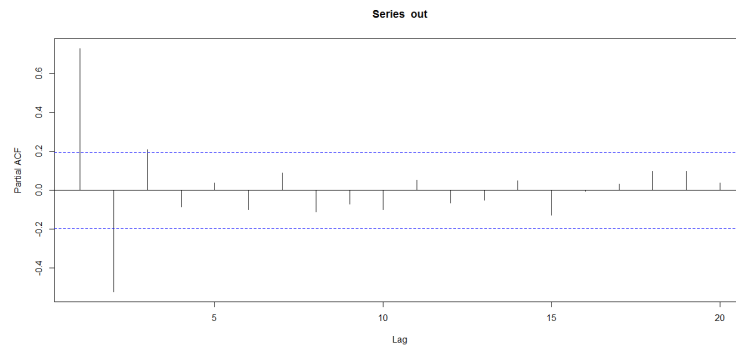


Figure 2.9 : Les auto-corrélations partielles de $X_t + 0.5X_{t-1} + 0.8X_{t-2} = \varepsilon_t + 0.9\varepsilon_{t-1} + 0.1\varepsilon_{t-2}$

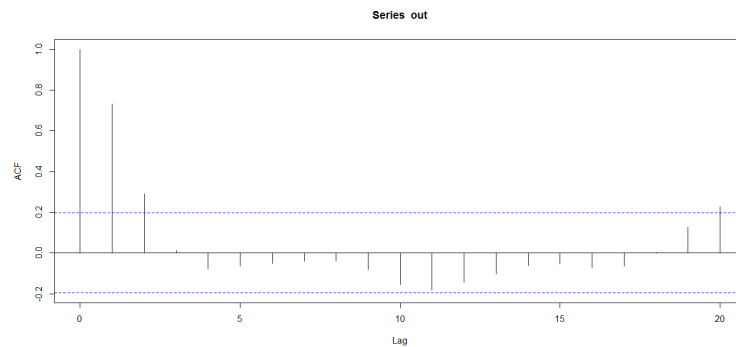


Figure 2.10 : Les auto-corrélations partielles de $X_t + 0.5X_{t-1} + 0.8X_{t-2} = \varepsilon_t + 0.9\varepsilon_{t-1} + 0.1\varepsilon_{t-2}$

Test sur les paramètres (p,q) : Si X est un processus $ARMA(p, q)$ vérifiant

$$\phi(L)(X_t) = \theta(L)(\varepsilon_t)$$

avec $\theta(z)$ et $\phi(z)$ sans racine commune, et ε_t bruit blanc d'innovation de variance σ^2 , alors il peut se mettre sous forme $MA(\infty)$

$$X_t = \varepsilon_t + \sum_{u \geq 1} \psi_u \varepsilon_{t-u}$$

ou $AR(\infty)$

$$X_t + \sum_{v \geq 1} \eta_v X_{t-v} = \varepsilon_t$$

Comme on a les inégalités

$$\mathbb{E} \left[\left(X_t - \varepsilon_t - \sum_{u=1}^{q'} \psi_u \varepsilon_{t-u} \right)^2 \right] \leq \sigma^2 \left(\sum_{u \geq q'+1} |\psi_u| \right)^2$$

$$\mathbb{E} \left[\left(X_t - \varepsilon_t + \sum_{v \geq 1}^{p'} \eta_v X_{t-v} \right)^2 \right] \leq \gamma_X(0) \left(\sum_{v \geq p'+1} |\eta_v| \right)^2$$

et que $\sum_{u \geq 0} |\psi_u|$ comme $\sum_{v \geq 0} |\eta_v|$ sont finis ; on en déduit qu'il est possible d'approcher d'aussi près qu'on le souhaite X par un processus de type $MA(q')$ ou $AR(p')$ avec $p' \leq p$ et $q' \leq q$. En pratique, on choisit pour valeurs de p' et q' celles que l'on détermine par les méthodes vues précédemment, en supposant le processus AR puis MA . Cela fournit une première modélisation de X comme processus $ARMA(p', q')$, avec en générale plusieurs couples (p', q') possibles.

Exemple 2.2.1. Si $\hat{r}_T(h)$ et $\hat{\rho}_T^i$ sont significativement non nuls pour $h \leq 3$ seulement et si $\hat{\rho}_T$ est significativement non nul pour $h \leq 4$ seulement, on est amené à estimer un $AR(3)$ et un $MA(4)$ ce qui pousse à choisir un $ARMA(p, q)$ avec $p \leq 3$ et $q \leq 4$.

Tableaux récapitulatifs des différentes situations :

	AR(p)
$\rho(h)$	décroît exponentiellement vers 0 avec h
$r(h)$	0 si $h > p$

	MA(q)
$\rho(h)$	0 si $h > q$
$r(h)$	décroît exponentiellement vers 0 avec h

	ARMA(p, q)
$\rho(h)$	0 si $h > q$
$r(h)$	0 si $h > p$

2.3 Estimation des paramètres d'un modèle ARMA

Une fois que l'on a décidé de chercher pour la série temporelle X_t un modèle $ARMA(p, q)$, reste à estimer les coefficients qui le caractérisent.

2.3.1 Estimation de la partie autorégressive

Si X est un processus ARMA(p, q), vérifiant $\phi(L)X = \theta(L)\varepsilon$, alors sa fonction d'autocovariance vérifie la relation de récurrence suivante :

$$\forall t \geq q + 1, \gamma_X(t) + \phi_1\gamma_X(t-1) + \dots + \phi_p\gamma_X(t-p) = 0$$

Les estimateurs de Yule Walker de ϕ_1, \dots, ϕ_p sont donc naturellement définis par

$$\begin{pmatrix} \hat{\rho}(q) & \dots & \hat{\rho}(q-p+1) \\ \vdots & \ddots & \vdots \\ \hat{\rho}(q+p-1) & \dots & \hat{\rho}(q) \end{pmatrix} \begin{pmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = - \begin{pmatrix} \hat{\rho}(q+1) \\ \vdots \\ \hat{\rho}(q+p) \end{pmatrix}$$

2.3.2 Estimation de la partie moyenne mobile

(1) **Méthode de Durbin** : Soit $Y_t = \phi(L)X_t = \theta(L)\varepsilon_t$ un processus MA(q) où $\theta(z) \neq 0$ pour $|z| < 1$. Alors :

$$\varepsilon_t = \sum_{k=1}^{+\infty} \psi_k z^k$$

Il existe p_0 , t.q. $\forall p \geq p_0$ l'équation récurrente :

$$\tilde{Y}_t = \varepsilon_t + \sum_{k=1}^p \psi_k X_{t-k}$$

définit un AR(p) causal et on a $E[|Y_t - \tilde{Y}_t|^2] \rightarrow 0$ où $p \rightarrow \infty$.

D'où l'idée d'approcher un processus MA inversible par un AR causal suffisamment long.

On choisit p suffisamment grand de telle façon que les observations puissent être considérées comme celles d'un AR(p). On en estime les coefficients $\hat{\psi}_1, \dots, \hat{\psi}_p$ et partant de l'identité, $\theta(z)\psi(z) = 1$, on calcule $\hat{\theta}_1, \dots, \hat{\theta}_q$ en fonction de $\hat{\psi}_1, \dots, \hat{\psi}_p$.

En conclusion :

- on choisit p d'autant plus grand que possible.
- on estime les p coefficients ψ_1, \dots, ψ_p , du prédicteur linéaire optimal.

– on estime les q coefficients du prédicteur linéaire optimal obtenu à partir des données : $1, \psi_1, \dots, \psi_p$.

(2) **Méthode du maximum de vraisemblance** : Le processus ARIMA s'écrit

$$\varepsilon_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Les ε_t sont indépendants et de loi normal $\mathcal{N}(0, \sigma^2)$, la fonction de densité de probabilité jointe de ε est donné par :

$$p(\varepsilon | \phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2\right]$$

la fonction de densité de probabilité jointe de X_t peut s'écrire :

$$L(\phi, \theta, \sigma^2 | X) = g_1(\phi, \theta, \sigma^2) \exp\left[-\frac{1}{2\sigma^2} S(\phi, \theta)\right]$$

où g_1 est la fonction des paramètres (ϕ, θ, σ^2) et

$$S(\phi, \theta) = \sum_{t=1-p-q}^n E^2(u_t | X)$$

$E(u_t | X) = E^2(u_t | X, \phi, \theta, \sigma^2)$ est l'esperance conditionnelle de u_t sur X, ϕ, θ et σ^2 et

$$u_t = \begin{cases} \varepsilon_t & t=1, 2, \dots, n \\ g_2(\varepsilon_*, X_*) & t \leq 0 \end{cases}$$

où g_2 est une fonction linéaire des valeurs initiales $a_* = (a_{1-p}, \dots, a_1, a_0)$ et $X_* = (X_{1-p}, \dots, X_{-1}, X_0)$

2.4 Validation

Il s'agit de vérifier notamment que les résidus du modèle *ARMA* estimé, résidus notés $\hat{\varepsilon}_t$, vérifient les propriétés requises pour que l'estimation soit valide, à savoir qu'ils suivent un processus \mathcal{BB} , non autocorrélé et de même variance, et qu'ils suivent une loi normale.

Si ces hypothèses ne sont pas rejetées, on peut alors mener des tests sur les paramètres.

2.4.1 Tests sur les résidus

Test de bruité :

(1) **Par l'étude de la fonction d'auto-corrélation empirique.** Lorsque n est assez grand, les auto-corrélations d'un bruit blanc sont approximativement indépendantes et de loi $\mathcal{N}(0, \frac{1}{n})$. Ainsi, 95 pour cent des auto-corrélations devraient se trouver dans l'intervalle $[\frac{-1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}]$, et en traçant les 40 premières auto-corrélations il ne devrait pas y en avoir plus de 2 voir 3 en dehors de ces limites. ie. les autocorrélations simples et partielles. Elles doivent être significativement nulles si les résidus sont un bruit blanc.

(2) **À l'aide du test de portemanteau.** Plutôt que de regarder si chaque auto-corrélation est dans les bornes de l'intervalle précédent, nous considérons la statistique définie par la somme des h premières auto-corrélations au carré

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

D'après la remarque précédente sur la normalité des auto-corrélations, la statistique Q suit une loi du χ^2 à h degrés de liberté. Il est donc possible de construire un test qui consistera à rejeter l'hypothèse nulle (la série est un bruit blanc) si Q est supérieur au quantile $\chi_{h,1-\alpha}^2$.

Ljung et Box (1978) ont amélioré ce test en considérant la statistique

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j}$$

(3) **Test de normalité.** Il s'agit de tester que les résidus estimés $\hat{\varepsilon}_t$ suivent une loi normale, c'est-à-dire ne présentent pas d'asymétrie (Skewness) ni d'aplatissement (kurtosis).

Le coefficient de Skewness est donné par :

$$\beta_1^{1/2} = \frac{\mu_3}{\mu_2^{3/2}}$$

et le coefficient de kurtosis est donné par :

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

où

$$\mu_k = \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^k$$

est le moment centré d'ordre k de la variable $\hat{\varepsilon}_t$.

Si la distribution est normale et le nombre d'observations est grand, alors :

$$\beta_1^{1/2} \sim \mathcal{N}(0, \sqrt{6/T}) \quad \text{et} \quad \beta_2 \sim \mathcal{N}(3, \sqrt{24/T})$$

On construit alors les statistiques :

$$\nu_1 = \frac{\beta_1^{1/2}}{\sqrt{6/T}} \quad \text{et} \quad \nu_2 = \frac{\beta_2 - 3}{\sqrt{24/T}}$$

qui suivent chacune une $\mathcal{N}(0,1)$. Le test de Jarque Bera permet de tester simultanément l'absence d'asymétrie et l'absence d'aplatissement. La statistique de test est donnée par :

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

Cette statistique suit, sous l'hypothèse nulle de normalité, une loi du χ_2^2 .

(4) **Test de stationnarité (test ADF)**. Dickey et Fuller (1979) proposent un test détectant la non stationnarité d'une série temporelle. En considérant une série temporelle notée y_t , le test de Dickey et Fuller (DF) est un test de racine unitaire qui estime l'hypothèse nulle de racine unitaire (ou de non stationnarité). Le test ADF consiste alors à estimer les modèles en introduisant des variables retardées. Par exemple, le modèle sans constante ni dérive temporelle est le modèle suivant :

$$\Delta y_t = \phi y_{t-1} + \sum_{j=2}^p \beta_j \cdot \Delta y_{t-j+1} + \varepsilon_t$$

L'hypothèse nulle du test ADF est l'hypothèse de racine unitaire (ou non stationnarité) de la variable y_t soit l'hypothèse $H_0 : \phi=0$.

Le test ADF consiste à comparer la valeur estimée du t de student associé au paramètre ϕ aux valeurs tabulées de cette statistique. L'hypothèse nulle est rejetée

au seuil de 5 % lorsque la valeur observée du t de student est inférieur à la valeur critique tabulée où $t < ADF_{0,5}$

Remarque 2.4.1. *Supposons que le processus est :*

$$\phi(L)X_t = \theta(L)\varepsilon_t$$

Cependant le processus ε_t n'est pas un bruit blanc mais un ARMA(\tilde{p}, \tilde{q}) donc :

$$\tilde{\phi}(L)\varepsilon_t = \tilde{\theta}(L)a_t, \forall t \in \mathbb{Z}$$

$$\varepsilon_t = \tilde{\phi}^{-1}(L)\tilde{\theta}(L)a_t$$

Quelques manipulations donnent :

$$\phi(L)X_t = \theta(L)\varepsilon_t$$

$$\phi(L)X_t = \theta(L)\tilde{\phi}^{-1}(L)\tilde{\theta}(L)a_t$$

Donc

$$\tilde{\phi}\phi(L)X_t = \theta(L)\tilde{\theta}(L)a_t$$

Ou encore

$$\phi^*(L)X_t = \theta^*(L)a_t$$

Ce modèle n'est rien qu'un ARMA(p^, q^*).*

2.4.2 Tests sur les paramètres

On vérifie tout d'abord que les racines des polynômes AR et MA ne sont pas égales à 1. Il s'agit de tester $H_0 : p' = p-1$ et $q' = q$ ou $H_0 : p' = p$ et $q' = q-1$.

En d'autres termes, on teste : $H_0 = ARMA(p-1, q)$ vs $H_1 : ARMA(p, q)$.

On teste la significativité des retards du modèle $ARMA$ par des tests de Student.

2.5 Choix d'un modèle parmi plusieurs

Comme on l'a vu précédemment l'étude des auto-covariances, auto-corrélations et auto-corrélations partielles peut conduire à certaines hypothèses sur la nature du modèle. Une fois quelques modèles choisis, et leur paramètres estimés, on utilise les critères suivants pour choisir le meilleur modèle :

2.5.1 Critères d'information

Pour choisir le modèle qui effectue le meilleur compromis entre :

- ajustement à la série de données.
- complexité du modèle.

Il est en effet très important de prendre en compte ce compromis, car si on ne s'intéressait qu'à coller au mieux aux données, on serait tenté de choisir un modèle ARMA avec un très grand nombre de paramètres. Or, plus il y a de paramètres, plus il faut de données pour les estimer. Et donc pour un nombre d'observations fixé de la série, plus le modèle sera complexe, moins bien seront estimés les paramètres. Les critères de choix de modèles les plus courants sont :

- le critère AIC (Akaike Information Criterion), qui sera généralement préféré si l'objectif de l'étude est de faire de la prévision, et qui est défini par :

$$\text{AIC}(p,q) = -2\log L + 2(p+q)$$

- le critère BIC (Bayesian Information Criterion) sera quant à lui généralement préféré si l'objectif de l'étude est de s'ajuster à la série observée, et est défini par :

$$\text{BIC}(p,q) = -2\log L + (p+q)\log T$$

où $\log L$ est la log-vraisemblance du modèle ARMA(p,q) estimé et T est le nombre d'observations. Les modèles ayant la plus petite valeur du critère devront être choisis. Ces deux critères conduisent donc à sélectionner des modèles dont la vraisemblance est grande, en la pénalisant par la complexité du modèle.

2.5.2 Critères d'erreur

Dans l'échantillon initial, (X_1, \dots, X_T) on considère seulement une partie d'observations. Le reste sera à prévoir par le modèle. On peut alors considérer les erreurs :

- Racine quadratique moyenne :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i)^2}$$

- Quadratique moyenne :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{X}_i - X_i)^2$$

– Absolue :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{X}_i - X_i|$$

2.6 Prévision dans les *ARMA* et les *ARIMA*

On suppose $X_t \rightsquigarrow \text{ARMA} : \phi(L)X_t = \mu + \theta(L)\varepsilon_t$ de représentation canonique minimal. On est à la date t et on cherche les meilleures prévisions possibles pour $(x_{t+1}, \dots, x_{t+h})$, prévision à l'ordre h . La prévision optimale de x_{t+h} à la date t est $EL(X_{t+h}|\underline{X}_t)$ (la projection de X_{t+h} sur X_t, X_{t-1}, \dots).

On suppose que les coefficients de ϕ et de θ , ainsi que μ sont connus. En pratique, ils seront remplacés par leurs estimateurs.

2.7 Prévision dans un *AR(p)*

(Prévision optimale). On note : ${}_t X_{t+h}^* = EL(X_{t+h}|\underline{X}_t)$

Soit : $X_t = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$ la représentation canonique.

On a :

(i) Au rang 1 :

$$\begin{aligned} X_{t+1} &= \mu + \varphi_1 X_t + \dots + \varphi_p X_{t+1-p} + \varepsilon_{t+1} \\ {}_t X_{t+1}^* &= EL(X_{t+1}|\underline{X}_t) = \mu + \varphi_1 X_t + \dots + \varphi_p X_{t+1-p} \end{aligned}$$

Cette prévision est calculable à partir des observations X_t, \dots, X_{t+1-p} .

(ii) Au rang 2 :

$$\begin{aligned} X_{t+2} &= \mu + \varphi_1 X_{t+1} + \dots + \varphi_p X_{t+2-p} + \varepsilon_{t+2} \\ {}_t X_{t+2}^* &= EL(X_{t+2}|\underline{X}_t) = \mu + \varphi_1 X_{t+1}^* + \dots + \varphi_p X_{t+2-p} \end{aligned}$$

(iii) Au rang h :

$$\begin{aligned} X_{t+h} &= \mu + \varphi_1 X_{t+h-1} + \dots + \varphi_p X_{t+h-p} + \varepsilon_{t+h} \\ {}_t X_{t+h}^* &= \mu + \sum_{k=1}^{h-1} \varphi_k {}_t X_{t+k}^* + \sum_{k=h}^p \varphi_k X_{t+h-k} \end{aligned}$$

Cette prédiction s'obtient donc par itérations successives.

On peut aller plus vite si on sait résoudre les équations de récurrence :

$$\phi(L)X_t = \mu + \varepsilon_t \Leftrightarrow \phi(L)(X_t - m) = \varepsilon_t, \quad \mu = \frac{m}{\phi(1)}$$

On pose : $Y_t = (X_t - m)$, et on a : ${}_t Y_{t+k}^* = {}_t X_{t+k}^* - m$.

$$\phi(L)Y_{t+h} = \varepsilon_{t+h} \Rightarrow \phi(L)Y_{t+h}^* = 0 \quad \text{avec la notation précédente.}$$

D'où l'équation de récurrence d'ordre p :

$$\phi(L)Y_{t+k}^* = Y_{t+k}^* - \sum_{k=1}^p \varphi_k Y_{t+h-k}^* = 0$$

Son polynôme caractéristique est $z^p \phi(\frac{1}{z})$. On obtient ainsi les ${}_t Y_{t+k}^*$ à partir de la forme générale de la solution de l'équation de récurrence.

Exemple 2.7.1.

$$(1 - \varphi L)X_t = \mu + \varepsilon_t \Leftrightarrow (1 - \varphi L)(X_t - m) = \varepsilon_t$$

On pose $Y_t = X_t - m$, et on a alors : $(1 - \varphi L)Y_{t+h}^* = 0 \Leftrightarrow {}_t Y_{t+h}^* = \varphi {}_t Y_{t+h-1}^*$, d'où la solution générale : ${}_t Y_{t+h}^* = \varphi^h Y_t$.

Donc : ${}_t X_{t+h}^* = m + \varphi^h (X_t - m)$

On ne va certes pas beaucoup plus vite qu'avec une récurrence ici, mais avec des polynômes plus complexes, cette méthode peut être bien plus rapide.

Exemple 2.7.2.

$$(1 - \varphi L)^2 X_t = \varepsilon_t \Rightarrow (1 - \varphi L)^2 {}_t X_{t+h}^* = 0$$

D'où une solution générale de la forme : $\varphi^p (ah + b)$.

On détermine les coefficients en utilisant les conditions initiales :

(i) $h = 0$: $b = X_t$;

(ii) $h = 1$: $\varphi(a + b) = {}_t X_{t+1}^* = 2\varphi X_t + \varphi X_{t-1} \rightarrow b$, on a : $a = X_t - \varphi X_{t-1}$.

Ainsi : ${}_t X_{t+1}^* = \varphi^p ((h + 1)X_t - \varphi h X_{t-1})$

2.8 Prévision dans un $MA(q)$

$$X_t = m + \theta(L)\varepsilon_t = m + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$$

Cas $h > q$

$$X_{t+h} = m + \varepsilon_{t+h} - \theta_1\varepsilon_{t+h-1} - \cdots - \theta_q\varepsilon_{t+h-q}$$

avec $t+h > t+h-1 > \cdots > t+h-q > t$.

D'où :

$${}_tX_{t+h}^* = EL(X_{t+h} | \underline{X}_t) = EL(m + \varepsilon_{t+h} - \theta_1\varepsilon_{t+h-1} - \cdots - \theta_q\varepsilon_{t+h-q} | \underline{X}_t)$$

et donc :

$${}_tX_{t+h}^* = m$$

Cas $h \leq q$

$$\begin{aligned} {}_tX_{t+h}^* &= EL(X_{t+h} | \underline{X}_t) \\ &= EL(m + \varepsilon_{t+h} - \theta_1\varepsilon_{t+h-1} - \cdots - \theta_h\varepsilon_t \cdots - \theta_q\varepsilon_{t+h-q} | \underline{X}_t) \\ &= m + 0 - \theta_h\varepsilon_t \cdots - \theta_q\varepsilon_{t+h-q} \end{aligned}$$

Cette forme théorique est inutilisable dans la pratique, faute de pouvoir observer les ε_t . On va donc utiliser la forme $AR(\infty)$:

$$\theta(L)^{-1}(X_t - m) = \varepsilon_t \Rightarrow \theta(L)^{-1} = \sum_{k=0}^{+\infty} a_k L^k, \quad a_0 = 1, \quad \sum |a_k| < \infty,$$

donc

$$\sum_{k=0}^{+\infty} a_k X_{t-k} - \mu - \varepsilon_t \Leftrightarrow X_t = \mu - \sum_{k=1}^{+\infty} a_k X_{t-k} + \varepsilon_t, \quad \text{où } \mu = \frac{m}{\theta(1)}.$$

D'où

(i) $h = 1$:

$$\begin{aligned} X_{t+1} &= \mu - \sum_{k=1}^{+\infty} a_k X_{t-k+1} + \varepsilon_{t+1} \\ {}_t X_{t+1}^* &= \sum_{k=1}^{+\infty} a_k X_{t-k+1} + 0 \end{aligned}$$

(ii) $h = 2$:

$$\begin{aligned} X_{t+2} &= \mu - a_1 X_{t+1} - \sum_{k=2}^{+\infty} a_k X_{t-k+2} + \varepsilon_{t+2} \\ {}_t X_{t+2}^* &= \mu - a_1 {}_t X_{t+1}^* - \sum_{k=2}^{+\infty} a_k X_{t-k+2} + 0 \end{aligned}$$

(iii) Cas général :

$$\begin{aligned} X_{t+h} &= \mu - \sum_{k=1}^{+\infty} a_k X_{t-k+h} + \varepsilon_{t+h} \\ {}_t X_{t+h}^* &= \mu + \sum_{k=1}^{+\infty} a_k {}_t X_{t+h-k}^* + 0 \\ {}_t X_{t+h}^* &= \mu - \sum_{k=1}^{h-1} a_k {}_t X_{t+h-k}^* - \sum_{k=h}^{+\infty} a_k X_{t-k+h} \end{aligned}$$

2.9 Prévision pour un $ARMA(p, q)$

2.9.1 Prévision optimale

On utilise la forme $AR(\infty)$:

$$\begin{aligned} \phi(L)X_t &= \mu + \theta(L)\varepsilon_t \\ \phi(L)(X_t - \mu) &= \theta(L)\varepsilon_t \\ \theta(L)^{-1}\phi(L)(X_t - \mu) &= \varepsilon_t \\ \theta(L)^{-1}\phi(L)X_t &= \mu_0 + \varepsilon_t \quad \mu_0 = \frac{\mu}{\theta(1)} \end{aligned}$$

D'où :

$$\sum_{k=0}^{+\infty} a_k X_{t-k} = \mu_0 + \varepsilon_t \Leftrightarrow X_t = -\sum_{k=1}^{+\infty} a_k X_{t-k} + \mu_0 + \varepsilon_t$$

On va donc faire comme dans le cas d'un $MA(q)$, sauf que les prévisions ne seront pas constantes à partir d'un certain rang :

$${}_t X_{t+h}^* = \mu_0 - \sum_{k=1}^{+\infty} a_k {}_t X_{t+h-k}^* = \mu_0 - \sum_{k=1}^{h-1} a_k {}_t X_{t+h-k}^* - \sum_{k=h}^{+\infty} a_k {}_t X_{t+h-k}$$

La prévision optimale se calcule de manière itérative.

2.9.2 Prévision approchée

De même que pour le $MA(q)$, si $t > 0$,

$$\begin{aligned} {}_t\hat{X}_{t+h} &= \mu_0 - \sum_{k=1}^{t+h} a_k {}_t\hat{X}_{t+h-k} \\ &= \mu_0 - \sum_{k=1}^{h-1} a_k {}_t\hat{X}_{t+h-k} - \sum_{k=h}^{t+h} a_k {}_tX_{t+h-k} \end{aligned}$$

On le calcule par itérations sur h .

2.9.3 Equation de récurrence

$$\phi(L)(X_t - m) = \theta(L)\varepsilon_t, \quad m = \frac{\mu}{\theta(1)}.$$

$$\phi(L)Y_t = \theta(L)\varepsilon_t \Leftrightarrow Y_t = -(\varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p}) + \varepsilon_t - (\theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q})$$

Si $h \geq q$,

$$Y_{t+h} = -\varphi_1 Y_{t+h-1} - \dots - \varphi_p Y_{t+h-p} + \underbrace{\varepsilon_{t+h}}_{>t} - \theta_1 \underbrace{\varepsilon_{t+h-1}}_{>t} - \dots - \theta_q \underbrace{\varepsilon_{t+h-q}}_{>t}$$

Donc,

$$\begin{aligned} {}_tY_{t+h}^* &= EL(Y_{t+h} | \underline{Y}_t) = EL(X_{t+h} | \underline{X}_t) \\ \Rightarrow {}_tY_{t+h}^* &= -\varphi_1 {}_tY_{t+h-1}^* - \dots - \varphi_q {}_tY_{t+h-q}^* + 0 \end{aligned}$$

Les ${}_tY_{t+h}^*$ vérifient donc l'équation de récurrence : $\phi(L)({}_tY_{t+h}^*) = 0$ de polynôme caractéristique $z^p \phi(\frac{1}{z})$. Ils peuvent être explicités avec p conditions initiales (observations ou prévisions) : On a donc ${}_tX_{t+h}^* = {}_tY_{t+h}^* + m$ vérifient la même équation de récurrence.

2.10 Prévision pour un $ARIMA(p, d, q)$

$$\begin{aligned} (1-L)^d \phi(L)X_t = \mu + \theta(L)\varepsilon_t &\Leftrightarrow \psi(L)X_t = \mu + \theta(L)\varepsilon_t \\ &\Leftrightarrow \psi(L)(X_t - m_t) = \theta(L)\varepsilon_t \\ &\Leftrightarrow \psi(L)Y_t = \theta(L)\varepsilon_t \end{aligned}$$

avec $m_t = \mathbb{E}(X_t)$.

2.10.1 Prédiction optimale

Une prédiction optimale de X_{t+h} sera de la forme (prédiction de $Y_{t+h} + m_{t+h}$).
 Pour les ARIMA, on a vu qu'on devait se donner les conditions initiales :
 $Z = (1, X_0, \dots, X_{-p-d+1}, \varepsilon_0, \dots, \varepsilon_{-q+1})$.

On a alors :

$${}_t Y_{t+h}^* = EL(Y_{t+h} | Y_t, \dots, Y_1, Z)$$

Approximation AR :

$$Y_t = - \sum_{j=1}^{t-1} a_j Y_{t-j} + g'(t)Z \quad \text{avec} \quad \lim_{t \rightarrow +\infty} g(t) = 0$$

Approximation MA :

$$Y_t = \sum_{j=0}^{t-1} b_j \varepsilon_{t-j} - \tilde{g}'(t)Z \quad \lim_{t \rightarrow +\infty} \tilde{g}(t) = 0$$

Alors :

$$\begin{aligned} Y_{t+h} &= - \sum_{j=1}^{t+h-1} a_j Y_{t+h-j} + g'(t+h)Z \\ {}_t Y_{t+h}^* &= EL(Y_{t+h} | Y_t, \dots, Y_1, Z) \\ &= - \sum_{j=1}^{t+h-1} a_j {}_t Y_{t+h-j}^* + g'(t+h)Z \end{aligned}$$

On retrouve ainsi un calcul itératif, problématique car il faudrait calculer explicitement g .

2.10.2 Prédiction approchée

On définit la prédiction approchée comme :

$${}_t \hat{Y}_{t+h} = - \sum_{j=1}^{t+h-1} a_j {}_t \hat{Y}_{t+h-j}$$

En effet, $\lim_{t \rightarrow +\infty} g(t) = 0$.

On peut donc également le calculer par itération sur h .

2.10.3 Equation de recurrence

$$\begin{aligned}\psi(L)Y_t &= \theta(L)\varepsilon_t \\ \psi(L)Y_t &= \varepsilon_{t+h} - \theta_1\varepsilon_{t+h-1} - \cdots - \theta_q\varepsilon_{t+h-q}\end{aligned}$$

donc, si $h > q$, $\psi_t Y_{t+h}^* = 0$, les ${}_t Y_{t+h}^*$ sont solution de l'équation de récurrence dont le polynôme caractéristique est $z^{p+d}\phi(\frac{1}{z})$, ce qui permet de les exprimer en fonction de h . On peut remarquer que 1 est racine à l'ordre d de ce polynôme. ${}_t Y_{t+h}^*$ est ainsi déterminé si on se donne $p + d$ conditions initiales.

On montre de même que les \hat{Y}_{t+h} vérifient la même équation de récurrence.

Chapitre 3

Simulation

3.1 Mise en oeuvre de la méthode de Box-Jenkins avec R

Pour les simulations on a utilisé le langage R et plus particulièrement les packages `tseries` et `forecast`.

3.1.1 Simuler une série chronologique

Créer un objet de type série temporelle :

```
serie = ts (data,start,end,frequency)
```

- **data** contient le vecteur des données.
- **start** et **end** mentionne les dates de début et de fin de la série.
- **frequency** mentionne le nombre de données par unité de temps.

Exemple 3.1.1. *Série `co2` : La concentration atmosphérique de CO₂ à Maunaloa*

`g=co2`

```
t=ts(g,1959,1994,f=12)
```

Représenter graphiquement un objet de type série temporelle :

Exemple 3.1.2. *`plot.ts(g)`*

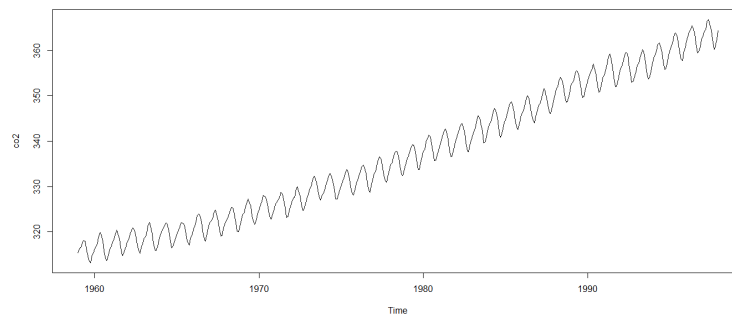


Figure 3.1 : La concentration atmosphérique de CO2 à Maunaloa

Exemple 3.1.3. Simuler un bruit blanc $\mathcal{BB}(0,1)$

```
bruitblanc = rnorm(200,0,1)
```

```
plot(bruitblanc,type='l')
```

On obtient la trajectoire suivante :

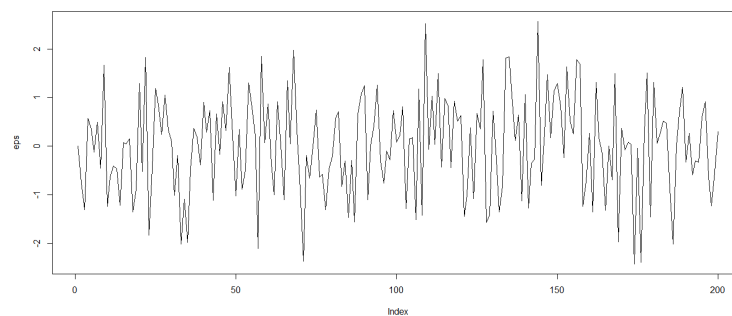


Figure 3.2 : bruit blanc

3.1.2 Transformation

La fonction **log** permet de stabiliser la variance.

La fonction **diff** permet d'appliquer l'opérateur de différenciation Δ_T^k en utilisant la fonction :

```
diff(serie,lag=T,difference=k)
```

Δ_T^k élimine une tendance et une saisonnalité d'ordre k (propo (1.3.2)).

On peut aussi utiliser la fonction **decompose** pour extraire les composantes d'une série temporelle (via la méthode de la moyenne mobile) :

```
decomp <- decompose(serie,type=c('additive','multiplicative'))
```

- **la composante saisonnière** : par `decomp$seasonal`, que l'on suppose additive ou multiplicative dans l'option `type`
- **la tendance** : par `decomp$trend`
- **la partie aléatoire stationnaire de la série** : par `decomp$random`.

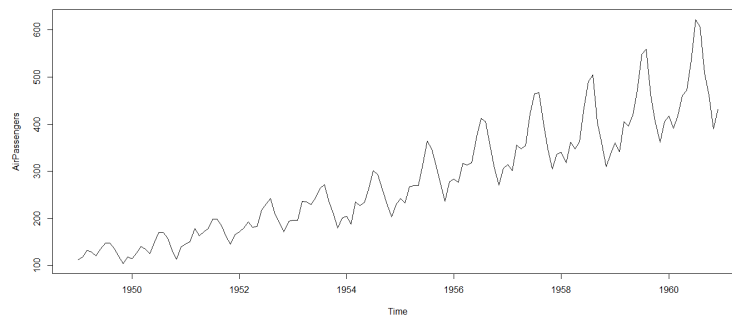


Figure 3.3 : Le graphe de la série AirPassengers

après avoir appliqué la fonction `log` on obtien :

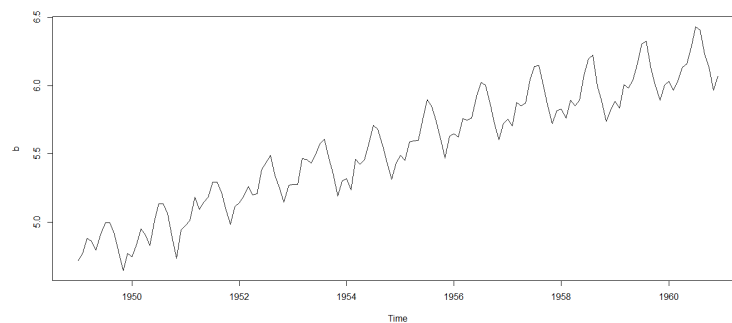


Figure 3.4 : Le graphe de `log AirPassengers`

après la différentiation de la série :

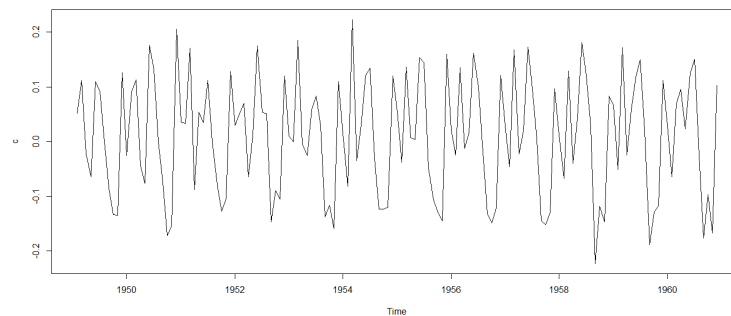


Figure 3.5 : La graphe de la $\text{diff}(\log \text{AirPassengers})$

Après avoir éliminer la tendance et la saisonnalité, l'étape suivante consiste à modéliser la série résiduelle obtenue. La première chose à faire est donc de tester s'il y a une dépendance entre les termes de cette série. Si cette série est un bruit blanc, il n'est donc pas utile d'aller plus loin dans la modélisation. Pour cela on utilise les fonctions suivantes :

(1) La fonction **Box.test(série)** teste la bruité de la série.

Si $p\text{-value} > 0.05$ la série est un bruit blanc.

Dans l'exemple précédent, **Box.test(t)** donne $p\text{-value} = 0.856 > 0.05$ donc c'est un bruit blanc.

(2) **qqplot**, **qqnorm** et **jarque.bera.test** ce sont des fonctions pour tester la normalité des résidus.

(3) La fonction **adf.test(série)** teste la stationnarité de la série. Si $p\text{-value} < 0.05$ la série est stationnaire. On peut également utiliser la fonction **KPSS.test(série)**.

Pour la série précédente **adf.test(t)** donne un $p\text{-value} = 0.01 < 0.05$ et par suite la série est stationnaire.

3.1.3 Identification

La fonction **acf(x)** calcule (et trace si l'option `plot` est à `TRUE`) les `lag.max` premières auto-corrélations empiriques et auto-covariances.

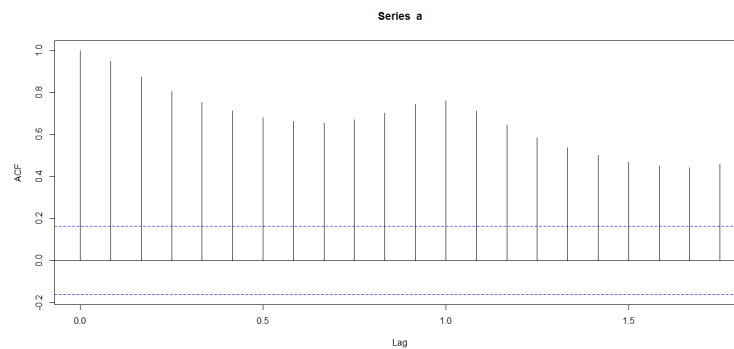


Figure 3.6 : Auto-corrélation

La fonction `pacf(x)` calcule et trace les autocorrélations partielles empiriques

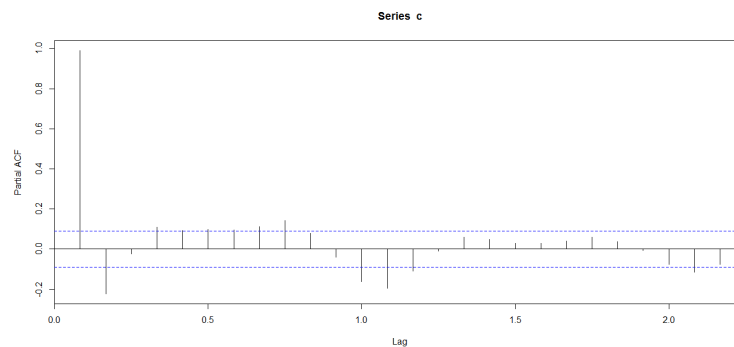


Figure 3.7 : Auto-corrélation partielle

3.1.4 Estimation des paramètres

(1) Estimation des paramètres d'un AR(p) :

Usage : `ar(x, aic = TRUE, order.max = NULL, method=c("yule-walker", "burg", "ols", "mle", "yw"))`

`out<-ar(data,aic=TRUE,order.max=NULL)`

(2) Estimation des paramètres du modèle : Dans le cas général la fonction `arima` permet d'estimer les paramètres :

– D'un ARMA(p,q) : `out=arima(serie,order=c(p,0,q))`

– D'un ARIMA(p,d,q) : `out=arima(serie,order=c(p,d,q))`

– D'un SARIMA(p,d,q,T) :

`out=arima(serie,order=c(p,d,q),seasonal=list(order=c(P,D,Q),period=T))`

Exemple 3.1.4. *Pour un modèle ARMA(2,2)*

```
out=arima(data,order=c(2,0,2))
```

3.1.5 Tests d'adéquation et choix du modèle

Après avoir estimé le modèle, il faut s'assurer encore une fois que les résidus vérifient les propriétés d'un bruit blanc sans oublier de tester aussi la normalité, on utilise alors :

```
Box.test(out$resid) # pour tester la bruité  
adf.test(out$resid) # pour tester la stationnarité  
qqplot(out$resid)  
qqnorm(out$resid) # pour tester la stationnarité graphiquement  
jarque.bera.test(out$resid) # test de stationnarité
```

3.1.6 Prévision

Si à la fin on a plusieurs modèles valides on choisit parmi eux celui qui donne les meilleurs critères d'ajustement et les meilleurs performances en prévision.

La fonction **predict** permet de faire les prévisions à l'horizon h :

```
prevision1=predict(out,n.ahead=h)
```

3.2 Prévision par lissage exponentiel

Afin de comparer la méthode de Box-Jenkins avec d'autres méthodes de prévision, on a choisi la méthode de lissage exponentiel disponible sous R, grâce à la fonction **HoltWinters**.

Pour une série temporelle x , cette procédure permet :

– **Un lissage exponentiel simple :**

```
xlisse <- HoltWinters(x,alpha= $\alpha$ , beta=F, gamma=F)
```

– **Un lissage de Holt-Winters sans composante saisonniere :**

```
xlisse <- HoltWinters(x,alpha= $\alpha$ , beta= $\beta$ , gamma=F)
```

– **Un lissage Holt-Winters additif :**

- ```
xlisse <- HoltWinters(x,alpha= α , beta= β , gamma= γ ,seasonal="add")
```
- **Un lissage Holt-Winters multiplicatif :**

```
xlisse <- HoltWinters(x,alpha= α , beta= β , gamma= γ , seasonal="mul").
```

  - `summary(xlisse)` : description de l'objet `xlisse` obtenu précédemment par la procédure `HoltWinters`
  - `plot(xlisse)` : représentation graphique des valeurs observées et des valeurs lissées
  - `plot(xlissefitted[,1])` : représentation graphique de l'ajustement de la série remis à jour à chaque observation.

Les prévisions à l'horizon  $h$  sont réalisées à l'aide de la fonction **predict** ou la fonction **forecast.HoltWinters** du package `forecast` :

```
p<-predict(xlisse,n.ahead=h)
p2=forecast.HoltWinters(xlisse, h=19)
```

#### Exemple 3.2.1. La série *skirts* :

```
skirts <- scan("http://robjhyndman.com/tsdldata/roberts/skirts.dat",skip=5)
skirtsseries <- ts(skirts,start=c(1866))
plot.ts(skirtsseries)
```

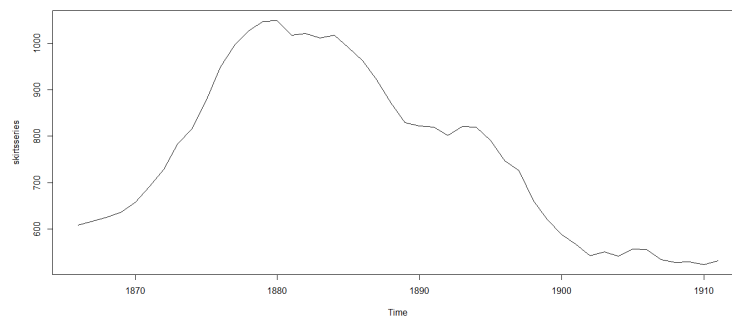


Figure 3.8

La série semble avoir une tendance quadratique et on remarque l'absence de saisonnalité, on utilise alors un lissage exponentiel double.

```
skirtw=window (skirtsseries,1866,1908,1)
skirtsseriesforecasts <- HoltWinters(skirtw, gamma=FALSE)
skirtsseriesforecasts
plot(skirtsseriesforecasts)
```

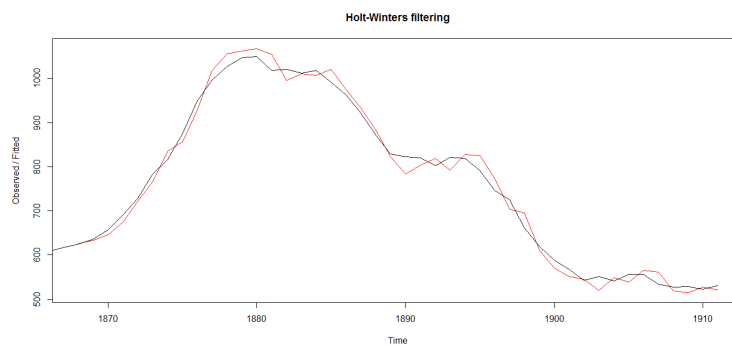


Figure 3.9 : La série skirts et prédiction par lissage

La figure (3.9) montre que les prévisions par lissage sont assez bonnes.

```
skirtsseriesforecasts2 <- forecast.HoltWinters(skirtsseriesforecasts, h=3)
skirtsseriesforecasts2
```

|      | Point forecast | Lo 80    | Hi 80    | Lo 95    | Hi 95    |
|------|----------------|----------|----------|----------|----------|
| 1909 | 514.7115       | 488.5947 | 540.8282 | 474.7694 | 554.6535 |
| 1910 | 502.8466       | 451.9460 | 553.7473 | 425.0008 | 616.6670 |
| 1911 | 490.9818       | 408.0019 | 573.9617 | 364.0750 | 617.8886 |

Ce tableau représente les valeurs de la série skirts, leurs prévisions par lissage et les bornes d'intervalles de confiance à 80 % et 95 %.

```
plot.forecast(skirtsseriesforecasts2)
```

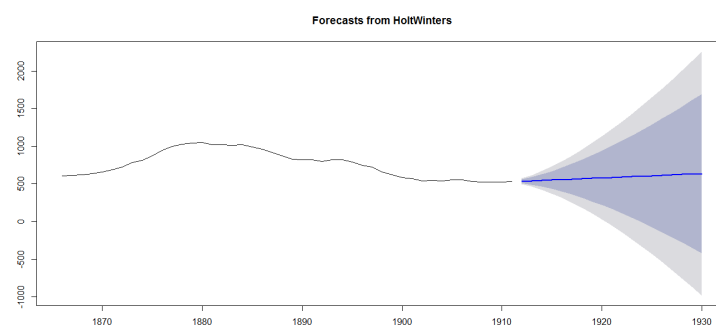


Figure 3.10 : Prédiction + intervalle de confiance



## 3.3 Exemples et comparaisons

### 3.3.1 Application sur des données simulées

**Modèle AR :** On considère le modèle :  $X_t = 0.6X_{t-1} + 0.2X_{t-2} + \varepsilon_t \dots$  (1)

```
modele<-list(ar=c(0.6,0.2))
```

```
ma=arima.sim(modele,200)
```

```
plot(ma)
```

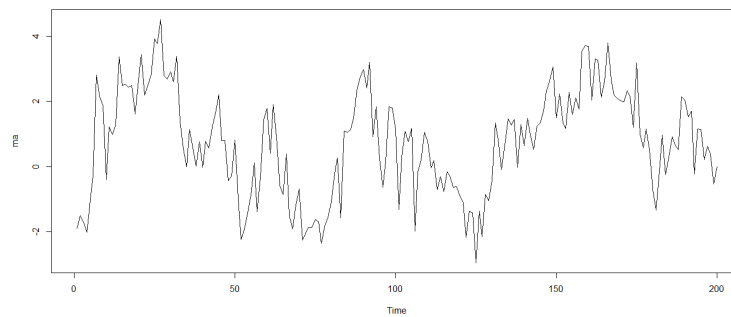


Figure 3.11 : Réalisation du modèle (1)

```
acf(ma)
```

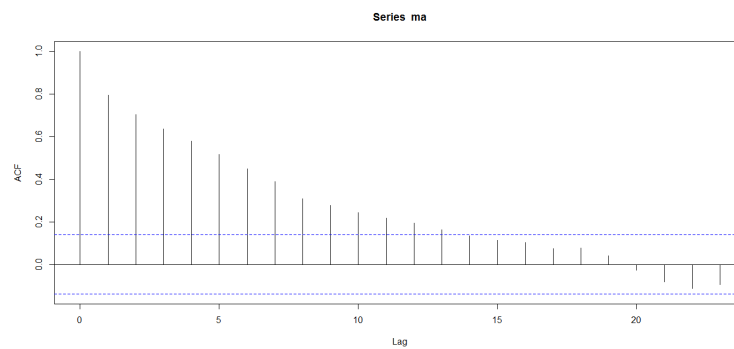


Figure 3.12 : Autocorrélation du modèle (1)

```
pacf(ma)
```

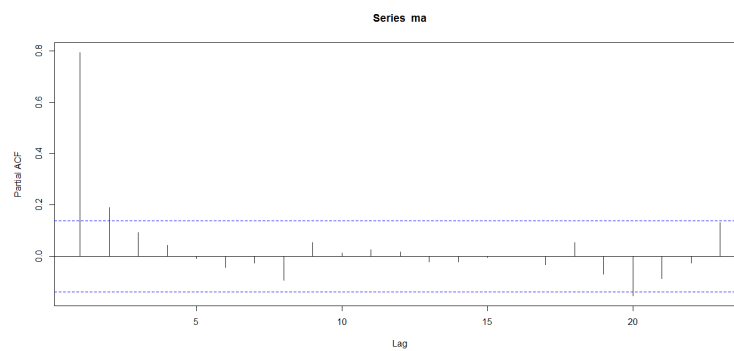


Figure 3.13 : Autocorrélation partielle du modèle (1)

```
out1=ar(ma,aic=TRUE,order.max=NULL)
```

```
out1
```

```
Call : ar(x = ma, aic = TRUE, order.max = NULL)
```

```
Coefficients : 1 2 0.6431 0.1908
```

```
Order selected 2 sigma2 estimated as 0.9133
```

### Interprétation :

Le modèle optimal sélectionné est un AR(2) avec  $\hat{a}_1=0.6431$  et  $\hat{a}_2=0.1908$

### Analyse des résidus :

```
o=out1$res
```

```
plot(o)
```

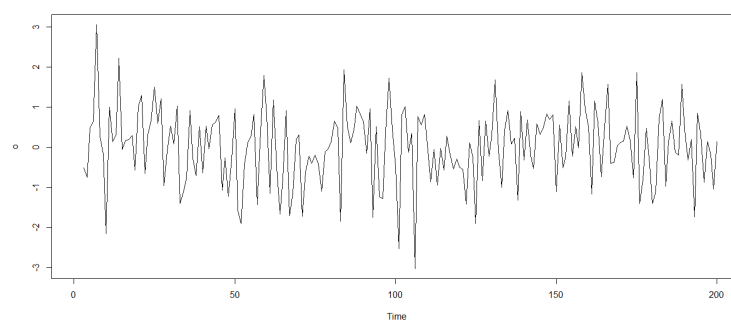


Figure 3.14

```
Box.test(o)
```

```
p-value = 0.6694 > 0.05 .
```

```
p1=predict(out1,n.ahead=1)
```

```
p1
```

```
$pred
```

```
Time Series :
```

```
Start = 201
```

```
End = 201
```

```
Frequency = 1
```

```
[1] 0.01902258
```

```
$se
```

```
Time Series :
```

```
Start = 201
```

```
End = 201
```

```
Frequency = 1
```

```
[1] 0.9556761
```

**Modèle 2 ARMA :**  $X_t = 0.6X_{t-1} + 0.1X_{t-2} + 0.7\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + \varepsilon_t \dots$  (2)

```
modele2<-list(ar=c(0.6,0.1), ma=c(0.7,0.3))
```

```
ma2=arima.sim(modele,200)
```

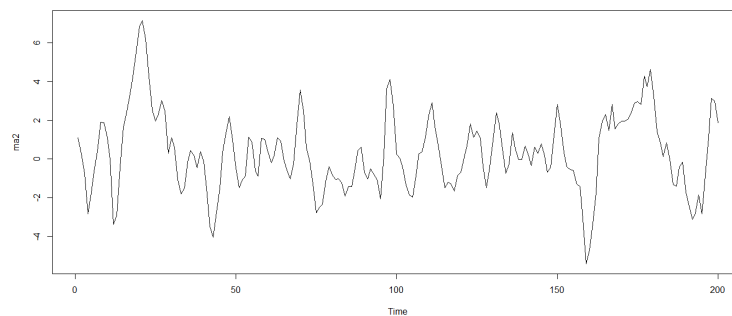


Figure 3.15 : Réalisation du modèle (2)

```
acf(ma2)
```

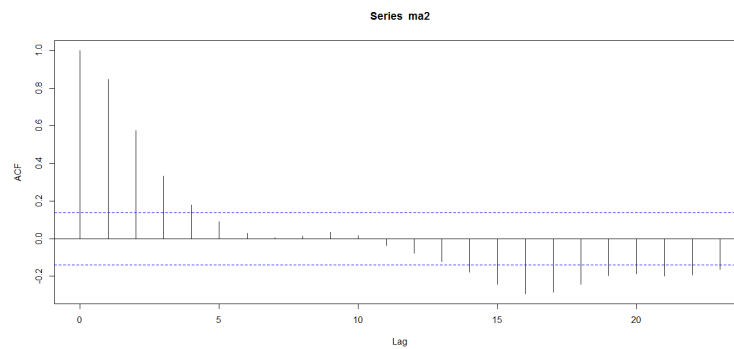


Figure 3.16 : Autocorrélations du modèle (2)

`pacf(ma2)`

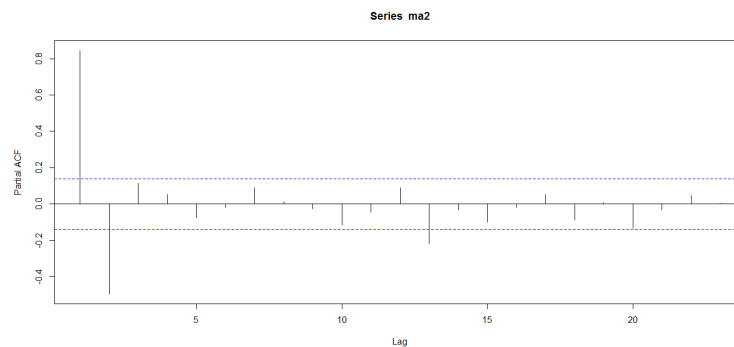


Figure 3.17 : Autocorrélations partielles du modèle (2)

```
outma2=arima(ma2,order=c(2,0,2))
```

```
outma2
```

```
Call : arima(x = ma, order = c(2, 0, 2))
```

|                | <i>ar1</i> | <i>ar2</i> | <i>ma1</i> | <i>ma2</i> | <i>intercept</i> |
|----------------|------------|------------|------------|------------|------------------|
| Coefficients : | 0.5743     | 0.0051     | 0.6684     | 0.3326     | 0.3606           |
| <i>s.e.</i>    | 0.2899     | 0.2325     | 0.2812     | 0.1307     | 0.3148           |

```
sigma2 estimated as 0.8928 : loglikelihood = -273.31, aic = 558.62
```

### Interprétation :

Le modèle sélectionné par la fonction `arima` a pour paramètres  $\hat{a}_1=0.5743$ ,  $\hat{a}_2=0.0051$ ,  $\hat{b}_1=0.6684$ ,  $\hat{b}_2=0.3326$ ,  $\mu=0.3606$

### Analyse des résidus :

```
o2=outma2$res
plot(o)
```

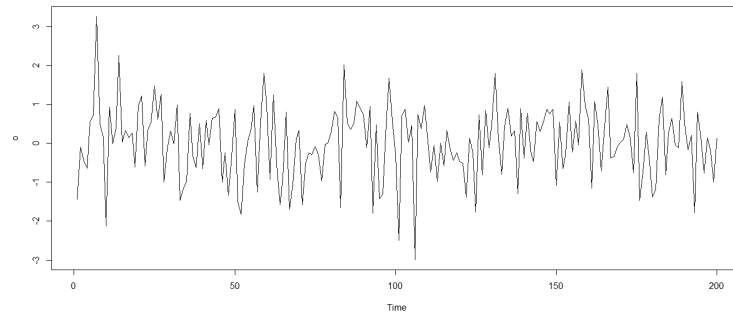


Figure 3.18

```
Box.test(o)
```

```
p-value = 0.9045 > 0.05.
```

Donc la partie résiduelle est bien un bruit blanc.

```
p2=predict(outma2,n.ahead=2)
```

```
p2
```

```
$pred
```

```
Time Series :
```

```
Start = 201
```

```
End = 202
```

```
Frequency = 1
```

```
[1] 3.518635 2.741299
```

```
$se
```

```
Time Series :
```

```
Start = 201
```

```
End = 202
```

```
Frequency = 1
```

```
[1] 0.9448677 1.5071695
```

### 3.3.2 Application sur des données réelles

(1) **La série UKgas** : représente la consommation trimestrielle de gas au Royaume-  
unis de 1960 à 1986 en million de thermes.

(a) **Prévision par méthode de Box-Jenkins**

```
gaz=window(UKgas,1960,c(1985,4))
```

```
plot(gaz)
```

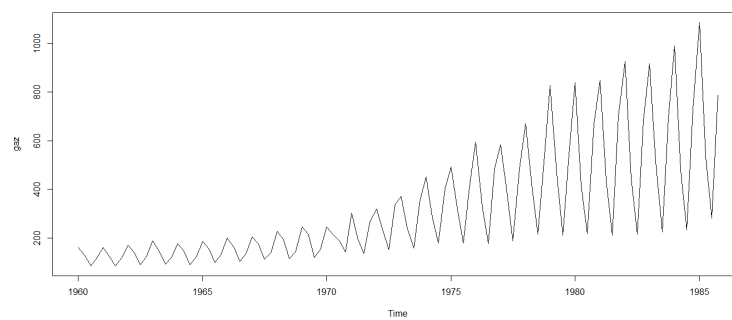


Figure 3.19 : Le graphe de la série gaz

```
acf(gaz)
```

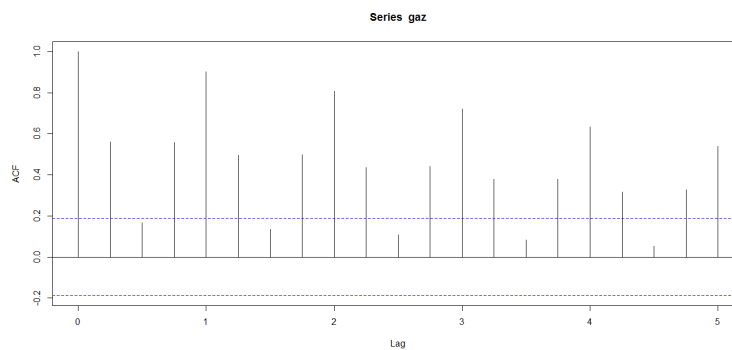


Figure 3.20 : Les ACF de la série gaz

```
pacf(gaz)
```

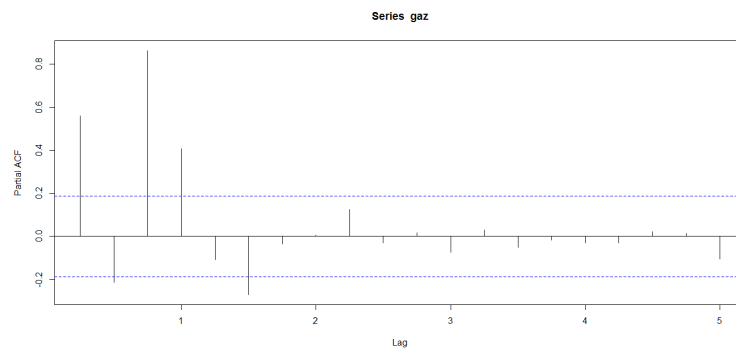


Figure 3.21 : Les PACF de la série gaz

On remarque l'existence d'une saisonnalité et d'une tendance donc l'aspect d'une série non stationnaire.

On passe alors à l'étape de la transformation

```
gaz1=diff(gaz,lag=4,diff=1)
```

```
plot(gaz1)
```

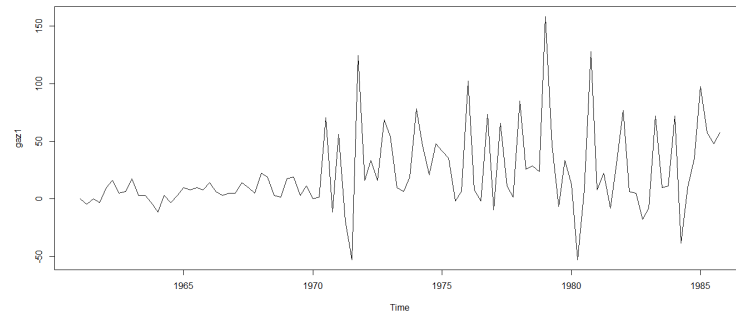


Figure 3.22 : Le graphe après la différentiation

```
acf(gaz1)
```

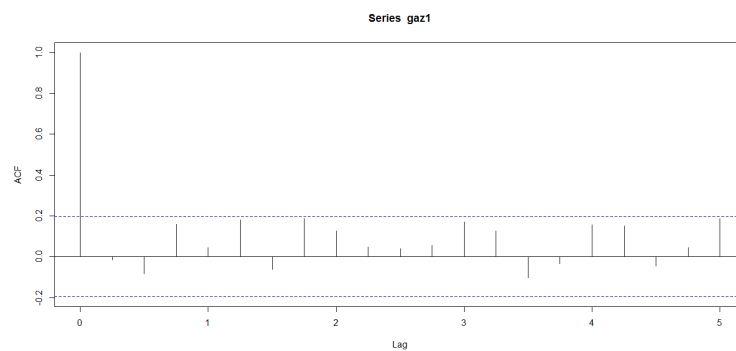


Figure 3.23 : Auto-corrélation de la série gaz1

`pacf(gaz1)`

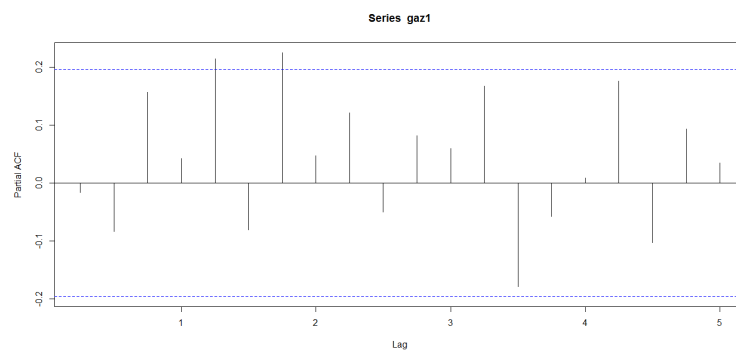


Figure 3.24 : Auto-corrélation partielle de la série gaz1

La série résiduelle semble stationnaire. Testons alors sa bruité.

`Box.test(gaz1)`

on a  $p\text{-value} = 0.8714 > 0.05$  donc c'est un bruit blanc.

Le modèle retenu pour cette série est  $(\mathbf{1-L}^4)\mathbf{UKgas} = \mu + \varepsilon_t$

**Prévisions trimestrielles :**

$$\mu = \text{mean}(\text{gaz1})$$

$$\mu = 21.964$$

$$\hat{X}_{t+1} = \mu + X_{t-3}$$

$$= 1108.964$$

$$\hat{X}_{t+2} = \mu + X_{t-2}$$

$$= 556.664$$



$$\hat{X}_{t+3} = \mu + X_{t-1}$$

$$= 303.764$$

$$\hat{X}_{t+4} = \mu + X_t$$

$$= 809.564$$

**(b) Par lissage exponentiel**

```
gaz=UKgas
```

```
gazw=window(gaz,1960,c(1985,4))
```

```
liss=HoltWinters(gazw,seasonal='mul')
```

```
k=fitted(liss)
```

```
pliss=predict(liss,n.ahead=4)
```

pliss donne les prévisions trimestrielles suivantes :

|      | <i>Qtr1</i> | <i>Qtr2</i> | <i>Qtr3</i> | <i>Qtr4</i> |
|------|-------------|-------------|-------------|-------------|
| 1986 | 1126.1733   | 558.7327    | 287.3435    | 828.2976    |

**(c) Comparaison**

```
gazw2=window(gaz,1986,c(1986,4))
```

```
g=c(1108.964,556.664,303.764,809.564)
```

```
gs=ts(g,1986,f=4)
```

```
Ereur1=(sqrt((sum((gazw2-g)^2))))/4
```

```
Erreur1
```

```
23.48324
```

```
Ereur2=(sqrt((sum((gazw2-pliss)^2))))/4
```

```
Erreur2
```

```
25.06982
```

```
plot(pliss,col='blue')
```

```
lines(gazw2)
```

```
lines(gs,col='red')
```

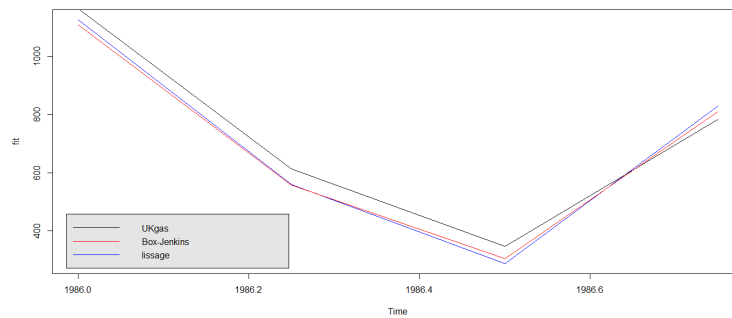


Figure 3.25 : courbes des prédicteurs par Box-Jenkins et par lissage exponentiel

La méthode de Box-Jenkins donne des prévisions meilleures.

(2) **La série température** : représente les données de la température d'Oran de 2010 à 2015.

(a) **Prévision par Box-Jenkins** :

```

temperature=c(24,33,27,32,30,34,37,37,34,33,25,28
,28,21,24,29,30,35,37,37,35,30,28,22,21,18,22,26,31,36,
35,39,36,31,29,20,29,23,25,27,29,33,34,37,32,34,29,24,24
,28,28,30,32,36,38,36,36,36,30,23,24,20,27,33,37,39,39,39,36)
c=ts(temperature,2010,f=12)
d=window(c,2010,c(2014,12),f=12)

```

plot(d)

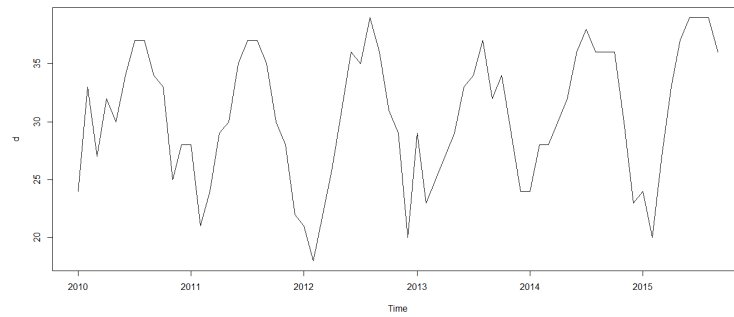


Figure 3.26 : Le graphe de la série température d'Oran

acf(d)

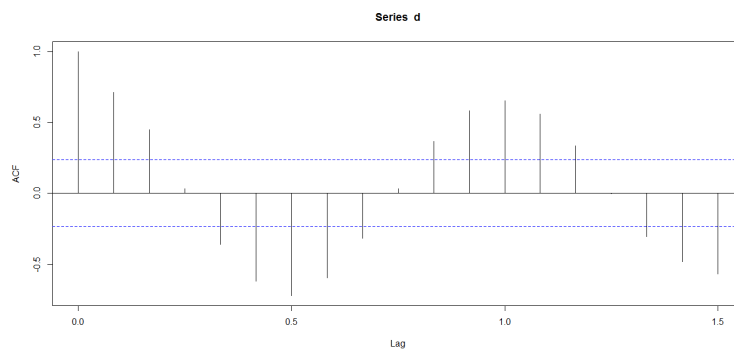


Figure 3.27 : Les ACF de la série température

pacf(d)

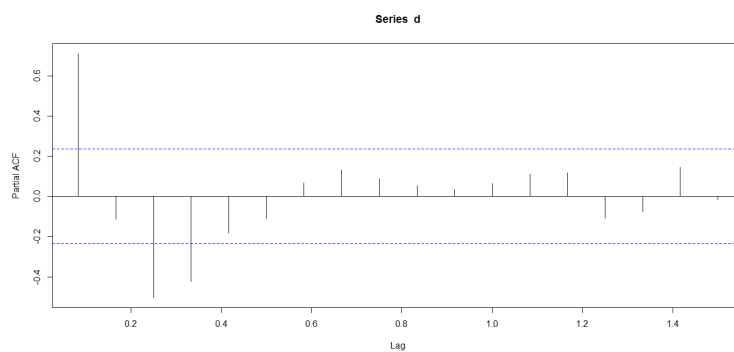


Figure 3.28 : Les PACF de la série température

La série n'est pas stationnaire car l'acf de la série montre l'existence d'une saisonnalité.

### Transformation

```
did=diff(d,dif=1,lag=12)
```

```
plot(did)
```

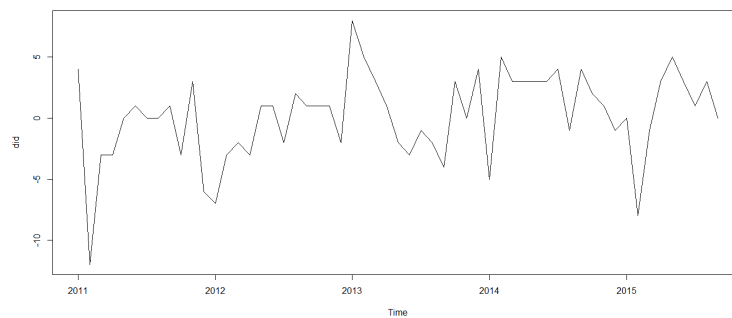


Figure 3.29 : Le graphe de la série température après la différentiation

```
acf(did)
```

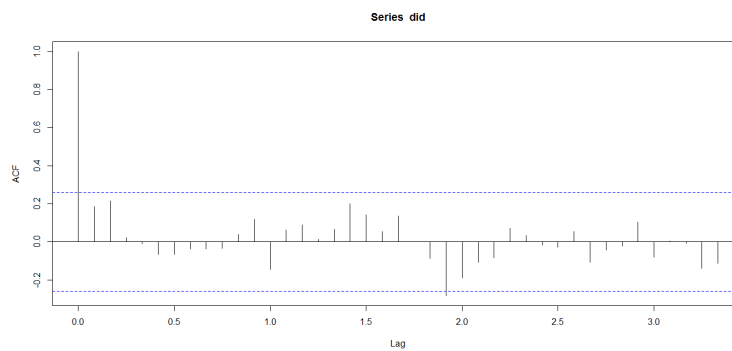


Figure 3.30 : Auto-corrélation de la série température

```
pacf(did)
```

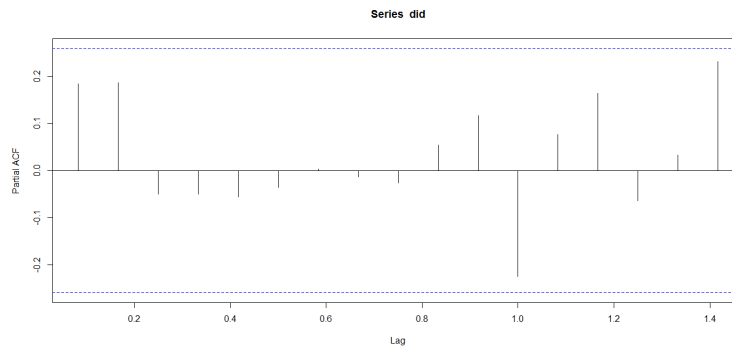


Figure 3.31 : Auto-corrélation partielle de la série température

### Test sur la série résiduelle

Box.test(did)

on a p-value = 0.2867 > à 0.05 donc c'est un bruit blanc.

**Donc le modèle retenu est  $(1-L^{12})X_t = \mu + \varepsilon_t$**

Les prévisions mensuelles de l'année 2015 de la température sont calculées par :

$$\mu = \text{mean}(\text{did})$$

$$\mu = 0.0625$$

$$\begin{aligned}\hat{X}_{t+1} &= \mu + X_{t-11} \\ &= 24.0625\end{aligned}$$

$$\begin{aligned}\hat{X}_{t+2} &= \mu + X_{t-10} \\ &= 28.0625\end{aligned}$$

$$\begin{aligned}\hat{X}_{t+3} &= \mu + X_{t-9} \\ &= 28.0625\end{aligned}$$

$$\begin{aligned}\hat{X}_{t+4} &= \mu + X_{t-8} \\ &= 30.0625\end{aligned}$$

$$\begin{aligned}\hat{X}_{t+5} &= \mu + X_{t-7} \\ &= 32.0625\end{aligned}$$

$$\begin{aligned}\hat{X}_{t+6} &= \mu + X_{t-6} \\ &= 36.0625\end{aligned}$$

$$\begin{aligned}\hat{X}_{t+7} &= \mu + X_{t-5} \\ &= 38.0625\end{aligned}$$

$$\hat{X}_{t+8} = \mu + X_{t-4}$$

```

= 36.0625
 $\hat{X}_{t+9} = \mu + X_{t-3}$
= 36.0625
(b) Par lissage exponentiel :
liss=HoltWinters(d,seasonal='mul')
k=fitted(liss)
pliss=predict(liss,n.ahead=9)
pliss= 26.36578, 24.43487, 25.65824, 28.39210, 30.61366, 34.93813, 36.03144, 36.56475,
34.44850.
(c) Comparaison
d2=window(d,2015,c(2015,9),f=12)
v=ts(c(24.0625,28.0625,28.0625,30.0625,32.0625,36.0625,38.0625,36.0625,36.0625),2015,f=12)
Erreur1=(sqrt((sum((d2-v)^2))))/9
Erreur1
1.203314
Erreur2=(sqrt((sum((d2-pliss)^2))))/9
Erreur2
1.230939
plot(d2)
lines(v,col='red')
lines(pliss,col='blue')

```

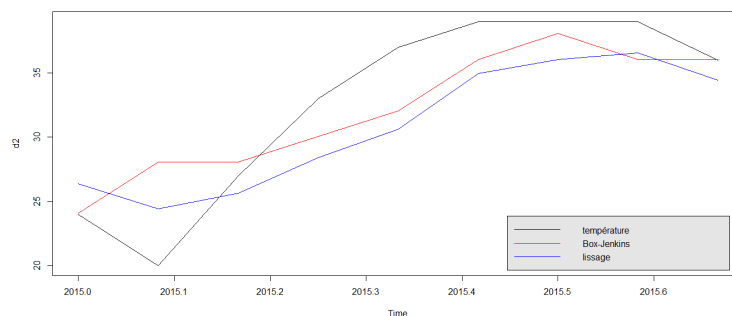


Figure 3.32 : courbes des prédicteurs par Box-Jenkins et par lissage exponentiel

La méthode de Box-Jenkins donne légèrement des prévisions meilleures.

(3) **La série ciment** : représente des données mensuelles des ventes du ciment au sein S.C.I.S de Saïda de 2015 au premier trimestre du 2017.

(a) **Prévision par méthode de Box-Jenkins**

```
ciment=c(26040,25260,33350,30750,11220,21750, 27540,30520,28310,30580,22910,24890,19460,
10660,28950,22250,25110,16300,22130,31420, 29960,32640,25030,25080 ,20260,25100,25200)
```

```
s=ts(ciment,2015,c(2017,3),f=12)
```

```
sw=window(s,2015,c(2016,12))
```

```
plot(sw)
```

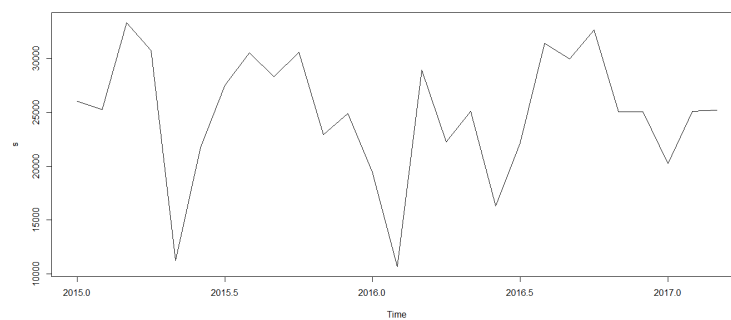


Figure 3.33 : Le graphe de la série ciment

```
acf(sw)
```

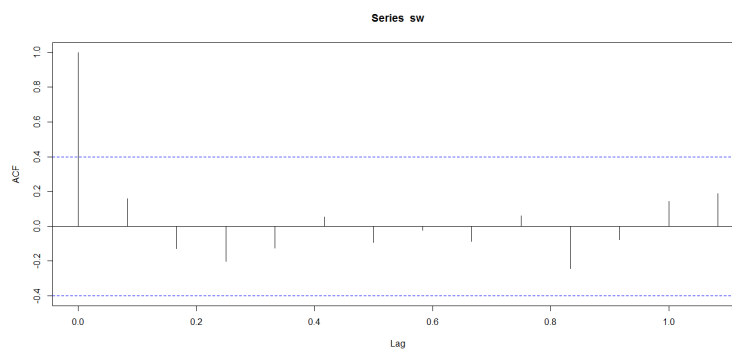


Figure 3.34 : Les ACF de la série ciment

```
pacf(sw)
```

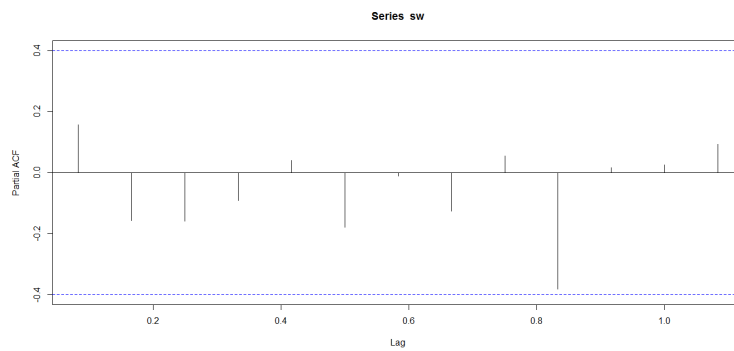


Figure 3.35 : Les PACF de la série ciment

La série n'est pas stationnaire selon les graphes précédents. On opte alors à la transformation suivante.

```
sis=diff(sw,lag=12,diff=1)
```

```
plot(sis)
```

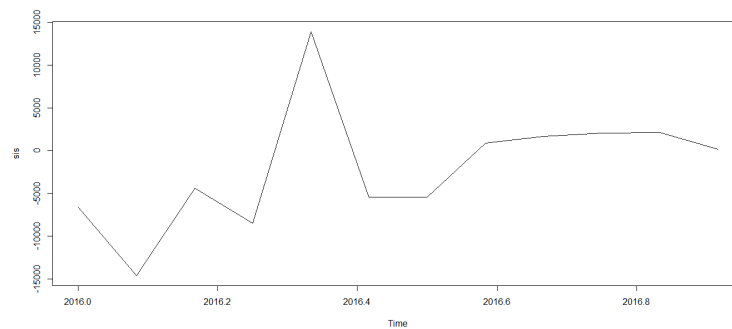


Figure 3.36 : Le graphe de la série ciment après la différentiation



acf(sis)

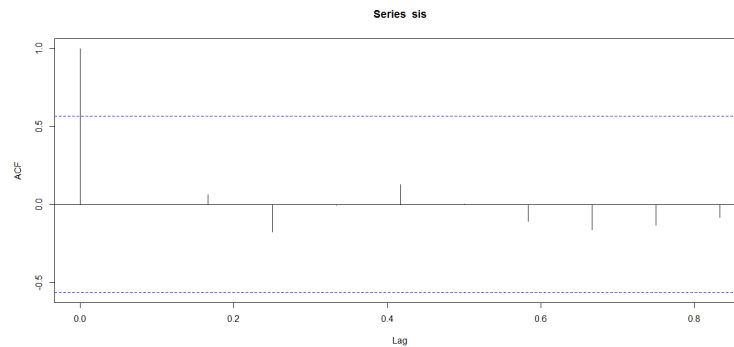


Figure 3.37 : Auto-corrélation de la série ciment

pacf(sis)

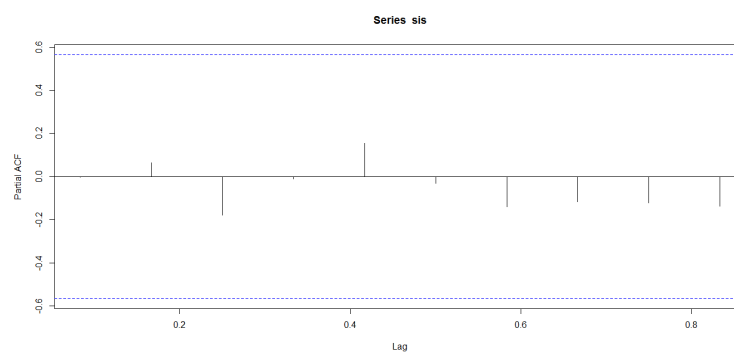


Figure 3.38 : Auto-corrélation partielle de la série ciment

### Analyse des résidus

Box.test(sis)

on a p-value = 0.9926 > à 0.05

Le test sur la série résiduelle montre alors que la série différencié est stationnaire.

**Le modèle retenu est  $(1-L^{12})X_t = \mu + \varepsilon_t$**

**Les prévisions mensuelles :**

$$\mu = \text{mean}(\text{did})$$

$$\mu = -2010.833$$

$$\hat{X}_{t+1} = \mu + X_{t-11}$$

$$\begin{aligned}
 &= 17449.17 \\
 \hat{X}_{t+2} &= \mu + X_{t-10} \\
 &= 8649.167 \\
 \hat{X}_{t+3} &= \mu + X_{t-9} \\
 &= 26939.17
 \end{aligned}$$

**Prévision par lissage exponentiel :**

```
liss=HoltWinters(sw)
```

```
k=fitted(liss)
```

```
pliss=predict(liss,n.ahead=3)
```

(c) **Comparaison**

```
sw2=window(s,2017,f=12)
```

```
a=c(17449.17,8649.167,26939.17)
```

```
a=ts(c(17449.17,8649.167,26939.17),2017,f=12)
```

```
Erreur=(sqrt((sum((sw2-a)^2))))/3
```

```
Erreur1
```

```
5593.204
```

```
Erreur2=(sqrt((sum((sw2-pliss)^2))))/3
```

```
Erreur2
```

```
4744.211
```

```
plot(a,col='red',type='l')
```

```
lines(sw2)
```

```
lines(pliss,col='blue')
```

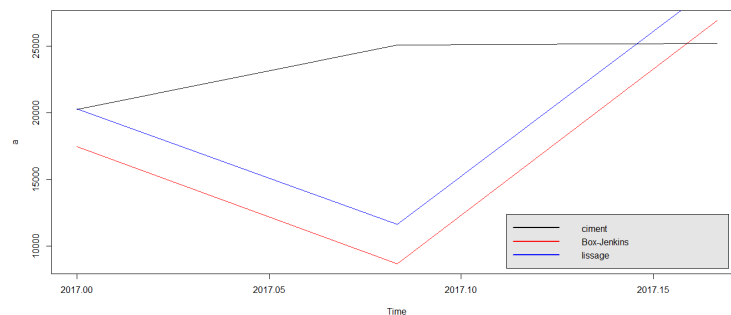


Figure 3.39 : courbes des prédicteurs par Box-Jenkins et par lissage exponentiel

La meilleure prévision est donnée par la méthode de lissage exponentiel.

(4) **La série skirts** : On reprend la série skirts déjà analysée par lissage exponentiel.

```
skirts <- scan("http://robjhyndman.com/tsdldata/roberts/skirts.dat",skip=5)
skirtsseries <- ts(skirts,start=c(1866))
plot(skirtsseries)
```

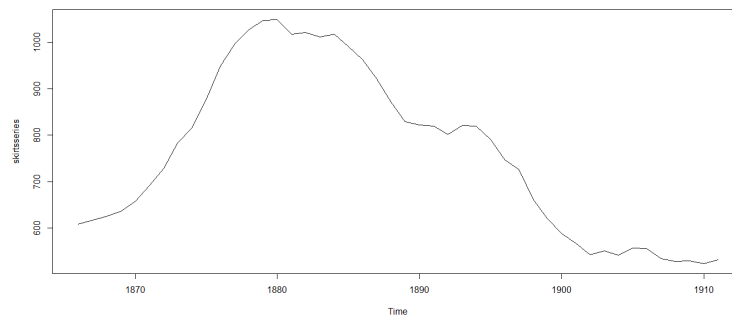


Figure 3.40 : La série skirtsseries

On remarque d'après le graphe qu'il y a une tendance.

```
skirtsseriesdiff1 <- diff(skirtsseries, differences=1)
plot(skirtsseriesdiff1)
```



Figure 3.41 : La série skirtsseries différenciée

On remarque qu'après une première différenciation la série n'est pas encore stationnaire.

```
skirtsseriesdiff2 <- diff(skirtsseries, differences=2)
plot(skirtsseriesdiff2)
```

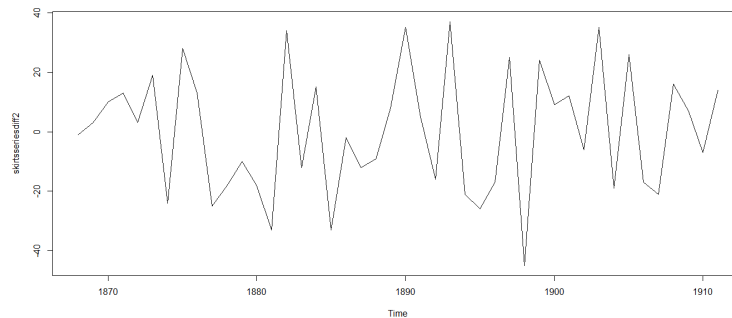


Figure 3.42 : La série skirtseries deux fois différenciée

La série semble stationnaire après une deuxième différenciation.

```
Box.test(skirtseriesdiff2)
```

p-value = 0.0442 < 0.05 donc le test sur la série résiduelle à montrer que cette dernière n'est pas un bruit blanc.

```
adf.test(skirtseriesdiff2)
```

p-value = 0.2817 < 0.5

Le test de stationnarité adf indique que la série résiduelle est stationnaire.

### L'estimation :

Puisque cette série est stationnaire on l'a modélise par un processus stationnaire en utilisant :

```
estimation=arima(skirtseriesdiff2)
```

Call :

```
arima(x = skirtseriesdiff2)
```

Coefficients :

intercept

-0.0227

s.e. 3.1209

sigma  $\hat{2}$  estimated as 428.6 : log likelihood = -195.76, aic = 395.53

**Le modèle retenu est  $(1-L)^2\text{skirtseries} = -0.0227 + \varepsilon_t$**

**Les prévisions annuelles :**

$\mu = -0.0227$

$\hat{X}_{t+1} = \mu + 2X_t - X_{t-1}$

$$= 521.9773$$

$$\hat{X}_{t+2} = \mu + 2_{t+1} - X_t$$

$$= 515.9319$$

$$\hat{X}_{t+3} = \mu + 2_{t+2} - X_{t+1}$$

$$= 509.8638$$

(b) **La prévision annuelle par lissage exponentiel :**

prevision=c(514.7115,502.8466,490.9818)

(c) **Comparaison**

skirtw2=window(skirtsseries,1909,f=1)

f=ts(c(521.9773,515.9319,509.8638),1909,f=1)

Erreur1 = (sqrt((sum((skirtw2 -f)<sup>2</sup>))))/3

Erreur1

7.78899

Erreur2=(sqrt((sum((skirtw2 -w)<sup>2</sup>))))/3

Erreur2

15.67651

plot(skirtw2,ylim=c(485,535))

lines(f,col='red')

lines(skirtsseriesforecasts,col='blue')

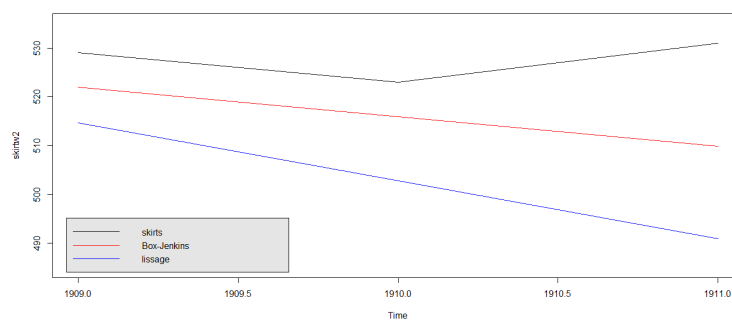


Figure 3.43 : courbes des prédicteurs par Box-Jenkins et par lissage exponentiel

La meilleure prévision est donnée par la méthode de Box-Jenkins.

# 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 part 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. Alors ce mémoire correspond à peu près au choix : "temps discret, scalaire, linéaire, domaine temporel, paramétrique".

# Bibliographie

- [1] **Abadie et Meslier.** Etude de l'utilisation des modèles ARIMA pour la prévision à très court terme de l'énergie journalière produite par Electricité de France. RAIRO Recherche opérationnelle Operations research *n*°1, vol. 13p37 – 54, février 1979.
- [2] **Yves. Aragon.** Introduction aux Séries temporelles, Septembre 2004.
- [3] **Yves. Aragon.** Séries temporelles avec R : Méthodes et cas, Springer science Business Media, 2011, ISSN : 281780208X, 9782817802084.
- [4] **G. Box, G. Jenkins :** Time Series analysis, Holden Day, 1976.
- [5] **Arthur. Charpentier.** Cours de séries temporelles, Volume1 : Introduction à la théorie des processus en temps discret Modèles ARIMA et méthode Box & Jenkins.
- [6] **Didier. Delignières.** Séries temporelles-Modèles ARIMA, Mars 2000.
- [7] **C.Do.** Prévisions dans les ARMA et les ARIMA, Mars 2003.
- [8] **C.Do.** Séies Temporelles Linéaires, Année 2004-2005.
- [9] **M. Gevers et L. Vandendorpe.** Processus stochastiques, estimation et prédiction, Université catholique de Louvain Faculté des Sciences Appliquées.
- [10] **Céline Lévy-Leduc.** Introduction à l'Etude des Séries Chronologiques, 13 mai 2008.
- [11] **Martin.** Filtrage de Kalman d'une série temporelle saisonnière, Application à la prévision de consommation d'électricité. Revue de statistique appliquée v. XLVII p 69 – 86, 1999
- [12] **Eric. Moulines et François Roueff.** Analyse des Séries Temporelles et Applications, 15 septembre 2010.

- [13] Corinne. **Perraudin**. Séries Chronologiques, Année 2004-2005.
- [14] Bernard. **Rapacchi**. Les séries chronologiques, Centre Interuniversitaire de Calcul de Grenoble, 18 août 1993.
- [15] **Axe**. **Sectoriel**. Analyses des Chroniques, 2007-2008.