

## *Dédicaces*

*Après avoir terminé ce mémoire de fin d'étude, Je remercie  
Dieu m'avoir donnée le courage et la force de bien  
mener mon travail. Je dédie ce modeste travail :*

*A celle qui me conseille.*

*Et grâce à elle que j'aime la vie, à ma mère, ma moitié et  
ma chère.*

*A celui que j'offre ma réussite, mon bonheur et mon  
respect, qu'il trouve ici*

*l'expression de son affection et une récompense des  
sacrifices consentis pour moi, à mon tendre, cher père.*

*Que Dieu les protège et leur préserve la santé et le bonheur.*

*A mes chers frères.*

*A mes soeurs.*

*A tous mes chers amis.*

*A toute la promotion : 2011-2012*

*Et à toute personne qui m'a aidé de près ou de loin.*

## *Remerciements*

*Je voudrais remercier chaleureusement ma directrice de mémoire Mme **Mokhtari Fatiha**, pour la façon dont elle a encadré mon travail. Ainsi que pour sa disponibilité, ses conseils et ses encouragements. Je la remercie pour la deuxième fois.*

*Je remercie également monsieur **Kandouci Abdeljabbar** qui m'a fait l'honneur de présider le jury de cette mémoire.*

*Je tiens à remercier aussi Mme **Benziadi Fatima** et Melle **Maref Fouzia** qui ont accepté d'évaluer mon travail.*

*Je remercie particulièrement les enseignants de probabilité, sans oublier les membres du laboratoire.*

*Je saisi cette occasion pour remercier l'ensemble des enseignants qui m'ont initié aux mathématiques.*

*Je pense aussi à ceux et celles avec qui j'ai étudié. Pour finir, je remercie tous les amis et collègues qui m'ont soutenu et encouragé.*

*Merci à tous.*

**TIGRINE Samiha**

## Résumé

On examine les caractéristiques principales des processus stochastiques univariés en nous attachant principalement aux processus stationnaires. On a détaillé quelques processus paramétriques usuels qui seront fort utiles par la suite. Le fil conducteur de ce mémoire concerne la non-stationnarité et les modèles SARIMA pour modéliser la partie résiduelle d'une série chronologique par la méthode de "Box & Jenkins". On a suggéré plusieurs méthodes pour obtenir la stationnarité. On verra que ces méthodes sont engendrées par la méthode de "Box & Jenkins". On termine ce travail par donner des applications sur les séries économiques (la série AirPassengers, et la chronique SNCF).

### *Mots clés*

*Processus (stationnaire et non stationnaire), tendance, saisonnalité, polynômes retard et avance, processus (AR, MA, ARMA, TS, DS ARIMA, SARIMA), autocovariance, autocorrélation, autocorrélation partiel, critère AIC, critère BIC, identification, estimation, validation et prévision.*

## ***Title : SARIMA models***

### **Abstract**

We have reviewed the main features of univariate stochastic processes in us focusing primarily on stationary processes. It has detailed some usual parametric processes that will be useful later. The theme of this memory concerns the non-stationarity and SARIMA models to model the residual part of a series by the method of "Box & Jenkins". It was suggested several methods to obtain stationarity. Will show that these methods generate by the method of "Box & Jenkins". We end this work by giving applications on economic series (the series AirPassenge, and chronicling SNCF).

### ***Key words***

*Process (stationary and nonstationary), trend, seasonal component, polynomial backwards and forwards, process (AR, MA, ARMA, TS, DS, ARIMA, SARIMA), autocovariance, autocorrelation, autocorrelation part, AIC criterion, BIC criterion, identification, estimation, validation and forecasting.*

# Table des matières

Introduction	9
<b>1 Généralité sur les séries chronologiques (time series)</b>	<b>11</b>
1.1 Définitions et Exemples . . . . .	11
1.2 Les composantes d'une série temporelle . . . . .	14
1.3 Quelques types de décomposition . . . . .	17
1.4 Stationnarité . . . . .	21
1.4.1 Définitions . . . . .	21
1.4.5 Exemples ( <i>Processus stationnaires</i> ) . . . . .	22
1.4.6 Définition ( <i>La fonction d'autocorrélation</i> ) . . . . .	23
1.4.8 Définition ( <i>La fonction d'autocorrélation partiel</i> ) . . . . .	23
1.4.11 Polynôme retard et avance . . . . .	24
1.5 Inversibilité de $I - \lambda L$ . . . . .	27
1.6 Inversibilité des polynômes en $L$ . . . . .	27
1.6.1 Définition ( <i>Inversibilité</i> ) . . . . .	27
<b>2 Modélisation ARMA des séries chronologiques</b>	<b>29</b>
2.1 Processus linéaires . . . . .	29
2.1.1 Définitions et propriétés . . . . .	29
2.1.5 Définition (innovation) . . . . .	30
2.1.6 Théorème (Wold) . . . . .	30
2.2 Processus auto-régressifs (AR : AutoRegressive) (Yule 1927) . . . . .	32
2.2.1 Définition et structure . . . . .	32
2.2.8 Ecriture moyenne mobile infinie d'un AR(p) . . . . .	34
2.2.9 Bruit blanc d'innovation d'un AR(p) . . . . .	34
2.2.10 Propriétés des processus AR(p) . . . . .	35
2.2.15 Exemple AR(1) . . . . .	36
2.3 Processus moyenne mobile (MA : Moving Average) (Slutsky 1927) . . . . .	40
2.3.1 Définitions et structures . . . . .	40
2.3.6 Forme autorégressive infinie d'un MA(q) . . . . .	41
2.3.7 Bruit blanc d'innovation d'un MA(q) . . . . .	41
2.3.9 Propriétés d'un MA(q) . . . . .	42

2.3.13	Prévision d'un MA(q)	44
2.4	Processus mixtes ARMA(Wold 1954)	44
2.4.1	Définition et représentation canonique minimale	44
2.4.5	Expression d'un ARMA(p, q)	46
2.4.7	Bruit blanc d'innovation d'un ARMA(p, q)	47
2.4.8	Autocovariances et propriétés d'un ARMA(p, q)	48
<b>3</b>	<b>Modélisation par les processus linéaires non stationnaires</b>	<b>51</b>
3.1	Non stationnarité	51
3.1.1	série intégrée	51
3.1.3	Exemples ( <i>Processus non stationnaires</i> )	51
3.2	Processus ARIMA	54
3.2.1	Définitions et propriétés	54
3.3	Les modèles ARIMA saisonniers (SARIMA)	58
3.3.1	Définitions et exemples	58
3.4	Modélisation des chroniques d'un modèle SARIMA	60
3.4.1	Identification et ajustement des modèles SARIMA aux données	60
3.4.8	Prévision des processus SARIMA	72
3.4.9	Exemple : (Modélisation de la série des voyageurs)	76
3.4.10	Exemple : (Application à la chronique SNCF)	82
	<b>Conclusion</b>	<b>93</b>
	<b>Bibliographie</b>	<b>95</b>

# Table des figures

1.1	Croissance de la productivité du travail dans le secteur privé, au Canada et aux États-Unis . . . . .	12
1.2	Naissances et Décès en Russie depuis 1959 . . . . .	13
1.3	Évolution du nombre de battements par minute en fonction du temps mesuré en seconde. . . . .	13
1.4	Températures . . . . .	14
1.5	Tendance linéaire . . . . .	15
1.6	Tendance polynômiale . . . . .	15
1.7	Tendance exponentielle . . . . .	16
1.8	Modèle Additif : Ventes d'un produit . . . . .	18
1.9	Modèle Multiplicatif : Ventes d'un produit . . . . .	19
1.10	Modèle Mixte : Nombre de pages vues par mois . . . . .	20
1.11	Simulation d'un Bruit blanc et sa fonction d'autocorrélation. . . . .	22
2.1	Simulation d'un $AR(1)$ : $X_t = 0.8X_{t-1} + \epsilon_t$ . . . . .	38
2.2	Simulation d'un $MA(1)$ : $X_t = \epsilon_t - 0.8\epsilon_t$ . . . . .	43
2.3	Simulation d'un $ARMA(1)$ : $X_t = 0.8X_{t-1} + 0.6\epsilon_{t-1} + \epsilon_t$ et sa fonction d'autocorrélation . . . . .	49
3.1	Simulation de processus TS . . . . .	52
3.2	Simulation de $X_t = X_{t-1} + \epsilon_t$ et sa fonction d'autocorrélation. . . . .	55
3.3	Simulation d'un $X_t = a \sin(\frac{\pi t}{6} + 4) + \epsilon_t$ , . . . . .	59
3.4	Évolution du prix de pétrole, et de son logarithme, entre Janvier 1989 et Avril 2009 . . . . .	61
3.5	Simulation de la série $CAC$ . . . . .	63
3.6	Schéma général de la modélisation d'une série temporelle par un modèle SARIMA. . . . .	74
3.7	Nombre total de passagers avion de 1949 à 1960. . . . .	76
3.8	Log de la série du nombre total de passagers avion de 1949 à 1960 . . . . .	77
3.9	La fonction d'autocorrélation de la série des voyageurs . . . . .	78
3.10	La fonction d'autocorrélation partiel de la série des voyageurs . . . . .	78
3.11	La série du nombre total de passagers avion sans saisonnalité . . . . .	80
3.12	La série du nombre total de passagers avion sans tendance . . . . .	81

---

3.13	Prédiction du nombre de du voyageurs pour les deux années suivante .	82
3.14	Chronique SNCF . . . . .	83
3.15	Corrélogramme de chronique SNCF . . . . .	83
3.16	Corrélogramme partiel de chronique SNCF . . . . .	84
3.17	Chronique des différences saisonnières . . . . .	85
3.18	Corrélogramme des différences saisonnières . . . . .	86
3.19	Corrélogramme partiel des différences saisonnières . . . . .	86
3.20	Chronique différence deux fois . . . . .	87
3.21	Corrélogramme de la chronique différence deux fois . . . . .	87
3.22	Corrélogramme partiel de la chronique différence deux fois . . . . .	88
3.23	Chronique des résidus . . . . .	88
3.24	Corrélogramme des résidus . . . . .	89
3.25	Corrélogramme partiel des résidus . . . . .	89
3.26	Prévisions par SARIMA $(0, 1, 1)(0, 1, 1)_{12}$ . . . . .	91

# Introduction

Depuis un demi siècle les processus aléatoires stationnaires jouent un rôle central dans la modélisation mathématique de nombreux phénomènes concrets.

Les domaines d'application incluent entre autre la théorie de signal (transmission de signaux en présence de bruit, modélisation de parole, reconnaissance des formes, etc...), la prédiction de quantités économiques à évolution fluctuante (indice de prix, cour de la bourse, taux de chômage, etc ...), la méthodologie (analyse de donnée climatiques séquentielles), la géologie (modélisation de la dépendance entre la composition d'échantillon de sol et leur localisation), la médecine (analyse d'électroencéphalogramme, d'électrocardiogramme, etc ...).

Trois approches mathématiques définissent l'utilisation actuelle des processus stationnaires : l'analyse spectrale, liée à la transformation de Fourier et largement implantée par N. Wiener ; l'utilisation des représentations markoviennes, la modélisation par processus autorégressifs et moyenne mobile.

Le but est de modéliser des séries de données (temporelles). Pour pouvoir se servir ensuite de ces modèles pour faire des prédictions, il faut que le comportement du processus, en particulier la dépendance entre passé et futur ait une certaine invariante, ou en d'autres termes qu'il y ait une "constance" dont on puisse tirer parti ; il faut que le processus se "reproduise" au cours du temps ce qui peut ailleurs avoir plusieurs sens mathématiques différents. Les processus stationnaires sont une classe plus large qui conserve une propriété du même genre.

L'analyse statistique d'une série chronologique peut être simplement de nature descriptive. Il s'agit alors de dégager des éléments de synthèse, en général sous forme de nouvelles séries, qui résument au mieux la grandeur observée ou d'exhiber certaines caractéristiques. À cette fin on a recours à des modèles probabilistes qui, le plus souvent, sont de nature paramétrique. C'est en particulier le cadre idéal pour faire de la prévision, qui reste un des objectifs principaux de l'étude d'une série chronologique.

Plusieurs travaux sont consacrés aux méthodes dans lesquelles la partie structurée de la grandeur étudiée est déterministe et représente la moyenne des variables aléatoires observées ou d'une transformation simple de celles-ci. La partie stochastique est alors une erreur additive et forme une séquence de variables aléatoires centrées, non corrélées

et de même variance.

L'étude des séries chronologiques ne se limite évidemment pas à cette première approche. La deuxième étape consiste à supposer que les observations, après transformation éventuelle, sont issues d'une série stationnaire au second ordre. La structure du phénomène se traduit donc par les corrélations entre les variables. Les modèles autorégressifs moyennes mobiles (ARMA) constituent la base de cette seconde approche.

Loin d'être exhaustif, ce travail a pour seule ambition d'en pointer quelques éléments. Cette étude est divisée en trois chapitres.

Le premier chapitre indique quelques généralités sur les séries temporelles. Ceci est fait surtout à l'aide de plusieurs exemples atypiques sur un niveau descriptif, du cadre général de la modélisation de séries temporelles. Par suite nous identifierons les composantes de cette série. On définit la "stationnarité", et on présente quelques exemples de telles séries et les propriétés de leurs moyennes et leurs fonctions de covariance. Plus encore dans ce chapitre on étudiera les notions des polynômes retard et avance qui seront très utiles dans les autres chapitres, ainsi l'invisibilité des polynômes en  $L$ .

Au second chapitre, on considère le théorème fondamental de l'analyse des séries stationnaires, le théorème de "Wold". D'une part, on s'intéresse plus particulièrement aux processus stationnaires (AR(p), MA(q) et ARMA(p, q)) donnera ses définitions et leurs propriétés ainsi ses caractérisations qui jouent un rôle très important dans la suite.

On illustre enfin, dans le dernier chapitre les processus non stationnaires (TS, DS), et plus précisément, processus ARIMA et SARIMA. Ces derniers représentent une classe très complexe, permettent de traiter les saisonnalités par différenciation. On modélise cette classe de processus, en utilisant la méthode "Box & Jenkins". Pour terminer on propose quelques applications et simulations avec le logiciel "R" sur ce modèle.

# Chapitre 1

## Généralité sur les séries chronologiques (time series)

### 1.1 Définitions et Exemples

**Définition 1.1.1.** *Un processus stochastique est une famille de variables aléatoires indiquées par le temps. Définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs réelles pour un processus réel, à valeurs dans  $\mathbb{R}^K$  pour un processus réel vectoriel. Par exemple : la position d'une mouche aux instants 0, 1, 2, 3...*

*Une série chronologique (ou temporelle) est la réalisation d'un processus aléatoire indiqué par le temps  $\{X_t, t \in I\}$ .*

*Une série chronologique donc est une suite d'observations indiquées par le temps, la date à la quelle l'observation est faite est une information importante sur le phénomène observé.*

*En d'autre mots tout ce qui est chiffrable et varie en fonction du temps.*

**Objectif principal :** *étudier les liaisons entre elles  $\{X_t, t \in I\}$  (ou  $\{X_t\}_{t \in I}$ ) où  $I$  est un intervalle de temps qui peut être discret, (dans ce cas,  $t = 1, 2, \dots, T$  où  $T$  est le nombre d'observations ( $I \subseteq \mathbf{N} =$  nombres entiers), à partir des observations équidistances, répétées avec certains fréquence par exemple une fois par (an, mois, jour, etc) ou continu ( $I \subset \mathbb{R}, \mathbb{R}^+$ ), par exemple variable observées au moyen d'appareils physiques ( températures ,...)*

*Une série temporelle peut concerner des données :*

- ▷ *macroéconomiques (Masse monétaire, PIB, inflation,...),*
- ▷ *microéconomiques (nombre d'employés d'une entreprise, ventes, ...),*
- ▷ *politiques (nombre de votants, nombre de votes nuls,...),*
- ▷ *financières ( Indice BRVM composite, cours d'une action,...).*

etc...

Les séries temporelles sont généralement représentées sur des graphiques de valeurs (ordonnées) en fonction du temps (abscisses). Une telle observation constitue un outil qui permet au modélisateur ayant un peu d'expérience de ce rendre compte tout de suite des propriétés dynamiques principales ; on peut aussi repérer les valeurs atypiques ou aberrantes.

### Exemple 1.1.2.

(1) *En Economie : Evolution d'indices (INSEE, chambre de commerce, bourse, ...)*

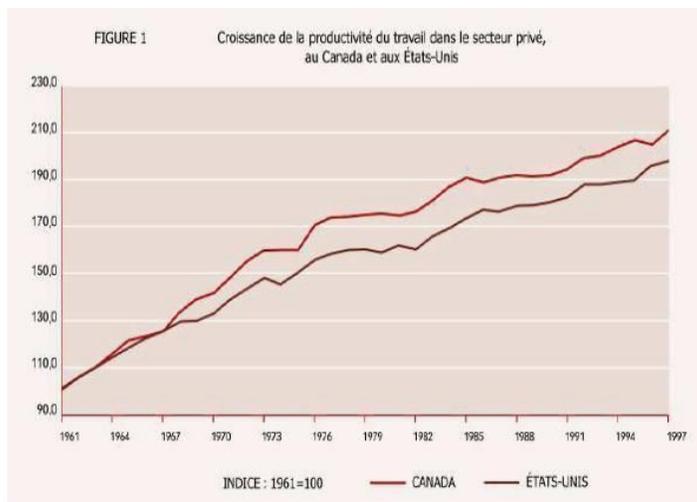


FIGURE 1.1 – Croissance de la productivité du travail dans le secteur privé, au Canada et aux États-Unis

(2) *En Démographie : Population urbaine, rurale, d'un pays, comportement des familles : mariages, naissances.*

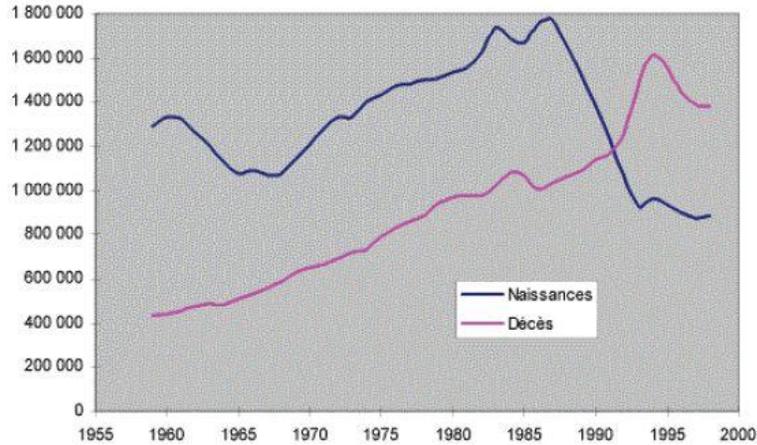


FIGURE 1.2 – Naissances et Décès en Russie depuis 1959

(3) *En Médecine : Battements cardiaques.*

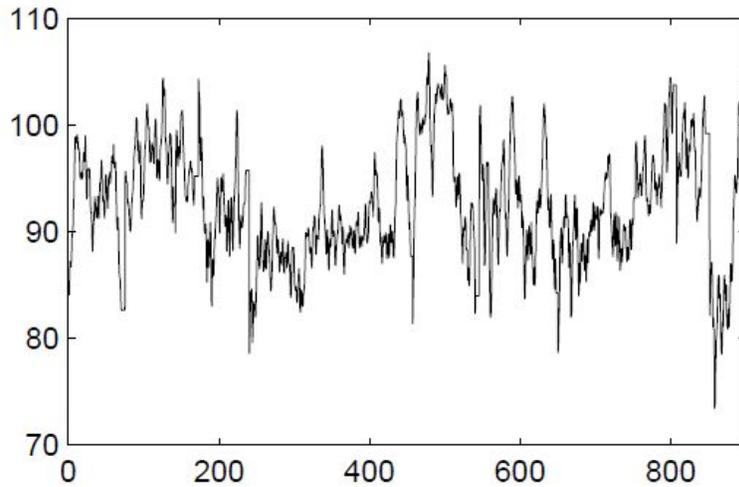


FIGURE 1.3 – Évolution du nombre de battements par minute en fonction du temps mesuré en seconde.

(4) En Météorologie : Pluie, température, débit des cours d'eaux,

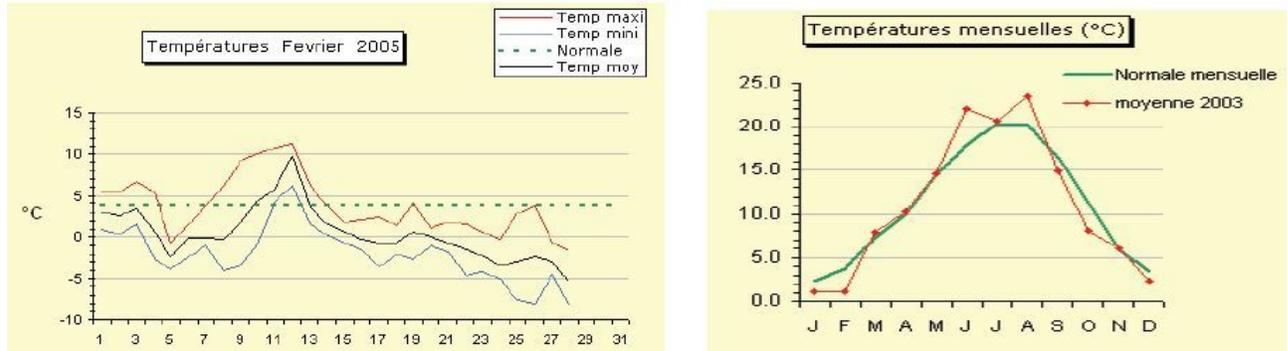


FIGURE 1.4 – Températures

## 1.2 Les composantes d'une série temporelle

Lorsqu'on souhaite analyser une série chronologique présentée sous la forme d'un tableau de chiffres, il faut tout d'abord représenter la série sous la forme d'un graphe. Une série chronologique peut également présente ces deux comportements en même temps. Bien entendu, cette tendance ou cette saisonnalité ne sont pas toujours faciles à déceler. Il peut même arriver que des séries ne contiennent aucune de ces composantes déterministes. Il se peut aussi que le trend n'apparaisse qu'après avoir transformé les données par une fonction, par exemple, logarithmique.

Les quatre composantes d'une série temporelle sont :

**Tendance(trend) :**

**La tendance** est traduit l'évolution globale du phénomène, elle décrit le mouvement à long terme de la série qui peut être croissante, décroissante ou stable.

**Quelques Modèles de tendance :** L'analyse graphique de la série doit permettre de se déterminer en faveur d'une courbe dont l'équation donnera un ajustement paramétrique de la tendance. Il est donc important de savoir caractériser le type de courbe en présence

**Droite**

On parle alors d'ajustement lineaire.

$$f_{\theta(t)} = at + b$$

avec  $\theta = (a, b)$

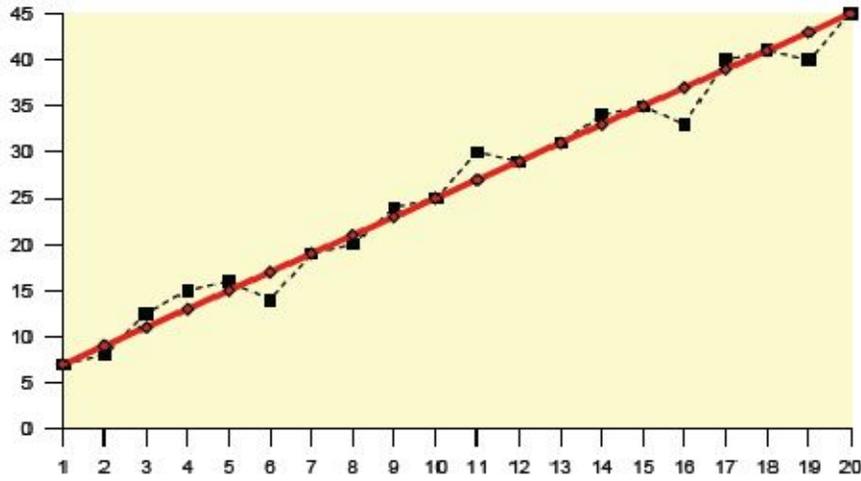


FIGURE 1.5 – Tendence linéaire

### Courbe polynômiale

$$f_{\theta(t)} = a_p t^p + a_{p-1} t^{p-1} + \dots + a_1 t + a_0$$

$$\text{avec } \theta = (a_0, a_1, \dots, a_p)$$

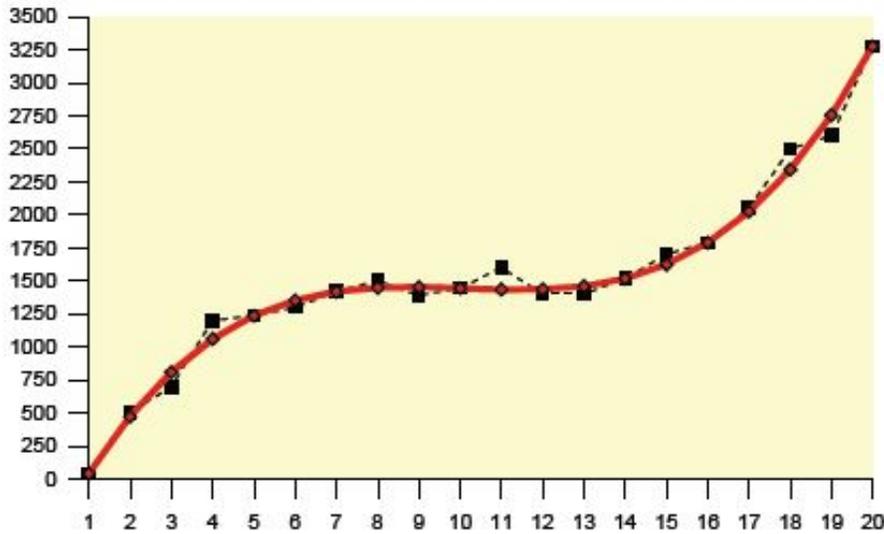


FIGURE 1.6 – Tendence polynômiale

### Courbe exponentielle

$$f_{\theta(t)} = a \exp(bt) + c$$

$$\text{avec } \theta = (a, b, c)$$

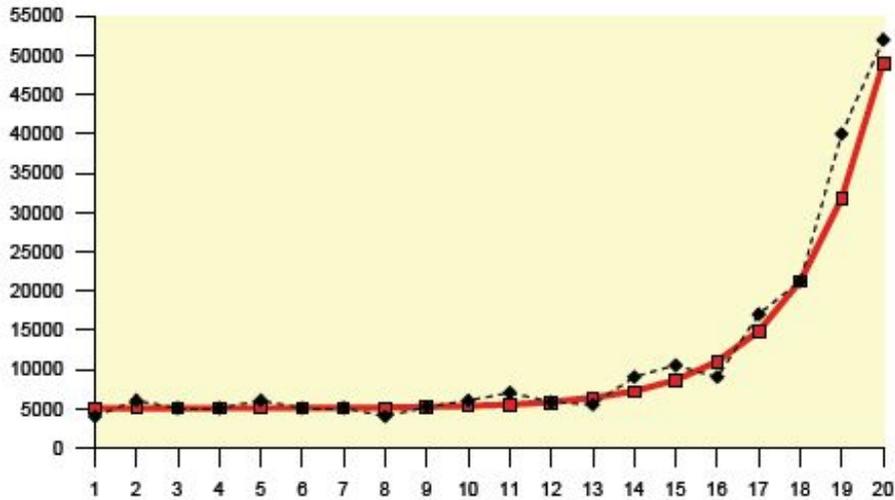


FIGURE 1.7 – Tendance exponentielle

### *Saisonnalités :*

La variation **saisonnaire** notée  $S_t$  représente les fluctuations périodique dues au rythme des saisons, a des facteurs humains (fêtes, jours fériés,...). Les conditions climatiques sont classées parmi les facteurs responsables des variations saisonnières ; elles représentent d'ailleurs le facteur le plus important. Le nombre de mises en chantier, les ventes de mazout et de crème solaire, la production agricole sont autant de variables influencées par la température. Cependant, la température n'est pas la seule responsable des variations saisonnières ; il faut aussi mentionner les coutumes propres à la population ainsi que les fêtes religieuses.

### *Les fluctuations cycliques :*

À l'instar des variations saisonnières, **les fluctuations cycliques** sont périodiques et caractérisent généralement un mouvement oscillatoire. Contrairement aux variations saisonnières, les cycles d'affaires s'étendent habituellement sur une période de plusieurs années et peuvent ne pas suivre exactement des modèles identiques après des intervalles égaux. Chaque nouveau cycle est influencé par des facteurs différents. De grands efforts ont été faits pour analyser les périodes de prospérité, de stabilité, de récession et de

dépression afin d'en arriver à un modèle qui décrit d'une façon satisfaisante un cycle complet. Malheureusement, il y a probablement à ce jour, autant de théories explicatives qu'il y a eu de cycles et aucun modèle mis au point pour expliquer et prévoir les fluctuations cycliques ne s'est avéré jusqu'à maintenant tout à fait satisfaisant.

### *Les variations irrégulières :*

On retrouve souvent dans les séries chronologiques des mouvements qui apparaissent irrégulièrement et généralement durant de courtes périodes. Ces mouvements ne suivent pas de modèle particulier et sont imprévisibles. **Les variations irrégulières** sont souvent attribuées à des facteurs aléatoires ou à des événements particuliers tels que les grèves, les guerres, les élections, pour ne nommer que quelques-uns. En pratique, toutes les variations d'une série temporelle qui ne peuvent être attribuées à l'influence cyclique, saisonnière ou à celle du trend sont classées dans les variations irrégulières.

## 1.3 Quelques types de décomposition

Après avoir détecté graphiquement quelles sont les composantes présentes, il faut proposer un modèle :

### *Le modèle additif*

$$X_t = Z_t + S_t + \epsilon_t$$

$\{X_t, t = 1, 2, \dots, T\}$  est la série observée,  $Z_t$  la tendance,  $S_t$  le facteur saisonnier et  $\epsilon_t$  la composante aléatoire (ou irrégulière, ou fluctuation résiduelle). Par convention, on impose à  $S_t$  d'avoir une moyenne nulle, pour séparer de façon non ambiguë la tendance et le facteur saisonnier :

$$\forall 1 \leq t \leq n - p \quad \sum_{k=1}^p S_k = 0$$

où  $p$  est la "période" de la fluctuation saisonnière. On impose également à  $\epsilon_t$  d'avoir une moyenne nulle, la moyenne de  $X_t$  étant alors également comprise dans la tendance  $Z_t$ .

Dans ce modèle, les amplitudes des variations saisonnières et de la composante irrégulière à un instant donné ne dépendent pas de la valeur de la composante tendance au même instant.

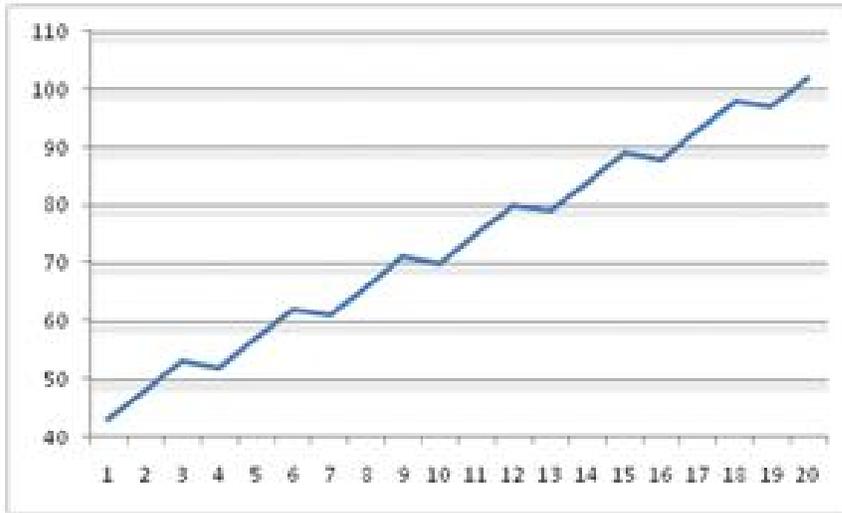


FIGURE 1.8 – Modèle Additif : Ventes d'un produit

### *Le modèle multiplicatif*

$$X_t = Z_t(1 + S_t)(1 + \epsilon_t)$$

où  $X_t (t = 1, 2, \dots, T)$  est la série observée,  $Z_t$  la tendance,  $(1 + S_t)$  le facteur saisonnier et  $(1 + \epsilon_t)$  la composante aléatoire (ou irrégulière, ou fluctuation résiduelle). Cette formulation conduit aux propriétés suivantes :

le facteur saisonnier ( $Z_t S_t$ ) qui s'ajoute à la tendance est proportionnel à la valeur de celle-ci, la composante irrégulière ( $Z_t(1 + S_t)\epsilon_t$ ) qui s'ajoute à la somme des deux termes précédents est elle-même proportionnelle à cette somme.

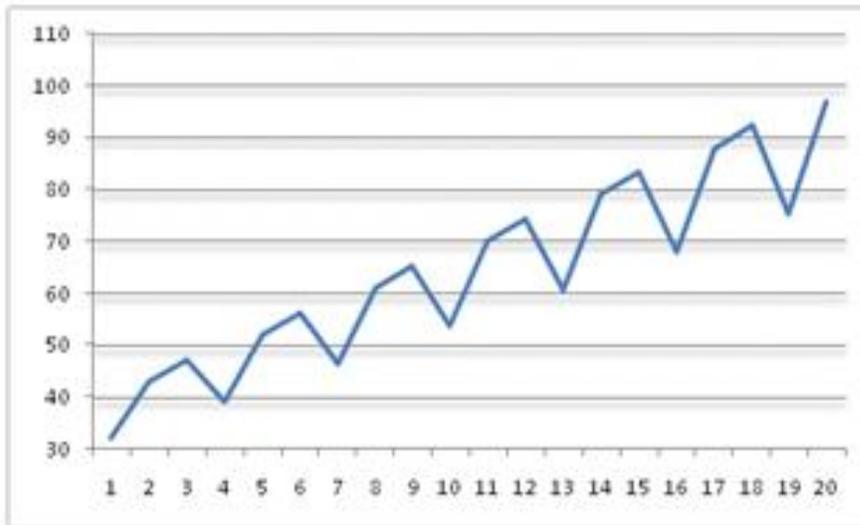


FIGURE 1.9 – Modèle Multiplicatif : Ventes d'un produit

### *Les modèles mixtes*

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

$$X_t = Z_t \times S_t + \epsilon_t$$

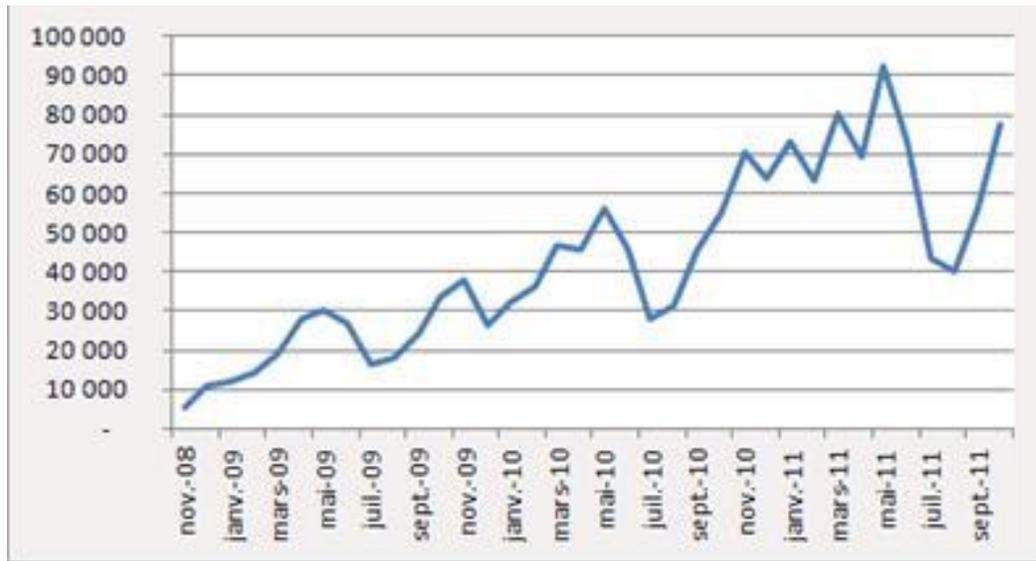


FIGURE 1.10 – Modèle Mixte : Nombre de pages vues par mois

La modélisation stochastique des séries temporelles commence en observant leurs graphiques et en cherchant une décomposition additive ou multiplicative. Nous étudierons en suite le modèle additif (le modèle multiplicatif revient à un modèle additif pour le log des données).

Le choix entre le modèle additif et multiplicatif pour représenter une série temporelle réelle se base sur l'aspect du graphique. Un graphique entre deux parallèles caractérise un modèle additif (voir la figure 1.8). Un graphique qui présente des parties vides qui vont en se rétrécissant, est caractéristique d'un modèle multiplicatif (voir la figure 1.9) ; et quand l'inspection des séries laisse à penser que la variation d'amplitude est constante en valeur ; un modèle additif convient, si la variation est constante en pourcentage de la moyenne annuelle, il est préférable de recourir à un modèle multiplicatif.

## 1.4 Stationnarité

La **stationnarité** joue un rôle central dans la théorie des processus, car elle remplace (de façon naturelle) l'hypothèse d'observation i.i.d en statistique.

### 1.4.1 Définitions

La notion de **stationnarité** est utilisée pour un phénomène dont les caractéristiques essentielles ne se modifient pas au cours du temps.

#### Définition 1.4.2. (*Stationnarité faible*)

$\{X_t, t \in Z\}$  une suite **stationnaire au sens faible** (ou stationnaire du second ordre) lorsque les trois propriétés suivantes sont simultanément vérifiées

$$(i) \quad EX_t = \mu < \infty \quad \forall t \in \mathbb{N}.$$

$$(ii) \quad EX_t^2 < \infty \quad \forall t \in \mathbb{N}.$$

$$(iii) \quad Cov(X_t, X_{t+h}) = Cov(X_{t-1}, X_{t-1+h}) = \dots = Cov(X_0, X_h) \quad \forall h, t \in \mathbb{N}.$$

Dans cette définition, la propriété (i) exprime la stationnarité en moyenne de la suite, (ii) assure que la variance de chaque variable reste finie, et (iii) précise ce qu'on entend par "invariance de la structure de covariance". Par cette propriété, on peut introduire la suite

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

qui est indépendante de  $t$ . Cette suite est la suite des autocovariances (acv) de  $\{X_t\}$  et n'est bien définie que si la suite est stationnaire du second ordre, donc respecte (iii).

Le graphe de cette fonction est appelé **variogramme**

**Propriété 1.4.1.** La fonction d'autocovariance d'un processus stationnaire vérifie :

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

**Définition 1.4.3. (Stationnarité stricte ou forte)**

La suite  $\{X_t, t \in \mathbb{Z}\}$  est **stationnaire au sens fort** (ou stricte) si

$$\mathcal{L}(X_{t+h_1}, X_{t+h_2}, \dots, X_{t+h_k}) = \mathcal{L}(X_{h_1}, X_{h_2}, \dots, X_{h_k}) \quad \forall k \in \mathbb{Z}$$

**Lemme 1.4.4.** [2] Si  $\{X_t\}$  est fortement stationnaire et  $EX_t^2 < \infty$ , alors  $\{X_t\}$  est faiblement stationnaire. La réciproque est fautive en général. Toute fois, si  $\{X_t\}$  est gaussien, les concepts du stationnarité faible et forte coïncident.

**Convention :** on dira qu'une suite est stationnaire si elle est faiblement stationnaire

**1.4.5 Exemples (Processus stationnaires)**

- **Bruit blanc (White noise) :** On appelle bruit blanc faible  $\epsilon_t$  un processus aléatoire stationnaire du second ordre, centré, de fonction d'autocovariance ,  $\gamma(s, t) = Cov(\epsilon_t, \epsilon_s) = \sigma^2 \delta_{t,s}$ . On le notera  $\{X_t\} \rightsquigarrow \mathcal{BB}(0, \sigma^2)$  .
- **Bruit blanc fort :** On appelle bruit blanc fort une suite du second ordre de variables aléatoires  $\{X_t\}$ , centrées, indépendantes et identiquement distribuées (i.i.d.) de variance  $\mathbb{E}[X_t^2] = \sigma^2 < \infty$ . On le notera  $\{X_t\} \rightsquigarrow IID(0, \sigma^2)$ .

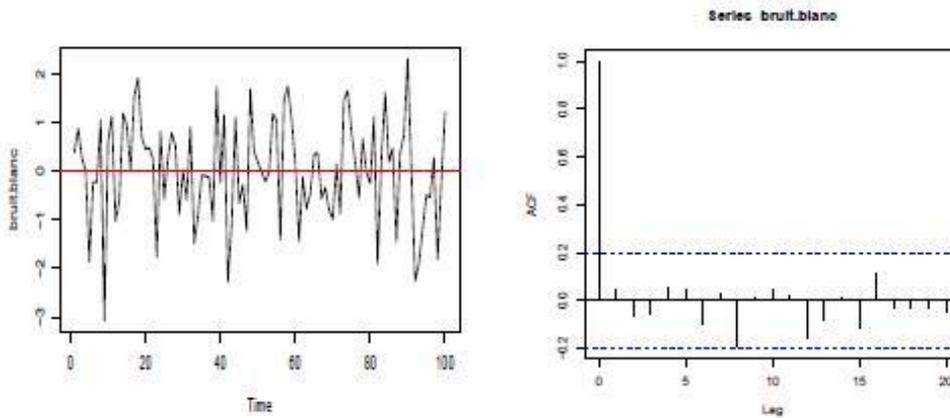


FIGURE 1.11 – Simulation d'un Bruit blanc et sa fonction d'autocorrélation.

### 1.4.6 Définition (*La fonction d'autocorrélation*)

La fonction d'**autocorrélation** est la fonction de  $\mathbb{Z}$  dans  $\mathbb{R}$  notée  $\rho(h)$  qui mesure la corrélation de la série avec elle-même décalée de  $h$  périodes définie par

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

Le graphe de cette fonction est appelé **corrélogramme**

#### **Propriété**

La fonction d'autocorrélation d'un processus stationnaire vérifie

1.  $\rho : h \mapsto \rho(h)$  est une fonction paire ( $\rho(h) = \rho(-h)$ ),
2.  $|\rho(h)| \leq 1$  (à valeurs dans  $[-1, 1]$ ),
3.  $\rho(0) = 1$ .

La fonction  $\rho(h)$  est l'expression du lien linéaire entre  $X_t$  et  $X_{t-h}$ . Si  $t$  est l'instant présent et  $h > 0$ ,  $\rho(h)$  est l'expression du lien linéaire entre le présent et le passé d'ordre  $h$  (mais aussi par parité de la fonction  $\gamma$  entre le présent et le futur d'ordre  $h$ , plus  $|\rho(h)|$  est proche de 1 et plus ce lien est fort).

#### **Définition 1.4.7. (*Auto-corrélogramme*)**

L'*auto-corrélogramme* de  $(X_t)_{t \in \mathbb{Z}}$  est le graphe de :

$$\begin{cases} \mathbb{N} \rightarrow ]-1, 1[ \\ h \mapsto \rho(h). \end{cases}$$

### 1.4.8 Définition (*La fonction d'autocorrélation partielle*)

On appelle **auto-corrélation partielle** d'ordre  $p$  est la fonction notée  $r(p)$  qui mesure la corrélation entre  $X_t$  et  $X_{t-p}$ , l'influence des autres variables  $(X_t, X_{t-1}, \dots, X_{t-p+1})$  ayant été retirée :

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

On montre que  $r(p) = a_p$  coefficient de  $X_{t-p}$  dans  $EL(X_t|X_{t-1}, \dots, X_{t-(p+1)})$ .

$$EL(X_t|X_{t-1}, \dots, X_{t-(p+1)}) = \sum_{j=1}^p a_j X_{t-j}$$

Le graphe de cette fonction est appelé **corrélogramme partiel**

**Proposition 1.4.9.** [2] L'autocorrélation partielle d'ordre  $h$ ,  $r(h)$  représente le coefficient de corrélation linéaire entre

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_{h-1} X_{t-h+1} + U$$

$$X_{t-h} = b_1 X_{t-1} + b_2 X_{t-2} + \dots + b_{h-1} X_{t-h+1} + V$$

et

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

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

**Définition 1.4.10. (Auto-corrélogramme partiel)**

L'auto-corrélogramme partiel de  $(X_t)_{t \in \mathbb{Z}}$  est le graphe de :

$$\begin{cases} \mathbb{N} \rightarrow ]-1, 1[ \\ p \mapsto r(p). \end{cases}$$

### 1.4.11 Polynôme retard et avance

**Définition 1.4.12.**

(1) **Opérateurs retard** : on note  $B$  (**back words**) ou  $L$  (**lag**), l'opérateur qui fait passer de  $x_t$  à  $x_{t-1}$  :

$$L : \{X_t\}_{t \in \mathbb{Z}} \mapsto \{Y_t\}_{t \in \mathbb{Z}} \text{ telle que } Y_t = X_{t-1}$$

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

De la même façon,

(2) **L'opérateurs avance**  $F$  (**forwards**) correspond à

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

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

**Proposition 1.4.13.** [2]

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

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

**Notations :**  $L_0 = Id$  est noté  $L_0 = 1$  i.e ( $L_0 X_t = X_t$ ).

(3) **L'opérateur différence**

La différence première est :

$$\begin{aligned}\nabla X_t &= (1 - L)X_t \\ &= X_t - X_{t-1}.\end{aligned}$$

De la même façon,

$$\begin{aligned}\nabla^2 X_t &= \nabla(\nabla X_t) \\ &= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

(4) **L'opérateur de désaisonnalisation  $\nabla_s$**

L'opérateur  $\nabla_s$  est défini par :

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

En d'autres termes :

$$\nabla_s = 1 - L^s$$

**Les effets des opérateurs de Box-Jenkins**

**L'opérateur  $\nabla$**

1. permet d'éliminer la tendance de la série.
2. peut être répété plusieurs fois si la tendance n'est pas linéaire.

Par exemple :

$$\begin{aligned}\nabla^2 X_t &= (1 - L)^2 X_t \\ &= (1 - 2L + L^2) X_t\end{aligned}$$

permet d'éliminer une tendance quadratique. Le nombre de fois où on applique  $\nabla$  est appelé ordre de différenciation.

### **L'opérateur $\nabla_s$**

1. permet d'éliminer la saisonnalité de période  $s$ .
2. On peut également l'appliquer plusieurs fois.

$$\nabla_s^2 = \nabla_s(X_t - X_{t-s})$$

Le nombre de fois où on applique  $\nabla_s$  est appelé ordre de désaisonnalisation

### **Définition 1.4.14. (Polynôme retard et avance)**

Soit  $P$  un polynôme,  $P(z) = \sum_{k=0}^p a_k z^k$ ,  $a_k \in \mathbb{R}$ , on lui associe le polynôme de retard  $P(L)$  défini comme suit :

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

De façon similaire on obtient le polynôme d'avance  $P(F)$ .

### **Définition 1.4.15. (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 1.4.16.** [2] On suppose  $\sum_{k=0}^{+\infty} |a_k| < +\infty$  et  $\sum_{k=0}^{+\infty} |b_k| < +\infty$  et

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

$$B(L) = \sum_k^{+\infty} b_k L^k.$$

Alors

$$(i) \quad \alpha A(L) = (\alpha A)(L) = \alpha(\sum_k^{+\infty} a_k L^k) = \sum_k^{+\infty} (\alpha a_k) L^k,$$

$$(ii) \quad A(L) + B(L) = (A + B)(L) = (\sum_k^{+\infty} a_k L^k) + (\sum_k^{+\infty} b_k L^k),$$

$$(iii) \quad A(L) \circ B(L) = (AB)(L) = B(L) \circ A(L) \text{ avec } (AB)(L) = (BA)(L) = C(L) \text{ et}$$

$$C(L) = \sum_k^{+\infty} c_k L^k,$$

$$\text{tel que } c_k = \sum_j^{+\infty} a_j b_{k-j}.$$

## 1.5 Inversibilité de $I - \lambda L$

**Proposition 1.5.1.** [2] *L'application  $I - \lambda B$  est **inversible** si et seulement si  $|\lambda| \neq 1$ . Son inverse est alors donnée par*

$$\begin{cases} \sum_{i=0}^{+\infty} \lambda^i L^i & \text{pour } |\lambda| < 1, \\ \sum_{i=1}^{+\infty} \frac{1}{\lambda^i} F^i & \text{pour } |\lambda| < 1. \end{cases} = \begin{cases} \sum_{i=0}^{+\infty} \lambda^i L^i & \text{pour } |\lambda| < 1, \\ -\sum_{i=-1}^{-\infty} \lambda^i L^i & \text{pour } |\lambda| < 1. \end{cases}$$

## 1.6 Inversibilité des polynômes en $L$

### 1.6.1 Définition (*Inversibilité*)

$A(L)$  est *inversible*  $\Leftrightarrow \exists B(L)$  tel que  $A(L) \circ B(L) = Id$

On suppose  $P(L) = \sum_{k=0}^p a^k L^k$  et  $Y_t = P(L)X_t$  et on désire savoir si  $X_t$  peut s'exprimer en fonction de  $Y_t$  donc ( $X_t = P(L)Y_t^{-1}$ ). On peut décomposer notre polynôme de la façon suivante :

$$\begin{aligned}
P(z) &= a_p \prod_{i=1}^p (z - z_i) \\
&= a_p \prod_{i=1}^p (-z_i) \prod_{i=1}^p \left(1 - \frac{z}{z_i}\right) \\
&= a_p \alpha \prod_{i=1}^p (1 - \lambda_i z) \quad \text{avec} \quad \lambda_i = \frac{1}{z_i} \quad \text{et} \quad \alpha = \prod_{i=1}^p (-z_i).
\end{aligned}$$

où les  $z_i \in \mathbb{C}$  sont les racines de  $P$ .

Finalement  $P(L) = \beta \prod_{i=1}^p (1 - \lambda_i L)$  tel que  $\beta = a_p \alpha$

**Proposition 1.6.2.** [2] *Considérons maintenant le cas le plus général de l'opérateur polynomial, soit  $\phi$  un polynôme de degré  $p$  à coefficients réels :*

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p.$$

(i) *L'opérateur  $\Phi(L)$  est **inversible** si et seulement si les racines du polynôme  $\Phi$  sont de module différent de 1.*

(ii) *Si  $|\lambda_j| < 1, \forall j \in [1, p]$ , alors  $\phi(L)$  est inversible et*

$$\phi(L)^{-1} = \sum_{k \geq 0} a_k L^k$$

où

$$a_0 = 1, a_k \in R \quad \text{et} \quad \sum_{k \in \mathbb{Z}} |a_k| < +\infty$$

# Chapitre 2

## Modélisation ARMA des séries chronologiques

### 2.1 Processus linéaires

Un processus linéaire est un processus stochastique  $X_t$  formé par une combinaison linéaire (non nécessairement finie) de bruits blancs forts. On définit également la classe des processus linéaires généraux, qui sont constitués de combinaisons linéaires de bruits blancs faibles.

#### 2.1.1 Définitions et propriétés

**Définition 2.1.2.**  $(X_t)_{t \in \mathbb{Z}}$  est un processus linéaire (resp. linéaire général) de moyenne  $\mu$  s'il peut être écrit sous la forme

$$X_t = \mu + \sum_{-\infty}^{+\infty} a_i \epsilon_{t-i}$$

où  $(\epsilon_t)_{t \in \mathbb{Z}}$  est un bruit blanc fort (resp. faible), de variance  $\sigma^2$  et où la suite des coefficients  $a_i$  est supposée telle que

$$\sum_{-\infty}^{+\infty} |a_i|^2 < +\infty$$

Le résultat suivant montre que les processus linéaires que nous avons définis sont des processus stationnaires.

**Proposition 2.1.3.** [2] Soit  $X = (X_t)_{t \in \mathbb{Z}}$  un processus stationnaire et soit  $(a_i)_{i \in \mathbb{Z}}$  une suite de réels absolument sommable (i.e. telle que  $\sum_{-\infty}^{+\infty} |a_i| < \infty$ ), Alors le processus  $Y = (Y_t)_{t \in \mathbb{Z}}$  défini par

$$Y_t = \sum_{-\infty}^{+\infty} a_i X_{t-i}$$

est stationnaire.

**Proposition 2.1.4.** [2]  $(X_t)_{t \in \mathbb{Z}}$  est un processus linéaire général alors  $(X_t)_{t \in \mathbb{Z}}$  est stationnaire et on a

$$\begin{cases} \gamma(0) = \text{Var}(X_t) = \sigma^2 \sum_{-\infty}^{+\infty} a_i^2 & \text{si } |h| > p \\ \gamma(h) = \text{Cov}(X_t, X_{t+h}) = \sigma^2 \sum_{-\infty}^{+\infty} a_i a_{i+h} & |h| = p. \end{cases}$$

### 2.1.5 Définition (innovation)

Le processus d'innovation du processus  $X_t$  est le processus  $\eta_t$  défini par

$$\eta_t = X_t - \sum_{i=1}^{\infty} \alpha^i X_{t-i} \text{ pour } |\alpha| < 1$$

### 2.1.6 Théorème (Wold)

Pour tout processus stationnaire centré  $X$ , il existe (de manière unique) un processus purement innovant (ou nul)  $Y$  et un processus purement déterministe  $\eta$  non corrélés tels que

$$X = Y + \eta.$$

De plus, la partie purement innovante s'exprime (de manière unique) comme moyenne mobile en fonction du bruit blanc d'innovation  $\epsilon$  de  $X$  :

$$Y_t = \sum_{j>0} c_j \epsilon_{t-j} \quad \forall t,$$

avec  $c_0 = 1$  et  $\sum_{j>0} c_j^2 < \infty$ . Le bruit blanc d'innovation de  $X$  est également bruit blanc d'innovation pour  $Y$ .

### Définition 2.1.7.

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

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{Z}} \phi_i Y_{t-i}$$

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

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{N}} \phi_i Y_{t-i}$$

Le théorème ci-dessous traite le cas fondamental des filtres causaux et de leur inversion.

**Théorème 2.1.7.1.** [2] Soit  $(X_t)_{t \in \mathbb{Z}}$  le processus stationnaire solution de l'équation suivante :

$$\forall t \in \mathbb{Z} \quad \Phi(L)X_t = Y_t,$$

où  $(Y_t)_{t \in \mathbb{Z}}$  est un processus stationnaire et  $\Phi(L)$  le polynôme en  $L$  de la forme

$$\Phi(L) = I - \phi_1 L - \dots - \phi_p L^p$$

avec  $p \in \mathbb{N}$ . Alors

1. Si  $\Phi$  n'a pas de racine de module égal à 1, alors il existe **une représentation inversible** du processus  $(X_t)_{t \in \mathbb{Z}}$ , c'est-à-dire qu'il existe une suite  $(\phi_i, i \in \mathbb{Z})$  telle que

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{Z}} \phi_i Y_{t-i}$$

2. Si de plus toutes les racines de  $\Phi$  sont de module supérieur à 1, alors il existe **une représentation causale** du processus  $(X_t)_{t \in \mathbb{Z}}$ , c'est-à-dire qu'il existe une suite  $(\phi_i, i \in \mathbb{Z})$  telle que

$$\forall t \in \mathbb{Z}, \quad X_t = \sum_{i \in \mathbb{N}} \phi_i Y_{t-i}$$

Autrement dit, dans le premier cas, on garantit que  $X_t$  puisse s'écrire en fonction du processus  $(Y_t)_{t \in \mathbb{Z}}$  et dans le second cas, que  $X_t$  puisse s'écrire en fonction seulement du passé de  $Y_t$ . Si  $\Phi$  a au moins une racine de module égale à 1, alors il n'existe pas de représentation inversible. Le lemme suivant permet de bien comprendre l'importance du rôle des racines de  $\Phi$  dans le cas particulier où  $\Phi(L) = I - \lambda L$ .

**Lemme 2.1.8.** Soit  $(X_t)_{t \in \mathbb{Z}}$  le processus stationnaire avec  $\Phi$  de la forme

$$\Phi(L) = I - \lambda L.$$

Alors il existe deux cas possibles :

1. Si  $|\lambda| < 1$ , alors il existe une représentation causale du processus  $(X_t)_{t \in \mathbb{Z}}$ .
2. Si  $|\lambda| > 1$ , alors il existe une représentation inversible du processus  $(X_t)_{t \in \mathbb{Z}}$ .

## 2.2 Processus auto-régressifs (AR : AutoRegressive) (Yule 1927)

### 2.2.1 Définition et structure

Ces processus forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps  $t$  s'explique linéairement par les observations précédentes.

**Définition 2.2.2.**  $(X_t)_{t \in \mathbb{Z}}$  est un processus  $AR(p)$  si

1.  $(X_t)$  est stationnaire,
2.  $(X_t)$  vérifie :  $\forall t \in \mathbb{Z}, X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$  avec  $\phi_p \neq 0$  et  $\epsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$ .

On note  $\Phi(L)X_t = \mu + \epsilon_t$  où  $\Phi(L) = 1 - (\phi_1 + \dots + \phi_p)$ .

**Exemple 2.2.3.** On dit que la série  $X_t$  suit un processus autorégressif d'ordre 1 ( $X_t \rightsquigarrow AR(1)$ ) si on peut écrire :

$$\begin{aligned} X_t &= \phi X_{t-1} + \epsilon_t \\ X_t - \phi X_{t-1} &= \epsilon_t \\ (1 - \phi L)X_t &= \epsilon_t. \end{aligned}$$

où la série est  $\epsilon_t$  un bruit blanc.

**Proposition 2.2.4.** [8] Si  $(X_t) \rightsquigarrow AR(p)$  tel que  $\Phi(L)X_t = \mu + \epsilon_t$ , alors

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

**Proposition 2.2.5.** [2] Si  $(X_t) \rightsquigarrow AR(p)$  tel que  $\Phi(L)X_t = \mu + \epsilon_t$  et si l'on pose  $Y_t = X_t - m$  (où  $m = \mathbb{E}X_t$ ), on a alors

$$\Phi(L)Y_t = \epsilon_t \text{ et } \mathbb{E}Y_t = 0.$$

**Proposition 2.2.6.** [15]

Si  $(X_t) \rightsquigarrow AR(p)$  tel que  $\Phi(L)X_t = \mu + \epsilon_t$ , alors

1.  $\bar{\mathcal{L}}(X_t) = \bar{\mathcal{L}}(\epsilon_t)$
2.  $\epsilon_t$  est l'innovation de  $X_t$

**Rappel de notation**

$$\bar{\mathcal{L}}(X_t) = \bar{\mathcal{L}}(1, X_t, X_{t-1}, \dots, X_{t-p}, \dots)$$

$$\bar{\mathcal{L}}(\epsilon_t) = \bar{\mathcal{L}}(1, \epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-p}, \dots)$$

**Définition 2.2.7. (représentation canonique)**

Soient  $(X_t) \rightsquigarrow AR(p)$  et  $\Phi$  un polynôme vérifiant :

- i.  $\Phi(L)X_t = \mu + \epsilon_t$
- ii.  $|z| \leq 1 \Leftrightarrow \Phi(z) \neq 0$

On dit que la représentation  $\Phi(L)X_t = \mu + \epsilon_t$  est la **représentation canonique** de  $(X_t)_{t \in \mathbb{Z}}$

**Théorème 2.2.7.1. (Existence des processus AR(p))**

Soit l'équation récurrente :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (2.1)$$

où  $\epsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$ . (2.1) admet une solution stationnaire au second ordre si et seulement si le polynôme

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \text{ pour } |z| = 1.$$

et cette solution est unique.

Elle a pour expression :

$$X_t = \sum_{k=-\infty}^{+\infty} \psi_k \epsilon_{t-k}$$

où  $\psi_k$  est la suite des coefficients du développement en série de Laurent de  $1/\Phi(z)$  au voisinage du cercle unité.

### 2.2.8 Ecriture moyenne mobile infinie d'un AR(p)

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

**Théorème 2.2.8.1.** [2]

1. Si  $\Phi$  n'a pas de racine de module égal à 1,  $\Phi(L)$  est inversible et on en déduit que l'équation  $\Phi(L)X_t = \eta_t, t \in \mathbb{Z}$ . a une solution unique, avec une écriture moyenne mobile infinie

$$\forall t \in \mathbb{Z}, \quad X_t = (\Phi(L)^{-1}) = \sum_{i \in \mathbb{Z}} \phi_i \eta_{t-i}$$

avec  $\sum_{i=-\infty}^{+\infty} \phi_i < \infty$

Les coefficients  $\phi_i$  convergent rapidement vers 0 lorsque  $i \rightarrow \infty$ , en effet ce sont les coefficients du filtre associé au polynôme  $1/\Phi$ .

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

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

$$\forall t \in \mathbb{Z}, \quad X_t = (\Phi(L)^{-1}) = \sum_j^{+\infty} \phi_j \eta_{t-j}$$

avec  $\sum_{j=0}^{+\infty} \phi_j < \infty$

C'est-à-dire que le processus  $X_t$  s'exprime en fonction de  $\eta_s, s \leq t$  et d'après la définition, on voit que  $\eta_t$  n'est pas corrélé avec  $X_{t-1}, X_{t-2}, \dots$

La variable  $\eta_t$  est donc l'innovation du processus à la date  $t$  et  $\sum_j^{+\infty} \phi_j \eta_{t-j}$  est la régression affine de  $X_t$  sur  $(X_s, s \leq t-1)$  ; il s'agit de la représentation canonique d'un processus AR(p).

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

**Théorème 2.2.9.1.** [2]

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

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

### 2.2.10 Propriétés des processus AR(p)

On suppose que  $\Phi(L)X_t = \mu + \epsilon_t$  où

les racines de  $\Phi$  sont de module strictement supérieur à 1,

$\epsilon_t$  suit un bruit blanc.

#### Proposition 2.2.11. [2]

On peut écrire que

$$X_t = \sum_{j=0}^{+\infty} \phi_j \epsilon_{t-j}$$

#### L'auto-covariance

est donnée par :

$$\begin{aligned} \gamma(h) &= \text{Cov}(X_t, X_{t+h}) = \lim_{n \rightarrow \infty} \mathbb{E} \left[ \left( \sum_{j=0}^n \phi^j \epsilon_{t-j} \right) \left( \sum_{k=0}^n \phi^k \epsilon_{t+h-k} \right) \right] \\ &= \sigma^2 \phi^{|h|} \sum_{j=0}^n \phi^{2j} \\ &= \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2} \end{aligned}$$

(série convergente car  $|\phi| < 1$ )

et donc

$$\text{Var}(X_t) = \gamma(0) = \frac{\sigma^2}{1 - \phi^2} > \sigma^2.$$

#### L'auto-corrélations :

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

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

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

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

**Proposition 2.2.12.** [2]  $r(h)$  est le coefficient de  $X_{t-h}$  dans  $EL(X_t | X_{t-1}, \dots, X_{t-h+1})$ .

**Proposition 2.2.13.** [8] Si  $(X_t) \rightarrow AR(p)$ , alors :

$$\begin{cases} 0 & \text{si } h > p \\ \neq 0 & \text{sinon.} \end{cases}$$

**Proposition 2.2.14.** [2] Si  $X_t \rightsquigarrow AR(p)$  alors les  $\rho(h)$  et les  $\gamma(h)$  décroissent vers 0 exponentiellement avec  $h$ .

### Les équations de Yule-Walker

Les équations de Yule-Walker fournissent une relation linéaire entre les paramètres  $\phi_1, \dots, \phi_p$  et  $\sigma^2$

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

$$1 = \varphi_1 \rho(1) + \dots + \varphi_p \rho(p) + \frac{\sigma_\epsilon^2}{\gamma(0)}$$

$$\Rightarrow \gamma(0) = \sigma_\epsilon^2 \frac{1}{1 - \varphi_1 \rho(1) + \dots + \varphi_p \rho(p)}$$

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

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

Les solutions de l'équation de récurrence sont complètement déterminées par la donnée de condition 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  $\hat{\rho}_T(1), \dots, \hat{\rho}_T(p)$ .

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

### 2.2.15 Exemple AR(1)

#### ACF et PACF d'un AR(1)

Prenons une série  $X_t$  suivant un processus autorégressif d'ordre 1.

Par exemple :

$$X_t = 0.9X_{t-1} + \epsilon_t$$

**ACF(1)** La corrélation entre  $X_t$  et  $X_{t+1}$  peut se calculer facilement

$$\begin{aligned} ACF(1) &= Cor(X_{t+1}, X_t) \\ &= Cor(0.9X_t + \epsilon_t, X_t) \\ &= 0.9Cor(X_t, X_t) + Cor(X_t, \epsilon_t) \\ &= 0.9 + 0 \\ &= 0.9 \end{aligned}$$

Ceci est vrai car  $X_t$  et  $\epsilon_t$  sont indépendants

**ACF(2)** En écrivant la relation entre une valeur à un instant et la valeur précédente à l'ordre  $t - 1$ , on a :

$$X_{t-1} = 0.9X_{t-2} + \epsilon_{t-1}$$

où

$$X_t = 0.9(0.9X_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

$$X_t = 0.81X_{t-2} + 0.9\epsilon_{t-1} + \epsilon_t$$

Comme  $X_{t-2}$  est indépendant à la fois de  $\epsilon_{t-1}$  et  $\epsilon_t$ , on en déduit facilement que :

$$Cor(X_t, X_{t-2}) = 0.81$$

**ACF(p)** En continuant ce type de raisonnement, on montre que :

$$Cor(X_t, X_{t-k}) = 0.9^k$$

**PACF(1)** L'autocorrélation partielle au décalage 1 est :

$$Cor(X_t, X_{t-k}) = 0.9^k$$

$$\begin{aligned} PACF(1) &= Cor(X_{t+1}, X_t) \\ &= 0.9 \end{aligned}$$

**PACF(2)** Il faut, pour calculer  $PACF(2)$ , faire la régression de  $X_{t+1}$  sur  $X_{t+2}$  et sur  $X_t$ .

On sait déjà que :

$$X_{t+2} = 0.9X_{t+1} + \epsilon_{t+2}$$

Du rappel précédent, on en déduit que, si on fait la régression de  $X_{t+1}$  sur  $X_t$  on obtient :

$$X_t = 0.9X_{t+1} + V_t$$

Pour le calcul de  $PACF(2)$ , on calcule la corrélation entre les 2 résidus de ces régressions :

$$\begin{aligned} PACF(2) &= Cor(X_{t+2} - 0.9X_{t+1}, X_t - 0.9X_{t+1}) \\ &= Cor(\epsilon_{t+2}, X_t - 0.9X_{t+1}) \\ &= Cor(\epsilon_{t+2}, X_t) - 0.9Cor(X_{t+1}, \epsilon_{t+2}) \\ &= 0 \end{aligned}$$

**PACF(k)** Par de calculs similaires on peut montrer que :

$$PACF(k) = 0 \text{ pour tout } k > p$$

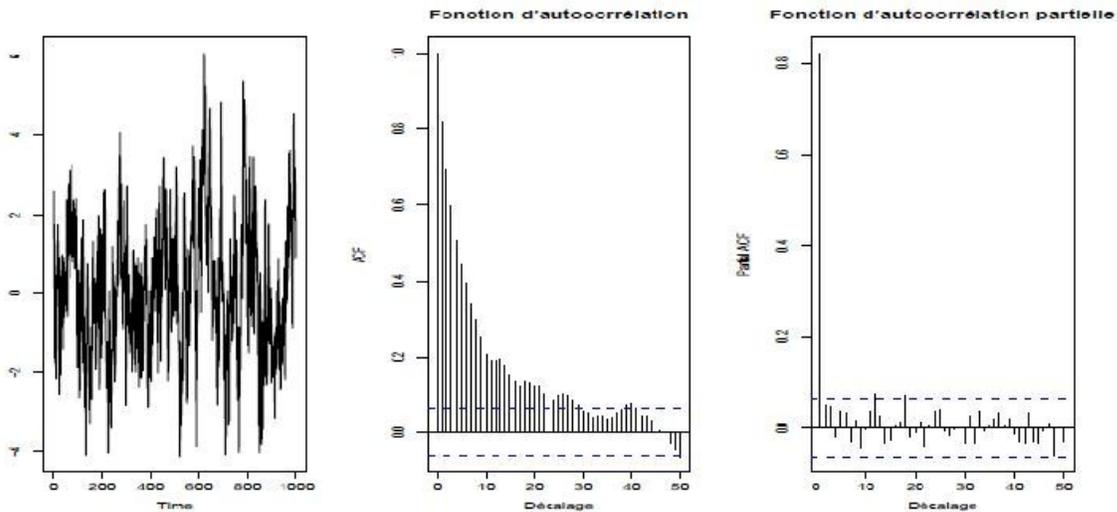


FIGURE 2.1 – Simulation d'un  $AR(1)$  :  $X_t = 0.8X_{t-1} + \epsilon_t$

### Prévision d'un AR(p)

Les  $AR(p)$  sont bien adaptés à la prédiction. Les prévisions optimales  $\hat{X}_t(h)$  pour  $h > 0$  sont des combinaisons linéaires de  $X_t, X_{t-1}, \dots, X_{t-p+1}$  comme d'habitude. On les calcule de façon récurrente.

– À l'horizon 1 : on a

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

que l'on prédit par

$$\hat{X}_t(1) = \phi_1 X_t + \phi_2 X_{t-1} + \dots + \phi_p X_{t+1-p}$$

– À l'horizon 2 : on a

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

que l'on prédit par

$$\hat{X}_t(2) = \phi_1 \hat{X}_t(1) + \phi_2 X_{t-1} + \dots + \phi_p X_{t+2-p}$$

et ainsi de suite. Ainsi on a le théorème suivant :

**Théorème 2.2.15.1.** [2] Soit  $X$  un  $AR(p)$  de bruit blanc d'innovation  $\eta$  et de polynôme canonique  $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ . Les prévisions optimales  $\hat{X}_t(h)$  pour  $h > 0$  sont des combinaisons linéaires de  $X_t, X_{t-1}, \dots, X_{t-p+1}$  :

$$\hat{X}_t(h) = \sum_{0 \leq j \leq p-1} a_j(h) X_{t-j}$$

qui s'obtiennent par récurrence selon

$$\hat{X}_t(h) = \sum_{0 \leq k \leq p} \phi_k \hat{X}_t(h-k)$$

## 2.3 Processus moyenne mobile(MA :Moving Average)(Slutsky 1927)

### 2.3.1 Définitions et structures

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

**Définition 2.3.2.** On appelle **processus moyenne mobile d'ordre  $q$**  ( $X_t \rightsquigarrow MA(q)$ ) tout processus  $(X_t)_{t \in \mathbb{Z}}$  stationnaire tel que

$$\forall t \in \mathbb{Z}, \quad X_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

où  $\theta_1, \theta_2, \dots, \theta_q$  sont des réels fixés et  $(\epsilon_t)_{t \in \mathbb{Z}}$  est un bruit blanc de variance  $\sigma^2$ . Un tel processus est dit  $MA(q)$  (Moving Average of order  $q$ ).

**Remarque 2.3.3.** Notons immédiatement que

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

Mais attention son bruit blanc innovant n'est pas nécessairement égal à  $\epsilon$ .

On peut étendre cette définition aux  $MA(\infty)$  en faisant croître  $q$ . On pourra alors vérifier que  $X$  est stationnaire ssi

$$\sum_{j \in \mathbb{Z}} \theta_j^2 < \infty.$$

**Définition 2.3.4. (représentation canonique)**

Soit  $(X_t) \rightsquigarrow MA(q)$ , On dit que la représentation

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

est la **représentation canonique**  $AR(\infty)$  de  $(X_t)_{t \in \mathbb{Z}}$ .

**Proposition 2.3.5.** [8] Si  $X_t \rightsquigarrow MA(q)$  alors les  $\rho(h)$  et les  $\gamma(h)$  décroissent vers 0 exponentiellement avec  $h$ .

### 2.3.6 Forme autorégressive infinie d'un MA(q)

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

$$\epsilon_t = \sum_{j \in \mathbb{Z}} \pi_j X_{t-j},$$

avec  $\sum_{i=-\infty}^{+\infty} |\pi_i| < \infty$ .

Dans ce cas, on dit que le processus est *inversible*.

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

$$\epsilon_t = \sum_{j \in \mathbb{N}} \pi_j X_{t-j},$$

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

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

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

avec

$$m_k = \frac{1}{\prod_{i \neq k} (1 - \frac{\mu_i}{\mu_k})}$$

et

$$\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q = \prod_j (1 - \mu_j z)$$

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

**Théoreme 2.3.6.1.** [2] *Si  $\Theta$  s'annule en un point du cercle unité, alors il n'existe aucun choix de coefficients  $\pi_j$  pour lesquels  $\sum_{j \geq 0} \pi_j X_{t-j}$  converge et coïncide avec  $\epsilon_t$ . Cependant, on peut représenter  $\epsilon_t$  comme limite de telles combinaisons linéaires des  $X_u$ ,  $u \leq t$ .*

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

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

**Théoreme 2.3.7.1.** [15] *La relation d'un MA(q) avec son bruit blanc d'innovation est aussi du type  $X_t = \tilde{\Theta}(L)\eta_t, t \in \mathbb{Z}$ . Le polynôme  $\tilde{\Theta}$  de degré q, dit polynôme canonique, s'obtient à partir de  $\Theta$  en remplaçant toutes les racines éventuelles à l'intérieur du cercle unité par leur inversion ( $\mu \rightarrow \frac{1}{\mu}$ ), les déplaçant ainsi à l'extérieur du cercle unité. Notons  $\mu_1, \dots, \mu_q$  les co-racines de  $\Theta(z)$  :*

$$\Theta(z) = \prod_j (1 - \mu_j z) = 1 - \theta_1 z - \dots - \theta_q z^q.$$

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

$$\tilde{\Theta}(z) = \prod_{|\mu_j| > 1} (1 - \mu_j z) \prod_{|\mu_j| < 1} \left(1 - \frac{1}{\bar{\mu}_j} z\right).$$

**Proposition 2.3.8.** [2] *La variance du bruit blanc d'innovation  $\eta$  est donnée par*

$$\text{Var}(\eta_t) = \gamma_\eta(0) = \gamma_\epsilon(0) = \prod \mu_j^2.$$

**A partir de maintenant, nous ne considérerons plus que des processus MA(q) donnés sous leur représentation canonique c.à.d vérifiant l'équation**

$$X_t = \Theta(L)\epsilon_t,$$

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

### 2.3.9 Propriétés d'un MA(q)

#### Exemple MA(1)

Soit  $X$  un MA(1).  $(X_t)$  est donc généré par un bruit blanc  $(\epsilon_t)$  sous la forme

$$X_t = \epsilon_t - \theta \epsilon_{t-1} = (I - \theta L)\epsilon_t.$$

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

1. Par simples calculs, on peut montrer que

$$\text{Var}(X_t) = (1 + \theta^2)\sigma^2 > \sigma^2, \quad \forall t \in \mathbb{Z}.$$

Ce qui signifie qu'en modélisant, on diminue la variance du phénomène ce qui est par nature le propre de toute modélisation

Et plus généralement,

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

2. De même, on peut montrer que

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

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

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

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

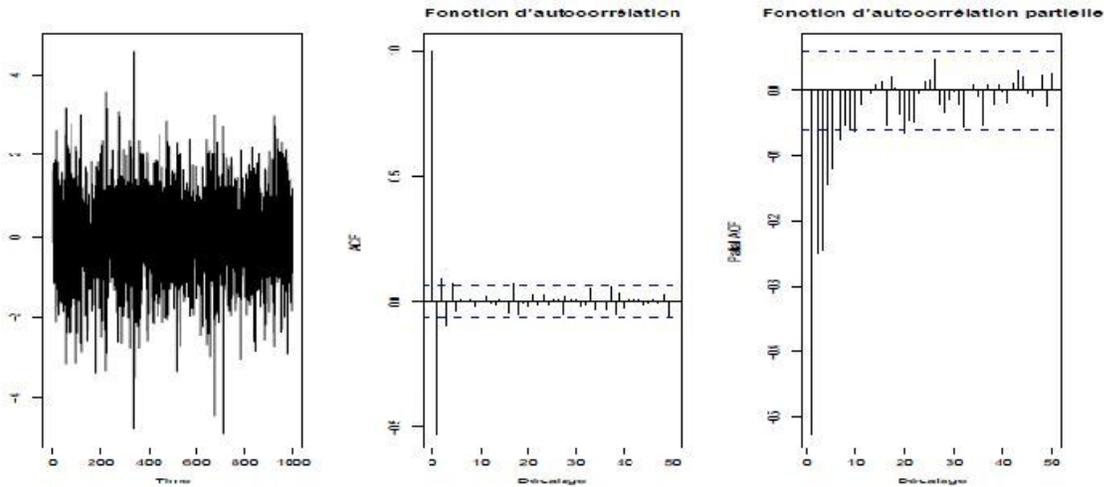


FIGURE 2.2 – Simulation d'un  $MA(1)$  :  $X_t = \epsilon_t - 0.8\epsilon_t$

**Proposition 2.3.10.** [15]

1. La variance de  $X_t$  est donnée par

$$\text{Var}(X_t) = \gamma(0) = (1 + \theta_1^2 + \dots + \theta_q^2)\gamma_\epsilon(0) > \gamma_\epsilon(0) = \sigma^2, \quad \forall t$$

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

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

3. La fonction d'autocorrélation est donnée par

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

En particulier,  $\rho(h) = \frac{-\theta_q}{1+\theta_1^2+\dots+\theta_q^2}$  ne s'annule que si  $\theta_q = 0$ , ce qui revient à dire que le processus n'est pas d'ordre  $q$ .

**Proposition 2.3.11.** [2] Soit  $(X_t)$  un processus linéaire stationnaire corrélé d'ordre  $q$ , c'est à dire  $\gamma(h) = 0$  pour tout  $|h| > q$ . Alors  $(X_t)$  possède une représentation comme processus  $MA(q)$ .

**Proposition 2.3.12.** [2] La somme d'un  $MA(q_1)$  et d'un  $MA(q_2)$  (non corrélés) est un  $MA(\leq \max(q_1, q_2))$ .

### 2.3.13 Prédiction d'un $MA(q)$

Soit  $X$  un  $MA(q)$  de polynôme  $\Phi$  et de bruit blanc d'innovation  $\eta$ . La relation qui le lie à son bruit blanc d'innovation  $\eta$  :

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

on a donc

$$\hat{X}_t(1) = -\theta_1\eta_t - \dots - \theta_q\eta_{t-q}$$

$$\hat{X}_t(2) = -\theta_2\eta_t - \dots - \theta_q\eta_{t-q}$$

...

$$\hat{X}_t(q) = -\theta_q\eta_t$$

$$\hat{X}_t(j) = 0, j > q$$

Par conséquent, les  $MA(q)$  ne sont pas bien adaptés à la prédiction puisque d'une part  $j > q \Rightarrow \hat{X}_t(j) = 0$  et d'autre part, les prévisions nécessitent la combinaison de toutes les valeurs du passé de  $X$ .

## 2.4 Processus mixtes ARMA (Wold 1954)

### 2.4.1 Définition et représentation canonique minimale

Nous définissons les séries ARMA qui sont des généralisations directes des deux exemples introductifs, la combinaison des processus autorégressifs et moyennes mobiles. Cette classe de processus ARMA est encore un cas particulier de processus linéaires et jouera un rôle prépondérant dans la modélisation concrète des processus stationnaires.

**Définition 2.4.2.** On appelle **processus autorégressif moyenne mobile** d'ordre  $(p, q)$   $(X_t \rightsquigarrow \text{ARMA}(p, q))$  tout processus  $(X_t)_{t \in \mathbb{Z}}$  stationnaire tel que

$$\forall t \in \mathbb{Z}, \quad X_t - \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

où  $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$  sont des réels fixés et  $(\epsilon_t)_{t \in \mathbb{Z}}$  est un bruit blanc de variance  $\sigma^2$ . Un tel processus est dit **ARMA**( $p, q$ ) (*AutoRegressive Moving Average of order* ( $p, q$ )).

Posons comme précédemment

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

et

$$\Theta(L) = I - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q.$$

On peut alors écrire

$$\forall t \in \mathbb{Z} \quad \Phi(L)X_t = \Theta(L)\epsilon_t.$$

**Définition 2.4.3.** Un processus stationnaire  $(X_t)_{t \in \mathbb{Z}}$  admet une représentation **ARMA**( $p, q$ ) canonique minimale s'il vérifie une équation :

$$\forall t \in \mathbb{Z} \quad \Phi(L)X_t = \mu + \Theta(L)\epsilon_t.$$

où

i)  $\epsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$

ii)  $\Phi(L) = I - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ , avec  $\phi_p \neq 0$

iii)  $\Theta(L) = I - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ , avec  $\theta_q \neq 0$

iv)  $\phi$  et  $\theta$  ont toutes leurs racines de module strictement supérieur à 1 (représentation canonique).

v)  $\phi$  et  $\theta$  n'ont pas de racines communes (représentation minimale).

**Remarque 2.4.4.** Il est immédiat qu'un **ARMA**( $p, 0$ ) est un **AR** pur et qu'un **ARMA**( $0, q$ ) est un **MA** pur. Les seuls processus admettant simultanément une représentation **MA** pure et une représentation **AR** pure correspondent au cas **ARMA**( $0, 0$ ), c'est à dire aux bruits blancs.

**Théorème 2.4.4.1. (Existence des processus ARMA( $p, q$ ))**

Soit l'équation récurrente :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

où  $\epsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$ . On pose  $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$  et  $\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q$ . On suppose que  $\Phi(z)$  et  $\Theta(z)$  n'ont pas de zéros communs. Alors cette équation admet une solution stationnaire au second ordre si et seulement si le polynôme  $\Phi(z) \neq 0$  pour  $|z| = 1$ . Cette solution est unique et a pour expression :

$$X_t = \sum_{k=-\infty}^{+\infty} \psi_k \epsilon_{t-k}$$

où  $\psi_k$  est la suite des coefficients du développement en série de Laurent de  $\Theta(z)/\Phi(z)$  au voisinage du cercle unité.

## 2.4.5 Expression d'un ARMA(p, q)

### Représentation moyenne mobile infinie

#### Propriété

i) Si le polynôme  $\Phi$  a toutes ses racines de module différent de 1, l'opérateur  $\Phi(L)$  est inversible et la relation admet une solution stationnaire donnée par

$$\forall t \in \mathbb{Z} \quad X_t = \frac{\Theta(L)}{\Phi(L)} \epsilon_t = \sum_{i \in \mathbb{Z}} \phi_i \epsilon_{t-i} \quad \text{avec} \quad \sum_{i \in \mathbb{Z}} |\phi_i| < +\infty,$$

c'est à dire sous une forme moyenne mobile infinie.

ii) Si de plus les racines du polynôme  $\Phi$  sont de module strictement supérieur à 1, seules les valeurs passées du bruit interviennent dans cette écriture  $MA(\infty)$ . Dans ce cas, les  $\phi_i$  de la représentation causale vérifient  $\phi_0 = 1$  et

$$\phi_k = \begin{cases} -\theta_k + \sum_{i=1}^k \psi_i \phi_{k-i} & \text{si pour } 1 \leq k < \max(p, q+1) \\ \sum_{i=1}^p \psi_i \phi_{k-i} & \text{si pour } k \leq \max(p, q+1). \end{cases}$$

en posant  $\theta_0 = -1, \theta_i = 0$  si  $i > q$  et  $\phi_i = 0$  si  $i > p$ .

Ces équations peuvent se résoudre successivement pour  $\phi_0, \phi_1, \dots$

**Exemple 2.4.6.** On considère  $X$  défini par

$$X_t - X_{t-1} + \frac{1}{4} X_{t-2} = \epsilon_t + \epsilon_{t+1}$$

avec  $\epsilon_t$  un bruit blanc de variance  $\sigma^2$ .

On a  $\psi_1 = 1, \psi_2 = -\frac{1}{4}, \theta_1 = -1, p = 2$  et  $q = 1$ .

En utilisant les équations précédentes, on obtient

$$\phi_0 = 1,$$

$$\phi_1 = -\theta_1 + \psi_1\phi_0 = -\theta_1 + \psi_1 = 2$$

### Représentation autorégressive

Une démarche analogue peut-être suivie pour le polynôme  $\Theta$  :

#### Propriété

i) Supposons que le polynôme  $\Theta$  a toutes ses racines de module différent de 1, l'opérateur  $\Theta(L)$  est inversible et on obtient la forme autorégressive infinie

$$\forall t \in \mathbb{Z} \quad \epsilon_t = \frac{\Theta(L)}{\Phi(L)} X_t = \sum_{i \in \mathbb{Z}} \pi_i X_{t-i} \quad \text{avec} \quad \sum_{i \in \mathbb{Z}} |\pi_i| < +\infty,$$

ii) Si de plus les racines du polynôme  $\Theta$  sont de module strictement supérieur à 1, cette représentation  $AR(\infty)$  ne fait intervenir que les valeurs présentes et passées du processus. Dans ce cas, les  $\pi_i$  de la représentation causale vérifient  $\pi_0 = 1$  et

$$\pi_k = \begin{cases} -\psi_k + \sum_{i=1}^k \theta_i \pi_{k-i} & \text{si pour } 1 \leq k < \max(p+1, q) \\ \sum_{i=1}^k \theta_i \pi_{k-i} & \text{si pour } k \leq \max(p+1, q). \end{cases}$$

### 2.4.7 Bruit blanc d'innovation d'un ARMA(p, q)

Comme pour les AR et les MA, le bruit blanc  $\epsilon$  n'est pas nécessairement le bruit blanc d'innovation :

**Théoreme 2.4.7.1.** [2] Pour que soit le bruit blanc  $\epsilon$  d'innovation de  $X$  il faut et il suffit que

$$\Phi(z) = \prod_{1 \leq k \leq p} (1 - \lambda_k z) \quad \text{et} \quad \Theta(z) = \prod_{1 \leq j \leq q} (1 - \lambda_j z)$$

avec  $\forall k, |\lambda_k| < 1$  et  $\forall j, |\lambda_j| \leq 1$ .

**Théoreme 2.4.7.2.** [2] Si  $X$  est un ARMA(p, q) donné selon  $\Phi(L)X = \Theta(L)\epsilon$ , alors la relation (dite minimale) qui le lie à son bruit blanc d'innovation est aussi du type  $\tilde{\Phi}(L)X = \tilde{\Theta}(L)\eta$ , où les polynômes  $\tilde{\Phi}$  et  $\tilde{\Theta}$  s'obtiennent à partir de  $\Phi$  et  $\Theta$  selon les mêmes règles que pour les AR(p) et les MA(q) puis en supprimant tout facteur commun éventuel.

## 2.4.8 Autocovariances et propriétés d'un ARMA(p, q)

### Propriété

La fonction d'autocovariance d'un processus ARMA(p, q) vérifie pour  $h \geq \max(p, q+1)$ ,

$$\gamma(h) = \sum_{i=1}^p \psi_i \gamma(h-i)$$

et pour  $0 \leq h < \max(p, q+1)$ ,

$$\gamma(h) = \sum_{i=1}^p \psi_i \gamma(h-i) + \sigma^2(\phi_0 - \phi_1 \theta_{h+1} - \dots - \phi_{q-h} \theta_q)$$

### Exemple 2.4.9.

En prenant  $p = 1$  et  $q = 1$ , la série  $X_t$  suit un processus ARMA(1, 1) si on peut écrire :

$$(1 - \phi L)X_t = (1 - \theta L)\epsilon_t$$

Les propriétés d'un ARMA(1, 1) sont :

- Pour que la série  $X_t$  soit stationnaire et inversible (c'est-à-dire qu'on puisse calculer  $\epsilon_t$  en fonction de  $X_t$  il faut que :

$$|\phi| < 1 \text{ et } |\theta| < 1$$

- L'ACF d'un ARMA(1, 1) décroît exponentiellement à partir de  $k = 1$ , alors que l'ACF d'un AR(1) décroît exponentiellement à partir de  $k = 0$ .
- La PACF d'un ARMA(1, 1) ressemble à celle d'un MA(1) à partir de  $k = 2$ .

On peut dire, en gros, que si on ne se retrouve pas d'une façon évidente en face d'un processus AR ou MA, on a de fortes chances de se trouver en face d'un processus ARMA

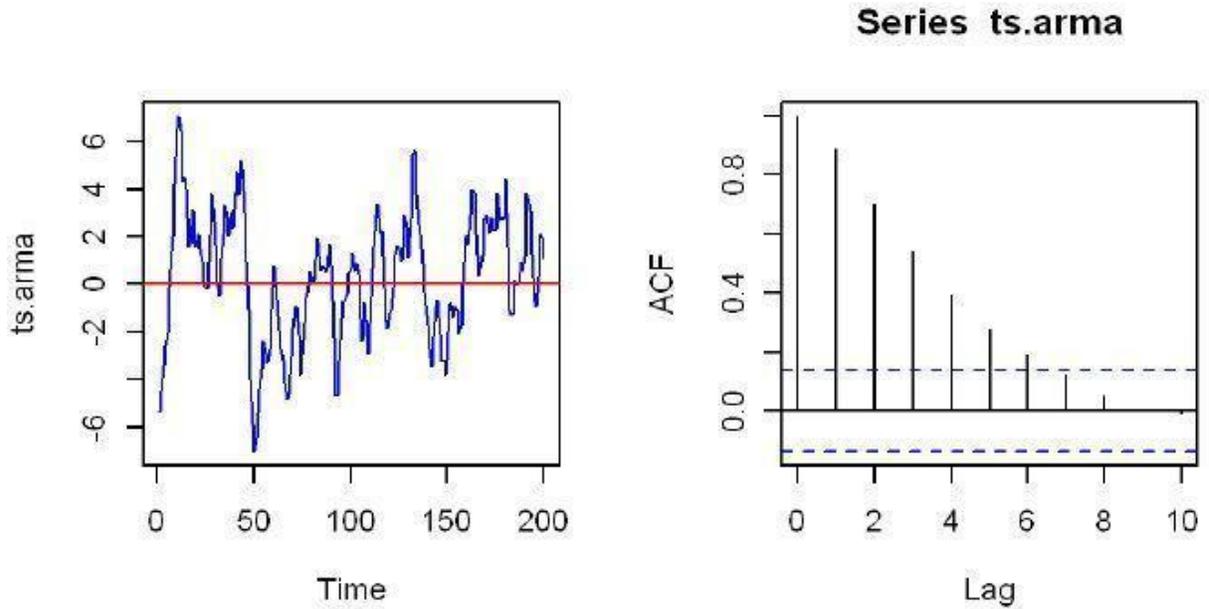


FIGURE 2.3 – Simulation d'un  $ARMA(1)$  :  $X_t = 0.8X_{t-1} + 0.6\epsilon_{t-1} + \epsilon_t$  et sa fonction d'autocorrélation

**Proposition 2.4.10.** [2] *La somme d'un  $ARMA(p_1, q_1)$  et d'un  $ARMA(p_2, q_2)$  (non corrélés) est un  $ARMA(\leq \max(p_1, p_2), \leq \max(q_1, q_2))$ .*

### Prévision pour un $ARMA(p, q)$

- **Prévision optimale**

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

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

$$\phi(L)(X_t - \mu) = \theta(L)\epsilon_t$$

$$\theta^{-1}(L)\phi(L)(X_t - \mu) = \epsilon_t$$

$$\theta^{-1}(L)\phi(L)X_t = \mu_0 + \epsilon_t \text{ avec } \mu_0 = \frac{\mu}{\theta(1)}$$

D'où :

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

Les prévisions dans ce cas données par :

$$\begin{aligned} X_{t+h}^* &= \mu_0 - \sum_{k=1}^{t+h} a_k X_{t+h-k} \\ &= \mu_0 - \sum_{k=1}^{h-1} a_k X_{t+h-k}^* - \sum_{k=h}^{+\infty} a_k X_{t+h-k}^* \end{aligned}$$

- **Prévision approchée**

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

On le calcule par itérations sur  $h$ .

# Chapitre 3

## Modélisation par les processus linéaires non stationnaires

### 3.1 Non stationnarité

#### 3.1.1 série intégrée

**Définition 3.1.2.** Une série  $\{X_t, t \in I\}$  est dit *intégrée* d'ordre  $d$ , On note  $X_t \rightarrow I(d)$ , si sa différence  $d$ -ième est stationnaire. La série stationnaire est alors *intégrée* d'ordre 0 et est notée  $X_t \rightarrow I(0)$ .

#### *Pourquoi des séries sont non stationnaires*

1. La première série non stationnaire à laquelle on pense est celle qui est croissante dans le temps.
2. Une série saisonnière est également non stationnaire puisque la valeur espérée dépend du temps dans la période de la saison.
3. Une série dont la dispersion varie dans le temps n'est pas stationnaire.

#### 3.1.3 Exemples (*Processus non stationnaires*)

- *processus TS (Trend Stationary)*

Commençons par définir ce qu'est un processus TS, selon la terminologie proposée par Nelson et Plosser (1982).

Le processus TS s'écrit :

$$X_t = f(t) + \epsilon_t,$$

où  $\epsilon_t$  représente l'erreur du modèle à la date  $t$ .

Il présente une non stationnarité de nature déterministe.

Le processus TS est non stationnaire car  $\mathbb{E}(X_t) = f(t)$  dépend du temps  $t$ .

Le processus TS  $X_t$  peut être stationnarisé en retranchant à  $X_t$  la valeur estimée par la méthode des Moindres Carrés Ordinaires.

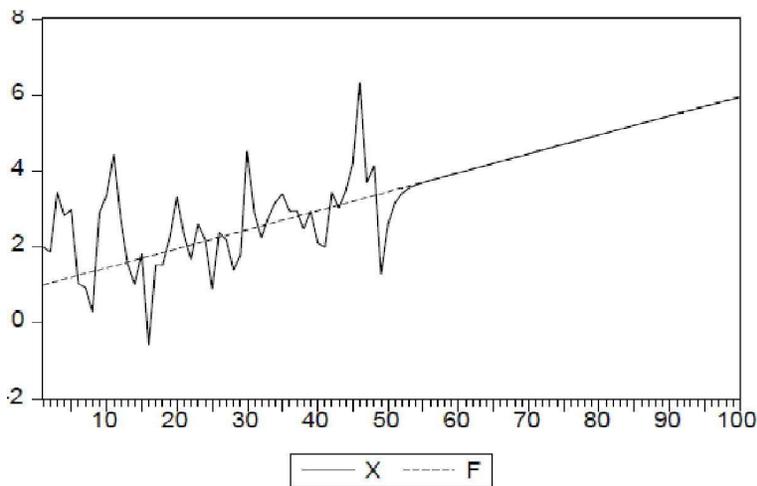


FIGURE 3.1 – Simulation de processus TS

- **processus DS ( *D*ifference *S*tationary )**

Le processus DS avec dérivée ( $\beta \neq 0$ ) s'exprime comme suit :

$$X_t = X_{t-1} + \beta + \epsilon_t$$

Le processus DS avec dérivé est appelé aussi marche au hasard (ou marche aléatoire) avec dérivé.

Il présente une non stationnarité de nature stochastique.

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

$$X_1 = X_0 + \beta + \epsilon_1 \quad X_2 = X_1 + \beta + \epsilon_2 = X_0 + \beta + \epsilon_1 + \epsilon_2 .$$

.

$$\begin{aligned} & \cdot \\ & \cdot \\ X_t &= X_0 + \beta t + \sum_{i=1}^t \epsilon_i \end{aligned}$$

où  $\epsilon_i \rightarrow iid(0, \sigma^2)$

$\epsilon_i$  est indépendante identiquement distribuée.

Le processus DS avec dérivé est non stationnaire car  $\mathbb{E}(X_t) = X_0 + \beta t$  qui dépend du temps  $t$ . Plus  $t \rightarrow \infty$  et Plus  $\mathbb{E}(X_t) \rightarrow \infty$ .

Le processus DS sans dérivée est applé aussi marche aléatoire .

Par récurrence , on obtient (dans le cas sans dérivé)

$$X_1 = X_0 + \epsilon_1 \quad X_2 = X_1 + \epsilon_2 = X_0 + \beta + \epsilon_1 + \epsilon_2 .$$

$\cdot$   
 $\cdot$   
 $\cdot$

$$X_t = X_0 + \sum_{i=1}^t \epsilon_i$$

où  $\epsilon_i \rightarrow iid(0, \sigma^2)$

Le processus DS sans dérivé est non stationnaire car on a :

$$\mathbb{V}ar(X_t) = \mathbb{V}ar\left(\sum_{i=1}^t \epsilon_i\right) = \sum_{i=1}^t \mathbb{V}ar\epsilon_i = \sum_{i=1}^t \sigma_\epsilon^2 = t\sigma_\epsilon^2$$

On constate que la variance du processus DS sans dérivée dépend du temps  $t$

Plus  $t \rightarrow \infty$  et Plus  $\mathbb{E}(X_t) \rightarrow \infty$ . et Plus  $\mathbb{V}ar(X_t) \rightarrow \infty$ .

Pour stationnariser le processus DS (avec ou sans dérivée), il suffit de le passer en différence première :

$$X_t - X_{t-1} = \beta + \epsilon_t$$

ou

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

**Remarque 3.1.4.**

1. Un processus *TS* est un processus que l'on peut rendre stationnaire par régression sur la tendance déterministe .
2. Pour stationnariser un processus *DS* d'ordre  $d$ , il convient d'appliquer le filtre  $(1 - L)^d$ .

## 3.2 Processus ARIMA

L'objectif de cette section est d'inclure les composantes non stationnaires de tendance (trend) dans la modélisation des séries chronologiques. Ces processus sont non stationnaires dès que  $d \leq 1$ . les séries économiques sont souvent non stationnaires, tel le PIB.

### 3.2.1 Définitions et propriétés

**Définition 3.2.2.** Soient  $p, d, q$  positifs ou nuls. Un processus  $(X_t)$  est dit **ARIMA**( $p, d, q$ ) (AutoRegressive Integrated Moving Average)

si le processus  $\nabla^d X_t = (1 - L)^d X_t$  est un processus ARMA( $p, q$ ) stationnaire, Ce sont les processus  $(X_t)$  de type :

$$\Phi(L)(1 - L)^d = \Theta(L)\epsilon_t.$$

où les racines de  $\Phi$  et  $\Theta$  sont de module supérieur à 1 et où  $(\epsilon_t)$  est un bruit blanc de variance  $\sigma^2$ .

Les processus ARIMA sont utiles pour des processus qui ont des corrélations positives et lentement décroissantes car cette propriété des autocorrélations peut être le signe d'une tendance dans la série.

En pratique, les modèles ARIMA sont préférés aux modèles AR dont les racines proches du cercle de l'unité. Par exemple, le modèle AR(2)

$$(1 - 0.815L)(1 - 0.989L)X_t = \epsilon_t$$

possède des racines proches du cercle unité, mais un modèle ARIMA

$$(1 - aL)(1 - L)X_t = \epsilon_t$$

est plus stable pour l'analyse et la prévision.

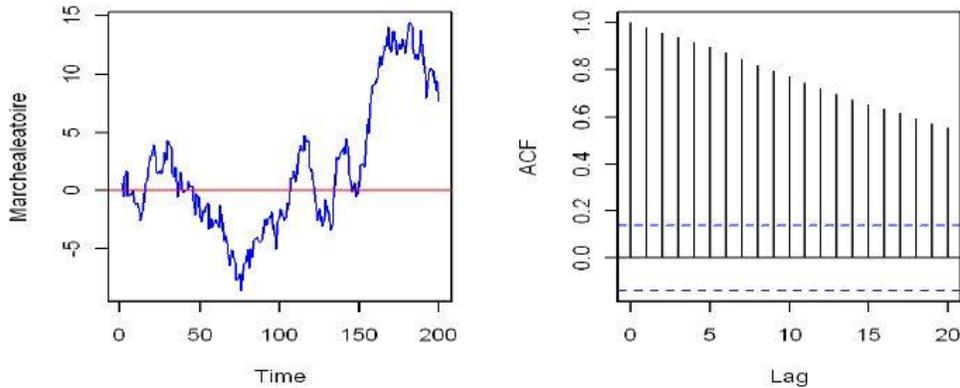


FIGURE 3.2 – Simulation de  $X_t = X_{t-1} + \epsilon_t$  et sa fonction d'autocorrélation.

### Exemple 3.2.3.

- Soit *une marche aléatoire (random walk)*. c'est-à-dire un processus  $X$  satisfaisant

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

a partir d'une certaine date par exemple  $t = 0$ . Nous avons déjà vu au Chapitre 1 que ce processus n'est pas stationnaire. Si une valeur initiale  $X_{-1}$  (déterministe ou aléatoire) est donnée alors

$$X_t = X_{t-1} + \sum_{j=0}^{t-1} \epsilon_j$$

Il est naturel d'autre part de supposer que  $X_{-1}$  est non corrélée avec les valeurs futures du bruit. Le processus  $(X_t)$  est alors défini sans ambiguïté i.e. de manière unique.

### Définition 3.2.4. (Représentation canonique minimale)

$(X_t)_{pd}$  est un processus ARIMA( $p, d, q$ ) en *représentation canonique minimale* s'il vérifie une équation de type :

$$\Phi(L)(1 - L)^d X_t = \mu + \Theta(L)\epsilon_t.$$

Et ceci avec :

i)  $\epsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$

ii)  $\Phi(L) = I - \phi_1 L - \phi_2 L^2 - \dots - \phi_P L^P$ , avec  $\phi_P \neq 0$

iii)  $\Theta(L) = I - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ , avec  $\theta_q \neq 0$

iv)  $\phi$  et  $\theta$  ont toutes leurs racines de module strictement supérieur à 1 et n'ont pas de racines communes.

**Notations :**

- Si  $d = 0$ ,  $X_t \rightsquigarrow ARMA(p, q)$  qui est un processus stationnaire.

On note  $X_t \rightsquigarrow I(0)$ .

- Si  $d = 1$ ,  $(X_t)$  est un processus intégré d'ordre 1.

On note  $X_t \rightsquigarrow I(1)$ .

- Si  $d = 2$ ,  $(X_t)$  n'est pas stationnaire,  $Y_t = (1 - L)X_t$  non plus,

$Z_t = (1 - L)Y_t = (1 - L)^2 X_t$  est asymptotiquement équivalent à un processus stationnaire. On note  $X_t \rightsquigarrow I(2)$ .

**Remarque 3.2.5.** Comme  $(1 - L)^d \Phi(L)X_t = \Theta(L)X_t$ , on pose  $Y_t = (1 - L)^d X_t$ .  $(Y_t)$  suit alors le processus :

$$\Phi(L)(1 - L)^d Y_t = \mu + \Theta(L)\epsilon_t.$$

**Prévision d'un modèle ARIMA(p, d, q)**

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

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

• **Prévision optimale**

Une prévision optimale de  $X_{t+h}$  sera de la forme (prévision de  $Y_{t+h} + m_{t+h}$ ). Pour les ARIMA, on a vu qu'on devait se donner les conditions initiales :  $Z = (1, X_0, \dots, X_{-p-d+1}, \epsilon_0, \dots, \epsilon_{-q+1})$ . On a alors :

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

– *Approximation AR* :

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

– *Approximation MA* :

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

On a  $\mathcal{L}(Y_t, \dots, Y_1, Z) = \mathcal{L}(\epsilon_t, \dots, \epsilon_1, Z)$  alors :

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

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

- **Prévision approchée**

On définit la prévision approchée comme :

$${}_t Y_{t+h}^* = - \sum_{j=1}^{t+h-1} a_j Y_{t+h-j}$$

En effet,  $\lim_{t \rightarrow +\infty} g(t) = 0$ . On peut donc également le calculer par itération sur  $h$ .

### 3.3 Les modèles ARIMA saisonniers (SARIMA)

Si l'on veut en même temps traiter les saisonnalités de période  $s$  (sans supposer une répétition exacte, déterministe des données), on est amené à définir les processus SARIMA.

#### 3.3.1 Définitions et exemples

**Définition 3.3.2.**  $X_t$  est appelé **processus autorégressif moyenne mobile et saisonnalité intégrés** avec période  $s$  ( $X_t \rightsquigarrow SARIMA(p, d, q) \times (P, D, Q)_s$ ) (Seasonal AutoRegressive Integrated Moving Average), si

$$Y_t = \Delta^d \Delta_s^D X_t = (1 - L)^d (1 - L^s)^D X_t$$

est un processus ARMA stationnaire de la forme

$$A(L) \times F(L^s) Y_t = \theta(L) \times G(L^s) \epsilon_t$$

où  $A(z)$  est le polynôme générateur d'un AR( $p$ )

$$A(z) = 1 - \sum_{k=1}^p a_k z^k$$

$\Theta(z)$  est le polynôme générateur d'un MA( $q$ )

$$\Theta(z) = 1 + \sum_{k=1}^q \theta_k z^k$$

et où, pour la saisonnalité  $Y_t - Y_{t-s}$ ,  $F(z)$  est le polynôme générateur d'un AR( $P$ )

$$F(z) = 1 - \sum_{k=1}^P \phi_k z^k,$$

et  $G(z)$  est le polynôme générateur d'un MA( $Q$ )

$$G(z) = 1 + \sum_{k=1}^Q \gamma_k z^k.$$

**Remarque 3.3.3.**  $Y_t$  est un processus particulier du type ARMA( $p + Ps, q + Qs$ ) stationnaire.

**Exemple 3.3.4.**

1) Soit le processus  $(X_t)$  défini par :

$$X_t = \mu + S_t + \epsilon_t,$$

où  $S_t = S_{t+s} \quad \forall t$ , est un processus SARIMA(0,0,0)  $\times$  (0,1,1)<sub>s</sub>. Donc ,

$$Y_t = (1 - B^s)X_t = \epsilon_t - \epsilon_{t-s}$$

est un ARMA(0, s) (non inversible).

2)  $s = 12$  ( données mensuelles ),  
 $p = 1 = P, q = 0 = Q, d = D = 1$  :  
 soit l'opérateur

$$A(z)F(z^s) = (1 - az)(1 - \phi z^{12}) = (1 - az - \phi z^{12} - a\phi z^{13})$$

$$\Rightarrow Y_t = aY_{t-1} + \phi Y_{t-12} - a\phi Y_{t-13} + \epsilon_t,$$

donc un AR( $p + Ps$ ) = AR(13) ou un AR(1) pour  $Y_{t-1} - Y_{t-12}$ .

3) Soit le processus  $(X_t)$  défini par :

$$X_t = a \sin\left(\frac{\pi t}{6} + 4\right) + \epsilon_t,$$

ou  $(\epsilon_t)$  est un bruit blanc.

On a

$$E(X_t) = a \sin\left(\frac{\pi t}{6} + 4\right).$$

$(X_t)$  n'est pas stationnaire en moyenne, la partie déterministe est périodique de période 12.

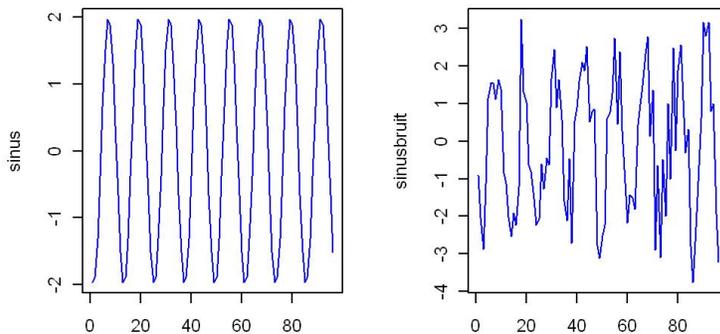


FIGURE 3.3 – Simulation d'un  $X_t = a \sin\left(\frac{\pi t}{6} + 4\right) + \epsilon_t$ ,

**Proposition 3.3.5.** [15]

Inclusion des trends ( $d > 0$ ) et des composantes déterministes périodiques ( $D > 0$ )

$$\Rightarrow SARIMA(p, d, q) \times (P, D, Q)_s.$$

**Proposition 3.3.6.** [15]

Des autocorrélations oscillantes  $\Rightarrow$  existence des saisonnalités (processus SARIMA).

Note

La notion de processus SARIMA peut s'étendre à plusieurs périodes différentes pour un même processus.

Par exemple, l'observation des températures en un lieu donné fait apparaître :

- des variations journalières,
- des variations annuelles.

Il est nécessaire de disposer de séries de grande taille pour analyser ces processus. Par exemple, observer une température pendant trente ans avec une période de une heure conduit à une taille de l'ordre de 260000.

## 3.4 Modélisation des chroniques d'un modèle SARIMA

L'ouvrage de Box et Jenkins "Time series analysis, forecasting and control", publié en 1970 a proposé une démarche de prévision pour les séries univariées, fondée sur l'utilisation de processus SARIMA .

Les étapes pour l'estimation des coefficients d'un processus SARIMA sont les suivantes :

### 3.4.1 Identification et ajustement des modèles SARIMA aux données

Plan :  $X_1, \dots, X_T$  données

#### 0. Transformation

Applications des transformations ("Box-Cox") pour stabiliser la variance  
 transformation de type :  $(\log, \exp, \sqrt{\cdot})$ .

### Rappel : Transformations de Box-Cox

Il arrive fréquemment que des séries chronologiques aient une variance de bruit et une amplitude de variations saisonnières qui sont des fonctions croissantes de  $m$ . Pour y remédier, une solution est d'utiliser une transformation de Box-Cox pour les séries strictement positives. Cela consiste à transformer une série  $x_t$  en  $x_t^{(a)}$  où :

$$x_t^{(a)} = \begin{cases} \frac{(x_t)^a - 1}{a} & \text{si } a \neq 0 \\ \log(x_t) & \text{si } a = 0. \end{cases}$$

où  $a$  est une constante à choisir. En pratique, un bon choix de  $a$  permet souvent de stabiliser les fluctuations des résidus et de la saisonnalité.

**Exemple 3.4.2.** *L'analyse de série de temps est plus facile sur des données stationnaires. (plus d'outils) quand une série temporelle est non stationnaire, vous pouvez essayer de trouver une autre série chronologique, liée à l'initiale, et qui est stationnaire. Dans de nombreux cas la prise de la différence sera assez. (La série chronologique intégrée avec l'ordre 1) parfois que vous devrez prendre le logarithme avant par exemple pour la série chronologique financière*

*La série de prix du baril de pétrole (son logarithme plus précisément dans la suite) sur la période (Janvier 1989 - Avril 2009) en dollars On constate clairement une tendance, la série  $\log(\text{prix de pétrole})$  n'est pas stationnaire*

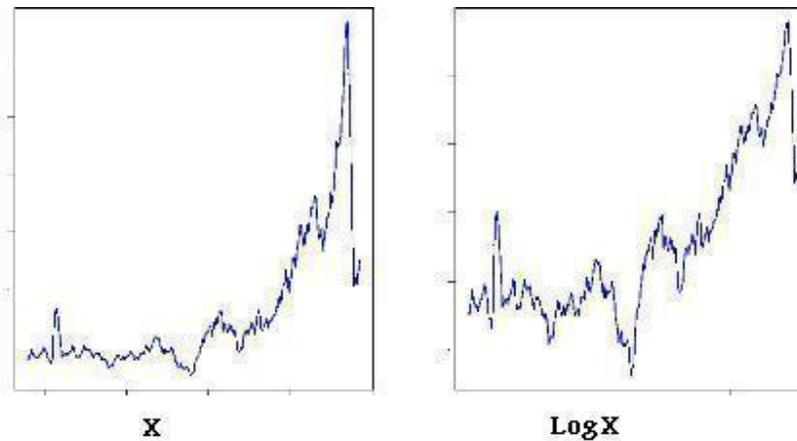


FIGURE 3.4 – Évolution du prix de pétrole, et de son logarithme, entre Janvier 1989 et Avril 2009

**Remarque 3.4.3.** *On peut utiliser les tests dits tests de racines unitaire pour vérifier si la série est stationnaire au non.*

### 1. Choisir $d$ , $D$ , $s$

(souvent  $0 \leq d \leq 2$ ,  $0 \leq D \leq 1$ ) de sorte que  $(1-L)^d(1-L^s)^D X_t = Y_t$  stationnaire.

#### Remarque 3.4.4.

**Role de  $D$**  (souvent  $0 \leq D \leq 1$ ) : Si  $D > 1$ , la tendance n'est pas seulement dans  $X_t$ , mais encore dans  $(\nabla^d X)_t - (\nabla^d X)_{t-s}$ . C.a.d., les différences de l'ordre  $s$ ,  $(\nabla^d \nabla_s X)_t$  ne sont pas stationnaires, il reste une tendance (faible, alors que  $D = 2$  suffira).

#### Détermination de l'ordre de différenciation $d$

1. Une série stationnaire fluctue autour d'une valeur moyenne et sa fonction d'auto-corrélation décline rapidement vers zéro. Si une série présente des auto-corrélations positives pour un grand nombre de décalages (par exemple 10 ou plus), alors elle nécessite d'être différenciée. La différenciation tend à introduire des auto-corrélations négatives.

**Exemple 3.4.5.** On considère un processus  $(X_t)$  tel que  $X_t - X_{t-1} = \epsilon_t$  où  $\epsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$  et

$$X_t = X_0 + \sum_{k=1}^t \epsilon_k$$

$$X_{t+h} = X_0 + \sum_{k=1}^{t+h} \epsilon_k$$

$$\begin{aligned} \rho(h) &= \frac{\text{Cov}(X_t, X_{t+h})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+h})}} \\ &= \frac{\text{Cov}(X_0 + \sum_{k=1}^t \epsilon_k, X_0 + \sum_{k=1}^{t+h} \epsilon_k)}{\sqrt{\text{Var}(X_0 + t\sigma^2)\text{Var}(X_0 + (t+h)\sigma^2)}} \\ &= \frac{\text{Var}(X_0) + t\sigma^2}{\sqrt{\text{Var}(X_0 + t\sigma^2)\text{Var}(X_0 + (t+h)\sigma^2)}} \end{aligned}$$

Pour  $t$  grand et  $h < t$ ,

$$\rho(h) \# \frac{t\sigma^2}{\sigma^2 \sqrt{t(t+h)}} = \frac{1}{\sqrt{1 + \frac{h}{t}}} \# 1 - \frac{h}{2t}$$

La décroissance est lente et linéaire en  $h$ . D'où une règle pratique :

si les  $\hat{\rho}_T(h)$  restent proches de 1 ou décroissent linéairement avec  $h$  alors le processus est sans doute non stationnaire.

**Exemple 3.4.6.** : Modélisation des indices boursiers (bourse de Paris). On considère les prix de clôture de l'indice boursier CAC40 sur la période 1991–1998 (1860 observations).

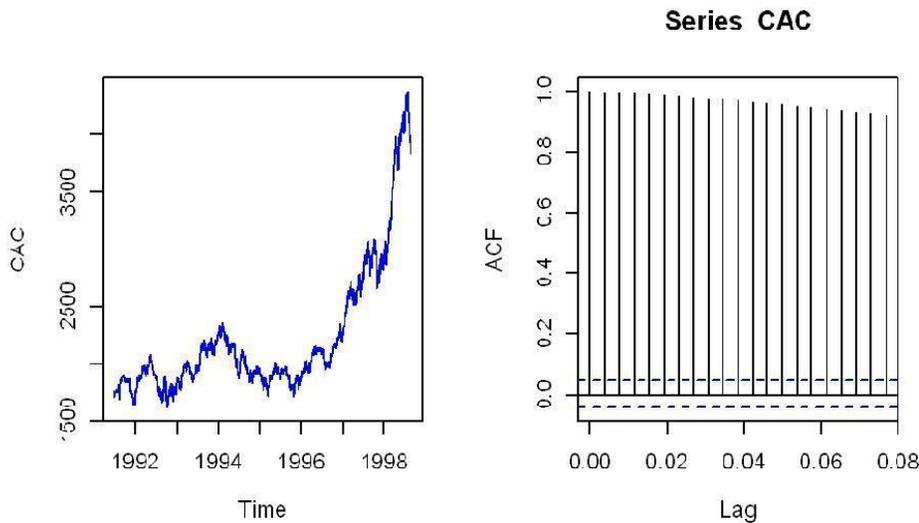


FIGURE 3.5 – Simulation de la série  $CAC$

**Remarque 3.4.7.** Si  $(X_t)$  admet une racine unité, la proposition  $\rho(h)$  décroît exponentiellement vers 0 avec  $h$  n'est plus vraie : c'est la persistance des chocs .

**choix de  $d$**  : combien de fois faut-il différencier pour obtenir une série stationnaire (autocorrélogrammes, tests statistiques...). Donc un estimateur de  $d$  est le nombre total de fois où on rejette la stationnarité.

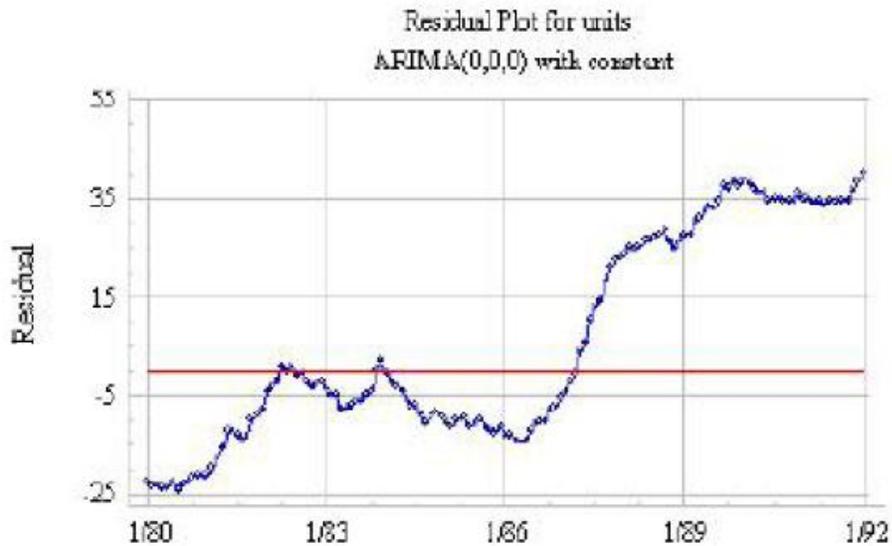
2. L'ordre optimal de différenciation est souvent celui pour lequel l'écart-type est minimal. Un accroissement de l'écart-type doit donc être considéré comme un symptôme de surdifférenciation.

3. Si l'auto-corrélation de décalage 1 est égale à 0 ou négative, la série n'a pas besoin d'être différenciée. Si l'auto-corrélation de décalage 1 est inférieure à  $-0.5$ , la série est sur-différenciée
4. Un troisième symptôme de sur-différenciation est un changement systématique de signe d'une observation à l'autre.

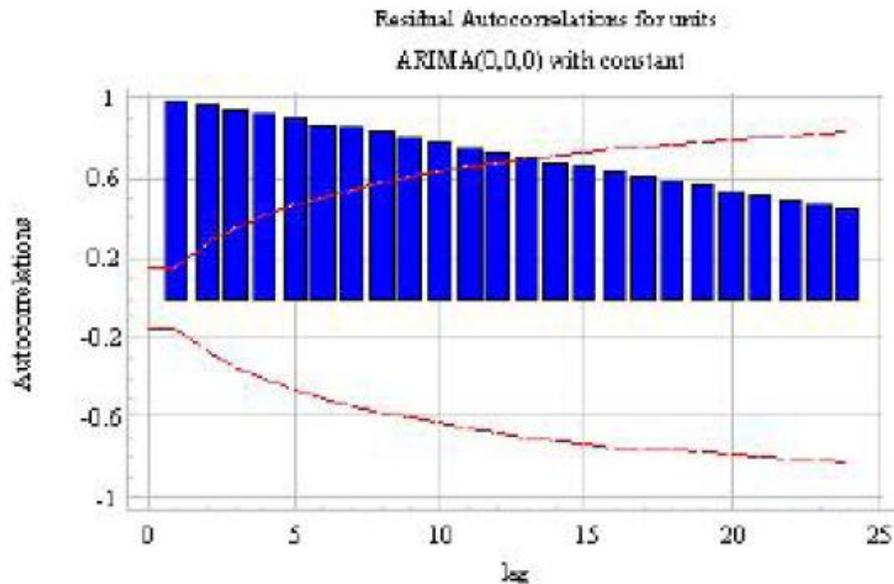
## 2. Stationnarisation par filtrage

- L'opérateur  $(1 - L^s)^D$ , élimine la périodicité (désaisonnalisation).
- L'opérateur de différentiation  $\nabla$ , ( ou  $\nabla^d$ ) élimine les tendances.

Par exemple, la figure suivante représente une série temporelle originale (il s'agit de l'évolution temporelle d'unités de vente).

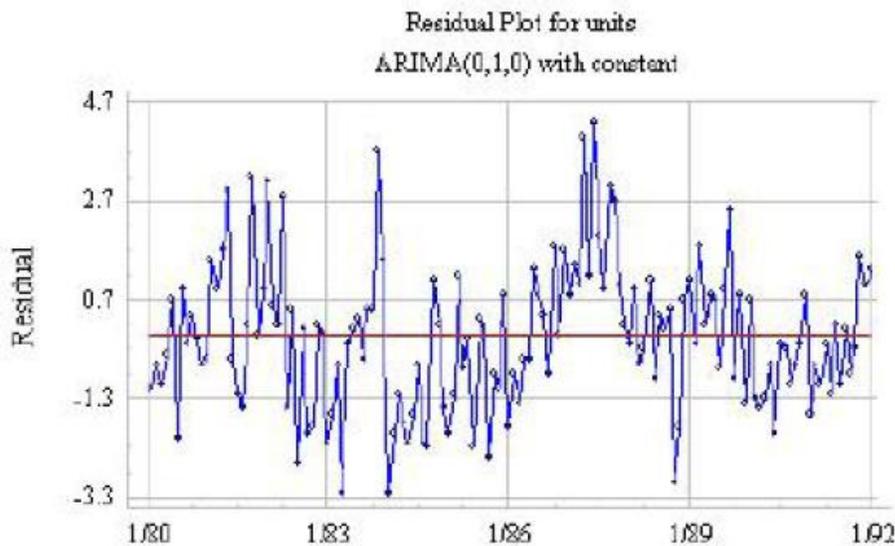


Le graphique de la fonction d'auto-corrélation présente une régression lente et linéaire, typique des séries non stationnaires :

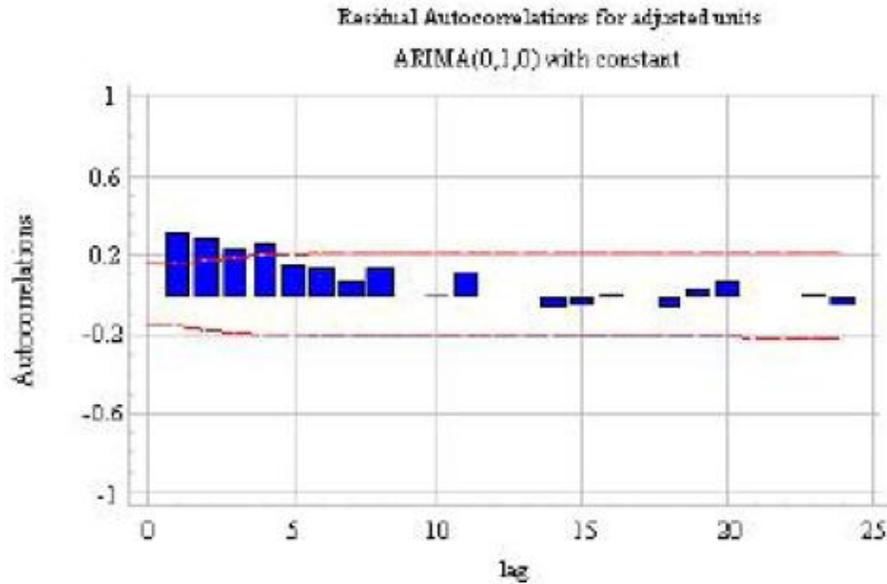


Enfin l'écart-type de la série est important : 17.56.

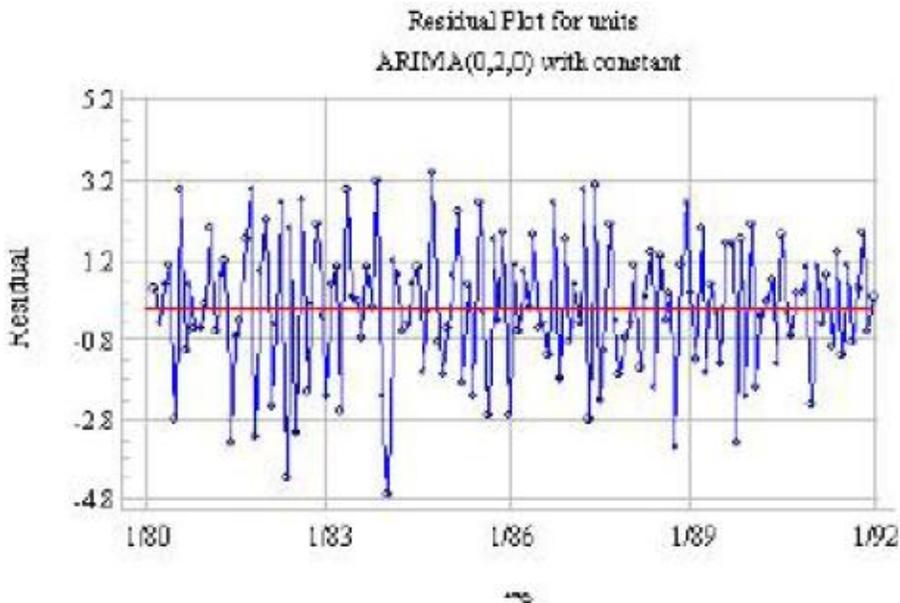
Cette série nécessite de toute évidence d'être différenciée. Une première différenciation (c'est-à-dire l'application d'un ajustement  $ARIMA(0, 1, 0)$  avec constante donne les résidus suivants :



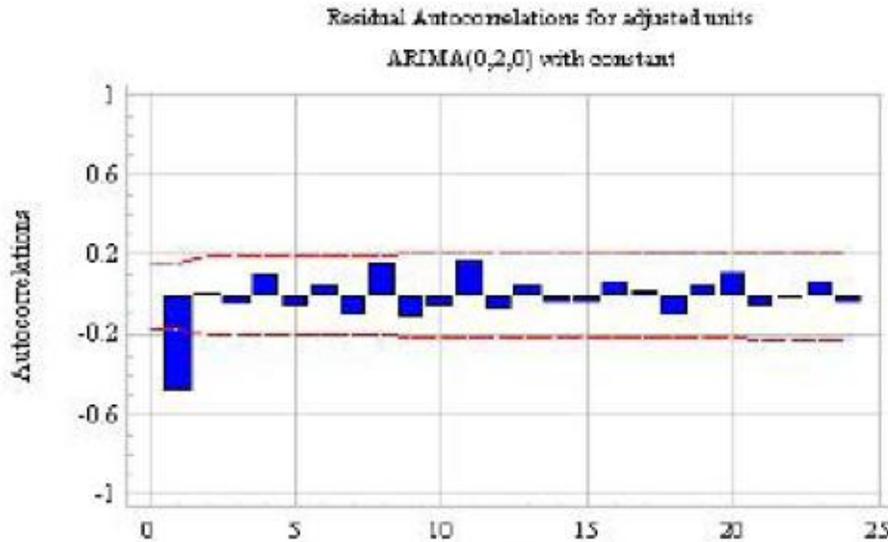
La série semble a peu près stationnaire,



L'écart-type a cependant été réduit de manière importante, de 17,56 à 1,54. Si l'on essaie une seconde différenciation (c'est-à-dire l'application d'un modèle  $ARIMA(0, 2, 0)$  avec constante, on obtient les résultats suivants :



Cette série montre de clairs signes de sur-différenciation. Les valeurs passent de manière quasiment systématique du signe positif au signe négatif. La fonction d'auto-corrélation montre un pic négatif au premier décalage proche de 5



Enfin, l'écart-type a augmenté de 1.54 à 1.81. Ceci semble indiquer que l'ordre optimal de différenciation, pour cette série, est de 1. Un modèle sans différenciation suppose que la série originale est stationnaire. Un modèle avec une différenciation d'ordre 1 suppose que la série originale présente une tendance constante. Un modèle avec une différenciation d'ordre 2 suppose que la série originale présente une tendance variant dans le temps.

### 3. Estimation des paramètres P, Q

Calculer ACF  $\hat{\rho}_t$  et PACF  $\hat{r}_t$  empirique et examiner  $\hat{\rho}_{ks}$ ,  $k \geq 0$ , et  $\hat{r}_{ks}$ ,  $k \geq 0$ , pour trouver P,Q de sorte que  $\hat{\rho}_{ks}$  et  $\hat{r}_{ks}$  correspond a ARMA(P,Q) (et décroissent exponentiellement).

Exemple :

ACF ( $s = 12$ ) :  $|\hat{\rho}_{12}| > IC \implies P \geq 1$ .

PACF  $|\hat{r}_{12}| > IC \implies P \geq 1$ .

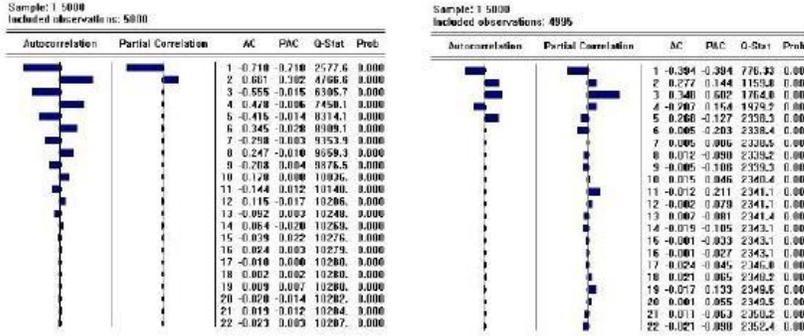
### 4. Estimation des paramètres p, q

Choix du p, q tel que  $\hat{\rho}_1, \dots, \hat{\rho}_{s-1}$  et  $\hat{r}_1, \dots, \hat{r}_{s-1}$  sont compatibles avec modèle ARMA(p, q).

#### Méthode pratique d'estimation des ordres p, q

pour estimer les ordres  $p$  ou  $q$ , on utilise les propriétés vues précédemment sur les formes autocorrélogrammes  $\rho(h)$  ou des autocorrélogrammes partiels  $r(h)$ .

En particulier



1. pour les processus  $AR(p)$  l'autocorrélogramme partiel s'annule à partir de  $p$ .
2. pour les processus  $MA(q)$  l'autocorrélogramme partiel s'annule à partir de  $q$ .

### Cas d'un processus $MA(q)$

Soit  $X = (X_t, t \in \mathbb{Z})$  un processus  $MA(q)$ . D'après le théorème 4.1.2, pour tout  $r > q$ , alors l'autocorrélation empirique  $\rho_X^{(N)}(r)$  est approximativement une gaussienne centrée de variance

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

En pratique, on fait donc l'approximation

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

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

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

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  $N$ .

### Cas d'un processus $AR(p)$

Soit  $X = (X_t, t \in \mathbb{Z})$  un processus  $AR(p)$ . D'après le théorème 4.1.2, pour tout  $r > p$ , alors l'autocorrélation partielle empirique  $r_X^{(N)}(r)$  est approximativement 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  $H_0$ .

### Cas d'un processus ARMA(p, q)

On pourrait s'inspirer du critère du coin pour construire un test sur les paramètres  $p$  et  $q$ . En pratique, on procède différemment et plus simplement en vertu par :

Si  $X$  un processus ARMA( $p, q$ ) vérifiant

$$\Phi(B)(X) = \Theta(B)(\epsilon),$$

avec  $\Theta(z)$  et  $\Phi(z)$  sans racine commune, et  $\epsilon$  bruit blanc d'innovation de variance  $\sigma^2$ , alors il peut se mettre sous forme MA( $\infty$ )

$$X_t = \epsilon_t + \sum_{u \geq 1} \psi_u \epsilon_{t-u}$$

ou AR( $\infty$ )

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

Comme on a les inégalités

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

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

et que  $\sum_{u>0} |\psi_u|$  comme  $\sum_{v>0} |\eta_v|$  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  $p' \geq p$  et  $q' \geq q$  : 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. On verra par la suite comment améliorer cette première estimation majoration des paramètres du processus ARMA.

## 5. Estimation des paramètres du filtre ARMA

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

### Estimation préliminaire

- (i) Estimation de la partie autorégressive : Si  $X$  est un processus  $ARMA(p, q)$  causal et inversible, vérifiant  $\Phi(B)X = \Theta(B)\epsilon$ , alors on a vu que sa fonction d'autocovariance vérifie la relation de récurrence suivante :

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

et que l'on peut en déduire le système d'équations dit aussi de Yule-Walker

$$\begin{pmatrix} \rho_X(q) & \dots & \rho_X(q - p + 1) \\ \vdots & \ddots & \vdots \\ \rho_X(q + p - 1) & \dots & \rho_X(q) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho_X(q + 1) \\ \vdots \\ \rho_X(q + p) \end{pmatrix}$$

Les estimateurs de Yule Walker de  $\phi_1, \dots, \phi_p$  sont donc naturellement définis par

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

- (ii) Estimation de la partie moyenne mobile et de la variance du bruit blanc : On s'appuie sur la remarque simple suivant :  $Y = \Phi(B)(X)$  est un processus  $MA(q)$ , de paramètres  $\theta_1, \dots, \theta_q$  : Une fois calculé  $\hat{\phi}^{(N)}(z) = 1 + \hat{\phi}_1^{(N)}z + \dots + \hat{\phi}_p^{(N)}z^p$ , on pose

$$Y = \hat{\phi}^{(N)}(B)(X),$$

et l'on détermine les estimateurs  $\hat{\theta}_1^{(N)}, \dots, \hat{\theta}_q^{(N)}$  on appliquant à  $Y$  l'algorithme des innovations.

On en déduit également l'estimateur  $\sigma_{(N)2}$  de la variance du bruit blanc  $\sigma^2$ .

## 6. Validation

Il s'agit de vérifier notamment que les résidus du modèle ARMA estimé, résidus notés  $\hat{\epsilon}_t$ , vérifient les propriétés requises pour que l'estimation soit valide, à savoir qu'ils suivent un processus 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.

### Tests de bruit blanc et de stationnarité

#### 1. Analyse des fonctions d'autocorrélation

- i. Regarder le graphique des résidus estimés pour voir s'il apparaît des points aberrants, une tendance, une rupture, de l'autocorrélation, etc. Ceci n'est évidemment qu'indicatif.

- ii. Regarder les autocorrélations simples et partielles. Elles doivent être significativement nulles si les résidus sont un bruit blanc.

## 2. Test du portemanteau

Afin de tester que les résidus estimés suivent un BB, on teste l'hypothèse d'absence d'autocorrélation jusqu'à l'ordre  $m$ . En pratique, on utilise deux tests :

- a) Le test de Box-Pierce :

Ce test a pour objet de tester le caractère non autocorrélé des résidus.

- b) Le test de Ljung et Box :

Ce test est à appliquer, de préférence au test de Box-Pierce, lorsque l'échantillon est de petite taille, la statistique de Ljung-Box, donnée par :

$$LB = T(T + 2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k}$$

où les coefficients d'autocorrélation  $\hat{\rho}_k$  sont calculés sur les résidus estimés  $\hat{\epsilon}_t$ . Cette statistique, sous l'hypothèse que les résidus suivent un BB, suit une loi du  $\chi_m^2$

## 3. Test de normalité

Il s'agit de tester que les résidus estimés  $\hat{\epsilon}_t$  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

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

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

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

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

### Tests sur les paramètres

On vérifie tout d'abord que les racines des polynômes AR et MA ne sont pas égales à 1.

Si les hypothèses testées sur les résidus ne sont pas rejetées, on teste la significativité des retards du modèle ARMA par des tests de Student.

### Choix d'un modèle parmi plusieurs

Si, à la suite de ces étapes, il reste plusieurs modèles valides, on peut choisir parmi ces modèles, soit celui qui donne les meilleurs critères d'ajustement, soit celui qui donne les meilleures performances en prévision. Concernant les critères d'ajustement, on retient le modèle qui minimise les critères d'information (AIC et BIC). Ces critères sont tous basés sur les deux idées suivantes :

1. Etant donné que  $\sigma^2$  est la variance d'erreur de prévision à horizon 1, on aimerait choisir, parmi les modèles estimés, celui qui fournit la plus petite valeur.
2. On ne peut pas, pour des raisons statistiques, choisir un modèle présentant un grand nombre de paramètres ( $p$  et  $q$  grands).

Concernant les performances en prévision des modèles, on utilise couramment les critères suivants (que l'on cherche bien entendu à minimiser) :

$$RMSE = \sqrt{\frac{1}{T - k + 1} \sum_{t=k}^T (Y_t - \hat{Y}_t(h))^2}$$

et

$$MAE(h) = \frac{1}{T - k + 1} \sum_{t=k}^T |Y_t - \hat{Y}_t(h)|$$

où  $K$  est le nombre d'observations minimales pour mener une estimation du modèle. On peut calculer ces critères, soit sur la base de prévisions in-sample (toutes les observations ont été utilisées pour estimer le modèle et on calcule les prévisions sur cet même ensemble d'observations), soit sur la base de prévisions out-of-sample (on estime le modèle sur un ensemble d'observations et on mène la prévision sur le reste).

### 3.4.8 Prévision des processus SARIMA

Si les données  $X_1, \dots, X_T$  suivent un modèle  $SARIMA(p, d, q) \times (P, D, Q)_s$ , on définit

$$Y_t = (I - L)^d (I - L^s)^D X_t$$

Si par exemple,  $(d, D, s) = (1, 1, 12)$ ,

$$Y_t = (X_t - X_{t-12}) - (X_{t-1} - X_{t-13})$$

On traite le problème de prédiction du processus  $Y_t$  du type  $ARMA(p + Ps, q + Qs)$ ,

(a) Traitement de partie  $AR(p + Ps)$  de  $Y_t$ , par exemple :

$$\begin{aligned} Y_t &= aY_{t-1} + \phi Y_{t-12} - a\phi Y_{t-13} + \epsilon_t \\ \implies \hat{Y}_{T+1} &= aY_T + \phi Y_{T-11} - a\phi Y_{T-12} \\ \implies \hat{Y}_{T+2} &= a\hat{Y}_{T+1} + \phi Y_{T-10} - a\phi Y_{T-11} \\ &\text{etc.} \\ \implies \hat{X}_{T+1} &= \hat{Y}_{T+1} + X_{t-11} + X_t - X_{t-12} \end{aligned}$$

(b) Traitement de partie  $MA(q + Qs)$  de  $Y_t$  comme vu avec les résidus, par exemple pour un  $SARIMA(1, 1, 1) \times (1, 1, 1)_{12}$  :

$$e_t = Y_t - aY_{t-1} + \phi Y_{t-12} - a\phi Y_{t-13} - \theta e_{t-1} - \gamma e_{t-12} - \theta\gamma e_{t-13}$$

Prévision selon Box-Jenkins si T est suffisamment grand et l'horizon h n'est pas trop grand. Sinon : méthodes non paramétriques, par exemple lissage exponentiel. et on rentre après au niveau du processus  $X_t$  du type SARIMA en exprimant  $\hat{X}_T(1)$  en fonction de  $\hat{Y}_T(1)$  et les valeurs observés  $X_t$  pour  $t \leq T$ .

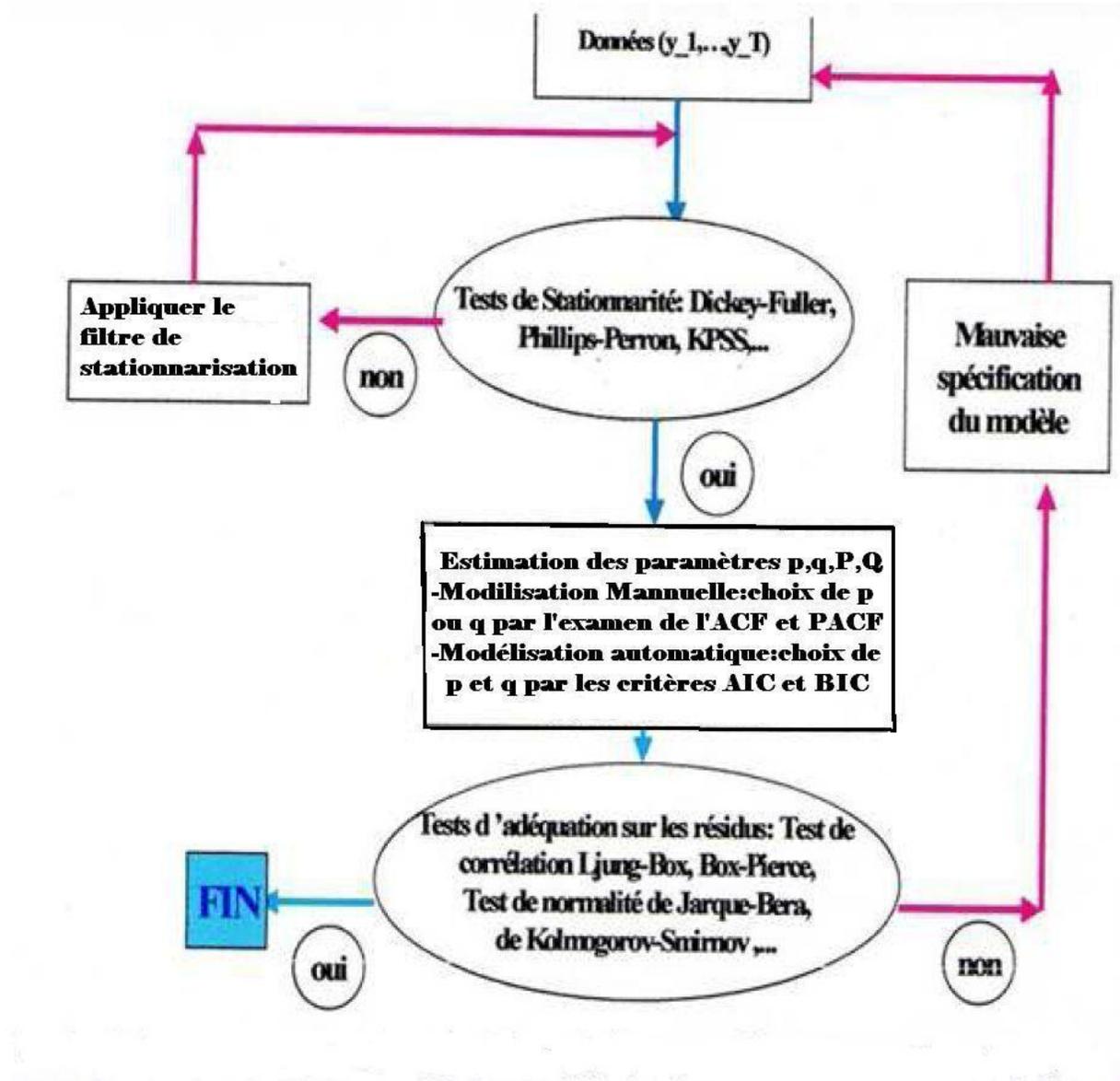


FIGURE 3.6 – Schéma général de la modélisation d'une série temporelle par un modèle SARIMA.

Phase d'analyse	Fonctions
Création et manipulation de séries	ts,rts,cts, its time start end frequency cycle window diff filter stl
Identification	plot, ts.plot acf pacf
Estimation	ar arima ARMAacf ARMAtoMA
Diagnostics	tsdiag
Calcul de prévisions	predict
Simulation	arima.sim

Principales fonctions R pour l'analyse de séries chronologiques.

### 3.4.9 Exemple : (Modélisation de la série des voyageurs)

La série AirPassengers est un classique de l'étude des séries chronologiques. Cette série mensuelle donne le nombre total de passagers avion de 1949 à 1960. Ouvrir R et charger cette série. En donner la représentation graphique. Constaté l'existence d'une saisonnalité mensuelle ( $d = 12$ ) et d'une croissance moyenne approximativement affine (prévision sans modèle).

On constate qu'il subsiste une non-stationnarité en variance qui nécessitera une transformation de Box-Cox des données afin de les stationnariser en variance

#### – *Etape 1 : Représentation des données*

On commence par analyse du graphique,

```
> plot(AirPassengers)
```

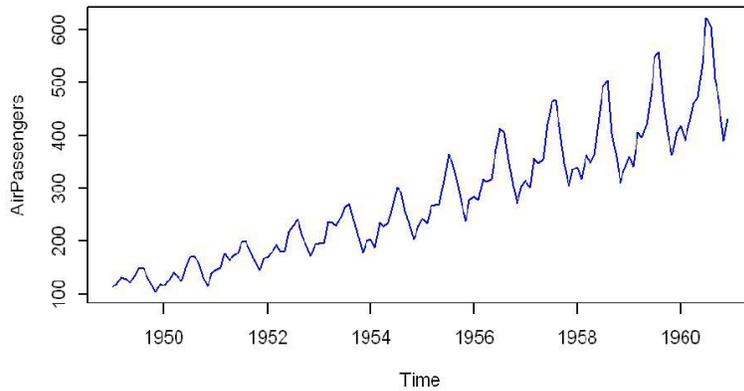


FIGURE 3.7 – Nombre total de passagers avion de 1949 à 1960.

Son graphe indique une forte dépendance entre la variabilité de la série et son niveau : lorsque le temps augmente, les valeurs prises par la série sont de plus en plus étendues. Pour éliminer cette variabilité, on applique une transformation aux données observées. Ici le logarithme népérien :

```
> plot(log(AirPassengers))
```

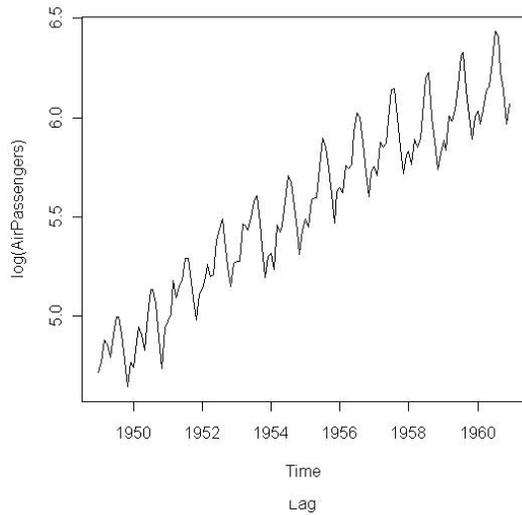


FIGURE 3.8 – Log de la série du nombre total de passagers avion de 1949 à 1960

– *Etape 2 : Tracé de la fonction d'autocorrélation et de la fonction d'autocorrélation partielle*

On donne dans le tableau ci-dessous un récapitulatif de la forme de la fonction des autocorrélations et des autocorrélations partielles pour différents processus stationnaires :

processus	autocorrélation $\rho_X(h)$	autocorrélation partielle $r_X(h)$
$AR(p)$	amortie	nulle pour $h > p$
$MA(q)$	nulle pour $ h  > q$	amortie
$ARMA(p, q)$	amortie	amortie

```
> acf(AirPassengers,type="correlation")
```

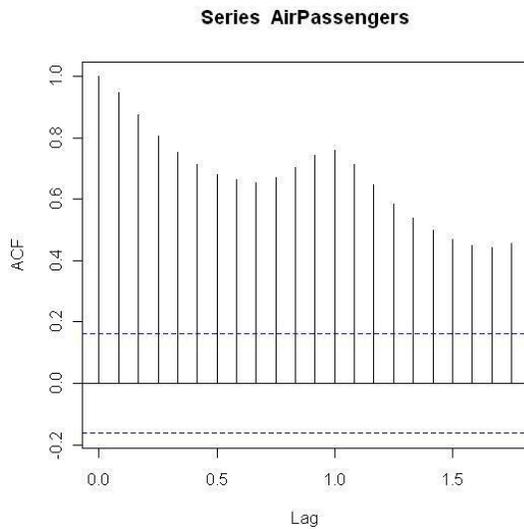


FIGURE 3.9 – La fonction d'autocorrélation de la série des voyageurs

```
> pacf(AirPassengers,lag=20,type="correlation")
```

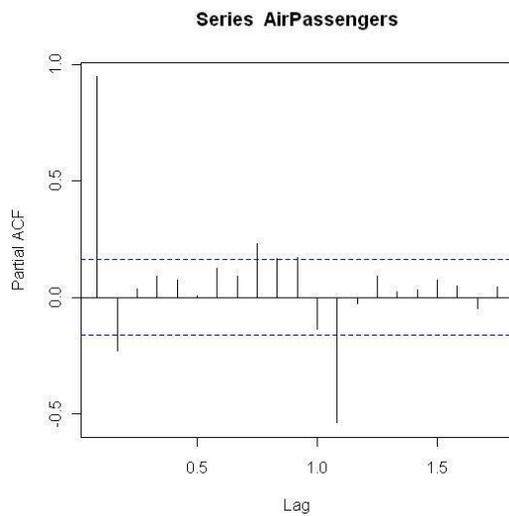


FIGURE 3.10 – La fonction d'autocorrélation partiel de la série des voyageurs

### Test d'autocorrélation

```
> Box.test(AirPassengers, 10)
```

Box-Pierce test

```
data : AirPassengers
X - squared = 816.4925, df = 10, p - value < 2.2e - 16
Les données sont corrélées.
```

### Test de Stationnarité

```
> adf.test(AirPassengers)
Augmented Dickey-Fuller Test
data : AirPassengers
Dickey - Fuller = -7.3186, Lagorder = 5, p - value = 0.01
alternative hypothesis : stationary
Message d'avis :
In adf.test(AirPassengers) : p-value smaller than printed p-value
Donc la série n'est pas stationnaire
> pp.test(AirPassengers)
Phillips-Perron Unit Root Test
data : AirPassengers
Dickey - Fuller Z(alpha) = -46.4056, Truncationlagparameter = 4, p-value =
0.01
alternative hypothesis : stationary
Message d'avis :
In pp.test(AirPassengers) : p-value smaller than printed p-value
> kpss.test(AirPassengers)
KPSS Test for Level Stationarity
data : AirPassengers
KPSSLevel = 4.3423, Truncationlagparameter = 2, p - value = 0.01
Message d'avis :
In kpss.test(AirPassengers) : p - value smaller than printed p - value
On rejette l'hypothèse nulle de stationnarité ( la série est non stationnaire)
```

### – Etape 3 : Stationnarisation

```
Avec le code :
> dAirPassengers <- diff(AirPassengers, lag = 12)
> plot(dAirPassengers)
on obtient
```

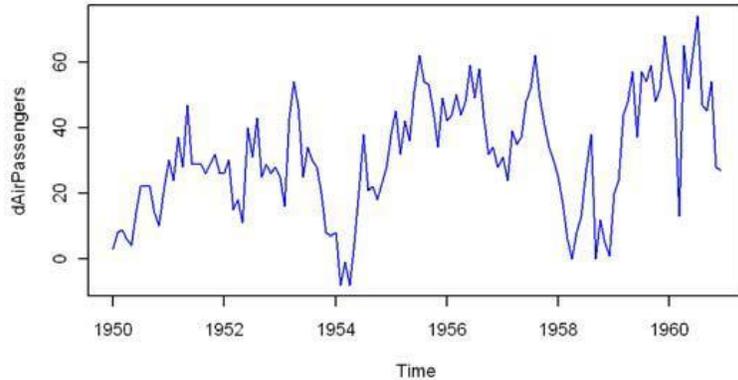


FIGURE 3.11 – La série du nombre total de passagers avion sans saisonnalité

L'effet saisonnier semble disparaître.

Avec le code : `library(tseries),`

`> adf.test(dCAC)`

on obtient :

Augmented Dickey-Fuller Test

data : dAirPassengers

*Dickey – Fuller* = -3.1519, *Lagorder* = 5, *p – value* = 0.09899

alternative hypothesis : stationary

Donc il y a une racine unitaire.

Avec le code : `fit8 <- arima0(log(AirPassengers), c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))` on obtient

Call :

`arima0(x = log(AirPassengers), order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))`

Coefficients :

*ma1* : -0.4013 *des.e* : 0.0897

*sma* 1 - 0.5544 *de s.e* : 0.0713 :

*sigma*<sup>2</sup> estimated as 0.00135 : *log likelihood* = 244.69, *aic* = -483.37

C'est à dire que le modèle proposé pour la série est donné par

$$(1 - B)(1 - B^{12})\log(\text{AirPassengerst}) = (1 - 0.4013B)(1 - 0.65544B^{12})\epsilon_t$$

Les résidus sont donnés par

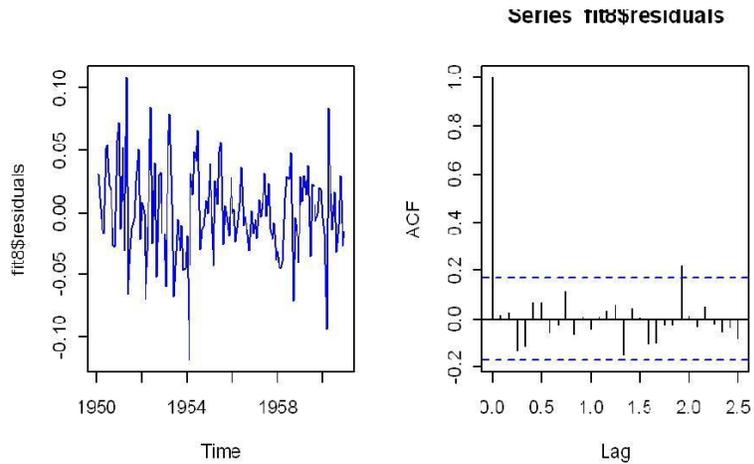


FIGURE 3.12 – La série du nombre total de passagers avion sans tendance

***Test du bruit blanc :***

Avec le code :

```
> Box.test(fit8$residuals, lag = 10, type = "Box-Pierce")
```

on obtient :

Box-Pierce test

```
data : fit8$residuals
```

```
X - squared = 58.4638, df = 10, p - value = 7.069e - 09
```

```
Les résidus sont non corrélés > Box.test (fit8$residuals, lag = 10, type = "Ljung-Box")
```

on obtient :

Ljung-Box test

```
data : fit8$residuals
```

```
X - squared = 61.3282, df = 10, p - value = 2.03e - 09
```

Donc la série *fit8\$residuals* n'est pas un bruit blanc.

– **Etape 4 : Prédiction**

On trace sur la figure les valeurs prédites ainsi que l'intervalle de confiance à 95% correspondant.

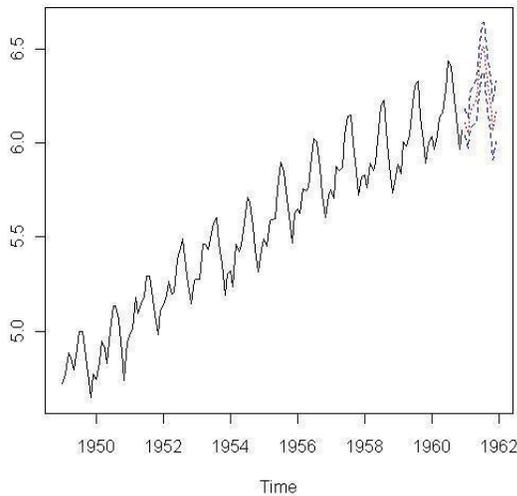


FIGURE 3.13 – Prédiction du nombre de du voyageurs pour les deux années suivante

### 3.4.10 Exemple : (Application à la chronique SNCF)

Il s'agit à la série mensuelle du nombre de voyageurs kilomètres, correspondant au trafic deuxième classe du réseau principal *SNCF* pour la période 1963 – 1980 (source : Gouriéroux et Monfort, 1983).

On enlève à la série les 12 dernières valeurs (soit l'année 1980) afin de faire de l'estimation sur les années 1963 à 1979 puis comparer les prévisions issu du modèle estimé aux vraies valeurs.

```
> n = 12
> index = 1 : (TY - n)
> annee = anneeY[index]
> T = length(annee)
> X = Y[index]
> X = as.ts(X)
> X = ts(X, start = 1963, frequency = 12)
> anneeY = seq(1963, 1980 + 11/12, by = 1/12)
> plot(annee, X, main = "Evolutiondu traficSNCFaucoursdutemps(1963-1980)", t =
"l", col = "blue", xlab = "temps", ylab = "nbvoyageurs")
```

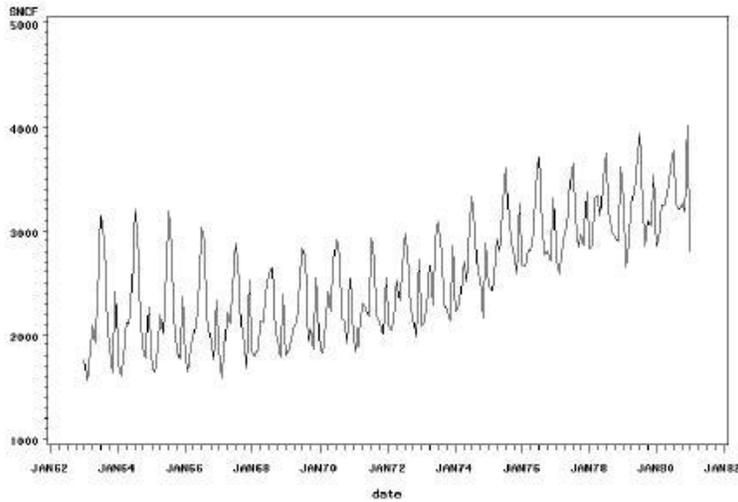


FIGURE 3.14 – Chronique SNCF

### 1. Stationnarité de la série

Traçons le corrélogramme et le corrélogramme partiel de  $X$ .

```
par(mfrow = c(1, 2))
```

```
> acf(X)
```

```
> pacf(X)
```

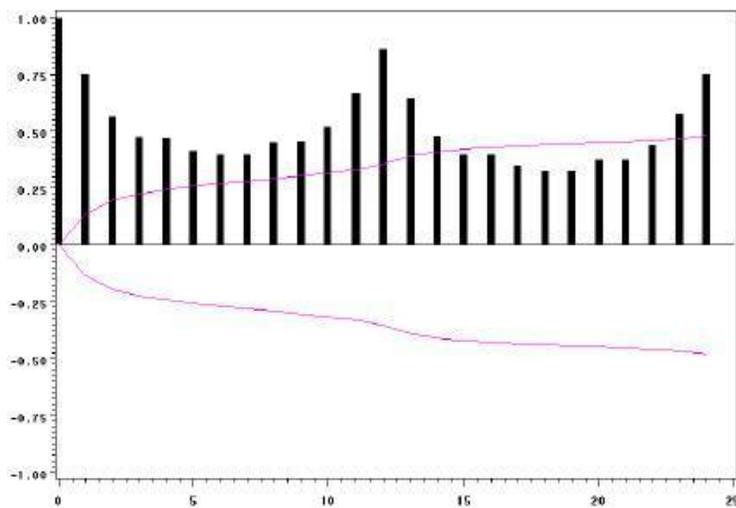


FIGURE 3.15 – Corrélogramme de chronique SNCF

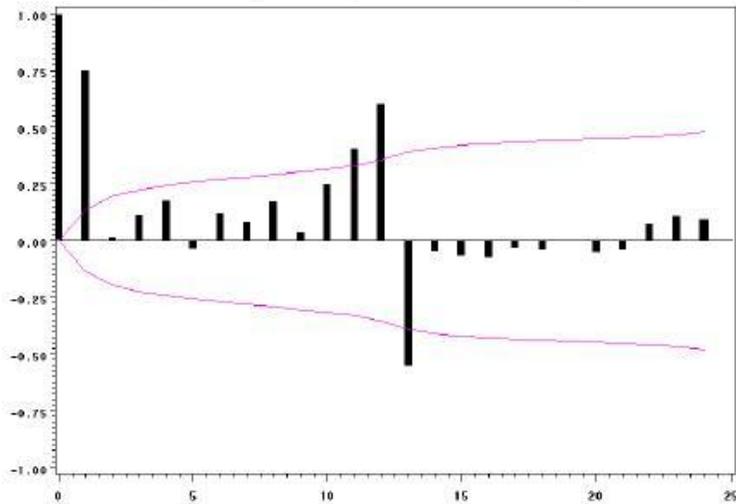


FIGURE 3.16 – Corrélogramme partiel de chronique SNCF

Les dessins *Figure3.15* et *Figure3.16* et *Figure3.17* montrent que la chronique présente une saisonnalité annuelle et suggèrent d'effectuer une différenciation saisonnière.

Le test du portmanteau (testant la non-corrélation de l'échantillon) rejette aussi la stationnarité de la série :

```
> Box.test(X)
```

## 2. Etude des séries $\text{diff}(X)$ et $\text{diff12}(\text{diff}(X))$

Nous allons maintenant travailler sur la série  $(D_t = X_t - X_{t-1})_{2 \leq t \leq T}$ . Sur le tracé de  $D_t$ , on constate que la série ne semble plus avoir de tendance (comportement très erratique).

```
> diffX = diff(X)
> T = length(annee)
> par(mfrow = c(1, 1))
```

```
> plot(annee[2 : T], diffX, main = "diffX", t = "l", col = "blue", xlab =
"temps", ylab = "diffX")
```

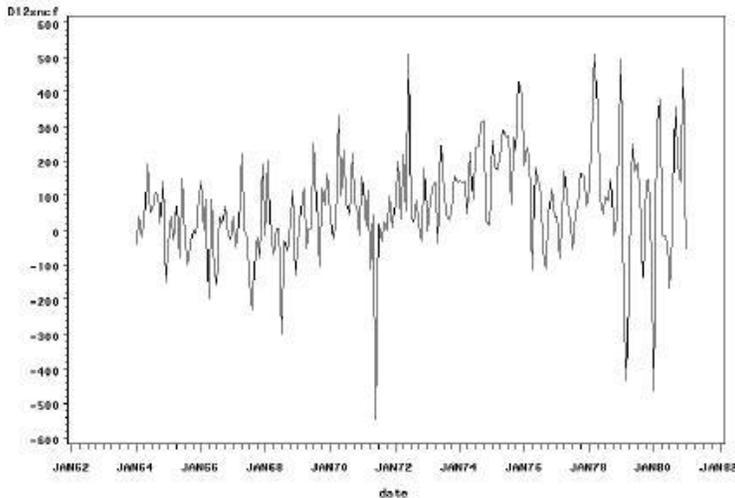


FIGURE 3.17 – Chronique des différences saisonnières

On peut aussi tracer les fonctions d'autocorrélation empiriques. Le corrélogramme estimé de la série différenciée  $(I - B)X$  montre de fortes corrélations pour les valeurs  $h$  multiples de 12. Ceci implique qu'il faut appliquer au moins une fois l'opérateur  $(I - B^{12})$  à la série des différences.

```
> par(mfrow = c(1, 2))
> acf(diffX, lag.max = 30)
> pacf(diffX, lag.max = 30)
> diff12X = diff(diff(X), lag=12)
> par(mfrow = c(1, 1))
> plot(annee[14 :T], diff12X, main = "S´erie Z", t = "l", col = "blue", xlab =
"temps", ylab = "Z")
> par(mfrow = c(1, 2))
> acf(diff12X, lag.max = 30)
> pacf(diff12X, lag.max = 30)
```

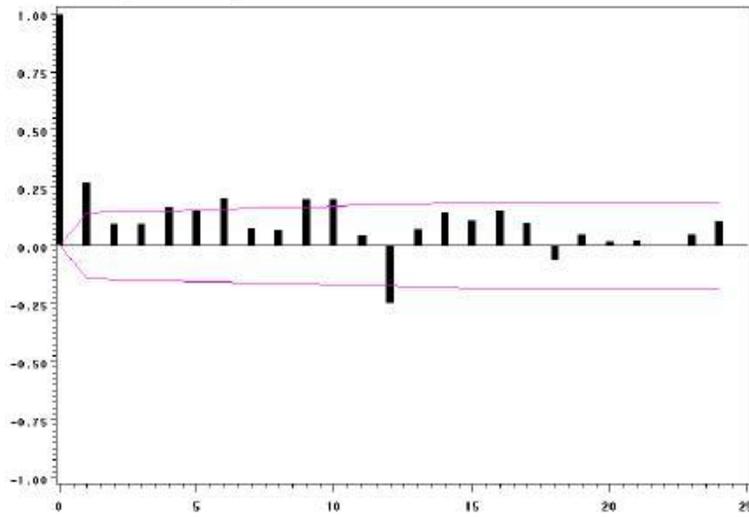


FIGURE 3.18 – Corrélogramme des différences saisonnières

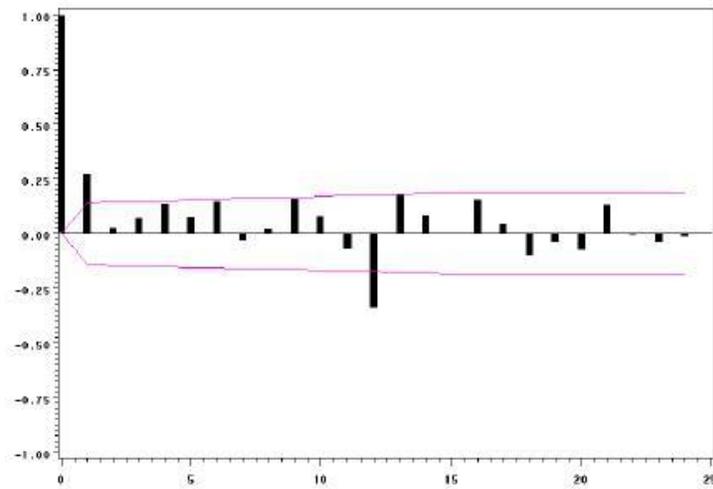


FIGURE 3.19 – Corrélogramme partiel des différences saisonnières

Les dessins *Figure3.18* et *Figure3.19* et *Figure3.20* après différenciation saisonnière d'ordre 12 montrent qu'il n'y a pas stationnarité et suggèrent une différenciation d'ordre 1 pour éliminer la tendance. Le dessin *Figure3.21* montre que ces différenciations sont suffisantes : il ne semble pas devoir mettre en cause la stationnarité. Les corrélogrammes *Figure3.22* et *Figure3.23* des différences d'ordres 1 et 12 (i-e  $(1 - B)(1 - B^{12})X_t$ ) montrent :

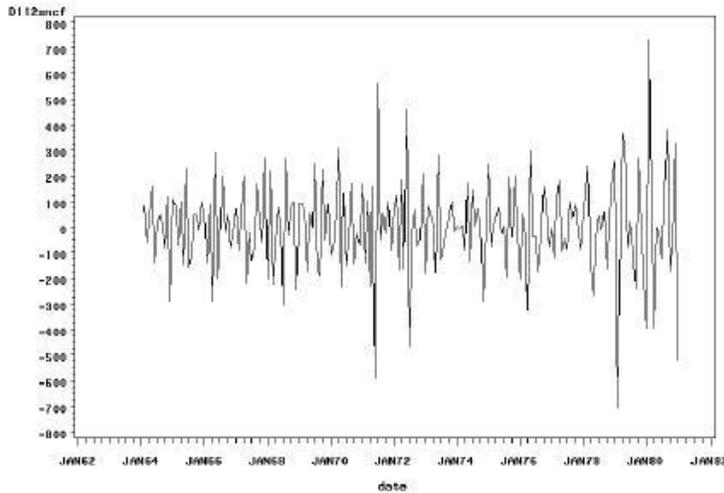


FIGURE 3.20 – Chronique différence deux fois

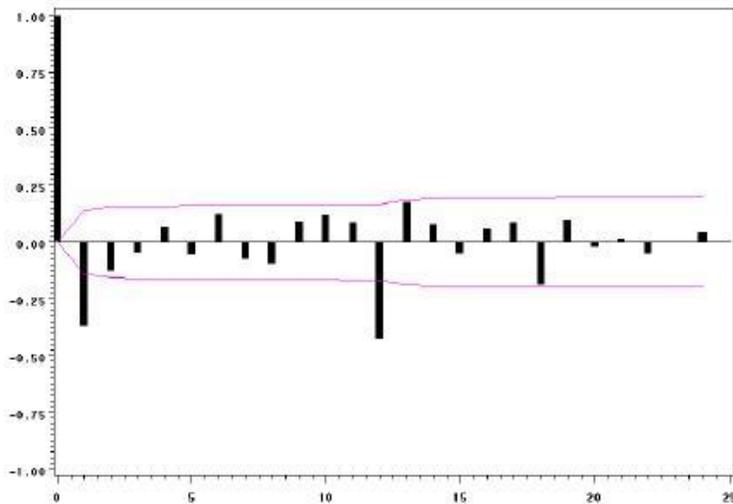


FIGURE 3.21 – Corrélogramme de la chronique différence deux fois

- une structure non saisonnière qui fait plutôt penser à un processus  $MA(1)$ ,
- une structure saisonnière qui semble plutôt correspondre à un processus  $MA$  saisonnier d'ordre 1. Si on hésitait entre  $AR$  et  $MA$ , il serait toujours possible de tester les deux. Finalement, on peut envisager un modèle  $(0, 1, 1)(0, 1, 1)_{12}$ , c'est-à-dire encore :

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \theta_2 B^{12})\epsilon_t.$$

### 3. Etude des résidus



FIGURE 3.22 – Corrélogramme partiel de la chronique différence deux fois

```

> residus = arima(X, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period =
12))residual
> par(mfrow = c(1, 1))
> plot(residus, main = "Residus au cours du temps", t = "l", col = "blue", xlab =
"temps", ylab = "residus")

```

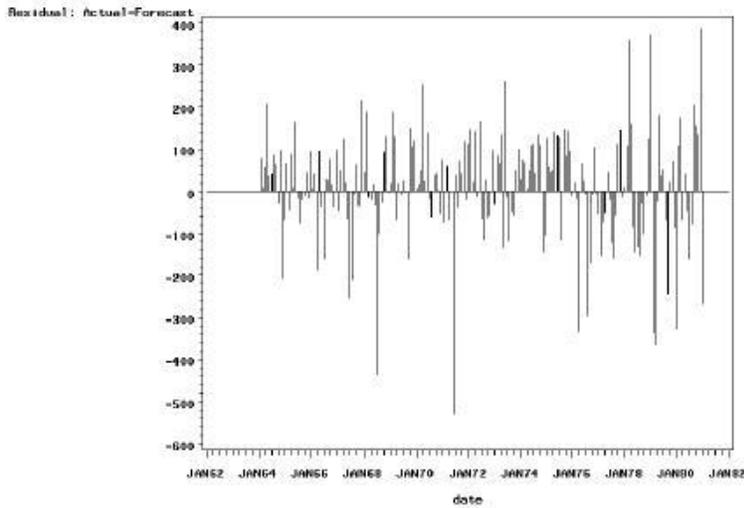


FIGURE 3.23 – Chronique des résidus

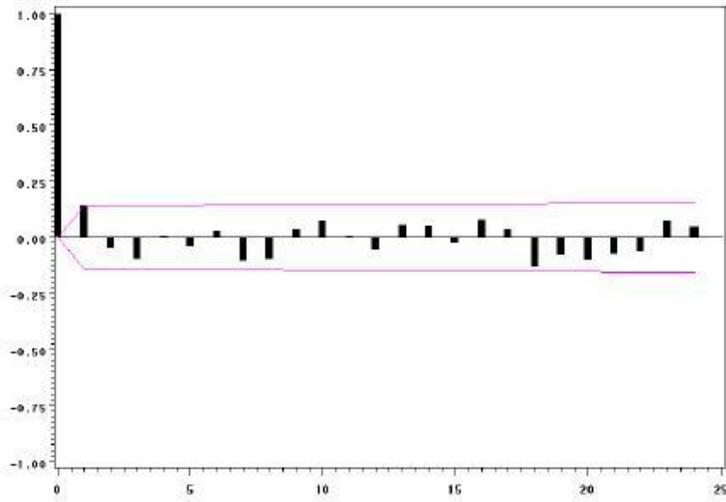


FIGURE 3.24 – Corrélogramme des résidus

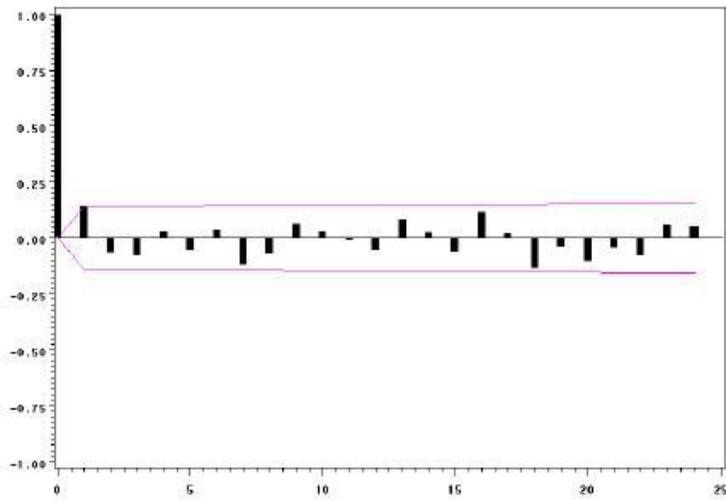


FIGURE 3.25 – Corrélogramme partiel des résidus

#### 4. Vérification

Le tableau ci-dessous donne les estimations des deux paramètres 1 et 2 du modèle Box et Jenkin  $ARIMA(0, 1, 1)(0, 1, 1)_{12}$  :

$$(1 - B)(1 - B^{12})X_t = (1 - 0.86B)(1 - 0.535B^{12})$$

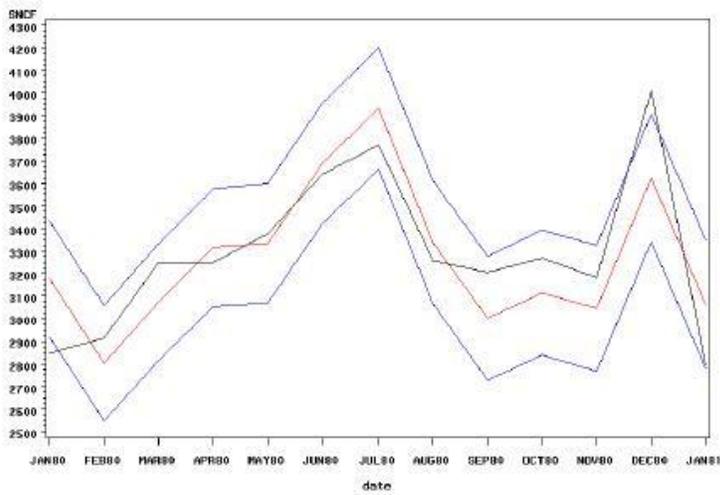
Modèle calculé sur les différences d'ordre 1 et 12				
Paramètre	Estimation	Ecart type	<i>t</i> de Student	Probabilité
MA[1]	0.860	0.036	23.91	0.000
MA[12]	0.535	0.064	8.30	0.000
Constante	0.561	0.644	0.87	0.386
Ecart type estimé des résidus : 130.6				
Portmanteau sur les 12 premières autocorrélations résiduelles : 12.18				
Probabilité d'une valeur plus grande pour un bruit blanc : 0.273				

Le portmanteau est égal à 12.25 (sur les 1 premiers termes du corrélogramme des résidus) et il lui correspond une probabilité 0.2687 largement supérieure à 0.05 : l'hypothèse d'indépendance des résidus n'est pas infirmée.

## 5. Prévisions

Cela n'a pas été fait ici pour calculer les prévisions. montre les prévisions jusqu'à l'horizon 12 et la plage de confiance à 95 leur correspond. On n'a pas fait intervenir les 12 dernières observations dans le calcul du modèle, ce qui permet de comparer prévisions et réalisations pour les 12 dernières observations.

```
> pred = predict(arima(X,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12),12)
> par(mfrow = c(1, 3))
> plot(anneeY[(TY - 4 * n) :TY], Y[(TY - 4 * n) :TY - 1], main = "prevision
ARMA(0,2)", t = "t", col = "blue", xlab = "temps", ylab = "X")
> lines(anneeY[(TY - n) :TY], c(X[T - n - 1], pred$pred),col="red")
> lines(anneeY[(TY - n) :TY], c(X[T - n - 1], pred$pred), c(0,pred$se)* 1.96,
lty = 2,col="green")lines(anneeY[(TY - n) :TY], c(X[T - n - 1], pred$pred) -
c(0,pred$se)*1.96, lty = 2,col="green")
```

FIGURE 3.26 – Prévisions par SARIMA  $(0, 1, 1)(0, 1, 1)_{12}$ .



# Conclusion

Au terme de cette étude sur la modélisation des séries chronologiques, par l'utilisation des méthodes de Box et Jenkins, nous avons épousé volontairement une approche pragmatique de Box Jenkins généralisé par les processus SARIMA. L'hypothèse de stationnarité est fondamentale dans la modélisation d'une série temporelle.

Le fait qu'une série soit stationnaire ou non, conditionne le choix de la modélisation que l'on doit adopter. En règle générale, si l'on s'en tient notamment à la méthodologie de Box et Jenkins ; si la série étudiée est stationnaire, on cherche alors le meilleur modèle parmi la classe des séries stationnaires pour la représenter puis on estime ce modèle.

En revanche, si la série est non stationnaire, on doit avant toutes choses chercher à la "stationnariser" c'est-à-dire trouver une transformation stationnaire de cette série. Puis la modéliser et estimer les paramètres associés au modèle.

Néanmoins, l'étude du problème de prévision et le rôle de la fonction d'auto-corrélation partielle dans l'étude des séries chronologiques, peuvent être étudiés en détail. Ainsi l'étude simultanée de plusieurs séries stationnaires (cas multidimensionnelle) et son importance dans la modélisation. Les tests de stationnarité méritent d'être développés car notre but de les mentionner n'a été qu'à des fins de simulation de séries.



# Bibliographie

- [1] Anderson O.D. (1976) :”Time Series Analysis and Forecasting : The BOX JENKINS Approach”, Butterworth, London.
- [2] Arthur. Charpentier. Cours de séries temporelles, Volume1 : Introduction à la théorie des processus en temps discret Modèles ARIMA et méthode *Box&Jenkins*.
- [3] Brockwell. P, & Davis. R. (1991). Time Series : Theory and Methods (Second ed.). New York : Springer.
- [4] Brockwell. P, & Davis. R. (1996) : Introduction to Time Series and Forecasting. New York : Springer
- [5] Box, G. E., & Jenkins, G. M. (1976). Time series analysis : forecasting and control (2nd ed.). San Francisco : Holden Day.
- [6] Clement M. P., Hendry D. F. (2000) : Forecasting Economic Time Series, Cambridge University Press, Cambridge.
- [7] Enders, W : Applied Econometric Time Series. 1995, Wiley, New York.
- [8] Eric. Moulines et François Roueff : Analyse des Séries Temporelles et Applications, 15 septembre 2010.
- [9] Gouriéroux, Ch., Monfort, A. : Séries Temporelles et Modèles Dynamiques. 1995, Economica, Paris.

- [10] Mills, T. : Time Series Techniques for Economists. 1990, University Press, Cambridge.
- [11] M. Gevers et L. Vandendorpe : Processus stochastiques, estimation et prédiction, Université catholique de Louvain Faculté des Sciences Appliquées.
- [12] 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
- [13] Pankartz, A : Forecasting with Univariate Box-Jenkins Models. 1983, Wiley, New York.
- [14] Pankartz, A. : Forecasting with Dynamic Regression Models. 1991, Wiley, New York.
- [15] Ralf Vandenhouten : Non-stationary Time Series Analysis of Complex Systems and Applications in Physiology, Aachen, Shaker Verlag GmbH 1998, (ISBN 3-8265-3814-5).
- [16] Shumway R.H. et Stoffer D.S : Times Series Analysis and Its Applications, With R Example, Springer, 2006.