

الجمهورية الجزائرية الديمقراطية الشعبية  
République algérienne démocratique et populaire  
وزارة التعليم العالي والبحث العلمي  
Ministère de l'enseignement supérieur et de la recherche scientifique  
جامعة عين تموشنت بلحاج بوشعيب  
Université –Ain Temouchent- Belhadj Bouchaib  
Faculté des Sciences et de Technologie  
Département de Mathématiques et Informatique



**Projet de Fin d'Etudes**  
**Pour l'obtention du diplôme de Master en Mathématiques**

**Domaine :** Mathématiques et Informatique  
**Filière :** Mathématiques  
**Spécialité :** Probabilités et Statistique Appliquées

**Thème**

**Les Séries Stationnaires Appliquées**

**Présenté Par :**

Mlle. BOUCHETA Anissa

**Devant le jury composé de :**

<b>M. MAMI Tawfik Fawzi</b>	<b>Pr</b>	UAT.B.B (Ain Temouchent)	Président
<b>M. HARIRI Mohammed</b>	<b>MCA</b>	UAT.B.B (Ain Temouchent)	Examineur
<b>M<sup>me</sup> . BENNAFLA Djamila</b>	<b>MCB</b>	UAT.B.B (Ain Temouchent)	Encadrante

*Année Universitaire 2024/2025*

## Résumé

---

Ce mémoire vise à étudier les séries chronologiques stationnaires, en particulier leurs propriétés, leur identification, leur modélisation et leur prévision. L'accent sera mis sur l'application des méthodes dans des domaines spécifiques. Le projet inclura une analyse pratique des données réelles et une implémentation sous R

## Abstract

---

This thesis aims to study stationary time series, with a particular focus on their properties, identification, modelling, and forecasting. Emphasis will be placed on the application of these methods in specific domains.

The project will include a practical analysis of real-world data and an implementation using R.

## ملخص

---

هدف هذا البحث هو دراسة معمقة للسلاسل الزمنية الثابتة مع التركيز على تحليل خصائصها الأساسية و تحديدها . بالإضافة الى التنبؤ بها . سيتم إيلاء اهتمام خاص للتطبيق العلمي لهذه الأساليب في السياقات الواقعية بحيث سيتضمن العمل تحليلاً تجريبياً لبيانات حقيقية و استخدام أدوات احصائية مناسبة ضمن برنامج R .

# Dédicaces

♡ *Je dédie* ce mémoire ♡  
*à ma chère mère et mon cher père*  
*qui ont toujours été à mes côtés,*  
*et n'ont pas cessés de m'encourager*  
*à ma chère sœur CHAIMAA*  
*à mes deux frères*  
*et Enfin à toutes*  
*mes amies*



# *Remerciements*

*Je remercie Dieu qui m'a aidé et donné la force pour terminer ce travail.*

*Je remercie ma famille, mon premier soutien, pour leur présence, leur aide et leurs encouragements.*

*Un merci spécial à mes parents, **mon père** et **ma mère**, pour leur amour et leur soutien, ainsi qu'à ma sœur **CHAIMAA** et mes deux frères **MOHAMED** et **BOUHAS** pour leur appui.*

*Je tiens à exprimer ma profonde gratitude à Dr. **BENNAFLA Djamil**a pour avoir accepté d'être mon encadrante.*

*Je la remercie sincèrement pour son soutien moral, ses conseils précieux et son accompagnement constant tout au long de la préparation de ce mémoire.*

*Je remercie les membres du jury Professeur **MAMI . T** et Dr. **HARIRI. M** pour avoir accepté d'évaluer ce travail et pour leurs remarques utiles.*

*Je remercie aussi mes enseignants pour leurs conseils et leur accompagnement tout au long de mon parcours.*

*Merci à mes camarades pour les bons moments et les efforts partagés.*

*Et enfin, merci à tous ceux qui m'ont aidé, même avec un simple mot ou un conseil.*

ANISSA

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Compréhension et Exploration des Séries Chronologiques</b>	<b>4</b>
1.1 Les Séries Chronologiques . . . . .	4
1.2 Exploration des Séries Chronologiques . . . . .	5
1.2.1 Structure et Décomposition des Séries chronologiques . . . . .	5
1.2.2 Les schémas de décomposition . . . . .	7
1.3 Outils et Techniques d'Exploration . . . . .	9
1.3.1 Correction des données . . . . .	10
1.3.2 La Modélisation . . . . .	13
1.3.3 Analyse de la série à partir de ses composantes . . . . .	15
1.3.4 Diagnostic du modèle . . . . .	16
1.3.5 Prédiction . . . . .	17
<b>2 Analyse et Identification des séries Stationnaires</b>	<b>18</b>
2.1 Nature et Propriétés des Séries Stationnaires . . . . .	18
2.1.1 Analyse de la Fonctions d'autocovariance . . . . .	20
2.1.2 Analyse de la Fonction d'Autocorrélation . . . . .	22
2.1.3 Analyse de la Fonction d'Autocorrélation partielle . . . . .	23
2.1.4 Densité spectrale . . . . .	25
2.2 Non stationnarité . . . . .	26
2.3 Les tests de Vérification de la Stationnarité . . . . .	27
2.3.1 Le test de Dickey-Fuller (DF) . . . . .	28
2.3.2 Le test de Dickey-Fuller augmentée (ADF) . . . . .	30
2.3.3 Le test de Phillips-Perron (PP) . . . . .	32
2.3.4 Le test de KPSS . . . . .	33
2.4 Techniques de Stationnarisation . . . . .	34
2.5 Modélisation des séries selon leur stationnarité . . . . .	39
2.6 Identification de modèle . . . . .	42
<b>3 Traitements et Modélisation Pratiques des Séries chronologiques</b>	<b>47</b>
3.1 Modélisation des données à l'aide du modèle ARMA . . . . .	47
3.1.1 Présentation et exploration des données . . . . .	47
3.1.2 Construction du modèle ARMA . . . . .	50
3.1.3 La prévision . . . . .	51
3.2 Modélisation des données à l'aide du modèle ARIMA . . . . .	52
3.2.1 Présentation et exploration des données . . . . .	52
3.2.2 Traitement de la non-stationnarité . . . . .	55

3.2.3	Construction du modèle ARIMA . . . . .	56
3.2.4	Prévision . . . . .	58
3.3	Modélisation des données à l'aide du modèle SARIMA . . . . .	59
3.3.1	Présentation et exploration des données . . . . .	59
3.3.2	Construction du modèle SARIMA . . . . .	61
3.3.3	Prévision . . . . .	62
3.4	Comparaison entre les modèles ARMA, ARIMA et SARIMA . . . . .	63
	<b>Conclusion</b>	<b>63</b>
	<b>Annexe</b>	<b>64</b>
	<b>Bibliographie</b>	<b>68</b>

# Table des figures

1.1	Les séries chronologiques $CO_2$ , Trafic aérien, Bitcoin, Température . . . .	5
1.2	Séries chronologiques avec différentes tendances . . . . .	6
1.3	Série avec saisonnalité . . . . .	6
1.4	Série comprenant des résidus . . . . .	7
1.5	Série avec événements imprévus . . . . .	7
1.6	Modèle additif . . . . .	8
1.7	Modèle multiplicatif . . . . .	8
1.8	Modèle mixte . . . . .	8
1.9	Comparaison entre le modèle additif et le modèle multiplicatif . . . . .	9
1.10	Représentation graphique des séries chronologiques . . . . .	11
1.11	ACF d'une série avec tendance . . . . .	11
1.12	ACF d'une série avec saisonnalité . . . . .	12
1.13	ACF d'une série avec absence d'autocorrélation . . . . .	12
1.14	PACF d'une série stationnaire . . . . .	13
1.15	PACF d'une série avec saisonnalité . . . . .	13
2.1	Trajectoire d'un bruit blanc . . . . .	19
2.2	Densité spectrale . . . . .	25
2.3	Exemples de séries non stationnaires . . . . .	27
2.4	Série avec tendance linéaire et sa différenciation première . . . . .	35
2.5	Transformation inverse : réduction de l'asymétrie extrême . . . . .	35
2.6	Série originale, transformation et différenciation logarithmique . . . . .	36
2.7	Transformation racine carrée et différenciation . . . . .	36
2.8	Graphiques de l'ACF et du PACF . . . . .	43
3.1	Évolution mensuelle de la production de ciment et sa décomposition . . . . .	48
3.2	ACF et PACF de la série de quantité mensuelle de ciment . . . . .	49
3.3	Résultats des tests de stationnarité (ADF, PP et KPSS) . . . . .	50
3.4	Ajustement du modèle ARMA à la série des quantités de ciment. . . . .	50
3.5	Les résidus du modèle ARMA(0,1) . . . . .	51
3.6	Prévisions des ventes de ciment à l'aide d'un modèle ARMA . . . . .	52
3.7	Évolution temporelle du taux de réussite entre 1976 et 2023 . . . . .	53
3.8	ACF et PACF de la série temporelle du taux de réussite . . . . .	54
3.9	Les tests de stationnarité (ADF, PP, KPSS) du taux de réussite . . . . .	54
3.10	La série différenciée du taux de réussite . . . . .	55
3.11	ACF et PACF de la La série différenciée du taux . . . . .	55
3.12	Les tests de stationnarité (ADF, PP, KPSS) après différenciation . . . . .	56
3.13	Ajustement du modèle ARIMA au taux de réussite. . . . .	57
3.14	Les résidus du modèle ARIMA(0,1,1) . . . . .	57

3.15	Prévisions du taux de réussite à l'aide d'un modèle ARIMA . . . . .	58
3.16	Résultats de la prévision du taux de réussite . . . . .	58
3.17	Évolution temporelle des importations de marchandises . . . . .	60
3.18	ACF et PACF de la série des des importations de marchandise . . . . .	60
3.19	Résultat du tests appliqué à la série des importations. . . . .	61
3.20	Ajustement du modèle SARIMA à la série des importations . . . . .	61
3.21	Les résidus du modèle SARIMA . . . . .	62
3.22	Les prévisions obtenues à l'aide du modèle SARIMA . . . . .	62

# Liste des tableaux

2.1	Valeurs critiques sous différents modèles ( $\alpha, \beta$ ) . . . . .	30
2.2	Valeurs critiques du test KPSS . . . . .	34
3.2	Quantités mensuelles de ciment (en tonnes) à Beni Saf (2013-2019) . . .	48
3.3	Taux de réussite au baccalauréat de 1976 à 2023 . . . . .	53
3.4	Les valeurs des importations de marchandises aux prix courants en Algérie	59

# Index des notations

Notation	Description
v.a.r.	Variable aléatoire réelle
i.e.	C'est à dire (id est)
i.i.d.'s	Indépendamment et identiquement distribué(e)s
MCO	Moindres carrés ordinaires
MV	Maximum de vraisemblance
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
DF	Dickey-Fuller
ADF	Augmented Dickey-Fuller
PP	Phillips-Perron
MM	Moyenne mobile
MA	Modèle à moyenne mobile (Moving Average)
AR	Modèle autorégressif (AutoRegressive)
ARMA	Modèle autorégressif à moyenne mobile
ARIMA	AutoRegressive Integrated Moving Average
SARIMA	Seasonal AutoRegressive Integrated Moving Average
AIC	Critère d'information d'Akaike
BIC	Critère d'information bayésien
BB	Bruit Blanc
SC	Critère de Schwarz
MAE	Erreur absolue moyenne
MSE	Erreur quadratique moyenne
RMSE	Racine carrée de l'erreur quadratique moyenne
MAPE	Écart absolu moyen en pourcentage

# Introduction

L'analyse des séries chronologiques constitue une discipline fondamentale de la statistique appliquée et des sciences des données. Elle a pour objectif l'étude de phénomènes évoluant dans le temps. Historiquement, dès le 19<sup>e</sup> siècle, des chercheurs tels que George Udny Yule ont commencé à formaliser les concepts associés aux données temporelles. Toutefois, ce n'est qu'au 20<sup>e</sup> siècle, avec l'essor des modèles autorégressifs et des méthodes d'identification de la stationnarité, que la modélisation rigoureuse des séries chronologiques s'est réellement structurée. Les travaux pionniers de Box et Jenkins ont joué un rôle majeur dans la diffusion et la vulgarisation de ces approches, aujourd'hui largement utilisées dans de nombreux domaines, tels que la finance, la climatologie, la production agricole ou encore l'ingénierie.

Le concept de stationnarité en analyse des séries temporelles ne s'est pas imposé par hasard. Il est apparu lorsque les chercheurs ont tenté de comprendre des phénomènes dynamiques tels que les prix de marché, les températures, la consommation ou la production. En manipulant ces données chronologiques, ils ont rapidement constaté que toutes les séries n'avaient pas un comportement similaire : certaines montraient des caractéristiques stables dans le temps, tandis que d'autres évoluaient de manière aléatoire ou suivaient des tendances marquées.

L'étude de la stationnarité a émergé au début du 20<sup>e</sup> siècle, notamment avec les avancées en statistique mathématique. Des chercheurs comme Yule et Slutsky ont mis en évidence l'existence de modèles capables de générer des séries aux propriétés régulières dans le temps. Cette distinction a permis de classer les séries temporelles en deux grandes catégories : les séries stationnaires, dont les propriétés statistiques (moyenne, variance, autocorrélation) restent constantes dans le temps, et les séries non stationnaires, pour lesquelles ces propriétés varient.

Dans l'analyse des séries chronologiques, l'un des défis majeurs réside dans la gestion des propriétés évolutives des données. La majorité des séries observées dans la réalité présente des caractéristiques non stationnaires, telles que des tendances, des variations saisonnières ou des ruptures de structure. Or, ces instabilités nuisent à la pertinence des outils classiques de modélisation statistique, qui reposent le plus souvent sur l'hypothèse de stationnarité.

C'est dans ce contexte que la méthodologie ARIMA, développée dans les années 1970 par Box et Jenkins, a marqué un tournant. Elle propose une démarche rigoureuse intégrant une phase de diagnostic indispensable : celle de la stationnarité. L'identification de la nature stationnaire ou non d'une série repose sur des tests statistiques, tels que les tests ADF ou KPSS, permettant d'évaluer la stabilité des caractéristiques du processus sous-jacent.

Lorsqu'une série se révèle non stationnaire, elle doit être transformée avant toute tentative de modélisation. Plusieurs techniques existent à cet effet : la différenciation permet d'éliminer les tendances, la transformation logarithmique atténue les effets de variance

---

croissante, et le centrage peut corriger certaines irrégularités. Ces étapes préalables ne sont pas de simples formalités, mais des conditions essentielles pour garantir la validité des modèles prévisionnels, tels que ARIMA ou SARIMA.

Ainsi, loin d'être une exigence théorique, la stationnarité conditionne directement la qualité des analyses et des prévisions produites à partir des séries temporelles.

Ainsi, la problématique à laquelle ce travail entend répondre est la suivante : Comment identifier, transformer et modéliser efficacement les séries chronologiques stationnaires afin d'en extraire des prévisions fiables et de mieux comprendre leur dynamique sous-jacente ? Cette problématique soulève plusieurs questions spécifiques :

- Quels outils permettent de diagnostiquer la stationnarité ?
- Quelles transformations peuvent être appliquées pour stationnariser une série ?
- Quelles méthodes d'estimation et de prévision conviennent aux séries stationnaires ?

Ce projet de fin d'études vise à :

- Fournir une compréhension approfondie des séries chronologiques, de leur structure interne et des outils d'exploration associés .
- Présenter les concepts fondamentaux relatifs à la stationnarité, ainsi que les méthodes permettant de l'identifier et d'y remédier .
- Développer des méthodes pratiques d'estimation et de modélisation adaptées aux séries stationnaires, illustrées par une application sur des données réelles.

Ce mémoire est structuré en trois chapitres principaux, afin de proposer une analyse cohérente et progressive des séries stationnaires et de leurs applications :

**Chapitre 1 :** Introduction aux séries chronologiques. Ce chapitre définit les séries temporelles et en explore les composantes fondamentales : tendance, saisonnalité, bruit aléatoire. Il propose également une première analyse graphique, permettant de mieux cerner le comportement des séries et d'évaluer leur stationnarité.

**Chapitre 2 :** Étude approfondie de la stationnarité. Ce chapitre expose les différents types de stationnarité, les tests statistiques permettant de la diagnostiquer, les méthodes de transformation des séries non stationnaires, et les modèles adaptés à chaque type de série.

**Chapitre 3 :** Application sur données réelles. Ce dernier chapitre met en œuvre l'ensemble des outils présentés à travers l'analyse d'une ou plusieurs séries chronologiques issues de contextes réels. Il inclut la vérification de la stationnarité, les transformations éventuelles, le choix d'un modèle adéquat, et les prévisions associées.

Ce mémoire s'achève par une conclusion générale qui synthétise les principaux résultats obtenus, met en évidence les contributions théoriques et pratiques du travail réalisé, et suggère des pistes de recherche futures, tant sur le plan méthodologique que sur celui des applications. La bibliographie, placée à la suite, recense l'ensemble des références, articles et sources mobilisés au cours de cette étude.

# Chapitre 1

## Compréhension et Exploration des Séries Chronologiques

Ce chapitre vise à introduire les fondements des séries chronologiques, en commençant par leur définition et leurs principales caractéristiques. Il développe ensuite les méthodes d'exploration graphique permettant d'identifier visuellement les tendances, les effets saisonniers ou les irrégularités. La suite du chapitre est consacrée à l'étude de la structure des séries, à leur décomposition, ainsi qu'aux outils de correction et de modélisation, en vue d'une meilleure compréhension et d'une estimation fiable des paramètres décrivant le comportement des données.

Les éléments développés s'appuient sur les références suivantes : [1],[3], [8],[9], [10],[14].

### 1.1 Les Séries Chronologiques

Étudier les séries chronologiques permet de comprendre l'évolution des données dans le temps, d'en identifier les principales caractéristiques, et de prévoir leurs valeurs futures. Pour assurer une modélisation rigoureuse, il est essentiel de commencer par définir le concept de processus stochastique.

**Définition 1.1.1.** Soit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ . Un processus stochastique est une famille  $X = \{X_t; t \in \mathcal{T}\}$  de variables aléatoires définies sur un même  $(\Omega, \mathcal{A}, \mathbb{P})$ .

**Remarque 1.1.1.** Si  $\mathcal{T} = \mathbb{N}$ , alors  $X$  correspond à un processus à temps discret.  
Si  $\mathcal{T} = \mathbb{R}$  ou  $\mathcal{T} = [0, 1]$ , alors  $X$  est un processus à temps continu.

**Définition 1.1.2.** Une série chronologique (ou série temporelle) est une suite d'observations d'un phénomène au cours du temps. Elle peut s'écrire sous la forme :

$$X_t = D_t + N_t.$$

Où :

$D_t$  est la composante déterministe.

$N_t$  est la composante aléatoire.

**Remarque 1.1.2.** Une série chronologique (temporelle) est une réalisation d'un processus stochastique  $\{X_t; t \in \mathcal{T}\}$  où  $\mathcal{T}$  est un ensemble dénombrable et totalement ordonné.

On trouve dans la figure 1.1 une représentation de plusieurs types de séries chronologiques.

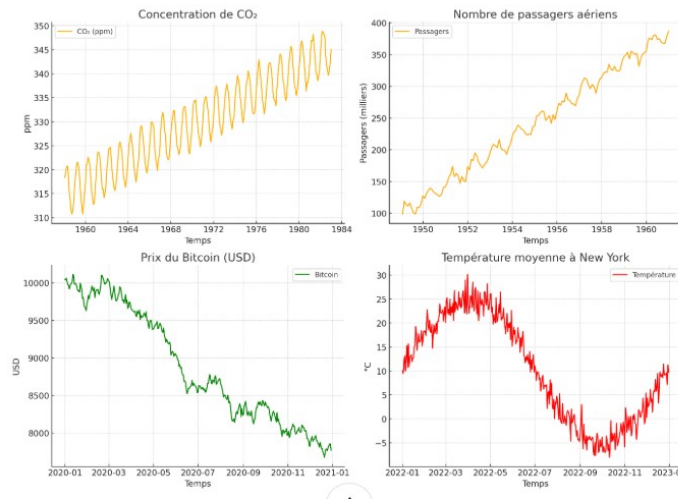


FIGURE 1.1 – Les séries chronologiques  $CO_2$ , Trafic aérien, Bitcoin, Température

Pour mieux comprendre la dynamique interne d'une série chronologique, il est nécessaire de passer à une exploration graphique et descriptive, afin d'en faire ressortir les principales composantes. C'est ce que nous abordons dans la section suivante.

## 1.2 Exploration des Séries Chronologiques

L'exploration des séries chronologiques constitue une étape préliminaire essentielle dans l'analyse des données temporelles. Elle permet de révéler la structure interne des séries, notamment la présence de tendances, de saisonnalités ou d'irrégularités, et d'en comprendre les caractéristiques principales. Cette phase exploratoire est indispensable pour orienter le choix des méthodes statistiques appropriées et préparer efficacement la modélisation des données.

### 1.2.1 Structure et Décomposition des Séries chronologiques

La plupart des séries présentent une structure commune constituée d'une composante déterministe (la tendance, la saisonnalité) et d'une composante aléatoire (les variations).

- **Tendance (trend)**

Une tendance ( $T_t$ ) représente la croissance ou décroissance à long terme de la série observée. Elle représente le comportement (moyen) de la série. Elle est généralement caractérisée par une fonction (linéaire, polynomiale, exponentielle, etc ...) en fonction du temps

a. **La tendance linéaire** est une forme la plus simple de tendance. la série évolue de manière linéaire, c'est-à-dire de façon croissante ou décroissante .

$$T_t = a + \beta t \quad \text{avec} \quad \beta, \alpha \in \mathbb{R}.$$

b. **La tendance polynomiale** est une forme plus flexible. Elle permet de modéliser l'évolution de la série à long terme à l'aide d'un polynôme en  $t$ , capturant des courbes ou des changements de direction.

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_q t^q \quad \text{avec} \quad \forall \beta_i \in \mathbb{R} \quad i = 0, \dots, q.$$

c. **La tendance exponentielle** utilisée pour les phénomènes avec une croissance ou décroissance rapide.

$$T_t = a \cdot e^{\beta t} \quad \text{ou} \quad T_t = a \cdot \beta^t \quad \text{avec} \quad \beta, \alpha \in \mathbb{R}..$$

La figure 1.2 permet de visualiser distinctement les différentes tendances.

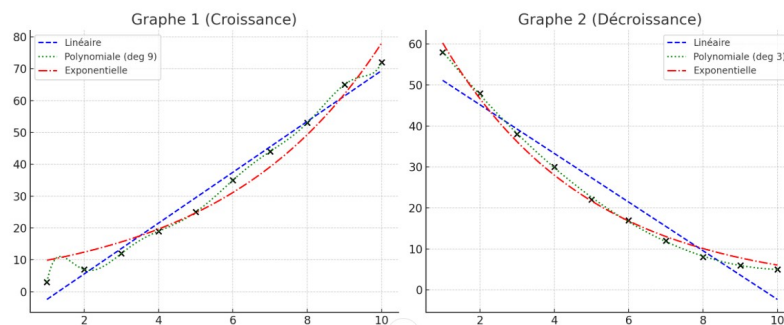


FIGURE 1.2 – Séries chronologiques avec différentes tendances

### • Saisonnalité (Seasonality)

Une saisonnalité ( $S_t$ ) est un phénomène qui se répète à intervalles réguliers, avec une composante constante à chaque période  $p$ , c'est-à-dire que  $S_{t+p_s} = S_t$ . Par exemple, une série est dite trimestrielle si  $p_s = 4$  et mensuelle si  $p_s = 12$ .

La figure 1.3 illustre les effets saisonniers de la série chronologique.

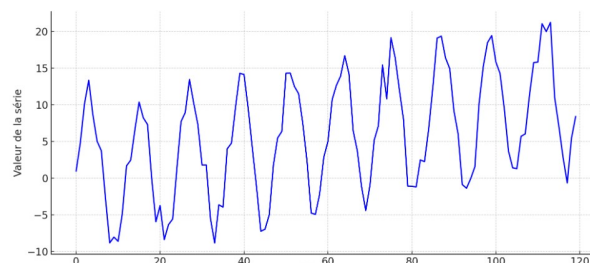


FIGURE 1.3 – Série avec saisonnalité

## • Résidus (Residual)

Les résidus ou bruits notés  $(\varepsilon_t)$  représentent la composante non structurée du phénomène. Ils sont modélisés par une série de variables aléatoires  $(\varepsilon_t)$  centrées, non corrélées et ayant une variance constante.

On observe dans la figure 1.4 la série chronologique accompagnée de ses résidus.

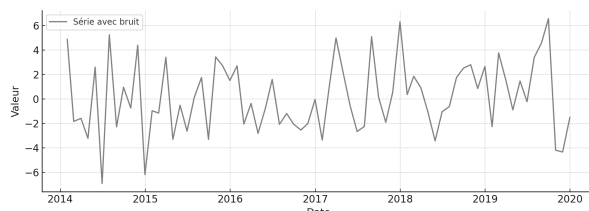


FIGURE 1.4 – Série comprenant des résidus

**Remarque 1.2.1.** *Il peut y avoir des événements imprévus, comme des grèves, des conditions climatiques extrêmes ou des crises économiques. Dans ce cadre, nous les incluons dans Les résidus  $\varepsilon_t$ .*

*L'impact des événements imprévus est bien visible dans la figure 1.5.*

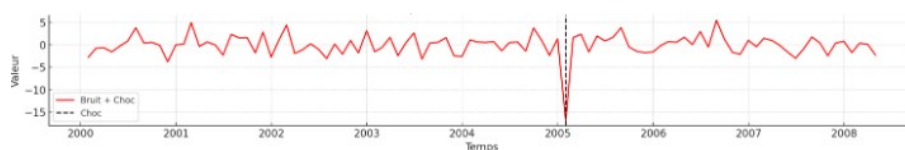


FIGURE 1.5 – Série avec événements imprévus

Après avoir identifié les composantes essentielles des séries chronologiques, il est utile de les illustrer à travers un schéma de composition, offrant une représentation claire et synthétique de leur influence respective sur la série observée.

## 1.2.2 Les schémas de décomposition

La décomposition des séries chronologiques s'effectue généralement selon trois schémas additif, multiplicatif ou mixte. Le choix du modèle dépend de la nature des données et de la manière dont la tendance et la saisonnalité interagissent. Ces approches permettent d'isoler distinctement les composantes de la série, facilitant ainsi son analyse et l'amélioration des prévisions.

### • Modèle additif

Ce modèle représente la série chronologique selon la formulation suivante :

$$X_t = T_t + S_t + \varepsilon_t.$$

La figure 1.6 illustre clairement la décomposition des différentes composantes.

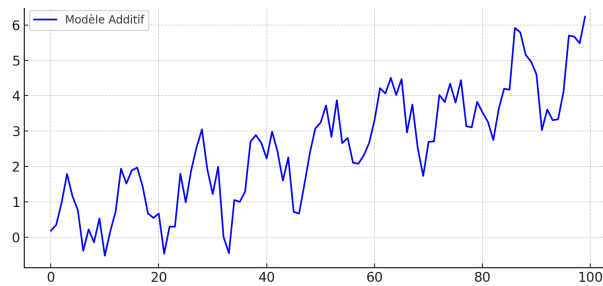


FIGURE 1.6 – Modèle additif

## • Modèle multiplicatif

La série chronologique est modélisée dans ce cadre selon l'expression suivante :

$$X_t = T_t \times S_t \times \varepsilon_t.$$

La décomposition des différentes composantes est mise en évidence dans la figure 1.7

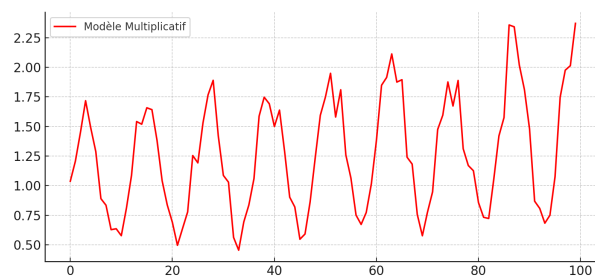


FIGURE 1.7 – Modèle multiplicatif

## • Modèle mixte

Selon ce modèle, la série chronologique prend la forme suivante :

$$X_t = T_t \times S_t + \varepsilon_t.$$

La figure 1.8 présente de manière claire les différentes composantes décomposées.

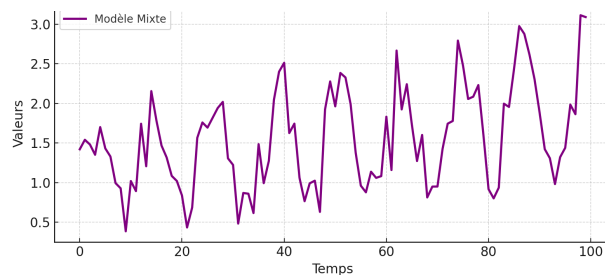


FIGURE 1.8 – Modèle mixte

**Remarque 1.2.2.** *L'identification du modèle repose sur l'observation du graphique et des lignes reliant les points minimum et maximum. Lorsque ces lignes sont proches d'être parallèles, le modèle est additif; sinon, il s'agit d'un modèle multiplicatif.*

La figure 1.9 montre comment distinguer entre le modèle additif et le modèle multiplicatif.

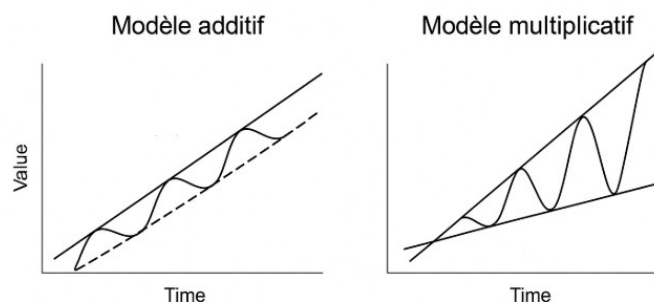


FIGURE 1.9 – Comparaison entre le modèle additif et le modèle multiplicatif

Ainsi, une fois les composantes de la série correctement identifiées à l'aide des schémas de décomposition (additif, multiplicatif ou mixte), il est nécessaire de renforcer cette analyse par des outils statistiques appropriés. Ces outils permettent de valider les observations préliminaires et d'orienter rigoureusement la modélisation. C'est dans ce cadre que s'inscrit la section suivante, dédiée aux principaux outils et techniques d'exploration des séries chronologiques.

### 1.3 Outils et Techniques d'Exploration

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

1. **Correction des données** : traitement des valeurs manquantes, aberrantes ou incohérentes.
2. **Observation de la série** : visualisation graphique pour détecter des tendances, des cycles ou une saisonnalité.
3. **Modélisation** : formulation d'un modèle avec un nombre fini de paramètres.
4. **Analyse de la série à partir de ses composantes** : décomposition en tendance, saisonnalité, bruit, etc.
5. **Diagnostic du modèle** : évaluation de l'adéquation entre le modèle et les données observées.
6. **Prédiction (prévision)** : utilisation du modèle pour anticiper les valeurs futures de la série.

### 1.3.1 Correction des données

Avant de se lancer dans l'étude d'une série chronologique, il est souvent nécessaire de traiter ou modifier les données brutes. Parmi les opérations couramment effectuées, on peut citer :

- **Évaluation et traitement des données manquantes** : remplacement de valeurs absentes ou accidentelles.
- **Découpage en sous-séries** : division de la série initiale en périodes homogènes ou significatives.
- **Standardisation** : transformation des données pour obtenir des intervalles de longueur fixe. Par exemple, pour des données mensuelles, on peut ramener chaque mois à un mois standard en calculant la moyenne journalière (total des observations du mois divisé par le nombre de jours dans le mois).
- **Transformation des données** : certaines analyses nécessitent l'utilisation de données transformées. En économie par exemple, on utilise fréquemment la famille de transformations de Box-Cox :

$$Y_t = \begin{cases} \frac{(X_t)^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \ln(X_t), & \text{si } \lambda = 0 \end{cases} \quad \text{avec } \lambda \in \mathbb{R}^*.$$

### 1.3.1 Visualisation graphique

La représentation graphique constitue une première étape essentielle dans l'analyse des séries chronologiques. Elle permet de visualiser l'évolution des données dans le temps et de détecter rapidement des caractéristiques importantes, telles que :

- Une tendance lorsque la série présente une croissance ou une décroissance globale au fil du temps.
- Une saisonnalité : lorsque certains motifs se répètent à intervalles réguliers.
- Des ruptures changements soudains dans la dynamique de la série.
- Des valeurs aberrantes points qui s'écartent fortement du comportement général.

Grâce à cette approche visuelle, on peut également poser un diagnostic préliminaire sur la stationnarité de la série. Une série est dite stationnaire si ses fluctuations s'organisent autour d'une moyenne constante, sans tendance ni saisonnalité marquée. Dans le cas contraire, on parlera de non-stationnarité, nécessitant des transformations avant toute modélisation.

Ainsi, le graphique d'une série temporelle ne se limite pas à une simple visualisation : il constitue un outil d'inspection précieux pour guider les choix méthodologiques et orienter les analyses statistiques à venir.

La figure 1.10 propose des exemples de séries chronologiques accompagnés de leur diagnostic visuel (stationnarité, tendance, saisonnalité, etc. . .), illustrant l'intérêt de cette représentation pour l'analyse exploratoire.

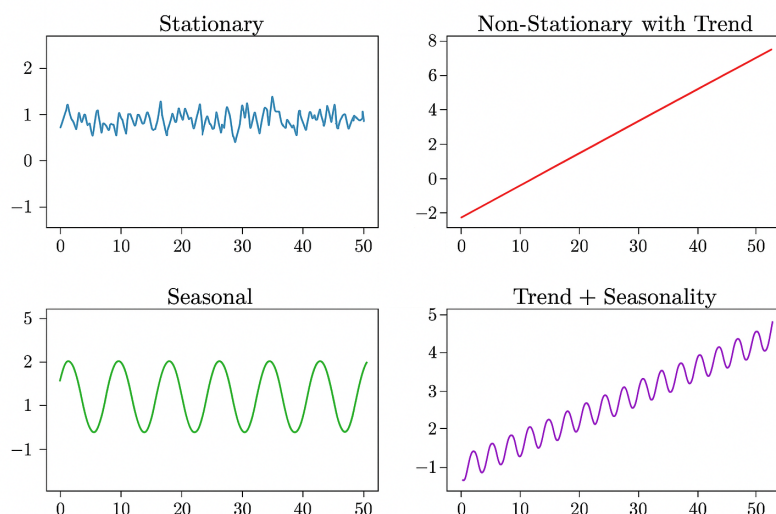


FIGURE 1.10 – Représentation graphique des séries chronologiques

Pour compléter cette première lecture visuelle, analysons maintenant la fonction d'autocorrélation.

• **Analyse graphique de la fonctions d'autocorrélation**

La fonction d'autocorrélation (ACF) mesure la corrélation entre les valeurs actuelles d'une série temporelle  $X_t$  et ses valeurs passées  $X_{t-k}$ , à différents décalages  $k$ . Elle permet de détecter les structures de dépendance dans les données.

1. Dans la figure 1.11, On observe que les valeurs de l'ACF diminuent progressivement et lentement, dépassant la zone hachurée (intervalle de confiance à 95%), ce qui indique une autocorrélation significative et suggère la présence d'une tendance dans la série.

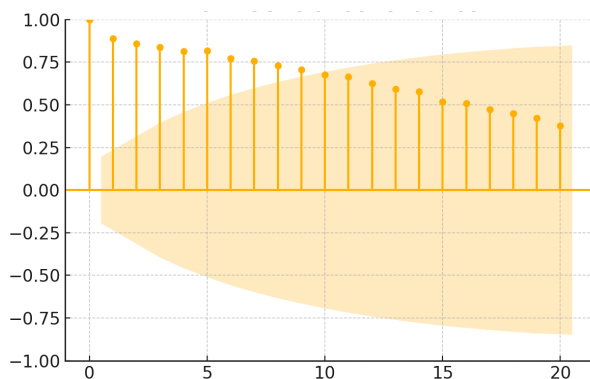


FIGURE 1.11 – ACF d'une série avec tendance

2. Dans la figure 1.12, On observe que l'ACF présente des pics périodiques à intervalles réguliers, dépassant la zone hachurée, ce qui traduit la présence d'une saisonnalité dans la série.

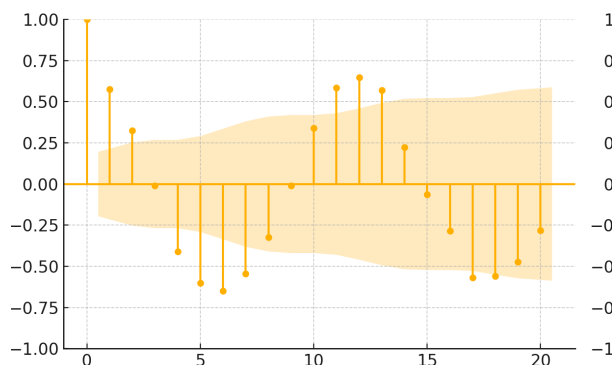


FIGURE 1.12 – ACF d'une série avec saisonnalité

3. Dans la figure 1.13, on remarque que la plupart des valeurs restent dans les bandes de confiance, ce qui signifie que la série n'a pas de dépendance significative à ces décalages.

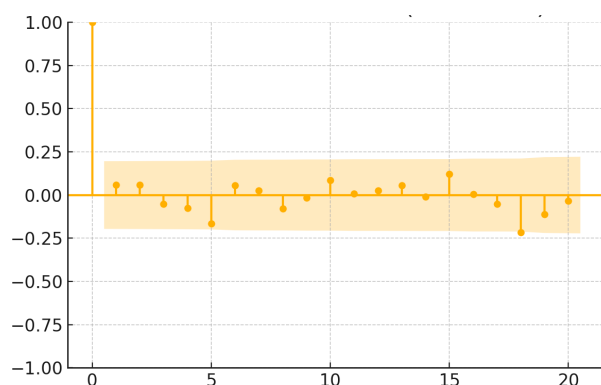


FIGURE 1.13 – ACF d'une série avec absence d'autocorrélation

**Remarque 1.3.1.** *Le lag 0 correspond à la corrélation de la série avec elle-même ; sa valeur est toujours égale à 1 par définition. Il est présent systématiquement et sert de référence pour comparer les autres lags (1, 2, 3, etc.), mais n'est pas interprété en termes de dépendance temporelle. La zone hachurée représente l'intervalle de confiance à 95%. Lorsque les barres de l'ACF dépassent cette zone, cela indique une autocorrélation significative aux lags correspondants, ce qui permet de détecter des structures dans la série, telles qu'une tendance ou une saisonnalité.*

En plus de l'autocorrélation, il est souvent nécessaire d'analyser les corrélations directes entre les observations, sans l'influence des valeurs intermédiaires. Dans ce contexte, la fonction d'autocorrélation partielle (PACF) est utilisée.

- **Analyse graphique de la fonctions d'autocorrélation partielle**

La fonction d'autocorrélation partielle (PACF) mesure la corrélation entre  $X_t$  et  $X_{t-k}$  en éliminant l'effet des valeurs intermédiaires  $X_{t-1}, \dots, X_{t-k+1}$ . Elle permet ainsi de capturer la relation directe entre deux points séparés par  $k$  périodes.

L'observation du graphique PACF permet d'analyser la structure de dépendance de la série.

1. Dans la figure 1.14, on remarque que toutes les valeurs restent dans les bandes de confiance après un certain décalage, ce qui signifie que les décalages supplémentaires n'apportent pas d'informations nouvelles, ce qui peut indiquer une série stationnaire.

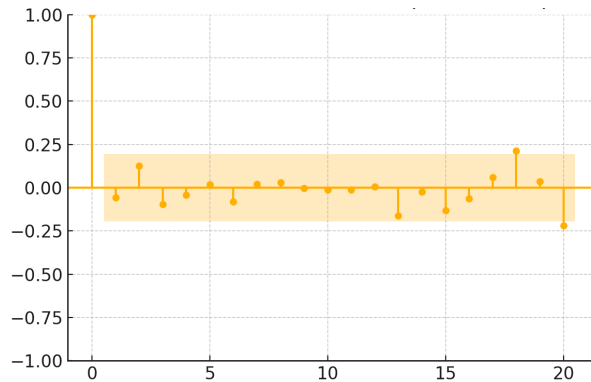


FIGURE 1.14 – PACF d'une série stationnaire

2. Dans la figure 1.15, on remarque que la PACF montre des variations périodiques, ce qui peut indiquer la présence d'une saisonnalité dans les données.

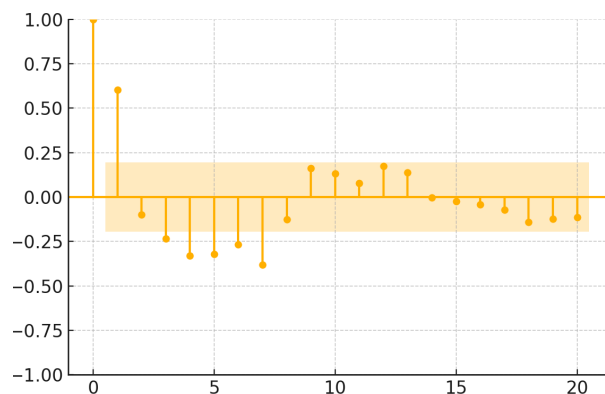


FIGURE 1.15 – PACF d'une série avec saisonnalité

La modélisation prend le relais de la visualisation pour décrire formellement la dynamique de la série.

### 1.3.2 La Modélisation

Au-delà des formes structurelles (additive, multiplicative ou mixte), les modèles de séries chronologiques peuvent également être classés selon leur nature : déterministes ou stochastiques. Cette distinction repose sur la manière dont les composantes aléatoires sont prises en compte dans la modélisation du phénomène étudié.

#### • Le modèles déterministes

Relèvent de la statistique descriptive. Ils considèrent que l'observation d'une série temporelle à un instant donné  $t$  est le résultat d'une fonction du temps et d'un terme

d'erreur  $\varepsilon_t$ , supposé centré et décorrélé :

$$X_t = f(t, \varepsilon_t).$$

La structure de ces modèles suppose que les composantes (tendance, saisonnalité) sont fixes et observables, tandis que les erreurs sont indépendantes dans le temps. Deux schémas déterministes classiques sont largement utilisés :

— **Modèle additif déterministe :**

$$X_t = T_t + S_t + \varepsilon_t.$$

Où  $T_t$  représente la tendance,  $S_t$  la saisonnalité, et  $\varepsilon_t$  les résidus. Ce modèle est adapté lorsque l'amplitude de la saisonnalité reste constante au fil du temps.

— **Modèle multiplicatif déterministe :**

$$X_t = T_t \times (1 + S_t) \times (1 + \varepsilon_t) \quad \text{ou bien} \quad X_t = T_t \times S_t \times \varepsilon_t.$$

Ce modèle convient lorsque l'effet saisonnier croît ou décroît proportionnellement à la tendance.

### • Les modèles stochastiques

Quant à eux, considèrent que le bruit  $\varepsilon_t$  n'est pas indépendant, mais corrélé dans le temps. Il est modélisé comme une fonction des valeurs passées et d'un ensemble de valeurs aléatoires indépendantes, centrées autour de zéro, ne présentant aucune dépendance temporelle apparente.  $\eta_t$  :

$$\varepsilon_t = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \eta_t).$$

Cela signifie que la composante aléatoire a une structure dynamique propre, que l'on cherche à estimer. La classe la plus utilisée est celle des modèles SARIMA (et ses sous-classes ARIMA, ARMA, etc...), particulièrement adaptée aux séries stationnaires ou rendues stationnaires par différenciation.

Quelle que soit la nature du modèle choisi, l'analyse des séries chronologiques suit généralement trois étapes :

1. Identification des composantes : mise en évidence de la tendance, saisonnalité, et bruit.
2. Estimation des composantes via ajustement du modèle choisi.
3. Élimination de ces composantes : détendancialisation et désaisonnalisation, afin d'isoler la série résiduelle  $\varepsilon_t$ .

### Remarque 1.3.2.

— *Dans les modèles déterministes, la série est décrite à l'aide de fonctions explicites du temps (comme une tendance linéaire, polynomiale ou une saisonnalité connue). Une fois ces composantes extraites, la série résiduelle est supposée décorrélée, c'est-à-dire qu'elle ne présente pas de dépendance temporelle significative. Elle est alors assimilée à un bruit blanc et ne nécessite pas de modélisation supplémentaire.*

*Exemple : une série modélisée par  $X_t = a + bt + \varepsilon_t$ , où  $\varepsilon_t$  est un bruit blanc.*

- Dans les modèles stochastiques, la série résiduelle ne se résume pas à un simple bruit. Elle est supposée stationnaire (ou rendue stationnaire), et doit être modélisée à son tour par un processus temporel, tel qu'un modèle autorégressif (AR), à moyenne mobile (MA), ou mixte (ARMA, ARIMA). L'analyse du comportement des résidus permet alors de capturer la structure aléatoire persistante de la série.

*Exemple : une série  $X_t$  avec tendance supprimée, dont les résidus suivent un modèle  $AR(1)$  :  $\varepsilon_t = \phi\varepsilon_{t-1} + u_t$ , avec  $u_t$  bruit blanc.*

Une fois l'étape de modélisation effectuée, l'analyse des séries temporelles se poursuit en étudiant séparément les différentes composantes du modèle adopté.

### 1.3.3 Analyse de la série à partir de ses composantes

L'analyse de la tendance permet de mettre en évidence l'évolution globale de la série au cours du temps (stabilité, croissance, décroissance, etc. ...).

Les facteurs saisonniers révèlent des variations périodiques systématiques selon la période de l'année (ou toute autre périodicité).

L'étude des fluctuations irrégulières permet d'identifier des perturbations exceptionnelles qui ne peuvent être attribuées ni à la tendance ni à la saisonnalité.

#### • Série corrigée de la tendance

La tendance introduit une corrélation forte entre les observations, mais cette corrélation n'a pas nécessairement de signification explicative. Il est donc important d'isoler la tendance, de l'étudier à part, puis de la retirer de la série afin d'analyser les relations restantes.

La série corrigée de la tendance, notée  $(X_{CST,t})$ , est définie par :

- Modèle additif :  $X_{CST,t} = S_t + \varepsilon_t$ .
- Modèle multiplicatif :  $X_{CST,t} = S_t(1 + \varepsilon_t)$ .

#### • Série corrigée des variations saisonnières

De manière analogue, il est nécessaire de corriger la série des variations saisonnières afin d'éliminer les effets liés aux comportements périodiques. Cela permet de mieux interpréter les variations conjoncturelles et d'éviter de fausses conclusions.

La série désaisonnalisée, notée  $(X_{CVS,t})$ , est définie par :

- Modèle additif :  $X_{CVS,t} = T_t + \varepsilon_t$ .
- Modèle multiplicatif :  $X_{CVS,t} = T_t(1 + \varepsilon_t)$ .

Cette série permet une meilleure comparaison des valeurs d'une période à une autre en neutralisant les effets saisonniers.

#### • Série lissée des prédictions

Enfin, la série lissée des prédictions, notée  $(X_{SLP,t})$ , est obtenue en supprimant les fluctuations irrégulières du modèle. Elle constitue la base des prévisions, car elle représente la composante régulière de la série (tendance + saisonnalité).

Elle est définie par :

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

Cette série sert de support aux anticipations futures dans les modèles de prévision.

### 1.3.4 Diagnostic du modèle

Une fois la structure du modèle choisie (déterministe ou stochastique, additif ou multiplicatif, etc.), il convient de procéder à l'estimation de ses paramètres, puis à la vérification de la qualité de l'ajustement.

#### Estimation des paramètres

L'estimation consiste à déterminer les valeurs numériques des paramètres du modèle (par exemple : coefficients de tendance, paramètres AR/MA, effets saisonniers, etc.) à partir des données observées.

Selon le type de modèle, différentes méthodes d'estimation sont utilisées :

- Pour les modèles déterministe (tendance linéaire ou polynomiale), les paramètres sont estimés à l'aide de la méthode des moindres carrés ordinaires (MCO), qui consiste à minimiser la somme des carrés des écarts entre les valeurs observées et ajustées.
- Pour les modèles stochastiques (tels que AR, MA, ARMA ou ARIMA), l'estimation repose généralement sur la méthode de la vraisemblance maximale (MV ou MLE), qui vise à maximiser la probabilité d'observer les données sous le modèle retenu.

Les détails techniques relatifs à ces deux méthodes d'estimation sont présentés en **annexe**.

#### Validation de l'ajustement

Une fois les paramètres estimés, il est indispensable de valider l'adéquation du modèle aux données. Cette validation repose sur deux approches complémentaires :

- Étude des résidus : Les résidus représentent les écarts entre les valeurs observées et les valeurs ajustées par le modèle. Un modèle bien ajusté doit produire des résidus :
  - non autocorrélés ;
  - de moyenne nulle ;
  - de variance constante (homoscédasticité).

Ce comportement est typique d'un bruit blanc, indiquant que toute la structure de la série a été correctement modélisée.

- Tests statistiques sur les résidus : Pour confirmer que les propriétés attendues des résidus sont bien respectées, plusieurs tests sont appliqués :
  - le test de *Ljung-Box* pour détecter une autocorrélation résiduelle ;
  - le test de *Jarque-Bera* pour vérifier la normalité des résidus ;
  - d'autres tests complémentaires peuvent être utilisés selon les hypothèses spécifiques du modèle.

Une présentation détaillée de ces tests et de leur interprétation est fournie en **annexe**.

### 1.3.5 Prédiction

Enfin, une fois l'ensemble des étapes précédentes complétées, nous sommes en mesure de passer à la phase de prédiction.

La prédiction constitue l'un des objectifs majeurs de la modélisation des séries chronologiques. Elle permet d'estimer les valeurs futures de la série sur la base de la structure temporelle identifiée dans les données historiques : tendance, saisonnalité et dépendances stochastiques.

#### Types de prévisions

Les prévisions peuvent être formulées selon différents horizons :

- À court terme, elles permettent d'anticiper l'évolution immédiate de la série (par exemple : pour le suivi opérationnel, la gestion de stocks ou la planification à très court terme).
- À moyen ou long terme, elles servent à dégager des tendances générales, utiles dans un cadre stratégique ou décisionnel.

#### Forme des prévisions

- Les modèles déterministes produisent généralement des prévisions sous forme de valeurs ponctuelles, fonction uniquement du temps (par exemple :  $Y_t = a + bt$ ).
- Les modèles stochastiques, en revanche, fournissent des prévisions probabilistes : ils permettent non seulement d'estimer une valeur future attendue, mais aussi de construire un intervalle de confiance autour de cette estimation, traduisant l'incertitude croissante à mesure que l'horizon de prévision s'allonge. Cela permet d'évaluer le risque associé aux différents scénarios futurs.

#### Qualité des prévisions

La qualité des prévisions dépend étroitement de plusieurs facteurs :

- la validité des hypothèses sur les résidus (absence d'autocorrélation, normalité, variance constante) ;
- la stabilité structurelle du modèle dans le temps (absence de ruptures ou de changements de régime) ;
- et la pertinence du modèle retenu au regard de la dynamique réelle de la série.

Dans la suite de ce travail, nous nous focaliserons plus particulièrement sur l'analyse et l'identification des séries stationnaires, condition préalable essentielle à la modélisation stochastique. En effet, la stationnarité constitue une hypothèse clé pour garantir la validité des prévisions à partir de modèles tels que ARMA ou ARIMA. Ainsi, une attention particulière sera portée à la vérification de cette propriété, à travers des outils graphiques et des tests statistiques dédiés.

# Chapitre 2

## Analyse et Identification des séries Stationnaires

Ce chapitre est dédié à l'analyse et à l'identification des séries stationnaires, qui jouent un rôle central dans la modélisation des séries chronologiques. Après avoir défini la nature et les propriétés de ces séries, nous étudierons les outils d'analyse tels que les fonctions d'autocovariance, d'autocorrélation, d'autocorrélation partielle et la densité spectrale. Nous aborderons ensuite les situations de non-stationnarité, ainsi que les principaux tests statistiques permettant de vérifier la stationnarité, notamment les tests de Dickey-Fuller, ADF, Phillips-Perron et KPSS. Le chapitre se termine par les techniques de transformation permettant de stationnariser une série et par les premières étapes de modélisation, en particulier l'identification des modèles adaptés aux données observées.

Les concepts fondamentaux abordés ici reposent essentiellement sur les sources suivantes : [4],[6],[7],[12],[13],[15],[17].

### 2.1 Nature et Propriétés des Séries Stationnaires

La stationnarité est une propriété d'une série chronologique qui signifie que son comportement (caractéristique) ne dépend pas du temps. Autrement dit, une série  $(X_t)_{t \in \mathcal{T}}$  est dite stationnaire si elle ne présente ni tendance saisonnière, ni tendance à la hausse ou à la baisse.

**Définition 2.1.1.** Une série chronologique  $(X_t)_{t \in \mathcal{T}}$  est dite stationnaire au second ordre (ou faiblement stationnaire, ou même simplement stationnaire) si sa moyenne  $\mu$  et sa fonction de covariance  $\gamma(t, s)$  sont invariantes par translation dans le temps. Autrement dit, pour tout  $t, s \in \mathcal{T}$  et tout décalage temporel  $k \in \mathbb{Z}$ , les conditions suivantes sont satisfaites :

1.  $\mathbb{E}[X_t] = \mu$ , constante (la moyenne ne dépend pas du temps)
2.  $\mathbb{E}[X_t^2] < \infty$  (la variance est finie)
3.  $\text{Cov}(X_t, X_{t+k}) = \gamma(k)$ , dépend uniquement du décalage  $k$ , et non des temps  $t$  ou  $s$

Il existe une notion plus forte de stationnarité.

**Définition 2.1.2.** Soit  $(X_t)_{t \in \mathcal{T}}$  une série chronologique. On dit que cette série est fortement (ou strictement) stationnaire si, pour tout entier  $n \in \mathbb{N}$ , pour tous instants  $t_1, \dots, t_n \in \mathcal{T}$ , et pour tout décalage  $k \in \mathbb{Z}$  tel que  $t_1 + k, \dots, t_n + k \in \mathcal{T}$ , le vecteur aléatoire  $(X_{t_1}, \dots, X_{t_n})$  a la même loi que  $(X_{t_1+k}, \dots, X_{t_n+k})$ .

**Lemme 2.1.1** Si la série  $(X_n)$  est fortement stationnaire et que  $\mathbb{E}(X_n^2) < \infty$ , alors  $(X_n)$  est faiblement stationnaire. La réciproque est fautive en général. Toutefois, si  $(X_n)$  est un processus gaussien, alors la loi jointe  $\mathcal{L}(X_{n_1}, X_{n_2}, \dots, X_{n_k})$  est une loi gaussienne, et les notions de stationnarité faible et forte coïncident.

Dans l'analyse des séries chronologiques, on utilisera une catégorie spécifique de processus stationnaires les bruits blancs.

**Définition 2.1.3.** On appelle bruit blanc un processus  $(\varepsilon_t)_{t \in \mathcal{T}}$  tel que :

$$\mathbb{E}(\varepsilon_t) = 0 \quad (\text{centré});$$

$$\text{Var}(\varepsilon_t) = \sigma^2 \quad (\text{homoscédasticité});$$

$$\text{Cov}(\varepsilon_s, \varepsilon_t) = 0 \quad \text{si } s \neq t \quad (\text{non corrélé}).$$

Si, en plus,  $\varepsilon_t$  suit une loi normale, on parle de bruit blanc gaussien, noté :

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2), \text{ i.i.d.}$$

**Exemple 2.1.1.** On considère des variables aléatoires indépendantes et identiquement distribuées (i.i.d.)  $(\varepsilon_n)_{n=1, \dots, N}$  suivant une loi normale standard  $\mathcal{N}(0, 1)$ . Cette suite constitue un bruit blanc gaussien. Par définition, une telle série est stationnaire, car les moments d'ordre 1 et 2 sont constants et indépendants du temps. La réalisation simulée de ce processus est illustrée dans la Figure 2.1.

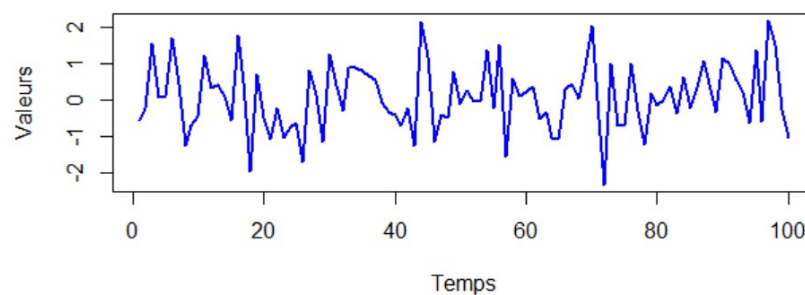


FIGURE 2.1 – Trajectoire d'un bruit blanc

Nous introduisons désormais le théorème de Wold, fondamental pour la représentation et la décomposition des séries stationnaires au sens faible.

**Théorème 2.1.2** ( Théorème de Wold). *Tout processus  $(X_t)_{t \in \mathbb{T}}$  faiblement stationnaire peut être représenté sous la forme :*

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

où :

- $j \in \mathbb{N}$  représente un indice de retard (ou décalage temporel) ;
- les coefficients  $\psi_0, \psi_1, \psi_2, \dots \in \mathbb{R}$  vérifient  $\psi_0 = 1$  et  $\sum_{j=0}^{+\infty} \psi_j^2 < +\infty$  ;
- $(\varepsilon_t)_{t \in \mathbb{T}}$  est un bruit blanc : suite de variables aléatoires centrées, non corrélées, et de variance constante ;
- $k_t$  est une composante déterministe, non aléatoire, qui n'est pas influencée par les termes aléatoires  $\varepsilon_{t-j}$ .

**Remarque 2.1.3.** *Selon le théorème de Wold, tout processus stationnaire d'ordre 2 peut être représenté comme une combinaison linéaire (ou somme infinie) de valeurs passées d'un bruit blanc.*

Désormais, il est important d'examiner les fonctions ACF et PACF, qui sont des outils essentiels pour comprendre la dépendance entre les valeurs d'une série stationnaire.

### 2.1.1 Analyse de la Fonctions d'autocovariance

**Définition 2.1.4.** *La fonction d'autocovariance d'un processus  $(X_t)_{t \in \mathcal{T}}$ , notée  $\gamma(k)$  ou  $\gamma_k$ , est définie, pour tout  $k \in \mathbb{Z}$ , par :*

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}) = \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)]$$

*lorsque  $\mathbb{E}(X_t) = \mu$  existe et est constante (cas stationnaire).*

**Proposition 2.1.4.** *La fonction d'autocovariance d'une série faiblement stationnaire vérifie les propriétés suivantes :*

- i) *La parité :  $\gamma(-k) = \gamma(k)$ . La fonction est paire.*
- ii) *La positivité : pour tout  $n \in \mathbb{N}$ , tout vecteur  $(a_i) \in \mathbb{R}^n$ , et tous instants  $(t_i) \in \mathcal{T}^n$ , on a :*

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(t_i - t_j) \geq 0$$

- iii) *La variance :  $\gamma(0) = \text{Var}(X_t)$ .*
- iv) *La bornitude :  $\forall k \in \mathbb{Z}, |\gamma(k)| \leq \gamma(0)$ .*

#### Démonstration

##### i) La parité

La fonction d'autocovariance  $\gamma(h)$  est définie comme la covariance entre  $X_t$  et  $X_{t+k}$ , c'est-à-dire :

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

On utilise la propriété de la covariance, qui est symétrique, c'est-à-dire que :

$$\text{Cov}(X_t, X_{t+k}) = \text{Cov}(X_{t-k}, X_t).$$

On remplace donc  $X_t$  par  $X_{t-k}$  dans l'expression :

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

Avec

$$\text{Cov}(X_t, X_{t-k}) = \text{Cov}(X_{t-k}, X_t).$$

Donc

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

D'où :

$$\gamma(h) = \gamma(-k).$$

### ii) La positivité

Soit  $\{X_t\}$  un processus aléatoire de moyenne nulle. Pour toute combinaison linéaire de la forme  $\sum_{i=1}^n a_i X_{t_i}$ , avec  $a_i \in \mathbb{R}$ , la variance s'écrit :

$$\text{Var} \left( \sum_{i=1}^n a_i X_{t_i} \right) = \text{Cov} \left( \sum_{i=1}^n a_i X_{t_i}, \sum_{j=1}^n a_j X_{t_j} \right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(t_i - t_j) > 0,$$

où  $\gamma(\cdot)$  désigne la fonction d'autocovariance du processus.

**Remarque 2.1.5.** *Comme la variance est toujours positive ou nulle pour toute combinaison linéaire non triviale, cette quantité est strictement positive si les  $a_i$  ne sont pas tous nuls et si le processus n'est pas dégénéré. Cela illustre que la matrice de covariance est définie positive, propriété essentielle des processus stationnaires.*

### iii) La variance

Soit  $\{X_t\}_{t \in \mathcal{T}}$  un processus stochastique faiblement stationnaire, de moyenne constante  $\mu = \mathbb{E}[X_t]$ . La fonction d'autocovariance est définie par :

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}) = \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)].$$

En particulier, pour  $k = 0$ , on obtient :

$$\gamma(0) = \text{Cov}(X_t, X_t) = \mathbb{E}[(X_t - \mu)^2] = \text{Var}(X_t).$$

Ainsi, la covariance au décalage nul correspond à la variance du processus :

$$\boxed{\gamma(0) = \text{Var}(X_t)}.$$

**Remarque 2.1.6.** *Cette propriété est fondamentale, notamment pour normaliser la fonction d'autocorrélation :*

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

*iv) La bornitude*

Par définition :

$$\gamma(k) = \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)] \quad \text{et} \quad \gamma(0) = \mathbb{E}[(X_t - \mu)^2].$$

Or, pour tous réels  $a, b$ , on a l'inégalité classique :

$$|ab| \leq \frac{a^2 + b^2}{2}.$$

En appliquant cette inégalité à  $a = X_t - \mu$  et  $b = X_{t+k} - \mu$ , on obtient :

$$|(X_t - \mu)(X_{t+k} - \mu)| \leq \frac{(X_t - \mu)^2 + (X_{t+k} - \mu)^2}{2}.$$

En prenant l'espérance des deux membres :

$$|\gamma(k)| \leq \frac{1}{2} [\mathbb{E}[(X_t - \mu)^2] + \mathbb{E}[(X_{t+k} - \mu)^2]].$$

Comme le processus est stationnaire, on a :

$$\mathbb{E}[(X_t - \mu)^2] = \mathbb{E}[(X_{t+k} - \mu)^2] = \gamma(0).$$

Donc :

$$|\gamma(k)| \leq \frac{1}{2}(\gamma(0) + \gamma(0)) = \gamma(0).$$

### 2.1.2 Analyse de la Fonction d'Autocorrélation

**Définition 2.1.5.** La fonction appelée fonction d'autocorrélation, notée  $\rho(k)$  ou  $\rho_k$ , est définie pour tout  $k \in \mathbb{Z}$  par :

$$\rho(k) = \rho_k = \frac{\gamma(k)}{\gamma(0)} \quad \text{avec} \quad \gamma(0) \neq 0.$$

**Proposition 2.1.7.** La fonction d'autocorrélation d'une série faiblement stationnaire vérifie les propriétés suivantes, pour tout  $k \in \mathbb{Z}$  :

- i) La symétrie :  $\rho(-k) = \rho(k)$ .
- ii) La normalisation :  $\rho(0) = 1$ .
- iii) La bornitude :  $|\rho(k)| \leq 1$ .

**Démonstration**

*i) La parité*

L'autocorrélation est définie par :

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \quad \text{avec } \gamma(0) \neq 0$$

comme on a

$$\gamma(-k) = \gamma(k),$$

donc

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

*ii) La normalisation*

L'autocorrélation en  $k = 0$ , on obtient :

$$\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1 \quad \text{avec } \gamma(0) \neq 0.$$

*iii) la bornitude*

Comme  $\gamma(0)$  représente la variance (strictement positive) du processus, on a, pour tout décalage  $k \in \mathbb{Z}$  :

$$|\rho(k)| = \left| \frac{\gamma(k)}{\gamma(0)} \right| \leq 1 \quad \text{avec } \gamma(0) \neq 0.$$

Ainsi, le coefficient de corrélation  $\rho(k)$  est toujours compris entre  $-1$  et  $1$ .

### 2.1.3 Analyse de la Fonction d'Autocorrélation partielle

**Définition 2.1.6.** La fonction appelée fonction d'autocorrélation partielle, notée  $\rho_p(k)$ , est définie pour tout  $k \in \mathbb{Z}$ , lorsque l'espérance  $\mathbb{E}(X_t) = \mu$  est constante.

$$\rho_p(0) = 1, \quad \text{et } \forall k \geq 1, \quad \rho_p(k) = \text{corr}(X_t, X_{t+k} \mid X_{t+1}, \dots, X_{t+k-1}).$$

**Proposition 2.1.8.** La fonction d'autocorrélation partielle d'une série faiblement stationnaire vérifie les propriétés suivantes :

$$\rho_p(1) = \rho(1) \quad \text{et} \quad \rho_p(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}.$$

**Démonstration**

Pour  $k = 1$  :

$$\rho_p(1) = \text{corr}(X_t, X_{t+1}).$$

Puisqu'il n'y a pas de variable intermédiaire entre  $X_t$  et  $X_{t+1}$ , on a directement :

$$\rho_p(1) = \rho(1).$$

**Cas**  $k = 2$  :

L'autocorrélation partielle d'ordre 2 est définie par :

$$\rho_p(2) = \text{corr}(X_t, X_{t+2} \mid X_{t+1})$$

On utilise la formule de la corrélation conditionnelle entre deux variables aléatoires  $A$  et  $B$  conditionnellement à une troisième  $C$  :

$$\text{corr}(A, B \mid C) = \frac{\text{corr}(A, B) - \text{corr}(A, C) \cdot \text{corr}(B, C)}{\sqrt{(1 - \text{corr}(A, C)^2)(1 - \text{corr}(B, C)^2)}}$$

Dans notre cas :

$$A = X_t, \quad B = X_{t+2}, \quad C = X_{t+1}$$

Par stationnarité, on a :

$$\text{corr}(X_t, X_{t+2}) = \rho(2), \quad \text{corr}(X_t, X_{t+1}) = \rho(1), \quad \text{corr}(X_{t+2}, X_{t+1}) = \rho(1)$$

Donc :

$$\rho_p(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}.$$

Quand on étudie une série temporelle stationnaire (comme une température enregistrée chaque jour, ou une production mensuelle), on s'intéresse souvent à la manière dont ses valeurs évoluent dans le temps. C'est ce qu'on appelle l'analyse dans le domaine temporel (avec l'autocorrélation, par exemple).

Mais il est aussi possible de voir la série d'un autre point de vue : au lieu de regarder comment les valeurs sont liées dans le temps, on cherche à savoir quelles sortes de "rythmes" ou de "cycles" sont présents dans la série. C'est ce qu'on appelle l'analyse fréquentielle.

La densité spectrale est un outil qui permet cela. Elle nous dit quelles fréquences (ou quels rythmes) sont dominants dans la série fréquentielle de la série.

### 2.1.4 Densité spectrale

La densité spectrale permet de caractériser différents types de processus stationnaires.

**Définition 2.1.7.** Soit  $(X_t)_{t \in \mathcal{T}}$  un processus stationnaire.

On appelle densité spectrale la fonction  $g$ , définie sur l'intervalle  $[-\pi, \pi]$ , par :

$$g(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) e^{-i\lambda k},$$

où :

- $\gamma(k)$  est la fonction d'autocovariance, et  $k$  représente le décalage,
- $\lambda$  est une fréquence appartenant à l'intervalle  $[-\pi, \pi]$ ,
- $e^{-i\lambda k}$  est un terme complexe provenant de la transformée de Fourier,
- La fonction  $g$  est paire et  $2\pi$ -périodique. Elle est continue si :

$$\sum_{k \in \mathbb{Z}} |\gamma(k)| < \infty.$$

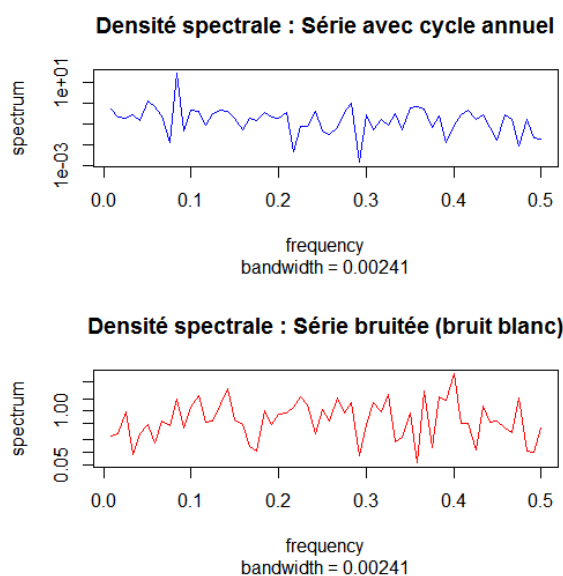


FIGURE 2.2 – Densité spectrale

**Remarque 2.1.9.** La densité spectrale est simplement la transformation mathématique (de Fourier) de la fonction d'autocorrélation. Cela veut dire qu'elle contient la même information, mais exprimée autrement : au lieu de voir comment les valeurs sont liées dans le temps, on voit quelles fréquences expliquent le comportement de la série.

- Si une série a un cycle annuel, la densité spectrale montrera un pic à la fréquence correspondante.
- Si les données sont très irrégulières et bruitées, la densité sera plus étalée.

La densité spectrale et la fonction d'autocovariance sont liées par une transformation de Fourier, ce qui permet de passer de l'une à l'autre de façon réversible.

**Proposition 2.1.10.** *Soit  $(X_t)_{t \in \mathcal{T}}$  un processus stationnaire. Alors, sa fonction d'autocovariance  $\gamma(k)$  peut être obtenue par la formule :*

$$\gamma(k) = \int_{-\pi}^{\pi} g(\lambda) e^{i\lambda k} d\lambda.$$

**Remarque 2.1.11.** *La densité spectrale permet de repérer les rythmes qui se répètent dans une série, même s'ils sont masqués par le bruit. Elle aide à décomposer la série en composantes simples et à choisir un modèle adapté, notamment en détectant la présence de saisonnalité ou de cycles.*

Bien que la stationnarité soit une hypothèse centrale, de nombreuses séries observées dans la réalité présentent des caractéristiques de non-stationnarité qu'il convient d'identifier et de traiter.

## 2.2 Non stationnarité

La non-stationnarité peut être de trois natures. Elle peut être de nature déterministe, c'est-à-dire due à l'existence d'une tendance liée au temps ou à une variance non constante. Elle peut aussi être de nature stochastique, comme c'est le cas par exemple pour les marchés aléatoires.

### 2.2.1 Non-stationnarité par Tendance linéaire

Une tendance déterministe signifie que la série évolue de manière prévisible au fil du temps selon une fonction déterminée.

$$X_t = a + \beta t + \varepsilon_t.$$

Avec

- $a$  et  $\beta$  sont des constantes.
- $\beta t$  représente une tendance linéaire.
- $\varepsilon_t$  est un bruit blanc .

### 2.2.2 Non-stationnarité par Tendance Stochastique

Dans ce cas, la série suit un processus de marche aléatoire :

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

- Chaque valeur dépend de la précédente et d'un choc aléatoire  $\varepsilon_t$ .
- Ce processus ne revient jamais à une moyenne fixe.

### 2.2.3 Non-stationnarité par Variance Non Constante (Hétéroscédasticité)

La volatilité de la série change au cours du temps.

$$X_t = \varepsilon_t \sqrt{h_t}, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1).$$

Avec

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2.$$

- Les fluctuations sont irrégulières.
- On observe des périodes de forte variabilité suivies de périodes calmes.

Dans la figure 2.3, nous observons les trois types de non-stationnarité.

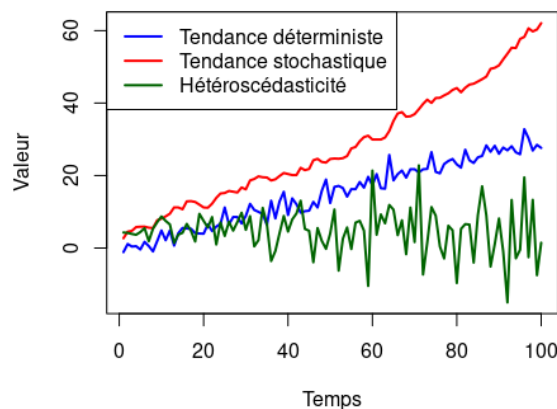


FIGURE 2.3 – Exemples de séries non stationnaires

La détection visuelle ou descriptive d'une stationnarité ou d'une non-stationnarité constitue une première étape, qui doit être complétée par une évaluation rigoureuse à l'aide de tests statistiques appropriés.

## 2.3 Les tests de Vérification de la Stationnarité

Pour déterminer si une série temporelle est stationnaire ou non, un ensemble de tests statistiques a été développé pour vérifier la présence d'une racine unitaire dans la série, ce qui constitue un indicateur principal de non-stationnarité. Parmi ces tests, on peut citer le test de Dickey-Fuller, qui est utilisé pour tester l'hypothèse de la présence d'une racine unitaire, et le test KPSS, qui suppose la stationnarité sous l'hypothèse nulle, ce qui fait de ces outils des instruments essentiels pour évaluer la stabilité des données.

De plus, le test de Phillips-Perron est particulièrement utile dans ce contexte, car il tient compte de la présence éventuelle d'autocorrélation dans les erreurs et propose une méthode fiable pour tester la stationnarité d'une série temporelle.

### 2.3.1 Le test de Dickey-Fuller (DF)

Le test de Dickey-Fuller permet de déterminer si une série temporelle contient une racine unitaire, ce qui signifierait qu'elle est non stationnaire. En effet, l'absence de stationnarité indique que la série peut suivre une tendance persistante à long terme.

$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad (2.1)$$

où le terme d'erreur  $\varepsilon_t$  est supposé non autocorrélé.

où  $0 \leq \varphi \leq 1$ , et le terme d'erreur  $\varepsilon_t$  vérifie :

$$\mathbb{E}(\varepsilon_t) = 0, \quad \mathbb{E}(\varepsilon_t^2) = \sigma^2, \quad \text{Cov}(\varepsilon_t, \varepsilon_j) = 0 \quad \text{si} \quad t \neq j.$$

Les hypothèses du test sont :

$$\begin{cases} H_0 : \varphi = 1 & (\text{la série est non stationnaire}) \\ H_1 : \varphi < 1 & (\text{la série est stationnaire}) \end{cases}$$

On réécrit le modèle (2.1) sous la forme des différences premières :

$$\Delta X_t = \lambda X_{t-1} + \varepsilon_t \quad \text{avec} \quad \lambda = \varphi - 1.$$

Les hypothèses du test deviennent alors :

$$\begin{cases} H_0 : \lambda = 0 & (\text{la série est non stationnaire}) \\ H_1 : \lambda < 0 & (\text{la série est stationnaire}) \end{cases}$$

#### Étapes du test

- Calculer la différence première de la série :  $\Delta X_t = X_t - X_{t-1}$ .
- Estimer  $\lambda$  par la méthode des moindres carrés ordinaires.
- Calculer la statistique de test :

$$\tau_{\text{emp}} = \frac{\hat{\lambda}}{\text{erreur standard de } \hat{\lambda}}.$$

L'erreur standard de  $\hat{\lambda}$  est donnée par :

$$\sqrt{\text{Var}(\hat{\lambda})} = \sqrt{\frac{\sigma^2}{\sum_{t=2}^n X_{t-1}^2}},$$

avec :

$$\sigma^2 = \frac{SCR}{n-2} = \frac{\sum_{t=2}^n (\Delta X_t - \widehat{\Delta X}_t)^2}{n-2}.$$

avec :

- $n$  : Nombre d'observations dans la série.
- $2$  : Nombre de paramètres estimés dans le modèle (par exemple :  $\lambda$  et éventuellement  $\alpha$ , la constante).
- $n-2$  : Degrés de liberté utilisés pour estimer la variance des résidus  $\sigma^2$ , c'est-à-dire le nombre d'observations moins le nombre de paramètres estimés.

On compare  $\tau_{\text{emp}}$  à la valeur critique  $\tau_{\text{crit}}$  de la table de Dickey-Fuller (voir tableau 2.1). Si  $\tau_{\text{emp}} < \tau_{\text{crit}}$ , on rejette  $H_0$  et on conclut que la série est stationnaire.

Plus précisément, Dickey et Fuller proposent trois versions du test, correspondant à trois modèles différents selon la forme de la tendance présente dans la série.

### 1. Modèle sans constante ni tendance

$$X_t = \varphi X_{t-1} + \varepsilon_t \quad \text{ou bien} \quad \Delta X_t = \lambda X_{t-1} + \varepsilon_t.$$

On utilise ce modèle lorsque la série chronologique oscille autour de zéro sans tendance claire, il n'y a pas de moyenne non nulle à prendre en compte.

### 2. Modèle avec constante mais sans tendance

$$X_t = \alpha + \varphi X_{t-1} + \varepsilon_t, \quad \text{ou bien} \quad \Delta X_t = \lambda X_{t-1} + \alpha + \varepsilon_t.$$

On utilise ce modèle si la série temporelle fluctue autour d'une moyenne différente de zéro mais sans tendance croissante ou décroissante. la constante  $\alpha$  représente cette moyenne à long terme.

### 3. Modèle avec constante et tendance déterministe

$$X_t = \beta t + \alpha + \varphi X_{t-1} + \varepsilon_t, \quad \text{ou bien} \quad \Delta X_t = \beta t + \alpha + \lambda X_{t-1} + \varepsilon_t.$$

On utilise ce modèle si la série montre une tendance croissante ou décroissante au fil du temps, la constante  $\alpha$  ajuste la moyenne, tandis que  $\beta t$  capture la tendance.

Les tests sont réalisés dans un ordre séquentiel, le premier modèle analysé étant :

$$\Delta X_t = \beta t + \alpha + \lambda X_{t-1} + \varepsilon_t.$$

Pour tester la présence d'une tendance linéaire, nous posons les hypothèses suivantes :

$$\begin{cases} H_0 : \beta t = 0 \\ H_1 : \beta t \neq 0 \end{cases}$$

La statistique de test de Student est donnée par :

$$\tau_\beta = \frac{\hat{\beta}_t}{\hat{\sigma}_{\beta_t}},$$

où  $\hat{\beta}_t$  est l'estimation de  $\beta_t$  et  $\hat{\sigma}_{\beta_t}$  son erreur standard.

- Si  $\tau_\beta$  est supérieur à la valeur critique de Student, alors  $H_0$  est rejeté et la tendance est significative, donc, nous procédons au test de racine unitaire.
- Sinon, on ne rejette pas  $H_0$  et on retire la tendance du modèle.

On considère le modèle :

$$\Delta X_t = \lambda X_{t-1} + \alpha + \varepsilon_t.$$

Les hypothèses du test sont :

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

La statistique de test est donnée par :

$$\tau_\alpha = \frac{\hat{\alpha}}{\hat{\sigma}_\alpha},$$

où  $\hat{\alpha}$  est l'estimation de  $\alpha$  et  $\hat{\sigma}_\alpha$  son erreur standard.

- Si  $\tau_\alpha$  est supérieur à la valeur critique,  $H_0$  est rejeté et la constante est significative. Donc, nous testons alors l'hypothèse de la présence d'une racine unitaire.
- Sinon, on ne rejette pas  $H_0$  et on enlève la constante du modèle, le test de racine unitaire est réalisé en utilisant le modèle :

$$\Delta X_t = \lambda X_{t-1} + \varepsilon_t.$$

Pour chacun des trois tests, les valeurs critiques du test de racine unitaire sont fournies en raison de la nature de ce test.

N	$\beta t = 0, \alpha = 0$				$\alpha \neq 0, \beta t = 0$				$\alpha \neq 0, \beta t \neq 0$			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-2.661	-2.273	-1.955	-1.609	-3.724	-3.318	-2.986	-2.633	-4.375	-3.943	-3.559	-3.218
50	-2.612	-2.246	-1.947	-1.612	-3.568	-3.213	-2.921	-2.599	-4.152	-3.791	-3.495	-3.181
100	-2.588	-2.234	-1.944	-1.614	-3.487	-3.146	-2.873	-2.572	-4.092	-3.722	-3.459	-3.142
250	-2.573	-2.227	-1.942	-1.616	-3.471	-3.136	-2.873	-2.570	-3.977	-3.760	-3.419	-3.118
500	-2.570	-2.224	-1.942	-1.616	-3.443	-3.127	-2.867	-2.570	-3.963	-3.660	-3.413	-3.112
>500	-2.567	-2.223	-1.941	-1.616	-3.443	-3.120	-2.863	-2.568	-3.963	-3.660	-3.413	-3.128

TABLE 2.1 – Valeurs critiques sous différents modèles ( $\alpha, \beta$ )

### 2.3.2 Le test de Dickey-Fuller augmentée (ADF)

Il peut y avoir auto-corrélation des résidus de l'équation estimée, les tests de Dickey-Fuller augmentés (Augmented Dickey Fuller) consistent à introduire des variables explicatives supplémentaires, les retards de la variable  $X_{t-1}$ , c'est-à-dire les variables  $\Delta X_{t-2}, \dots, \Delta X_{t-p}$ .

L'hypothèse sous-jacente est que l'auto-corrélation des résidus a pour cause l'absence de variables retardées différenciées dans l'équation.

Considérons la série chronologique :

$$X_t = \varphi X_{t-1} + \sum_{j=1}^m \phi_j \Delta X_{t-j} + \varepsilon_t, \quad (2.2)$$

ou sous forme équivalente :

$$\Delta X_t = \lambda X_{t-1} + \sum_{j=1}^m \phi_j \Delta X_{t-j} + \varepsilon_t. \quad (2.3)$$

où :

- Les termes  $\sum_{j=1}^m \phi_j \Delta X_{t-j}$  permettent d'éliminer l'autocorrélation dans les erreurs.
- $m$  indique le nombre de valeurs passées de la série  $X_t$  qui sont prises en compte pour expliquer son évolution actuelle.
- $\phi_j$  représente les coefficients d'ajustement des retards dans un modèle autorégressif.

On distingue trois variantes du test ADF en fonction de la structure de la série chronologique :

### 1. Modèle sans constante ni tendance

$$\Delta X_t = \lambda X_{t-1} + \sum_{j=1}^m \phi_j \Delta X_{t-j} + \varepsilon_t.$$

### 2. Modèle avec constante mais sans tendance

$$\Delta X_t = \alpha + \lambda X_{t-1} + \sum_{j=1}^m \phi_j \Delta X_{t-j} + \varepsilon_t.$$

### 3. Modèle avec constante et tendance

$$\Delta X_t = \beta t + \alpha + \lambda X_{t-1} + \sum_{j=1}^m \phi_j \Delta X_{t-j} + \varepsilon_t.$$

Ces modèles permettent de tester la présence d'une racine unitaire dans une série temporelle, c'est-à-dire de déterminer si la série est stationnaire ou non stationnaire.

$$\begin{cases} H_0 : \lambda = 0 & \text{(La série est non stationnaire)} \\ H_1 : \lambda < 0 & \text{(La série est stationnaire)} \end{cases}$$

La statistique du test de Dickey-Fuller Augmenté (ADF) est définie comme suit :

$$\tau_{\text{ADF}} = \frac{\hat{\lambda}}{\text{erreur standard de } \hat{\lambda}}.$$

où :

- $\hat{\lambda}$  représente l'estimateur du paramètre  $\lambda$  obtenu à l'aide de la méthode des moindres carrés,
- L'erreur standard de  $\hat{\lambda}$  est notée  $\sqrt{\widehat{\text{Var}}(\hat{\lambda})}$ .

La variance estimée de  $\hat{\lambda}$  est généralement calculée à l'aide de la formule suivante :

$$\widehat{\text{Var}}(\hat{\lambda}) = \frac{\hat{\sigma}^2}{\sum_{t=m+2}^n X_{t-1}^2},$$

où :

- $\hat{\sigma}^2$  : estimation de la variance des erreurs du modèle,
- $n$  : nombre total d'observations dans la série,
- $m$  : nombre de retards  $\Delta X_{t-j}$  introduits dans le modèle.

De plus, la variance des résidus  $\hat{\sigma}^2$  est donnée par :

$$\hat{\sigma}^2 = \frac{1}{n-r} \sum_{t=m+2}^n (\Delta X_t - \widehat{\Delta X}_t)^2,$$

où :

- $r$  : nombre total de paramètres estimés dans le modèle (incluant éventuellement une constante, une tendance et les coefficients des retards),
- $\widehat{\Delta X}_t$  : valeur ajustée de  $\Delta X_t$  à partir du modèle estimé.

Si la statistique observée vérifie :  $\tau_{ADF} < \tau_{critique}$ , on rejette l'hypothèse nulle  $H_0 : \lambda = 0$ , et l'on conclut que la série est stationnaire.

### 2.3.3 Le test de Phillips-Perron (PP)

Le test de Phillips-Perron (PP) est une amélioration du test de Dickey-Fuller, développé pour détecter la présence d'une racine unitaire dans une série temporelle, tout en corrigeant l'autocorrélation et l'hétéroscédasticité possibles dans les résidus du modèle.

Contrairement au test ADF qui ajoute des retards différenciés dans l'équation, le test PP ajuste la statistique de test à l'aide d'une méthode non paramétrique, sans modifier la structure du modèle.

Le test de Phillips-Perron repose sur l'estimation du modèle suivant :

$$\Delta X_t = \alpha + \beta t + \lambda X_{t-1} + \varepsilon_t$$

Les hypothèses sont :

$$\begin{cases} H_0 : \lambda = 0 & (\text{série non stationnaire}) \\ H_1 : \lambda < 0 & (\text{série stationnaire}) \end{cases}$$

On ne peut pas conclure à la stationnarité uniquement à partir de la valeur de  $\lambda$  ; il est nécessaire de calculer la statistique  $\tau_{PP}$  et de la comparer à une valeur critique pour décider du rejet de l'hypothèse nulle.

La variance à court terme est estimée par :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2.$$

ou  $n$  : Nombre d'observations (longueur de la série)

La variance à long terme :

$$S^2 = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2 + 2 \sum_{j=1}^{\bar{b}} \left(1 - \frac{j}{\bar{b}+1}\right) \cdot \left(\frac{1}{n} \sum_{t=j+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}\right)$$

où le nombre de retards  $\bar{b}$  est défini par la règle :

$$\bar{b} \approx 4 \left(\frac{n}{100}\right)^{2/9}$$

La statistique du test de Phillips-Perron s'écrit :

$$\tau_{PP} = \sqrt{B} \cdot \frac{(\hat{\lambda} - 1)}{\hat{\sigma}\hat{\lambda}} + n(\sqrt{B} - 1) \cdot \frac{1}{\hat{\sigma}\hat{\lambda}}, \quad \text{où } B = \frac{\hat{\sigma}^2}{S^2}.$$

On compare la statistique  $\tau_{PP}$  aux valeurs critiques 2.1 pour décider du rejet de l'hypothèse nulle  $H_0$ .

Si la statistique de test  $\tau_{pp}$  est inférieure à la valeur critique, on rejette l'hypothèse nulle et on conclut que la série est stationnaire.

### 2.3.4 Le test de KPSS

Kwiatkowski, Phillips, Schmidt et Shin ont créé ce test en 1992 pour vérifier si une série temporelle est stationnaire appelé KPSS, ce test utilise le multiplicateur de Lagrange et considère que la série est stationnaire au départ. En réalité, il cherche à savoir si la partie non stationnaire d'une série a une variance de zéro.

Les équations du modèle sont les suivantes :

$$X_t = w_t + \beta t + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2).$$

$$w_t = w_{t-1} + v_t, \quad v_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_v^2).$$

où :

- $w_t$  est un composant aléatoire évoluant dans le temps.
- $\beta t$  représente une tendance linéaire.
- $\varepsilon_t$  est un bruit blanc indépendant et identiquement distribué.
- $v_t$  est un bruit blanc avec une moyenne nulle et une variance  $\sigma_v^2$ .

Le test vérifie l'hypothèse nulle suivante :

$$\begin{cases} H_0 : \sigma_v^2 = 0 \text{ ( la série est stationnaire),} \\ H_1 : \sigma_v^2 > 0 \text{ (la série est non stationnaire).} \end{cases}$$

Il ne suffit pas d'examiner uniquement la variance ; la statistique complète du test KPSS, qui intègre cette variance dans son calcul, doit être calculée. La décision finale repose sur la comparaison de cette statistique à la valeur critique correspondante.

La statistique du test KPSS est donnée par :

$$\tau_{KPSS} = \frac{1}{\hat{\sigma}_\varepsilon^2} \sum_{t=1}^n \left( \sum_{i=1}^n \hat{\varepsilon}_i \right)^2.$$

La statistique  $\tau_{KPSS}$  peut aussi être écrite comme :

$$\tau_{KPSS} = \frac{1}{\hat{\sigma}_\varepsilon^2} \sum_{t=1}^n R_t^2 \quad \text{où } R_t = \sum_{i=1}^n \hat{\varepsilon}_i.$$

Autrement dit, c'est la somme des carrés des sommations cumulées des résidus, divisée par la variance des résidus.

Après calcul de la statistique  $\tau_{KPSS}$  on compare avec les valeurs critiques (voir la table 2.2) :

- Si la statistique calculée est plus grande que la valeur critique alors la série n'est pas stationnaire.
- Si la statistique calculée est plus petite que la valeur critique alors la série est stationnaire.

Niveau de signification	Valeur critique
10 %	0,119
5 %	0,146
2,5 %	0,176
1 %	0,216

TABLE 2.2 – Valeurs critiques du test KPSS

En présence de non-stationnarité détectée par les tests, la suite logique consiste à appliquer des techniques de stationnarisation afin de satisfaire les conditions requises pour l'analyse et la modélisation.

## 2.4 Techniques de Stationnarisation

Ces techniques incluent la différenciation, la transformation logarithmique, l'élimination des effets saisonniers et la décomposition de la série en ses composants de base. Chacune de ces méthodes vise à éliminer les tendances ou la variabilité non stationnaire, ce qui rend la série temporelle apte à être utilisée pour la modélisation et la prévision.

### 2.4.1 Différenciation

La différenciation est une technique utilisée pour transformer une série temporelle non stationnaire en une série stationnaire. On applique la méthode des différences premières pour éliminer la racine unitaire, en substituant à chaque observation sa valeur observée à la période précédente :

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

Une fois la série transformée, il est important de vérifier l'hypothèse de stationnarité. Si la méthode des différences premières ne réussit pas à produire une série stationnaire, il faut répéter le processus jusqu'à ce qu'une série stationnaire soit obtenue. Par exemple, pour une série temporelle intégrée d'ordre deux, il faut utiliser les différences secondes :

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

Pour atteindre la stationnarité, plus généralement. La différenciation d'ordre  $p$  d'une série temporelle  $X_t$  est donnée par

$$\Delta^k X_t = \Delta^{k-1} X_t - \Delta^{k-1} X_{t-1}.$$

La principale force de la méthode des différences est sa simplicité. Cependant, elle a un inconvénient important en exprimant la variable analysée en termes de différence, on perd obligatoirement la première observation.

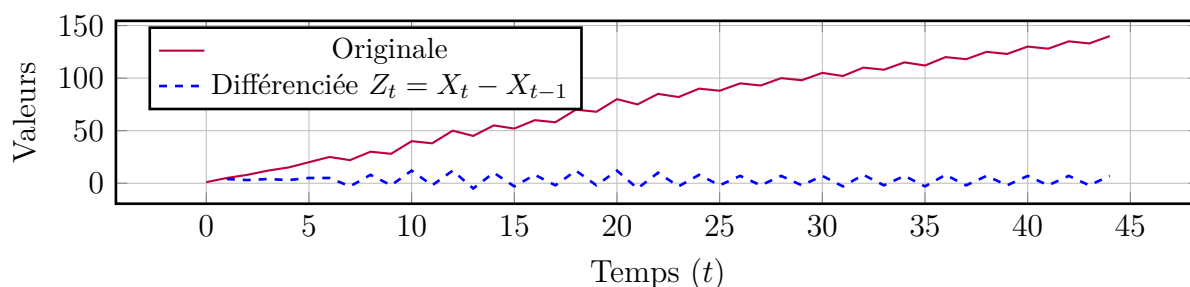


FIGURE 2.4 – Série avec tendance linéaire et sa différenciation première

### 2.4.2 L'inverse

L'inverse est utilisé si les valeurs élevées ont un impact très fort et qu'il faut les réduire brutalement. Il est rarement utilisé, mais peut être utile en présence d'asymétries extrêmes.

$$X'_t = \frac{1}{X_t}.$$

- $X_t$  représente la série temporelle originale.
- $X'_t$  est la série transformée.

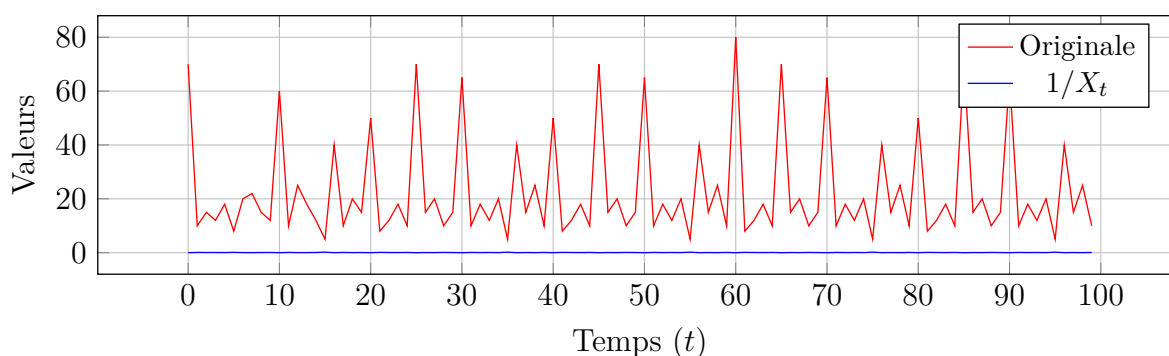


FIGURE 2.5 – Transformation inverse : réduction de l'asymétrie extrême

### 3.4.3 Transformation Logarithmique

Le logarithme est utilisé lorsque la série présente une croissance exponentielle et que la variance augmente avec la moyenne.

$$X'_t = \log(X_t).$$

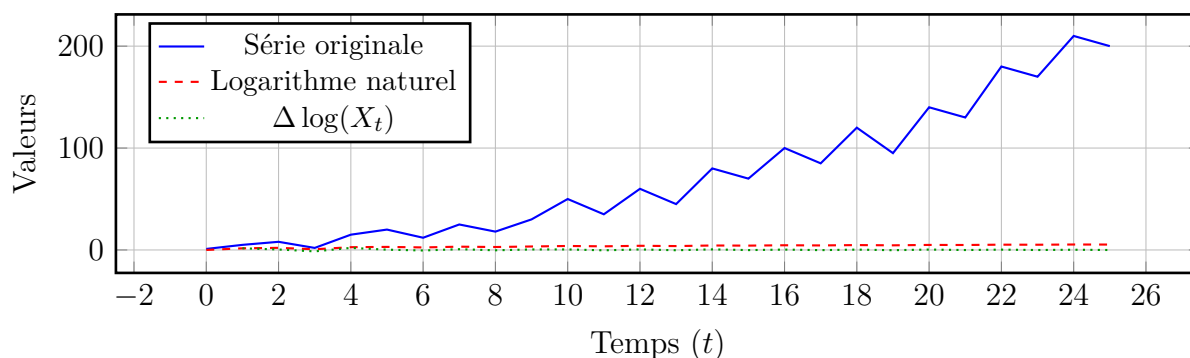


FIGURE 2.6 – Série originale, transformation et différenciation logarithmique

### 2.4.4 La racine

La racine carrée est utilisée lorsque la variance augmente légèrement avec la moyenne, mais que l'effet exponentiel n'est pas dominant. Elle est utile pour des données comptant des occurrences, comme le nombre de ventes ou de client.

$$X'_t = \sqrt{X_t}.$$

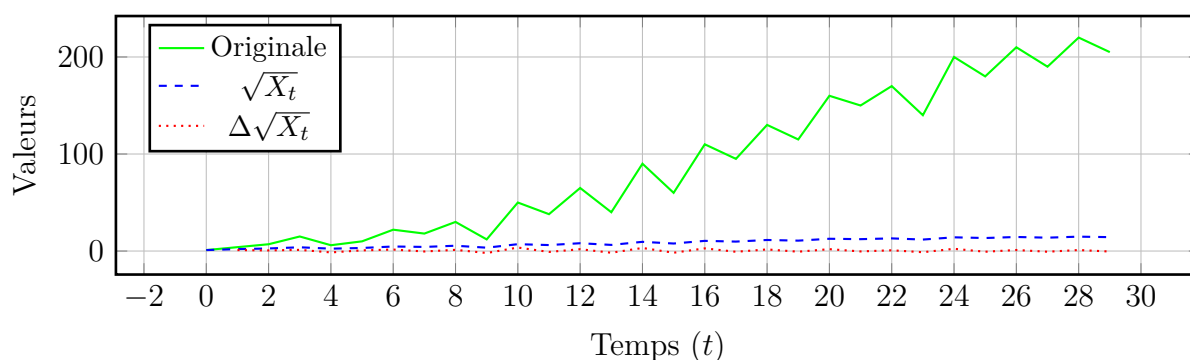


FIGURE 2.7 – Transformation racine carrée et différenciation

**Remarque 2.4.1.** *La différenciation est une méthode utilisée pour supprimer la tendance dans une série temporelle, en stabilisant sa moyenne dans le temps. Cependant, lorsqu'une série contient des valeurs très variables ou des effets de croissance rapide, il est souvent préférable d'appliquer d'abord une transformation, comme la racine carrée ou le logarithme. Ces transformations permettent de réduire l'écart entre les grandes et les petites valeurs, tout en facilitant l'analyse.*

*En particulier, le logarithme est souvent utilisé car il conserve les valeurs strictement positives tout en réduisant les amplitudes, ce qui évite d'introduire des valeurs négatives ou incohérentes lors de la différenciation. Cela rend la série plus adaptée aux méthodes statistiques basées sur la stationnarité.*

### 2.4.5 Décomposition Saisonnière

Les méthodes de décomposition reposent sur l'identification et l'évaluation des différentes composantes (tendance, élément saisonnier, etc.) d'une série temporelle. Ces techniques

sont particulièrement pertinentes lorsqu'il s'agit de retirer une ou plusieurs composantes d'une série temporelle. Par exemple, dans l'analyse des séries d'indices économiques, lorsqu'on cherche à supprimer les fluctuations saisonnières. Toutefois, elles peuvent aussi s'avérer utiles pour décrire les séries et, dans certaines situations, pour effectuer des calculs de prévisions.

• **Décomposition par moyennes mobiles :**

La méthode de moyenne mobile est utilisée dans le lissage des séries chronologiques . Elle permet d' éliminer les fluctuations aléatoires et d'isoler la tendance de long terme, et elle peut réduire la variabilité de la série.

La moyenne mobile centrée d'ordre  $p_s$  à la date  $t$  le nombre noté  $MM_{p_s}(t)$  ou  $T_t$  est défini par :

— Si  $p_s$  est impair, soit  $p_s = 2\ell + 1$  avec  $\ell \in \mathbb{N}$  :

$$MM_{p_s}(t) = \frac{1}{p_s} \sum_{i=-\ell}^{\ell} X_{t+i}.$$

— Si  $p_s$  est pair, soit  $p_s = 2\ell$  :

$$MM_{p_s}(t) = \frac{1}{p_s} \left( \frac{X_{t-\ell} + X_{t+\ell}}{2} + \sum_{i=-\ell+1}^{\ell-1} X_{t+i} \right).$$

Les coefficients saisonniers  $S_j$  sont calculés comme suit :

$$S_j = \frac{1}{n} \sum_{i=1}^n (X_{ij} - T_{ij}), \quad (\text{le modèle additif}).$$

$$S_j = \frac{1}{n} \sum_{i=1}^n \left( \frac{X_{ij}}{T_{ij}} \right), \quad (\text{le modèle multiplicatif}).$$

où,

- $T_{ij}$  : Composante de tendance estimée à cette même date.
- $i, j$  : indices pour accéder aux valeurs autour de  $t$  .
- $n$  : nombre d'observations .

• **Décomposition par le lissage exponentiel**

Le lissage des séries chronologique permet de stabiliser les données en réduisant le bruit et en détectant les tendances actuelles, tandis que la décomposition sépare la série en tendance, saisonnalité et résidus afin d'éliminer les effets non stationnaires, l'objectif étant d'obtenir une série aussi stationnaire que possible, sans tendance ni saisonnalité.

a) **Lissage exponentiel simple**

Une méthode utilisée pour lisser une série temporelle en réduisant le bruit tout en donnant plus de poids aux observations récentes.

$$\tilde{X}_t = \beta X_t + (1 - \beta)\tilde{X}_{t-1}.$$

Avec

- $\tilde{X}_t$  : la valeur lissée à l'instant  $t$ .
- $X_t$  : la valeur d'origine à l'instant  $t$ .
- $\beta \in [0, 1]$  : le coefficient de lissage .
  - Si  $\beta$  est proche de 1 alors on accorde plus de poids à la valeur actuelle.
  - Si  $\beta$  est proche de 0 alors on accorde plus de poids aux valeurs passées.

### b) Lissage exponentiel double

Une méthode utilisée pour lisser une série temporelle tout en prenant en compte sa tendance, mais sans considérer la saisonnalité.

$$\tilde{X}_t = \beta X_t + (1 - \beta)(\tilde{X}_{t-1} + T_{t-1}).$$

$$T_t = \nu(\tilde{X}_t - \tilde{X}_{t-1}) + (1 - \nu)T_{t-1}.$$

Avec

- Ce modèle ajoute un composant de tendance ( $T_t$ ) au modèle.
- $T_t$  : représente la tendance (la pente ) à l'instant  $t$ .
- $\nu$  : le coefficient de lissage de la tendance (généralement similaire à  $\beta$ ).

### c) Lissage exponentiel de Holt-Winters

Une méthode permettant de lisser une série temporelle tout en prenant en compte à la fois sa tendance et sa saisonnalité. Ce modèle repose sur trois équations principales .

$$\tilde{X}_t = \beta X_t + (1 - \beta)(\tilde{X}_{t-1} + T_{t-1}).$$

$$T_t = \nu(\tilde{X}_t - \tilde{X}_{t-1}) + (1 - \nu)T_{t-1}.$$

$$S_t = \delta(X_t - \tilde{X}_t) + (1 - \delta)S_{t-p_s}.$$

Avec

- $S_t$  : le composant saisonnier à l'instant  $t$ .
- $p_s$  : la longueur du cycle saisonnier (par exemple,  $p_s = 12$  pour les mois).
- $\delta$  : le coefficient de lissage saisonnier.

**Calcul des données sans saisonnalité et sans tendance :**

$$\varepsilon_t = X_t - T_t - S_t, \quad (\text{modèle additif}).$$

$$\varepsilon_t = \frac{X_t}{S_t \times T_t}, \quad (\text{modèle multiplicatif}).$$

Une fois la stationnarité des séries temporelles assurée grâce aux techniques appropriées, il devient possible de passer à l'étape de modélisation, en s'appuyant sur les propriétés de la série stabilisée pour identifier le modèle statistique le plus adapté à sa dynamique et à la prévision de son évolution future.

## 2.5 Modélisation des séries selon leur stationnarité

Modéliser une série chronologique consiste à choisir une structure adaptée à son comportement dans le temps. Une série stationnaire peut être décrite par des modèles AR, MA ou ARMA. Si elle présente une tendance ou une saisonnalité, on utilise des modèles ARIMA ou SARIMA après transformation. Le but est d'expliquer la dynamique temporelle et de prévoir les valeurs futures.

### 2.5.1 Le modèle ARMA

Les processus ARMA (autoregressive moving average) processus sont des modèles linéaires qui jouent un rôle important dans la modélisation pratique de série stationnaires. Ils offrent l'avantage d'être plus souples à utiliser et de fournir généralement de bonnes approximations des séries réelles avec moins de paramètres que les modèles purs. Le modèle est composé de deux parties une partie autorégressive (AR) et une partie moyenne-mobile (MA).

**Définition 2.5.1.** *Un processus stationnaire  $\{X_t, t \in \mathcal{T}\}$  est dit autorégressif d'ordre  $p$  (ou  $AR(p)$ ) s'il obéit à une équation du type :*

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad \forall t \in \mathcal{T},$$

où  $\varepsilon_t$  est un bruit blanc.

En utilisant l'opérateur de retard  $L$ , défini par  $L^k X_t = X_{t-k}$ , on peut réécrire l'équation comme :

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) X_t = c + \varepsilon_t,$$

où  $c \in \mathbb{R}$  est une constante, et  $\phi = (\phi_1, \dots, \phi_p) \in \mathbb{R}^p$  est le vecteur des coefficients autorégressifs.

**Définition 2.5.2.** Un processus stationnaire  $\{X_t\}$  est dit **inversible** s'il existe une suite réelle  $\{b_k\}_{k \geq 0} \subset \mathbb{R}$  telle que :

$$\sum_{k=0}^{\infty} |b_k| < \infty, \quad \text{et} \quad \varepsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}$$

où  $\varepsilon_t$  représente un bruit blanc.

La condition de convergence absolue garantit l'existence de cette représentation.

Une autre façon de dire qu'un processus est inversible est de dire qu'il possède une représentation AR( $\infty$ ).

**Théorème 2.5.1.** Un processus autorégressif AR( $p$ ) est causal et stationnaire si et seulement si son polynôme  $\phi(L)$  est tel que :

$$\phi(L) \neq 0 \quad \text{pour tout } L \in \mathbb{C} \text{ tel que } |L| \leq 1,$$

avec

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

**Définition 2.5.3.** Un processus stationnaire  $\{X_t, t \in \mathcal{T}\}$  est dit **moyenne mobile d'ordre  $q$**  (ou MA( $q$ )) s'il obéit à une équation du type :

$$X_t = d + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad \forall t \in \mathcal{T},$$

où  $\varepsilon_t$  est un bruit blanc.

En introduisant l'opérateur de retard  $L$ , défini par :

$$L^k \varepsilon_t = \varepsilon_{t-k},$$

on peut réécrire le modèle MA( $q$ ) sous forme opératorielle :

$$X_t = d + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t.$$

avec

- $d \in \mathbb{R}$  : constante,
- $\theta = (\theta_1, \dots, \theta_q) \in \mathbb{R}^q$  : vecteur des coefficients de moyenne mobile,
- $\varepsilon_t \sim i.i.d. \mathcal{N}(0, \sigma^2)$  : bruit blanc, centré et homoscédastique.

**Définition 2.5.4.** Un processus  $\{X_t\}$  est dit **causal** s'il existe une suite réelle  $\{a_k\}_{k \geq 0} \subset \mathbb{R}$  telle que :

$$\sum_{k=0}^{\infty} |a_k| < \infty \quad \text{et} \quad X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$$

où  $\varepsilon_t$  est un bruit blanc centré, non corrélé et de variance constante.

La condition de convergence absolue assure que la série est bien définie.

Parfois, lorsque l'on parle d'un processus causal, on dit que celui-ci a une représentation MA( $\infty$ ).

**Théorème 2.5.2.** *Un processus à moyenne mobile MA( $q$ ) est inversible si et seulement si son polynôme  $\theta(z)$  est tel que :*

$$\theta(L) \neq 0 \quad \text{avec } L \in \mathbb{C} \text{ tel que } |L| \leq 1.$$

**Définition 2.5.5.** *Un processus stationnaire  $\{X_t, t \in \mathcal{T}\}$  est dit **autorégressif et moyenne mobile d'ordre** ( $p, q$ ) (ou ARMA( $p, q$ )) s'il vérifie l'équation suivante :*

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \forall t \in \mathcal{T},$$

où  $\varepsilon_t$  est un bruit blanc.

En utilisant l'opérateur de retard  $L$ , défini par  $L^k X_t = X_{t-k}$ , le modèle peut s'écrire sous forme opératorielle :

$$\Phi(L)X_t = c + \Theta(L)\varepsilon_t,$$

avec :

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p, \quad \Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q.$$

Avec

- $c \in \mathbb{R}$  : constante,
- $\phi = (\phi_1, \dots, \phi_p) \in \mathbb{R}^p$  : coefficients autorégressifs (AR),
- $\theta = (\theta_1, \dots, \theta_q) \in \mathbb{R}^q$  : coefficients de moyenne mobile (MA),
- $\varepsilon_t \sim i.i.d. \mathcal{N}(0, \sigma^2)$  : bruit blanc centré et homoscedastique.

**Remarque 2.5.3.** *Un processus ARMA est stationnaire si toutes les racines du polynôme caractéristique  $\phi(L)$  ont un module strictement supérieur à 1.*

### Cas particuliers du modèle ARMA :

- Si  $p = q = 0$ , alors  $X_t = \varepsilon_t$  et donc un processus ARMA(0,0) est un bruit blanc faible.
- Si  $p = 0$  et  $q > 0$ , les processus ARMA(0, $q$ ) sont appelés des processus MA( $q$ ) (moyenne mobile).
- Si  $p > 0$  et  $q = 0$ , les processus ARMA( $p$ ,0) sont appelés des processus AR( $p$ ) (auto-régressifs).

Les processus suivants ce sont des généralisation des processus ARMA aux cas non stationnaires, avec tendance polynomiale (ARIMA) ou avec une saisonnalité (SARIMA).

## 2.5.2 Le modèle ARIMA

Un modèle ARIMA, qui signifie Auto-Régressif Intégré à Moyenne Mobile, est un modèle utilisé pour représenter une série temporelle non stationnaire qui peut être rendue stationnaire par différenciation d'ordre  $d$ .

**Définition 2.5.6.** Une série temporelle  $\{X_t\}$  suit un processus ARIMA (AutoRegressive Integrated Moving Average) d'ordre  $(p, d, q)$  si elle peut s'écrire sous la forme :

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

où :

- $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  est un polynôme autorégressif (AR) de degré  $p$ .
- $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  est un polynôme de moyenne mobile (MA) de degré  $q$
- $L$  est l'opérateur de retard défini par  $LX_t = X_{t-1}$ .
- $\varepsilon_t$  est un bruit blanc centré, non corrélé, de variance constante.

### 2.5.3 Le modèle SARIMA

Les modèles SARIMA (S pour saisonnalité) permettent de rendre compte des variations saisonnières dans la série considérée, ces variations pouvant elles-mêmes présenter un caractère aléatoire.

**Définition 2.5.7.** Un processus  $(X_t)_{t \in \mathcal{T}}$  suit un modèle SARIMA  $(p, d, q)(P, D, Q)_s$  (Seasonal AutoRegressive Integrated Moving Average) s'il vérifie l'équation suivante :

$$\Phi(L)\Phi_s(L^s)(1 - L)^d(1 - L^s)^D X_t = \Theta(L)\Theta_s(L^s)\varepsilon_t,$$

où  $L$  est l'opérateur de retard défini par  $LX_t = X_{t-1}$ , et  $\varepsilon_t$  est un bruit blanc.

où :

- $p, d, q$  : ordres de l'AR, différenciation et MA non saisonniers,
- $P, D, Q$  : ordres de l'AR, différenciation et MA saisonniers,
- $s$  : périodicité de la saisonnalité (ex. : 12 pour des données mensuelles),
- $\Phi(L)$  : polynôme AR non saisonnier,  $\Theta(L)$  : polynôme MA non saisonnier,
- $\Phi_s(L^s)$  : polynôme AR saisonnier,  $\Theta_s(L^s)$  : polynôme MA saisonnier,
- $(1 - L)^d$  : différenciation ordinaire d'ordre  $d$ ,
- $(1 - L^s)^D$  : différenciation saisonnière d'ordre  $D$ .

Après avoir envisagé les différentes formes de modélisation possibles, l'étape suivante consiste à déterminer précisément quel modèle correspond le mieux à la série étudiée.

## 2.6 Identification de modèle

L'identification du modèle de séries chronologiques constitue une étape essentielle dans l'analyse des données temporelles. Elle consiste à déterminer la structure et le type de modèle les plus appropriés pour représenter les observations disponibles. Ce processus permet de mieux comprendre la dynamique des données et d'établir des prévisions fiables.

La méthode la plus couramment utilisée pour identifier les paramètres d'un modèle repose sur l'analyse des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle

(PACF). Ces représentations graphiques permettent d'examiner, pour chaque décalage  $k$ , la corrélation globale (ACF) et la corrélation partielle (PACF) entre les valeurs de la série.

- L'ordre  $p$  du modèle AR (Auto-Régressif) est généralement déterminé à partir du PACF, en repérant le premier décalage à partir duquel les corrélations chutent brusquement à zéro.

- L'ordre  $q$  du modèle MA (Moyenne Mobile) est estimé à l'aide de l'ACF, en observant le premier décalage pour lequel les autocorrélations deviennent non significatives.

**Exemple 2.6.1.** Dans la figure 2.8 les graphiques de l'ACF et de la PACF, nous pouvons déterminer ces points de "rupture" pour estimer les valeurs de  $p$  et  $q$ .

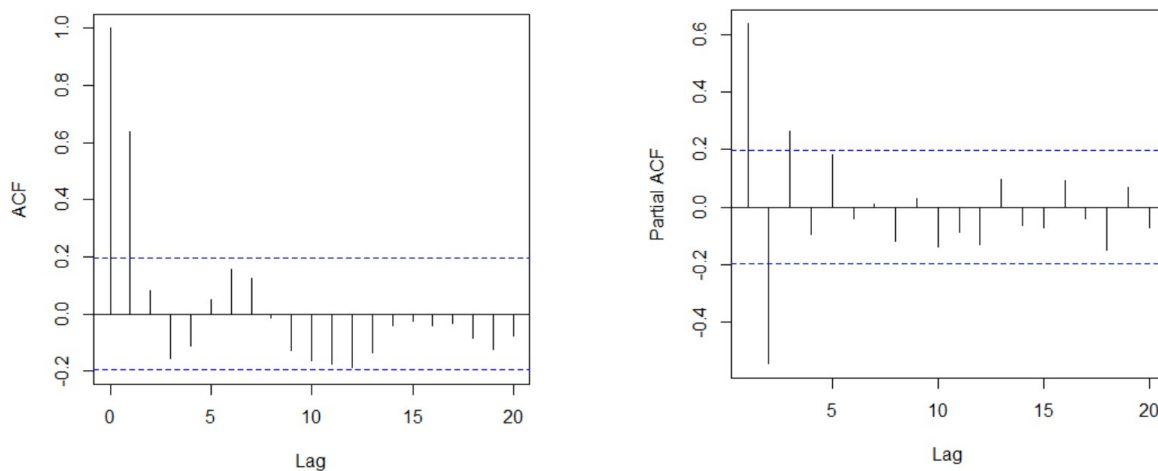


FIGURE 2.8 – Graphiques de l'ACF et du PACF

Dans cet exemple, on remarque que sur le graphe ACF, le retard à partir duquel les corrélations deviennent non significatives est 2, suggérant un ordre  $q$  de 2 pour la moyenne mobile. De même, d'après le graphe PACF, les corrélations significatives sont présentes jusqu'au retard 1, suggérant un ordre  $p$  de 1 pour le processus autorégressif.

Afin d'optimiser la performance du modèle, il est proposé d'explorer l'utilisation de valeurs inférieures pour les paramètres  $p$  et  $q$  par rapport aux estimations initiales. Cette démarche s'appuie sur des critères d'information tels que le Critère d'Information d'Akaike (AIC) et le Critère d'Information Bayésien (BIC).

### 2.6.1 Critères d'information (AIC, BIC)

Les critères d'information d'Akaike (AIC) et bayésien (BIC) sont utilisés pour sélectionner le modèle le plus adapté en équilibrant la précision des prévisions et la complexité du modèle.

L'idée des critères AIC et BIC est de fournir des fonctions permettant d'évaluer la qualité de l'ajustement d'un modèle. On sait que si le nombre de paramètres augmente, la variance  $\hat{\sigma}^2$  diminue, mais cela peut entraîner une surparamétrisation du modèle. Le critère AIC ajoute un facteur de pénalité basé sur le nombre de paramètres afin de trouver un équilibre entre complexité et ajustement optimal. De son côté, le critère BIC applique

une pénalisation plus stricte, tenant compte également de la taille de l'échantillon, favorisant ainsi les modèles plus simples, particulièrement lorsque les données sont nombreuses. Ainsi, l'application de ces critères permet de déterminer la meilleure structure à adopter, garantissant un équilibre optimal entre fidélité aux données et généralisation des prévisions.

$$AIC = \log(\hat{\sigma}^2) + 2K.$$

$$BIC = \log(\hat{\sigma}^2) + K \log(n).$$

Où :

- $\log(\hat{\sigma}^2)$  est une mesure de l'ajustement basée sur la variance des résidus.
- $K$  est le nombre total de paramètres estimés (paramètres AR, MA, SAR, SMA et éventuellement une constante).
- $n$  est le nombre d'observations dans la série temporelle.

**Remarque 2.6.2.** *Il faut tester plusieurs modèles ARMA avec différentes valeurs de  $p$  et  $q$ , calculer AIC et BIC pour chaque modèle, puis choisir celui qui a la plus petite valeur dans l'un des deux critères.*

Aux côtés des critères d'information tels que l'AIC et le BIC, qui sont utilisés pour sélectionner le modèle le plus approprié en équilibrant qualité d'ajustement et complexité, il existe d'autres mesures permettant d'évaluer la précision prédictive du modèle qui servent à juger les performances du modèle, en particulier sur des données de validation ou de test.

Les critères standards utilisés sont les suivants :

- Le critère de Schwarz (1978) :

$$SC(K) = n \log(\hat{\sigma}_\varepsilon^2) + K \log(n).$$

- L'erreur absolue moyenne (Mean Absolute Error) :

$$MAE = \frac{1}{n} \sum_{t=1}^n |\varepsilon_t|.$$

Où  $\varepsilon_t$  est le résidu du modèle étudié et  $n$  est le nombre d'observations.

- L'erreur quadratique moyenne (Mean Squared Error) :

$$MSE = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2.$$

- La racine carrée de l'erreur quadratique moyenne (Root Mean Squared Error) :

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n \varepsilon_t^2}.$$

- L'écart absolu moyen en pourcentage (Mean Absolute Percentage Error) :

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{\varepsilon_t}{X_t} \right|.$$

### 2.6.2 Ajustement du modèle et estimation des paramètres

On peut effectuer l'estimation des paramètres d'un modèle ARMA(p,q), en supposant que les ordres p et q sont connus, en utilisant diverses techniques dans le domaine temporel :

- Moindres Carrés Ordinaires (MCO). Lorsque le modèle ne contient pas de composante MA ( $q = 0$ ).
- Maximum de Vraisemblance Approché .
- Maximum de Vraisemblance Exacte .

Dans un modèle ARMA(p, q), on suppose que les erreurs  $\varepsilon_t$  suivent une loi normale  $N(0, \sigma^2)$ .

Il s'ensuit que :

$$\varepsilon_t = X_t - \sum_{i=1}^p \phi_i X_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j},$$

Puisque  $\varepsilon_t \sim N(0, \sigma^2)$ , la probabilité d'obtenir une erreur  $\varepsilon_t$  donnée est :

$$P(\varepsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right).$$

### 2.6.3 Validation de modèle

Lorsque le modèle est bien ajusté, les résidus  $\hat{\varepsilon}_t$  doivent se comporter comme un bruit blanc. Les principaux tests permettant de vérifier cette hypothèse seront brièvement présentés dans la suite, tandis que leurs développements sont reportés en annexe.

#### • Test de nullité de la moyenne des résidus

Soit  $n$  le nombre de données disponibles (après avoir enlevé les retards correspondant aux termes AR et MA). Si le processus  $\{\varepsilon_t; t \in \mathcal{T}\}$  est i.i.d.  $(0, \sigma_\varepsilon^2)$ , on doit avoir :

$$\bar{\varepsilon}_t = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t \xrightarrow[n \rightarrow \infty]{} 0.$$

Par application du théorème central limite, on montre que :

$$\frac{\bar{\varepsilon}_n}{\hat{\sigma}_\varepsilon/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

On peut tester la nullité de la moyenne des résidus en construisant l'intervalle de confiance sur  $\bar{\varepsilon}_n$  au seuil standard de 95% :

$$P\left(\bar{\varepsilon}_n \in \left[-1,96 \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{n}}, 1,96 \cdot \frac{\hat{\sigma}_\varepsilon}{\sqrt{n}}\right]\right) = 0,95.$$

### • Test d'autocorrélation des résidus

Si les résidus  $\{\varepsilon_t, t \in \mathcal{T}\}$  obéissent à un bruit blanc, il ne doit pas exister d'autocorrélation dans la série. On peut alors utiliser les différents tests suivants :

1. Test de Durbin-Watson il permet de tester l'autocorrélation d'ordre 1.
2. Étude de l'ACF et de la PACF on doit vérifier qu'il n'existe aucune autocorrélation ou autocorrélation partielle significativement non nulle pour le processus étudié. Cette étude est prolongée par les tests dits ( porte-manteau ).
3. Tests du porte-manteau (tests d'adéquation globale du modèle) ces tests reposent sur l'idée que l'ACF d'un bruit blanc ne doit pas révéler d'autocorrélations non nulles. Les plus couramment utilisés sont les tests de Box-Pierce et de Ljung-Box.

### • Test d'homoscédasticité

Un bruit blanc est par définition homoscédastique. Tous les tests d'hétéroscédasticité peuvent ici être employés. Test de Chow (comparaison des variances des résidus sur des sous périodes de la chronique).

### • Tests de normalité

Pour vérifier si le processus des résidus  $\{\varepsilon_t, t \in \mathcal{T}\}$  est un bruit blanc gaussien, plusieurs tests peuvent être utilisés, mais le test le plus courant est celui de Jarque et Bera. Ce dernier est fondé sur la notion de skewness (moment d'ordre 3 et asymétrie) et de Kurtosis (moment d'ordre 4 et queue de distribution).

L'analyse théorique développée dans ce chapitre a permis de clarifier la nature des séries stationnaires, d'en examiner les principales propriétés, et d'établir les fondements nécessaires à leur modélisation. Nous abordons à présent l'aspect pratique, en appliquant ces concepts à des séries temporelles issues de données empiriques. Ce nouveau chapitre mettra en œuvre les différentes étapes de construction, d'ajustement et d'évaluation des modèles ARMA, ARIMA et SARIMA, en insistant sur leur capacité à fournir des prévisions robustes et pertinentes.

# Chapitre 3

## Traitements et Modélisation Pratiques des Séries chronologiques

Ce chapitre illustre, à l'aide de données réelles, la démarche d'analyse et de modélisation de séries temporelles. Il mobilise les notions fondamentales de stationnarité, les transformations nécessaires en cas de non-stationnarité, ainsi que la mise en œuvre de modèles adaptés (ARMA, ARIMA, SARIMA) permettant de comprendre les structures temporelles sous-jacentes et de produire des prévisions fiables.

Les traitements ont été réalisés à l'aide du logiciel R (version 4.2.2), en mobilisant principalement les packages `forecast`, `tseries`, `urca` et `ggplot2`, pour la modélisation, les tests de stationnarité, et la visualisation graphique. La fonction `auto.arima()` a notamment été utilisée pour identifier automatiquement la structure optimale du modèle, et `checkresiduals()` pour en évaluer la qualité.

Ce travail s'appuie sur des références fondamentales, parmi lesquelles [3],[5],[10],[16]et [18], utilisées pour la base théorique, les méthodes de test, ainsi que la mise en œuvre pratique des modèles de séries temporelles. La liste complète figure en bibliographie.

### 3.1 Modélisation des données à l'aide du modèle ARMA

#### 3.1.1 Présentation et exploration des données

Les données représentent les quantités mensuelles de ciment (en tonnes) dans la wilaya de Aïn Témouchent (Bni saf) pour la période de janvier 2013 à décembre 2019.

Mois	2013	2014	2015	2016	2017	2018	2019
Janvier	100420.22	80376.50	85027.82	120738.26	91813.78	93620.20	93823.68
Février	89345.52	102045.56	90435.52	122636.02	100863.40	81844.61	87833.18
Mars	62667.60	109028.66	99488.96	94289.24	97118.36	92570.46	100155.64
Avril	52566.06	107694.36	88904.42	61334.98	84567.24	75954.26	92957.06
Mai	84857.36	90244.12	92905.53	122744.04	54822.32	80014.76	84364.06
Juin	97804.30	92711.14	59421.38	89380.84	62488.60	54292.32	69596.20
Juillet	100770.99	80119.62	85327.80	75795.36	115036.54	90237.34	83904.96
Août	73317.54	99935.14	109594.98	109212.58	111639.94	66123.54	65968.04
Septembre	108246.68	109815.58	110110.60	99051.68	95643.64	94661.88	83242.56
Octobre	99214.38	107858.76	117541.76	116242.44	109205.50	106955.04	99584.52
Novembre	87335.30	113212.20	122844.36	119725.14	108098.20	100655.72	69515.66
Décembre	66012.96	107807.94	76393.96	105062.24	86702.16	113634.72	41650.94

TABLE 3.2 – Quantités mensuelles de ciment (en tonnes) à Beni Saf (2013-2019)

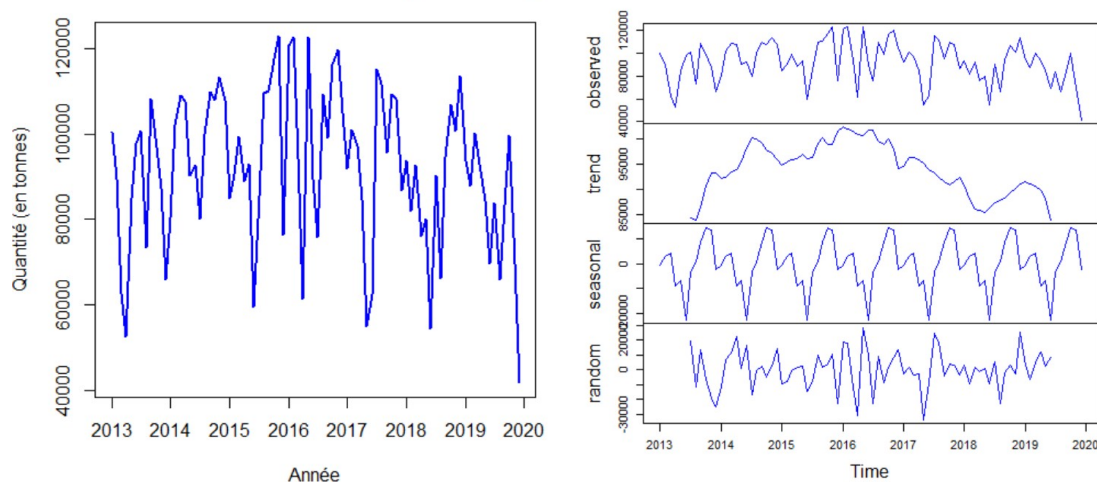


FIGURE 3.1 – Évolution mensuelle de la production de ciment et sa décomposition

L'analyse de la série chronologique (2013–2019) a été réalisée sous R à l'aide des fonctions `plot.ts()` pour la visualisation globale et `decompose()` pour la décomposition en composantes tendance, saisonnière et irrégulière, après avoir converti les données en objet de type `ts`.

Le graphique de la série montre une évolution autour d'une moyenne stable, avec une légère baisse à partir de 2016, mais sans tendance statistiquement significative.

La décomposition révèle une composante saisonnière peu régulière et faiblement marquée, ce qui ne permet pas de conclure à une saisonnalité significative. L'ensemble des observations visuelles suggère que la série est stationnaire au second ordre, c'est-à-dire qu'elle présente une moyenne et une variance constantes au cours du temps.

L'analyse est ensuite approfondie par l'examen des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF), obtenues à l'aide des fonctions `acf()` et `pacf()` de R, appliquées directement à la série d'origine afin d'évaluer visuellement sa stationnarité.

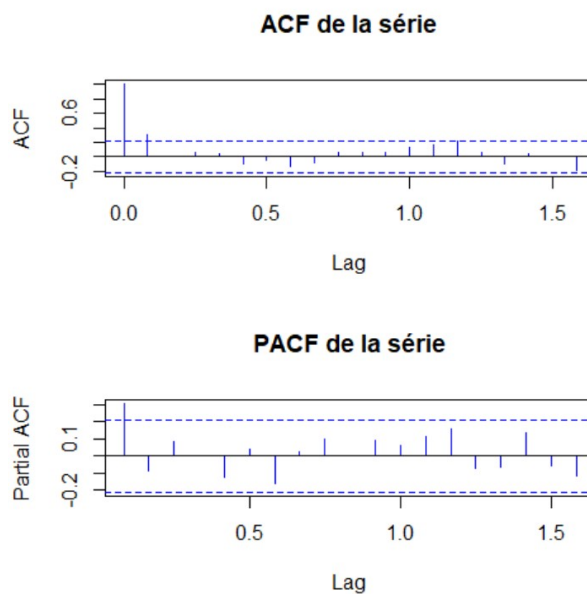


FIGURE 3.2 – ACF et PACF de la série de quantité mensuelle de ciment

L'ACF montre une chute rapide après le premier retard (retard 1), ce qui suggère une stationnarité de la série. La PACF présente également un seul pic significatif à ce même retard, renforçant cette hypothèse.

**Remarque 3.1.1.** *Il est possible qu'une série temporelle soit stationnaire autour d'un cycle saisonnier, ce qui ne constitue pas une contradiction. Par conséquent, le test ADF peut conclure à la stationnarité de la série, même si celle-ci présente une composante saisonnière marquée.*

Afin de valider cette première analyse visuelle, des tests statistiques de stationnarité seront réalisés à l'aide des fonctions `adf.test()`, `kps.test()` et `pp.test()` disponibles dans les packages `tseries` et `urca`.

• Vérification de la stationnarité

```

> adf.test(data_ts)

        Augmented Dickey-Fuller Test

data: data_ts
Dickey-Fuller = -4.0518, Lag order = 4, p-value = 0.01119
alternative hypothesis: stationary

> kpss.test(data_ts)

        KPSS Test for Level Stationarity

data: data_ts
KPSS Level = 0.27184, Truncation lag parameter = 3, p-value = 0.1

Message d'avis :
Dans kpss.test(data_ts) : p-value greater than printed p-value
> pp.test(data_ts)

        Phillips-Perron Unit Root Test

data: data_ts
Dickey-Fuller Z(alpha) = -54.297, Truncation lag parameter = 3, p-value
= 0.01
alternative hypothesis: stationary

```

FIGURE 3.3 – Résultats des tests de stationnarité (ADF, PP et KPSS) .

Les tests ADF et Phillips-Perron (PP) présentent des p-values inférieures au seuil de 5 %, ce qui conduit au rejet de l'hypothèse nulle de racine unitaire et indique que la série est stationnaire. Par ailleurs, le test KPSS, avec une p-value supérieure à ce seuil, ne rejette pas l'hypothèse nulle de stationnarité, confirmant ainsi la stationnarité de la série.

Ces conclusions signifient que nous pouvons continuer à modéliser la série en utilisant des modèles ARMA.

### 3.1.2 Construction du modèle ARMA

L'identification automatique du modèle le plus adapté a été réalisée à l'aide de la fonction `auto.arima()` du package `forecast`, qui sélectionne les paramètres optimaux en minimisant le critère d'information (AIC).

```

Series: ts_data
ARIMA(0,0,1) with non-zero mean

Coefficients:
      mal      mean
    0.3546  91984.52
s.e.  0.0983  2517.43

sigma^2 = 2.99e+08: log likelihood = -937.92
AIC=1881.83  AICc=1882.13  BIC=1889.12

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set -25.41511 17084.42 13552.96 -4.377535 16.72309 0.7878067
      ACF1
Training set 0.01417224

```

FIGURE 3.4 – Ajustement du modèle ARMA à la série des quantités de ciment.

Le modèle ARMA(0,1), identifié à l'aide de la fonction `auto.arima()` dans R, suggère que la dynamique de la série est bien capturée par une composante moyenne mobile d'ordre 1, sans terme autorégressif ni différenciation. Les coefficients estimés sont significatifs, et les mesures d'erreur (RMSE = 17084.42, MAE = 13552.96, MAPE = 16.72%) indiquent une qualité de prévision raisonnable. L'autocorrélation des résidus (ACF1 = 0.014) est proche de zéro, ce qui renforce l'adéquation du modèle.

Après l'identification du modèle ARMA(0,1), l'analyse des résidus permet de valider que le modèle capture correctement la dynamique de la série sans laisser de structure prévisible non expliquée. Cette étape est essentielle pour s'assurer que les hypothèses du modèle sont respectées, notamment l'absence d'autocorrélation et la stabilité des erreurs.

### • Diagnostic des résidus

L'analyse des résidus a été réalisée à l'aide de la fonction `checkresiduals()` dans R de modèle ARMA(0,1)

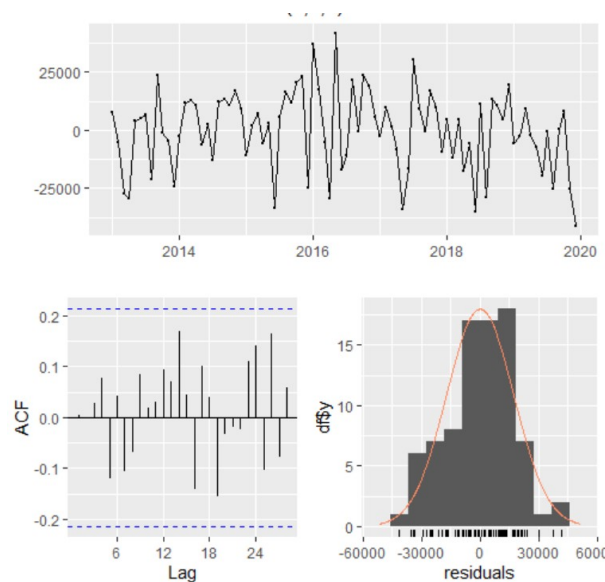


FIGURE 3.5 – Les résidus du modèle ARMA(0,1)

Elle montre qu'ils sont centrés autour de zéro sans tendance apparente, ce qui indique une bonne stationnarité des erreurs. L'ACF des résidus ne révèle aucune autocorrélation significative, suggérant que le modèle capture correctement la dynamique temporelle de la série. Par ailleurs, l'histogramme des résidus présente une distribution proche de la normale, avec une légère asymétrie tolérable dans un cadre empirique. L'ensemble de ces éléments confirme la validité de l'ajustement du modèle.

### 3.1.3 La prévision

La prévision a été réalisée à partir du modèle ajusté à l'aide de la fonction `forecast()`, permettant de générer les valeurs attendues. La visualisation obtenue via `plot()` met en évidence la tendance projetée ainsi que les intervalles de confiance, facilitant l'interprétation des résultats.

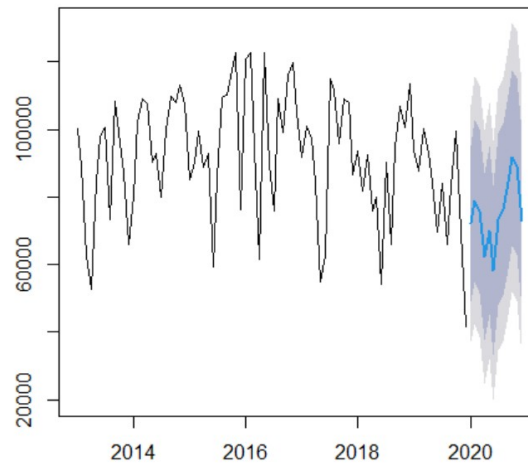


FIGURE 3.6 – Prévisions des ventes de ciment à l'aide d'un modèle ARMA

Les courbes bleues représentent les valeurs prévues pour les périodes futures. On observe que ces prévisions suivent une dynamique cohérente avec le comportement de la série d'origine, ce qui indique que le modèle ARMA utilisé est bien adapté et fournit des prévisions fiables concernant la demande future de ciment.

## 3.2 Modélisation des données à l'aide du modèle ARIMA

### 3.2.1 Présentation et exploration des données

Les données utilisées dans cette étude représentent les taux de réussite au baccalauréat en Algérie sur une période de 56 ans, allant de (1976 - 2023). Ces taux, exprimés en pourcentage (%), reflètent la proportion des candidats ayant réussi l'examen par rapport au nombre total des inscrits chaque année. <https://www.ons.dz>

La série temporelle obtenue est de fréquence annuelle (périodicité 1), ce qui signifie qu'une observation est enregistrée pour chaque année. Elle constitue une série univariée qui sera analysée afin d'étudier son comportement au fil du temps et d'identifier un modèle de prévision adéquat.

Le tableau ci-dessous présente l'évolution annuelle des taux de réussite :

Année	Taux (%)	Année	Taux (%)	Année	Taux (%)
1976	20.65	1991	19.05	2006	50.65
1977	22.49	1992	22.83	2007	51.21
1978	23.50	1993	11.07	2008	52.21
1979	21.75	1994	16.54	2009	45.04
1980	23.67	1995	20.72	2010	61.23
1981	36.61	1996	24.07	2011	62.85
1982	19.77	1997	27.52	2012	59.26
1983	18.93	1998	24.00	2013	44.78
1984	24.48	1999	24.73	2014	44.87
1985	29.06	2000	32.28	2015	51.35
1986	24.55	2001	33.71	2016	49.79
1987	16.72	2002	32.93	2017	56.07
1988	24.21	2003	29.13	2018	55.88
1989	17.97	2004	42.66	2019	54.28
1990	20.56	2005	36.27	2020	55.30
				2021	61.30
				2022	58.75
				2023	50.63

TABLE 3.3 – Taux de réussite au baccalauréat de 1976 à 2023

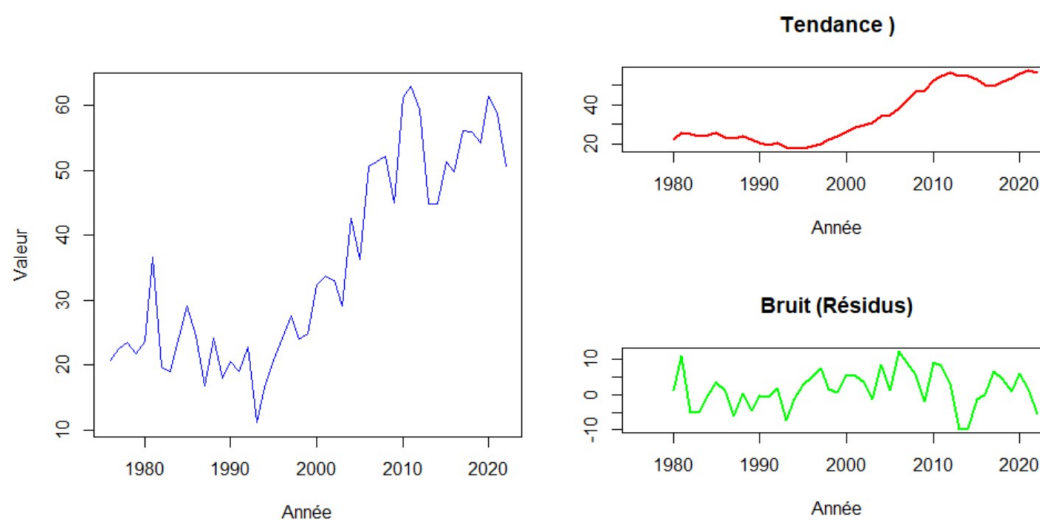


FIGURE 3.7 – Évolution temporelle du taux de réussite entre 1976 et 2023

L’observation initiale, effectuée grâce à la fonction `plot.ts()`, met en évidence une tendance haussière marquée, en particulier à partir de l’année 2016, traduisant une croissance progressive du phénomène étudié. Cette évolution suggère une instabilité au niveau de la moyenne, dans la mesure où la série ne fluctue pas autour d’une moyenne constante au fil du temps. Afin de mieux cerner la tendance globale, un lissage a été appliqué à travers la technique de la moyenne mobile simple avec une fenêtre de cinq périodes  $SMA(n = 5)$ , ce qui a permis d’atténuer les variations aléatoires et de faire ressortir plus clairement

la composante tendancielle. L'ensemble de ces éléments indique que la série analysée présente une non-stationnarité en moyenne.

Afin de confirmer ces premières observations sur la nature de la série, une analyse des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) a été réalisée, dans le but de révéler la structure temporelle sous-jacente de la série.

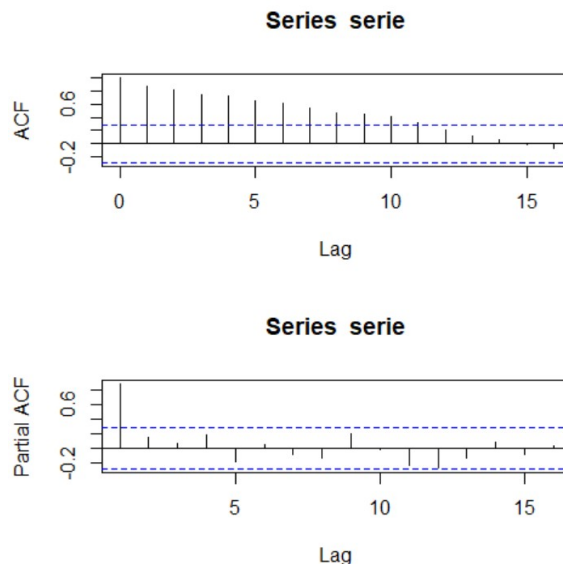


FIGURE 3.8 – ACF et PACF de la série temporelle du taux de réussite

Les graphes ACF et PACF révèlent une forte autocorrélation dans la série, notamment aux faibles retards, ce qui indique une non-stationnarité probable. L'ACF décroît lentement, tandis que la PACF présente un pic marqué au premier retard suivi de valeurs négligeables, ce qui est caractéristique d'une série ayant une structure de dépendance à court terme. Ces observations suggèrent qu'une différenciation est nécessaire pour stabiliser la série avant d'envisager l'ajustement d'un modèle ARIMA approprié.

### • Vérification de la stationnarité

```
> adf.test(serie)
Augmented Dickey-Fuller Test
data: serie
Dickey-Fuller = -1.6295, Lag order = 3, p-value = 0.7217
alternative hypothesis: stationary

> kpss.test(serie)
KPSS Test for Level Stationarity
data: serie
KPSS Level = 1.0684, Truncation lag parameter = 3, p-value = 0.01
Message d'avis :
Dans kpss.test(serie) : p-value smaller than printed p-value

> pp.test(serie)
Phillips-Perron Unit Root Test
data: serie
Dickey-Fuller Z(alpha) = -14.23, Truncation lag parameter = 3, p-value = 0.2474
alternative hypothesis: stationary
```

FIGURE 3.9 – Les tests de stationnarité (ADF, PP, KPSS) du taux de réussite

Les résultats des tests complémentaires (ADF, KPSS, PP) indiquent que la série n'est pas stationnaire. En effet, le test ADF ne rejette pas l'hypothèse de racine unitaire ( $p > 5\%$ ), le test KPSS rejette l'hypothèse de stationnarité ( $p < 1\%$ ), et le test PP ne rejette pas l'hypothèse de non-stationnarité ( $p > 5\%$ ). Ainsi, tous les tests confirment que la série étudiée est non stationnaire à son niveau.

Une fois la non-stationnarité confirmée à travers l'analyse statistique, il devient alors nécessaire d'appliquer des méthodes de transformation appropriées.

### 3.2.2 Traitement de la non-stationnarité

Pour traiter la non-stationnarité observée, nous avons appliqué une différenciation de la série à l'aide de la commande `diff()` sous R.

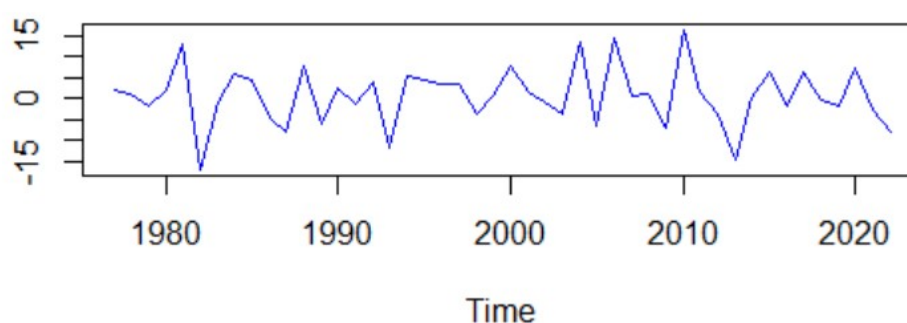


FIGURE 3.10 – La série différenciée du taux de réussite

L'application de la différenciation a permis de stabiliser la série temporelle, en éliminant la tendance détectée précédemment. La nouvelle série obtenue oscille désormais de manière aléatoire autour d'une moyenne constante

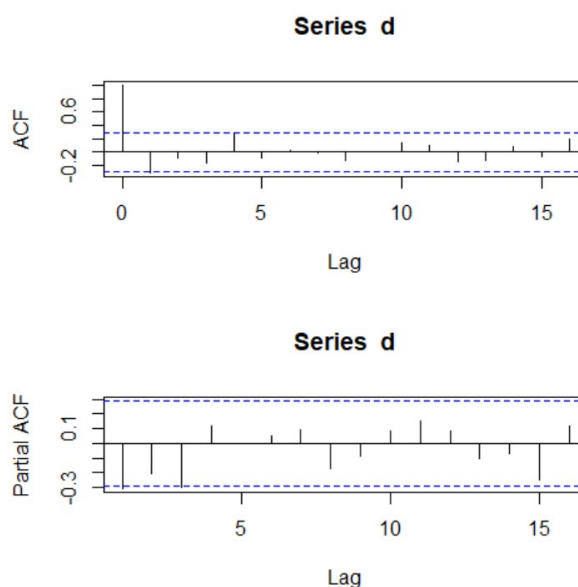


FIGURE 3.11 – ACF et PACF de la La série différenciée du taux

Les graphes ACF et PACF de la série différenciée révèlent une structure aléatoire, caractérisée par l'absence d'autocorrélations significatives persistantes, ce qui est compatible avec l'hypothèse de stationnarité. Cette impression visuelle sera confirmée par l'application du test de Dickey-Fuller augmenté (ADF), permettant de vérifier formellement la présence ou non de racine unitaire.

- **Vérification de la stationnarité**

```
Augmented Dickey-Fuller Test

data: d
Dickey-Fuller = -3.7076, Lag order = 3, p-value = 0.03487
alternative hypothesis: stationary

> pp.test(d)

Phillips-Perron Unit Root Test

data: d
Dickey-Fuller Z(alpha) = -49.823, Truncation lag parameter = 3, p-value = 0.01
alternative hypothesis: stationary

Message d'avis :
Dans pp.test(d) : p-value smaller than printed p-value
> kpss.test(d)

KPSS Test for Level Stationarity

data: d
KPSS Level = 0.10479, Truncation lag parameter = 3, p-value = 0.1
```

FIGURE 3.12 – Les tests de stationnarité (ADF, PP, KPSS) après différenciation

Le test de Dickey-Fuller augmenté (ADF) appliqué à la série différenciée donne une p-valeur de 0,03487, ce qui permet de conclure à la stationnarité de la série au seuil de 5%. Cela confirme que la série est intégrée d'ordre 1 ( $I(1)$ ). Par ailleurs, la fonction `auto.arima()` de R prend en charge automatiquement cette étape de transformation en détectant la nécessité de différencier la série, puis en identifiant et ajustant un modèle ARIMA optimal sur la version stationnaire obtenue.

### 3.2.3 Construction du modèle ARIMA

L'ajustement du modèle ARIMA(0,1,1), obtenu automatiquement à l'aide de la fonction `auto.arima()` du logiciel R, a été suivi du calcul d'indicateurs d'évaluation (tels que RMSE, MAE, MAPE) permettant de juger de l'efficacité du modèle en termes de prévision.

```

ARIMA(0,1,1)

Coefficients:
      ma1
      -0.4203
s.e.      0.1349

sigma^2 = 42.29: log likelihood = -150.99
AIC=305.98  AICc=306.26  BIC=309.63

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 1.215639 6.363334 4.866942 0.2285989 16.08708 0.9326751 -0.0184589
    
```

FIGURE 3.13 – Ajustement du modèle ARIMA au taux de réussite.

Les résultats des différents indicateurs montrent que le modèle ARIMA est bien ajusté à la série temporelle analysée. L'erreur moyenne (ME = 1.2156) suggère une légère sous-estimation. La précision globale est acceptable avec un RMSE de 6.3633 et un MAE relativement faible (4.8669). Le biais est quasi nul, comme l'indique le MPE (0.2286%). Le MAPE (16.087%) étant inférieur à 20%, cela reflète une bonne précision. Le MASE (0.9327 < 1) montre que le modèle surpasse un modèle naïf de référence. Enfin, l'ACF1 (0.0185) confirme l'absence d'autocorrélation significative des résidus.

L'analyse des résidus vise à vérifier que les erreurs du modèle se comportent comme un bruit blanc, c'est-à-dire qu'elles sont indépendantes, non autocorrélées et centrées, ce qui constitue une hypothèse fondamentale pour garantir la fiabilité des résultats issus du modèle ARIMA(0,1,1).

- **Diagnostic des résidus**

La fonction `checkresiduals()` du package `forecast` dans R permet d'évaluer de manière synthétique la qualité des résidus du modèle ajusté.

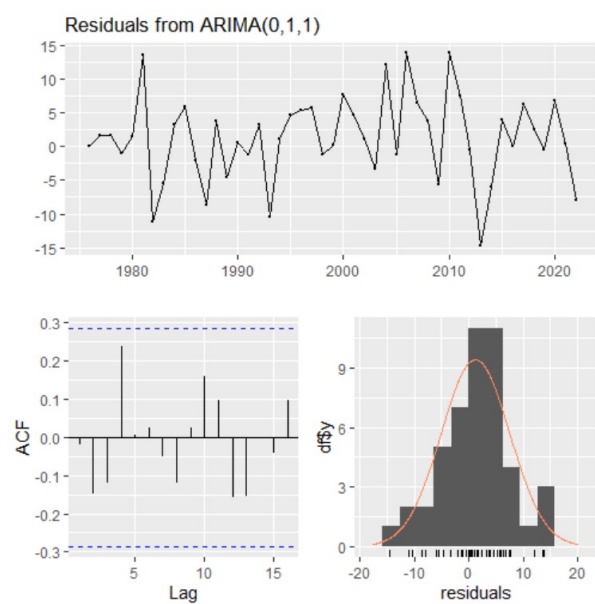


FIGURE 3.14 – Les résidus du modèle ARIMA(0,1,1)

L'analyse des résidus du modèle ARIMA(0,1,1) révèle des résultats globalement satisfaisants. Les résidus oscillent autour de zéro sans tendance marquée, ce qui suggère l'absence de biais systématique. L'ACF des résidus ne montre aucune autocorrélation significative, indiquant que les erreurs sont statistiquement indépendantes.

En ce qui concerne la distribution des résidus, elle s'approche globalement d'une distribution normale, bien qu'on observe une légère asymétrie et quelques écarts à la normalité parfaite. Cela implique que, même si la condition de normalité stricte n'est pas entièrement vérifiée, elle est suffisamment approchée pour garantir la validité du modèle à des fins de prévision.

Ces éléments confirment que le modèle est correctement ajusté à la série temporelle étudiée, et qu'il peut être raisonnablement utilisé pour générer des prévisions sur les périodes futures.

### 3.2.4 Prévision

La fonction `forecast()` a été appliquée pour projeter les valeurs sur les périodes futures définies.

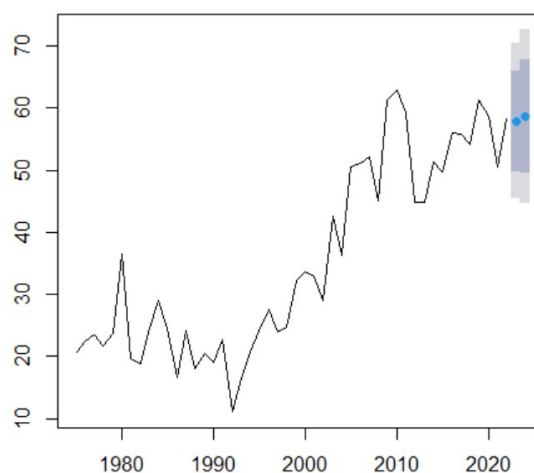


FIGURE 3.15 – Prévisions du taux de réussite à l'aide d'un modèle ARIMA

```
> forecast_arima <- forecast(modele_auto, h = 2);forecast_arima
  Point Forecast   Lo 80   Hi 80   Lo 95   Hi 95
2024      57.90306 49.74220 66.06391 45.42211 70.38401
2025      58.66945 49.54756 67.79134 44.71873 72.62018
```

FIGURE 3.16 – Résultats de la prévision du taux de réussite

Le modèle ARIMA prévoit une légère amélioration du taux de réussite au baccalauréat entre 2024 et 2025, avec des valeurs centrales proches de 58%. Pour l'année 2024, la prévision ponctuelle (57,90%) se révèle très proche de la valeur réelle observée (58,28%), ce qui renforce la pertinence du modèle à court terme.

Cependant, les intervalles de confiance relativement larges témoignent d'une incertitude élevée, notamment pour 2025, et imposent de rester prudent dans l'interprétation des résultats.

Ces prévisions offrent une base utile pour l'analyse prospective, mais doivent être enrichies par des données qualitatives ou contextuelles telles que les réformes éducatives, les conditions d'enseignement ou les dynamiques sociales afin de mieux comprendre les facteurs susceptibles d'influencer les évolutions futures du taux de réussite.

### 3.3 Modélisation des données à l'aide du modèle SA-RIMA

#### 3.3.1 Présentation et exploration des données

Les données utilisées dans cette étude représentent l'évolution de la valeur des importations de marchandises aux prix courants en Algérie, sur la période allant de janvier 2014 à décembre 2023. Elles sont exprimées en milliards de dinars algériens (DA) et sont structurées sous forme d'une série chronologique comprenant des valeurs mensuelles et annuelles. L'analyse de ces données révèle une saisonnalité marquée avec une périodicité de 12 mois. <https://www.ons.dz>

Le tableau ci-dessous présente l'évolution annuelle des importations :

	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Janvier	355,3	395,4	417,4	446,7	443,6	465,4	372,4	376,4	411,1	456,7
Février	356,0	356,3	408,4	432,9	416,7	421,0	365,1	379,6	448,6	425,0
Mars	405,6	484,9	453,7	430,6	428,7	458,1	365,5	445,8	454,3	496,1
Avril	471,7	497,0	390,2	448,1	443,0	437,3	371,0	391,0	442,1	467,5
Mai	436,7	466,8	384,5	434,5	495,0	437,4	324,8	416,1	443,9	519,4
Juin	411,9	474,7	395,5	405,0	403,7	406,3	392,0	462,1	484,6	477,2
Juillet	354,7	408,9	363,6	426,7	465,1	450,8	358,1	365,8	433,3	453,1
Août	396,1	408,5	394,0	451,2	410,0	409,3	371,3	423,7	500,9	523,2
Septembre	392,9	397,6	359,3	346,1	430,6	412,7	368,1	434,3	523,2	430,1
Octobre	399,4	426,1	428,0	472,2	496,2	374,9	341,6	404,3	481,3	540,1
Novembre	398,8	432,8	432,6	434,5	436,4	339,5	398,7	419,8	473,8	484,0
Décembre	428,3	459,2	475,9	450,5	512,1	429,6	395,5	529,8	481,3	519,1

TABLE 3.4 – Les valeurs des importations de marchandises aux prix courants en Algérie

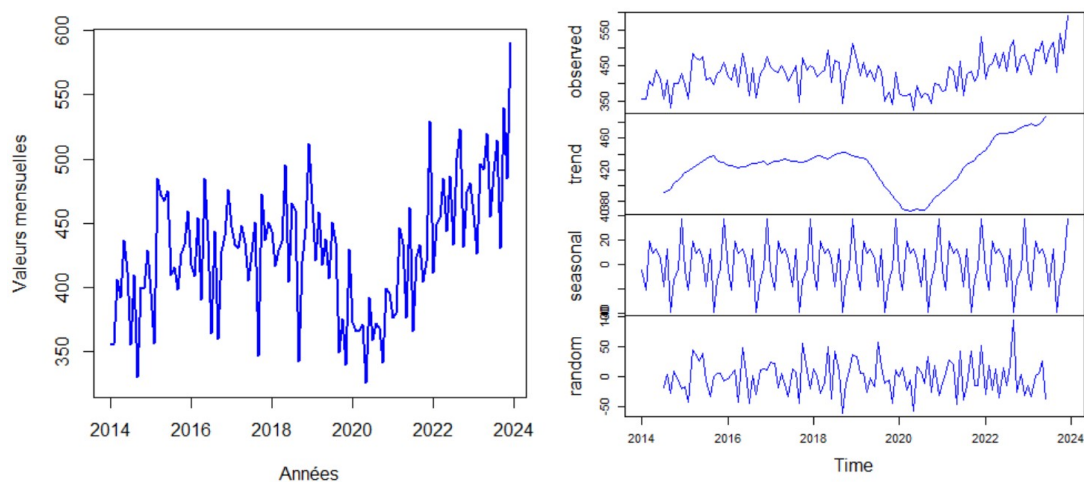


FIGURE 3.17 – Évolution temporelle des importations de marchandises

La série présente une forte variabilité initiale entre 2014 et 2018, suivie d’une baisse significative à partir de 2019, puis d’une reprise progressive dès 2021, culminant en 2023 avec une tendance haussière marquée. On observe également des fluctuations saisonnières régulières, suggérant un effet saisonnier récurrent.

Ces observations empiriques traduisent une non-stationnarité manifeste, résultant à la fois d’une tendance de fond et d’une saisonnalité structurelle.

Pour confirmer cette hypothèse, nous avons analysé les fonctions d’autocorrélation (ACF) et d’autocorrélation partielle (PACF).

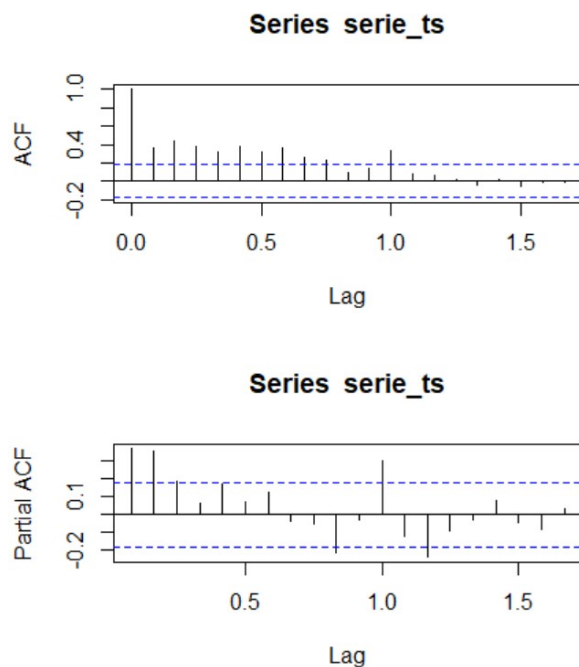


FIGURE 3.18 – ACF et PACF de la série des des importations de marchandise

Le graphique ACF présente une décroissance lente, ce qui indique que la série n’est pas stationnaire.

- Vérification de la stationnarité

```

> adf.test(serie_ts)

Augmented Dickey-Fuller Test

data: serie_ts
Dickey-Fuller = -1.6217, Lag order = 4, p-value = 0.7337
alternative hypothesis: stationary

> kpss.test(serie_ts)

KPSS Test for Level Stationarity

data: serie_ts
KPSS Level = 0.59012, Truncation lag parameter = 4, p-value = 0.02353
    
```

FIGURE 3.19 – Résultat du tests appliqué à la série des importations.

Deux tests complémentaires ont été appliqués : le test de Dickey-Fuller augmenté (ADF) et le test KPSS. Le test ADF a fourni une statistique de 1.6217 avec une p-value de 0.7337, ce qui ne permet pas de rejeter l’hypothèse nulle de non-stationnarité. En parallèle, le test KPSS a donné une statistique de 0.59012 et une p-value de 0.02353, indiquant un rejet de l’hypothèse nulle de stationnarité. Ainsi, ces deux tests convergent vers la même conclusion la série n’est pas stationnaire en niveau.

Ainsi, étant donné que la série temporelle étudiée n’est pas stationnaire et présente une composante saisonnière, elle a été modélisée à l’aide d’un modèle SARIMA.

### 3.3.2 Construction du modèle SARIMA

Le choix des ordres du modèle a été effectué automatiquement à l’aide de la commande `auto.arima()`, qui intègre la différenciation nécessaire et sélectionne les paramètres optimaux.

```

Series: serie_ts
ARIMA(1,1,1) (2,0,0) [12]

Coefficients:
      ar1      ma1      sar1      sar2
    -0.2867  -0.5989  0.3548  0.1825
s.e.   0.1175   0.0979  0.0956  0.1016

sigma^2 = 1307:  log likelihood = -595.82
AIC=1201.64  AICc=1202.17  BIC=1215.54

Training set error measures:
              ME  RMSE  MAE  MPE  MAPE  MASE
Training set 2.803242 35.3852 27.80921 0.09091497 6.541181 0.6631181
              ACF1
Training set -0.007774813
    
```

FIGURE 3.20 – Ajustement du modèle SARIMA à la série des importations

Le RMSE de 35,39 et le MAPE de 6,54% témoignent d’une précision de prévision raisonnable. De plus, l’ACF du résidu au premier retard, proche de zéro (0,0078), indique l’absence d’autocorrélation significative, ce qui renforce la validité du modèle ajusté.

• Diagnostic des résidus

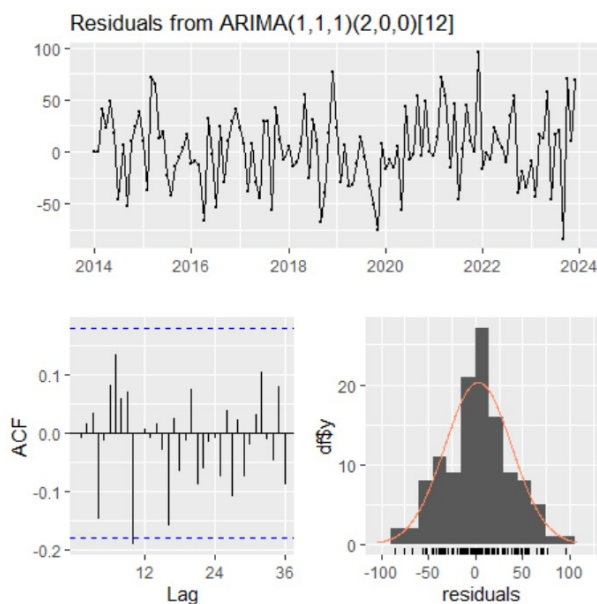


FIGURE 3.21 – Les résidus du modèle SARIMA

L'analyse des résidus confirme la validité du modèle SARIMA : les erreurs sont stationnaires, non autocorrélées et approximativement normales, indiquant une bonne adéquation aux hypothèses du modèle et une modélisation fiable de la série. Comme l'analyse des résidus a confirmé la validité du modèle SARIMA, nous pouvons maintenant procéder à la prévision.

### 3.3.3 Prévision

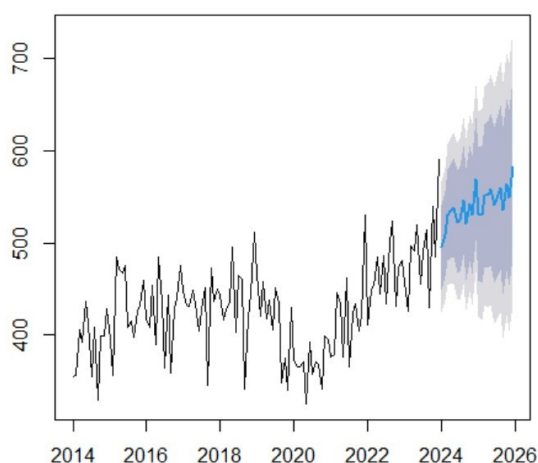


FIGURE 3.22 – Les prévisions obtenues à l'aide du modèle SARIMA

Le modèle SARIMA prévoit une hausse régulière de la série en 2024 et 2025. Bien que l'incertitude augmente avec le temps, les prévisions restent cohérentes avec les données passées. Cela montre que le modèle est fiable et que la série devrait continuer à croître modérément.

### 3.4 Comparaison entre les modèles ARMA, ARIMA et SARIMA

Les modèles ARMA, ARIMA et SARIMA sont largement utilisés pour la modélisation et la prévision des séries temporelles. Les modèles ARMA (AutoRegressive Moving Average) sont efficaces pour les séries stationnaires en capturant les dépendances temporelles à court terme. Lorsqu'une série présente une tendance, les modèles ARIMA (AutoRegressive Integrated Moving Average) permettent d'intégrer une étape de différenciation pour rendre la série stationnaire, ce qui élargit leur champ d'application. Enfin, les modèles SARIMA (Seasonal ARIMA) constituent une extension des ARIMA en intégrant des composantes saisonnières, ce qui les rend particulièrement adaptés aux séries présentant des schémas récurrents sur des périodes régulières (comme les données mensuelles ou trimestrielles).

Cependant, ces modèles présentent également certaines limites. Les modèles ARMA nécessitent que la série soit strictement stationnaire, ce qui peut limiter leur utilisation en pratique. De plus, les modèles ARIMA supposent que la structure de la série peut être entièrement captée par des composantes autorégressives, de moyenne mobile et de différenciation, ce qui n'est pas toujours le cas pour des séries plus complexes ou non linéaires. Quant aux modèles SARIMA, bien qu'ils soient capables de traiter la saisonnalité, ils deviennent rapidement complexes à paramétrer et à interpréter dès que le nombre de composantes augmente. Par ailleurs, ces modèles ne prennent pas en compte les effets exogènes ou les ruptures structurelles éventuelles dans les données.

# Conclusion

Ce mémoire, a été consacré à l'étude, l'analyse et la modélisation des séries chronologiques avec une attention particulière portée sur les séries stationnaires. La stationnarité constitue en effet une condition centrale dans de nombreuses méthodes statistiques de prévision, car elle garantit la stabilité des caractéristiques statistiques dans le temps (espérance, variance, structure de dépendance). Ainsi, une large partie de ce travail a été dédiée à la compréhension de la stationnarité, à sa vérification empirique, ainsi qu'aux techniques permettant d'y ramener les séries non stationnaires.

Nous avons d'abord exploré les séries chronologiques sous leurs différentes formes, en identifiant les composantes structurelles telles que la tendance et la saisonnalité. L'analyse des fonctions d'autocorrélation, les tests de stationnarité (ADF, KPSS, PP), et les transformations nécessaires (différenciation, décomposition) ont permis de mieux cerner la nature stationnaire ou non des données.

Les séries identifiées comme stationnaires ont été modélisées à l'aide de modèles ARMA, qui se sont révélés performants pour capturer les dépendances internes dans des séries stables. Pour les séries non stationnaires, nous avons utilisé les modèles ARIMA (après différenciation) et SARIMA pour les séries présentant à la fois une non-stationnarité et une saisonnalité marquée.

Le principal avantage des modèles stationnaires comme ARMA réside dans leur simplicité, leur interprétabilité et leur efficacité prédictive lorsque les hypothèses sont respectées. Toutefois, ces modèles présentent des limites lorsque les séries évoluent dans le temps de manière complexe ou non linéaire, ce qui nécessite alors des approches plus avancées.

Ce travail ouvre plusieurs perspectives :

- approfondir l'analyse des séries stationnaires à travers des tests plus robustes et des méthodes de détection automatique.
- Combiner les modèles linéaires stationnaires avec des techniques d'apprentissage automatique pour modéliser des phénomènes plus complexes.
- Etendre l'étude aux modèles multivariés (VAR, VECM) dans le cas de plusieurs séries interdépendantes.

En somme, l'étude des séries stationnaires n'est pas seulement un prérequis pour de nombreuses méthodes de modélisation, elle en est le fondement. Mieux comprendre et exploiter la stationnarité permet d'améliorer significativement la qualité des modèles et la précision des prévisions.

# Annexe

## 1. Les méthodes d'estimation

L'estimation d'un paramètre consiste à utiliser les informations contenues dans une série chronologique observée pour inférer une valeur approchée de ce paramètre inconnu. Lorsque cette approximation est réduite à un seul nombre, on parle d'estimation ponctuelle. Ce nombre est le résultat de l'application d'un estimateur, qui est une règle ou une fonction mathématique définie avant l'observation des données. L'estimation est la valeur numérique spécifique obtenue lorsque cette règle est appliquée à la série chronologique que nous avons réellement observée.

### a) Les Moindres Carrés Ordinaires (MCO)

La méthode des moindres carrés repose sur la réduction de la somme des carrés des différences entre les données mesurées et le modèle utilisé, dans le cadre de la régression. L'objectif est de trouver le modèle théorique  $f(x_i; \theta_i)$  qui correspond le mieux aux mesures.

Les paramètres du modèle,  $\theta_i$ , sont trouvés en réduisant une fonction appelée fonction-objectif  $S(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_i)$ . Le but est donc de réduire au minimum ces différences.

Les estimations  $\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_i\}$  représentent le minimum de la fonction-objectif. On illustre cela avec le modèle :

$$\hat{y}_i = \theta_1 x_i + \theta_2.$$

La fonction-objectif est définie comme suit dans le cas d'un modèle linéaire simple :

$$S(\theta_1, \theta_2) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \theta_1 x_i - \theta_2)^2.$$

Pour minimiser  $S(\theta_1, \theta_2)$ , on dérive cette fonction par rapport à  $\theta_1$  et  $\theta_2$ , puis on pose ces dérivées égales à zéro :

$$\begin{cases} \frac{\partial S}{\partial \theta_1} = 0, \\ \frac{\partial S}{\partial \theta_2} = 0. \end{cases}$$

En résolvant ce système, on obtient les estimateurs des moindres carrés  $\hat{\theta}_1$  et  $\hat{\theta}_2$ .

### b) La méthode du maximum de vraisemblance (MV)

La méthode du maximum de vraisemblance cherche plutôt la valeur qui a le plus de chances de correspondre aux valeurs réelles que l'on a observées. En d'autres termes, elle essaie de trouver la valeur la plus probable pour une série chronologique donnée.

Soit  $X_1, X_2, \dots, X_n$  une série chronologique indépendante et identiquement distribuée (i.i.d) suivant une loi de probabilité  $f(X; \theta)$  où  $\theta$  est un paramètre inconnu à estimer. La fonction de vraisemblance est donnée par :

$$L(\theta) = \prod_{i=1}^n f(X_i; \theta).$$

Puisque le produit est souvent complexe à manipuler, on utilise le logarithme de la vraisemblance (log-vraisemblance) :

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(X_i; \theta).$$

Cela simplifie les calculs et conserve la même valeur maximale que  $L(\theta)$  car le logarithme est une fonction croissante.

L'estimateur du maximum de vraisemblance  $\hat{\theta}$  est obtenu en résolvant :

$$\frac{d\ell(\theta)}{d\theta} = 0.$$

Cette équation donne une estimation  $\hat{\theta}$  du paramètre  $\theta$ .

On vérifie que  $\hat{\theta}$  correspond bien à un maximum en s'assurant que la dérivée seconde est négative :

$$\frac{d^2\ell(\theta)}{d\theta^2} < 0.$$

## 2. Tests Paramétriques

Les tests paramétriques des résidus ont pour objectif de vérifier si les hypothèses classiques du modèle sont respectées, notamment après estimation. Ils permettent d'analyser si les résidus se comportent comme un bruit blanc, ce qui implique qu'ils sont indépendants, de variance constante (homoscédasticité) et suivent une distribution normale. Parmi les tests les plus couramment utilisés figurent le test de Ljung-Box (autocorrélation des résidus), le test de Jarque-Bera (normalité), et les tests d'homoscédasticité comme celui de Breusch-Pagan. Ces tests sont essentiels pour valider la qualité du modèle ajusté et assurer la fiabilité des prévisions.

### a) Test du portemanteau (Box-Pierce)

Plutôt que d'examiner individuellement les autocorrélations, on considère la statistique suivante :

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

où :

- $T$  est le nombre d'observations.
- $h$  est le nombre de retards (lags).
- $\hat{\rho}(k)$  est l'autocorrélation empirique au retard  $k$ .

La région de rejet approximative au niveau  $\alpha$  est donnée par :

$$\mathcal{R}_\alpha = \left\{ T \sum_{k=1}^h \hat{\rho}^2(k) \geq q(\alpha; h) \right\}$$

où  $q(\alpha; h)$  est le quantile d'ordre  $\alpha$  de la loi  $\chi_h^2$ .

## b) Test de Ljung-Box

Amélioration du test précédent, la statistique de Ljung-Box est :

$$Q_{LB} = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}^2(k)}{T-k} \sim \chi_h^2$$

Ce test est mieux ajusté à la loi  $\chi^2$  que celui de Box-Pierce, et est implémenté dans R via la fonction `Box.test`.

## c) Test de McLeod-Li

Ce test permet de détecter une non-indépendance même si l'ACVF empirique semble indiquer un bruit blanc.

$$Q_{ML} = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}_{XX}^2(k)}{T-k} \sim \chi_h^2$$

où  $\hat{\rho}_{XX}(k)$  est l'autocorrélation des carrés ( $X_i^2$ ).

## d) Test de normalité

### • Droite de Henry (QQ-plot)

Soit  $(e_i)_{i=1}^n$  un échantillon i.i.d. ordonné. On considère :

$$\text{qq}(e) = \{(m_j, e_{(j)}), j = 1, \dots, n\}$$

avec  $m_j = \mathbb{E}(u_{(j)})$  les espérances des quantiles théoriques. Si les points s'alignent sur une droite, cela indique une normalité.

On peut formaliser par :

$$r^2 = \frac{\sum (e_i - \bar{e}) m_i}{\sqrt{\sum (e_i - \bar{e})^2 \cdot \sum m_i^2}}$$

On rejette l'hypothèse de normalité si  $r^2 < s(\alpha)$ . Par exemple :  $s(0.05) = 0.987$  pour  $n = 200$ .

### • Test de Jarque-Bera (JB)

Sous l'hypothèse de normalité, l'asymétrie  $k_3 = 0$  et l'aplatissement  $k_4 = 3$ . Soit  $S$  et  $K$  les estimateurs de  $k_3$  et  $k_4$ , alors :

$$JB = \frac{n-p}{6} \left( S^2 + \frac{1}{4} (K-3)^2 \right)$$

Sous  $H_0$ ,  $JB \sim \chi_2^2$ . La région de rejet est :

$$\mathcal{R}_\alpha = \{JB \geq q(2; \alpha)\}$$

### 3. La transformation de Fourier

La transformation de Fourier est un outil mathématique essentiel qui permet de passer d'une représentation temporelle d'un signal ou d'une série temporelle à une représentation fréquentielle.

Autrement dit, au lieu d'analyser les variations d'un signal au cours du temps, elle permet d'étudier la contribution des différentes fréquences qui composent ce signal. Elle met ainsi en évidence les composantes périodiques cachées.

Soit  $f \in L^1(\mathbb{R})$ . La **transformée de Fourier** de  $f$ , notée  $\hat{f}(x)$  ou  $\mathcal{F}(f)(x)$ , est définie par :

$$\hat{f}(x) = \int_{-\infty}^{+\infty} e^{-ixt} f(t) dt$$

- $f(t)$  : la fonction d'origine (le signal dans le temps)
- $x$  : la fréquence
- $e^{-ixt}$  : noyau harmonique complexe (combinant sinus et cosinus)

### 4. Le multiplicateur de Lagrange

Le **multiplicateur de Lagrange** est une méthode d'optimisation utilisée pour trouver les extrema (maximum ou minimum) d'une fonction  $f(x_1, x_2, \dots, x_n)$  sous une ou plusieurs contraintes  $g_i(x_1, x_2, \dots, x_n) = 0$ .

Le principe consiste à construire une fonction auxiliaire appelée *fonction de Lagrange* :

$$\mathcal{L}(x_1, \dots, x_n, \lambda) = f(x_1, \dots, x_n) - \lambda \cdot g(x_1, \dots, x_n)$$

où  $\lambda$  est le **multiplicateur de Lagrange**, représentant le taux de variation de l'optimum de  $f$  par rapport à une variation marginale de la contrainte  $g$ . Dans des tests comme KPSS, le multiplicateur de Lagrange est utilisé pour construire une statistique de test, qui vérifie si une contrainte (comme la stationnarité) est violée ou non.

# Bibliographie

- [1] Bigot, J. (2017). *Séries chronologiques*. Licence 3 MIASHS, Université de Bordeaux. Polycopié rédigé à partir des notes de cours de Frédéric Proia.
- [2] Boutahar, M. (2019). *Méthodes en séries temporelles et applications avec R*. Éditions Marketing.
- [3] Brockwell, P. J., & Davis, R. A. (1991). *Time Series : Theory and Methods* (2<sup>e</sup> éd.). Springer.
- [4] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis : Forecasting and Control* (5<sup>e</sup> éd.). Wiley.
- [5] Casin, P. (2011). *Économétrie : Méthode et application avec EViews*. Éditions Technip.
- [6] Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431.
- [7] Fortier, S. (2013). *Les modèles MA, AR et ARMA multidimensionnels*. Université de Sherbrooke.
- [8] Fuller, W. A. (1976). *Introduction to Statistical Time Series*. John Wiley & Sons.
- [9] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- [10] Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- [11] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting : Principles and Practice* (3<sup>e</sup> éd.). OTexts. <https://otexts.com/fpp3>
- [12] El Karoui, M., Lagnoux, A., & Institut de Mathématiques de Bordeaux. (2017). *Cours de séries temporelles – Chapitre 4 : Estimation, tests de validation et prévisions des processus ARMA*. Université de Bordeaux.
- [13] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3), 159–178.

- [14] Lagnoux, A. (2011). *Séries chronologiques*. ISMAG1 - MIS243Y, Université Toulouse II.
- [15] Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346.
- [16] R Core Team. (2023). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>
- [17] Roche, A. (2019). *Processus stationnaires – Modèles ARMA*. Executive Master Statistique et Big Data.
- [18] Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications : With R Examples* (4<sup>e</sup> éd.). Springer.