

الجمهورية الجزائرية الديمقراطية الشعبية

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE SAAD DAHLAB DE BLIDA

Faculté des sciences

Département de Mathématiques



**MEMOIRE DE MASTER**

**Spécialité : Mathématiques**

**Option : MODELISATION STOCHASTIQUE ET  
STATISTIQUE**

établi par :

-BETOU SOUAD

-ZEGHOUANI HAKIMA

Intitulé

**MODELISATION ET PREVISION PAR LE MODELE  
AUTOREGRESSIF MOYENNE MOBILE VECTORIEL  
(VARMA). APPLICATION AUX PRIX DU PÉTROLE ET DE  
L'OR**

Devant le jury composé de :

Omar Tami	Président	MCB.	Univ.Blida 1
Redouane Frihi	Encadreur	MCB.	Univ. Blida1
Abdelaziz Rassoul	Examineur	Pro	Univ. Blida 1

JUILLET 2023

## REMERCIEMENTS

Nous remercions Allah le tout-puissant de nous avoir donné la force et le courage pour bien mener et finir ce travail.

Tout d'abord, nous souhaitons exprimer notre gratitude envers notre directeur de mémoire, Monsieur **Redouane Frihi**, pour son encadrement précieux, ses conseils avisés et sa disponibilité tout au long de ce projet. Son expertise et son soutien ont été essentiels pour mener à bien cette recherche.

Mes remerciements s'adressent également aux membres du jury, qui nous ont fait l'honneur de bien vouloir étudier avec attention notre travail.

Nous tenons également à remercier le corps professoral du département de mathématiques, pour la qualité de l'enseignement dispensé durant ces années d'études. Leurs connaissances et leur dévouement ont grandement contribué à notre formation académique et à la réalisation de ce mémoire.

Enfin, nous tenons à exprimer notre gratitude à tous les membres de nos familles pour leur confiance permanente en nos capacités et leur soutien tout au long de nos études, en particulier nos parents, nos frères et nos sœurs.

---

## Dédicace

Je dédie cet humble travail à :

**Mon père**, mon cher père, et cher à mon cœur, mon soutien, ma force et ma fierté, à celui qui m'a tout donné dans cette vie, au plus grand homme de ma vie, merci pour tout.

**Ma mère**, ma chère mère, qui m'est chère, est mon premier soutien dans la poursuite de mes études et l'obtention d'un diplôme universitaire. Merci pour tout.

**Mes frères** : Youcef , Khireddine, Abdellah, Yasser.

**Mes soeurs** :Nessrine, Nawal, Manal.

**Les enfants de mes soeurs** :Rania, Djihene, Abderahmane, Lina .

**Mes amis proches** :Hakima, Abir, Imane, Chaimaa.

Et je dédie à toute ma famille et mes collègues, sans exception, et à tous ceux qui m'ont soutenu dans mon parcours académique.

**Enfin**, j'offre le plus grand dévouement et merci à moi-même, qui a fait face à de nombreux écueils pour arriver à ce stade.

**SOUAD BETOU**

---

## DEDICACE

Je dédie ce modeste travail : Aux personnes les plus chers à mon coeur

A ma source de tendresse, **ma très chère mère**

A mon symbole de la vie ,**mon chér père**

Que Allah vous garde!

A mon compagnon et partenaire de ma vie, mon mari

A mes chers frères et mes adorables soeurs qui sont mon support à toute heure

A toute ma famille

Je tiens aussi à dédier ce travail et remercier du fond du coeur mes amis et a l'ensemble de mes collègues sans exception

Enfin

A tous ceux qui ont participé, de près ou loin à la réalisation de ce travail

A tous ceux qui m'ont aimé et aidé durant toute la période de réalisation de ce travail.

**zeghouani hakima**



# Table des matières

- Introduction Générale** **1**
  
- 1 Modèles de Prévision Stochastique et Statistique** **4**
  - 1.1 Introduction . . . . . 4
  - 1.2 Modèles de régression linéaire . . . . . 5
  - 1.3 Modèles ARCH-GARCH . . . . . 5
  - 1.4 Méthodes de machine learning . . . . . 7
    - 1.4.1 Réseaux de neurones . . . . . 7
    - 1.4.2 Forêts aléatoires . . . . . 8
    - 1.4.3 SVM (Support Vector Machines) . . . . . 8
  - 1.5 Méthodes de simulation . . . . . 8
    - 1.5.1 Chaînes de Markov : . . . . . 9
    - 1.5.2 Processus de Monte Carlo : . . . . . 9
    - 1.5.3 Méthode de l'inversion : . . . . . 9
    - 1.5.4 Méthode de rejet : . . . . . 9
  - 1.6 Modèle ARMA . . . . . 10
  
- 2 Séries chronologiques Univariées** **11**
  - 2.1 Introduction . . . . . 11
    - 2.1.1 Exemples . . . . . 11
    - 2.1.2 Questions posées au statisticien pour l'étude d'une série chronologique . . . . . 14
  - 2.2 Modélisation déterministe . . . . . 14
    - 2.2.1 Le modèle additif : . . . . . 14
      - 2.2.1.1 Tendance . . . . . 15
      - 2.2.1.2 Composantes Périodiques, Saisonnalités . . . . . 15

2.2.1.3	Stationnarité . . . . .	16
2.2.2	Le modèle multiplicatif . . . . .	17
2.2.3	Les modèles mixtes . . . . .	18
2.2.4	Opérateurs définis sur une série chronologique : . . . . .	18
2.3	Types de modèles . . . . .	19
2.3.1	Processus autorégressif $AR(p)$ . . . . .	19
2.3.2	Processus à moyenne mobile $MA(q)$ . . . . .	20
2.3.3	Processus $ARMA(p,q)$ . . . . .	22
2.3.4	Critères d'identification de processus $ARMA$ : . . . . .	26
2.3.5	Principes d'identification : . . . . .	26
2.3.6	Identification de l'ordre . . . . .	26
2.3.6.1	Processus $ARIMA$ et $SARIMA$ : . . . . .	28
2.3.6.2	Modèles avec co-variables : $ARMAX$ . . . . .	28
2.3.7	Estimation des paramètres $\Phi$ , $\Theta$ et $\sigma_\varepsilon^2$ . . . . .	28
2.3.8	Application du modèle $ARMA$ . . . . .	29
2.4	Tests de Validation . . . . .	30
2.4.1	Tests de redondance . . . . .	30
2.4.2	Tests de significativité des coefficients . . . . .	30
2.4.3	Coefficient de détermination . . . . .	31
2.4.4	Autocorrélations simple et partielle . . . . .	31
2.4.4.1	Fonction d'autocovariance et d'autocorrélation . . . . .	31
2.4.4.2	Fonction d'autocorrélation partielle . . . . .	32
2.4.5	Autocorrelation partielle (Méthode de Walker) . . . . .	34
<b>3</b>	<b>Série chronologiques Multivariées</b> . . . . .	<b>36</b>
3.1	Introduction . . . . .	36
3.2	Modèle Autoregressif moyenne Mobile vectoriel bidimensionnel . . . . .	38
3.2.1	Définitions de causalité au sens de Granger . . . . .	38
3.2.2	Processus autorégressif bidimensionnel ( $VAR$ ) . . . . .	40
3.2.3	Processus à moyenne mobile bidimensionnel( $VMA$ ) . . . . .	42
3.2.4	Processus $ARMA$ bidimensionnel . . . . .	44
3.3	Stationnarité . . . . .	44
3.3.1	Tests de stationnarité . . . . .	45
3.3.1.1	Test Dukey Fuller augmenté . . . . .	45
3.3.1.2	Test de Johansen . . . . .	46
3.3.2	Estimation . . . . .	47
3.3.3	Sélection de l'ordre . . . . .	52

---

<b>4 APPLICATIONS</b>	<b>55</b>
4.1 Introduction . . . . .	55
4.1.1 Etude de la série . . . . .	56
4.1.2 Modélisation par VARMA . . . . .	56
4.2 Etapes de la modélisation . . . . .	57
4.3 Analyse de la série (Or, Pétrole) . . . . .	58
4.3.1 Graphe de la série du prix de Pétrole et du prix de l'Or . . . . .	58
4.3.2 Statistique descriptive de la série du prix de Pétrole et la série du prix de l'Or . . . . .	59
4.3.3 Correlation : . . . . .	59
4.3.4 Stationnarité : . . . . .	59
4.3.5 Sélection de l'ordre : . . . . .	59
4.3.6 Modèles sélectionnés . . . . .	60
4.3.7 Prévision . . . . .	61
4.4 Conclusion générale . . . . .	63

## TABLE DES FIGURES

1.1	Perceptrons multicouches . . . . .	8
2.1	Taches solaires . . . . .	12
2.2	Ventes de voitures . . . . .	12
2.3	Évolution de la consommation d'électricité en Algérie (1990-2018) . . .	13
2.4	Trafic aérien . . . . .	13
2.5	Population des États-Unis . . . . .	14
2.6	Modèle multiplicatif. Amplitude proportionnelle à la tendance . . . . .	18
2.7	Modèle $ARMA(1,1)$ . . . . .	24
2.8	Principes généraux d'identification des processus $ARMA$ . . . . .	26
2.9	Autocorrélation simple et Autocorrélation Partielle . . . . .	34
4.1	Prix de l'Or contre le prix de Pétrole . . . . .	56
4.2	Évolution mensuelle du prix de Pétrole et du prix de l'Or de janvier 1990 au décembre 2022. . . . .	58
4.3	Prévision du prix de l'Or des 5 premiers mois de 2023 . . . . .	62
4.4	Prévision du prix du Pétrole des 5 premiers mois de 2023 . . . . .	63

## LISTE DES TABLEAUX

4.1	Statistique description de la série (Pétrole, Or) . . . . .	59
4.2	Sélection de l'ordre par les critères AIC et BIC. . . . .	59
4.3	Trois premiers modèles sélectionnés par critères AIC et BIC . . . . .	60
4.4	Prévision des prix de l'Or et du prix de Pétrole de Janvier au Décembre 2023 par le modèle VARMA(1,1) . . . . .	61
4.5	Prévision des prix de l'Or et du Pétrole de Janvier au Décembre 2023 par le modèle VARMA(1,0) . . . . .	61
4.6	Prévision des prix de l'Or et du Pétrole de Janvier au Décembre 2023 par le modèle VARMA(2,0) . . . . .	61

ملخص :

يعد تحليل السلاسل الزمنية والتنبؤ بها ، مثل أسعار النفط والذهب ، أمرًا ضروريًا للعديد من المشاركين في السوق المالية. في هذا السياق ، يعد نموذج متوسط الانحدار الذاتي المتحرك للمتجه (VARMA) طريقة شائعة الاستخدام لنمذجة هذه السلاسل الزمنية والتنبؤ بها. يجمع نموذج VARMA بين مفاهيم الانحدار التلقائي (VAR) ونماذج المتوسط المتحرك (MA). إنه قادر على التقاط التبعيات الخطية بين سلاسل زمنية متعددة ، مما يجعله أداة قوية لتحليل التفاعلات بين أسعار النفط والذهب.

**Résumé :**

L'analyse et la prévision des séries temporelles, telles que les prix du pétrole et de l'or, sont essentielles pour de nombreux acteurs du marché financier. Dans ce contexte, le modèle autorégressif moyenne mobile vectoriel (VARMA) est une méthode couramment utilisée pour modéliser et prévoir ces séries temporelles.

Le modèle VARMA combine les concepts de modèles autorégressifs (VAR) et de modèles moyenne mobile (MA). Il est capable de capturer les dépendances linéaires entre plusieurs séries temporelles, ce qui en fait un outil puissant pour analyser les interactions entre les prix du pétrole et de l'or.

**Abstract :**

The analysis and forecasting of time series, such as oil and gold prices, is essential for many financial market players. In this context, the Vector Autoregressive Moving Average (VARMA) model is a commonly used method for modeling and forecasting such time series.

The VARMA model combines the concepts of autoregressive (VAR) and moving average (MA) models. It is able to capture linear dependencies between several time series, making it a powerful tool for analyzing the interactions between oil and gold prices.

## Abréviations et Notations

### Abréviations

<b>ARIMA</b>	AutoRegressive Integrated Moving Average.
<b>ARCH</b>	Autoregressive Conditional Heteroskedasticity.
<b>GARCH</b>	Autoregressive Conditional Heteroskedasticity-Generalized.
<b>MCO</b>	La méthode des moindres carrés ordinaires.
<b>IGARCH</b>	Integrated GARCH.
<b>TGARCH</b>	Threshold GARCH.
<b>ANN</b>	Les réseaux de neurones artificiels.
<b>MLP</b>	Perceptrons multicouches.
<b>CNN</b>	Les réseaux de neurones convolutifs.
<b>RNN</b>	Les réseaux de neurones récurrents.
<b>LSTM</b>	Les réseaux de neurones de mémoire à court terme.
<b>BIC</b>	Bayesian Information Criterion.
<b>AIC</b>	Akaike Information Criterion.
<b>HQIC</b>	Hannan-Quinn Information Criterion.
<b>AICC</b>	Corrected Akaike Information Criterion.
<b>CAIC</b>	Consistent Akaike Information Criterion.
<b>ARMA</b>	AutoRegressive Moving Average .
<b>VARMA</b>	Vector Autoregressive Moving Average.
<b>ADF</b>	Test de Dickey-Fuller augmenté .
<b>SVM</b>	Support Vector Machines.

### Notations

$\varepsilon$	Le vecteur d'erreurs
$\sigma^2(t)$	La variance conditionnelle au temps t.
$m_t$	La tendance.
$s_t$	La fonction de saisonnalité.
$\phi_k$	constituent les paramètres du modèle AR.
$\theta_k$	sont les paramètres du modèle MA.

## INTRODUCTION GÉNÉRALE

Une série chronologique est une suite d'observations temporelles, souvent à des intervalles réguliers. L'analyse des séries chronologiques est un outil couramment utilisé de nos jours pour la prédiction de données futures.

Ils ont de nombreuses applications dans de nombreux domaines, notamment en économie où les séries chronologiques sont souvent utilisées pour étudier l'évolution des indicateurs économiques tels que le PIB, l'inflation, le chômage, les ventes au détail, les prix des actions, etc. Ces données peuvent aider les analystes à prédire les tendances économiques et à prendre des décisions éclairées, en sciences sociales où les séries chronologiques sont souvent utilisées pour étudier l'évolution des comportements sociaux et des attitudes au fil du temps. Par exemple, les données d'enquête peuvent être utilisées pour étudier l'évolution des opinions politiques, des habitudes de consommation, des habitudes alimentaires, etc.

En météorologie où les séries chronologiques sont utilisées pour étudier l'évolution des conditions météorologiques et climatiques au fil du temps, telles que la température, les précipitations, la pression atmosphérique, etc. Ces données peuvent aider les scientifiques à comprendre les changements climatiques et à prédire les conditions météorologiques à venir.

En santé où les séries chronologiques sont souvent utilisées pour étudier l'évolution des taux de maladies, des taux de mortalité et de survie, et pour évaluer l'impact des politiques de santé publique sur ces indicateurs.

En L'ingénierie où les séries chronologiques sont souvent utilisées pour étudier l'évolution des performances des machines et des équipements, ainsi que pour surveiller les processus de production au fil du temps.

En les sciences de l'environnement où les séries chronologiques sont utilisées pour étudier l'évolution des variables environnementales telles que la qualité de l'air, la qualité de l'eau, la concentration de  $CO_2$ , etc.

Le modèle VARMA est largement utilisé pour analyser les séries temporelles qui présentent des dépendances linéaires complexes et des interactions entre plusieurs variables. Il combine les concepts de deux autres modèles populaires, à savoir le modèle autorégressif (VAR) et le modèle moyenne mobile (MA), pour capturer efficacement les dynamiques temporelles.

L'une des principales motivations de l'utilisation du modèle VARMA réside dans sa capacité à modéliser les interactions entre différentes séries temporelles. Par exemple, lors de l'analyse des prix du pétrole et de l'or, il est important de comprendre comment ces deux variables influencent mutuellement leurs évolutions. Le modèle VARMA permet d'explorer ces interdépendances et de quantifier leur impact.

Une autre motivation clé du modèle VARMA réside dans sa flexibilité. En permettant l'inclusion de termes autorégressifs et de termes de la moyenne mobile, il peut prendre en compte diverses dynamiques présentes dans les séries temporelles étudiées. Cela permet de modéliser les retards dans les variables, d'examiner les effets de court et de long terme, et de capturer les tendances et les cycles temporels.

De plus, le modèle VARMA est un outil précieux pour la prévision des séries temporelles. Une fois ajusté aux données historiques, il peut être utilisé pour effectuer des prévisions à court et à long terme, offrant ainsi une vision prospective des évolutions des variables étudiées. Il permet également d'évaluer l'incertitude associée aux prévisions, fournissant ainsi des informations sur la fiabilité des résultats.

En conclusion, le modèle VARMA est un cadre analytique puissant pour modéliser, analyser et prévoir les séries temporelles complexes telles que les prix du pétrole et de l'or. En combinant les avantages des modèles autorégressifs et moyenne mobile, il permet de capturer les interactions, les retards et les tendances, fournissant ainsi des informations précieuses pour les décisions économiques et financières.

L'objectif principal de ce mémoire est de développer un modèle VARMA robuste et efficace pour modéliser les prix du pétrole et de l'or, en prenant en compte leurs interdépendances et leurs dynamiques temporelles. Plus précisément, nous cherchons à répondre aux questions suivantes :

Comment les prix du pétrole et de l'or sont-ils liés ? Quelles sont les interactions et les dépendances entre ces deux variables ?

Est-ce que l'évolution des prix du pétrole influence celle de l'or, ou vice versa ?

Comment le modèle VARMA peut-il être utilisé pour modéliser ces interdépendances ? Comment les retards, les tendances et les cycles temporels sont-ils pris en compte dans le modèle ?

Quelles sont les spécifications optimales du modèle VARMA pour notre étude ?

Comment le modèle VARMA peut-il être utilisé pour effectuer des prévisions précises des prix du pétrole et de l'or?

Comment évaluer l'incertitude associée à ces prévisions?

Quelles sont les mesures d'évaluation appropriées pour évaluer la performance du modèle?

Pour atteindre ces objectifs, ce mémoire adoptera une approche méthodologique rigoureuse. Nous recueillerons des données historiques sur les prix du pétrole et de l'or, que nous prétraiterons pour satisfaire aux hypothèses nécessaires à l'application du modèle VARMA. Ensuite, nous ajusterons et estimerons les paramètres du modèle VARMA en utilisant des techniques statistiques appropriées.

Nous utiliserons ensuite le modèle VARMA ajusté pour effectuer des prévisions à court et à long terme des prix du pétrole et de l'or. Nous évaluerons la précision des prévisions en utilisant des mesures de performance standard et en construisant des intervalles de confiance pour évaluer l'incertitude entourant les prévisions.

En conclusion, ce mémoire de Master vise à contribuer à la littérature existante en proposant une modélisation et une prévision précises des prix du pétrole et de l'or en utilisant le modèle VARMA. Les résultats obtenus seront d'une grande valeur pour les professionnels de la finance, les décideurs économiques et les chercheurs intéressés par l'évolution des marchés du pétrole et de l'or.

Ce mémoire est répartie en quatre chapitres.

Dans le premier chapitre, nous exposons quelques modèles de prévision stochastiques et statistique. Le chapitre deux est consacré pour le modèle ARMA univarié. Par contre dans le chapitre trois, nous avons exposé le modèle autoregressif moyenne mobile vectoriel (VARMA). Dans le dernier chapitre, nous appliquons le modèle VARMA pour prédire le prix de pétrole et le le prix de l'or. Nous clôturons ce travail par une conclusion générale.

# CHAPITRE 1

## MODÈLES DE PRÉVISION STOCHASTIQUE ET STATISTIQUE

### 1.1 Introduction

Dans de nombreux domaines tels que l'économie, la finance, la météorologie, la démographie et d'autres, la capacité à prédire les valeurs futures des variables est cruciale. Les modèles stochastiques et statistiques de prévision jouent un rôle essentiel en utilisant des données historiques pour formuler des prédictions sur les évolutions futures. Ces modèles reposent sur l'idée que les variables étudiées sont influencées par des facteurs aléatoires et utilisent des techniques statistiques et probabilistes pour modéliser ces influences afin de générer des prédictions fiables. Ils captent les motifs, les tendances et les comportements récurrents dans les données historiques afin de les extrapoler dans le futur. Les séries temporelles, qui sont des ensembles de données ordonnées chronologiquement, sont souvent utilisées pour développer ces modèles. Différents modèles, tels que la régression linéaire, les modèles ARIMA, les modèles ARCH/GARCH, les modèles de séries temporelles non linéaires, les réseaux de neurones artificiels et les processus de Markov, sont disponibles pour s'adapter aux différentes caractéristiques des données et aux objectifs de prévision spécifiques. Le choix du modèle approprié dépend de facteurs tels que la nature des données, le contexte d'application, la qualité et la disponibilité des données, ainsi que les objectifs de prévision. Les performances des différents modèles doivent être évaluées et comparées à l'aide de mesures d'évaluation telles que l'erreur quadratique moyenne, l'erreur absolue moyenne et le coefficient de détermination, afin de sélectionner le modèle le plus adapté.

Dans ce chapitre, nous explorerons quelques modèles stochastiques et statistiques de prévision, en examinant leurs fondements théoriques,

## 1.2 Modèles de régression linéaire

Les modèles de régression linéaire sont largement utilisés dans les domaines de l'économie, de la finance, de la sociologie et d'autres sciences sociales pour établir une relation linéaire entre une variable à prédire (variable dépendante) et une ou plusieurs variables explicatives (variables indépendantes). Ces modèles supposent que la variable à prédire peut être prédite par une combinaison linéaire des variables explicatives, avec une certaine erreur aléatoire (voir [Greene \[2005\]](#)).

Le modèle de régression linéaire est basé sur l'équation générale :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

où  $Y$  représente la variable à prédire,  $X_i$  sont les variables explicatives,  $\beta_i$  sont les coefficients de régression correspondants,  $\beta_0$  est l'intercepte, et  $\varepsilon$  est le terme d'erreur aléatoire.

Ces modèles supposent qu'une combinaison linéaire des variables explicatives peut prédire la variable cible, avec une certaine erreur aléatoire. Les étapes clés pour construire un modèle de régression linéaire comprennent

1. La spécification du modèle.
2. L'estimation des coefficients à l'aide de méthodes d'estimation comme les moindres carrés ordinaires.
3. Le diagnostic du modèle pour évaluer s'il satisfait les hypothèses de base.
4. L'interprétation des coefficients pour comprendre la relation entre les variables.
5. L'utilisation du modèle pour effectuer des prévisions.

Cependant, il convient de noter que les modèles de régression linéaire présentent également certaines limitations. Ils supposent une relation linéaire constante entre les variables, ce qui peut ne pas être réaliste dans de nombreux cas réels.

## 1.3 Modèles ARCH-GARCH

Le modèle ARCH-GARCH (Autoregressive Conditional Heteroskedasticity-Generalized Autoregressive Conditional Heteroskedasticity) est un modèle économétrique utilisé pour analyser et prévoir la volatilité des séries temporelles financières, telles que les rendements des actifs ou les prix des actions.

Le modèle ARCH-GARCH est une extension du modèle ARCH, qui a été proposé par (voir e.g., ([Brooks \[2014b\]](#), [Alexander \[2001\]](#), [Francq \[2010\]](#)) [Engle \[1982\]](#)) Le modèle ARCH est basé sur l'idée que la variance d'une série temporelle peut varier au fil du temps et dépendre des erreurs de prédiction précédentes. Il introduit une

composante de volatilité conditionnelle qui est une fonction des résidus au carré du modèle de régression.

Le modèle GARCH est une généralisation du modèle ARCH qui permet de capturer des caractéristiques plus complexes de la volatilité conditionnelle. En plus de la composante ARCH, le modèle GARCH inclut également une composante GARCH qui dépend également des résidus de volatilité conditionnelle passée.

Le modèle ARCH-GARCH est spécifié par des équations qui décrivent l'évolution de la volatilité conditionnelle au fil du temps. Il est estimé à l'aide de techniques d'estimation statistique, telles que la méthode des moindres carrés ordinaires ou la méthode de maximisation de la vraisemblance.

Le modèle ARCH-GARCH est couramment utilisé en finance pour modéliser et prédire la volatilité des actifs financiers. Il permet aux analystes financiers et aux traders de prendre des décisions éclairées en matière de gestion des risques, de construction de portefeuille et de valorisation d'options, en tenant compte de la volatilité changeante du marché.

L'expression générale du modèle ARCH-GARCH peut être formulée comme suit :

ARCH(p) : Un processus  $(X_t)_t$  satisfait une représentation  $ARCH(p)$  si

$$X_t = \epsilon_t h_t$$

$$h^2(t) = \alpha_0 + \alpha_1 h_{t-1}^2 + \alpha_2 h_{t-2}^2 + \dots + \alpha_p h_{t-p}^2$$

et  $\epsilon_t \sim \mathcal{N}(0, 1)$  De manière équivalente, la variance  $\sigma^2(t)$  de  $X$

$$\sigma^2(t) = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \sigma_{t-2}^2 + \dots + \alpha_p \sigma_{t-p}^2$$

GARCH(p,q) : Un processus  $(X_t)_t$  satisfait une représentation  $GARCH(p, q)$  si

$$h^2(t) = \alpha_0 + \alpha_1 h_{t-1}^2 + \alpha_2 h_{t-2}^2 + \dots + \alpha_p h_{t-p}^2 + \beta_0 + \beta_1 X_{t-1}^2 + \beta_2 X_{t-2}^2 + \dots + \beta_q X_{t-q}^2$$

De même

$$\sigma^2(t) = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \sigma_{t-2}^2 + \dots + \alpha_p \sigma_{t-p}^2 + \beta_0 + \beta_1 X_{t-1}^2 + \beta_2 X_{t-2}^2 + \dots + \beta_q X_{t-q}^2$$

$\sigma^2(t)$  est la variance conditionnelle au temps  $t$ .  $\alpha_0, \alpha_1, \dots, \alpha_p$  sont les paramètres de l'ARCH qui mesurent l'impact des erreurs de prédiction précédentes sur la volatilité conditionnelle.  $\beta_0, \beta_1, \dots, \beta_q$  sont les paramètres du GARCH qui mesurent l'impact des résidus de volatilité conditionnelle passée sur la volatilité conditionnelle.  $\epsilon^2(t-1), \epsilon^2(t-2), \dots, \epsilon^2(t-p)$  sont les résidus au carré de la volatilité conditionnelle au temps  $t-1, t-2, \dots, t-p$ .  $\sigma^2(t-1), \sigma^2(t-2), \dots, \sigma^2(t-q)$  sont les volatilités conditionnelles au carré passées au temps  $t-1, t-2, \dots, t-q$ . La somme des termes ARCH représente la

dépendance de la volatilité conditionnelle par rapport aux erreurs de prédiction passées, tandis que la somme des termes GARCH représente la dépendance de la volatilité conditionnelle par rapport aux résidus de volatilité conditionnelle passée.

Il convient de noter que les valeurs de  $p$  et  $q$  peuvent varier selon le modèle spécifique que vous utilisez. Différentes variantes du modèle ARCH-GARCH existent, telles que l'IGARCH (Integrated GARCH), le TGARCH (Threshold GARCH), etc, qui introduisent des spécifications supplémentaires pour capturer des caractéristiques spécifiques de la volatilité conditionnelle.

## 1.4 Méthodes de machine learning

Les techniques de machine learning, comme les réseaux de neurones, les forêts aléatoires, les SVM (Support Vector Machines), etc. , peuvent également être utilisées pour effectuer des prévisions stochastiques et statistiques. Ces modèles apprennent à partir des données historiques et cherchent à identifier les schémas et les relations cachées pour prédire les résultats futurs (voir e.g., (Hastie [2009], Bishop [2006], Goodfellow [2016], Murphy [2012])).

### 1.4.1 Réseaux de neurones

Les réseaux de neurones sont des modèles d'apprentissage automatique qui s'inspirent du fonctionnement du cerveau humain. Ils sont composés de couches de neurones interconnectés qui traitent les informations en les passant à travers des fonctions d'activation. Les réseaux de neurones peuvent être utilisés pour des tâches de classification, de régression et de prédiction en analysant les relations complexes entre les variables d'entrée et de sortie. Ils sont entraînés en ajustant les poids des connexions entre les neurones à l'aide d'algorithmes d'optimisation tels que la rétropropagation du gradient. Les architectures populaires de réseaux de neurones comprennent les réseaux de neurones multicouches (Multi-Layer Perceptron, MLP), les réseaux de neurones convolutifs (Convolutional Neural Networks, CNN) pour la vision par ordinateur, et les réseaux de neurones récurrents (Recurrent Neural Networks, RNN) pour les données séquentielles (voir e.g., (Goodfellow [2016], Bishop [2006], Haykin [2016])).

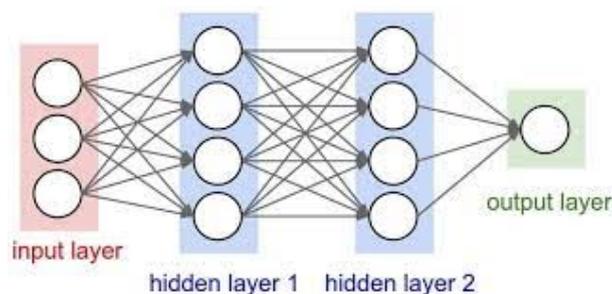


FIGURE 1.1 – Perceptrons multicouches

## 1.4.2 Forêts aléatoires

Les forêts aléatoires sont des modèles d'apprentissage automatique basés sur l'agrégation de plusieurs arbres de décision. Chaque arbre de décision est construit en utilisant un sous-ensemble aléatoire des données d'entraînement et un sous-ensemble aléatoire de variables d'entrée. Cela permet d'introduire de l'aléa et de réduire le surajustement (overfitting). Les forêts aléatoires peuvent être utilisées pour des tâches de classification et de régression. Elles fournissent également des mesures d'importance des variables, qui permettent de déterminer l'influence de chaque variable sur la prédiction. Lors de la prédiction, chaque arbre de décision de la forêt donne une prédiction et le résultat final est obtenu en agrégeant les prédictions de tous les arbres (voir e.g., (Cutler [2007], Strobl [2007], Pal [2008])).

## 1.4.3 SVM (Support Vector Machines)

Les SVM sont des modèles d'apprentissage automatique utilisés pour la classification et la régression. Ils sont basés sur le concept de trouver un hyperplan optimal qui sépare les données en différentes classes ou qui représente la meilleure régression. Les SVM cherchent à maximiser la marge entre les points de données les plus proches de chaque classe. Ces points de données sont appelés vecteurs de support (support vectors). Les SVM peuvent utiliser différentes fonctions noyau (kernel) pour transformer les données d'entrée dans un espace de dimension supérieure, permettant ainsi de résoudre des problèmes non linéaires. Les SVM sont des modèles à apprentissage supervisé et nécessitent l'étiquetage approprié des données d'entraînement (voir e.g., (Wang [(2022)], Chaudhuri [2022], Yan [2022])).

## 1.5 Méthodes de simulation

Les méthodes de simulation sont utilisées pour modéliser des scénarios futurs en générant des échantillons aléatoires à partir de distributions de probabilité spécifiées.

Voici quelques méthodes de simulation couramment utilisées avec leurs formules associées : (voir e.g., (Asmussen [(2007)], Gelman [2013], Fishman [2013]))

### 1.5.1 Chaînes de Markov :

Les chaînes de Markov sont utilisées pour modéliser des processus aléatoires qui évoluent au fil du temps. Elles sont caractérisées par une propriété appelée "propriété de Markov" qui stipule que l'état futur dépend uniquement de l'état présent, et non des états antérieurs. La probabilité de transition d'un état à l'autre est définie par une matrice de transition. Formule de transition de la chaîne de Markov :

$$P(X_{n+1} = j | X_n = i) = p_{ij}$$

### 1.5.2 Processus de Monte Carlo :

Les méthodes de Monte Carlo utilisent des échantillons aléatoires pour estimer des résultats probables. Ces méthodes sont particulièrement utiles lorsque les calculs exacts sont difficiles ou impossibles. La méthode de Monte Carlo peut être utilisée pour simuler des variables aléatoires à partir de leurs distributions de probabilité et pour générer des scénarios futurs. Formule de Monte Carlo pour estimer l'espérance d'une variable aléatoire :

$$E(X) = \frac{1}{N} \sum_{i=1}^N X_i$$

### 1.5.3 Méthode de l'inversion :

La méthode de l'inversion est utilisée pour générer des échantillons aléatoires à partir d'une distribution de probabilité spécifiée en utilisant la fonction de répartition inverse. Elle est souvent utilisée lorsque la fonction de répartition cumulative (*CDF*) est connue, mais la fonction de densité de probabilité (*PDF*) ne l'est pas. Formule de l'inversion pour générer des échantillons aléatoires à partir de la *CDF* :

$$X = F^{-1}(U)$$

### 1.5.4 Méthode de rejet :

il est important de noter que leur application pratique peut varier en fonction du contexte spécifique et des distributions de probabilité utilisées. La méthode de rejet est utilisée pour générer des échantillons aléatoires à partir d'une distribution de probabilité en utilisant une distribution plus simple à générer. Elle repose sur le rejet des

échantillons générés qui ne satisfont pas certaines conditions prédéfinies.

La Formule de la méthode de rejet :

1. Générer un échantillon aléatoire  $X$  à partir d'une distribution simple.
2. Générer un autre échantillon aléatoire  $U$  à partir d'une distribution uniforme.
3. Si  $U$  est inférieur à la densité de probabilité de  $X$ , accepter  $X$ ; sinon, le rejeter et revenir à l'étape 1.

Ces formules fournissent un aperçu des méthodes de simulation.

## 1.6 Modèle ARMA

Le modèle de prévision ARMA (AutoRegressive Moving Average) est un modèle statistique utilisé pour modéliser et prévoir les séries chronologiques. Il combine à la fois des termes autorégress (AR) et des termes de moyenne mobile (MA) (voir e.g., [Brockwell \[2016\]](#), [Tsay \[\(2013\)\]](#), [Hamilton \[\(2021\)\]](#)).

Ainsi, un modèle ARMA combine ces deux composantes dans une équation pour modéliser la série temporelle. L'équation générale d'un modèle ARMA est donnée par :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

où :

- $X_t$  est la valeur de la série temporelle à l'instant  $t$ .
- $c$  est une constante.
- $\phi_1, \phi_2, \dots, \phi_p$  sont les coefficients autorégressifs correspondant aux  $p$  termes AR.
- $\varepsilon_t$  est l'erreur à l'instant  $t$ .
- $\theta_1, \theta_2, \dots, \theta_q$  sont les coefficients de la moyenne mobile correspondant aux  $q$  termes MA.

**Ce modèle sera examiné en détail dans le prochain chapitre.**

## CHAPITRE 2

# SÉRIES CHRONOLOGIQUES UNIVARIÉES

### 2.1 Introduction

Une série chronologique est une suite d'observations temporelles, souvent à des intervalles réguliers. L'analyse des séries chronologiques est un outil couramment utilisé de nos jours pour la prédiction de données futures.

Elles ont de nombreuses applications dans de nombreux domaines, notamment :

en finance, en médecine, en économétrie et en météorologie et dans bien d'autres domaines. Par exemple, en finance, on s'intéresse à modéliser le taux de change d'une devise. En météorologie, on modélise par exemple la température, les précipitations, vitesse de vent.

**Définition 2.1** Une série temporelle est une suite  $(X_1, \dots, X_n)$  des variables aléatoires représentant l'évolution d'une quantité spécifique au cours du temps. L'indice temps  $t$  peut être selon les cas la minute, l'heure, le jour ou l'année ...

Le nombre  $n$  est appelé la longueur de la série.

**Remarque 2.1** Dans la plupart du temps il est bien utile de représenter la série temporelle sur un graphe construit de la manière suivante :

en abscisse le temps, en ordonnée la valeur de l'observation à chaque instant. Le graphe apparaît donc comme une ligne brisée.

Dans la pratique, une réalisation sur une durée finie d'un tel processus est également appelée série chronologique par abus de langage; cette réalisation correspond à l'observation à des instants discrets d'une quantité variable, réelle ou complexe.

#### 2.1.1 Exemples

**Exemple 2.1** : Les taches solaires. Il s'agit du nombre annuel de taches observées à la surface du soleil entre les années 1700 et 1980. Il y a une donnée par année, soit donc 281 points

reliés par des segments. C'est une des séries temporelles les plus célèbres. On y distingue la superposition de deux phénomènes périodiques. Pendant les périodes de forte activité solaire, le nombre de taches solaires peut atteindre des centaines par jour. Pendant les périodes de faible activité solaire, il peut y avoir plusieurs jours sans aucune tache solaire observable. En moyenne, sur une période de 11 ans, on peut observer plusieurs milliers de taches solaires

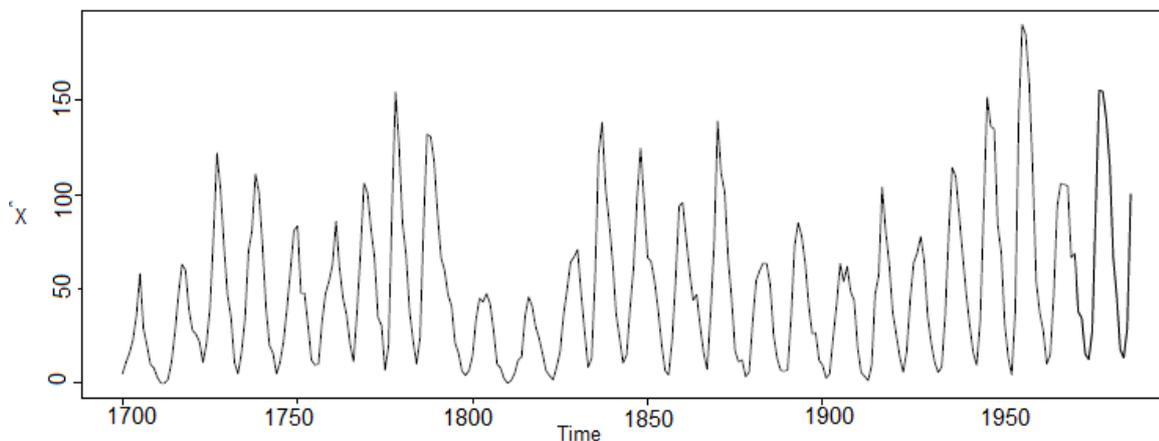


FIGURE 2.1 – Taches solaires

**Exemple 2.2 :** Les ventes de voitures. Les données courent sur 35 ans, et il y en a une par mois. Soit 420 points. On y voit une croissance non linéaire (peut être de type logarithmique). Les deux saisonnalités d'un an et de 6 mois se voient moins que dans la série précédente à cause du plus grand resserrement des abscisses. On voit aussi une rupture très nette aux environs de l'année 1983. Cette rupture fait que la série est extrêmement difficile à modéliser et donne de très mauvaises prévisions quelle que soit la méthode utilisée.

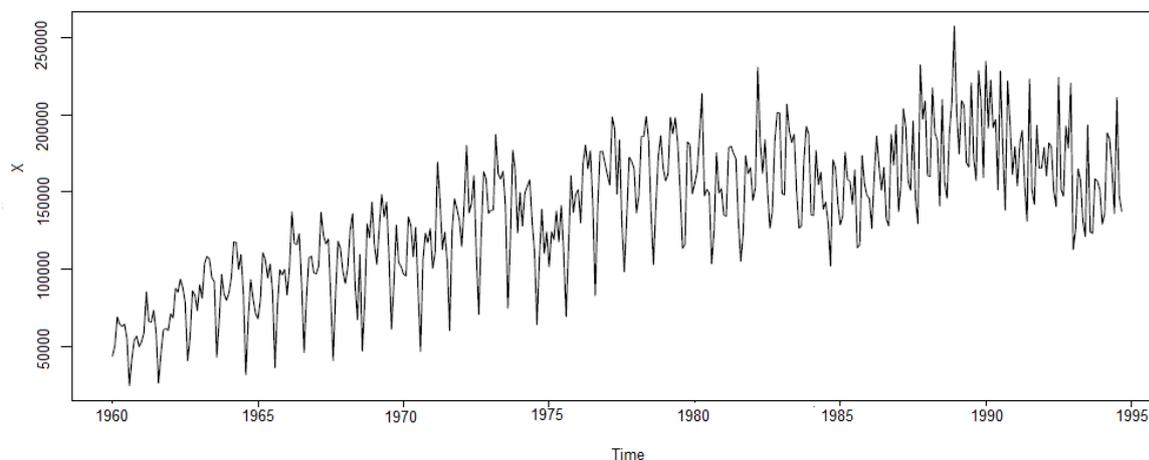


FIGURE 2.2 – Ventes de voitures

**Exemple 2.3** : Le graphe représente l'évolution de la consommation d'électricité en Algérie pendant toute la période (1990 – 2018) et surtout à partir de l'année 2010 où la demande d'électricité 60500 Gwh pour atteindre 140000 Gwh en 2018 soit un taux de croissance moyen de 56% , cet accroissement de la demande d'électricité est dû au développement économique et social du pays.

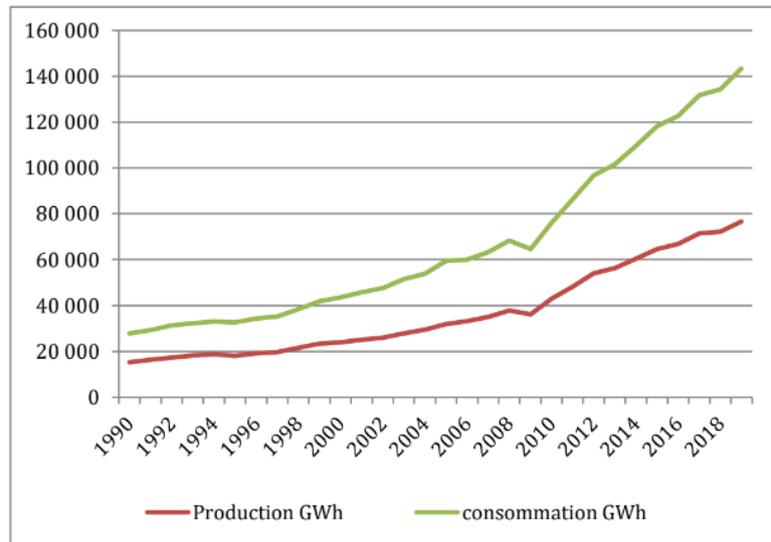


FIGURE 2.3 – Évolution de la consommation d'électricité en Algérie (1990-2018)

**Exemple 2.4** : Le nombre de passagers (en milliers) dans les transports aériens. Une donnée par mois de 1949 à 1960. Tendence linéaire (ou plus) très marquée, ainsi qu'une forte augmentation de la variabilité et une composante périodique correspondant à l'année.

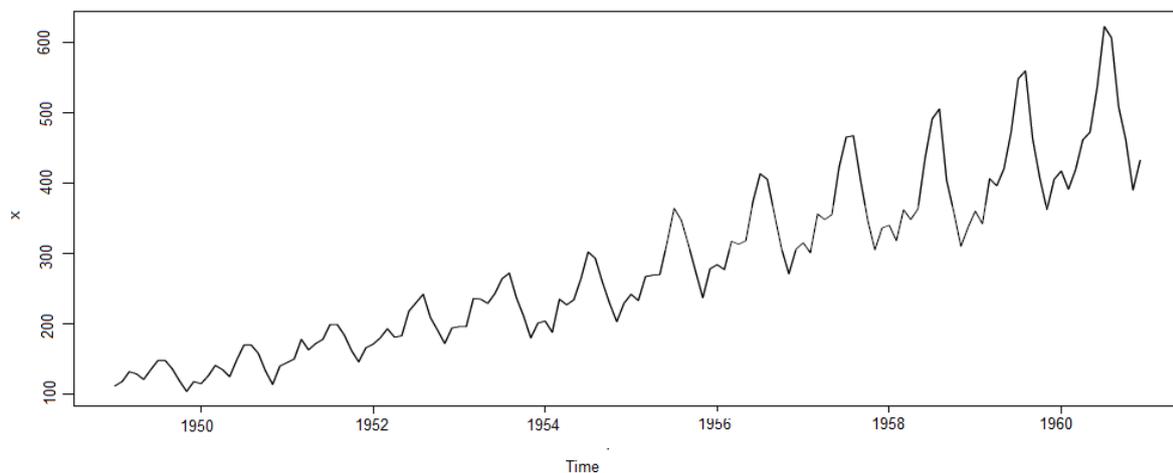


FIGURE 2.4 – Trafic aérien

**Exemple 2.5** : La population des États-Unis (une donnée tous les 10 ans, de 1790 à 1980, soit 21 points). Le seul phénomène évident est la croissance qui semble non-linéaire.

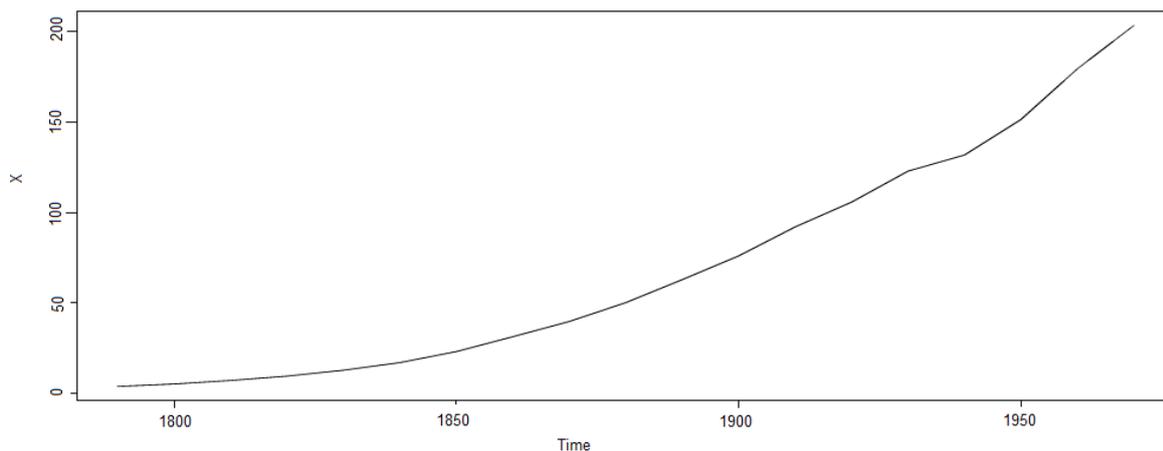


FIGURE 2.5 – Population des États-Unis

### 2.1.2 Questions posées au statisticien pour l'étude d'une série chronologique

Le statisticien est d'abord confronté au problème de la modélisation de la série temporelle. Il souhaite trouver et ajuster un modèle qui décrit ou représente au mieux les données. Il doit estimer les paramètres, faire les tests d'adéquation pour le modèle choisi.

En particulier, le statisticien cherche à déceler la présence d'une tendance, d'une saisonnalité dans les données .

Le statisticien a aussi une mission qui est la prévision des valeurs futures de la série.

## 2.2 Modélisation déterministe

Dans la littérature des modèles stochastiques des séries chronologiques, nous trouvons deux principaux modèles, le modèle additif et le modèle multiplicatif ( Voir e.g., [Robert H. Shumway \[2015\]](#), [Douglas C. Montgomery \[2015\]](#))

### 2.2.1 Le modèle additif :

Nous considérons dans cette section une série  $X = (X_t)_t$  admettant une décomposition additive

$$X_t = m_t + S_t + \varepsilon_t, \quad t = 1 \dots T,$$

où

- $t$  est l'indice du temps, à valeurs dans  $T$  sous ensemble de  $\mathbb{N}$  ou  $\mathbb{Z}$ ,
- $m_t$  est une fonction déterministe à variation appelée tendance, qui décrit les variations de niveau et variations à long terme,
- $S_t$  est une fonction déterministe périodique appelée saisonnalité de période  $r$  telle que :  

$$\sum_{i=1}^r s_{t+i} = 0, \forall t \in T$$
- $\varepsilon_t$  est un bruit aléatoire stationnaire. On l'appelle parfois le résidu.

### 2.2.1.1 Tendances

Les fonctions des tendances  $m_t$  ont les les formes possibles suivantes :

- Tendances linéaire :  $m_t = a_0 + a_1 t$ ;
- Tendances polynômiale :  $m_t = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n$ ;
- Tendances logarithmiques, exponentielles :  $m_t = \exp(a_0 + a_1^t)$ ;
- Tendances quadratique :  $m_t = a_0 + a_1 t + a_2 t^2$ ;
- Tendances Logistique :  $m_t = 1/a_0 + a_1 a^t$ ;

### 2.2.1.2 Composantes Périodiques, Saisonnalités

La composante saisonnière exprime un phénomène qui se reproduit de manière analogue sur chaque intervalle de temps successif :

$$S_{t+P} = S_t, \forall t.$$

Où  $P$  est l'étendue de cet intervalle.

Le modèle le plus simple dans lequel le saisonnier est caractérisé par  $P$  coefficients est :  $a_1, \dots, a_P$ .

On dit que la série est trimestrielle lorsque  $P = 4$ , mensuelle l'orsque  $P = 12$  .

On suppose que l'effet du saisonnier est en moyenne nul sur une période, ce qui signifie que :

$$\sum_{i=1}^P a_i = 0$$

Les erreurs sont des v.a centrées. On considère le plus souvent un bruit blanc, c'est-à-dire une suite de v.a.r. telles que :

$$E(\varepsilon_t) = 0 \quad \text{et} \quad E(\varepsilon_t \varepsilon_{t'}) = \sigma^2$$

Les v.a.r. sont alors non corrélées et lorsque le bruit blanc est gaussien c'est-à-dire que  $X_t \sim \mathbb{N}(0, \sigma^2)$  on a de plus l'indépendance des  $\varepsilon_t$ .

### 2.2.1.3 Stationnarité

La stationnarité d'une série chronologique fait référence à une propriété importante de cette série, où les caractéristiques statistiques de la série ne changent pas au fil du temps. En d'autres termes, la série chronologique présente des propriétés constantes ou invariantes dans le temps.

Une série chronologique peut être considérée comme stationnaire si elle satisfait les trois conditions suivantes :

1. Stationnarité au sens strict : La distribution conjointe des observations reste constante indépendamment du temps.

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n) = P(X_{t_{\sigma(1)}} \leq x_1, X_{t_{\sigma(2)}} \leq x_2, \dots, X_{t_{\sigma(n)}} \leq x_n)$$

2. Stationnarité en termes de moyenne : La moyenne des observations reste constante au fil du temps.

$$E(X_t) = \mu$$

où  $E(X_t)$  représente l'espérance mathématique de la variable aléatoire  $X_t$  à l'instant  $t$ , et  $\mu$  est une constante.

3. Stationnarité en termes de covariance (stationnarité en autocovariance) : L'autocovariance entre deux observations à des instants de temps différents ne dépend que de la différence entre ces instants de temps et non de leur position absolue dans le temps.

$$Cov(X_t, X_{t+h}) = Cov(X_{t+k}, X_{t+k+h})$$

pour tous les instants de temps  $t$  et pour tous les décalages  $h$  et  $k$ .

La stationnarité d'une série chronologique est importante pour plusieurs raisons :

La modélisation et la prévision de séries stationnaires sont souvent plus simples et plus fiables que celles des séries non stationnaires.

De nombreux modèles statistiques et techniques de prévision supposent la stationnarité des données.

La stationnarité d'une série chronologique permet de détecter et de quantifier les relations entre les variables, ainsi que d'effectuer des tests d'hypothèse statistiques précis. Différentes méthodes existent pour tester la stationnarité, telles que le test de Dickey-Fuller augmenté (ADF) (Voir [Lütkepohl \[2006\]](#)), le test de KPSS (Kwiatkowski-Phillips-Schmidt-Shin) et le test de Phillips-Perron (Voir [Kwiatkowski \[1992\]](#)). Ces tests utilisent des statistiques de test et des p-values pour déterminer si une série est stationnaire ou non. Lorsqu'on effectue le test de Dickey-Fuller augmenté (ADF), la p-value mesure la probabilité d'observer les résultats du test (ou des résultats encore plus extrêmes) si l'hypothèse nulle est vraie. Voici les étapes principales du test de Dickey-Fuller augmenté :

$H_0$  : La série temporelle possède une racine unitaire (Non stationnaire).

$H_1$  : La série temporelle ne possède pas une racine unitaire (stationnaire).

L'interprétation du p-value dans le contexte du test ADF est la suivante : Si le p-value est inférieur à un seuil de significative prédéfini (généralement 0,05 ou 0,01), cela signifie que la probabilité d'obtenir les résultats observés sous l'hypothèse nulle est très faible. Dans ce cas, on rejette l'hypothèse nulle et on conclut que la série temporelle est stationnaire.

Si le p-value est supérieur au seuil de significative, cela signifie que la probabilité d'obtenir les résultats observés sous l'hypothèse nulle est relativement élevée. Dans ce cas, on ne peut pas rejeter l'hypothèse nulle et on conclut que la série temporelle est non stationnaire.

### 2.2.2 Le modèle multiplicatif

Nous considérons dans cette section une série  $X = (X_t)_t$  admettant une décomposition multiplicative

$$X_t = m_t(1 + S_t)(1 + \varepsilon_t), \quad t = 1 \dots T,$$

où  $m_t$  est la composante tendancielle,  $S_t$  la composante saisonnière et  $\varepsilon_t$  représente l'erreur ou l'écart du modèle.

**Remarque 2.2** *L'amplitude de la série n'est plus constante au cours du temps : elle varie au cours du temps proportionnellement à la tendance  $m_t$  au bruit près. Dans ce modèle, on considère que les amplitudes des fluctuations dépendent du niveau*

Considérons par exemple le nombre d'entrées quotidiennes à la bouche de métro de la belle vue. Les valeurs  $S_4 = -0.5$  et  $S_6 = +0.8$  signifient ici que la fréquentation diminue de 50% le jeudi et augmente de 80% le samedi (par rapport à l'ensemble de la semaine).

Une valeur  $\varepsilon_9 = +0.2$  signifie que le nombre d'entrées du deuxième mardi a été de 20% supérieur au chiffre attendu pour ce jour là.

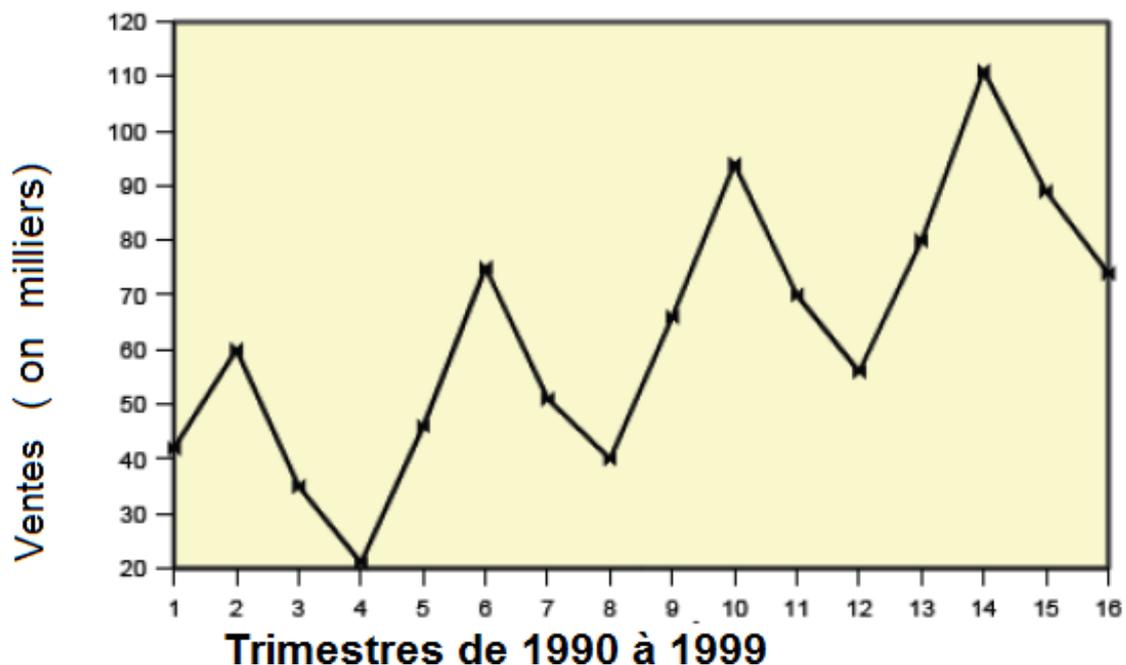


FIGURE 2.6 – Modèle multiplicatif. Amplitude proportionnelle à la tendance

**Remarque 2.3** *Le modèle multiplicatif est généralement utilisé pour des données de type économique.*

### 2.2.3 Les modèles mixtes

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

$$X_t = m_t S_t + \varepsilon_t, \quad t = 1 \dots T,$$

### 2.2.4 Opérateurs définis sur une série chronologique :

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

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

Si on compose  $B$  avec lui-même on obtient  $B^2 = B \circ B$  tel que :

$$\forall t \in \mathbb{Z}, \quad B^2 X_t = X_{t-2}.$$

On peut itérer cette application et définir par récurrence.

$$B^k X_t = X_{t-k}, \quad t \in \mathbb{N},$$

Par convention,  $B^0$  est l'opérateur identité I.

L'opérateur B est linéaire et inversible. Son inverse  $B^{-1} = F$  est défini par :

$$\forall t \in \mathbb{Z}, \quad F X_t = X_{t+1}.$$

L'opérateur F est appelé **opérateur avance**.

**Définition 2.2** :(Bruit blanc)

On dira que le processus  $\varepsilon_t$  est un "bruit blanc" s'il forme une suite de variables indépendantes et identiquement distribuées. Il sera dit centré si  $E[\varepsilon_t] = 0$  et réduit si  $Var[\varepsilon_t] = 1$ .

**Proposition 2.1** : Soit  $X_t$  un bruit blanc vérifiant  $E[X_t] = m$  et  $Var[X_t] = \sigma^2$ . Alors  $X_t$  est un processus stationnaire, telle que : l'espérance du processus et la fonction d'autocovariance sont indépendantes du temps. Mais elle implique en outre que les termes de la covariance (pour  $h \geq 0$ ) sont tous nuls. Seule la variance est non nulle. .

## 2.3 Types de modèles

Dans cette section, nous considérons certains modèles fréquemment utilisés pour une série chronologique  $\{X_t\}$ . Nous commençons par deux modèles simples :

Le modèle autorégressif AR et le modèle à moyenne mobile MA. On continue avec le modèle ARMA qui combine ces deux modèles (voir e.g., ([Robert H. Shumway \[2021\]](#), de [Paul S.P. Cowpertwait \[2019\]](#))).

### 2.3.1 Processus autorégressif AR(p)

Un processus autorégressif est un processus où l'on écrit une observation au temps  $t$  comme une combinaison linéaire des observations passées plus un certain bruit blanc.

La suite  $\{X_t : t \geq 0\}$  est un processus autorégressif d'ordre  $p$  ( $p > 0$ ) s'il peut s'écrire sous la forme suivante :

$$X_t = \sum_{k=1}^p \phi_k X_{t-k} + \varepsilon_t, \quad \text{où } \varepsilon_t \sim \mathbb{N}(0, \sigma_\varepsilon^2)$$

Les  $\phi_k (k = 1, 2, \dots, p)$  constituent les paramètres du modèle.

Dans ce cas, on note  $\{X_t\} \sim AR(p)$ , De la même façon, on peut réécrire un processus  $AR(p)$  avec un polynôme  $\phi(B)$  qui multipliera  $X_t$  cette fois-ci :

$$\phi(B)X_t = \varepsilon_t \quad \text{avec} \quad \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

**Exemple 2.6** Un processus  $AR(1)$  prend la forme suivante :

$$X_t = \phi X_{t-1} + \varepsilon_t \quad \text{où} \quad \varepsilon_t \sim \mathbb{N}(0, \sigma^2)$$

### 2.3.2 Processus à moyenne mobile $MA(q)$

Dans son article ( voir [Slutzky \[1927\]](#)) a introduit pour la première fois les processus à moyenne mobile. La définition suivante présente ce processus.

**Définition 2.3** On dit que la suite  $\{X_t : t \geq 0\}$  est un processus à moyenne mobile d'ordre  $q$  ( $q > 0$ ) si celui-ci peut s'écrire sous la forme suivante :

$$X_t = \sum_{k=1}^q \theta_k \varepsilon_{t-k} + \varepsilon_t, \quad \text{où} \quad \varepsilon_t \sim \mathbb{N}(0, \sigma^2)$$

où les  $\theta_k (k = 1, 2, \dots, q)$  sont les paramètres du modèle.

Dans ce cas, on note  $\{X_t\} \sim MA(q)$ .

**Exemple 2.7** Un processus  $MA(1)$  prend la forme suivante :

$$X_t = \theta \varepsilon_{t-1} + \varepsilon_t \quad \text{où} \quad \varepsilon_t \sim \mathbb{N}(0, \sigma_\varepsilon^2).$$

On peut utiliser l'opérateur de retard  $B$  pour écrire ce processus sous une autre forme.

On aura donc un polynôme en  $B$  qui multipliera  $\varepsilon_t$  :  $X_t = \theta(B)\varepsilon_t$

avec

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

**Remarque 2.4** : Si  $\{X_t\} \sim MA(q)$ , alors :

$$\begin{cases} X_t \text{ est stationnaire} \\ \rho_X(h) = 0 \text{ pour tout } h > q. \end{cases}$$

**Définition 2.4** : Un processus est dit causal s'il existe une suite  $\{a_k\}$  réelle telle que :

$$\sum_{k=0}^{\infty} a_k < \infty$$

et que

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$$

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

**Remarque 2.5** Tout processus MA( $q$ ) est causal.

**Définition 2.5** Un processus est dit inversible s'il existe une suite  $\{b_k\}$  réelle telle que

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

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

**Remarque 2.6** : Avec cette Définition, tout processus AR( $p$ ) est inversible.

**Théorème 2.1** : Un processus autoregressif AR( $p$ ) est causal et stationnaire si et seulement si son polynôme  $\phi(z)$  avec  $z \in \mathbb{C}$  tel que  $|z| \leq 1$ .

En d'autres mots, toutes les racines de  $\phi(z)$  sont de norme plus grande que 1.

**Exemple 2.8** : Le processus AR(2) suivant est stationnaire et causal :

$$X_t = -\frac{1}{6}X_{t-1} + \frac{1}{6}X_{t-2} + \varepsilon_t, \quad \text{où} \quad \{\varepsilon_t\} \sim \mathbf{N}(0, \sigma_\varepsilon^2).$$

En effet, on peut le réécrire comme :

$$\left(\frac{-1}{6}B^2 + \frac{1}{6}B + 1\right)X_t = \varepsilon_t.$$

Ainsi, on a que :

$$\phi(z) = \frac{1}{6}z^2 + \frac{1}{6}z + 1 = \frac{1}{6}(z^2 - z - 6) = \frac{1}{6}(z+2)(z-3).$$

Les racines sont à l'extérieur du cercle unité, donc le processus est stationnaire et causal (prévisible).

### 2.3.3 Processus ARMA (p,q)

Les modèles ARMA permettent de représenter un grand nombre de processus aléatoires stationnaires.

Il est supposé à ce niveau que l'on a isolé, dans la série chronologique d'origine, sa partie aléatoire, en lui retirant tendances et facteurs saisonniers .

On ne décrit pas ici les techniques liées aux modèles ARMAX, utilisées en particulier dans le domaine de l'automatique, où la séquence d'entrée notée  $\varepsilon$  n'est pas en général un bruit blanc. , où  $q$  est l'ordre de processus MA ,  $p$  est l'ordre de processus AR .

**Définition 2.6 :** Le processus  $X_t$ , ou  $t \in Z$ , est un ARMA(p,q) de moyenne nulle si

$$\begin{cases} (X_t) \text{ est stationnaire,} \\ X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \end{cases}$$

où  $\varepsilon_t$  est un bruit blanc faible, c'est-à-dire ici un processus aléatoire stationnaire de moyenne nulle avec  $\gamma_\varepsilon(0) = \sigma^2$  et  $\gamma_\varepsilon(h) = 0$  pour tout  $h \neq 0$ .

Une représentation équivalente et plus concise est la suivante :

$$\Phi(B)X_t = \Theta(B)\varepsilon_t,$$

où  $\Phi$  et  $\Theta$  sont des polynômes de degrés respectifs  $p$  et  $q$  :

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

$$\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

et  $B$  est l'opérateur "retard" défini par :

$$B^k X_t = X_{t-k}$$

L'équation qui lie  $X$  à  $\varepsilon$  est linéaire et à coefficients constants.

Ce modèle est donc linéaire au sens où, si

$$\varepsilon_t = \varepsilon_{1,t} + \varepsilon_{2,t},$$

alors

$$X_t = X_{1,t} + X_{2,t} \quad \text{pour tout } t ,$$

avec

$$\begin{cases} \Phi(B)X_{1,t} = \Theta(B)\varepsilon_{1,t} \\ \Phi(B)X_{2,t} = \Theta(B)\varepsilon_{2,t}. \end{cases}$$

Le processus  $X_t$  est un ARMA( $p, q$ ) de moyenne  $m$  si le processus  $X_{t-m}$  est un ARMA( $p, q$ ) de moyenne nulle.

Les processus AR et MA sont des ARMA particuliers.

**Définition 2.7 :** Un processus stochastique  $(X_t)_{t \in \mathbb{Z}}$  est dit un mélange autorégressif se déplaçant modèle moyen d'ordre  $p$  et  $q$ , ARMA( $p, q$ ), s'il vérifie la condition suivante :

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

$$\Phi(B)X_t = \mu + \Theta(B)\varepsilon_t$$

où

$$\theta_q \neq 0, \quad \phi_p \neq 0,$$

est un terme constant,

$(\varepsilon_t)_{t \in \mathbb{Z}}$  est un bruit blanc faible avec espérance nulle et variance  $\sigma_\varepsilon^2$  ( $\varepsilon_t \sim \mathbb{N}(0, \sigma_\varepsilon^2)$ ).

$$\Phi(B) = 1 - \phi_1 L - \dots - \phi_p L^p \quad \text{et} \quad \Theta(B) = 1 + \theta_1 L + \dots + \theta_q L^q. \quad (2.1)$$

**Exemple 2.9** (ARMA(1,1))

$$X_t = 0,5X_{t-1} + 0,1\varepsilon_t$$

où :

- .  $X_t$  est la valeur de la série temporelle au temps  $t$
- . 0,5 est le coefficient d'autorégression
- . 0,1 est le coefficient de la moyenne mobile
- .  $\varepsilon_t$  est un processus de bruit blanc

Le coefficient d'autorégression de 0,5 signifie que la valeur de la série temporelle à l'instant  $t$  dépend à 50 % de sa valeur au pas de temps précédent. Le coefficient de moyenne mobile de 0,1 signifie que la valeur de la série chronologique à l'instant  $t$  dépend à 10 % des fluctuations aléatoires de la série chronologique qui ne sont pas expliquées par le coefficient autorégressif.

**Définition 2.8** Un processus stochastique  $(X_t)_{t \in \mathbb{Z}}$  est dit un mélange autorégressif se déplaçant modèle moyen d'ordre 1, ARMA(1,1), s'il vérifie la condition suivante :

$$X_t = \mu + \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad \forall t$$

$$\Phi(B)X_t = \mu + \Theta(B)\varepsilon_t$$

où  $\theta \neq 0$ , est un terme constant.  $(\varepsilon_t)_{t \in \mathbb{Z}}$  est un bruit blanc faible processus avec espérance nulle et variance

$$\sigma_\varepsilon^2 (\varepsilon_t \sim \mathbb{N}(0, \sigma_\varepsilon^2))$$

$$\begin{cases} \Phi(B) = 1 - \phi B \\ \Theta(B) = 1 + \theta B. \end{cases} \quad (2.2)$$

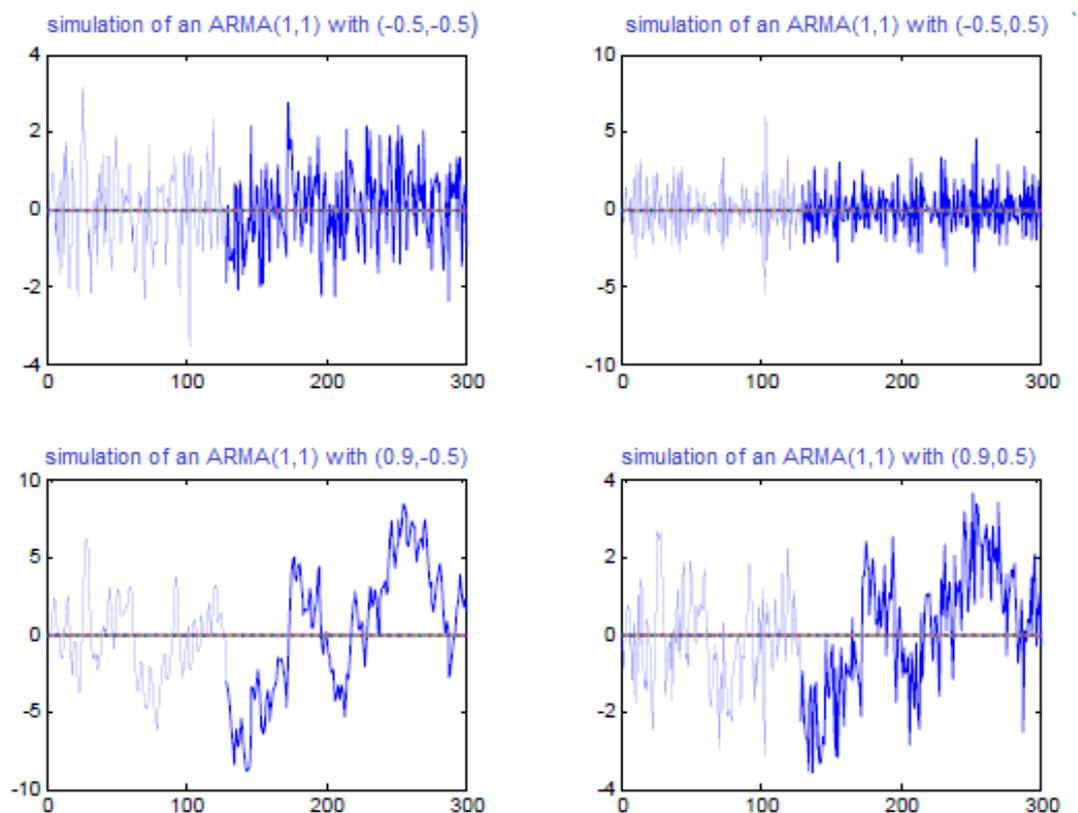


FIGURE 2.7 – Modèle ARMA(1,1)

**Proposition 2.2 (Fonction de covariance)**

Pour tout processus aléatoire stationnaire dont la fonction de corrélation  $\gamma(\cdot)$  tend vers 0 à l'infini, et pour tout entier  $k > 0$ , il est possible de trouver un processus ARMA dont la fonction de corrélation est égale à celle du processus aléatoire jusqu'à l'ordre  $k$ .

**Remarque 2.7** Cette propriété est l'une des raisons de l'intérêt pratique des modèles ARMA; une autre raison est leur facilité de synthèse.

**Proposition 2.3 (mémoire courte)**

Pour un processus ARMA, la fonction de corrélation est bornée géométriquement :

$$|\rho(k)| \leq Cr^k \quad ; k = 1, 2, \dots, \quad C > 0 \quad \text{et} \quad 0 < r < 1$$

on dit qu'un processus ARMA est à mémoire courte (décroissance "rapide" de  $\rho$ ).

**Proposition 2.4 (symétrie temporelle)**

Lorsque l'on observe un processus ARMA en retournant le temps, on obtient un nouveau processus qui a les mêmes propriétés statistiques; la représentation ARMA ne permet donc pas de modéliser un processus dont les décroissances sont en moyenne plus rapides que les croissances.

**Proposition 2.5 (ergodicité)**

Un processus ARMA est un processus aléatoire stationnaire ergodique, ce qui entraîne que pour toute application  $f$  de  $\mathbb{R}^z$  dans  $\mathbb{R}$  mesurable et telle que  $E|f(X)| < \infty$ , alors :

$$\frac{1}{n} \sum_{k=1}^n f \circ B^k(X_t) = \frac{1}{n} \sum_{k=1}^n f(X_{t-k}) \rightarrow E(f(X)) \quad \text{quand } n \rightarrow \infty \text{ et pour tout } t.$$

Cette propriété justifie l'évaluation des caractéristiques du processus à partir d'une réalisation.  $X_t$  est un processus MA d'ordre  $q$  si  $\Phi(z) \equiv 1$ , soit

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

$X_t$  est un processus AR d'ordre  $p$  si  $\Theta(z) \equiv 1$ , soit :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t$$

$X_t$  est un processus MA d'ordre infini,  $MA(\infty)$ , si :

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

avec :

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

Un processus ARMA est dit causal s'il peut être écrit sous la forme d'un  $MA(\infty)$ , soit :

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

( $X_t$  ne dépend que de  $\varepsilon_t$  et des valeurs précédentes de  $\varepsilon$ ).

La causalité est une propriété non pas du processus lui-même, mais de la relation entre  $X$  et  $\varepsilon$ .

### 2.3.4 Critères d'identification de processus ARMA :

Le processus ARMA est identifié si les conditions suivantes sont satisfaites :

- Le processus est stationnaire.
- Pas de tendance.
- Pas de saisonnalité.
- Variance constante.
- La fonction de corrélation empirique est à décroissance pas trop lente.
- Sans pics périodiques.

### 2.3.5 Principes d'identification :

L'identification se fait selon un principe global, illustré par la ( figure 2.8). Décrivons ces étapes, en finissant par la plus délicate, celle de l'hypothèse sur  $p$  et  $q$ .

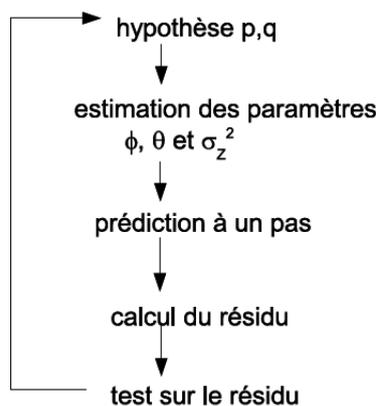


FIGURE 2.8 – Principes généraux d'identification des processus ARMA

### 2.3.6 Identification de l'ordre

Les critères d'information BIC (Bayesian Information Criterion) et AIC (Akaike Information Criterion) sont couramment utilisés pour la sélection de l'ordre du modèle ARMA . Ces critères prennent en compte à la fois l'ajustement du modèle et la complexité du modèle, et ils visent à trouver un équilibre entre la précision et la simplicité(voir e.g., ([Graham Elliott \[2004\]](#)), ([Akaike \[1974\]](#))).

L'expression du critère BIC est donnée par :

$$BIC = -2\log(L) + k \log(n) \quad (2.3)$$

où

$\log(L)$  est la valeur du logarithme de la fonction de vraisemblance maximisée du modèle,  $k = p + q$  est le nombre total de paramètres du modèle,  $n$  est la taille de l'échantillon. L'expression du critère AIC est donnée par :

$$AIC = -2\log(L) + 2(k) \quad (2.4)$$

où

$\log(L)$  est la valeur du logarithme de la fonction de vraisemblance maximisée du modèle,  $k = p + q$  est le nombre total de paramètres du modèle. Dans les deux expressions, le terme  $-2\log(L)$  représente l'ajustement du modèle aux données, et les termes  $k\log(n)$  dans le BIC et  $2k$  dans l'AIC pénalisent la complexité du modèle en fonction du nombre de paramètres.

En plus des critères BIC (Bayesian Information Criterion) et AIC (Akaike Information Criterion), il existe d'autres critères couramment utilisés pour la sélection de modèles dans l'analyse des séries temporelles. Voici quelques-uns d'entre eux :

HQIC (Hannan-Quinn Information Criterion) : Le critère HQIC est similaire au critère AIC, mais il pénalise davantage les modèles avec un grand nombre de paramètres. Il est défini comme :

$$HQIC = -2\log(L) + 2k\log(\log(n)) \quad (2.5)$$

L'interprétation et l'utilisation du HQIC sont similaires à celles de l'AIC, où une valeur plus faible indique un meilleur modèle.

AICC (Corrected Akaike Information Criterion) : Le critère AICC est une version corrigée de l'AIC qui prend en compte la taille de l'échantillon pour les petits échantillons. Il est défini comme :

$$AICC = AIC + (2k(k + 1))/(n - k - 1) \quad (2.6)$$

Le AICC est préféré lorsque la taille de l'échantillon est petite.

CAIC (Consistent Akaike Information Criterion) : Le critère CAIC est une modification de l'AIC qui estime la variance des erreurs de prédiction. Il est défini comme :

$$CAIC = -2\log(L) + k\log(n)(1 + (\log(n)/(n - k))) \quad (2.7)$$

Le CAIC est recommandé lorsque l'on souhaite prendre en compte la taille de l'échantillon et estime la variance des erreurs de prédiction.

Ces critères supplémentaires offrent des alternatives à considérer lors de la sélection de l'ordre du modèle VARMA. Il est généralement recommandé de comparer plusieurs critères et de prendre en compte d'autres facteurs tels que la pertinence théorique, les diagnostics du modèle et la validation croisée pour obtenir une évaluation

globale et robuste de l'ordre du modèle.

### 2.3.6.1 Processus ARIMA et SARIMA :

les modèles *ARIMA* sont bien adaptés au cas des séries temporelles présentant une décroissance très lente de la fonction d'auto-covariance . En pratique, on modélise le processus des accroissements  $X_t - X_{t-d}$  par un processus *ARMA* qui s'appelle *ARIMA*( $p, 1, q$ ).

Le plus souvent  $I = 1$ , ce qui permet d'éliminer une tendance linéaire. Le  $I$  de *ARIMA* signifie "intégré".

En utilisant l'opérateur de retard, on écrit :

$$\phi(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t$$

La marche aléatoire est un processus *ARIMA*(0,1,0) :

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

Le modèle *SARIMA* permet de prendre plus en compte la saisonnalité. Elle est éliminée par différenciation :

$$\phi(B)(1 - B)^d (1 - L^s)X_t = \Theta(B)\varepsilon_t$$

avec  $s$  longueur de la saison (par exemple,  $s=12$  si on a des données mensuelles avec une période d'un an).

### 2.3.6.2 Modèles avec co-variables : ARMAX

$$\phi(B)X_t = \Theta(B)\varepsilon_t + D(B)Z_t$$

soit encore

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_i + \sum_{i=1}^b \beta_i Z_{t-d}$$

où  $(Z_t)_t$  est processus exogène.

### 2.3.7 Estimation des paramètres $\Phi, \Theta$ et $\sigma_\varepsilon^2$

L'estimation des paramètres d'un modèle *ARMA*( $p, q$ ) lorsque les ordres  $p$  et  $q$  sont identifiés peut se réaliser par différentes méthodes :

- Moindres Carrés Ordinaires (modèle sans composante *MA*,  $q = 0$ ).

Dans ce cas, on retrouve les équations de Yule Walker. En remplaçant les auto-corrélations théoriques par leurs estimateurs, on peut retrouver les estimateurs

des MCO des paramètres du modèle par la résolution des équations de Yule Walker.

- Maximum de Vraisemblance approché.
- Maximum de Vraisemblance exacte.

**Définition 2.9** : (maximum de vraisemblance) :

Soit  $X$  une variable aléatoire à valeur dans  $(X, a)$  de loi  $P_\theta$  : On note  $f(x, \theta)$  la densité de  $P_\theta$  et  $f(x_1, \dots, x_n, \theta)$  la densité empirique correspondante. On appelle vraisemblance du paramètre  $\theta$  l'application  $\vartheta \rightarrow \mathbb{R}_+$  définie par :

$$\forall \varphi \in \vartheta \rightarrow L(x_1, \dots, x_n, \varphi) = \prod_{i=1}^n f(x_i, \varphi)$$

Le principe de la vraisemblance est de déterminer la valeur du paramètre  $\theta$  en fonction des observations  $(x_1, \dots, x_n)$ ; qui assure la plus grande probabilité d'apparition de ces observations

**Corollaire 2.1** : Lorsque l'on suppose que l'ensemble  $X$  est indépendant de  $\theta$  et que la fonction de vraisemblance  $L(\cdot)$  est deux fois continûment différentiable par rapport à  $\varphi$ ;  $\forall \varphi \in \vartheta$  alors l'estimateur du maximum de vraisemblance  $\hat{\varphi}$  est solution du système :

$$\left( \frac{\partial L(x, \varphi)}{\partial \theta} \Big|_{\varphi = \hat{\varphi}} = 0 \right)$$

$$\left( \frac{\partial^2 L(x, \varphi)}{\partial \varphi^2} \Big|_{\varphi = \hat{\varphi}} < 0 \right)$$

### 2.3.8 Application du modèle ARMA

On considère un processus stationnaire  $(X_t)_t$  satisfaisant une représentation  $ARMA(p, q)$  telle que :

$$\phi(B)X_t = c + \theta(B)\varepsilon_t, c \in \mathbb{R},$$

avec

$$\theta(B) = \sum_{j=0}^q \theta_j L^j, \quad \phi(B) = \sum_{j=0}^p \phi_j L^j$$

$\forall j < q, \theta_j \in \mathbb{R}, \forall j < p, \phi_j \in \mathbb{R}, \theta_0 = \phi_0 = 1$  et  $(\phi, \theta) \in \mathbb{R}^{p+q}$   
où  $\phi = (\phi_1, \dots, \phi_p)$  et  $\theta = (\theta_1, \dots, \theta_p)$ .

En plus de la définition standard d'un processus  $ARMA$ , on fait l'hypothèse de la normalité des résidus afin de spécifier une forme fonctionnelle à la vraisemblance du modèle.

## 2.4 Tests de Validation

Les tests de validation d'un modèle *ARMA* (AutoRegressive Moving Average) sont des méthodes statistiques utilisées pour évaluer la qualité d'ajustement et la performance prédictive d'un modèle *ARMA* par rapport à un ensemble de données observées. Ces tests sont essentiels pour vérifier si le modèle *ARMA* est approprié pour les données et s'il fournit des prévisions précises.

### 2.4.1 Tests de redondance

Le but est de vérifier si les composantes *AR* et *MA* n'ont pas de racines communes. Si tel est le cas, on peut alors se ramener à une représentation minimale excluant ces racines.

**Exemple 2.10** : On considère un processus stationnaire  $\{X_t\}_{t \in \mathbb{Z}}$  satisfaisant une représentation *ARMA*( $p; q$ ) telle que :

$$\phi(B)x_t = \theta(B)\varepsilon_t$$

Soient  $\lambda_i \in \mathbb{C}$ ,  $i \in [1, \tilde{p}]$ ,  $\tilde{p} \leq p$  les racines de  $\phi(B) = 0$  et

Soient  $\mu_i \in \mathbb{C}$ ,  $i \in [1, \tilde{q}]$ ,  $\tilde{q} \leq q$   $\exists j \in (1, \min(\tilde{p}, \tilde{q}))$  tel que  $\lambda_j = \mu_j$ .

Alors on peut réexprimer les deux polynômes sous la forme :

$$\phi(B) = \prod_{i=1}^{\tilde{p}} \left(1 - \frac{B}{\lambda_i}\right) = \left(1 - \frac{B}{\lambda_j}\right) \prod_{i=1, i \neq j}^{\tilde{p}} \left(1 - \frac{B}{\lambda_i}\right)$$

$$\theta(B) = \prod_{i=1}^{\tilde{q}} \left(1 - \frac{B}{\mu_i}\right) = \left(1 - \frac{B}{\mu_j}\right) \prod_{i=1, i \neq j}^{\tilde{q}} \left(1 - \frac{B}{\mu_i}\right)$$

Dès lors, en divisant les deux polynômes par  $(1 - B\lambda_j^{-1}) = (1 - B\mu_j^{-1})$ , le processus  $x_t, t \in \mathbb{Z}$  peut se réécrire sous la forme :

$$\tilde{\phi}(B)x_t = \tilde{\theta}(B)\varepsilon_t \iff \prod_{i=1, i \neq j}^{\tilde{p}} \left(1 - \frac{B}{\lambda_i}\right)x_t = \prod_{i=1, i \neq j}^{\tilde{q}} \left(1 - \frac{B}{\mu_i}\right)\varepsilon_t$$

La représentation de cette equation correspond à la représentation minimale du processus  $\{x_t, t \in \mathbb{Z}\}$  dans cette représentation les degrés de la représentation *ARMA* seront strictement inférieurs à ceux de la représentation initiale.

### 2.4.2 Tests de significativité des coefficients

Pour n'importe quelle méthode on peut calculer la matrice de variance covariance des  $p+q+1$  estimateurs des paramètres d'un modèle *ARMA*( $p, q$ ), notés  $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$ .

**Théorème 2.2** *L'estimateur du maximum de vraisemblance du vecteur des paramètres du modèle ARMA(p, q); noté  $\hat{\delta} = (\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ , est asymptotiquement distribué suivant une loi normale de moyenne  $\delta$  et de variance covariance*

$$E(\hat{\delta}\hat{\delta}') = F^{-1}$$

où  $F$  désigne la matrice d'information de Fischer.

### 2.4.3 Coefficient de détermination

On considère un processus stationnaire  $x_t, t \in \mathbb{Z}$  satisfaisant une représentation ARMA(p; q). On note  $\hat{\varepsilon}_t$  le résidu d'estimation du modèle. Les coefficients de détermination  $R^2$  et  $\bar{R}^2$  sont alors définis par :

$$R^2 = 1 - \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

$$\bar{R}^2 = 1 - \frac{T-1}{T-p-q} \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

### 2.4.4 Autocorrélations simple et partielle

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

#### 2.4.4.1 Fonction d'autocovariance et d'autocorrélation

La fonction d'autocovariance  $\{\gamma(h)\}_{h \in \mathbb{Z}}$  mesure la covariance entre une variable et cette même variable à des dates différentes, pour un délai  $h$  :

$$\gamma(h) = Cov(Y_t, Y_{t-h}) = E[(Y_t - E(Y_t))(Y_{t-h} - E(Y_{t-h}))]$$

Ainsi

$$\gamma(0) = Var(Y_t) = E[(Y_t - E(Y_t))^2] = \sigma_Y^2.$$

Elle fournit une information sur la variabilité de la série et sur les liaisons temporelles qui existent entre les diverses composantes de la série  $Y_t$ .

La fonction d'autocovariance d'un processus stationnaire est une fonction :

\* Paire :  $\gamma(-h) = \gamma(h) \quad \forall h$

\* Semi-définie positive :

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma(t_j - t_k) > 0 \quad \forall n \in \mathbb{N}, \forall a_j \in \mathbb{R}, \forall t_j \in \mathbb{Z}$$

puisque cette quantité est égale à  $Var(\sum_{k=1}^n a_j Y_{t_j})$

**La fonction d'autocorrélation** est définie par :

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

avec  $\rho(0) = 1$  et  $|\rho(h)| < 1$  (donc  $|\gamma(h)| \leq \Gamma(0)$ )

On appelle coefficient d'autocorrélation d'ordre 1 (resp. d'ordre k) le coefficient de corrélation linéaire  $\rho(1)$  (resp.  $\rho(k)$ ) calculé entre la série et cette série décalée d'une période (resp. k périodes).

On définit la matrice de corrélation (de dimension m) de la manière suivante :

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

Puisque la fonction  $\rho(h)$ ,  $h \in \mathbb{Z}$  est de type positif, on a la propriété suivante :

$$\det R(m) \geq 0 \quad m \in \mathbb{N}^*$$

Ainsi, on a les contraintes suivantes :

- $\det R(1) \geq 0$
- $\det R(2) \geq 0 \iff 1 - \rho(1)^2 \geq 0$
- $\det R(3) \geq 0 \iff [1 - \rho(2)][1 + \rho(2) - 2\rho(1)^2] \geq 0$ .

Ainsi, comme  $\rho(2) < 1$ , on a  $\rho(2) \geq 2\rho(1)^2 - 1$ . Si la corrélation d'ordre 1 est élevée, il en est de même de la corrélation d'ordre 2. Il ne peut donc y avoir de chute brutale de valeur entre  $\rho(1)$   $\rho(2)$  lorsque  $\rho(1)$  est grand.

#### 2.4.4.2 Fonction d'autocorrélation partielle

Elle mesure la liaison (linéaire) entre  $Y_t$  et  $Y_{t-h}$  une fois retirés les liens transitant par les variables intermédiaires  $Y_{t-1}, \dots, Y_{t-h+1}$ .

Le coefficient d'autocorrélation partielle d'ordre h, noté  $r(h)$ , est le coefficient de corrélation entre :

\*  $Y_t - E(Y_t/Y_{t-1}, \dots, Y_{t-h+1})$

\* et  $Y_{t-h} - E(Y_{t-h}/Y_{t-1}, \dots, Y_{t-h+1})$

On a donc :

$$r(h) = \text{corr}(Y_t, Y_{t-h}/Y_{t-1}, \dots, Y_{t-h+1})$$

C'est donc le coefficient de  $Y_{t-h}$  dans la régression de  $Y_t$  sur  $Y_{t-1}, \dots, Y_{t-h+1}, Y_{t-h}$ .

Si  $Y_t$  est un processus stationnaire centré, la prédiction optimale de  $Y_t$  sachant son passé jusqu'à  $t-h$  est donnée par :

$$E(Y_t/Y_{t-1}, \dots, Y_{t-h}) = a_1 Y_{t-1} + \dots + a_h Y_{t-h}$$

que l'on peut réécrire matriciellement :

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_h \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{h-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{h-2} \\ \vdots & & \ddots & \\ \gamma_{h-1} & \gamma_{h-2} & \dots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_h \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{h-1} \\ \rho_1 & 1 & \dots & \rho_{h-2} \\ \vdots & & \ddots & \\ \rho_{h-1} & \rho_{h-2} & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_h \end{pmatrix}$$

Le coefficient d'autocorrélation partielle d'ordre  $h$  d'un processus stationnaire est alors  $a_h$  se calcule de la manière suivante :

$$r(h) = \frac{|R(h)^*|}{|R(h)|}$$

$$R(h) = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{h-1} \\ \rho_1 & 1 & \dots & \rho_{h-2} \\ \vdots & & \ddots & \\ \rho_{h-1} & \rho_{h-2} & \dots & 1 \end{pmatrix}$$

et  $R(h)^*$  la matrice  $R(h)$  dans laquelle on a remplacé la colonne  $h$  par  $\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_h \end{pmatrix}$ ,

soit :

$$R(h)^* = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{h-1} \\ \rho_1 & 1 & \dots & \rho_{h-2} \\ \vdots & & \ddots & \\ \rho_{h-1} & \rho_{h-2} & \dots & \rho_h \end{pmatrix}$$

Ainsi,

$$r(1) = \rho(1) \quad r(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} \dots$$

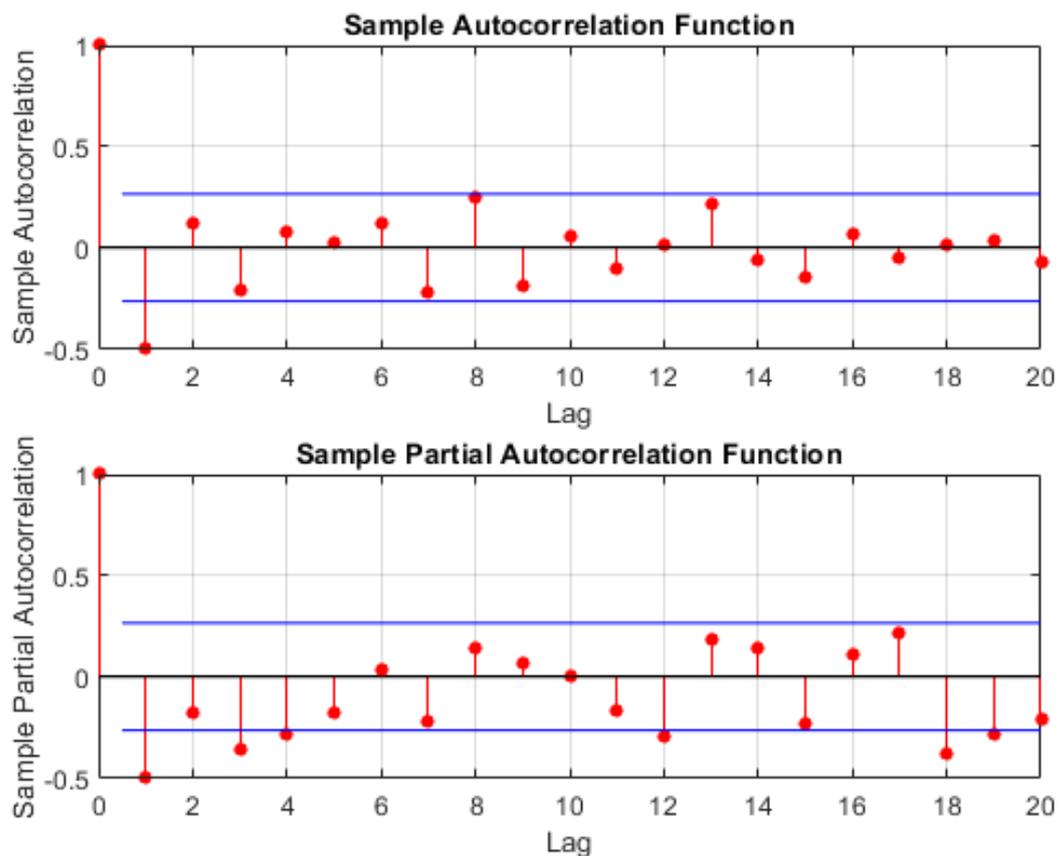


FIGURE 2.9 – Autocorrélation simple et Autocorrélation Partielle

### 2.4.5 Autocorrélation partielle (Méthode de Walker)

La méthode d'autocorrélation partielle (également connue sous le nom de méthode Wal-Kock) est une technique statistique utilisée pour estimer l'ordre du modèle auto-régressif.

La méthode est basée sur l'idée que la fonction d'autocorrélation partielle (PACF) d'une série temporelle décroîtra jusqu'à zéro après un certain nombre de décalages. L'ordre du modèle est ensuite choisi comme le point auquel le PACF franchit pour la première fois le seuil de signification.

La méthode d'autocorrélation partielle est une méthode plus robuste pour estimer l'ordre de décalage que la méthode d'autocorrélation, car elle est moins sensible aux valeurs aberrantes et à la non-stationnarité dans les données. Cependant, il peut être plus gourmand en calcul que la méthode d'autocorrélation.

La signification des autocorrélations partielles peut être évaluée à l'aide d'un test t. L'hypothèse nulle du test t est que l'autocorrélation partielle est égale à zéro. La valeur p du test t peut être utilisée pour déterminer si l'autocorrélation partielle est signifi-

tive.

L'ordre de décalage du modèle autorégressif est choisi comme le point auquel le PACF franchit pour la première fois le seuil de signification. Par exemple, si le PACF est significatif jusqu'au décalage 2, alors l'ordre de décalage du modèle serait choisi comme 2.

La méthode d'autocorrélation partielle est un outil utile pour estimer l'ordre de décalage d'un modèle autorégressif. Il s'agit d'une méthode plus robuste que la méthode d'autocorrélation et elle peut être utilisée pour estimer l'ordre de décalage des séries chronologiques non stationnaires.

### 3.1 Introduction

Le modèle VARMA (Vector Autoregressive Moving Average) est une approche puissante pour analyser les séries chronologiques multivariées, en combinant deux composantes essentielles : l'autorégression (AR) et la moyenne mobile (MA).

La composante AR permet de capturer la dépendance linéaire entre les valeurs passées et actuelles de chaque série chronologique, tandis que la composante MA rend compte de la dépendance linéaire entre les erreurs passées et actuelles du modèle.

En utilisant la notation VARMA(p,q), où p représente l'ordre de l'autorégression et q représente l'ordre de la moyenne mobile, il devient possible de représenter le modèle de manière concise.

Le modèle VARMA offre plusieurs avantages pratiques. Il permet de prévoir les valeurs futures d'une série chronologique multivariée, ce qui est précieux pour la planification et la prise de décision. De plus, il est utile pour identifier les relations de causalité entre différentes séries chronologiques, permettant ainsi de comprendre les liens entre les variables et d'explorer les dynamiques du système étudié.

Pour estimer les coefficients du modèle VARMA, différentes méthodes peuvent être utilisées en fonction des données disponibles.

Les méthodes couramment utilisées incluent la méthode des moindres carrés ordinaires et la méthode du maximum de vraisemblance. Ces approches permettent de déterminer les valeurs optimales des coefficients du modèle en ajustant le modèle aux données observées.

En résumé, le modèle VARMA est un outil statistique polyvalent pour analyser les séries chronologiques multivariées. Grâce à sa capacité à capturer la dépendance temporelle et à prévoir les valeurs futures, il offre de nombreuses possibilités pour l'analyse et la modélisation des données chronologiques.

**Définition 3.1** Un modèle VARMA est représenté par la notation VARMA( $p, q$ ), où  $p$  est l'ordre de l'autorégression et  $q$  est l'ordre de la moyenne mobile. L'ordre de l'autorégression,  $p$ , indique combien de valeurs passées de chaque variable sont prises en compte pour prédire sa valeur actuelle, tandis que l'ordre de la moyenne mobile,  $q$ , indique combien d'erreurs passées des autres variables sont prises en compte. La formule d'un modèle VARMA( $p, q$ ) est donnée par l'équation suivante :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.1)$$

la  $i$ -ème composante est donnée par :

$$X_{it} = \sum_{j=1}^{p_i} \sum_{k=1}^K \phi_{ik,j} X_{K,t-j} - \sum_{j=1}^{q_i} \theta_{ii,j} \varepsilon_{i,t-j} + \varepsilon_{it}$$

où  $X_t$  est un vecteur de dimension  $K$ , (i.e.  $X_t = (X_{1t}, \dots, X_{Kt})^t$ ), contenant les valeurs actuelles de chaque série chronologique dans le modèle, c'est un vecteur de constantes,  $\phi_1, \phi_2, \dots, \phi_p$  et  $\theta_1, \theta_2, \dots, \theta_q$  sont des matrices de coefficients autorégressifs et de moyenne mobile, respectivement,

et  $\varepsilon_t$  est un vecteur d'erreurs de dimension  $K$  qui est un bruit blanc multivarié avec une matrice de covariance constante  $\Sigma$ .

**Remarque 3.1** Il existe plusieurs façons de réduire l'équation du modèle VARMA à un modèle d'espace d'état incorporant une équation de transition qui correspond à un processus de Markov du premier ordre.

L'une des formulations les plus courantes est la représentation dite canonique donnée par :

$$\begin{aligned} \begin{pmatrix} X_{1t} \\ \vdots \\ X_{Kt} \end{pmatrix} &= \begin{pmatrix} \phi_{11}(1) & \dots & \phi_{1K}(1) \\ \vdots & \ddots & \vdots \\ \phi_{K1}(1) & \dots & \phi_{KK}(1) \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ \vdots \\ X_{Kt-1} \end{pmatrix} + \dots + \begin{pmatrix} \phi_{11}(p) & \dots & \phi_{1K}(p) \\ \vdots & \ddots & \vdots \\ \phi_{K1}(p) & \dots & \phi_{KK}(p) \end{pmatrix} \begin{pmatrix} X_{1t-p} \\ \vdots \\ X_{Kt-p} \end{pmatrix} \\ &+ \begin{pmatrix} \theta_{11}(1) & \dots & \theta_{1K}(1) \\ \vdots & \ddots & \vdots \\ \theta_{K1}(1) & \dots & \theta_{KK}(1) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \vdots \\ \varepsilon_{Kt-1} \end{pmatrix} + \dots + \begin{pmatrix} \theta_{11}(q) & \dots & \theta_{1K}(q) \\ \vdots & \ddots & \vdots \\ \theta_{K1}(q) & \dots & \theta_{KK}(q) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-q} \\ \vdots \\ \varepsilon_{Kt-q} \end{pmatrix} \\ &+ \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Kt} \end{pmatrix} \quad (3.2) \end{aligned}$$

Les coefficients du modèle peuvent être estimés à partir des données disponibles à l'aide des méthodes telles que la méthode des moindres carrés ordinaires ou la méthode de maximum de vraisemblance.

## 3.2 Modèle Autoregressif moyenne Mobile vectoriel bi-dimensionnel

Le choix d'un modèle *VARMA* dépend de l'analyse exploratoire des données et de la sélection du modèle approprié pour ajuster les données de manière optimale. Une proposition de modèle *VARMA* pourrait être la suivante :

Supposons que nous ayons deux variables endogènes,  $X_1$  et  $X_2$ , qui sont liées l'une à l'autre dans le temps. Nous voulons utiliser un modèle *VARMA* pour prédire les valeurs futures des deux variables.

Après avoir effectué une analyse exploratoire des données, nous constatons que les deux variables sont corrélées et présentent une certaine causalité. Si nous utilisons un modèle *VARMA(2,1)* qui comprend deux termes autorégressifs et un terme de moyenne mobile.

A titre exemple, l'équation du modèle *VARMA(2,1)* est la suivante :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$X_t = (X_{1t}, X_{2t})^T$  sont les valeurs des variables  $X$  à l'instant  $t$ , respectivement.  $c = (c_1, c_2)^T$  sont des constantes de décalage pour chaque variable.  $\phi_1, \phi_2$  sont des matrices carrées de coefficients autorégressifs d'ordre 2 pour chaque variable.  $\theta_1$  est une matrice de coefficient de moyenne mobile d'ordre 2 pour la relation entre les variables  $X_1$  et  $X_2$ .  $\varepsilon_{t-1}$  et  $\varepsilon_t$  sont les erreurs de prédiction pour chaque variable.

En résumé, le modèle *VARMA(2,1)* utilise les informations sur les variables passées et les moyennes mobiles des erreurs passées pour faire des prédictions précises sur les variables endogènes dans le futur, en tenant compte de la relation entre les deux variables.

Cependant, il est important de noter que le choix des ordres de modèle doit être validé par des techniques de validation de modèle et des analyses de diagnostic afin de s'assurer que le modèle sélectionné est adéquat pour ajuster les données de manière optimale.

### 3.2.1 Définitions de causalité au sens de Granger

La définition de causalité au sens de Granger a été introduite pour la première fois dans un article de Clive Granger [Granger].

**Définition 3.2** Soient :

1.  $X_1 = \{X_{1t}\}$  et  $X_2 = \{X_{2t}\}$  deux séries stationnaires et inversibles ;
2.  $\bar{X}_{1t} = \{X_{1s} \mid s < t\}$  et  $\bar{X}_{2t} = \{X_{2s} \mid s < t\}$  (Historique de  $X_{it}$  pour  $i = 1, 2$ );
3.  $\{\varepsilon_t\}$  un bruit blanc faible à deux dimensions avec matrice de covariance  $\Sigma$ .
4.  $A_t =$  Ensemble dénombrable de variables aléatoires de variances finies et comprenant  $X_{1t}$  et  $X_{2t}$  (Ensemble d'information).
5.  $\bar{A}_t = \bigcup_{s < t} A_s =$  Ensemble de variables aléatoires de variances finies comprenant  $\bar{X}_{1t}$  et  $\bar{X}_{2t}$ .
6.  $\bar{A}_t \setminus \bar{X}_{1t}$  constitue l'information disponible au temps  $(t - 1)$  sans l'historique de  $X_{1t}$ .
7.  $A = \bigcup_{t \in \mathbb{Z}} A_t$ .
8.  $B$  est un sous-ensemble de  $A$ .
9.  $\hat{X}_{it} = E(X_{it} \mid B)$  est le meilleur prédicteur linéaire sans biais de  $X_{it}$  sachant l'information contenu dans  $B$ .
10.  $\varepsilon(X_{it} \mid B) = X_{it} - \hat{X}_{it}$  est l'écart entre la véritable valeur et la prédiction.
11.  $\sigma_\varepsilon^2(X_{it} \mid B) = E[(\varepsilon(X_{it} \mid B))^2]$ .

**Définition 3.3** On dit que  $X_1$  cause  $X_2$  si et seulement si

$$\sigma_\varepsilon^2(X_{2t} \mid \bar{A}_t) < \sigma_\varepsilon^2(X_{2t} \mid \bar{A}_t \setminus \bar{X}_{1t}) \text{ pour au moins une valeur de } t.$$

On note alors  $X_1 \rightarrow X_2$ . On dit qu'il y a rétroaction entre  $X_1$  et  $X_2$  si et seulement si  $X_1$  cause  $X_2$  et  $X_2$  cause  $X_1$ .

**Définition 3.4** On dit que  $X_1$  ne cause pas  $X_2$  si et seulement si

$$\sigma_\varepsilon^2(X_{2t} \mid \bar{A}_t) = \sigma_\varepsilon^2(X_{2t} \mid \bar{A}_t \setminus \bar{X}_{1t}) \text{ pour tout } t \quad (\text{Causalité en variance})$$

ou, de façon équivalente,

$$E(X_{2t} \mid \bar{A}_t) = E(X_{2t} \mid \bar{A}_t \setminus \bar{X}_{1t}) \text{ presque sûrement pour tout } t.$$

On note alors  $X_1 \nrightarrow X_2$

**Définition 3.5** (Causalité en moyenne) *En d'autres termes,  $X_1$  ne cause pas  $X_2$  si et seulement si la prédiction de  $X_{2t}$  sachant l'historique de  $X_{1t}$  et  $X_{2t}$  reste la même que si on connaît seulement l'historique de  $X_{2t}$ . Pour les démonstrations qui suivent, on utilisera cette définition de causalité utilisant ces prédictions. Dans les sections qui suivent,  $\varepsilon_t$  est un bruit blanc faible à deux dimensions tel que :*

1.  $E[\varepsilon_t] = 0$
2.  $E[\varepsilon_t' \varepsilon_s] = \begin{cases} I_2 & \text{si } t = s \\ 0 & \text{si } t \neq s \end{cases}$ ,

où  $I_2$  est la matrice identité d'ordre deux et  $\varepsilon_t'$  est le vecteur  $\varepsilon_t$  transposé.

Dans la suite, nous présentons la forme bidimensionnelle des modèles vus précédemment.

### 3.2.2 Processus autorégressif bidimensionnel (VAR)

On dit que  $(X_1, X_2)'$  est un processus autorégressif s'il peut s'écrire sous la forme

$$\sum_{j=0}^p \phi_j \begin{pmatrix} X_{1,t-j} \\ X_{2,t-j} \end{pmatrix} = \begin{pmatrix} \phi_{11}(B) & \phi_{12}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

où  $\phi_j$  sont des matrices de format  $2 \times 2$  telles que  $\phi_0 = I_2$  et où  $\phi_{11}, \phi_{12}, \phi_{21}$  et  $\phi_{22}$  sont des polynômes en  $B$  définissant les coefficients autorégressifs.

Le théorème suivant nous donne les conditions nécessaires et suffisantes pour que  $X_1$  ne cause pas  $X_2$ .

#### **Théorème 3.1 :**

*Soit  $(X_1, X_2)'$ , un processus autorégressif, alors  $X_1 \nrightarrow X_2$  si et seulement si  $(X_1, X_2)$  est tel que*

$$\begin{pmatrix} \phi_{11}(B) & \phi_{12}(B) \\ 0 & \phi_{22}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

*Démonstration.* On reprend ici une démonstration de (voir [Brooks \[2014a\]](#)) pour le cas bidimensionnel. Premièrement, établissons certaines équations qui nous serviront pour la démonstration. On ne démontre que la condition suffisante dans cet article, mais le reste de la démonstration peut se trouver dans ( voir [Brooks \[2014a\]](#)). Soit le processus  $(X_1, X_2)'$  autorégressif d'ordre deux défini par :

$$\phi(B) \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11}(B) & \phi_{12}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

où  $\phi$  est une matrice de format  $2 \times 2$ .

La définition du processus bidimensionnel introduite dans 3 exige que la matrice  $\phi_0$  soit la matrice identité. Ceci nous garantit que les coefficients de  $X_{1t}$  et  $X_{2t}$  seront toujours un. On peut donc écrire les polynômes  $\phi_{21}(B)$  et  $\phi_{22}(B)$  sous les formes suivantes :

$$\phi_{21}(B) = - \sum_{k=1}^{\infty} \phi_{21(k)} B^k \quad \text{et} \quad \phi_{22}(B) = 1 - \sum_{k=1}^{\infty} \phi_{22(k)} B^k$$

Si l'on multiplie la deuxième ligne de la matrice  $\phi$  avec la colonne  $\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}$ , on peut isoler  $X_{2t}$  et l'écrire sous la forme :

$$X_{2t} = \sum_{k=1}^{\infty} \phi_{21(k)} X_{1,t-k} + \sum_{k=1}^{\infty} \phi_{22(k)} X_{2,t-k} + \varepsilon_{2t}$$

En posant

$$Z_{it} = \sum_{k=1}^{\infty} \phi_{2i(k)} X_{i,t-k},$$

on peut écrire

$$X_{2t} = Z_{1t} + Z_{2t} + \varepsilon_{2t}.$$

On peut donc dire que

$$\begin{aligned} E(X_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) &= E(Z_{1t} + Z_{2t} + \varepsilon_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) \\ &= E(Z_{1t} | \bar{X}_{1t}, \bar{X}_{2t}) + E(Z_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) \\ &\quad + E(\varepsilon_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) = Z_{1t} + Z_{2t}. \end{aligned}$$

La dernière égalité vient du fait que comme  $Z_{it}$  est composé des valeurs de l'historique de  $X_{it}$ , la prédiction de  $Z_{it}$  restera  $Z_{it}$ . Ensuite, la prédiction de  $\varepsilon_{2t}$  est 0, car celui-ci est non corrélé avec  $\bar{X}_{2t}$ . De même, on peut dire que

$$\begin{aligned} E(X_{2t} | \bar{X}_{2t}) &= E(Z_{1t} + Z_{2t} + \varepsilon_{2t} | \bar{X}_{2t}) \\ &= E(Z_{1t} | \bar{X}_{2t}) + E(Z_{2t} | \bar{X}_{2t}) + E(\varepsilon_{2t} | \bar{X}_{2t}) \\ &= E(Z_{1t} | \bar{X}_{2t}) + Z_{2t}. \end{aligned}$$

Si l'on suppose que  $\phi_{21} = 0$ , il s'ensuit que  $Z_{1t} = 0$ , ce qui implique que  $E(Z_{1t} | \bar{X}_{2t}) = 0$ . Alors, on a que  $E(X_{2t} | \bar{X}_{1t}, \bar{X}_{2t}) = E(X_{2t} | \bar{X}_{2t})$  et par conséquent,  $X_1 \rightarrow X_2$

**Exemple 3.1 :**

Soit le processus  $(X_1, X_2)$  ayant pour modèle :

$$\begin{cases} X_{1t} = -0,5X_{1,t-1} + X_{1,t-2} + 2X_{2,t-1} + \varepsilon_{1t} \\ X_{2t} = 0,2X_{2,t-1} + 0,5X_{2,t-2} + \varepsilon_{2t} \end{cases}$$

On peut réécrire le modèle comme suit :

$$\begin{pmatrix} 1 + 0,5B - B^2 & -2B \\ 0 & 1 - 0,2B - 0,5B^2 \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Nous avons que  $\phi_{21}(B) = 0$  donc, en vertu du théorème précédent,  $X_1 \rightarrow X_2$ .

### 3.2.3 Processus à moyenne mobile bidimensionnel(VMA)

On dit que  $(X_1, X_2)'$  possède une représentation de processus à moyenne mobile si l'on peut écrire :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \sum_{j=0}^q \theta_j \begin{pmatrix} X_{1,t-j} \\ X_{2,t-j} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

où, les  $\theta_j$  sont des matrices  $2 \times 2$  telles que  $\theta_0 = I_2$  et  $\theta_{11}, \theta_{12}, \theta_{21}$  et  $\theta_{22}$  sont des polynômes en B définissant les coefficients du processus à moyenne mobile. Le théorème suivant nous donne les conditions nécessaires et suffisantes pour que  $(X_1, X_2)'$  soit inversible.

**Théorème 3.2 :** Si  $(X_1, X_2)'$  est un processus à moyenne mobile bidimensionnel tel que :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Alors  $(X_1, X_2)$  est inversible si et seulement si

$$\begin{vmatrix} \theta_{11}(z) & \theta_{12}(z) \\ \theta_{21}(z) & \theta_{22}(z) \end{vmatrix} \neq 0 \text{ pour tout } z \in \mathbb{C} \text{ telque } |z| \leq 1.$$

Le théorème qui suit nous donne des conditions nécessaires et suffisantes pour vérifier la causalité entre deux séries possédant une structure MA bidimensionnelle

**Théorème 3.3** Soit  $X = (X_1, X_2)'$  un processus à moyenne mobile bidimensionnel inversible, alors  $X_1 \rightarrow X_2$  si et seulement si  $(X_1, X_2)'$  est tel que

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ 0 & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

*Démonstration.* Soit le processus à moyenne mobile donné par

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Comme le processus est inversible, la matrice

$$\theta = \begin{pmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{pmatrix}$$

est inversible et on a

$$\theta^{-1} = \frac{1}{\det(\theta)} \begin{pmatrix} \theta_{22}(B) & -\theta_{12}(B) \\ -\theta_{21}(B) & \theta_{11}(B) \end{pmatrix}$$

On peut donc multiplier par l'inverse des deux côtés de l'équation. On obtient donc :

$$\frac{1}{\det(\theta)} \begin{pmatrix} \theta_{22}(B) & -\theta_{12}(B) \\ -\theta_{21}(B) & \theta_{11}(B) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

qui est un processus autorégressif. On peut donc appliquer le théorème précédent et on montre que  $X_1 \rightarrow X_2$  si et seulement si  $\phi_{21} = 0$ .

**Exemple 3.2** Soit le processus  $(X_1, X_2)$  ayant pour modèle :

$$\begin{cases} X_{1t} = \varepsilon_{1,t} - 0,25\varepsilon_{1,t-2} + \varepsilon_{2,t-1} - 0,25\varepsilon_{2,t-2} \\ X_{2t} = \varepsilon_{2,t} - 0,2\varepsilon_{2,t-2} \end{cases}$$

On peut réécrire le modèle comme suit :

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 1 - 0,25B^2 & B - 0,25B^2 \\ 0 & 1 - 0,2B^2 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Comme  $\phi_{21}(B) = 0$ , en vertu du théorème précédent,  $X_1 \rightarrow X_2$ .

### 3.2.4 Processus ARMA bidimensionnel

Pour la suite du document, nous écrivons  $\theta_{ij} = \theta_{ij}(B)$  et  $\phi_{ij} = \phi_{ij}(B)$ .  
Le processus  $(X_1, X_2)'$  possède une représentation ARMA s'il peut s'écrire sous la forme :

$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

où  $\phi_{ij}$  et  $\theta_{ij}$  sont des polynômes en  $B$  pour  $i, j = 1, 2$ .

#### Théorème 3.4 :

Soit  $(X_1, X_2)'$  un processus suivant un modèle ARMA inversible donné par

$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

a) Condition suffisante : Si  $\phi_{21} = 0$  et  $\theta_{12} = \theta_{21} = 0$ , alors  $X_1 \rightarrow X_2$ .

Cette condition correspond à la représentation suivante :

$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11} & 0 \\ 0 & \theta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

b) Condition nécessaire et suffisante :  $X_1 \rightarrow X_2$  si et seulement si

$$\phi_{11}\theta_{21} - \theta_{21}\theta_{11} = 0.$$

Pour démontrer ces résultats, on utilise le fait que le processus est inversible et on multiplie par  $\theta^{-1}$  comme dans la démonstration précédente. On utilise ensuite la condition de causalité du processus autorégressif.

Théoriquement, ces conditions de causalité sont plutôt faciles à vérifier lorsque l'on possède déjà un modèle. Cependant, en pratique, nous possédons un échantillon du vecteur  $X_t$ . Nous devons d'abord déterminer quel type de modèle utiliser et l'ordre de celui-ci.

## 3.3 Stationarité

Lorsque nous disons qu'une série temporelle multivariée est stationnaire, nous faisons référence à une propriété où les propriétés statistiques de la série restent constantes dans le temps pour chaque variable de la série. En d'autres termes, chaque variable présente individuellement la stationnarité.

Pour qu'une série temporelle multivariée soit considérée comme stationnaire, elle doit satisfaire aux conditions suivantes :

- **Moyenne constante** : La moyenne de chaque variable reste constante dans le temps.
- **Covariance constante** : La covariance entre deux variables de la série reste constante dans le temps. Cela signifie que la relation entre les variables ne change pas systématiquement au fil du temps.
- **Autocovariance constante** : L'autocovariance de chaque variable reste constante à travers différents décalages temporels.
- **Absence de tendance ou de saisonnalité** : L'absence de tendance systématique ou de schémas saisonniers dans la série. Cela signifie que la moyenne et la variance de chaque variable ne dépendent pas du moment précis où elles sont observées.

La stationnarité est une propriété importante pour l'analyse des séries temporelles multivariées, car de nombreux modèles statistiques et méthodes de prévision supposent des données stationnaires ou fonctionnent mieux avec celles-ci.

Elle simplifie l'analyse et permet une inférence et une prédiction plus fiables.

Pour évaluer la stationnarité d'une série temporelle multivariée, différents tests peuvent être utilisés, tels que le test de Johansen ([Johansen \[1988\]](#)), le test de Phillips-Perron (voir [Kwiatkowski \[1992\]](#)) ou d'autres extensions multivariées du test de Dickey-Fuller augmenté (ADF) (voir [Lütkepohl \[2006\]](#)). Ces tests examinent la présence de racines unitaires ou d'autres formes de non-stationnarité entre les variables de la série.

### 3.3.1 Tests de stationnarité

#### 3.3.1.1 Test Dukey Fuler augmenté

Le test ADF (Augmented Dickey-Fuller) est couramment utilisé pour évaluer la stationnarité d'une série temporelle univariée. Cependant, il n'existe pas de test ADF direct pour les séries temporelles multivariées dans R. Le test ADF est spécifiquement conçu pour les données univariées.

Pour les séries temporelles multivariées, nous pouvons utiliser le test de Johansen, qui est un test populaire de cointégration et de stationnarité dans un contexte multivarié.

Le test de Johansen examine s'il existe une relation à long terme, ou cointégration, entre les variables de la série.

S'il n'y a pas de cointégration, cela signifie que les variables sont stationnaires individuellement.

L'hypothèse nulle du test de Dickey-Fuller augmenté (ADF) est que la série temporelle a une racine unitaire. L'hypothèse alternative est que la série temporelle n'a pas de racine unitaire. La statistique du test est un nombre négatif, et plus il est négatif, plus la preuve contre l'hypothèse nulle est forte (voir e.g., ([Lütkepohl \[2006\]](#), [Hamilton](#)

[(2021)].

### 3.3.1.2 Test de Johansen

Le test de Johansen est un test largement utilisé pour vérifier la cointégration et la stationnarité des séries temporelles multivariées. Il permet de déterminer s'il existe une relation à long terme entre les variables de la série.

Le test de Johansen est utilisé pour tester l'hypothèse de cointégration entre les variables d'une série temporelle multivariée. La cointégration implique l'existence d'une relation à long terme ou d'un équilibre entre les variables, malgré les fluctuations à court terme.

Le test de Johansen permet de déterminer si une telle relation existe et fournit des informations supplémentaires sur le rang de la matrice de cointégration.

Le test de Johansen consiste à tester deux séries d'hypothèses :

**Hypothèse nulle** : L'hypothèse nulle du test de Johansen est qu'il n'y a pas de cointégration entre les variables. En d'autres termes, il n'existe pas de relations à long terme ou de relations d'équilibre entre les variables de la série temporelle multivariée.

**Hypothèse alternative** : L'hypothèse alternative dépend du rang de la matrice de cointégration testée.

L'expression générale du test de Johansen est basée sur la théorie des processus de vecteurs autorégressifs (VAR) et repose sur la décomposition spectrale des matrices de covariance.

Il existe plusieurs versions du test de Johansen, dont les tests de trace (trace test) et les tests des valeurs propres maximales (maximum eigenvalue test). Voici les expressions de base pour ces deux tests :

**Définition 3.6 Test de trace (Trace test)** : Le test de trace examine l'hypothèse nulle selon laquelle il y a au plus  $r$  relations de cointégration, contre l'hypothèse alternative selon laquelle il y a  $r + 1$  relations de cointégration.

L'expression générale du test de trace est donnée par :

$$\text{Tracetest} = -T \sum (\log(1 - \lambda_i))$$

Où :

- $T$  est la taille de l'échantillon (le nombre d'observations).
- $\lambda_i$  représente les valeurs propres (eigenvalues) des tests.
- La somme ( $\sum$ ) est effectuée pour toutes les valeurs propres ( $\lambda_i$ ) du test.

**Test des valeurs propres maximales (Maximum eigenvalue test) :** Le test des valeurs propres maximales examine l'hypothèse nulle selon laquelle il y a au plus  $r$  relations de cointégration, contre l'hypothèse alternative selon laquelle il y a  $r$  relations de cointégration. L'expression générale du test des valeurs propres maximales est donnée par :

$$\text{Max eigenvalue test} = -T \log(1 - \lambda_i)$$

Où :

- $T$  est la taille de l'échantillon (le nombre d'observations).
- $\lambda_i$  représente les valeurs propres (eigenvalues) du test.

Dans les deux tests, les valeurs propres ( $\lambda_i$ ) sont obtenues à partir de la décomposition spectrale des matrices de covariance. Les tests comparent les valeurs propres observées avec les valeurs propres critiques pour déterminer la significativité statistique des relations de cointégration.

Il est important de noter que ces expressions sont simplifiées et qu'il existe différentes versions du test de Johansen qui peuvent inclure des ajustements pour des problèmes tels que la non-stationnarité des données et l'autocorrélation. Les versions spécifiques du test peuvent différer légèrement dans les expressions et les calculs, mais les principes fondamentaux restent les mêmes.

Il convient également de souligner que ces expressions sont fournies à titre informatif et qu'il est recommandé de se référer à la littérature spécialisée ou aux packages logiciels appropriés pour obtenir des détails complets et précis sur la mise en œuvre du test de Johansen.

### 3.3.2 Estimation

Nous introduisons ensuite des éléments de notation pour les paramètres de notre modèle. Tout d'abord, quelle que soit la représentation VARMA utilisée, nous divisons l'ensemble du vecteur de paramètres  $\gamma$  en deux parties  $\gamma_1$  (les paramètres de la partie AR) et  $\gamma_2$  (partie MA) :

$$\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$

$$\begin{cases} \phi_{i\bullet}(L) = [\phi_{i1}(L), \dots, \phi_{iK}(L)] \\ \phi_{\bullet k}(L) = [\phi_{1k}(L), \dots, \phi_{Kk}(L)] \end{cases}$$

Pour un modèle VARMA sous forme d'équation diagonale MA

$$\begin{aligned} \gamma_1 &= [\phi_{1\bullet,1}, \dots, \phi_{1\bullet,p}, \dots, \phi_{k\bullet,1}, \dots, \phi_{k\bullet,p}]^t, \\ \gamma_2 &= [\theta_{11,1}, \dots, \theta_{11,q_1}, \dots, \theta_{kk,1}, \dots, \theta_{kk,q_K}]^t, \end{aligned}$$

tandis que pour un modèle VARMA sous forme d'équation MA finale

$$\gamma_2 = [\theta_1, \dots, \theta_q]$$

Pour les modèles VARMA sous forme d'équation diagonale AR, nous inversons simplement  $\gamma_1$  et  $\gamma_2$

$$\begin{aligned}\gamma_1 &= [\phi_{11,1}, \dots, \phi_{11,p_1}, \dots, \phi_{kK,1}, \dots, \phi_{kK,p_K}]^t, \\ \gamma_2 &= [\theta_{1\bullet,1}, \dots, \theta_{1\bullet,q}, \dots, \theta_{k\bullet,1}, \dots, \theta_{k\bullet,q}]^t.\end{aligned}$$

Tandis que pour un modèle VARMA sous forme d'équation AR finale

$$\gamma_1 = [\phi_1, \dots, \phi_p]^t$$

La méthode d'estimation comporte trois étapes. Les observations vont de  $t = 1, \dots, T$ .

**Etape 1** : *Estimer un modèle VAR( $n_T$ ) pour approximer le modèle VARMA( $p, q$ ) et calculer les résidus*

$$\hat{\varepsilon}_t = X_t - \sum_{j=1}^{n_T} \hat{\Pi}_j^{(n_T)} X_{t-j} \quad \text{pour } t = n_T + 1, \dots, T.$$

**Etape 2** : *À l'aide des résidus  $\hat{\varepsilon}_t$ , nous calculons une estimation de la matrice de covariance de  $\varepsilon_t$ ,*

$$\hat{\Sigma}_\varepsilon = \frac{1}{T} ; \quad \sum_{t=n_t+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

et

$$\phi(L)X_t = [\Theta(L) - I_K] \hat{\varepsilon}_t + e_t.$$

pour obtenir des estimation  $\tilde{\phi}(L)$  et  $\tilde{\Theta}(L)$  de  $\phi(L)$  et  $\Theta(L)$ , l'estimateur est

$$\tilde{\gamma} = \left[ \sum_{t=l}^T \hat{Z}'_{t-1} \hat{\Sigma}_\varepsilon^{-1} \hat{Z}_{t-1} \right]^{-1} \left[ \sum_{t=l}^T \hat{Z}'_{t-1} \hat{\Sigma}_\varepsilon^{-1} X_t \right] \quad (3.3)$$

ou  $l = n_T + \max(p, q) + 1$ .

$$X_{t-1}^{(p)} = [X_{1,t-1}, \dots, X_{k,t-1}, \dots, X_{1,t-p}, \dots, X_{k,t-p}],$$

$$\hat{\varepsilon}_{t-1}^{(p)} = [\hat{\varepsilon}_{1,t-1}, \dots, \hat{\varepsilon}_{k,t-1}, \dots, \hat{\varepsilon}_{1,t-q}, \dots, \hat{\varepsilon}_{k,t-q}],$$

$$X_{t-1}^{(k)} = [X_{k,t-1}, \dots, X_{k,t-p_k}],$$

**Etape 3** : *En utilisant les estimations de la deuxième étape, nous formons d'abord de nou-*

---

veaux résidus

$$\tilde{\varepsilon}_t = X_t - \sum_{i=1}^p \tilde{\Phi}_i X_{t-i} - \sum_{j=1}^q \tilde{\Theta}_j \tilde{U}_{t-j}$$

initiant avec  $\tilde{\varepsilon}_t = 0$  pour  $t \leq \max(p, q)$ , et nous définissons

$$M_t = \sum_{j=1}^q \tilde{\Theta}_j M_{t-j} + X_t, \quad W_t = \sum_{j=1}^q \tilde{\Theta}_j W_{t-j} + \tilde{\varepsilon}_t$$

initiant avec  $X_t = W_t = 0$  pour  $t \leq \max(p, q)$ .

Nous calculons également une nouvelle estimation de  $\Sigma_\varepsilon$ ,  $\tilde{\Sigma}_\varepsilon = \frac{1}{T} \sum_{t=l}^T \tilde{\varepsilon}_t \tilde{\varepsilon}_t'$  avec  $l = \max(p, q) + 1$ . Ensuite, nous effectuons une régression  $\tilde{\varepsilon}_t + X_t - W_t$  sur  $\tilde{V}_{t-1}$  par GLS, avec

$$\tilde{V}_t = \sum_{j=1}^q \tilde{\Theta}_j \tilde{V}_{t-j} + \tilde{Z}_t$$

où  $\tilde{Z}_t$  est calculé comme  $\hat{Z}_t$  à l'étape 2,  $\tilde{\varepsilon}_t$  étant remplacé par  $\hat{\varepsilon}_t$ .

On obtient ainsi

$$\tilde{\gamma} = \left[ \sum_{t=l}^T \hat{V}_{t-1}' \tilde{\Sigma}_\varepsilon^{-1} \tilde{V}_{t-1} \right]^{-1} \left[ \sum_{t=l}^T \tilde{V}_{t-1}' \tilde{\Sigma}_\varepsilon^{-1} [\tilde{\varepsilon}_t + X_t - W_t] \right]$$

Nous extrayons les estimations finales  $\hat{\Phi}_i$  et  $\hat{\Theta}_j$ .

Les propriétés des estimateurs ci-dessus sont résumées dans les trois théorèmes suivants. Le théorème (3.5) est une généralisation des résultats de (voir Lewis [1985]) où la convergence est démontrée pour des innovations mixtes plutôt que des innovations i.i.d.

Nous désignons la norme euclidienne par  $\|B\|^2 = \text{tr}(B'B)$ .

**Théorème 3.5** (Première étape d'estimations du modèle VARMA)

Supposons que  $X_t$  satisfasse au modèle VARMA en (3.1) ainsi que les hypothèses (2.1), (3.1) et (2.2). Si  $n_T / \log(T) \rightarrow \infty$  alors les estimateurs  $\hat{\Pi}_j^{(n_T)}$  satisfaits

$$\sum_{j=1}^{n_T} \|\hat{\Pi}_j^{(n_T)} - \Pi_j\| = O_p(\sqrt{n_T/T}),$$

**Théorème 3.6** (Deuxième étape d'estimation du modèle VARMA ).

Sous les hypothèses du Théorème (3.5), supposons que les paramètres du modèle soient identifiés.

Alors l'estimateur de seconde étape  $\tilde{\gamma}$  dans (3.1) converge en probabilité vers la vraie valeur  $\gamma$ , et

$$\sqrt{T}(\tilde{\gamma} - \gamma) \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}\left[0, J_{(2)}^{-1} I_{(2)} J_{(2)}^{-1}\right]$$

où

$$I_{(2)} = \sum_{j=-\infty}^{\infty} E\left[Z'_{t-1} \sum_{\varepsilon}^{-1} \varepsilon_t \left[ Z'_{t-1-j} \sum_{\varepsilon}^{-1} \varepsilon_{t-j} \right]^t\right], \quad J_{(2)} = E\left[Z'_{t-1} \sum_{\varepsilon}^{-1} Z_{t-1}\right]$$

et  $Z_{t-1}$  est égale à la matrice  $\hat{Z}_{t-1}$  où  $\hat{\varepsilon}_t$  remplacer par  $\varepsilon_t$ , en plus si

$\omega(j, m_T) = 1 - |j|/(m_T + 1)$  avec  $m_T^4/T \rightarrow 0$  et  $m_T \rightarrow \infty$  avec  $T \rightarrow \infty$ , nous avons :

$$\hat{I}_{(2)} := \frac{1}{T} \sum_{j=-m_T}^{m_T} \omega(j, m_T) \sum_{t=l+|j|}^T \left[ \hat{Z}'_{t-1} \sum_{\varepsilon}^{-1} \hat{\varepsilon}_t \left[ \hat{Z}'_{t-1-j} \sum_{\varepsilon}^{-1} \hat{\varepsilon}_{t-j} \right]^t \right] \xrightarrow[T \rightarrow \infty]{p} I_{(2)},$$

$$\hat{J}_{(2)} := \frac{1}{T} \sum_{t=l}^T \hat{Z}'_{t-1} \sum_{\varepsilon}^{-1} \hat{Z}_{t-1} \xrightarrow[T \rightarrow \infty]{p} J_{(2)},$$

**Théorème 3.7** (Troisième étape d'estimations du modèle VARMA)

Sous les hypothèses du théorème (3.5), l'estimateur de troisième étape  $\hat{\gamma}$  converge en probabilité vers la vraie valeur  $\gamma$  et

$$\sqrt{T}(\hat{\gamma} - \gamma) \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}\left[0, J_{(3)}^{-1} I_{(3)} J_{(3)}^{-1}\right]$$

où

$$I_{(3)} = \sum_{j=-\infty}^{\infty} E\left\{ \left[ V'_{t-1} \sum_{\varepsilon}^{-1} \varepsilon_t \left[ V'_{t-1-j} \sum_{\varepsilon}^{-1} \varepsilon_{t-j} \right]^t \right] \right\}, \quad J_{(3)} = E\left[ V'_{t-1} \sum_{\varepsilon}^{-1} V_{t-1} \right]$$

et  $V_{t-1}$  est égale à la matrice  $\hat{V}_{t-1}$  où  $\hat{\varepsilon}_t$  remplacer par  $\varepsilon_t$  en plus si  $m_T^4/T \rightarrow 0$  et  $m_T \rightarrow \infty$  comme  $T \rightarrow \infty$ , on a

$$\hat{I}_{(3)} := \frac{1}{T} \sum_{j=-m_T}^{m_T} \omega(j, m_T) \sum_{t=l'+|j|}^T \left[ \tilde{V}'_{t-1} \sum_{\varepsilon}^{-1} \tilde{\varepsilon}_t \left[ \tilde{V}'_{t-1-j} \sum_{\varepsilon}^{-1} \tilde{\varepsilon}_{t-j} \right]^t \right] \xrightarrow[T \rightarrow \infty]{p} I_{(3)},$$

$$\hat{J}_{(3)} = \frac{1}{T} \sum_{t=l'}^T \tilde{V}'_{t-1} \sum_{\varepsilon}^{-1} \tilde{V}_{t-1} \xrightarrow[T \rightarrow \infty]{p} J_{(3)},$$

où  $\omega(j, m_T) = -1 - |j|/(m_T + 1)$  et  $\bar{\varepsilon}_t$  sont les résidus calculés avec  $\hat{\gamma}$ .

A l'étape 1, si l'ordre  $n_T$  du VAR croît au taux approprié, l'erreur d'approximation résultant de l'estimation d'un VAR d'ordre fini, lorsque la vérité est  $VAR(\infty)$  se réduira à zéro avec la taille de l'échantillon.

À l'étape 2, l'erreur de mesure problème d'utilisation des résidus au lieu des véritables innovations car les régresseurs disparaissent asymptotiquement ce qui rend les estimations cohérentes des coefficients de régression. Remplacer l'innovation  $\varepsilon_t$  par un résiduel induit une structure MA dans l'erreur de mesure.

L'étape 3, construit un filtre MA pour le corriger et améliorer l'efficacité de l'estimateur.

Vu la simplicité de cette méthode d'estimation, seules trois régressions sont nécessaires, ce qui évite les difficultés associées aux optimisations non linéaires. Cela est particulièrement important dans les modèles VARMA, où la convergence numérique peut être difficile à obtenir en raison du grand nombre de paramètres à traiter.

De plus, la distribution asymptotique des estimateurs peut être une mauvaise approximation de la distribution d'échantillons finis dans les modèles dynamiques de grande dimension.

Par conséquent, une méthode d'estimation basée sur des méthodes linéaires est intéressante, car elle suggère l'utilisation de procédures basées sur la simulation, telles que les techniques de bootstrap. Cela serait impraticable si l'estimation était basée sur des optimisations non linéaires.

De plus, il est souligné que cette procédure n'est pas spécifique aux représentations utilisées dans ce travail, et peut être adaptée à d'autres représentations identifiées.

Étant donné que la méthode d'estimation repose uniquement sur des régressions linéaires, il est possible d'utiliser une représentation moins parcimonieuse, alors que cela serait essentiellement limité lors de l'utilisation de méthodes d'estimation non linéaires.

Un avantage des représentations MA Diagonal et MA Final proposées est que si les estimations de la deuxième étape ne correspondent pas à une représentation MA inversible (avec des racines à l'intérieur du cercle unité), il est facile d'obtenir la représentation inversible correspondante pour effectuer la troisième étape.

**Théorème 3.8** (*Efficacité des estimations de la troisième étape sous gaussianité*)

*Sous les hypothèses de Théorème (3.7), supposons que les innovations  $\varepsilon_t$  soient i.i.d. Gaussien. Ensuite, la variance asymptotique de l'estimateur de troisième étape est égal à la variance asymptotique de l'estimateur du maximum de vraisemblance.*

**Remarque 3.2** (*Efficacité des estimations de troisième étape sous innovations faibles*)

Sous les hypothèses du théorème (3.7), la variance asymptotique de l'estimateur de troisième étape est égale à la variance asymptotique de l'estimateur du quasi-maximum de vraisemblance.

Nous pouvons voir que la variance asymptotique de cet estimateur linéaire en trois étapes est la même que celle des moindres carrés non linéaires si les innovations ne sont non corrélées ou si elles sont *i.i.d* Gaussien.

Pour avoir une idée de la perte d'efficacité dans les échantillons finis due au remplacement des véritables innovations par des résidus d'un long VAR.

Nous pouvons interpréter la troisième étape comme une étape de **Newton-Raphson** dans la minimisation de la somme des carrés des résidus  $\sum_{t=1}^T \varepsilon_t \varepsilon_t'$ .

Il peut également être considéré comme une correction induite par le MA structure dans l'erreur que nous commettons, lorsque le vrai terme d'erreur  $\varepsilon_t$  est remplacé par le résidu de première étape  $\hat{\varepsilon}_t$  (voir [Reinsel \[1992\]](#)).

### 3.3.3 Sélection de l'ordre

Nous avons encore des inconnues dans notre modèle, les ordres des opérateurs AR et MA. Si aucune théorie ne spécifie ces paramètres, nous devons utiliser une procédure statistique pour les choisir. La méthode suivante une méthode de critère d'information pour choisir les ordres des modèles VARMA dans les différentes zones identifiées.

Dans la deuxième étape de l'estimation, nous calculons pour toutes les valeurs de  $p \leq P$  et  $q \leq Q$  le critère d'information suivant :

$$DP(p, q, T) := \log \left( \det \sum_{\varepsilon} \right) + \dim(\gamma) \frac{(\log T)^{1+\delta}}{T}$$

pour un certain  $\delta < 0$ , nous choisissons alors  $\hat{p}$  et  $\hat{q}$  comme l'ensemble qui minimise le critère d'information.

Nous supposons que les limites supérieures  $P$  et  $Q$  des ordres de la partie AR et MA sont plus grandes que les vraies valeurs de  $p$  et  $q$  (ou qu'elles augmentent lentement avec la taille de l'échantillon). Les propriétés de  $\hat{p}$  et  $\hat{q}$  sont résumées dans le théorème suivant.

**Théorème 3.9** (*Estimation de l'ordre  $(p, q)$  du modèle VARMA*)

Sous les hypothèses de Théorème , supposons que  $0 \leq p \leq P$  et  $0 \leq q \leq Q$ .

Si  $n_T = O((\log T)^{(1+\delta_1)})$  de certains  $\delta_1 > 0$  avec  $\delta_1 < \delta$ , et  $(\hat{p}, \hat{q})$  minimisent  $DP(p, q, T)$ , alors  $\hat{p}$  et  $\hat{q}$  convergent en probabilité vers leur vraies valeurs.

**Remarque 3.3** *En pratique, cette procédure peut conduire à une recherche sur trop de mo-*

dèles pour les représentations diagonales. Une alternative valable consiste à rechercher les vrais ordres en procédant équation par équation.

Dans la deuxième étape de l'estimation, plutôt qu'une estimation simultanée, effectuez simplement des régressions. Pour un modèle VARMA sous forme d'équation diagonale MA, nous estimons les régressions.

$$X_{it} = \sum_{j=1}^{p_i} \sum_{k=1}^K \phi_{ik,j} X_{K,t-j} - \sum_{j=1}^{q_i} \theta_{ii,j} \varepsilon_{i,t-j} + \varepsilon_{it}$$

pour  $i = 1 \dots K$  tandis que pour un modèle VARMA sous forme d'équation diagonale AR, nous considérons

$$X_{it} = \sum_{j=1}^{p_i} \phi_{ii,j} X_{i,t-j} - \sum_{j=1}^{q_i} \sum_{k=1}^K \theta_{iK,j} \varepsilon_{K,t-j} + \varepsilon_{it}$$

Nous avons alors choisi  $\hat{p}_i$  et  $\hat{q}_i$  comme ordres minimisant le critère d'information suivant :

$$\log(\hat{\sigma}_i^2) + g(p_i, q_i) \frac{(\log T)^{1+\delta}}{T} \quad \text{où } \delta > 0$$

et

$$g(p_i, q_i) = p_i K + q_i \quad \text{où } g(p_i, q_i) = p_i + q_i K$$

pour la représentation diagonal MA ou AR sous forme d'équation respectivement. L'ordre global de l'opérateur autorégressif est alors  $\hat{p} = \max(\hat{p}_1, \dots, \hat{p}_K)$  pour la représentation diagonale MA, et de même  $\hat{q} = \max(\hat{q}_1, \dots, \hat{q}_K)$ .

Pour la représentation AR diagonale. On voit que cette procédure de sélection équation par équation n'est pas seulement plus facile à appliquer, mais peut conduire à des représentations plus parcimonieuses en identifiant des rangées de zéros coefficients en  $\phi_j$  ou  $\theta_j$ .

**Théorème 3.10** (Estimation de l'ordre  $(p, q)$  du modèle VARMA diagonal)

Sous l'hypothèse du théorème 3.9, si le modèle VARMA est sous la forme d'équation diagonale MA ou AR et les ordres sont choisis en minimisant  $DP(p, q, T)$ , alors  $\hat{p}_i$  et  $\hat{q}_i$ ,  $i = 1, \dots, K$ , convergent en probabilité vers leur vraie valeur.

**Remarque 3.4** Le critère  $DP(p, q, T)$  est une généralisation du critère d'information proposé par (voir Hannan [1982]).

Cependant, ces auteurs ont reconnu plus tard que ce critère devait être modifié pour obtenir des estimations cohérentes des ordres  $p$  et  $q$ . Le critère initial était :

$$\log(\tilde{\sigma}^2) + (p + q) \frac{(\log T)^\delta}{T}$$

avec  $\delta > 0$  (voir [Hannan \[1983\]](#)) ont montré que  $\tilde{\sigma}^2 - \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2$  est  $O_p(n_T T^{-1})$  plutôt que  $O_p(T^{-1})$ , de sorte que  $(\log T)^\delta / T$  n'est pas assez élevé.

Deux modifications possibles sont alors considérés.

La première est relativement simple et consiste à prendre  $(\log T)^{1+\delta}$  au lieu de  $(\log T)^\delta$  dans le critère d'information, de sorte que la pénalité sur  $p + q$  domine  $\log(\hat{\sigma}^2)$  dans le critère.

La seconde qu'ils privilégient et utilisent dans des travaux ultérieurs (voir [Hannan \[1984\]](#)), considère une modification de la première étape de la procédure :

au lieu de prendre  $n_T = O(\log T)$ .

Un autre critère d'information est utilisé pour choisir l'ordre de la longue autorégression en choisissant des valeurs  $p$  et  $q$  potentiellement différentes à chaque itération.

Une approche similaire est également proposé par (voir [Poskitt \[1987\]](#)).

## 4.1 Introduction

Lorsque l'économie mondiale croît, la demande de Pétrole et d'Or augmente en raison de leur utilité et de leur statut d'investissement sûr. L'Or est considéré comme un actif refuge en période d'incertitude économique, tandis que le Pétrole alimente les transports et d'autres industries. Lorsque l'inflation est élevée, la valeur du dollar diminue, ce qui rend l'Or plus attractif en tant qu'un investissement stable.

Les événements géopolitiques tels que les guerres ou l'instabilité politique peuvent perturber la production et le transport du pétrole, ainsi qu'augmenter la demande d'or en tant que refuge. Le Pétrole et l'Or sont des matières premières essentielles avec une grande négociabilité, une forte liquidité et des mouvements synchronisés. Le Pétrole est la matière première la plus échangée avec une volatilité élevée, tandis que l'Or, principal métal précieux, présente une volatilité plus faible.

Au cours des dernières décennies, l'Or et le Pétrole ont généralement évolué ensemble, montrant une corrélation positive de 80%. Les fluctuations du prix du pétrole ont été identifiées comme influençant le prix de l'Or, avec des rendements du Pétrole représentant 1,7% de ceux de l'Or. Une explication possible de cette relation est l'utilisation de l'Or comme moyen de paiement par de nombreux pays importateurs de Pétrole. Les statistiques économiques ont prouvé que le prix de l'Or et le prix du Pétrole sont fortement liés (voir le graphe 4.1), d'après le cite [macrotrends.net](https://www.macrotrends.net/1334/gold-prices-vs-oil-prices-historical-correlation) (Voir le lien du site internet (<https://www.macrotrends.net/1334/gold-prices-vs-oil-prices-historical-correlation>))

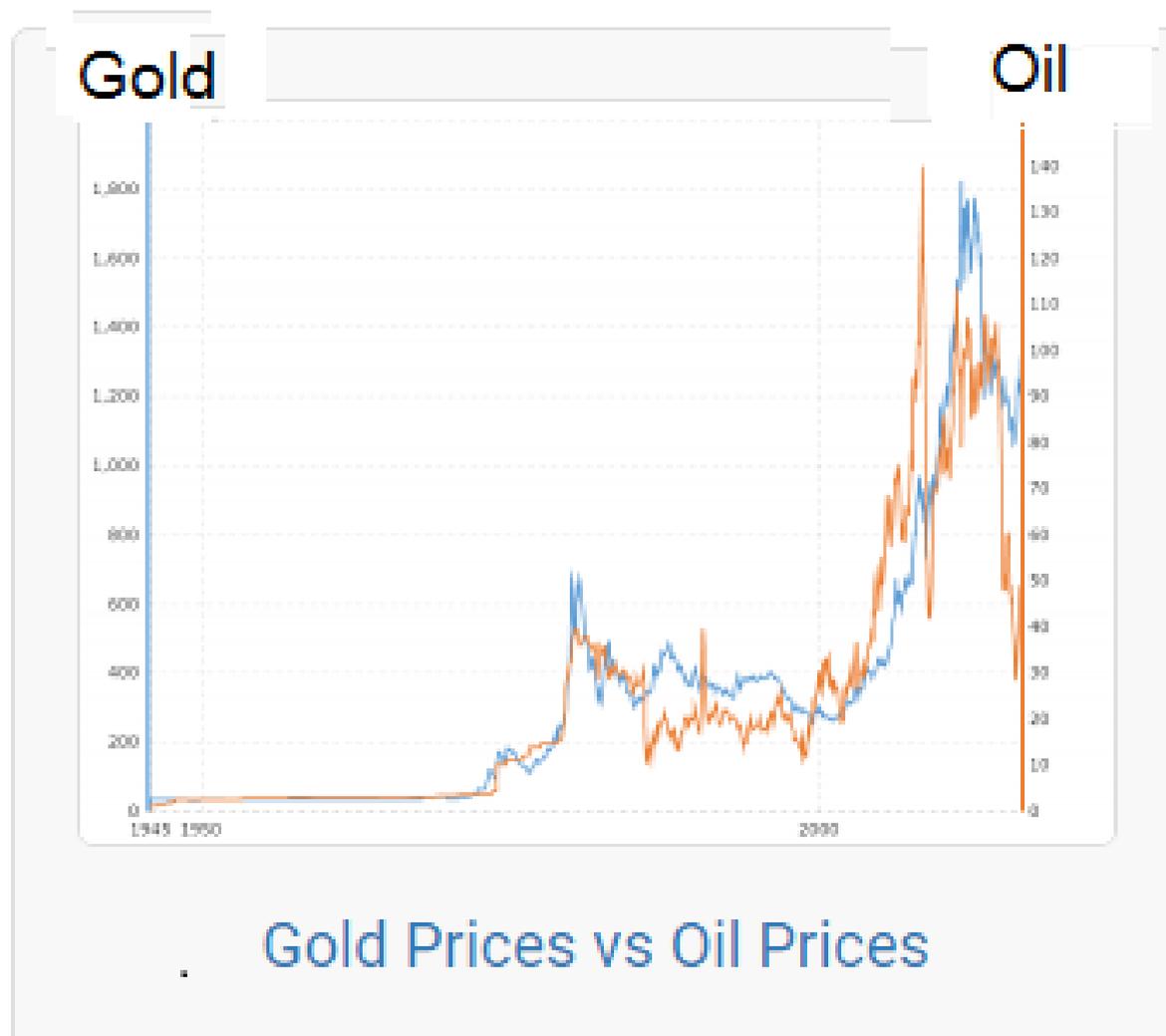


FIGURE 4.1 – Prix de l’Or contre le prix de Pétrole

#### 4.1.1 Etude de la série

Nous avons étudié une série temporelle de deux composantes l’Or et le Pétrole durant la période janvier 1990 à décembre 2022 par le modèle *VARMA*.

Tous les calculs sont faits avec le logiciel R. Nous avons utilisé en particulier les packages *MTS* (Multivariate time series package), le package *VAR* (Vectorial autoregressive package) et le package *tseries* (time series package).

Avant de faire la prévision, nous allons identifier le modèle *VARMA* qui modélise notre série.

#### 4.1.2 Modélisation par VARMA

Le modèle *VARMA* peut être utilisé pour prévoir les valeurs futures de (Or, Pétrole). Il peut également être utilisé pour analyser la relation entre le prix de l’Or et le prix

du Pétrole.

$$\begin{bmatrix} Or(t) \\ Pétrole(t) \end{bmatrix} = c + \phi_1 \begin{bmatrix} Or(t-1) \\ Pétrole(t-1) \end{bmatrix} + \dots + \phi_p \begin{bmatrix} Or(t-p) \\ Pétrole(t-p) \end{bmatrix} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4.1)$$

## 4.2 Etapes de la modélisation

L'étude de la série du prix de L'Or et le Pétrole par le *VARMA* passe par ces étapes :

1. **Préparation des données** : La première étape consiste à collecter les données que nous souhaitons modéliser. Les données doivent être présentées sous forme de séries chronologiques, chaque observation représentant un seul instant.
2. **Test de stationnarité** : L'étape suivante consiste à vérifier si les données sont stationnaires en utilisons le test *ADF* (Augmented Dickey-Fuller)

**Hypothèse nulle** : La série temporelle n'a pas de racine unitaire (stationnaire).

**Hypothèse alternative** : La série temporelle a une racine unitaire (non stationnaire).

3. **Identification de l'ordre** : L'étape suivante consiste à sélectionner l'ordre  $(p, q)$  du modèle *VARMA*. L'ordre du modèle fait référence au nombre de paramètres autorégressifs  $(\phi_1, \dots, \phi_p)$  et de moyennes mobiles  $(\theta_1, \dots, \theta_q)$  (Voir la formule (3.1)). Il peut être choisi à l'aide de diverses méthodes, telles que le critère d'information d'Akaike (*AIC*) (voir la formule(2.4)) et le critère d'information bayésien (*BIC*)(voir la formule(2.3)), ou les autres critères similaires à savoir *HQIC* (Hannan-Quinn Information Criterion),(voir la formule(2.5)) *AICC* (Corrected Akaike Information Criterion)(voir la formule(2.6)), *CAIC* (Consistent Akaike Information Criterion)(voir la formule(2.7)).
4. **Estimation des paramètres du modèle** : Une fois qu'on sélectionne l'ordre du modèle, on peut estimer les coefficients du modèle. Les coefficients du modèle peuvent être estimés à l'aide de diverses méthodes statistiques, telles que l'estimation du maximum de vraisemblance que nous avons utilisé dans notre étude.
5. **Diagnostics du modèle** : Une fois les coefficients du modèle estimés, on doit vérifier les diagnostics du modèle. Les diagnostics du modèle peuvent nous aider à identifier tout problème lié au modèle, tel qu'une spécification erronée ou un ajustement excessif.
6. **Prévision** : une fois le modèle diagnostiqué, on peut l'utiliser pour prévoir les valeurs futures. La prévision peut être faite en utilisant les coefficients estimés du modèle et les données observées.

$$\begin{aligned} \begin{bmatrix} \hat{O}r(t+l) \\ \hat{P}et\hat{r}ole(t+l) \end{bmatrix} &= \hat{c} + \hat{\phi}_1 \begin{bmatrix} \hat{O}r(t+l-1) \\ \hat{P}et\hat{r}ole(t+l-1) \end{bmatrix} + \dots + \hat{\phi}_p \begin{bmatrix} \hat{O}r(t+l-p) \\ \hat{P}et\hat{r}ole(t+l-p) \end{bmatrix} \\ &+ \hat{\theta}_1 \varepsilon_{t+l-1} + \dots + \hat{\theta}_q \varepsilon_{t+l-q} + \varepsilon_{t+l} \end{aligned} \quad (4.2)$$

7. **Évaluation** : La dernière étape consiste à évaluer les prévisions. La prévision peut être évaluée à l'aide de diverses méthodes, telles que l'erreur quadratique moyenne (MSE) ou l'erreur absolue moyenne (MAE).

### 4.3 Analyse de la série (Or, Pétrole)

#### 4.3.1 Graphe de la série du prix de Pétrole et du prix de l'Or

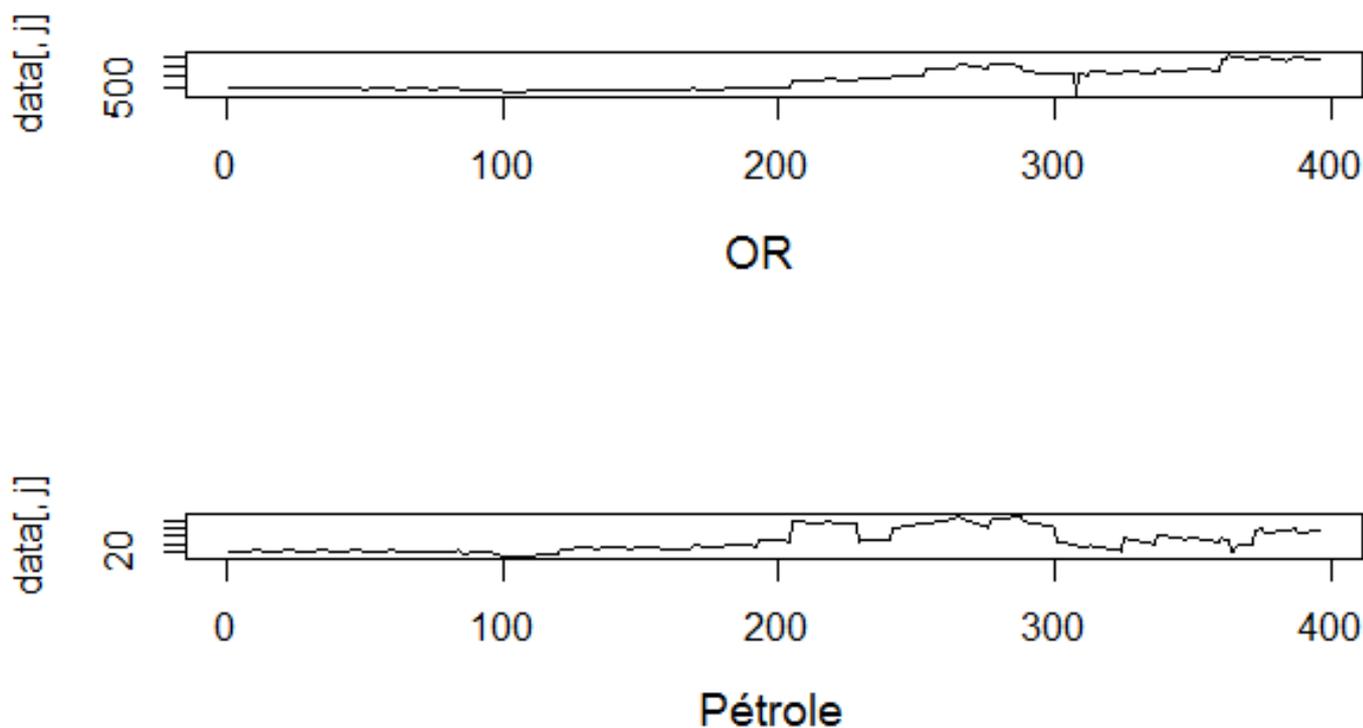


FIGURE 4.2 – Évolution mensuelle du prix de Pétrole et du prix de l'Or de janvier 1990 au décembre 2022.

### 4.3.2 Statistique descriptive de la série du prix de Pétrole et la série du prix de l'Or

TABLE 4.1 – Statistique description de la série (Pétrole, Or)

	Moyenne	Variance	Min	Max
Pétrole	46,79513	823,3732	9.25	109,94
Or	816,7209	280237,8	225	2075,4

### 4.3.3 Correlation :

La matrice de corrélation  $\rho(Petrole, Or)$ , entre l'Or et le Pétrole est donnée par :

$$\rho(Petrole, Or) = \begin{pmatrix} 1 & 0,6763511 \\ 0,6763511 & 1 \end{pmatrix}$$

La corrélation entre le prix de l'Or et celui du Pétrole est établie à 0,6763511, dépassant le seuil de 0,5. Cela confirme la corrélation entre les deux prix, soutenant ainsi les affirmations des institutions financières selon lesquelles cette corrélation se situe autour de 80%.

### 4.3.4 Stationnarité :

L'application du test de Dukkey Fuller augmenté (ADF) a donné le  $P - value = 0.4664 > 0.05$ , ce qui signifié qu'on rejette l'hypothèse que le polynome  $\Phi(z)$  admet une racine unitaire. Donc, on peut conclure que la série est stationnaire.

### 4.3.5 Sélection de l'ordre :

TABLE 4.2 – Sélection de l'ordre par les critères AIC et BIC.

Critères Modèle	VARMA (1,1)	VARMA(2,1)	VARMA(2,2)	VAR(1,0)	VAR (2,0)	VAR(3,0)
AIC	12,34	12,39	12,38	12,44	12,35	12,36
BIC	12,42	12,51	12,54	12,48	12,43	12,49

Les valeurs minimales d'AIC et BIC correspondent au modèle VARMA(1,1). Le deuxième modèle qui a AIC et BiC minimale est modèle VARMA(1,0) ou bien VAR(1). On peut retenir le modèle VARMA(2,0) ou bien VAR(2), car son AIC et son BIC n'est pas loin de celle de VARMA(1,0) ou bien VAR(1). Les résultats sont présentés dans le tableau (4.3)

TABLE 4.3 – Trois premiers modèles sélectionnés par critères AIC et BIC

Critères Modèle	VAR(1,0)	VAR (2,0)	VARMA(1,1)
AIC	12,44	12,35	12,34
BIC	12,48	12,43	12,42

d'après le tableau (4.2) il y a trois valeurs minimales d'AIC et BIC le meilleur modèle qu'on peut choisit est VARMA(1,1)

### 4.3.6 Modèles sélectionnés

Nous avons retenu trois modèles VARMA qui ont donné valeurs des AIC et BIC minimale.

1. **Modèle VARMA(1,1)** : Le modèle VARMA(1,1) peut s'écrire comme suit :

$$\begin{bmatrix} Or(t) \\ Petrole(t) \end{bmatrix} = \begin{bmatrix} 1,00331 & 0,00781 \\ 0,00134 & 0,97531 \end{bmatrix} \begin{bmatrix} Or(t-1) \\ Petrole(t-1) \end{bmatrix} + \begin{bmatrix} 0,4635 & -3,7024 \\ 0,0115 & -0,0549 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1} \\ \varepsilon_t \end{bmatrix}$$

2. **Modèle VARMA(1,0)** : la formule du modèle VARMA(1,0) est la suite :

$$\begin{bmatrix} Or(t) \\ Petrole(t) \end{bmatrix} = \begin{bmatrix} 0,982594 & 0,333 \\ 0,000877 & 0,983 \end{bmatrix} \begin{bmatrix} Or(t-1) \\ Petrole(t-1) \end{bmatrix} + \varepsilon_t$$

3. **Modèle VARMA(2,0)** : Le modèle VARMA(2,0) peut s'écrire comme suit :

$$\begin{bmatrix} Or(t) \\ Petrole(t) \end{bmatrix} = \begin{bmatrix} 0,60538 & 4,20 \\ -0,000832 & 1,02 \end{bmatrix} \begin{bmatrix} Or(t-1) \\ Petrole(t-1) \end{bmatrix} + \begin{bmatrix} 0,3887 & -4,0372 \\ 0,0094 & -0,0407 \end{bmatrix} \begin{bmatrix} Or(t-2) \\ Petrole(t-2) \end{bmatrix} + \varepsilon_t$$

### 4.3.7 Pr evision

Finalement, apr es avoir test e le mod ele estim e et test e sa validit e, nous pouvons pr edire les prix de l'Or et de P etrole pour cette ann ee . Nous obtenons les pr evisions suivantes :

TABLE 4.4 – Pr evision des prix de l'Or et du prix de P etrole de Janvier au D ecembre 2023 par le mod ele VARMA(1,1)

Mois	Jan	F�ev	Mar	Avr	Mai	Juin	jui	Ao�ut	Sept	Oct	Nov	D�ec
OR	1800	1803	1807	1810	1813	1817	1820	1823	1827	1830	1834	1837
P�etrole	75,43	75,91	76,47	77,03	77,57	78,10	78,63	79,15	79,66	80,16	80,65	81,14

TABLE 4.5 – Pr evision des prix de l'Or et du P etrole de Janvier au D ecembre 2023 par le mod ele VARMA(1,0)

Mois	Jan	F�ev	Mar	Avr	Mai	Juin	jui	Ao�ut	Sept	Oct	Nov	D�ec
OR	1772	1763	1755	1746	1738	1730	1722	1714	1706	1699	1692	1684
P�etrole	74,54	74,57	74,60	74,62	74,63	74,64	74,64	74,63	74,62	74,60	74,58	74,56

TABLE 4.6 – Pr evision des prix de l'Or et du P etrole de Janvier au D ecembre 2023 par le mod ele VARMA(2,0)

Mois	Jan	F�ev	Mar	Avr	Mai	Juin	jui	Ao�ut	Sept	Oct	Nov	D�ec
OR	1788	1785	1785	1783	1782	1782	1781	1780	1779	1778	1777	1776
P�etrole	74,79	74,91	75,08	75,23	75,38	75,52	75,65	75,78	75,91	76,03	76,14	76,25

Nous remarquons que les prévisions du prix de l'Or des cinq premiers mois de l'année 2023, est acceptable par le modèle  $VARMA(1,1)$  et le modèle  $VARMA(2,0) = VAR(2)$ . Par contre le modèle  $VARMA(1,0) = VAR(1)$  ne donne pas une bonne prévision (Voir le graphe 4.3).

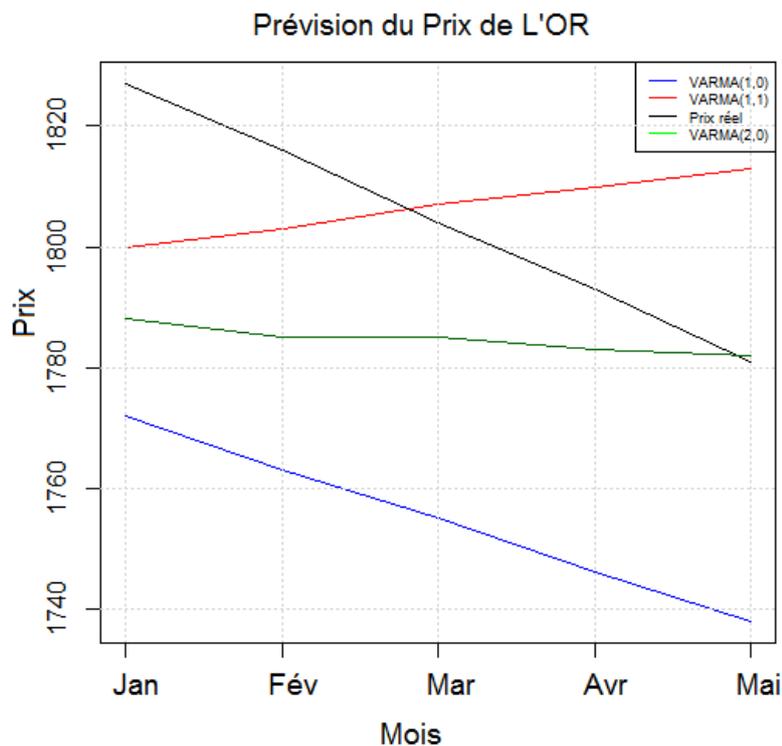


FIGURE 4.3 – Prévision du prix de l'Or des 5 premiers mois de 2023

Les prévisions des prix du pétrole par les trois modèles sont systématiquement sous-estimées (Voir le graphe 4.4), ce qui soulève des interrogations quant à leur précision. Une explication possible réside dans le constat que la variance des prix du pétrole n'est pas constante, ce qui contredit une hypothèse fondamentale nécessaire pour utiliser le modèle VARMA (Vector Autoregressive Moving Average). Cette hypothèse suppose en effet une variance stable des données sur lesquelles le modèle est appliqué.

La volatilité des prix du pétrole est influencée par de nombreux facteurs complexes et changeants, tels que les tensions géopolitiques, l'offre et la demande mondiales, les politiques gouvernementales, ainsi que d'autres variables économiques. Cette volatilité accrue rend difficile la prédiction précise des prix à l'aide de modèles qui supposent une variance constante.

Pour mieux appréhender cette variabilité, il serait intéressant de considérer des modèles plus sophistiqués qui intègrent explicitement la notion de volatilité hétéroscédastique, tels que les modèles ARCH (AutoRegressive Conditional Heteroscedasticity) ou GARCH (Generalized AutoRegressive Conditional Heteroscedasticity). Ces modèles permettent de prendre en compte les variations de la volatilité au fil du temps, offrant ainsi une meilleure adéquation aux caractéristiques complexes des données de prix du pétrole.

En résumé, les prévisions sous-estimées des prix du pétrole par les modèles examinés peuvent être attribuées en partie à une variance non constante des prix. Pour obtenir des prévisions plus précises, il est recommandé d'explorer des modèles qui tiennent compte de cette variabilité, tels que les modèles ARCH ou GARCH.

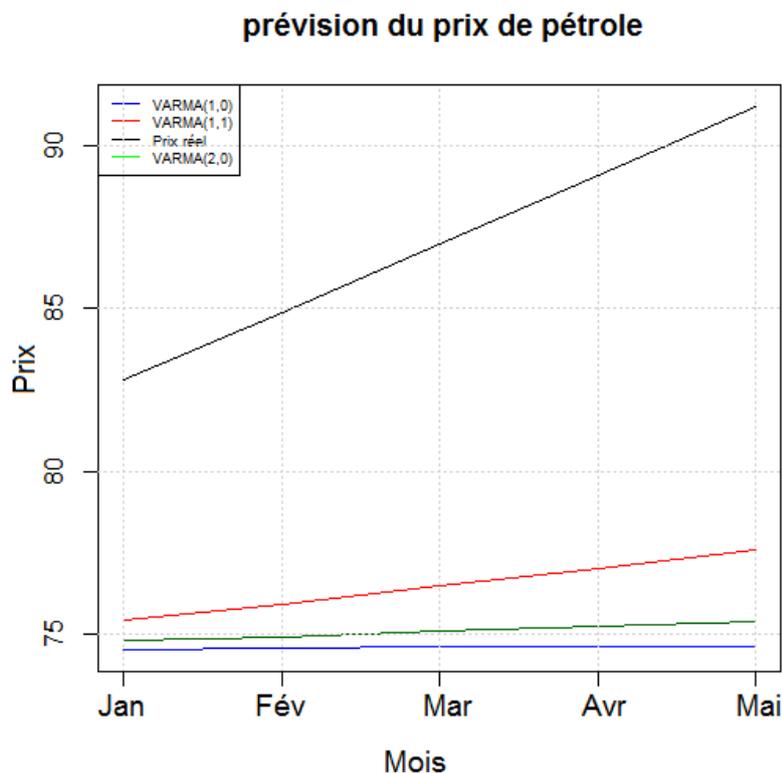


FIGURE 4.4 – Prévision du prix du Pétrole des 5 premiers mois de 2023

## 4.4 Conclusion générale

L'utilisation du modèle VARMA (Vector Autoregressive Moving Average) pour prédire le prix du pétrole et de l'or au niveau mondial offre une approche prometteuse pour les analystes et les investisseurs. Ce modèle est basé sur l'idée que les variables économiques et financières sont interconnectées et influencent mutuellement leurs comportements au fil du temps.

L'une des principales forces du modèle VARMA réside dans sa capacité à capturer les relations complexes et dynamiques entre le prix du pétrole, le prix de l'or et d'autres variables pertinentes.

En utilisant des séries temporelles passées, le modèle peut fournir des estimations probabilistes des prix futurs, ce qui peut être extrêmement précieux pour la prise de décision éclairée dans les marchés mondiaux.

Cependant, il est important de noter que la prédiction des prix des matières premières, en particulier le pétrole et l'or, est un défi complexe.

Ces marchés sont influencés par une multitude de facteurs, tels que l'offre et la demande mondiales, les fluctuations économiques, les événements géopolitiques, les politiques gouvernementales, les conditions météorologiques, etc. Par conséquent, malgré les avantages du modèle VARMA, il est difficile d'obtenir des prévisions précises à long terme.

De plus, il est essentiel de prendre en compte les limites et les hypothèses du modèle VARMA lors

de son utilisation. Par exemple, il suppose que les données sont stationnaires, que les relations entre les variables sont linéaires et que les résidus sont indépendants et identiquement distribués. Si ces hypothèses ne sont pas respectées, les prédictions du modèle peuvent être biaisées ou inefficaces.

Il convient également de souligner que les marchés financiers et les prix des matières premières sont influencés par des facteurs imprévisibles et souvent volatils.

Les événements géopolitiques soudains, les catastrophes naturelles, les crises économiques ou les changements majeurs dans les politiques monétaires peuvent avoir un impact significatif sur les prix du pétrole et de l'or, rendant ainsi les prédictions basées sur les modèles VARMA moins fiables.

En conclusion, l'utilisation du modèle VARMA pour prédire le prix du pétrole et de l'or au niveau mondial offre une approche intéressante et potentiellement utile pour les analystes et les investisseurs.

Cependant, il est important de reconnaître les limites du modèle et de compléter les prévisions avec d'autres informations et analyses pour obtenir une perspective globale plus solide.

Les décisions d'investissement doivent toujours être prises avec prudence et en tenant compte de la nature complexe et incertaine des marchés financiers et des matières premières.

Le modèle VARMA présente certaines limitations, notamment :

**Non-modélisation de l'hétéroscédasticité :** Le modèle VARMA ne prend pas explicitement en compte la possibilité de volatilité conditionnelle dans les données, ce qui peut être problématique lorsque les séries temporelles présentent une variance non constante.

**Dépendance temporelle :** Le modèle VARMA suppose une dépendance linéaire entre les variables sur toute la période d'observation, ce qui peut être limitant lorsque la dépendance entre les variables évolue au fil du temps.

**Prévision de la volatilité :** Le modèle VARMA ne permet pas de fournir directement des prévisions de la volatilité future, ce qui est important pour de nombreuses applications financières telles que la gestion des risques ou le trading.

## Bibliographie

- Hirotsugu Akaike. *Model Selection and Akaike's Information Criterion (AIC). The General Theory and Its Analytical Extensions*, 1974.
- Alexander. *Market Models. A Guide to Financial Data Analysis*. Wiley., 2001.
- Asmussen, S. *Stochastic Simulation. Algorithms and Analysis*. Springer., (2007).
- C. M Bishop. *Pattern Recognition and Machine Learning*. Springer., 2006.
- Brockwell, P. J. *Introduction to Time Series and Forecasting (3rd ed.)*. Springer., 2016.
- Brooks. *BDR92*. Cambridge University Press, 2014a.
- Brooks. *Introductory Econometrics for Finance (3rd ed.)*. Cambridge University Press, 2014b.

Chatterjee S. amp ; Dutta S. Chaudhuri, S. ri, s., chatterjee, s., amp ; dutta, s. (2022). enhancing the performance of support vector machines using harmony search optimization algorithm, 2022.

Cutler A. amp ; Stevens J. R. Cutler, D. R. Random forests in r. documentation de la librairie random forest, disponible à l'adresse, 2007.

Andrew V de Paul S.P. Cowpertwait. *Introductory Time Series with R*. Metcalfe (2e édition), 2019.

Cheryl L Douglas C. Montgomery. *Introduction to Time Series Analysis , Forecasting*. Jennings et Murat Kulahci, 2015.

Robert F. Engle. *ECONOMETRIC ANALYSIS*. Newton-Raphson, 1982.

G. S. Fishman. *Monte Carlo*. Concepts, Algorithms, and Applications. Springer., 2013.

amp ; Zakoïan J. r Francq, C. *GARCH Models*. Structure, Statistical Inference and Financial Applications. Wiley., 2010.

Carlin J. B. Stern H. S. Dunson D. B. Vehtari A. amp ; Rubin D. B. Gelman, A. *Bayesian Data Analysis (3rd ed.)*. CRC Press., 2013.

Bengio Y. amp ; Courville A. Goodfellow, I. *Deep Learning*. MIT Press., 2016.

Allan Timmermann Graham Elliott. *Information Criteria and Statistical Modeling*. Information Criteria and Statistical Modeling, 2004.

Clive Granger. *Clive Granger*. Clive Granger.

William H. Greene. *ECONOMETRIC ANALYSIS*. New York University, 2005.

J. D. Hamilton. *Time Series Analysis*. Princeton University Press., (2021).

Kavalieris Hannan. *ECONOMETRIC ANALYSIS*. Stock et Watson, 1984.

Rissanen Hannan. *Testing causality between two vectors in multivariate autoregressive moving average models*. Stock et Watson, 1982.

Rissanen Hannan. *Testing causality between two vectors in multivariate autoregressive moving average models*. Stock et Watson, 1983.

Tibshirani R. amp ; Friedman J. Hastie, T. *The Elements of Statistical Learning*. Data Mining, Inference, and Prediction. Springer., 2009.

S. Haykin. *Neural Networks*. A Comprehensive Foundation. Prentice Hall., 2016.

- S. (1988) Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3), 231-254, 1988.
- Phillips P. C. Schmidt P. amp ; Shin Y Kwiatkowski, D. Testing the null hypothesis of stationarity against the alternative of a unit root, 1992.
- Reinsel Lewis. *ECONOMETRIC ANALYSIS*. Newton-Raphson, 1985.
- H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer., 2006.
- K. P. Murphy. *Machine Learning*. A Probabilistic Perspective. MIT Press., 2012.
- amp ; Foody G. M. Pal, M. Feature selection for classification of hyperspectral data by svm, 2008.
- Poskitt. *ECONOMETRIC ANALYSIS*. Poskitt, 1987.
- Basu et Yap Reinsel. *ECONOMETRIC ANALYSIS*. Newton-Raphson, 1992.
- David Stoffer Robert H. Shumway. *With R Examples*quot. Time Series Analysis , Its Applications., 2015.
- David S Robert H.Shumway. *Time Series Analysis, Its Applications*. Stoffer (4e édition)., 2021.
- Eugen Slutsky. *ECONOMETRIC ANALYSIS*. New York University, 1927.

Boulesteix A. L. Zeileis A. amp; Hothorn T. Strobl, C. Bias in random forest variable importance measures, 2007.

R. S. Tsay. *Analysis of Financial Time Series (3rd ed.)*. Wiley, (2013).

Zhang C. Li X. amp; Liu W. Wang, L. A novel hierarchical decision tree framework for support vector machine, (2022).

Huang H. Peng L. amp; Yin Y. Yan, Z. Support vector machine with quaternion kernels for classification of hyperspectral images, 2022.