

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université M'hamed BOUGUERA Boumerdes  
Faculté des sciences



### **Mémoire**

Présenté pour l'obtention du diplôme de Master en  
Mathématiques Financières

## **Étude prévisionnelle de l'activité de l'exploration pétrolière en Algérie**

**Réalisé par :**

**TAMAZIRT Amira**

**CHETTIR Safia**

Soutenue le 08/07/2023 devant le jury composé de :

Mme AIT HOCINE.F      MAA      U.M.B.B      Président

Mme. GATT.F              MAA      U.M.B.B      Promotrice

Mme. BENMANSOUR.M    MAA      U.M.B.B      Examinatrice

**Année universitaire 2022/2023**

### **Remerciements**

Avant tout, nous remercions Dieu le tout puissant de nous avoir donné la force d'accomplir ce travail et de le mener jusqu'au bout

Nous tenons à exprimer toute nos gratitude envers notre promotrice Mme GATT Fella pour le soutien, l'aide et les conseils qu'elle nous a dispensé pour l'élaboration du présent mémoire de Master.

Nous remercions le président et les membres du jury qui nous feront l'honneur De juger notre travail.

Nous remercions respectueusement les cadres SONATRACH.

Et nous voudrions aussi remercier Mr BOUMEDINE Nabil et mlle DISSAOUI Soraya pour leurs aide et ses conseils.

Nous remercions nos très chers parents, nos frères et soeurs, qui ont toujours été là pour nous, merci pour leurs soutien constant et leurs encouragements.

Nous remercions également toutes personnes ayant contribué, de près et ou de loin, à la réalisation de ce projet, qui nous ont apporté leur attention, leurs encouragements, leur appui et leur assistance pour que nous puissions mener à bien ce travail.

# Dédicaces

Je dédie les résultats de ce mémoire

A mes parents qui me sont très chers, qui m'ont toujours donné espoir dans cette vie et qui m'ont entouré de leur soutien et leur affection, que dieu me les garde ;

A ma chère sœur : Imene et son mari ;

A mes frères : Khaled et Noradine ;

A mon binôme ;

A mes amis, Zineb, Imene, Cilya, Lamia, Radia, Fatoma ;

A toute la promotion 2022 / 2023 de mathématique financière ;

**Amira.**

# Dédicaces

Je dédie les résultats de ce mémoire

A mes parents qui me sont très chers, qui m'ont toujours donné espoir dans cette vie et qui m'ont entouré de leur soutien et leur affection, que dieu me les garde ;

A mes chères sœurs : Kahina , Linda, Wahiba, Zazi;

A mes frères : Wassim , Farid et sa femme Warda ;

A mon binôme ;

A mes amis, Nadjat, Imene, Hiba ;

A toute la promotion 2022 / 2023 de mathématique financière ;

**Safia.**

## **Résumé**

L'équilibre entre les réserves potentielles d'hydrocarbures et la production annuelle est crucial pour les pays producteurs d'énergie. L'Algérie, riche en hydrocarbures, cherche à renforcer son secteur pétrolier en collaborant avec des partenaires internationaux. Face à la demande croissante, SONATRACH vise à doubler son taux de production. Les prévisions dans l'exploration pétrolière jouent un rôle clé en anticipant les tendances futures, évaluant les risques et prenant des décisions éclairées pour maximiser les réserves et assurer la durabilité du secteur.

## **Abstract**

The balance between potential hydrocarbon reserves and annual production is crucial for energy-producing countries. Algeria, rich in hydrocarbons, seeks to strengthen its oil sector by collaborating with international partners. With increasing demand, SONATRACH aims to double its production rate. Forecasts in oil exploration play a key role in anticipating future trends, assessing risks, and making informed decisions to maximize reserves and ensure the sector's sustainability.

# Table des matières

<b>Introduction</b>	<b>8</b>
<b>1 Présentation de l'organisme d'accueil SONATRACH</b>	<b>10</b>
1.1 Présentation de l'entreprise . . . . .	10
1.1.1 Organigramme de la macrostructure de SONATRACH : . . . . .	11
1.1.2 Division Exploration . . . . .	12
1.2 Définitions et notions de base sur les hydrocarbures . . . . .	14
<b>2 Théorie des séries temporelles et la méthodologie de Box &amp; Jenkins</b>	<b>16</b>
2.1 Définition et notions de base . . . . .	16
2.1.1 Série chronologique . . . . .	16
2.1.2 Composantes d'une série chronologique . . . . .	17
2.1.3 Caractéristiques d'une série chronologique . . . . .	18
2.1.4 Opérateurs B et $\Delta$ . . . . .	20
2.1.5 Processus aléatoire . . . . .	21
2.2 Méthodologie de Box & Jenkins . . . . .	26
2.2.1 Test des composantes saisonnières et tendanciennes . . . . .	27
2.2.2 Etapes de la méthodologie de Box & Jenkins . . . . .	29
2.2.3 Prévision . . . . .	33
<b>3 Application de la méthodologie de BOX &amp; JENKINS</b>	<b>35</b>
3.1 Etude de la série des investissements financier pour l'exploration $INV_t$ . . . . .	35
3.1.1 Identification . . . . .	35
3.1.2 Estimation des paramètres . . . . .	42

3.1.3	Validation du modèle . . . . .	44
3.1.4	Prévisions . . . . .	48
3.2	Etude de la série prix du pétrole $PRIX_t$ . . . . .	49
3.2.1	Identification . . . . .	49
3.3	Etude de la série des volumes de découvertes (VD) . . . . .	51
3.3.1	Identification . . . . .	51
3.3.2	Estimations des paramètres . . . . .	53
3.3.3	Prévisions . . . . .	57
3.4	Etude de la série prix du pétrole $SIS3D_t$ . . . . .	58
3.4.1	Identification . . . . .	58
3.4.2	Estimation des paramètres . . . . .	61
3.4.3	Prévisions . . . . .	65
<b>4</b>	<b>Approche de la cointégration</b> . . . . .	<b>67</b>
4.1	Approche de la cointégration bivariée . . . . .	67
4.1.1	Cointégration et les modèles à correction d'erreur . . . . .	67
4.1.2	Estimation des modèles a correction d'erreur et tests de cointégration (au sens d'Engel et Granger) . . . . .	69
4.2	Approche multivariée de la cointégration : l'analyse de Johansen . . . . .	72
<b>5</b>	<b>Application de l'approche de cointégration</b> . . . . .	<b>79</b>
5.1	Cointégration bivariée . . . . .	80
5.1.1	Cointégration entre l'investissement ( $INV_t$ ) et le volume de découvertes ( $VD_t$ ) . . . . .	80
5.1.2	Test de causalité au sens de Granger . . . . .	86
5.1.3	Cointégration entre le volume de découvertes ( $VD_t$ ) et sismique trois dimension $SIS3D_t$ . . . . .	87
5.2	Cointégration multivariée . . . . .	92
5.3	Comparaison entre la méthodologie de Box-Jenkins et de la méthodologie de cointégration en terme de prévision . . . . .	101
	<b>Conclusion</b> . . . . .	<b>102</b>

# Liste des tableaux

3.1	Choix du modèle . . . . .	44
3.2	Les prévisions de INV . . . . .	48
3.3	Les prévisions de VD . . . . .	58
3.4	Les prévisions de SIS3D . . . . .	65
5.1	Choix de retard P . . . . .	86
5.2	Prevision de cointégration INV et VD . . . . .	87
5.3	Prevision de cointégration VD et SIS3D . . . . .	92
5.4	choix du nombre de retard p . . . . .	94
5.5	Prevision de cointégration multivarié . . . . .	100

# Table des figures

1.1	Source : Direction générale de SONATRACH . . . . .	11
1.2	Organigramme de La Division Exploration . . . . .	13
2.1	Diagramme de box & Jenkins . . . . .	27
3.1	Graphe de la série $INV_t$ . . . . .	36
3.2	Corrélogramme de la série $INV_t$ . . . . .	36
3.3	ADF modèle 3 de la série $INV_t$ . . . . .	37
3.4	ADF modèle [2] de la série $INV_t$ . . . . .	38
3.5	ADF modèle [1] de la série $INV_t$ . . . . .	39
3.6	ADF modèle [3] de la série $DINV_t$ . . . . .	40
3.7	ADF modèle [2] de la série $DINV_t$ . . . . .	41
3.8	Corrélogramme de la série $DINV_t$ . . . . .	42
3.9	Estimation des paramètres du modèle ARIMA(1,1,1) . . . . .	42
3.10	Estimation des paramètres du modèle ARIMA(1,1,0) . . . . .	43
3.11	Estimation des paramètres du modèle ARIMA(0,1,1) . . . . .	43
3.12	Racines inverses du(des) polynôme(s) MA/AR . . . . .	44
3.13	Graphe des séries résiduelles actuelles et estimée . . . . .	45
3.14	Corrélogramme des résidus . . . . .	45
3.15	L'Histogramme . . . . .	46
3.16	Table d'effet ARCH . . . . .	47
3.17	Graphe de prévision . . . . .	48
3.18	Evolution de la série $PRIX_t$ . . . . .	49
3.19	Corrélogramme de la série $PRIX_t$ . . . . .	49
3.20	Corrélogramme de la série $DPRIX_t$ . . . . .	50

3.21	Évolution de la série $VD_t$ . . . . .	51
3.22	Corrélogramme de la série $VD_t$ . . . . .	52
3.23	Corrélogramme de la série $DVD_t$ . . . . .	53
3.24	Estimation des paramètres du modèle ARIMA(1,1,1) . . . . .	53
3.25	Racines inverses du(des) polynôme(s) MA/AR . . . . .	54
3.26	Graphe des séries résiduelles actuelles et estimées . . . . .	54
3.27	Corrélogramme des résidus . . . . .	55
3.28	Histogramme de normalité . . . . .	56
3.29	Table d'effet Arch . . . . .	57
3.30	Graphe de prévision . . . . .	58
3.31	Evolution de la série $SIS3D_t$ . . . . .	59
3.32	Corrélogramme de la série $SIS3D_t$ . . . . .	59
3.33	Corrélogramme de la série $DSIS3D_t$ . . . . .	60
3.34	Estimation de paramètre ARIMA (1,1,1) . . . . .	61
3.35	Racines inverses du(des) polynôme(s) MA/AR . . . . .	61
3.36	Graphe des séries résiduelles actuelles et estimées . . . . .	62
3.37	Corrélogramme des résidus . . . . .	63
3.38	Histogramme de normalité . . . . .	63
3.39	Teste d'effet ARCH . . . . .	64
3.40	Graphe de prévision . . . . .	65
5.1	Evolution de $INV_t$ et $VD_t$ . . . . .	80
5.2	Estimation de la relation $INV_t$ et $VD_t$ . . . . .	81
5.3	Test ADF des résidus $\hat{Z}_t$ . . . . .	82
5.4	Corrélogramme des résidus $\hat{Z}_t$ . . . . .	82
5.5	Histogramme des résidus $\hat{Z}_t$ . . . . .	83
5.6	Test d'effet ARCH . . . . .	83
5.7	Estimation du Modèle à Correction d'Erreurs . . . . .	84
5.8	Graphe des séries réelles estimées et résiduelles . . . . .	85
5.9	Corrélogramme des residus $\varepsilon_t$ . . . . .	86
5.10	Test de causalité au sens de Granger . . . . .	86
5.11	Evolution de $VD_t$ et $SIS3D_t$ . . . . .	88
5.12	Estimation de la relation entre $VD_t$ et $SIS3D_t$ . . . . .	88
5.13	Corrélogramme des résidus $\hat{Z}_t$ . . . . .	89
5.14	L'estimation du Modèle à Correction d'Erreurs . . . . .	90
5.15	Graphe des séries réelles estimées et résiduelles . . . . .	91
5.16	Corrélogramme des residus $\varepsilon_t$ . . . . .	91
5.17	Graphes des trois séries non stationnaires . . . . .	93

---

5.18	Estimation du modèle VAR(1)	94
5.19	Test de la trace	95
5.20	Test de la valeur maximale	96
5.21	Estimation de la relation de long terme	96
5.22	Estimation du VECM	98
5.23	Test de causalité au sens de Granger	100

# Introduction générale

En matière d'énergie, l'équilibre entre les réserves potentielles d'hydrocarbures et le volume de production annuelle est crucial pour les pays producteurs tel-que l'Algérie. Cet équilibre stratégique repose sur la nécessité que chaque année, les volumes de découvertes de nouveaux gisements couvrent suffisamment les volumes de production, afin de maintenir le ratio stratégique du secteur des hydrocarbures.

D'autre part, la hausse des prix du pétrole au cours des dernières années a incité les entreprises à investir davantage dans l'exploration pétrolière, car de nouvelles découvertes permettent de renforcer le volume des réserves et d'augmenter le rythme de production pour maximiser les bénéfices. Dans ce contexte, il est essentiel pour SONATRACH de comprendre en profondeur les mécanismes régissant le domaine de l'exploration pétrolière et les paramètres qui l'influencent.

Dans les années à venir, SONATRACH s'est fixé des objectifs ambitieux, notamment celui de doubler son taux de production. Pour atteindre ces objectifs, la compréhension des prévisions en matière d'exploration pétrolière devient d'une importance capitale, car elles orienteront les décisions stratégiques et les actions à entreprendre pour assurer une croissance durable dans le secteur des hydrocarbures.

En effet, les découvertes d'hydrocarbures, considérées comme le résultat final du processus d'exploration, demeurent un élément clé pour répondre aux défis futurs et assurer la sécurité énergétique de l'Algérie.

Face à cette préoccupation, notre étude s'est basée sur l'objectif d'analyser les séries chronologiques des découvertes en hydrocarbures (pétrole et gaz) de l'Algérie réalisées par SONATRACH seule ou en partenariat ainsi que celle des efforts d'exploration en terme physique et financier. Pour cela, nous disposons des données annuelles sur la période 1980 à 2022 correspondantes à :

L'investissement financier pour l'exploration pétrolière (\$ U.S.);

aux volumes de découvertes. (tes);

aux Sismique réalisés 3D ( $km^2$ );

et au prix du pétrole en millions de dollar.

Notre travail s'articulera donc autour de cinq chapitres :

Dans le premier chapitre nous présentons l'organisme d'accueil (SONATRACH).

Dans le deuxième chapitre nous présentons la théorie des séries temporelles ainsi que la méthodologie de Box et Jenkins. Le troisième chapitre est consacré à l'application du deuxième chapitre.

Le quatrième chapitre sera consacré à la la théorie de cointégration bivarie et multivarie.

En fin nous terminons notre travail par une application sur le chapitre quatre.

# Chapitre 1

## Présentation de l'organisme d'accueil SONATRACH

### introduction

Ce chapitre est consacré à la présentation de l'entreprise SONATRACH. EN particulier, la direction "Exploration". Nous présenterons également des généralités sur les hydrocarbures et l'amont pétrolier utiles pour la compréhension de la suite de notre travail.

### 1.1 Présentation de l'entreprise

SONATRACH est une compagnie étatique algérienne et un acteur international majeur dans l'industrie des hydrocarbures.

Le groupe pétrolier et gazier est classé 1ère en Afrique et 12ème dans le monde en 2013, toutes activités confondues, avec un chiffre d'affaires à l'exportation de plus de 63 milliards de US\$.

Née le 31 décembre 1963, la compagnie intervient dans l'exploration, la production, le transport par canalisations, la transformation et la commercialisation des hydrocarbures et de leurs dérivés. Elle est 4ème exportateur mondial de GNL, 3ème exportateur mondial de GPL et 5ème exportateur de Gaz Naturel.

Adoptant une stratégie de diversification, Sonatrach se développe aussi bien dans les activités de génération électrique, d'énergies nouvelles et renouvelables, de dessalement d'eau de mer, de recherche et d'exploitation minière.

Poursuivant sa stratégie d'internationalisation, Sonatrach opère en Algérie et dans plusieurs régions du monde : Afrique (Mali, Niger, Libye, Egypte), Europe (Espagne, Italie, Portugal, Grande Bretagne), Amérique Latine (Pérou) et USA.



### 1.1.1 Organigramme de la macrostructure de SONATRACH :

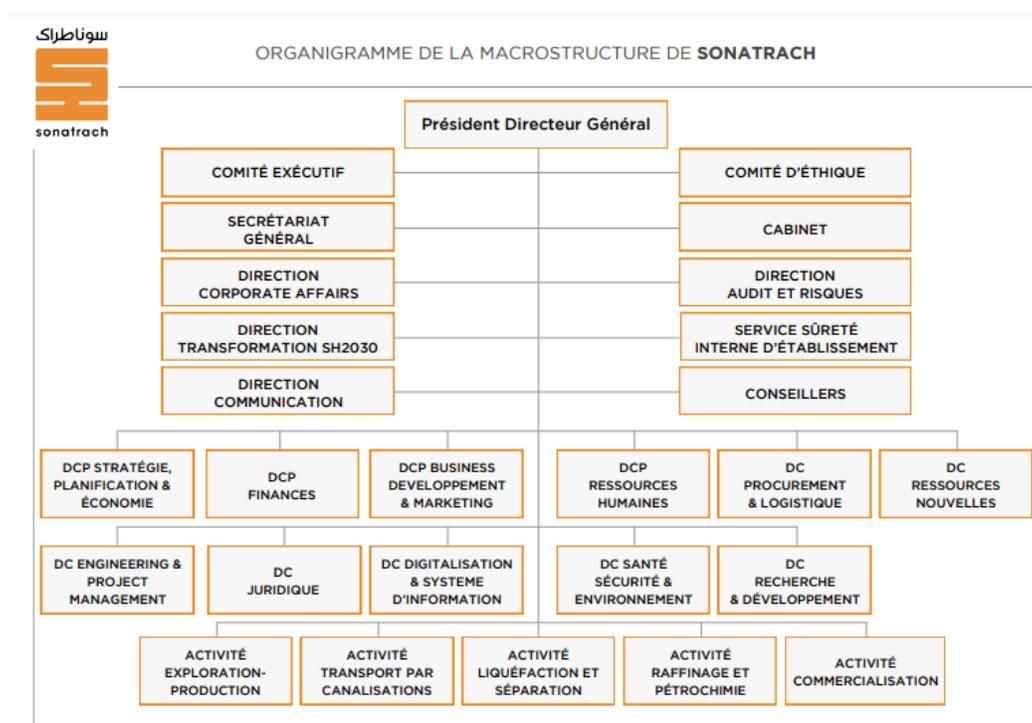


FIGURE 1.1 – Source : Direction générale de SONATRACH

### 1.1.2 Division Exploration

La division exploration fait partie de l'activité amont, elle a été créée en 1972 et elle avait le titre de la direction d'exploration, suite à la réorganisation de la SONATRACH en date du 4/4/1987, la direction exploration est élevée au rang de la division.

L'Activité Exploration-Production(EP) couvre les activités de recherche, d'exploration, de développement et de production d'hydrocarbures. Elles sont assurées par Sonatrach seule, ou en association avec d'autres compagnies pétrolières.



#### Missions de la Division exploration

##### L'organisation d'exploration a pour mission :

- La conduite et le développement des activités de prospection et de recherche des hydrocarbures ;
- La participation avec les autres Divisions aux appels d'offres d'exploration en Algérie et à l'étranger ;
- La participation à l'évaluation des offres de partenariat sur des projets d'exploration en Algérie et à l'étranger ;
- La mise en œuvre de la stratégie de la Société en matière d'exploration ;
- La préparation, l'établissement et la recommandation des programmes techniques d'exploration et leur suivi ;
- Le développement et la conduite des travaux d'analyse en matière de géologie et de géophysique ;
- La gestion et le suivi des contrats en effort propre et en association ;
- Le développement d'expertise dans le domaine de l'exploration,

##### Organigramme de La Division Exploration :

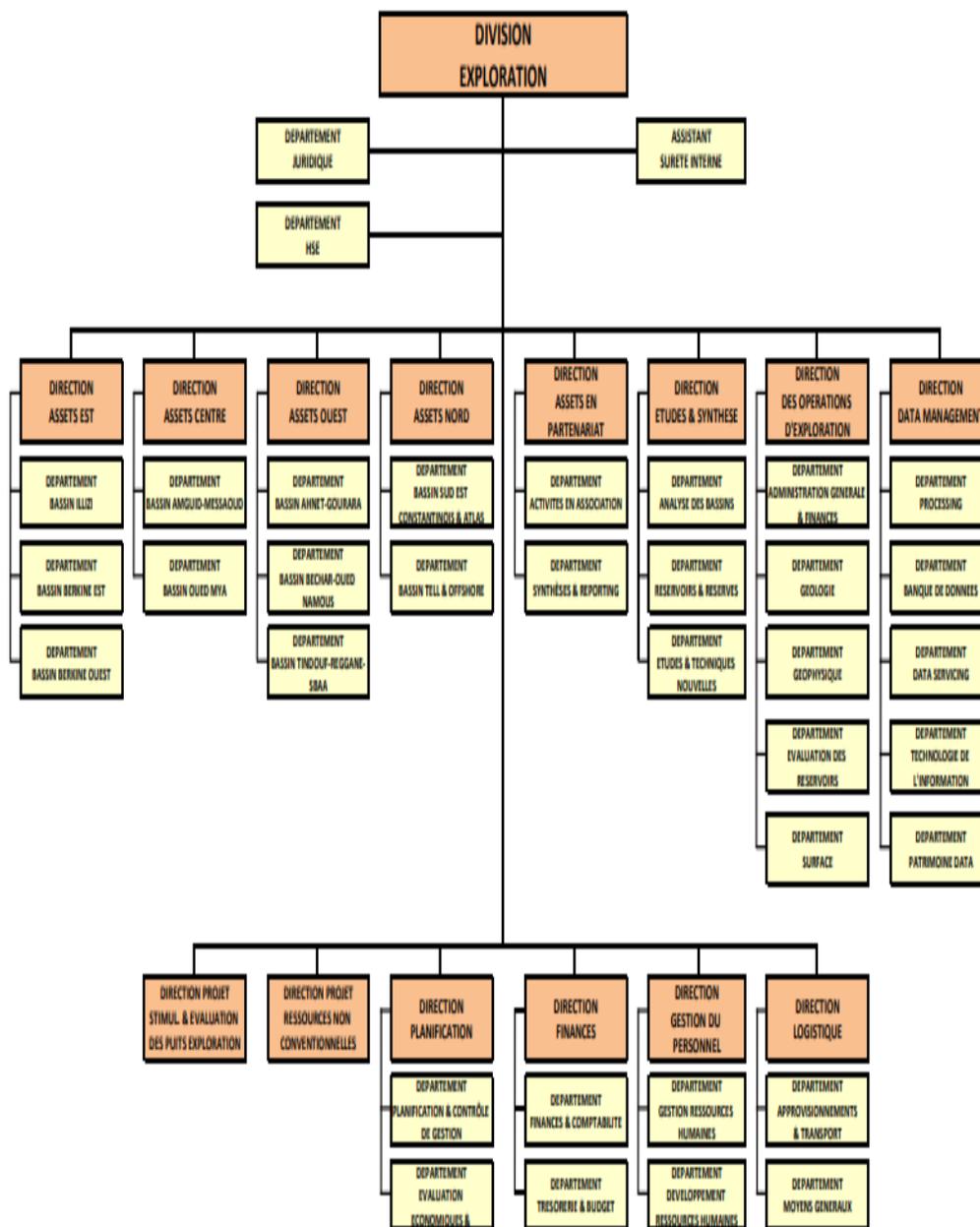


FIGURE 1.2 – Organigramme de La Division Exploration

## **1.2 Définitions et notions de base sur les hydrocarbures**

### **Hydrocarbures**

Les hydrocarbures sont, comme l'indique leur nom, des composés de carbone et d'hydrogène. Ils peuvent être solides, liquides ou gazeux selon que leur teneur est plus ou moins faible en carbone.

### **Gisement**

Un gisement peut être défini comme un réceptacle, constitué d'une roche réservoir perméable, clos à la fois par une couverture formée d'une roche imperméable et par une configuration favorable du sous-sol appelé piège.

### **Les pièges**

Pour qu'il y ait gisement il faut qu'au cours de leurs déplacements à l'intérieur des roches réservoirs où ils se trouvent disséminés, les hydrocarbures aient rencontré des points ou lieux particuliers, appelés « pièges », dans lesquels ils auront été arrêtés et où ils se seront accumulés en quantités plus ou moins considérables. C'est la recherche de ces pièges qui constitue l'objectif principal de la prospection pétrolière.

### **Délinéation**

Après une première étape de forage ayant montré la présence d'hydrocarbures dans une structure en cours d'exploration, la Délinéation est un programme de forages ultérieurs permettant de définir et de délimiter les formations éventuellement productives.

### **Roche réservoir**

Zone d'un gisement où sont accumulés les hydrocarbures (pétrole ou gaz); il peut comporter plusieurs couches productrices successives séparées par des couches stériles.

**Réserve**

Les réserves constituent l'ensemble des ressources disponibles pour les besoins présents et futurs.

**Roche mère**

Les roches mères sont les roches qui contiennent les matières organiques en quantité suffisante. Ce sont surtout des argiles et parfois des carbonates.

**Profil de production**

Evolution, au cours du temps, du niveau de production d'un gisement. Au début de la mise en production d'un gisement, la production augmente fortement avant de se stabiliser quelques années, puis décline progressivement.

**Sismique en trois dimension**

C'est une technique d'imagerie géophysique utilisée dans l'industrie pétrolière et gazière pour visualiser les structures géologiques souterraines. Elle permet d'obtenir des informations détaillées sur la composition, la géométrie et les propriétés des couches de roches situées sous la surface terrestre. Elle offre plusieurs avantages par rapport aux techniques sismiques conventionnelles en deux dimensions (2D).

**Volume de découverte**

Représente la quantité estimée de pétrole nouvellement découvert dans un gisement ou une région. Il indique la taille des réserves potentielles de pétrole identifiées lors des activités d'exploration. Ce volume est exprimé en barils ou en mètres cubes et peut varier selon le stade de l'exploration.

# Chapitre 2

## Théorie des séries temporelles et la méthodologie de Box & Jenkins

### Introduction

L'analyse des séries temporelles repose sur l'utilisation de données historiques collectées sur un phénomène particulier sur une période donnée. Cette méthode vise à trouver un modèle mathématique capable de refléter l'évolution de ces données. Elle se fonde sur l'hypothèse que le passé peut être un indicateur de l'avenir, permettant ainsi de réaliser des prévisions. Dans ce chapitre, nous présenterons les concepts de base et les éléments de l'analyse statistique des séries temporelles nécessaires à notre application.

### 2.1 Définition et notions de base

#### 2.1.1 Série chronologique

Une série chronologique, appelée aussi série temporelle ou chronique, notée  $(X_t, t \in T)$ , correspond à une suite d'observations d'une variable  $X$  à différentes dates  $t$ . Le temps  $T$  est généralement dénombrable, de sorte que  $t = \{t_1, t_2, \dots, t_n\}$ . L'objectif de l'analyse des séries temporelles est d'étudier la dynamique d'une variable, car cela est crucial pour pouvoir établir des liens entre différentes variables ayant des propriétés similaires, notamment en termes de stabilité ou d'instabilité. Les séries temporelles peuvent inclure des données macroéconomiques (telles que la masse monétaire, l'inflation, ... etc.), des données microéconomiques (comme le

nombre d'employés d'une entreprise, les ventes, . . . etc.), des données politiques (comme le nombre de votants ou de votes nuls), des données démographiques (comme l'âge moyen ou la taille des habitants d'une région) ou encore des données financières (comme le cours d'une action) ou climatologiques (telles que la pluviométrie, l'évaporation ou la température). La périodicité de la série importe peu, qu'elle soit annuelle, semestrielle, mensuelle, . . .etc.

Si  $T$  est dénombrable (généralement  $T \in \mathbb{Z}$ ), on parle de série temporelle à temps discret.

Si  $T$  n'est pas dénombrable (généralement un intervalle de  $\mathbb{R}$ ), on parle de série temporelle à temps continu.

### 2.1.2 Composantes d'une série chronologique

Les premières analyses de séries chronologiques visaient à en décomposer les principales caractéristiques. La représentation graphique de l'évolution de la série temporelle permet de mettre en évidence ses différentes composantes, à savoir :

#### **Composante tendancielle ( $T_t$ )**

Elle traduit l'allure générale du phénomène observé et représente son évolution à long terme. Cette tendance peut être croissante, décroissante ou stable (pas de tendance).

#### **Composante cyclique ( $C_t$ )**

Elle représente l'ensemble des fluctuations et variations de la série autour de la tendance sur une longue période (supérieure à un an).

#### **Remarque**

La tendance et le cycle sont parfois regroupés en une seule composante appelée l'extra-saisonnier.

#### **Composante saisonnière ( $S_t$ )**

Cette composante est liée essentiellement au rythme imposé par les saisons. Comme le cycle, la saisonnalité est périodique. Elle est dépendante de l'environnement qui correspond souvent à des phénomènes de mode, de coutume, de climat . . .etc.

### Composante résiduelle ( $R_t$ )

Regroupe tout ce qui n'est pas pris en compte par les trois composantes précédentes. Elle contient de nombreuses fluctuations qui proviennent souvent de circonstances non prévisibles (variations accidentelles : grèves, crises mondiales, catastrophes naturelles). Elle est résiduelle car elle reste quand on élimine toutes les autres composantes.

### 2.1.3 Caractéristiques d'une série chronologique

#### Moyenne et variance

Soit une série temporelle  $\{X_t; t \in Z\}$ ; les expressions de la moyenne et la variance sont :

- Moyenne :

$$E(X_t) = \frac{1}{n} \sum_{t=1}^n X_t.$$

- Variance :

$$Var(X_t) = \frac{1}{n} \sum_{t=1}^n [X_t - E(X_t)]^2.$$

- écart-type :

$$\sigma = \sqrt{VAR(X_t)}$$

#### Fonction d'autoconvariance

Soit  $\{X_t; t \in R\}$  un processus aléatoire de variance finie, on note l'autocovariance de  $X_t$  la fonction  $\gamma_h$  donnée par :

$$\gamma_h = cov(X_t, X_{t+h}) = E[(X_t - E(X_t)) \cdot (X_{t+h} - E(X_{t+h}))]$$

Cette fonction mesure la covariance entre deux variables séparées par un certain délai (h).

Le graphe de cette fonction est appelé variogramme. La fonction d'auto covariance d'un processus stationnaire vérifie :

1.  $\gamma_0 = cov(X_t, X_t) = Var(X_t)$
2.  $\|\gamma_h\| \leq \gamma_0$
3.  $\gamma_h = \gamma_{-h}$  (la fonction est paire)

### Fonction d'autocorrélation simple

La fonction d'autocorrélation simple notée FAC (ACF en Anglais) permet de mesurer les liaisons temporelles entre les différentes composantes de la série stationnaire  $X_t$ . Elle est définie par :

$$\rho_h = \frac{\text{cov}(X_t, X_{t+h})}{\sigma(X_t) \cdot \sigma(X_{t+h})} = \frac{\gamma_h}{\sqrt{\gamma_0} \cdot \sqrt{\gamma_0}} = \frac{\gamma_h}{\gamma_0}$$

La fonction d'autocorrélation d'un processus stationnaire vérifie les propriétés suivantes :

1.  $\rho_0 = 1$
2.  $|\rho_h| \leq \rho_0$
3.  $\rho_0 = \rho_{-h}$  (la fonction est paire)

### Fonction d'auto-corrélation partielle

La fonction d'auto-corrélation partielle notée FAP (PACF en Anglais) mesure la corrélation entre  $X_t$  et  $X_{t-h}$ ,  $h \in \mathbb{N}$  l'influence des variables  $X_{t-h+i}$  (pour  $i < h$ ) ayant été retirés.

Étant donnée la matrice des corrélations symétrique formée des  $(h-1)$  premières auto-corrélations.

$$P_h = \begin{bmatrix} 1 & \rho_1 & \cdots & \cdots & \rho_{h-1} \\ \rho_1 & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \cdots & 1 \end{bmatrix}$$

La fonction d'auto-corrélation partielle est donnée par :

$$\phi_{hh} = \frac{|P_h^*|}{|P_h|}$$

Où  $|P_h|$  est le déterminant de la matrice  $P_h$  et  $|P_h^*|$  est le déterminant de la matrice  $P_h$  en remplaçant la dernière colonne par le vecteur  $(\rho_1, \rho_2, \dots, \rho_h)$

Ainsi

$$P_h^* = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{h-2} & \rho_1 \\ \rho_1 & 1 & & & \rho_2 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \cdots & \rho_h \end{bmatrix}$$

## 2.1.4 Opérateurs B et Δ

### Opérateur de retard B

L'opérateur retard B (Backward noté aussi pour L) décale le processus d'une unité de temps vers le passé  $B(X_t) = X_{t-1}$

Si nous appliquons h fois cet opérateur, nous décalons le processus de h unité de temps

$$B(B(\cdots B(X_t) \cdots)) = B^h(X_t) = X_{t-h}$$

L'opérateur retard possède les propriétés suivantes :

$$B^j X_t = X_{t-j} \text{ en particulier } B^0 X_t = X_t$$

Si  $X_t = C \in R$  pour tout  $t \in Z$ , alors  $B^j X_t = B^j C = C$  pour tout  $j \in Z$

$$B^j (B^k X_t) = B^{j+k} X_t = X_{t-j-k}$$

$$B^{-j} X_t = X_{t+j}$$

$$(B^j + B^k) X_t = B^j X_t + B^k X_t = X_{t-j} + X_{t-k}$$

Si nous appliquons le polynôme retard  $\phi(B)$  défini comme suit :

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

à une série  $X_t$ , nous obtenons :

$$\phi(B) X_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p}$$

Où  $\phi_i$  des nombres réels.

### Opérateur d'avance F

L'opérateur d'avance F (Forward) décale le processus d'une unité de temps vers le futur  $F(X_t) = X_{t+1}$ .

Si nous appliquons h fois cet opérateur, nous décalons le processus de h unités de

temps comme suit :

$$F^h(X_t) = X_{t+h}$$

### Opérateur différence $\Delta$

L'opérateur  $\Delta$  fait la différence entre le processus et sa version décalée d'une unité de temps. L'opérateur est défini par :

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

En effet, par écriture purement formelle nous pouvons écrire :

$$\Delta X_t = X_t - BX_t = (1 - B)X_t$$

Si nous appliquons cet opérateur  $h$  fois nous obtenons :

$$\Delta^h(X_t) = X_t - B^h(X_t) = (1 - B)^h(X_t)$$

### 2.1.5 Processus aléatoire

D'un point de vue sémantique, un processus peut être considéré comme une série d'événements interconnectés dans le temps, formant une chaîne causale progressive. D'un point de vue mathématique, un processus peut être défini comme une solution à une équation impliquant des variables qui varient dans le temps. Si l'évolution temporelle de ces variables est influencée par des facteurs aléatoires, on parle alors de processus stochastique.

#### Processus stationnaire

Une classe très spéciale des processus stochastiques dite processus stationnaire démarre de l'hypothèse que le processus repose sur un équilibre statistique. Ceci implique que la série ne comporte ni de tendance ni de saisonnalité. Le plus simple processus stationnaire en analyse des séries temporelles est appelé : processus bruit blanc.

- **Processus bruit blanc**

Un processus de bruit blanc est une suite de variables aléatoires ( $\varepsilon_t$ ) de même distribution (Moyenne nulle et variance constante) et mutuellement indépendantes. C'est un terme qui a été emprunté aux physiciens dans le domaine de la lumière. Il fait allusion au spectre de la lumière blanche. Dans

le cas de l'analyse des séries chronologiques un bruit blanc  $\{\varepsilon_t, t \in T\}$  est tel que :

- $\forall t \in T : E(\varepsilon_t) = 0$
- $\text{var}(\varepsilon_t) = \sigma^2 \forall t \in T$
- $\text{cov}(\varepsilon_t, \varepsilon_{t+h}) = \begin{cases} \sigma_\varepsilon^2 & \text{si}(h = 0) \\ 0 & \text{sinon} \end{cases}$

• **Processus Auto Régressif d'ordre p**

Dans le processus autorégressif d'ordre p, noté AR(p), l'observation présente  $X_t$  est générée par une moyenne pondérée des observations passées jusqu'à la p-ième période sous la forme suivante :

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

ou encore, en utilisant le polynôme retard :

$$\phi(B)X_t = \varepsilon_t$$

Avec :

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

$\phi_1, \phi_2, \dots, \phi_p$  sont des paramètres réels à estimer et  $\varepsilon_t, t \in Z$  est un bruit blanc.

Le degré p est appelé le degré d'autocorrélation, il indique la profondeur de la mémoire.

• **Processus Moyen mobile d'ordre q**

Dans le processus moyenne mobile d'ordre q, noté MA(q), chaque observation  $X_t$  est générée par une moyenne pondérée d'aléas jusqu'à la q<sup>ième</sup> période comme suit :

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

ou encore, en utilisant le polynôme retard :

$$\theta(B)\varepsilon_t = X_t$$

Avec :

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

$\theta_1, \theta_2, \dots, \theta_q$  sont des paramètres réels à estimer et  $\varepsilon_t, t \in Z$  est un bruit blanc. Le degré  $q$  est nommé le degré de la moyenne mobile.

- **Processus autorégressif moyenne mobile**

Les processus autorégressifs moyenne mobile ARMA(p,q) sont donc représentatifs d'un processus généré par une combinaison des valeurs passées et des erreurs passées.

Un processus stationnaire  $X_t$  suit un ARMA (p,q) s'il vérifie l'équation stochastique suivante :

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

Avec :

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

le polynôme autorégressif d'ordre  $p$ .

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

le polynôme moyen mobile d'ordre  $q$ .

$\{ \varepsilon_t, t \in Z \}$  processus bruit blanc  $(0, \sigma_\varepsilon^2)$ .

Notons que :

ARMA (p, 0) est le processus AR (p) et ARMA (0, q) est le processus MA (q).

### Processus aléatoires non stationnaires

Les chroniques économiques sont rarement des réalisations de processus aléatoires stationnaires. Pour analyser la non-stationnarité, deux types de processus sont distingués :

- Les processus TS (Trend Stationary) qui représentent une non-stationnarité de type déterministe.
- Les processus DS (Differency Stationary) pour les processus non stationnaires aléatoires.

### Processus TS

Le processus TS (Trend Stationary) s'écrit :

$$X_t = \alpha + \beta t + \varepsilon_t$$

Où  $\varepsilon_t$  représente l'erreur du modèle à la date  $t$ .

Il présente une non stationnarité de nature déterministe.

Le processus TS est non stationnaire car  $E(X_t) = \alpha + \beta t$  dépend du temps  $t$ .

Le processus  $X_t$  peut être stationnarisé en retranchant à  $X_t$  la valeur estimée  $\hat{\alpha} + \hat{\beta}t$  par la méthode des Moindres Carrés Ordinaires.

### Processus DS (Differency Stationary)

#### Le processus DS avec dérive

C'est le cas d'un processus DS avec ( $\beta \neq 0$ ). Il s'écrit sous la forme suivante :

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

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

Il présente une non stationnarité de nature stochastique.

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

$$X_1 = X_0 + \beta + \varepsilon_1$$

$$X_2 = X_1 + \beta + \varepsilon_2 = X_0 + \beta + \varepsilon_1 + \beta + \varepsilon_2 = X_0 + 2\beta + \varepsilon_1 + \varepsilon_2$$

⋮

$X_t = X_0 + \beta t + \sum_{i=1}^t \varepsilon_i$  ou  $\varepsilon_i \sim \text{iid}(0, \sigma_\varepsilon^2)$  Autrement dit  $\varepsilon_i$  identiquement et indépendamment distribuée.

Le processus DS avec dérive est non stationnaire car on a  $E(X_t) = X_0 + \beta t$  qui dépend du temps  $t$ .

Plus  $t \rightarrow \infty$  et plus  $E(X_t) \rightarrow \infty$ .

#### Le processus DS sans dérivé

Dans ce cas ( $\beta = 0$ ) est le processus devient :

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

Par récurrence, on obtient (dans le cas sans dérive) :

$$X_1 = X_0 + \varepsilon_1$$

$$X_2 = X_1 + \varepsilon_2 = X_0 + \varepsilon_1 + \varepsilon_2$$

$$\vdots$$

$$X_t = X_0 + \sum_{i=1}^t \varepsilon_i \text{ ou } \varepsilon_i \sim \text{iid}(0, \sigma_\varepsilon^2)$$

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

$$\text{Var}(X_t) = \text{Var}(\sum_{i=1}^t \varepsilon_i) = \sum_{i=1}^t \text{Var}(\varepsilon_i) = \sum_{i=1}^t \sigma_\varepsilon^2 = t \sigma_\varepsilon^2$$

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

Plus  $t \rightarrow \infty$  et plus  $\text{Var}(X_t) \rightarrow \infty$ .

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

$$X_t - X_{t-1} = \beta + \varepsilon_t \text{ (cas avec dérive).}$$

ou

$$X_t - X_{t-1} = \varepsilon_t \text{ (cas sans dérive).}$$

### Processus autorégressifs moyenne mobile intégrés

Lorsque l'on a une série  $\{X_t\}$  à non stationnarité stochastique, il convient de la modéliser à l'aide d'un processus autorégressif moyenne mobile intégré ARIMA(p,d,q) où  $d$  désigne l'ordre de différenciation (ou d'intégration).

Un processus ARIMA (p,d,q) ou "Autoregressive Integrated Moving Average" d'ordre  $p$ ,  $d$ , et  $q$  pour la série  $X_t$  est un processus de la forme suivante :

$$(1 - \phi_1 B - \dots - \phi_p B^p) \nabla^d X_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

Ou encore

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

où  $\varepsilon_t \sim \text{BB}(0, \sigma_\varepsilon^2)$ ,  $B$  est l'opérateur de retard tel que  $BX_t = X_{t-1}$  et  $B^p X_t = X_{t-p}$ ,  $\nabla^d$  est l'opérateur de différence de degré  $d$  ( $d \geq 0$  est un entier positif),  $(\phi_1, \dots, \phi_p)$  et  $(\theta_1, \dots, \theta_q)$  sont des coefficients à estimer.

La série  $X_t$  est une série non stationnaire alors que la série  $W_t = \nabla^d X_t$  est une série stationnaire.

Estimer les paramètres du processus ARIMA(p,d,q) pour la série  $X_t$  non stationnaire revient à estimer les coefficients du processus ARMA(p,q) pour la série  $W_t$  stationnaire.

### Processus autorégressif moyenne mobile saisonnier

Ce sont des extensions des processus ARMA et ARIMA. Ils représentent généralement des séries marquées par une saisonnalité comme c'est le plus souvent, le cas pour des séries économiques voir financières. Ces séries peuvent mieux s'ajuster par des modèles saisonniers.

Ce sont les processus SARIMA (p,d,q)(P,D,Q) qui répondent au modèle :

$$\phi(B)\phi_s(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\theta_s(B^s)\varepsilon_t$$

Où

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  : polynôme autorégressif non saisonnier d'ordre p.

$\phi_s(B) = 1 - \phi_s B - \phi_{2s} B^{2s} - \dots - \phi_{ps} B^{ps}$  : polynôme autorégressif saisonnier d'ordre P.

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  : polynôme moyenne mobile non saisonnier d'ordre q.

$\theta_s(B) = 1 - \theta_s B - \theta_{2s} B^{2s} - \dots - \theta_{Qs} B^{Qs}$  : polynôme moyenne mobile saisonnier d'ordre Q.

$(1-B)^d$  : opérateur de différence d'ordre d.

$(1 - B^s)^D$  : opérateur de différence d'ordre D.

$\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$ , et s correspond à la saison (s=12 ou 4 ou ...).

- Les paramètres (p;d;q) correspondant à la partie non saisonnière, ils sont identifiés en examinant la coupure des fonctions d'autocorrélations simple et partielle.
- Pour la partie saisonnière (P;D;Q), les paramètres P et Q, ils sont identifiés en examinant la coupure des fonctions d'autocorrélations simple et partielle à des ordres où la différence saisonnière d'ordre "S" a été appliqué.

## 2.2 Méthodologie de Box & Jenkins

Après avoir décrit divers modèles de séries temporelles, la question suivante qui nous intéresse est de savoir comment sélectionner un modèle approprié qui peut produire des prévisions précises basées sur une description des données d'une série chronologique, et comment déterminer les ordres de modèle optimaux ?

Les statisticiens **George Box** et **Gwilym Jenkins** en 1976 ont proposés une démarche de prévision pour les séries temporelles univariées, fondée sur l'utilisation des modèles ARIMA : Leur concept est d'une importance fondamentale dans le domaine de l'analyse et de la prévision des séries chronologiques.

Le diagramme qui suit illustre la démarche générale de Box & Jenkins :

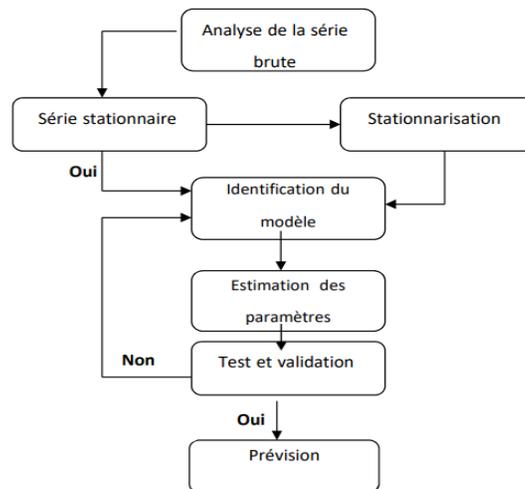


FIGURE 2.1 – Diagramme de box & Jenkins

### 2.2.1 Test des composantes saisonnières et tendancielle

#### Test de Fisher

Avant toute étude d'une chronique, il convient d'élaborer un test permettant de détecter l'existence d'une saisonnalité. Le test le plus communément employé est celui de Fisher par analyse de la variance du facteur période (mensuel, trimestrie, ...).

On considère  $n$  : Le nombre d'années,  $P$  : Le nombre d'observations dans l'année,  $X_{ij}$  : La valeur de la série pour la  $i^{\text{ème}}$  année et la  $j^{\text{ème}}$  période.

La moyenne générale  $\bar{X}$ , la moyenne de l'année  $i$   $\bar{X}_i$ , la moyenne de la période  $j$   $\bar{X}_j$ .

La variance année et la variance période sont définies respectivement par :

$$var_A = \frac{P \sum_{i=1}^n (\bar{X}_i - \bar{X})^2}{n-1}, \quad Var_P = \frac{n \sum_{j=1}^p (\bar{X}_j - \bar{X})^2}{p-1}$$

La variance résiduelle :

$$var_R = \frac{\sum_{i=1}^n \sum_{j=1}^p (X_{ij} - \bar{X}_i - \bar{X}_j - \bar{X})^2}{(n_1)(p-1)}$$

L'équation de la variance totale :

$$var_T = var_A + var_P + var_R = \frac{\sum_{i=1}^n \sum_{j=1}^p (X_{ij} - \bar{X})^2}{n-1}$$

L'hypothèse est :

$H_0$  : pas de saisonnalité

$H_1$  : il existe une saisonnalité

La valeur calculée  $F_0 = \frac{var_P}{var_R}$  est comparée à la valeur tabulée  $F_{v_1 v_2}^\alpha$

Avec  $v_1 = (p-1)$ ,  $v_2 = (n-1)(p-1)$  degré de liberté.

Si  $F_0 > F_{v_1 v_2}^\alpha$  on rejette  $H_0$ , la série est saisonnière.

Soient les hypothèses  $\left\{ \begin{array}{l} H_0 : \text{la série n'est pas affectée d'une tendance} \\ H_1 : \text{la série est affectée d'une tendance} \end{array} \right.$

Nous calculons  $F_1 = \frac{var_A}{var_B}$  que l'on compare avec  $F_{v_3 v_2}^\alpha$

Avec  $v_3 = (n-1)$ ,  $v_2 = (n-1)(p-1)$  degré de liberté.

Si  $F_1 > F_{v_3 v_2}^\alpha$  on rejette l'hypothèse, la série est affectée d'une tendance.

Concernant l'existence de la tendance, le test de Fisher s'avère faible, il convient d'effectuer un autre test.

### Test de Dickey-Fuller

Il permet de mettre en évidence le caractère stationnaire ou non d'une chronique par la détermination d'une tendance déterministe ou stochastique.

Pour la construction de ce test, Dickey et Fuller ont proposé trois modèles de base :

(1-  $\phi_1 B$ )  $X_t = \varepsilon_t$  : Modèle autorégressif d'ordre 1 ... **Modèle[1]**

(1-  $\phi_1 B$ )  $(X_t - c) = \varepsilon_t$  : Modèle autorégressif d'ordre 1 avec constante ... **Modèle[2]**

(1-  $\phi_1 B$ )  $(X_t - c - bt) = \varepsilon_t$  : Modèle autorégressif d'ordre 1 avec tendance ... **Modèle[3]**

Les hypothèses à tester sont  $\begin{cases} H_0 : \phi_1 = 1 \\ H_1 : |\phi_1| < 1 \end{cases}$

C'est un test itératif qui se déroule comme suit :

Si dans l'un des modèles la statistique relative à  $\phi_1$  et supérieure aux valeurs tabulées  $t_{0,05}$ , nous déduisons qu'il existe une racine unitaire. Nous estimons en premier le modèle (3), nous commençons par tester la significativité de la tendance :

Si la tendance n'est pas significative alors nous testons la significativité de la constante.

Si elle est significative le processus est de type DS avec dérive, pour stationnariser la chronique on applique le filtre aux différences et recommencer la procédure.

Si elle n'est pas significative, le processus est donc une marche aléatoire dit DS sans dérive, on différencie la série une fois et on recommence jusqu'à ce que la série soit stationnaire.

Sinon, le processus est déterministe ; pour le stationnariser on effectue une régression sur le temps et on recommence le test.

Sinon Si dans le modèle (3), le coefficient de la tendance est significativement différent de 0 alors le processus est TS ; Sinon la série est stationnaire.

## 2.2.2 Etapes de la méthodologie de Box & Jenkins

### Identification

La méthodologie de Box et Jenkins se résume en trois étapes :

### Estimation

On utilise pour cette phase la méthode d'estimation du maximum de vraisemblance ou bien celle des moindres carrés ordinaires. Elle consiste à identifier les coefficients autorégressifs et moyenne mobiles saisonniers ou non.

### Validation

Après la phase précédente, si on dispose de plusieurs modèles candidats, il est nécessaire d'effectuer des tests sur les paramètres et sur les résidus afin de les départager.

### Test sur les paramètres

C'est un test classique de Student qui permet de tester la significativité de chaque paramètre du processus en utilisant les hypothèses :

$H_0$  : « Le coefficient est nul », contre  $H_1$  : « le coefficient est significativement différent de zéro »

La statistique de Student représente le ratio du coefficient à estimer sur son écart type  $\frac{\hat{B}_p}{\hat{\sigma}_{\hat{B}_p}}$ .

Si  $|\frac{\hat{B}_p}{\hat{\sigma}_{\hat{B}_p}}| > 1.96(\alpha = 5\%)$  nous acceptons  $H_1$ , dans le cas contraire on rejette le modèle et on envisage une nouvelle spécification (retour à l'étape identification). Après avoir validé plusieurs modèles, il convient de choisir le modèle optimal en se basant sur une comparaison de leur qualité en utilisant les critères d'informations et /ou le principe de parcimonie.

### Critères d'informations

Ce sont des statistiques aidant à décider de l'ordre d'un modèle dans le cas où il en existerait plusieurs. Entre autres ; on peut citer :

#### Critère d'Akaike (AIC)

Il tient compte de la qualité d'ajustement du modèle à la série observée et du nombre de paramètres utilisés dans l'ajustement.

$$AIC(p,q) = n \text{Log } \hat{\sigma}_\varepsilon^2 + 2(p + q)$$

#### Critère de Schwartz 1978 (SC)

Il est définie par :

$$SC(p,q) = n \text{Log } \hat{\sigma}_\varepsilon^2 + (p + q) \text{Log} n$$

### Teste sur les résidus

Lorsque le modèle est bien estimé, les résidus entre les valeurs observées et les valeurs estimées doivent se comporter comme un bruit blanc (gaussien ou pas). Les résidus (ou erreurs de prévision) sont notés  $\varepsilon_t$ ,  $t \in \mathbb{Z}$ .

### Test de Box-Ljung

Appelé aussi test de « porte manteau », il permet de vérifier l'hypothèse de bruit blanc des résidus, ce qui signifie qu'il n'existe aucune autocorrélation significativement non nulle, il se base sur la fonction d'autocorrélation en utilisant deux

hypothèses :

$$H_0 : \rho_1 = 0, \rho_2 = 0, \dots, \rho_k = 0;$$

$H_1$  : Il existe au moins un  $\rho_i$  significativement différent de zéro .

Sa statistique est :

$$Q = n(n+1) \sum_{h=1}^k \frac{r_h^2(\varepsilon_i)}{n-h}$$

avec :

K est le nombre de retards choisis et n est le nombre d'observations.

En l'absence d'autocorrélation la statistique q obéit à une  $\chi^2$  (K-p-q-P-Q) degrés de liberté.

L'hypothèse  $H_0$  est rejetée au seuil de 5% si Q est supérieure au quantile 0.95 de la loi de  $\chi^2$ .

### Test de normalité

Le test le plus fréquent qui permet de vérifier la normalité d'une distribution statistique est celui de Jarque et Bera (1984) ce dernier est fondé sur la notion de Skewness (moment d'ordre 3, asymétrie) et de Kurtosis (moment d'ordre 4, aplatissement). Ces deux tests permettront par la suite le calcul des intervalles de confiance prévisionnels et aussi d'effectuer les tests de Student sur les paramètres.

Soit  $U_k = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^k$  le moment centré d'ordre k.

$$\text{Le coefficient de Skewness est : } B_1^{\frac{1}{2}} = \frac{U_3}{U_2^{\frac{3}{2}}}$$

$$\text{Le coefficient de kurtosis est : } B_2 = \frac{U_4}{U_2^2}$$

Si la distribution est normale et le nombre d'observation est grand. Alors,

$$B_1^{\frac{1}{2}} \sim (0, \sqrt{\frac{6}{n}}) \text{ et } B_2 \sim (3, \sqrt{\frac{24}{n}}).$$

On construit alors les statistiques :

$$|V_1| = \left| \frac{B_1^{\frac{1}{2}} - 0}{\sqrt{\frac{6}{n}}} \right| \text{ et } |V_2| = \left| \frac{B_2 - 3}{\sqrt{\frac{24}{n}}} \right|$$

Que l'on compare à la statistique de Student 1.96 au seuil 5%. Si les hypothèses suivantes :

$$\left\{ \begin{array}{l} H_0 \quad V_1=0(\text{symétrie}) \\ \quad \quad \quad \text{contre} \\ H_1 \quad V_2=0(\text{aplatissement normal}) \end{array} \right.$$

On accepte  $H_0$  si :  $|V_1| < 1.96$  et  $|V_2| < 1.96$

### Test de Jarque Bera

Le test de normalité de Jarque-Bera est également fondé sur les coefficients d'asymétrie et d'aplatissement.

Il évalue les écarts simultanés de ces coefficients avec les valeurs de référence de la loi normale. La statistique de Jarque-Bera (JB) qui s'écrit

$$JB = n \left( \frac{B_1^{\frac{1}{2}}}{6} + \frac{(B_2 - 3)^2}{24} \right)$$

Elle est distribuée asymptotiquement selon une loi de  $\chi^2$  à 2 degrés de liberté.

Si  $JB \geq \chi_{1-\alpha}^2(2)$  nous rejetons l'hypothèse  $H_0$  de normalité des résidus (au seuil  $\alpha$ ).

### Test de Durbin et Watson

Les modèles ajustés à des séries chronologiques manifestent parfois un certain degré de corrélation entre les valeurs successives des erreurs. En terme probabiliste, cela signifie que les erreurs sont autocorrélées, ou encore qu'une erreur produite en  $t-1$  à une influence sur l'erreur en  $t$ . Le test de Durbin et Watson (1951) permet de détecter l'autocorrélation des résidus pour un ordre un (corrélation entre  $\varepsilon_t$  et  $\varepsilon_{t-1}$ ) sous la forme :

$$\varepsilon_t = \rho \varepsilon_{t-1} + V_t \text{ ou } V_t \rightarrow N(0; \sigma_v^2) \text{ nous testons}$$

$H_0 : \ll \rho = 0 \gg$  (absence d'autocorrélation à l'ordre 1 des résidus).

contre

$H_1 : \ll \rho \neq 0 \gg$  (présence d'autocorrélation à l'ordre 1 des résidus).

La statistique de Durbin et Watson, notée **DW**, est donnée par :

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} \text{ avec } \varepsilon_t : \text{ sont les résidus de l'estimation du modèle}$$

De par sa construction, cette statistique varie entre 0 et 4 et nous avons  $DW = 2$  lorsque  $\hat{\rho} = 0$  ( $\hat{\rho}$  est l'estimateur de  $\rho$ ) :

$DW = 0$ , il existe une autocorrélation positive.

$DW = 4$ , il existe une autocorrélation négative.

$DW \approx 2$ , indique l'absence d'autocorrélation.

### Test d'hétéroscédasticité (l'existence d'effet ARCH)

Pour ce faire, nous utilisons les corrélogrammes des carrés des résidus, si un ou plusieurs termes sont significativement différents de zéro, nous déduisons qu'il y'a effet ARCH qui est détecté également par la statistique de Box et Ljung, et qui est confirmé par la statistique du Multiplicateur de Lagrange  $LM = nR^2$  avec  $n$  le nombre d'observations servant au calcul de la régression et  $R^2$  est le coefficient de détermination. Soit une spécification de type ARCH pour les erreurs  $\varepsilon_t$  tel que :

$$\varepsilon_t = U_t h_t$$

avec

$$U_t \sim N(0,1)$$

et

$$h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

Soit l'hypothèse  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$  contre  $H_1 : \alpha_i$  non tous nul »

Si  $LM < \chi_1^2$  nous acceptons  $H_0$ . La variance de l'erreur est constante  $h_t^2 = \alpha_0$

Dans le cas contraire  $LM > \chi^2(p)$  à  $p$  degré de liberté compris entre 1 et 3 nous rejetons  $H_0$  et le processus est justifiable d'un modèle ARCH ( $p$ ).

si  $p > 3$  le modèle sera justifié d'un modèle de type GARCH.

### 2.2.3 Prévision

C'est l'ultime étape de cette méthode, voire de la quasi-totalité des analyses statistiques. C'est une extrapolation des observations d'une série en se basant sur ses observations et ses erreurs passées. Lors de cette étape, il est indispensable de prendre en considération les transformations effectuées sur la chronique afin d'aboutir à des valeurs prévisionnelles appropriées. Par définition :

$\hat{X}_{t+1} = E(X_{t+1}/I_t)$  Ou  $I_t$  est l'information disponible à  $t$

$$I_t = (X_1, X_2, \dots, X_t, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_t)$$

Considérons un ARMA (1,1) et un horizon de prévision 1

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

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

$$\hat{X}_{t+1} = E(X_{t+1}/I_t) = \phi_1 X_t - \theta_1 \varepsilon_t$$

$$X_{t+2} = \phi_1 X_{t+1} + \varepsilon_{t+2} - \theta_1 \varepsilon_{t+1}$$

$$\hat{X}_{t+2} = E(X_{t+2}/I_t) = \phi_1 E(X_{t+1}/I_t)$$

$$\hat{X}_{t+2} = \phi_1 E((\phi_1 X_t + \varepsilon_{t+1} - \theta_1 \varepsilon_t)/I_t) = \phi_1 E((\phi_1 X_t - \theta_1 \varepsilon_t)/I_t)$$

$$\hat{X}_{t+2} = \phi_1 (\phi_1 X_t - \theta_1 \varepsilon_t) = \phi_1 \hat{X}_{t+1}$$

La généralisation pour un horizon h :

$$\phi(B)X_t = \theta(B)\varepsilon_t \iff X_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t = \psi(B)\varepsilon_t$$

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \text{ et } \hat{X}_{t+h} = E(X_{t+h}/I_t) = E(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t+h-i}/I_t)$$

$$\text{Avec } E(\varepsilon_{t+h-i}/I_t) = \begin{cases} \varepsilon_{t+h-i} & \text{si } h-i \leq 0 \Rightarrow i \geq h \\ 0 & \text{sinon} \end{cases}$$

D'où :

$$X_{t+h} = \sum_{i=h}^{\infty} \psi_i \varepsilon_{t+h-i}$$

Comme la prévision n'est pas identique à la valeur réelle, il existe un écart entre ces deux valeurs appelé erreur de prévision  $\hat{\varepsilon}_{t+1}$

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

$$\hat{\varepsilon}_{t+2} = X_{t+2} - \hat{X}_{t+2} = \varepsilon_{t+1} + \psi_1 \varepsilon_{t+1}$$

⋮

$$\hat{\varepsilon}_{t+h} = X_{t+h} - \hat{X}_{t+h} = \sum_{i=0}^{h-1} \psi_i \varepsilon_{t+h-i} \text{ avec } \psi_0 = 1$$

Sous réserve que les erreurs forment un bruit blanc gaussien, on pourra construire un intervalle de prévision tel que :

$$V(\hat{\varepsilon}_{t+h}) = V(\sum_{i=0}^{h-1} \psi_i \varepsilon_{t+h-i}) = E(\sum_{i=0}^{h-1} \psi_i \varepsilon_{t+h-i})^2 - 0 = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \psi_i^2$$

$$\text{D'où l'intervalle de prévision : } \hat{X}_{t+h} \pm 1.96 \sigma_\varepsilon (\sum_{i=0}^{h-1} \psi_i^2)^{\frac{1}{2}}$$

Pour clore ce chapitre, nous pouvons dire pour cette méthode, qu'en dépit de sa simplicité dans son principe de base, elle semble complexe dans sa mise en œuvre, d'autant plus qu'elle ne prend pas en compte les informations externes. Néanmoins, elle permet non seulement de traiter et d'analyser les processus stationnaires mais également les processus qui ne le sont pas.

# Chapitre 3

## Application de la méthodologie de BOX & JENKINS

### Introduction

Nous nous intéressons, dans ce chapitre à l'étude prévisionnelle des séries :

- Investissement en million de dollars (noté **INV**).
- Prix de pétrole (noté **PRIX**).
- Volume découvertes (noté **VD**).
- Sismique de trois dimension (noté **SIS3D**).

Toutes les séries comportent 42 observations, De données annuelles à partir de l'année 1980 à 2022, sauf la série sis3D qui contient 28 observations (1995-2022).

### 3.1 Etude de la série des investissements financier pour l'exploration $INV_t$

#### 3.1.1 Identification

La série  $INV_t$  représente l'évolution annuelle des investissements financiers pour l'exploration pétrolière sur une période allant de 1980 à 2022.

#### Analyse préliminaire de la série $INV_t$

La représentation graphique est donnée par le graphe suivant :

D'après l'analyse préliminaire, Nous remarquons qu'il y'a des fluctuations dans son parcours de 1980 jusqu'à 2022, ensuite l'investissement a enregistré

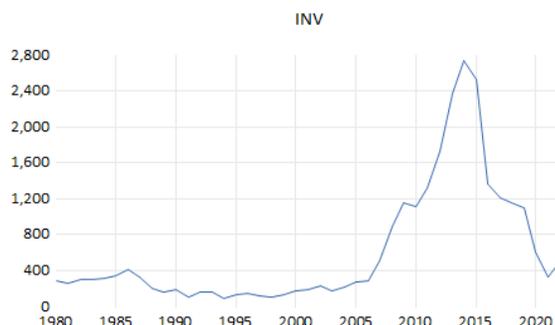


FIGURE 3.1 – Graphe de la série  $INV_t$

une tendance en 2005. Donc, la série n'est pas stationnaire. L'examen graphique ne permet pas toujours de déterminer avec certitude l'existence d'une tendance. Afin d'éloigner l'incertitude, nous utilisons le corrélogramme et le test de Dickey-Fuller.

**Corrélogramme de la série  $INV_t$  :**

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.924	0.924	39.342	0.000
		2	0.785	-0.473	68.408	0.000
		3	0.653	0.196	89.055	0.000
		4	0.548	-0.020	103.96	0.000
		5	0.426	-0.350	113.20	0.000
		6	0.281	-0.027	117.33	0.000
		7	0.144	0.043	118.44	0.000
		8	0.049	0.045	118.58	0.000
		9	-0.007	0.040	118.58	0.000
		10	-0.047	0.007	118.71	0.000
		11	-0.087	-0.070	119.16	0.000
		12	-0.120	-0.014	120.05	0.000
		13	-0.139	-0.056	121.30	0.000
		14	-0.154	-0.115	122.88	0.000
		15	-0.169	0.048	124.86	0.000
		16	-0.184	-0.015	127.28	0.000
		17	-0.195	-0.042	130.12	0.000
		18	-0.205	-0.018	133.38	0.000
		19	-0.215	-0.032	137.12	0.000
		20	-0.222	-0.017	141.28	0.000

FIGURE 3.2 – Corrélogramme de la série  $INV_t$

D'après le corrélogramme la fonction d'autocorrélation simple et partielle diminue lentement vers zéro, on remarque aussi de nombreux pics sortant de la bande de confiance. Donc, la série brute nous indique une non stationnarité de la série.

### Test de stationnarité de la série $INV_t$

#### Test de Dickey-Fuller Augmenté (ADF)

Nous appliquons le test d'ADF pour confirmer ou infirmer notre hypothèse concernant la stationnarité de ce processus, nous avons besoin de déterminer le retard qui minimise le critère d'Akaike et Schwarz qui est égal à 0. Par la suite, nous estimons par la méthode des moindres carrés les trois modèles.

Nous commençons par Modèle [3] :

#### Modèle [3] :

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.258589	0.8845
Test critical values:		
1% level	-4.192337	
5% level	-3.520787	
10% level	-3.191277	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(INV)  
 Method: Least Squares  
 Date: 04/27/23 Time: 11:51  
 Sample (adjusted): 1981 2022  
 Included observations: 42 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INV(-1)	-0.096092	0.076349	-1.258589	0.2157
C	24.64900	83.84486	0.293983	0.7703
@TREND("1980")	1.871233	4.304424	0.434723	0.6662

R-squared	0.043674	Mean dependent var	5.380952
Adjusted R-squared	-0.005368	S.D. dependent var	264.1384
S.E. of regression	264.8464	Akaike info criterion	14.06493
Sum squared resid	2735602.	Schwarz criterion	14.18905
Log likelihood	-292.3635	Hannan-Quinn criter.	14.11042
F-statistic	0.890546	Durbin-Watson stat	1.086253
Prob(F-statistic)	0.418613		

FIGURE 3.3 – ADF modèle 3 de la série  $INV_t$

Nous testons les hypothèses suivante :

**Test d'existence d'une racine unitaire :**

- $H_0$  : "il existe une racine unitaire "
- $H_1$  : "il n'existe pas une racine unitaire "

**Test d'existence de la tendance :**

- $H_0$  : " la tendance est non significative "
- $H_1$  : " la tendance est significative "

La statistique de Dickey-Fuller Augmenté égale -1.258589 est supérieure aux différentes valeurs critiques relatives aux seuils 1% 5% et 10% on accepte donc l'hypothèse nulle de l'existence de racine unitaire.

La t-statistique de la tendance égale à 0.434723 est inférieur aux valeurs critiques 3.53 ; 2.79 ; 2.38 tabulées par Dickey-Fuller respectivement aux seuils 1%, 5% et 10%. Donc, nous acceptons  $H_0$  ; et la tendance n'est pas significative. Le processus étudié n'est pas de type TS.

Nous passons au **modèle [2]** :

	t-Statistic	Prob.*		
Augmented Dickey-Fuller test statistic	-1.274777	0.6324		
Test critical values:				
1% level	-3.596616			
5% level	-2.933158			
10% level	-2.604867			
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(INV)				
Method: Least Squares				
Date: 04/27/23 Time: 11:55				
Sample (adjusted): 1981 2022				
Included observations: 42 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
INV(-1)	-0.075459	0.059194	-1.274777	0.2097
C	52.10439	54.58583	0.954541	0.3455
R-squared	0.039040	Mean dependent var		5.380952
Adjusted R-squared	0.015016	S.D. dependent var		264.1384
S.E. of regression	262.1478	Akaike info criterion		14.02214
Sum squared resid	2748858.	Schwarz criterion		14.10489
Log likelihood	-292.4650	Hannan-Quinn criter.		14.05247
F-statistic	1.625056	Durbin-Watson stat		1.101795
Prob(F-statistic)	0.209744			

FIGURE 3.4 – ADF modèle [2] de la série  $INV_t$

Nous testons l'hypothèse suivante :

**Test d'existence de la constante :**

- $H_0$  : " la constante est non significative "
- $H_1$  : " la constante est significative "

La statistique de Dickey-Fuller Augmenté égale -1.274777 est supérieure aux différentes valeurs critiques relatives aux seuils 1% 5% et 10% on accepte donc l'hypothèse nulle de l'existence de racine unitaire.

D'après la table, nous avons la valeur de la t-statistique de la constante qui est égale à 0.954541 est inférieure aux valeurs critiques 3.22 ; 2.54 et 2.17 tabulées par Dickey-Fuller respectivement aux seuils 1%, 5% et 10%. Donc, la constante est non significative.

Nous passons au **modèle [1]** :

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.856267	0.3391
Test critical values:		
1% level	-2.621185	
5% level	-1.948886	
10% level	-1.611932	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(INV)  
 Method: Least Squares  
 Date: 04/27/23 Time: 11:57  
 Sample (adjusted): 1981 2022  
 Included observations: 42 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INV(-1)	-0.037519	0.043817	-0.856267	0.3968

R-squared	0.017151	Mean dependent var	5.380952
Adjusted R-squared	0.017151	S.D. dependent var	264.1384
S.E. of regression	261.8635	Akaike info criterion	13.99705
Sum squared resid	2811473.	Schwarz criterion	14.03842
Log likelihood	-292.9380	Hannan-Quinn criter.	14.01221
Durbin-Watson stat	1.116404		

FIGURE 3.5 – ADF modèle [1] de la série  $INV_t$

La statistique de Dickey-Fuller Augmenté égale -0.856267 est supérieure aux différentes valeurs critiques relatives aux seuils 1% 5% et 10% on accepte donc

l'hypothèse  $H_0$ , il existe une racine unitaire (non stationnaire), la meilleure façon de la rendre stationnaire est de la différencier d'ordre 1.

La nouvelle série sera notée  $DINV_t$ . Donc, notre série est de type DS.

### Etude de la série $DINV_t$

#### Test de Dickey-Fuller Augmenté sur la série $DINV_t$

##### Modèle [3]

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.868789	0.0226
Test critical values:		
1% level	-4.198503	
5% level	-3.523623	
10% level	-3.192902	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(INV,2)  
 Method: Least Squares  
 Date: 04/27/23 Time: 11:58  
 Sample (adjusted): 1982 2022  
 Included observations: 41 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1))	-0.573385	0.148208	-3.868789	0.0004
C	22.35833	82.05604	0.272476	0.7867
@TREND("1980")	-0.752827	3.288306	-0.228941	0.8201

R-squared	0.282972	Mean dependent var	5.292683
Adjusted R-squared	0.245234	S.D. dependent var	285.2383
S.E. of regression	247.8073	Akaike info criterion	13.93354
Sum squared resid	2333521.	Schwarz criterion	14.05892
Log likelihood	-282.6375	Hannan-Quinn criter.	13.97919
F-statistic	7.498262	Durbin-Watson stat	1.713096
Prob(F-statistic)	0.001800		

FIGURE 3.6 – ADF modèle [3] de la série  $DINV_t$

Nous remarquons que :

La statistique de Dickey-Fuller Augmenté égale -3.868789 est supérieure à valeurs critique à 1% et inférieure aux 5% et 10% on accepte donc l'hypothèse nulle de l'existence de racine unitaire à 1% seulement.

La t-statistique de la tendance égale à -0.228941 est inférieur aux valeurs critiques 3.53 ; 2.79 ; 2.38 tabulées par Dickey-Fuller respectivement aux seuils 1%, 5% et

10%. Donc, nous acceptons  $H_0$ ; et la tendance n'est pas significative.

Nous passons au **modèle [2]**

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.913598	0.0043
Test critical values:		
1% level	-3.600987	
5% level	-2.935001	
10% level	-2.605836	

\*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(INV,2)  
 Method: Least Squares  
 Date: 04/27/23 Time: 11:59  
 Sample (adjusted): 1982 2022  
 Included observations: 41 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1))	-0.569892	0.145618	-3.913598	0.0004
C	5.793076	38.22816	0.151539	0.8803

R-squared	0.281983	Mean dependent var	5.292683
Adjusted R-squared	0.263572	S.D. dependent var	285.2383
S.E. of regression	244.7783	Akaike info criterion	13.88613
Sum squared resid	2336740.	Schwarz criterion	13.96972
Log likelihood	-282.6657	Hannan-Quinn criter.	13.91657
F-statistic	15.31625	Durbin-Watson stat	1.715124
Prob(F-statistic)	0.000354		

FIGURE 3.7 – ADF modèle [2] de la série  $DINV_t$

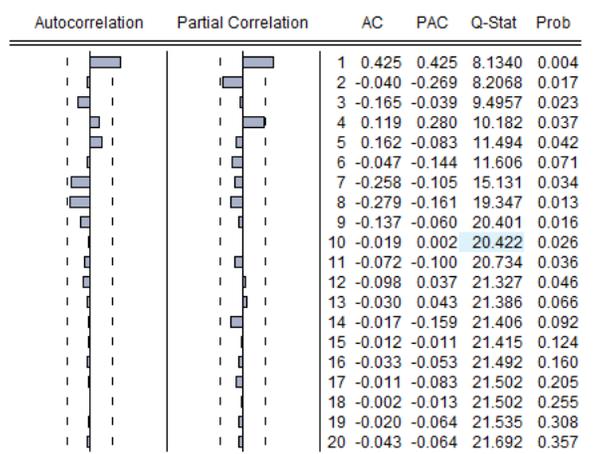
La statistique de Dickey-Fuller Augmenté égale -3.913598 est inférieure aux différentes valeurs critiques relatives aux seuils 1% 5% et 10% on accepte l'hypothèse nulle de non existence de racine unitaire (la série est stationnaire).

Nous avons la valeur de la t-statistique de la constante qui est égale à 0.151539 est inférieure aux valeurs critiques 3.22; 2.54 et 2.17 tabulées par Dickey-Fuller respectivement aux seuils 1%, 5% et 10%. Donc, la constante et non significative.

### Interprétation

D'après les résultats précédents, nous rejetons l'hypothèse nulle de racine unitaire, alors nous acceptons l'hypothèse de stationnarité  $H_0$ . La série est stationnaire.

### Identification du modèle :

**Corrélogramme de la série  $DINV_t$  :**FIGURE 3.8 – Corrélogramme de la série  $DINV_t$ 

L'analyse du corrélogramme simple de la série montre qu'au retard 1 le terme est significativement différent de zéro et pour le corrélogramme partiel, on remarque qu'au retard 1, le terme est aussi à l'extérieur de l'intervalle de confiance, par conséquent nous avons plusieurs modèles candidats parmi lesquels nous avons estimé les modèles suivants : ARIMA(1, 1, 0); ARIMA(0, 1, 1); ARIMA(1, 1, 1).

**3.1.2 Estimation des paramètres****Estimation des paramètres du modèle ARIMA(1,1,1)**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.922283	0.031791	29.01041	0.0000
MA(1)	0.449014	0.120246	3.734124	0.0006
SIGMASQ	51539.27	7185.963	7.172214	0.0000
R-squared	0.887064	Mean dependent var		616.7674
Adjusted R-squared	0.881417	S.D. dependent var		683.5377
S.E. of regression	235.3821	Akaike info criterion		13.89308
Sum squared resid	2216188.	Schwarz criterion		14.01595
Log likelihood	-295.7012	Hannan-Quinn criter.		13.93839
Durbin-Watson stat	1.748913			
Inverted AR Roots	.92			
Inverted MA Roots	-.45			

FIGURE 3.9 – Estimation des paramètres du modèle ARIMA(1,1,1)

Les t-statistiques en valeurs absolue de ce modèle sont supérieures à 1.96 (valeur tabulée de la table de student) alors les coefficients sont significativement différents de zéro chose confirmée par la probabilité de nullité de ce coefficient qui est inférieur à 0.05. La statistique de Durbin-Watson est égale à  $1.7489 \approx 2$ , ce qui laisse pressentir un bon ajustement.

### Estimation des paramètres du modèle ARIMA(1,1,0)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.947599	0.021974	43.12387	0.0000
SIGMASQ	65765.26	8011.339	8.209022	0.0000
R-squared	0.855891	Mean dependent var		616.7674
Adjusted R-squared	0.852376	S.D. dependent var		683.5377
S.E. of regression	262.6277	Akaike info criterion		14.07782
Sum squared resid	2827906.	Schwarz criterion		14.15974
Log likelihood	-300.6732	Hannan-Quinn criter.		14.10803
Durbin-Watson stat	1.098146			
Inverted AR Roots	.95			

FIGURE 3.10 – Estimation des paramètres du modèle ARIMA(1,1,0)

Les t-statistiques en valeurs absolue de ce modèle son supérieures à 1.96 (valeur tabulée de la table de student) alors les coefficients sont significativement différents de zéro chose confirmée par la probabilité de nullité de ce coefficient qui est inférieur à 0.05. La statistique de Durbin-Watson est égale à 1.098146 pas proche de 2.

### Estimation des paramètres du modèle ARIMA(0,1,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.917731	0.119059	7.708231	0.0000
SIGMASQ	261999.4	46101.39	5.683114	0.0000
R-squared	0.425891	Mean dependent var		616.7674
Adjusted R-squared	0.411888	S.D. dependent var		683.5377
S.E. of regression	524.1945	Akaike info criterion		15.44993
Sum squared resid	11265976	Schwarz criterion		15.53185
Log likelihood	-330.1735	Hannan-Quinn criter.		15.48014
Durbin-Watson stat	0.524907			
Inverted MA Roots	-.92			

FIGURE 3.11 – Estimation des paramètres du modèle ARIMA(0,1,1)

Les t-statistiques en valeurs absolue de ce modèle son supérieures à 1.96 (valeur tabulée de la table de student) alors les coefficients sont significativement différents de zéro ce qui est confirmée par la probabilité de nullité de ce coefficient qui est inférieur à 0.05. La statistique de Durbin-Watson est égale à 0.524907 très loin de 2.

### 3.1.3 Validation du modèle

#### Choix du modèle

	$R^2$	AIC	SCH
ARIMA(1,1,1)	<b>0.887064</b>	<b>13.89308</b>	<b>14.01595</b>
ARIMA(1,1,0)	0.855891	14.07782	14.15974
ARIMA(0,1,1)	0.425891	15.44993	15.53185

TABLE 3.1 – Choix du modèle

Suivant les critères de choix mentionnés dans le tableau ci-dessus, le modèle ARIMA(1, 1, 1) sera choisi pour représenter le processus. En effet, le coefficient  $R^2$  du modèle ARIMA(1, 1, 1) est supérieur, et les critères AIC et SCH sont inférieurs en comparaison avec les autres modèles.

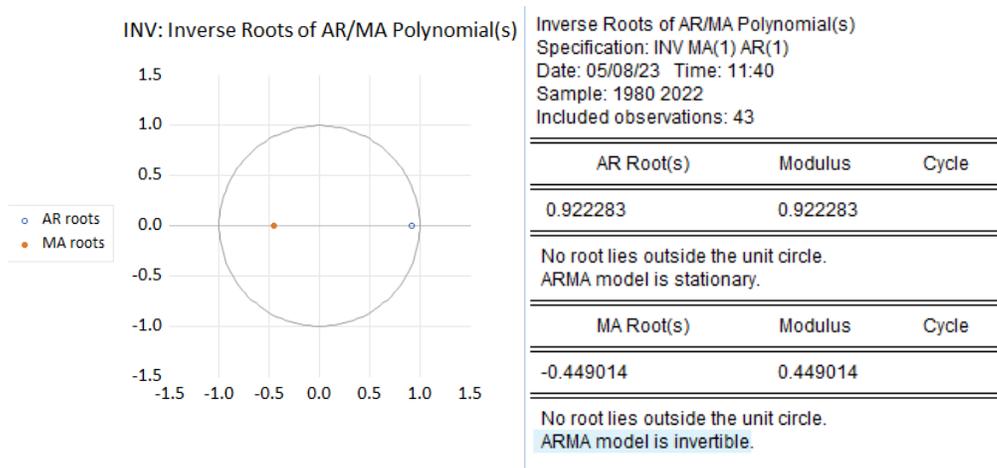


FIGURE 3.12 – Racines inverses du(des) polynôme(s) MA/AR

### Tests sur les résidus du modèle

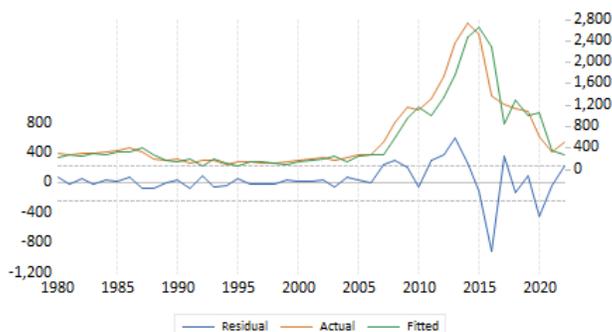


FIGURE 3.13 – Graphe des séries résiduelles actuelles et estimée

A partir de la représentation graphique des séries résiduelles actuelles et estimée, nous constatons que le graphe de la série estimée est presque semblable à celui de la série réelle à quelques pics près. En effet, le  $R^2$  est égal à 0.887. Donc, le modèle a bien expliqué la série soit 88,7% de la variation totale de la variable dépendante.

### Corrélogramme des résidus

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.138	0.138	0.8760	0.349
		2 0.044	0.026	0.9688	0.616
		3 0.327	0.324	6.1527	0.104
		4 0.315	0.262	11.066	0.026
		5 0.062	0.006	11.265	0.046
		6 0.003	-0.125	11.265	0.081
		7 0.066	-0.133	11.496	0.118
		8 0.058	-0.053	11.683	0.166
		9 0.047	0.083	11.809	0.224
		10 -0.037	0.021	11.887	0.293
		11 -0.033	-0.014	11.953	0.367
		12 -0.022	-0.074	11.982	0.447
		13 -0.031	-0.063	12.045	0.524
		14 -0.045	-0.009	12.178	0.592
		15 -0.046	0.024	12.326	0.654
		16 -0.047	0.014	12.485	0.710
		17 -0.050	-0.020	12.671	0.758
		18 -0.054	-0.051	12.897	0.798
		19 -0.058	-0.044	13.166	0.830
		20 -0.059	-0.023	13.460	0.857

FIGURE 3.14 – Corrélogramme des résidus

D'après l'analyse du corrélogramme des résidus, nous remarquons que tous les pics sont à l'intérieur de l'intervalle de confiance et les probabilités corres-

pondantes sont supérieures à 0.05. Donc les résidus forment un bruit blanc, ce qui est confirmée par la statistique de LJUNG-BOX (Q-stat) qui est inférieure à la valeur théorique  $\chi^2(h-3)$  de quel que soit le retard h, on particulier pour h=10 on a :

Q-stat (10) = 11.887 <  $\chi^2(7)$ = 14.06 au seuil 0.05, donc les résidus forment un bruit blanc.

Les tests étant tous vérifiés. Donc, **le modèle ARIMA(1,1,1) est validé.**

### Test de normalité sur les résidus de ARIMA(1,1,1)

L'histogramme de la distribution et les valeurs empiriques des Skewness-Kurtosis et la statistique de Jarque-Bera sont données par cette figure :

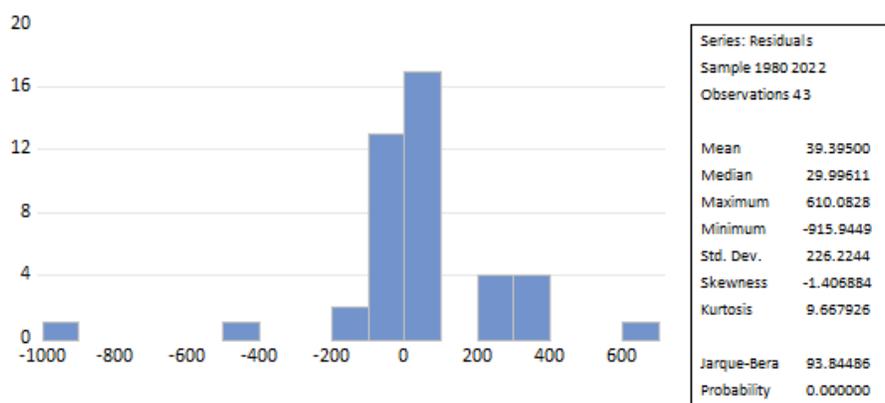


FIGURE 3.15 – L'Histogramme

Nous testons les hypothèses suivantes :

$$\begin{cases} H_0 : V_1 = 0 \text{ et } V_2 = 0 \\ H_1 : V_1 \neq 0 \text{ ou } V_2 \neq 0 \end{cases}$$

**Test de Skewness :**

$$V_1 = \frac{|\hat{\beta}_1^{\frac{1}{3}} - 0|}{\sqrt{\frac{6}{n}}} = \frac{|-1.406884 - 0|}{\sqrt{\frac{6}{43}}} = 3.7663171 > 1.96$$

On rejette alors l'hypothèse  $H_0 : V_1 > 1.96$  : la distribution est symétrique. Elle est donc asymétrique.

#### Test de kurtosis :

$$V_2 = \frac{|\hat{\beta}_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|9.667926 - 3|}{\sqrt{\frac{24}{43}}} = 8.9252292 > 1.96$$

On rejette l'hypothèse  $H_0 : V_2 = 0$  : "aplatissement normale". Donc elle a un aplatissement leptokurtique.

#### Test de Jarque-Bera

$JB = 93.84486 > \chi_2^2 = 5.99$ , Nous rejettons  $H_0$  : normalité des résidus.

#### Test d'hétéroscédasticité (l'existence d'effet ARCH)

Heteroskedasticity Test: ARCH				
F-statistic	0.777356	Prob. F(1,40)	0.3832	
Obs*R-squared	0.800664	Prob. Chi-Square(1)	0.3709	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/08/23 Time: 12:48				
Sample (adjusted): 1981 2022				
Included observations: 42 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	45517.56	23516.77	1.935536	0.0600
RESID^2(-1)	0.137897	0.156403	0.881678	0.3832
R-squared	0.019063	Mean dependent var	52620.03	
Adjusted R-squared	-0.005460	S.D. dependent var	142796.3	
S.E. of regression	143185.6	Akaike info criterion	26.62812	
Sum squared resid	8.20E+11	Schwarz criterion	26.71086	
Log likelihood	-557.1905	Hannan-Quinn criter.	26.65845	
F-statistic	0.777356	Durbin-Watson stat	2.009687	
Prob(F-statistic)	0.383218			

FIGURE 3.16 – Table d'effet ARCH

A partir de la figure nous remarquons que la probabilité de  $RESID^2$  est supérieure à 0.05, donc nous acceptons l'hypothèse nulle d'hétéroscédasticité ; il n'existe pas d'effet ARCH.

Par conséquent, l'estimation du modèle ARIMA (1,1,1) est validée, et il s'écrit comme suit :

$$INV_t = 1.92INV_{t-1} - 0.92INV_{t-2} - 0.44\varepsilon_{t-1} + \varepsilon_t$$

Où  $\varepsilon_t$  est un processus bruit blanc.

### 3.1.4 Prévisions

Sachant que nous avons arrêter notre étude en 2022, nous allons faire nos prévisions sur les 4 années qui suivent :

Sachant que :

$$DINV_t = INV_t - INV_{t-1}$$

Nous remplaçons t par t+h dans l'équation nous obtenons :

$$INV_{t+h} = 1.92INV_{t+h-1} - 0.92INV_{t+h-2} - 0.44\varepsilon_{t+h-1} + \varepsilon_{t+h}$$

Avec  $h=1, \dots, 4$

Nous obtenons les prévisions suivantes :

INV	Prévision
2023	578.2947
2024	533.3512
2025	491.9005
2026	453.6713

TABLE 3.2 – Les prévisions de INV

Le graphe des prévisions est donné comme suit :

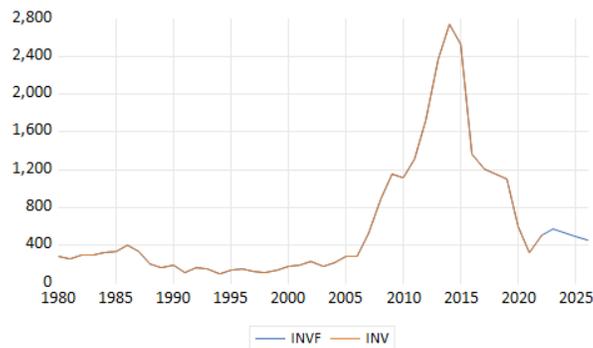


FIGURE 3.17 – Graphe de prévision

## 3.2 Etude de la série prix du pétrole $PRIX_t$

### 3.2.1 Identification

La série  $PRIX_t$  représente l'évolution annuelle des prix du pétrole sur une période allant de 1980 à 2022.

#### Analyse préliminaire de la série $PRIX_t$

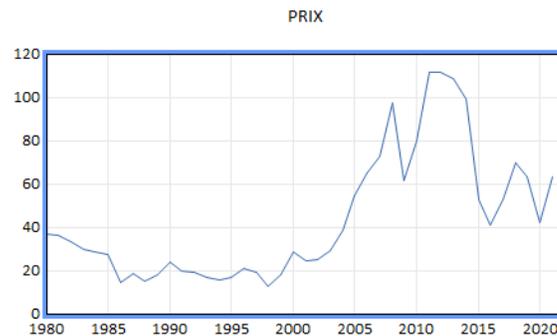


FIGURE 3.18 – Evolution de la série  $PRIX_t$

La représentation graphique de la série montre qu'elle est affectée d'une tendance qui peut être stochastique ou déterministe.

#### Corrélogramme de la série $PRIX_t$

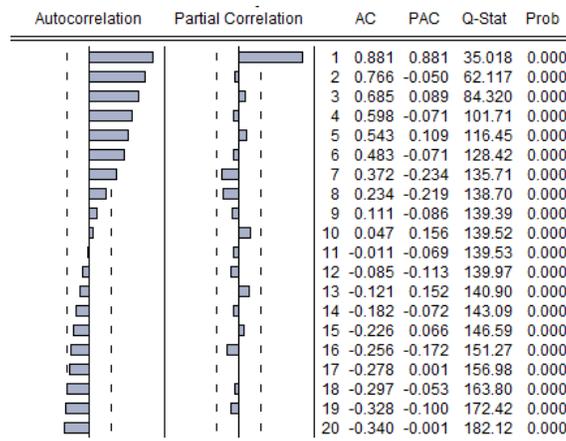


FIGURE 3.19 – Corrélogramme de la série  $PRIX_t$

D'après le corrélogramme la fonction d'autocorrélation simple et partiel diminue lentement vers zéro, on remarque aussi de nombreux pics sortant de la bande de confiance. Donc la série brute nous indique une non stationnarité de la série.

### Test de la stationnarité de la série $PRIX_t$

#### Test de Dickey-Fuller de la série $PRIX_t$

La meilleure façon de la rendre stationnaire est de la différencier d'ordre 1. La nouvelle série sera notée  $DPRIX_t$ .

Après la 1<sup>ère</sup> différenciation : nous rejetons l'hypothèse nulle de racine unitaire, alors nous acceptons l'hypothèse de stationnarité (La série est stationnaire).

La tendance et la constante ne sont pas significatives.

#### Corrélogramme de la série $DPRIX_t$ :

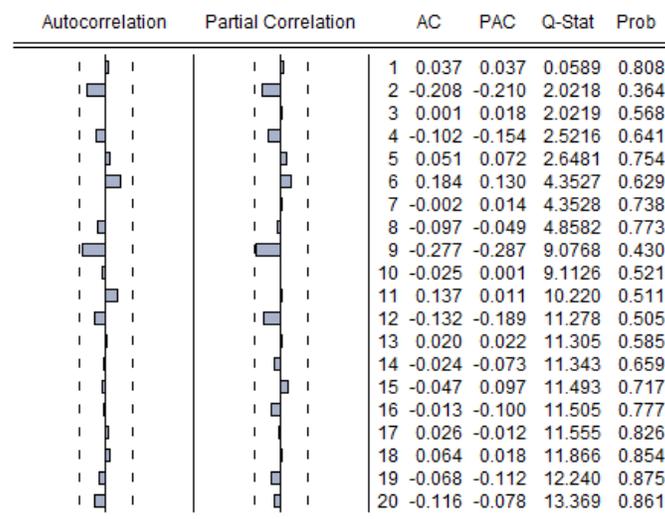


FIGURE 3.20 – Corrélogramme de la série  $DPRIX_t$

L'analyse du corrélogramme simple et partiel de la série montre que tous les pics sont à l'intérieur de l'intervalle de confiance, donc la série  $DPRIX_t$  représente un processus bruit blanc (sans mémoire).

### 3.3 Etude de la série des volumes de découvertes (VD)

#### 3.3.1 Identification

La série  $VD_t$  représente l'évolution annuelle des volumes de découvertes pétrolière sur une période allant de 1980 à 2022.

##### Analyse préliminaire de la série $VD_t$

La représentation graphique est donnée par le graphe suivant :

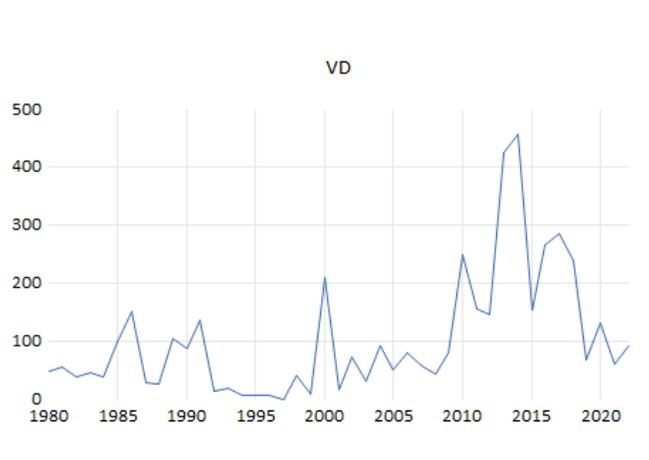
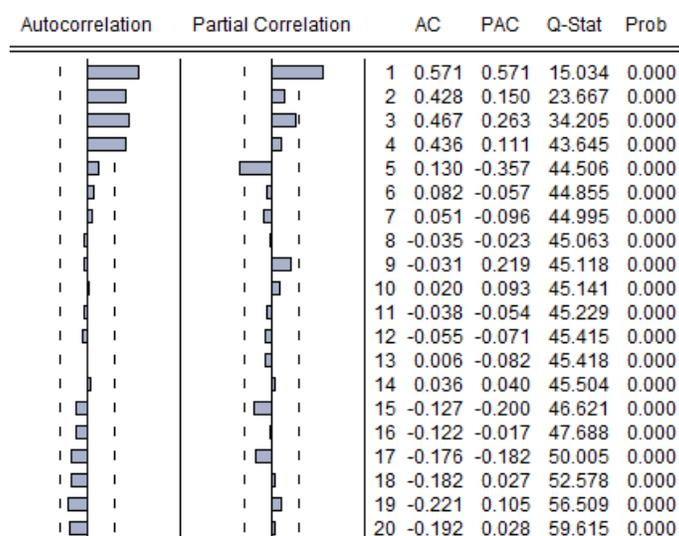


FIGURE 3.21 – Évolution de la série  $VD_t$

La représentation graphique de la série montre qu'elle est affectée d'une tendance qui peut être stochastique ou déterministe.

##### Corrélogramme de la série $VD_t$

Le corrélogramme d'autocorrélation simple et partiel diminue lentement vers zéro, on remarque aussi de nombreux pics sortant de la bande de confiance. Donc, la série brute paraît non stationnaire. Donc, pour confirmer ou infirmer notre hypothèse nous appliquant le test d'ADF.

FIGURE 3.22 – Corrélogramme de la série  $VD_t$ **Test de stationnarité de la série  $VD_t$** **Test de Dickey-Fuller Augmenté de la série  $VD_t$** 

D'après la table d'ADF nous avons trouvé que la tendance et la constante ne sont pas significatives et la série possède une racine unitaire. Donc, la série est non stationnaire de type DS.

La meilleure façon de la rendre stationnaire est de différencier d'ordre 1.

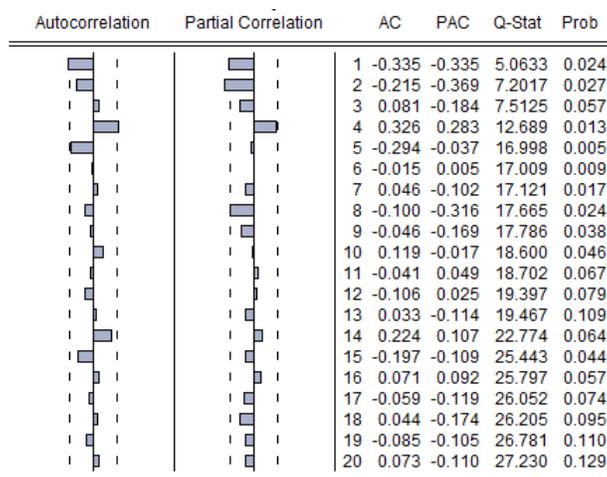
La nouvelle série sera notée  $DVD_t$ .

**Étude de la série  $DVD_t$** **Test de Dickey-Fuller sur la série  $DVD_t$** 

D'après la table ADF du **modèle 3** et **modèle 2** nous rejetons l'hypothèse nulle de racine unitaire, alors nous acceptons l'hypothèse de stationnarité  $H_0$ . La série est stationnaire. Notons que la tendance et la constante ne sont pas significatives.

**Identification du modèle :****Corrélogramme de la série  $DVD_t$  :**

En observant le corrélogramme suivant de la série  $DVD_t$  nous remarquons qu'il y'a quelques pics qui ne sont pas significativement différents de zéro. Par conséquent nous avons proposé le modèle : ARIMA(1,1,1).

FIGURE 3.23 – Corrélogramme de la série  $DVD_t$ 

### 3.3.2 Estimations des paramètres

#### Estimation des paramètres du modèle ARIMA (1,1,1)

Le tableau suivant contient l'estimation du modèle ARIMA (1,1,1), qui a été choisi parmi d'autres modèles estimés en se basant sur les critères de pouvoir prédictif, à savoir :  $R^2$  : maximum; AIC, SCH : minimum et sur la statistique de Durbin Watson proche de 2. ce qui présage un bon ajustement.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.931213	0.077023	12.09000	0.0000
MA(1)	-0.473376	0.171651	-2.757776	0.0087
SIGMASQ	7366.664	1081.120	6.813918	0.0000
R-squared	0.331452	Mean dependent var		102.4998
Adjusted R-squared	0.298025	S.D. dependent var		106.2133
S.E. of regression	88.98968	Akaike info criterion		11.90794
Sum squared resid	316766.6	Schwarz criterion		12.03081
Log likelihood	-253.0207	Hannan-Quinn criter.		11.95325
Durbin-Watson stat	1.952761			
Inverted AR Roots	.93			
Inverted MA Roots	.47			

FIGURE 3.24 – Estimation des paramètres du modèle ARIMA(1,1,1)

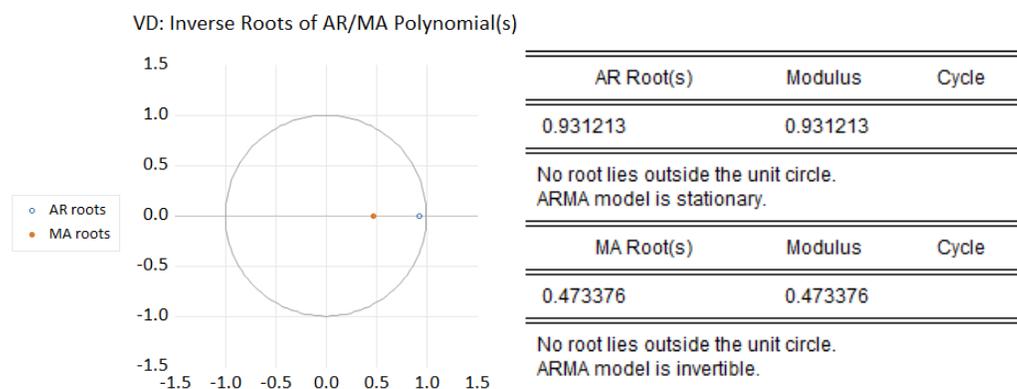


FIGURE 3.25 – Racines inverses du(des) polynôme(s) MA/AR

### Tests sur les résidus du modèle

La représentation graphique de la série résiduelle, actuelle et estimée est la suivante :

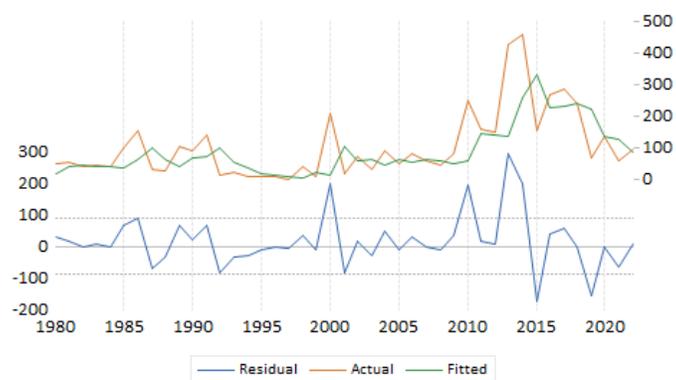


FIGURE 3.26 – Graphe des séries résiduelles actuelles et estimées

A partir de la représentation graphique, nous constatons que le graphe de la série estimée est presque semblable à celui de la série réelle à quelques pics près. En effet, le modèle explique 33.14% de la variation total de la variable dépendante ( $R^2 = 0.3314$ ).

### Corrélogramme des résidus

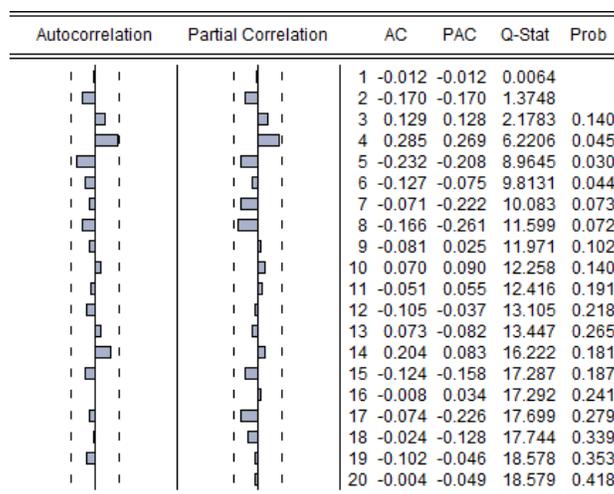


FIGURE 3.27 – Corrélogramme des résidus

D'après l'analyse du corrélogramme des résidus, nous remarquons que tous les pics sont à l'intérieure de l'intervalle de confiance et les probabilités correspondantes sont supérieures à 0.05. Donc les résidus forment un bruit blanc, ce qui est confirmée par la statistique de LJUNG-BOX (Q-stat) qui est inférieure à la valeur théorique  $\chi^2$  (h-3) de quel que soit le retard h, on particulier pour h=10 on a : Q-stat (10) = 12.258 <  $\chi^2$  (7) = 14.06 au seuil 0.05, donc les résidus forment un bruit blanc.

Les tests étants tous vérifiés. Donc, **le modèle ARIMA(1,1,1) est validé.**

#### Test de normalité sur les résidus de ARIMA(1,1,1)

L'histogramme de la distribution et les valeurs empiriques des Skewness-Kurtosis et la statistique de Jarque-Bera sont données par cette figure :

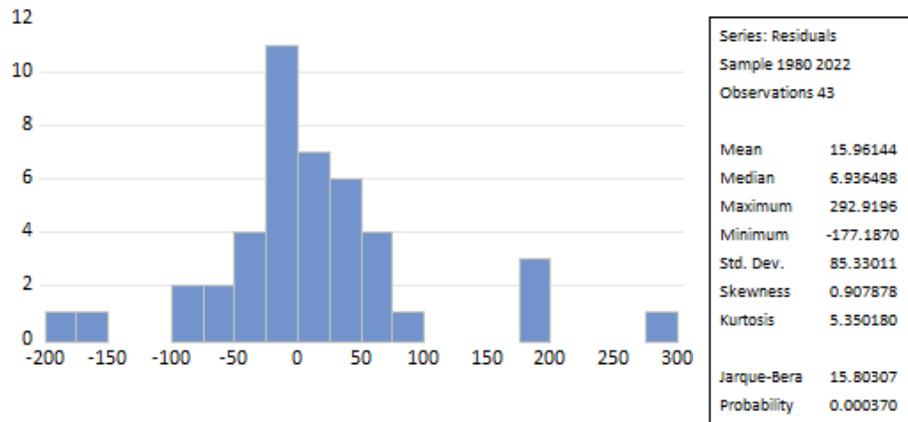


FIGURE 3.28 – Histogramme de normalité

**Test de Skewness :**

$$V_1 = \frac{|\hat{\beta}_1^{\frac{1}{2}} - 0|}{\sqrt{\frac{6}{n}}} = \frac{|0.907878 - 0|}{\sqrt{\frac{6}{43}}} = 2.43044 > 1.96$$

On rejette alors l'hypothèse  $H_0 : |V_1| > 1,96$  : la distribution est symétrique. Elle est donc asymétrique.

**Test de kurtosis :**

$$V_2 = \frac{|\hat{\beta}_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|5.350180 - 3|}{\sqrt{\frac{24}{43}}} = 3.14579 > 1.96$$

On rejette l'hypothèse  $H_0 : V_2 = 0$  : "aplatissement normale".  
Donc elle a un aplatissement leptokurtique.

**Test de Jarque-Bera**

$JB = 15.80307 > \chi_2^2$ , Nous rejettons  $H_0$  : normalité des résidus.

**Test d'hétéroscédasticité (l'existence d'effet ARCH)**

Heteroskedasticity Test: ARCH				
F-statistic	2.618535	Prob. F(1,40)	0.1135	
Obs*R-squared	2.580531	Prob. Chi-Square(1)	0.1082	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 07/02/23 Time: 15:40				
Sample (adjusted): 1981 2022				
Included observations: 42 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5650.596	2758.014	2.048792	0.0471
RESID^2(-1)	0.248005	0.153261	1.618189	0.1135
R-squared	0.061441	Mean dependent var	7520.781	
Adjusted R-squared	0.037977	S.D. dependent var	16546.19	
S.E. of regression	16228.96	Akaike info criterion	22.27343	
Sum squared resid	1.05E+10	Schwarz criterion	22.35618	
Log likelihood	-465.7420	Hannan-Quinn criter.	22.30376	
F-statistic	2.618535	Durbin-Watson stat	2.005538	
Prob(F-statistic)	0.113482			

FIGURE 3.29 – Table d'effet Arch

A partir de la figure nous remarquons que la probabilité de RESID<sup>2</sup> est supérieure à 0.05, donc nous rejetons l'hypothèse nulle d'hétéroscédasticité ; il n'existe pas d'effet ARCH ou GARCH.

Par conséquent, l'estimation du modèle ARIMA (1,1,1) est validée, et il s'écrit comme suit :

$$VD_t = 1.93VD_{t-1} - 0.93VD_{t-2} + 0.47\varepsilon_{t-1} + \varepsilon_t$$

Où  $\varepsilon_t$  est un processus bruit blanc.

### 3.3.3 Prévisions

Sachant que nous allons arrêter notre étude en 2022, nous devons faire nos prévisions sur les 4 années qui suivent.

Sachant que :

$$DVD_t = VD_t - VD_{t-1}$$

Nous remplaçons t par t+h dans l'équation nous obtenons :

$$VD_{t+h} = 1.93VD_{t+h-1} - 0.93VD_{t+h-2} + 0.47\varepsilon_{t+h-1} + \varepsilon_{t+h}$$

Avec  $h=1, \dots, 4$

Nous obtenons les prévisions suivantes :

VD	Prévision
2023	83.3086
2024	77.5780
2025	72.2416
2026	67.2723

TABLE 3.3 – Les previsions de VD

Le graphe des prévisions est donné comme suit :

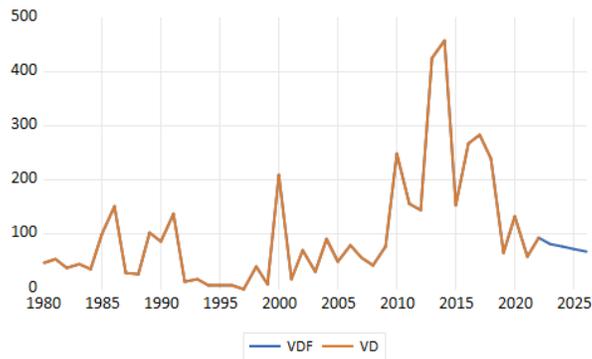


FIGURE 3.30 – Graphe de prévision

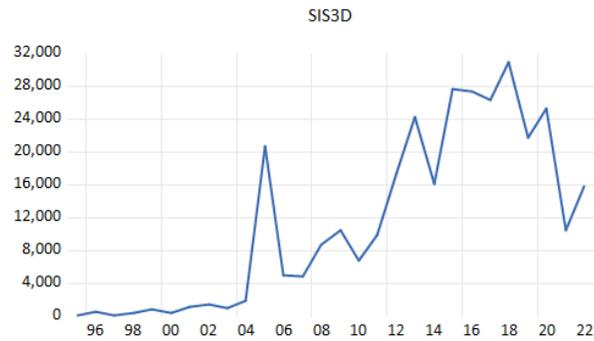
## 3.4 Etude de la série prix du pétrole $SIS3D_t$

### 3.4.1 Identification

La série  $SIS3D_t$  représente l'évolution annuelle de la sismique 3D pour l'exploration pétrolière sur une période allant de 1995 à 2022.

#### Analyse préliminaire de la série $SIS3D_t$

La représentation graphique est donnée par le graphe suivant :

FIGURE 3.31 – Evolution de la série  $SIS3D_t$ 

La représentation graphique de la série montre qu'elle est affectée d'une tendance qui peut être stochastique ou déterministe.

#### Corrélogramme de la série $SIS3D_t$

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.755	0.755	17.719	0.000	
2	0.714	0.336	34.203	0.000	
3	0.620	0.023	47.131	0.000	
4	0.512	-0.112	56.321	0.000	
5	0.361	-0.234	61.080	0.000	
6	0.225	-0.185	63.012	0.000	
7	0.188	0.159	64.421	0.000	
8	0.071	-0.011	64.630	0.000	
9	-0.059	-0.213	64.782	0.000	
10	-0.081	0.045	65.089	0.000	
11	-0.087	0.159	65.464	0.000	
12	-0.166	-0.110	66.907	0.000	
13	-0.197	-0.082	69.080	0.000	
14	-0.281	-0.320	73.814	0.000	
15	-0.292	-0.093	79.315	0.000	
16	-0.406	-0.085	90.828	0.000	
17	-0.390	0.174	102.42	0.000	
18	-0.401	0.011	115.91	0.000	
19	-0.357	0.142	127.82	0.000	
20	-0.345	-0.002	140.29	0.000	

FIGURE 3.32 – Corrélogramme de la série  $SIS3D_t$

D'après le corrélogramme la fonction d'autocorrélation simple et partielle diminue lentement vers zéro, on remarque aussi de nombreux pics sortant de la bande de confiance. Donc la série brute nous indique une non stationnarité de la série.

#### Test de stationnarité de la série $SIS3D_t$

##### Test de Dickey-Fuller Augmenté de la série $SIS3D_t$

D'après la table de ADF nous avons trouvé que la tendance et la constante ne sont pas significatives et la série possède une racine unitaire (série non stationnaire de type DS).

La meilleure façon de la rendre stationnaire est de différencier d'ordre 1.

La nouvelle série sera notée  $DSIS3D_t$ .

#### Étude de la série $DSIS3D_t$

##### Test de Dickey-Fuller sur la série $DSIS3D_t$

D'après la table ADF du **modèle 3** et **modèle 2** nous rejetons l'hypothèse nulle de racine unitaire, alors nous acceptons l'hypothèse de stationnarité.  $H_0$  La série est stationnaire. Il y'a lieu de noter que la tendance et la constante ne sont pas significatives.

#### Identification du modèle :

##### Corrélogramme de la série $DSIS3D_t$ :

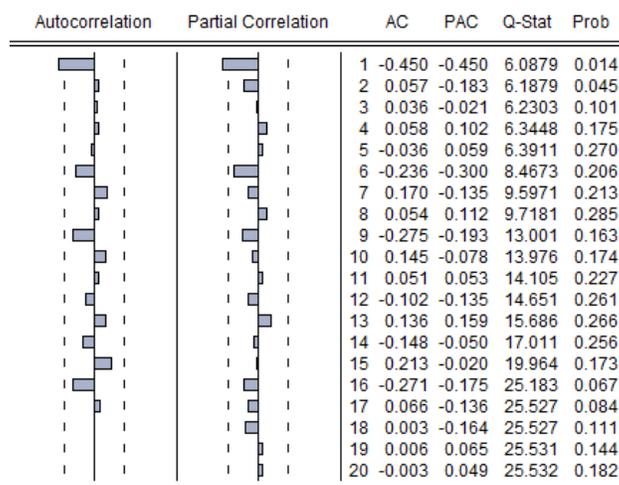


FIGURE 3.33 – Corrélogramme de la série  $DSIS3D_t$

En observant le corrélogramme de la série  $DSIS3D_t$  stationnaire, on remarque que seul le 1er terme du corrélogramme simple et partiel est significativement différent de zéro. Par conséquent nous avons proposé le modèle : ARIMA(1,1,1).

### 3.4.2 Estimation des paramètres

#### Estimation de paramètre ARIMA (1,1,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.960724	0.055198	17.40510	0.0000
MA(1)	-0.402856	0.246283	-1.635743	0.1144
SIGMASQ	38488966	7417995.	5.188595	0.0000
R-squared	0.640747	Mean dependent var		11309.92
Adjusted R-squared	0.612006	S.D. dependent var		10540.59
S.E. of regression	6565.641	Akaike info criterion		20.58097
Sum squared resid	1.08E+09	Schwarz criterion		20.72371
Log likelihood	-285.1336	Hannan-Quinn criter.		20.62461
Durbin-Watson stat	2.094230			
Inverted AR Roots	.96			
Inverted MA Roots	.40			

FIGURE 3.34 – Estimation de paramètre ARIMA (1,1,1)

Le tableau contient l'estimation du modèle ARIMA (1,1,1), qui a été retenu parmi d'autres modèles estimés en se basant sur les critères de pouvoir prédictif, à savoir :  $R^2$  : maximum ; AIC, SCH : minimum et sur la statistique de Durbin Watson qui est égale à 2.09, ce qui présage d'un bon ajustement.

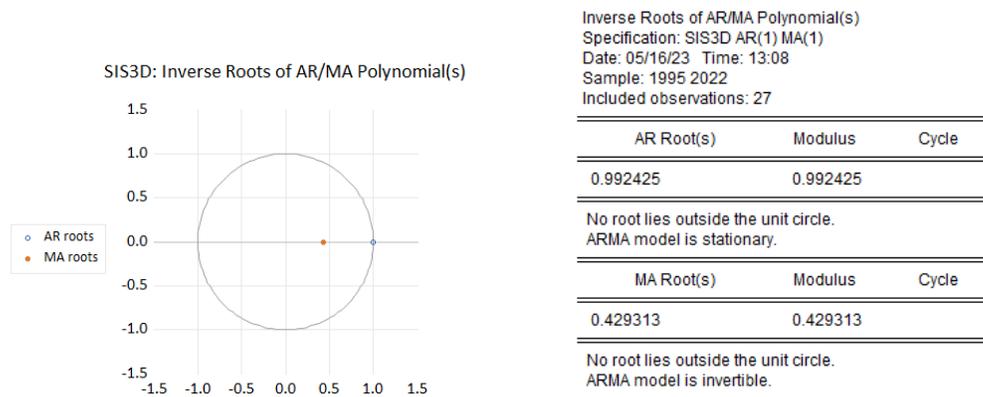


FIGURE 3.35 – Racines inverses du(des) polynôme(s) MA/AR

### Tests sur les résidus du modèle

La représentation graphique de la série résiduelle, actuelle et estimée est la suivante :

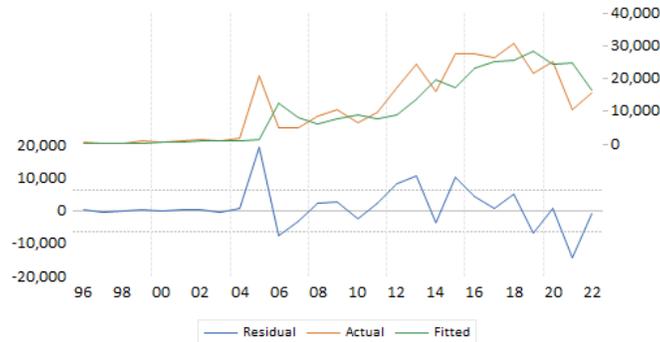


FIGURE 3.36 – Graphe des séries résiduelles actuelles et estimées

A partir de la représentation graphique, nous constatons que le graphe de la série estimée est presque semblable à celui de la série réelle à quelques pics près, donc le modèle explique 64.07% de la variation totale de la variable dépendante ( $R^2=0.640747$ ).

### Corrélogramme des résidus

D'après l'analyse du corrélogramme des résidus, nous remarquons que tous les pics sont à l'intérieur de l'intervalle de confiance et les probabilités correspondantes sont supérieures à 0.05.

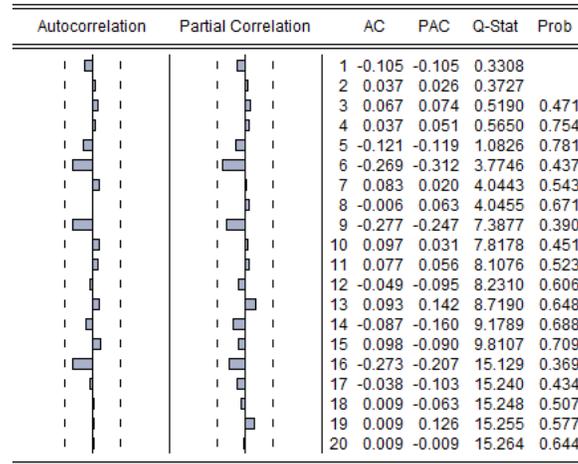


FIGURE 3.37 – Corrélogramme des résidus

**Test de normalité sur les résidus de ARIMA(1,1,1)**

L'histogramme de la distribution et les valeurs empiriques des Skewness-Kurtosis et la statistique de Jarque-Bera sont données par cette figure :

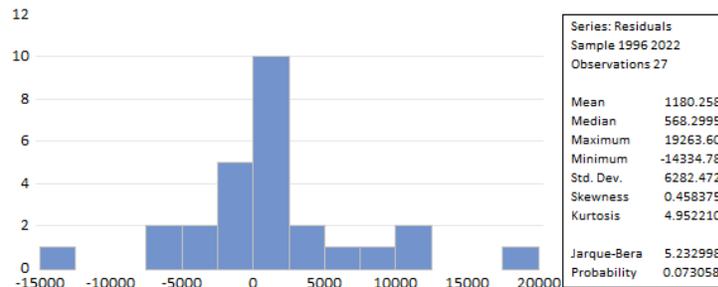


FIGURE 3.38 – Histogramme de normalité

**Test de Skewness :**

$$V_1 = \frac{|\hat{\beta}_1^{\frac{1}{3}} - 0|}{\sqrt{\frac{6}{n}}} = \frac{|0.458375 - 0|}{\sqrt{\frac{6}{28}}} = 0.97236 < 1.96$$

On rejette alors l'hypothèse  $H_0 : |V_1| > 1,96$  : la distribution est symétrique. Elle est donc asymétrique.

**Test de kurtosis :**

$$V_2 = \frac{|B_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|4.952210 - 3|}{\sqrt{\frac{24}{28}}} = 2.070631 > 1.96$$

On rejette l'hypothèse  $H_0 : V_2 = 0$  : "aplatissement normale".  
Donc elle a un aplatissement leptokurtique.

**Test de Jarque-Bera**

$JB = 5.232998 < \chi_2^2 = 5.99$  Nous acceptons  $H_0$  : normalité des résidus.

**Test d'hétéroscédasticité (l'existence d'effet ARCH)**

Heteroskedasticity Test: ARCH				
F-statistic	0.048212	Prob. F(1,24)	0.8281	
Obs*R-squared	0.052124	Prob. Chi-Square(1)	0.8194	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/16/23 Time: 13:29				
Sample (adjusted): 1997 2022				
Included observations: 26 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	42742130	18562337	2.302626	0.0303
RESID^2(-1)	-0.044771	0.203902	-0.219571	0.8281
R-squared	0.002005	Mean dependent var	40911117	
Adjusted R-squared	-0.039578	S.D. dependent var	82935578	
S.E. of regression	84560879	Akaike info criterion	39.41765	
Sum squared resid	1.72E+17	Schwarz criterion	39.51442	
Log likelihood	-510.4294	Hannan-Quinn criter.	39.44551	
F-statistic	0.048212	Durbin-Watson stat	2.007707	
Prob(F-statistic)	0.828062			

FIGURE 3.39 – Teste d'effet ARCH

La probabilité de  $RESID^2$  dans la figure précédente est supérieure à 0.05, donc nous acceptons l'hypothèse nulle d'hétéroscédasticité ; il n'existe pas d'effet ARCH ou GARCH.

Par conséquent, l'estimation du modèle ARIMA (1,1,1) est validée, et il s'écrit comme suit :

$$SIS3D_t = 1.96SIS3D_{t-1} - 0.96SIS3D_{t-2} - 0.4\varepsilon_{t-1} + \varepsilon_t$$

Ou  $\varepsilon_t$  est un processus bruit blanc.

### 3.4.3 Prévisions

Sachant que nous avons un historique de données jusqu'à 2022, nous proposons de faire nos prévisions sur les 4 années qui suivent : Sachant que :

$$DSIS3D_t = SIS3D_t - SIS3D_{t-1}$$

Nous remplaçons t par t+h dans l'équation nous obtenons :

$$SIS3D_t = 1.96SIS3D_{t-1} - 0.96SIS3D_{t-2} - 0.4\varepsilon_{t-1} + \varepsilon_t$$

Avec  $h=1, \dots, 4$  Nous obtenons les prévisions suivantes :

SIS3D	Prévision
2023	14623.37
2024	14096.23
2025	13588.10
2026	13098.28

TABLE 3.4 – Les prévisions de SIS3D

Le graphe des prévisions est donné comme suit :

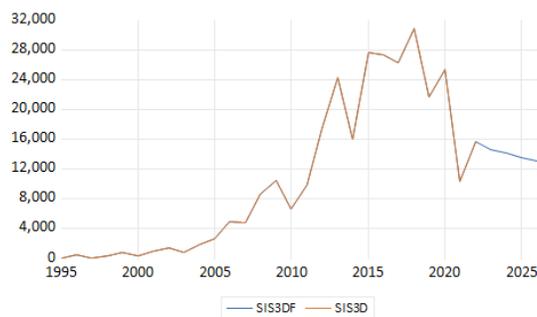


FIGURE 3.40 – Graphe de prévision

**Conclusion :**

D'après l'étude des trois séries  $INV_t$ ,  $Prix_t$ ,  $VD_t$  et  $SIS3D_t$  nous concluons pour :

- La série investissement (INV) :
  - INV est intégrée d'ordre 1 (non stationnaire) :  $INV \rightarrow I(1)$ .
  - DINV est stationnaire :  $DINV \rightarrow I(0)$ .
  - Après la différentiation la série  $INV \sim ARIMA(1,1,1)$
- La série PRIX (PRIX) :
  - PRIX est intégrée d'ordre 1 (non stationnaire) :  $PRIX \rightarrow I(1)$ .
  - DPRIX est stationnaire :  $DPRIX \rightarrow I(0)$ .
  - Après la différentiation la série est sans mémoire.
- La série volume de découvertes (VD) :
  - VD est intégrée d'ordre 1 (non stationnaire) :  $VD \rightarrow I(1)$ .
  - DVD est stationnaire :  $DVD \rightarrow I(0)$ .
  - Après la différentiation la série  $VD \sim ARIMA(1,1,1)$
- La série sismique trois dimensions (SIS3D) :
  - SIS3D est intégrée d'ordre 1 (non stationnaire) :  $SIS3D \rightarrow I(1)$ .
  - DSIS3D est stationnaire :  $DSIS3D \rightarrow I(0)$ .
  - Après la différentiation la série  $INV \sim ARIMA(1,1,1)$

# Chapitre 4

## Approche de la cointégration

### Introduction :

Dans cette section, on définira la notion de séries intégrées et le concept de cointégration. Il y a la cointégration au sens d'Engel et Granger, la cointégration au sens de Johansen et les tests de cointégration avec le modèle à correction d'erreur. La théorie de la cointégration permet d'étudier des séries non stationnaires mais dont une combinaison linéaire est stationnaire. Elle permet ainsi de spécifier des relations stables à long terme tout en analysant conjointement la dynamique de court terme des variables considérées.

### 4.1 Approche de la cointégration bivariée

#### 4.1.1 Cointégration et les modèles à correction d'erreur

##### Séries intégrées :

On dit qu'une série  $X_t$  est intégrée d'ordre  $d$ , on note  $X_t \sim I(d)$ , s'il convient de la différencier  $d$  fois afin de la rendre stationnaire.

Les propriétés les plus importantes des séries intégrées sont :

- Si la série  $X_t$  est intégrée d'ordre zéro,  $I(0)$ , elle est stationnaire (car dans ce cas elle n'est pas affectée d'une tendance).

- Si  $X_t \rightarrow I(d)$  alors  $\alpha + \beta X_t \rightarrow I(d)$  où  $\alpha$  et  $\beta$  sont des constantes, avec  $\beta \neq 0$ .
- Si  $\begin{cases} X_t \rightarrow I(0) \\ Y_t \rightarrow I(0) \end{cases}$  alors  $\alpha X_t + \beta Y_t \rightarrow I(0)$

Où  $\alpha$  et  $\beta$  sont des constantes non nulles.

- Si  $\begin{cases} X_t \rightarrow I(0) \\ Y_t \rightarrow I(1) \end{cases}$  alors  $\alpha X_t + \beta Y_t \rightarrow I(1)$

La série  $z_t$  est non stationnaire puisqu'elle résulte de la sommation d'une série affectée d'une tendance et une autre série stationnaire.

- Si  $\begin{cases} X_t \rightarrow I(d_1) \\ Y_t \rightarrow I(d_2) \end{cases}$  alors  $\alpha X_t + \beta y_t \rightarrow I(\max(d_1, d_2))$

Où  $\alpha$  et  $\beta$  sont des constantes non nulles.

Toute combinaison linéaire des séries intégrées d'ordre différent est généralement intégrée à l'ordre le plus élevé. Ceci veut dire qu'à long terme, l'écart entre les séries n'ira qu'en augmentation, et les trajectoires sur lesquelles se situent sont totalement différentes.

### Cointégration

Le traitement des séries chronologiques longues impose de tester une éventuelle cointégration (relation) entre les variables. En effet, le risque d'estimer des relations **fallacieuses** et d'interpréter les résultats de manière erronées est très élevés, car il est connu qu'une régression appliquée à des variables non stationnaires peut fournir des résultats curieux, appelés dans la littérature Engelo-saxonne **spurious régressions** ou **régressions fallacieuses**.

L'analyse de la cointégration permet d'identifier la relation entre plusieurs variables. Cette notion a été introduite dès 1974 par Engle et Newbold, puis formalisé par Engle et Granger en 1987 et enfin par Johansen en 1991 et 1995.

Dans la solution aux **régressions fallacieuses** est **la cointégration**.

### Conditions de cointégration

Deux séries  $X_t$  et  $Y_t$  sont dites cointégrées si les deux conditions suivantes sont vérifiées :

- Elles sont intégrées d'ordre  $d$ ;

- La combinaison linéaire de ces deux séries permet de se ramener à une série d'ordre d'intégration inférieur.

$$\text{Soit : } \begin{cases} X_t \rightarrow I(d) \\ Y_t \rightarrow I(b) \end{cases}$$

tel que  $\alpha_1 X_t + \alpha_2 Y_t \rightarrow I(d - b)$  avec  $d \geq 0$ ,  $b \geq 0$ .

On note  $X_t, Y_t \rightarrow CI(d, b)$  ou  $[\alpha_1, \alpha_2]$  est le vecteur de cointégration

Dans le cas général à  $k$  variables, on a :

$$X_{1,t} \rightarrow I(d)$$

$$X_{2,t} \rightarrow I(d)$$

⋮

$$X_{k,t} \rightarrow I(d)$$

On note  $X_t = [X_{1,t}, X_{2,t}, \dots, X_{k,t}]$

Il existe un vecteur de cointégration  $[\alpha_1, \alpha_2, \dots, \alpha_k]$  de dimension  $(k \times 1)$  tel que  $\alpha X_t \rightarrow I(d - b)$ , les  $k$  variables sont cointégrées et le vecteur de cointégration est  $\alpha$ .

### 4.1.2 Estimation des modèles à correction d'erreur et tests de cointégration (au sens d'Engel et Granger)

#### Tests de cointégration entre deux variables :

Nous représentons ici l'algorithme en deux étapes d'Engle et Granger.

**Étape 1 :** tester l'ordre d'intégration des deux variables : Une condition nécessaire de cointégration est que les séries doivent être intégrées de même ordre.

Si les séries ne sont pas intégrées de même ordre, elles ne peuvent être cointégrées.

Il convient donc de vérifier l'ordre d'intégration des chroniques étudiées à l'aide par exemple du test de Dickey-Fuller (simple ou augmenté).

Si les séries considérées ne sont pas intégrées de même ordre, il n'y a alors pas de risque de cointégration et la procédure s'arrête à cette première étape.

**Étape 2 :** estimation de la relation de long terme

Si on a :

$$X_t \sim I(1) \text{ et } Y_t \sim I(1)$$

On estime par les MCO la relation de long terme :

$$Y_t = \alpha X_t + \beta + \varepsilon_t$$

Pour qu'il y ait cointégration, il faut que le résidu issu de la régression soit stationnaire :

$$e_t = Y_t - \hat{\alpha}X_t - \hat{\beta} \sim I(0)$$

La stationnarité du résidu est testée à l'aide du test de Dickey-Fuller (simple et augmenté).

Dans tous les tests on a :

$$\begin{cases} H_0 : \text{absence de cointégration} \\ H_1 : \text{existence de cointégration} \end{cases}$$

On remarque ici que la relation porte sur les résidus estimés et non pas sur les « vrais » résidus de l'équation de cointégration. Par conséquent, nous ne pouvons pas nous référer aux tables de Dickey-Fuller pour mener le test de stationnarité. Il faut se référer aux tables de McKinnon.

Si le résidu est stationnaire nous pouvons alors estimer un modèle à correction d'erreur (MCE) qui intègre les variables en différence et en niveau (le théorème de la représentation de Granger met en évidence le lien entre cointégration et modèle à correction d'erreur). L'emploi d'un modèle à correction d'erreur dans le cas de la cointégration permet d'obtenir des prévisions plus fiables que si on avait utilisé la relation statique qui est une régression fallacieuse causée par la non stationnarité des séries.

### **Modèle à correction d'erreur(MCE)**

Lorsque les variables sont non stationnaires et cointégrées, les résultats d'Engel et Granger(1987) montrent que la bonne spécification du modèle consiste à utiliser une forme à correction d'erreur, qui permet de se ramener à une écriture ne faisant intervenir que des variables stationnaires, et dans lesquels il est possible d'effectuer des tests sur les paramètres du modèle. Les modèles à correction d'erreur permettent de corriger les erreurs ce qui conduit à l'équilibre à long terme en prenant compte de la relation à court terme pour atteindre une cible prédéterminée.

La représentation des modèles à correction d'erreur combine deux types de variables :

- Des variables en niveau qui s'interprètent comme les déterminants de l'équilibre de long terme.
- Des variables en différence première (stationnaires) qui représentent les fluctuations de court terme.

Statistiquement, un modèle ECM est un processus qui peut se mettre sous la forme :

**Étape 1** : estimation par méthode MCO de la relation de long terme :

$$Y_t = \hat{\alpha} + \hat{\beta}X_t + e_t$$

**Étape 2** : estimation par méthode MCO de la relation du modèle dynamique (court terme).

$$\Delta Y_t = \alpha_1 \Delta X_t + \alpha_2 e_{t-1} + v_t, \quad \alpha_2 < 0$$

Le coefficient  $\alpha_2$  (force de rappel vers l'équilibre) doit être significativement négatif ; dans le cas contraire, il convient de rejeter une spécification de type ECM. En effet, le mécanisme de correction d'erreur (rattrapage qui permet de tendre vers la relation de long terme) irait alors en sens contraire et s'éloignerait de la cible de long terme.

La procédure en deux étapes conduit à une estimation convergente des coefficients du modèle et les écarts types des coefficients peuvent s'interpréter de manière classique.

L'estimation d'un modèle à correction d'erreur ne pose donc pas de problème particulier et nécessite simplement le recours à une technique de Moindres Carrées Ordinaires MCO.

Le fait que les séries soient cointégrées et non stationnaires soulève un problème d'estimation. La bonne qualité statistique du modèle ( $R^2$  élevé et coefficients significatifs) est due au fait que les séries sont non stationnaires (puisqu'ont été différenciées), en régression directe de  $Y_t$ , sur  $X_t$ , lorsque  $Y_t, X_t \rightarrow CI(1,1)$ , l'utilisation de ce modèle à des fins prévisionnelles s'avère désastreuse. En effet, la relation mise en évidence par cette régression n'est pas réelle, elle découle simplement d'une relation entre deux tendances. Le problème est donc ; d'une part de retirer la relation commune de cointégration (la tendance commune), d'autre part, de rechercher la liaison réelle entre les variables ; **c'est l'objectif du modèle à correction d'erreur ECM.**

### Conclusion

La méthode d'Engel et Granger (1987) nous permet d'estimer aisément un modèle à correction d'erreur en deux étapes. Elle fournit également un certain nombre de tests de cointégration faciles à mettre en œuvre. L'inconvénient de cette approche est qu'elle ne permet pas de distinguer plusieurs vecteurs de cointégration, ceci est problématique dès lors que l'on étudie simultanément  $k$  variables, avec  $k > 2$ . On sait en effet que si l'on analyse le comportement de  $k$  variables (avec  $k > 2$ ), on peut avoir jusqu'à  $(k-1)$  relations de cointégration : l'approche d'Engel-Granger ne nous permettant d'obtenir qu'une seule relation de cointégration. Afin de pallier à cette difficulté, Johansen (1988) a proposé une approche multivariée de la cointégration fondée sur la méthode du maximum de vraisemblance. Cette approche, développée par la suite par Johansen et Juselius (1990) et Johansen (1991).

## 4.2 Approche multivariée de la cointégration : l'analyse de Johansen

Considérons un vecteur  $X_t$ , contenant  $k$  variables toutes  $I(1)$ , la représentation VAR( $p$ ) de  $X$ , est donnée par :

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

Où  $\varepsilon_t$  est un bruit blanc de moyenne nulle et de matrice variance-covariance  $\Omega$  :  $\varepsilon_t \sim BB(0, \Omega)$  et les  $\phi_i$  ( $i = 1, \dots, p$ ) sont des matrices de paramètres de taille  $(k \times k)$ .

On s'intéresse à l'hypothèse nulle suivante : il existe  $r$  relations de cointégration entre les  $k$  variables. En d'autres termes, sous l'hypothèse nulle  $H_0$ ,  $X_t$  est cointégré de rang  $r$ . on commence par déterminer les estimateurs des paramètres contraints par l'hypothèse de cointégration (c'est-à-dire que l'on cherche les estimateurs des paramètres sous l'hypothèse nulle).

### Estimateur du maximum de vraisemblance contraint :

On réécrit l'équation (1) sous la forme d'un modèle à correction d'erreur :

$$\Delta X_t = \Pi_1 \Delta X_{t-1} + \dots + \Pi_{p-1} \Delta X_{t-p+1} + \Pi_p \Delta X_{t-p} + \varepsilon_t \quad (2)$$

Où les matrices  $\Pi_i$  ( $i = 1, \dots, p$ ) sont de taille  $(k \times k)$ . Tous les termes de l'équation (2) sont  $I(0)$  sauf  $X_{t-p}$  qui est  $I(1)$ . Il y a donc un déséquilibre entre l'ordre

d'intégration du membre de gauche et les termes du membre de droite. Pour que, les deux membres de l'équation (2) soient I(0), une condition nécessaire qui est  $\Pi_p X_{t-p}$  soit I(0).

On pose :

$$\Pi_p = -\beta \alpha' \quad (3)$$

Où :

$\alpha'$  est une matrice (r, k) qui contient les r vecteurs de cointégration (r est donc le rang de cointégration);

$\beta$  est une matrice (k, r) qui contient les poids associés à chaque vecteur de cointégration. S'il y a r relations de cointégration alors :

$$Rg(\Pi_p) = r \quad (4)$$

Les tests de cointégration proposés par Johansen reposent sur cette condition. En vertu de (3), il est possible de réécrire l'équation (2) comme suit :

$$\Delta X_t = \Pi_1 \Delta X_{t-1} + \dots + \Pi_{p-1} \Delta X_{t-p+1} - \beta \alpha' X_{t-p} + \varepsilon_t \quad (5)$$

Pour estimer les différentes matrices, Johansen propose d'utiliser la méthode du maximum de vraisemblance Sous l'hypothèse de normalité de  $\varepsilon_t$ , la log-vraisemblance s'écrit :

$$\text{Log}L(\beta, \alpha, \Pi_1, \dots, \Pi_{p-1}, \Omega) = -\frac{kT}{2} \log(2\Pi) - \frac{T}{2} \log[\det(\Omega)] - \frac{1}{2} \sum_{t=1}^T \varepsilon_t' \Omega^{-1} \varepsilon_t \quad (6)$$

Où T est le nombre d'observations, k est le nombre de variables contenues dans X et  $\det(\Omega)$  désigne le déterminant de la matrice variance-covariance de  $\varepsilon_t$ .

Après l'obtention des estimateurs des matrices  $\Pi_i$  pour  $i = (1, \dots, p)$ ; la log vraisemblance devient :

$$\text{Log}(\beta, \alpha, \Omega) = -\frac{kT}{2} \log(2\Pi) - \frac{T}{2} \log[\det(\Omega)] - \frac{1}{2} \sum_{t=1}^T (e_{0t} + \beta \alpha e_{pt})' \Omega^{-1} (e_{0t} + \beta \alpha e_{pt}) \quad (7)$$

On maximise cette expression par rapport à  $\beta$  et  $\Omega$  pour  $\alpha$  fixé, on en déduit, alors :

$$\hat{\beta}(\alpha) = -S_{0p}\alpha(\alpha'S_{pp}\alpha)^{-1} \quad (8)$$

$$\hat{\Omega}(\alpha) = S_{00} - S_{0p}\alpha(\alpha'S_{pp}\alpha)^{-1}\alpha'S_{p0} \quad (9)$$

Où

$$S_{ij} = \frac{1}{T} \sum_{t=1}^T e_{it}e'_{jt}e'_{jt}$$

pour  $i, j = 0, p$ .

L'expression de  $S_{ij}$  nous donne les matrices de variance-covariance de  $e_{0t}$  et  $e_{pt}$ .

Johansen a montré que la concentration de la log-vraisemblance par rapport à  $\Omega$  permet d'obtenir l'expression suivante :

$$\text{Log}L(\alpha) = -\frac{kT}{2}\log(2\Pi) - \frac{T}{2}\log[\det(\hat{\Omega}(\alpha))] - \frac{kT}{2} \quad (10)$$

La log-vraisemblance ne dépend plus que de  $\hat{\Omega}$ .

### Estimation des vecteurs de cointégration :

Maximiser la log-vraisemblance par rapport à  $(\hat{\Omega}(\alpha))$  est équivalent à résoudre le problème de minimisation suivant :

$$\text{Min } \det(\hat{\Omega}(\alpha)) \iff \text{Min } \det[S_{00} - S_{0p}\alpha(\alpha'S_{pp}\alpha)^{-1}\alpha'S_{p0}] \quad (11)$$

La solution de ce problème est donnée par Anderson (1984) et se déduit d'une recherche de valeurs propres et de vecteurs propres. Ainsi pour obtenir les vecteurs de cointégration recherchés, il faut trouver les  $k$  valeurs propres de l'équation :

$$\det[\lambda S_{pp} - S_{p0}S_{00}^{-1}S_{0p}] = 0 \quad (12)$$

La résolution de cette équation nous donne les  $k$  valeurs propres estimées  $(\hat{\lambda}_1, \dots, \hat{\lambda}_k)$ ,  $\hat{\lambda}_1 > \dots > \hat{\lambda}_k$ , et les  $k$  vecteurs propres associés  $(\hat{V}_1, \dots, \hat{V}_r)$  avec la condition de normalisation  $\hat{V}'S_{pp}\hat{V} = I$  où  $\hat{V}$  est la matrice des vecteurs propres estimés.

On peut alors montrer que les estimateurs du maximum de vraisemblance des

vecteurs de cointégration sont les vecteurs propres  $(\hat{V}_1, \dots, \hat{V}_r)$  associés aux  $r$  plus grandes valeurs propres solutions de (12).

Plus précisément, l'estimateur de  $\beta$  est donné par :

$$\hat{\beta} = -S_{0p}\hat{\alpha}(\hat{\alpha}'S_{pp}\hat{\alpha}) = -S_{0p}\hat{\alpha} \quad (13)$$

Et

$$\hat{\Omega} = S_{00} - S_{0p}\hat{\alpha}\hat{\alpha}'S_{p0} = S_{00} - \hat{\beta}\hat{\beta}' \quad (14)$$

et la log-vraisemblance maximale s'écrit :

$$\text{Log}L_{\max} = -\frac{kT}{2}\log(2\Pi) - \frac{T}{2}\log[\det(S_{00})] - \frac{T}{2}\sum_{i=1}^r \log(1 - \hat{\lambda}_i) - \frac{kT}{2} \quad (15)$$

Les  $r$  vecteurs de cointégration sont donnés par les  $r$  vecteurs propres les plus significatifs

### Détermination du nombre de vecteurs de cointégration $r$ :

Afin de déterminer le nombre de vecteurs de cointégration  $r$ , Johansen a proposé deux tests : le test de la trace et le test de la valeur propre maximale.

#### Test de la trace :

Il s'agit d'un test du rapport de maximum de vraisemblance consistant à calculer la statistique suivante :

$$TR = -T \sum_{i=q+1}^k \log(1 - \hat{\lambda}_i)$$

L'hypothèse nulle testée est :  $r \leq q$ , c'est-à-dire qu'il existe au plus  $r$  vecteurs de cointégration. Ce test revient à tester le rang de la matrice  $(\Pi_p)$  puisque tester l'existence de  $r$  vecteurs de cointégration revient à tester l'hypothèse nulle :

$$\text{Rg}(\Pi_p) = r.$$

On rejette l'hypothèse nulle de  $r$  relations de cointégration lorsque la statistique TR est supérieur à sa valeur critique. Les valeurs critiques de la statistique TR ont été calculées par Johansen et Juselius (1990) puis par Osterwald-Lenum (1992).

Trois cas peuvent se présenter :

- $Rg(\Pi_p) = 0$ . Cela signifie que  $r = 0$  : il n'existe pas de relation de cointégration. Dans ce cas,  $X_t$  est intégré d'ordre 1 mais non cointégré. il est alors possible d'estimer un modèle VAR sur  $\Delta X_t$
- $Rg(\Pi_p) = r$ , avec  $0 < r < k$  Cela signifie que  $X_t$  est cointégré de rang  $r$  et qu'il existe donc  $r$  relations de cointégration. Un modèle à correction d'erreur peut alors être estimé.
- $Rg(\Pi_p) = k$  : En d'autres termes,  $r = k$  c'est-à-dire que  $\Pi_p$  est de plein rang. Dans ce cas,  $X_t$  est stationnaire et il n'existe pas de relation de cointégration. Un modèle VAR peut être estimé directement sur  $X_t$ .

**Remarque :** Dans la pratique, le logiciel EVIEWS donne directement la valeur de TR (calculée) et sa valeur tabulée.

### Test de la valeur propre maximale

La statistique du test est donnée par :

$$VP_{max} = -T \log(1 - \hat{\lambda}_{q+1})$$

On teste l'hypothèse nulle  $H_0 : r = q$  contre l'hypothèse alternative  $H_1 : r = q+1$ . Le test de la trace est plus utilisé que celui de la valeur propre maximale.

### Introduction d'une constante et/ou d'un trend et valeurs critiques pour le test de la trace :

Il est fondamental de remarquer que l'on a supposé jusqu'ici :

- L'absence de constante dans le modèle à correction d'erreur (équation (2));
- L'absence de constante et de tendance déterministe dans les relations de cointégration.

Or, les valeurs critiques des tests changent dès lors que l'on introduit une constante et/ou un trend. On a en fait quatre possibilités :

- **Cas1** : absence de constante dans le modèle à correction d'erreur (les séries n'ont pas de trend linéaire) et dans les relations de cointégration ;
- **Cas2** : constante dans le modèle à correction d'erreur (les séries sont caractérisées par des trends linéaires) et dans les relations de cointégration ;
- **Cas3** : constante dans les relations de cointégration mais pas dans le modèle à correction d'erreur (les séries n'ont pas de trend linéaire) ;

- **Cas4** : constante dans le modèle à correction d'erreur (les séries sont caractérisées par un trend linéaire) et constante et tendance dans les relations de cointégration.

### Synthèse de la procédure d'estimation

Nous essayons ici de synthétiser les grandes étapes relatives à l'estimation d'un modèle VECM.

- **Etape 1** : détermination du nombre de retards  $p$  du modèle (en niveau ou en Log) selon les critères AIC et SC.
- **Etape 2** : estimation de la matrice et test de Johansen permettant de connaître le nombre de relations de cointégration.
- **Etape 3** : identification des relations de cointégration, c'est-à-dire des relations de long terme entre les variables.
- **Etape 4** : estimation par la méthode du maximum de vraisemblance du modèle vectoriel à correction d'erreur et validation à l'aide des tests usuels : significativité des coefficients et vérification que les résidus sont des bruits blancs (test de Ljung-Box).

### Conclusion :

La notion de cointégration est étroitement liée à l'estimation du modèle à correction d'erreur au sens d'Engel et Granger et à l'estimation du modèle vectoriel à correction d'erreur au sens de Johansen (cas d'existence des relations de cointégration). La méthode d'Engel et Granger consiste hors de la première étape à montrer qu'il existe une relation de long terme entre les variables, puis hors de la seconde étape, à estimer le modèle à correction d'erreur exprimé par les variables cointégrées.

La méthode du maximum de vraisemblance de Johansen est une approche multivariée, elle permet d'identifier plusieurs vecteurs de cointégration et de les estimer.

On peut résumer la procédure d'estimation en quatre étapes :

La première étape est de rechercher le nombre de retard optimal à inclure dans le modèle VAR en niveau ou en logarithme, selon les critères de Akaike et Schwarz ; la deuxième étape est la détermination du nombre de relations de cointégration

grâce à l'application du test de Johansen (test de la trace), hors de la troisième on identifie les relations de cointégration, la quatrième étape donne lieu à estimer par la méthode du maximum de vraisemblance le modèle vectoriel à correction d'erreur VECM.

# Chapitre 5

## Application de l'approche de cointégration

### Introduction

Dans le chapitre 2 nous avons vu que les séries :

**INV**(Investissement), **VD**(Volume de découvertes), **SIS3D** (Sismique trois dimensions). sont toutes non stationnaires de même type (DS) et de même ordre (intégrées d'ordre 1), ce qui implique qu'il y a la possibilité d'avoir une relation de cointégration entre ces trois variables.

Après avoir fait plusieurs combinaisons et on peut vous poser la question comment vous avez tester l'existence de la relation de cointégration pour aboutir à ces combinaisons suivantes :

- L'investissement (**INV**) et volume de découvertes (**VD**)
- volume de découvertes (**VD**) et sismique trois dimension(**SIS3D**).
- L'investissement (**INV**) et sismique trois dimension(**SIS3D**).

Dans cette partie, nous allons présenter l'application de la méthode d'estimation de Engel et Granger en deux étapes sur les séries : **INV**, **VD** et **SIS3D**, prises deux à deux : C'est l'approche de cointégration bivariée. De plus, nous appliquons la méthode du maximum de vraisemblance de Johansen sur le vecteur (**INV,VD,SIS3D**) et nous estimons le modèle vectoriel à correction d'erreur **VECM** : C'est l'approche de cointégration multivariée.

## 5.1 Cointégration bivariée

### 5.1.1 Cointégration entre l'investissement ( $INV_t$ ) et le volume de découvertes ( $VD_t$ )

Considérons les séries annuelles  $INV_t$  et  $VD_t$  sur la période allant de l'année 1980 à 2022. Ces deux séries, comme on l'a déjà vu dans le chapitre 3, exhibent une tendance commune à la hausse. En effet, les deux séries semblent non stationnaire en moyenne et ayant une évolution presque similaire sur tout l'échantillon.

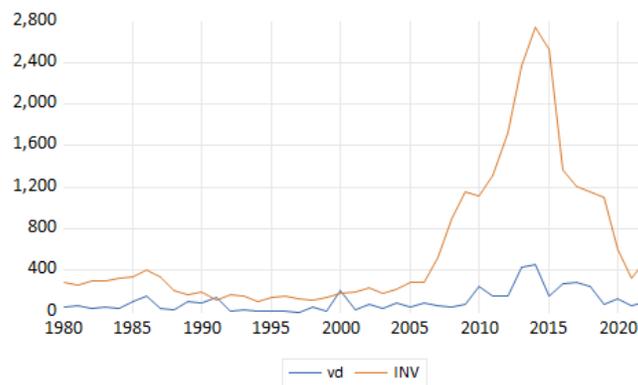


FIGURE 5.1 – Evolution de  $INV_t$  et  $VD_t$

Afin de trouver la relation de cointégration entre ces deux séries, nous avons appliqué la procédure en deux étapes de Engle et Granger.

#### Détermination de l'ordre d'intégration des séries

D'après l'application de Box-Jenkins sur les deux séries  $INV_t$  et  $VD_t$ , on a trouvé que les deux séries avaient le même ordre d'intégration qui est d'égal à 1. Donc il y a un risque de cointégration. On va donc chercher à estimer un MCE.

#### Estimation de la relation statistique entre INV et VD

Cette étape consiste à estimer la relation de long terme. Autrement dit, estimer l'équation d'équilibre. On l'appelle aussi relation cointégrante de long terme ou

encore équation structurelle de long terme. Cette équation s'écrit comme suit :

$$INV_t = \beta + \alpha VD_t + Z_t$$

L'estimation de cette relation, par moindres carrés ordinaire (MCO), est donnée par la table suivante :

Dependent Variable: INV Method: Least Squares Date: 06/06/23 Time: 15:02 Sample: 1980 2022 Included observations: 43				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
VD	5.512923	0.449374	12.26802	0.0000
R-squared	0.599959	Mean dependent var		616.7674
Adjusted R-squared	0.599959	S.D. dependent var		683.5377
S.E. of regression	432.3292	Akaike info criterion		14.99923
Sum squared resid	7850160.	Schwarz criterion		15.04019
Log likelihood	-321.4835	Hannan-Quinn criter.		15.01434
Durbin-Watson stat	1.785192			

FIGURE 5.2 – Estimation de la relation  $INV_t$  et  $VD_t$

Nous remarquons que la constante  $\beta$  était non significative. Donc, elle est omise du modèle et que le coefficient alpha égal à 5.512923 est significativement différent de zéro au seuil 5%. En effet, la statistique de student correspondante égale à 12.26802 est en valeur absolue supérieure à 1.96. La relation de long terme est écrite sous la forme suivante :

$$\hat{INV}_t = 5.512923 * \hat{VD}_t + \hat{Z}_t$$

Avec  $\hat{Z}_t$  les résidus estimés de la relation de long terme. Notons que la condition nécessaire pour que la relation de long terme soit une relation de cointégration est que la série de ces résidus soit stationnaires (dans le cas contraire on dit qu'il s'agit d'une régression fallacieuse).

#### Test de stationnarité sur les résidus estimés de long terme :

Comme nous venons de le mentionner, les résidus estimés doivent être stationnaire. Théoriquement, stationnaires sans tendance et parfois sans constante pour conclure qu'il existe une relation de cointégration entre les deux séries. Pour cela,

Augmented Dickey-Fuller test statistic	-6.184127	0.0000
Test critical values:		
1% level	-4.192337	
5% level	-3.520787	
10% level	-3.191277	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(Z)  
 Method: Least Squares  
 Date: 07/04/23 Time: 10:00  
 Sample (adjusted): 1981 2022  
 Included observations: 42 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.992774	0.160536	-6.184127	0.0000
C	-168.5284	136.7469	-1.232411	0.2252
@TREND("1980")	10.25537	5.683016	1.804564	0.0789

R-squared	0.495119	Mean dependent var	-0.525751
Adjusted R-squared	0.469227	S.D. dependent var	584.6415
S.E. of regression	425.9356	Akaike info criterion	15.01520
Sum squared resid	7075425.	Schwarz criterion	15.13932
Log likelihood	-312.3193	Hannan-Quinn criter.	15.06070
F-statistic	19.12293	Durbin-Watson stat	1.992908
Prob(F-statistic)	0.000002		

FIGURE 5.3 – Test ADF des résidus  $\hat{Z}_t$ 

nous avons appliqué le test d'ADF sur les résidus estimés. Les résultats sont affichés dans la table suivante :

Nous constatons que les résidus estimés sont stationnaire sans tendance et sans constante. Alors nous acceptons l'hypothèse que  $INV_t$  est  $VD_t$  sont cointégrées.

### Corrélogramme des résidus

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.094	0.094	0.4066	0.524	
2	-0.043	-0.053	0.4950	0.781	
3	0.217	0.229	2.7778	0.427	
4	0.327	0.297	8.0853	0.089	
5	-0.108	-0.152	8.6788	0.123	
6	0.103	0.128	9.2325	0.161	
7	0.126	-0.038	10.084	0.184	
8	-0.104	-0.183	10.686	0.220	
9	-0.151	-0.095	11.975	0.215	
10	0.073	-0.029	12.289	0.266	
11	-0.014	0.014	12.301	0.341	
12	-0.083	0.069	12.729	0.389	
13	-0.017	0.025	12.746	0.468	
14	0.070	0.078	13.070	0.521	
15	-0.152	-0.132	14.670	0.475	
16	0.045	0.083	14.816	0.538	
17	-0.027	-0.145	14.872	0.605	
18	-0.023	-0.016	14.914	0.668	
19	-0.083	-0.008	15.469	0.692	
20	-0.001	-0.083	15.469	0.749	

FIGURE 5.4 – Corrélogramme des résidus  $\hat{Z}_t$

D'après l'analyse du corrélogramme des résidus estimés,  $\hat{Z}_t$  est un bruit blanc avec toutes les probabilités  $> 0.05$ . Donc stationnaire.

Par conséquent, les deux séries sont cointégrées et puisque elles sont chacune suit I(1), elles sont de même ordre d'intégration I(1), elles sont alors cointégrées CI(1,1).

### Test de normalité des résidus estimés

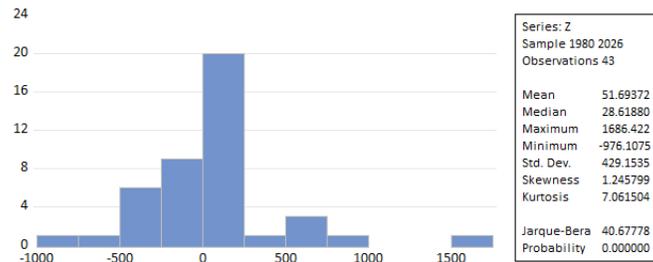


FIGURE 5.5 – Histogramme des résidus  $\hat{Z}_t$

$JB = 40.67778 > \chi_2^2$  La distribution résiduelle ne suivent pas une distribution normale.

### Test d'effet ARCH)

Heteroskedasticity Test: ARCH				
F-statistic	0.022392	Prob. F(1,40)	0.8818	
Obs*R-squared	0.023499	Prob. Chi-Square(1)	0.8782	
Test Equation:				
Dependent Variable: WGT_RESID^2				
Method: Least Squares				
Date: 07/04/23 Time: 11:22				
Sample (adjusted): 1981 2022				
Included observations: 42 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.047897	0.425457	2.462993	0.0182
WGT_RESID^2(-1)	-0.023657	0.158090	-0.149641	0.8818
R-squared	0.000559	Mean dependent var	1.023661	
Adjusted R-squared	-0.024427	S.D. dependent var	2.519105	
S.E. of regression	2.549686	Akaike info criterion	4.756265	
Sum squared resid	260.0360	Schwarz criterion	4.839012	
Log likelihood	-97.88157	Hannan-Quinn criter.	4.786595	
F-statistic	0.022392	Durbin-Watson stat	2.003407	
Prob(F-statistic)	0.881800			

FIGURE 5.6 – Test d'effet ARCH

A partir de la figure nous remarquons que la probabilité de  $RESID^2$  est supérieure à 0.05, donc nous acceptons l'hypothèse nulle d'hétéroscédasticité ; il n'existe pas d'effet ARCH.

Après avoir estimé la relation de long terme entre les deux séries, il convient à présent d'estimer le Modèle à Correction d'Erreur (MCE).

### Estimation du Modèle à Correction d'Erreurs

Dans cette section nous avons modélisé les deux séries différenciées :  $DINV_t$  en fonction de  $DVD_t$ , des résidus retardés d'une période ( $Z_{t-1}$ )  $DVD_t$  retardé d'une période ( $DINV_{t-1}$ ), c'est-à-dire que nous estimons le modèle suivant :

$$DINV_t = \sigma \hat{Z}_{t-1} + \alpha DVD_t + \gamma DINV_{t-1} + \varepsilon_t$$

A partir du résultat de cette estimation, nous aurons la table suivante :

Dependent Variable: DINV  
Method: Least Squares  
Date: 06/06/23 Time: 15:14  
Sample (adjusted): 1982 2022  
Included observations: 41 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DVD	1.122332	0.389995	2.877809	0.0065
Z(-1)	-0.384809	0.089824	-4.284020	0.0001
DINV(-1)	0.523316	0.123272	4.245221	0.0001
R-squared	0.451413	Mean dependent var		6.170732
Adjusted R-squared	0.422540	S.D. dependent var		267.3696
S.E. of regression	203.1763	Akaike info criterion		13.53638
Sum squared resid	1568664.	Schwarz criterion		13.66176
Log likelihood	-274.4958	Hannan-Quinn criter.		13.58204
Durbin-Watson stat	1.654451			

FIGURE 5.7 – Estimation du Modèle à Correction d'Erreurs

D'après le tableau, on constate que le coefficient associé à la force de rappel est négatif (-0.384809) est significativement différent de zéro au seuil statistique de 5% (son t de Student est supérieur à 1.96 en valeur absolue) il existe donc bien un mécanisme à correction d'erreurs : à long terme les déséquilibres entre  $INV_t$  et  $VD_t$  se compensent de sorte que les deux séries ont une évolution similaire.

Nous retenons alors le modèle à correction d'erreurs qui s'écrit sous la forme

suivante :

$$DINV_t = -0.384809\hat{Z}_{t-1} + 1.122332DVD_t + 0.523316DINV_{t-1} + \varepsilon_t$$

Avec :

$$\hat{Z}_t = INV_t - 5.512923 * VD_t$$

Car :

$$\hat{Z}_{t-1} = INV_{t-1} - 5.512923 * VD_{t-1}$$

Et

$$DINV_t = INV_t - INV_{t-1}$$

$$DVD_t = VD_t - VD_{t-1}$$

finalement :

$$INV_t = 1.138507INV_{t-1} + 1.122332VD_t + 3.243754VD_{t-1} - 0.523316INV_{t-2} + \varepsilon_t$$

Avec les  $\varepsilon_t$  les résidus du Modèle à Correction d'Erreurs.

Le graphe qui suit montre que les résidus sont des bruits blancs :

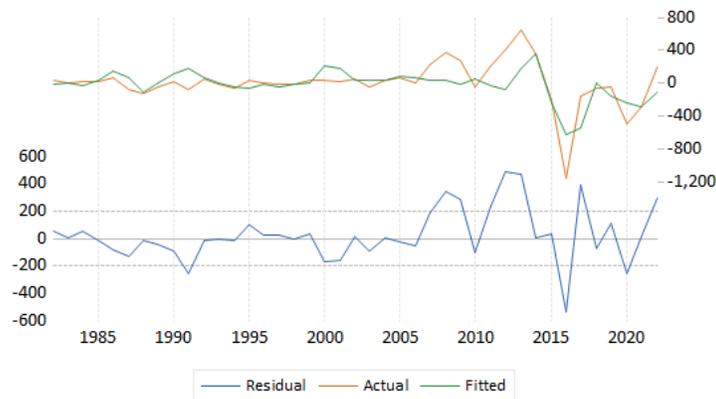
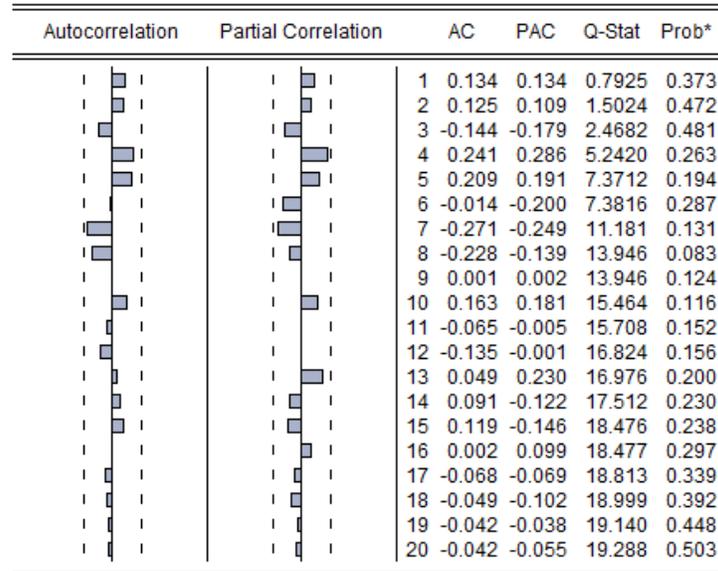


FIGURE 5.8 – Graphe des séries réelles estimées et résiduelles

Nous allons le confirmer par le corrélogramme : Tous les pics sont à l'intérieur de l'intervalle de confiance, ce qui indique que les résidus sont des bruits blancs.

FIGURE 5.9 – Corrélogramme des résidus  $\varepsilon_t$ 

### 5.1.2 Test de causalité au sens de Granger

Pour mener le test de causalité au sens de Granger, on a besoin de déterminer le nombre de retards  $p$  du modèle VAR( $p$ ) avec les 2 variables stationnaires. Pour cela, on utilise les critères AIC et SC.

Le retard $p$	AIC	SCH
P=1	25.67006	25.92083
P=2	25.46597	25.88819
P=3	25.42334	26.02052

TABLE 5.1 – Choix de retard P

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
DVD does not Granger Cause DINV	40	5.67309	0.0073
DINV does not Granger Cause DVD		3.27299	0.0498

FIGURE 5.10 – Test de causalité au sens de Granger

Dans notre cas la probabilité  $< 0.05$ , l'hypothèse nulle est rejetée. On constate que DVD influence DINV, et DINV influence DVD.

### Prévisions

Le modèle à correction d'erreurs de la série  $INV_t$  s'écrit sous la forme suivante :

$$INV_t = 1.138507INV_{t-1} + 1.122332VD_t + 3.243754VD_{t-1} - 0.523316INV_{t-2} + \varepsilon_t$$

D'où les prévisions de la série  $INV_t$  à un horizon  $l \geq 1$  sont données comme suit :

Pour  $l = 1$

$$\hat{INV}_1 = 1.138507INV_t + 1.122332VD_{t+1} + 3.243754VD_t - 0.523316INV_{t-1}$$

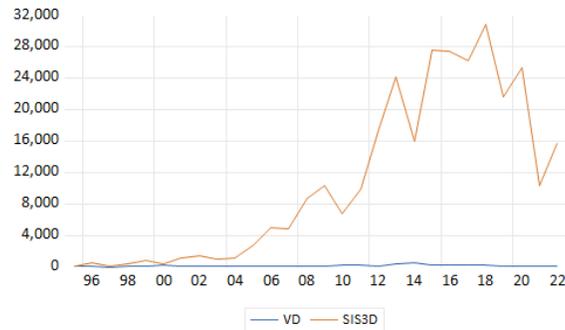
prevision	INV	VD
2023	655.9936	99.4422
2024	729.5048	111.1843
2025	771.2830	122.5778
2026	792.5998	125.0350

TABLE 5.2 – Prevision de cointégration INV et VD

### 5.1.3 Cointégration entre le volume de découvertes ( $VD_t$ ) et sismique trois dimension $SIS3D_t$

Considérons les séries annuelles  $VD_t$  et  $SIS3D_t$  sur la période allant 1995 à 2022. Ces deux séries, exhibent une tendance commune à la hausse.

Afin de trouver la relation de cointégration entre ces deux séries, nous avons appliqué la procédure en deux étapes de Engle et Granger.

FIGURE 5.11 – Evolution de  $VD_t$  et  $SIS3D_t$ 

### Détermination de l'ordre d'intégration des séries

D'après l'application de Box-Jenkins sur les deux séries  $VD_t$  et  $SIS3D_t$ , on a trouvé que les deux séries avaient le même ordre d'intégration qui est d'égal à 1.

### Estimation de la relation statique entre $VD_t$ et $SIS3D_t$

Cette étape consiste à estimer la relation de long terme suivante :

$$VD_t = \beta + \alpha SIS3D_t + Z_t$$

L'estimation de cette relation est donnée par la table suivante :

Dependent Variable: VD  
 Method: Least Squares  
 Date: 06/12/23 Time: 00:51  
 Sample: 1995 2022  
 Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SIS3D	0.009583	0.001293	7.410842	0.0000

R-squared	0.307422	Mean dependent var	125.6429
Adjusted R-squared	0.307422	S.D. dependent var	121.9193
S.E. of regression	101.4628	Akaike info criterion	12.11232
Sum squared resid	277957.0	Schwarz criterion	12.15990
Log likelihood	-168.5725	Hannan-Quinn criter.	12.12687
Durbin-Watson stat	1.530410		

FIGURE 5.12 – Estimation de la relation entre  $VD_t$  et  $SIS3D_t$

Nous remarquons que le coefficient est significatif différent de zéro au seuil 5% car sa statistique de Student est en valeur absolue supérieure à 1.96. Donc la relation de long terme est écrite sous la forme suivante :

$$\hat{V}D_t = 0.009583 * SIS3D_t + \hat{Z}_t$$

Avec  $\hat{Z}_t$  les résidus estimés de la relation de long terme. Notons que la condition nécessaire pour que la relation de long terme soit une relation de cointégration et que la série de ces résidus soit stationnaires (dans le cas contraire on dit qu'il s'agit d'une régression fallacieuse).

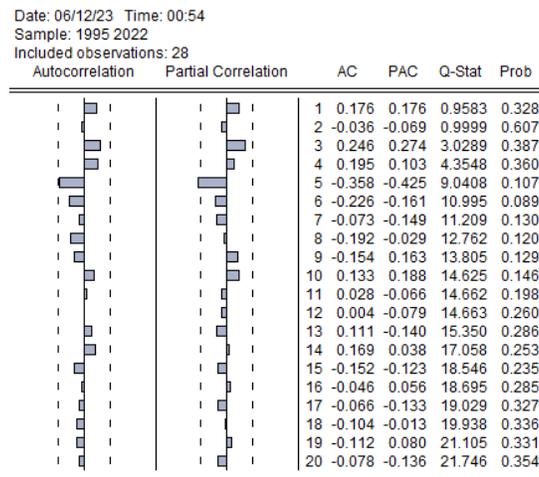


FIGURE 5.13 – Corrélogramme des résidus  $\hat{Z}_t$

D'après l'analyse de corrélogramme des résidus estimés,  $\hat{Z}_t$  est un bruit blanc avec toutes les probabilités  $> 0.05$ . Donc stationnaire

Par conséquent, les deux séries sont cointégrées et puisque elles sont chacune suit I(1), elles sont alors  $\sim CI(1,1)$ .

Après avoir estimé la relation de long terme entre les deux séries, il convient à présent d'estimer le Modèle à Correction d'Erreur (MCE).

### Estimation du Modèle à Correction d'Erreurs

Dans cette section nous avons modélisé les deux séries différenciées :  $DVD_t$  en fonction de  $DSIS3D_t$ , des résidus retardés d'une période ( $Z_{t-1}$ )  $DSIS3D_{t-1}$

retardé d'une période  $DSIS3D_{t-2}$ , c'est-à-dire que nous estimons le modèle suivant :

$$DVD_t = \sigma \hat{Z}_{t-1} + \alpha DSIS3D_{t-1} + \beta DSIS3D_{t-2} + \varepsilon_t$$

Le résultat de cette estimation est donné par la table suivante :

Dependent Variable: DVD				
Method: Least Squares				
Date: 06/12/23 Time: 01:00				
Sample (adjusted): 1998 2022				
Included observations: 25 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.748660	0.196218	-3.815456	0.0009
DSIS3D(-1)	0.008569	0.003815	2.245956	0.0351
DSIS3D(-2)	0.011824	0.005059	2.337221	0.0289
R-squared	0.465366	Mean dependent var	3.720000	
Adjusted R-squared	0.416762	S.D. dependent var	121.6910	
S.E. of regression	92.93544	Akaike info criterion	12.01385	
Sum squared resid	190013.9	Schwarz criterion	12.16012	
Log likelihood	-147.1732	Hannan-Quinn criter.	12.05442	
Durbin-Watson stat	1.914416			

FIGURE 5.14 – L'estimation du Modèle à Correction d'Erreurs

D'après le tableau, on constate que le coefficient associé à la force de rappel est négatif (-0.748660) est significativement différent de zéro au seuil statistique de 5% (son t de Student est supérieur à 1.96 en valeur absolue) il existe donc bien un mécanisme à correction d'erreurs : à long terme les déséquilibres entre  $VD_t$  et  $SIS3D_t$  se compensent de sorte que les deux séries ont une évolution similaire. Nous retenons alors le modèle à correction d'erreurs qui s'écrit sous la forme suivante :

$$DVD_t = -0.748660 \hat{Z}_{t-1} + 0.008569 DSIS3D_{t-1} + DSIS3D_{t-2} + \varepsilon_t$$

Avec :

$$\hat{Z}_t = VD_t - 0.009583 * SIS3D_t$$

Car :

$$\hat{Z}_{t-1} = VD_{t-1} - 0.009583 * SIS3D_{t-1}$$

Et

$$DVD_t = VD_t - VD_{t-1}$$

$$DSIS3D_t = SIS3D_t - SIS3D_{t-1}$$

finalement :

$$VD_t = 0.25134VD_{t-1} + 0.015743SIS3D_{t-1} + 0.003255SIS3D_{t-2} - 0.0011824SIS3D_{t-3} + \varepsilon_t$$

Avec les  $\varepsilon_t$  les résidus du Modèle à Correction d'Erreurs. Le graphe qui suit montre que les résidus sont des bruits blancs.



FIGURE 5.15 – Graphe des séries réelles estimées et résiduelles

Nous allons le confirmer par le corrélogramme :

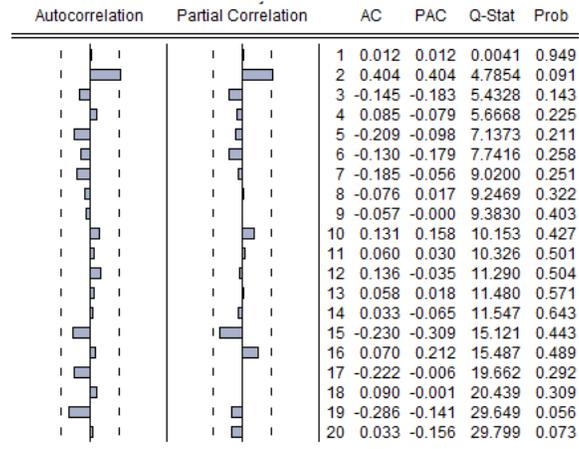


FIGURE 5.16 – Corrélogramme des résidus  $\varepsilon_t$

Les pluparts des pics sont à l'intérieur de l'intervalle de confiance, ce qui indique que les résidus sont des bruits blancs.

### Prévisions

Le modèle à correction d'erreurs de la série  $VD_t$  s'écrit sous la forme suivante :

$$VD_t = 0.25134VD_{t-1} + 0.015743SIS3D_{t-1} + 0.003255SIS3D_{t-2} - 0.0011824SIS3D_{t-3} + \varepsilon_t$$

D'où les prévisions de la série  $VD_t$  à un horizon  $h \geq 1$  sont données comme suit :

Pour  $h = 1$  :

$$\hat{VD}(1) = 0.25134VD_t + 0.015743SIS3D_t + 0.003255SIS3D_{t-1} - 0.0011824SIS3D_{t-2}$$

prevision 1	VD	SIS3D
2023	171.3214	10337.26
2024	112.5215	12059.82
2025	140.033	12981.64
2026	148.9406	12059.50

TABLE 5.3 – Prevision de cointégration VD et SIS3D

## 5.2 Cointégration multivariée

### Identification

On considère les séries déjà étudiées dans le premier chapitre. On constate graphiquement que ces séries sont caractérisées par un trend général à la hausse.

Par ailleurs, des tests de racine unitaire ont été effectués et ont montré que les trois séries sont intégrées d'ordre 1.

### Test de Johansen

#### Détermination du nombre de retard optimal

Pour déterminer le nombre de retards optimal qui correspond à la spécification du modèle VAR nous allons recourir aux critères d'Akaike et/ou Schwarz pour des décalages  $p$  ( $p$  étant le retard maximum) allant de 1 à 3. En d'autres termes, il s'agit d'estimer 3 modèles différents et choisir celui dont les critères AIC et/ou SC sont les plus petits.

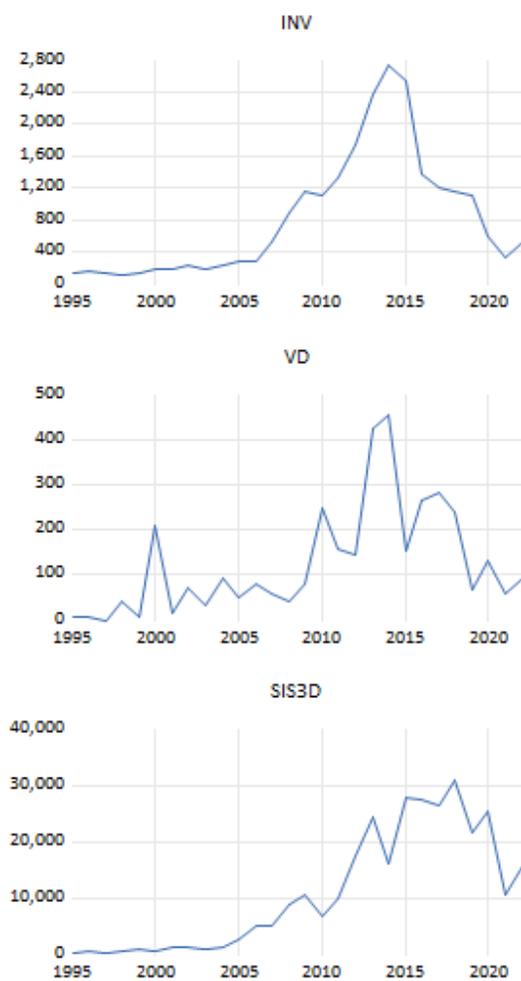


FIGURE 5.17 – Graphes des trois séries non stationnaires

Le retard p	AIC	SCH
p=1	14.45186	14.64383
p=2	14.33033	14.66905
p=3	14.30587	14.79342

TABLE 5.4 – choix du nombre de retard p

Choix de retard qui minimise les critères d' Akaike et Schwarz  
Après la comparaison entre les deux critères on va choisi p=1.

### Estimation du modèle VAR(1)

On estime le modèle VAR(1) on aura la table suivant :

	INV	SIS3D	VD
INV(-1)	0.887742 (0.14346) [ 6.18820]	5.286734 (2.11457) [ 2.50015]	0.112362 (0.04004) [ 2.80634]
SIS3D(-1)	-0.013875 (0.00853) [-1.62601]	0.598919 (0.12578) [ 4.76174]	0.001256 (0.00238) [ 0.52753]
VD(-1)	1.236883 (0.79466) [ 1.55649]	-3.218034 (11.7133) [-0.27473]	-0.088315 (0.22179) [-0.39819]
C	94.95613 (91.1856) [ 1.04135]	800.8128 (1344.08) [ 0.59581]	35.09782 (25.4498) [ 1.37910]

R-squared	0.859494	0.831995	0.552340
Adj. R-squared	0.841167	0.810081	0.493950
Sum sq. resids	2221117.	4.83E+08	173016.9
S.E. equation	310.7576	4580.584	86.73219
F-statistic	46.89815	37.96684	9.459433
Log likelihood	-191.1001	-263.7454	-156.6430
Akaike AIC	14.45186	19.83299	11.89948
Schwarz SC	14.64383	20.02497	12.09146
Mean dependent	842.0741	11028.32	130.0741
S.D. dependent	779.7440	10510.82	121.9224
Determinant resid covariance (dof adj.)		1.15E+16	
Determinant resid covariance		7.08E+15	
Log likelihood		-607.6332	
Akaike information criterion		45.89875	
Schwarz criterion		46.47468	
Number of coefficients		12	

FIGURE 5.18 – Estimation du modèle VAR(1)

D'après ce tableau le modèle VAR(1) s'écrit sous forme de 3 équations différentes :

$$INV_t = 0.887742INV_{t-1} - 0.013875SIS3D_{t-1} + 1.236883VD_{t-1} + 94.95613 + \varepsilon_t$$

$$SIS3D_t = 5.286734INV_{t-1} - 0.598919SIS3D_{t-1} + 3.218034VD_{t-1} + 800.8128 + \varepsilon_t$$

$$VD_t = 0.112362INV_{t-1} - 0.002156SIS3D_{t-1} - 0.088315VD_{t-1} + 35.09782 + \varepsilon_t$$

### Test de la trace

Pour effectuer le test de la trace, la spécification à retenir dépend de :

- L'absence ou la présence de constante dans le modèle à correction d'erreur.
- L'absence ou la présence de constante et tendance dans la relation de cointégration.

Nous effectuons ici le test de la trace en supposant l'absence de tendance dans la relation de cointégration et la présence d'une constante dans le modèle à correction d'erreur. Ce choix peut être justifié économiquement en supposant que les relations d'équilibre de long terme entre ces séries (INV, VD, SIS3D) ne comportent pas de trend. La présence d'une constante dans le modèle à correction d'erreur vient du fait que les trois séries en logarithme semblent être caractérisées par un trend linéaire à la baisse.

On obtientra le tableau ci-dessous :

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.716741	47.95605	29.79707	0.0002
At most 1 *	0.397451	16.42125	15.49471	0.0362
At most 2	0.139519	3.756590	3.841465	0.0526

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level  
 \* denotes rejection of the hypothesis at the 0.05 level  
 \*\*Mackinnon-Haug-Michelis (1999) p-values

FIGURE 5.19 – Test de la trace

D'après ce tableau, on remarque que  $(47.95605 > 29.79707)$  et  $(16.42125 > 15.49471)$  au seuil statistique de 5%. On rejette l'hypothèse nulle d'absence de cointégration. En revanche, On accepte l'hypothèse nulle selon laquelle il existe une relation de cointégration entre les trois variables  $(3.75659 < 3.841465)$ .

### Test de la valeur propre maximale

Dans la même table du test de la trace on trouvera celle du test de la valeur propre maximale suivante :

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.716741	31.53480	21.13162	0.0012
At most 1	0.397451	12.66466	14.26460	0.0881
At most 2	0.139519	3.756590	3.841465	0.0526

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level  
 \* denotes rejection of the hypothesis at the 0.05 level  
 \*\*MacKinnon-Haug-Michelis (1999) p-values

FIGURE 5.20 – Test de la valeur maximale

On a ( $31.5348 > 21.13162$ ) on accepte l'hypothèse  $H_0$ , il y a une relation de cointégration entre les trois variables aux seuils de 5%.

### Estimation de la relation de long terme

L'estimation de la relation de long terme se fait à l'aide de la table suivante :

1 Cointegrating Equation(s):			Log likelihood	-567.9035
Normalized cointegrating coefficients (standard error in parentheses)				
DINV	DVD	DSIS3D		
1.000000	-12.77034 (1.66968)	0.075534 (0.04453)		

FIGURE 5.21 – Estimation de la relation de long terme

Les valeurs entre parenthèses sont des écarts types estimés associés aux coefficients estimés. On a ici normalisé le coefficient de  $INV_t$ , c'est à dire que l'on a choisi  $INV_t$  comme variable endogène,  $VD_t$  et  $SIS3D_t$  étant les deux variables exogènes. En d'autres termes, la relation de long terme s'écrit :

$$DINV_t = 12.77034DVD_t - 0.075534DSIS3D_t + Z_t$$

Où  $Z_t$  est le terme résiduel.

Il est alors possible d'estimer un Modèle à Correction d'Erreur Vectoriel (VECM).

Dans la mesure où nous avons trois variables, le VECM comportera trois équations.

Les résultats de l'estimation du VECM sont reportés dans les tableaux qui suivent :

Error Correction:	D(INV)	D(VD)	D(SIS3D)
CointEq1	-0.012919 (0.01008) [-1.28136]	-0.010043 (0.00357) [-2.81564]	0.414989 (0.14020) [ 2.96005]
D(INV(-1))	0.521674 (0.18928) [ 2.75604]	0.101435 (0.06696) [ 1.51476]	2.377836 (2.63195) [ 0.90345]
D(VD(-1))	1.678965 (0.61377) [ 2.73548]	0.048866 (0.21714) [ 0.22505]	-29.34160 (8.53436) [-3.43806]
D(SIS3D(-1))	-0.010820 (0.01109) [-0.97524]	0.003771 (0.00392) [ 0.96078]	-0.405415 (0.15426) [-2.62805]
C	11.28207 (54.7577) [ 0.20604]	0.989781 (19.3719) [ 0.05109]	789.7931 (761.393) [ 1.03730]
R-squared	0.421315	0.427270	0.562484
Adj. R-squared	0.311089	0.318178	0.479148
Sum sq. resids	1626889.	203616.8	3.15E+08
S.E. equation	278.3359	98.46847	3870.193
F-statistic	3.822289	3.916617	6.749570
Log likelihood	-180.4655	-153.4491	-248.9035
Akaike AIC	14.26658	12.18839	19.53104
Schwarz SC	14.50852	12.43033	19.77298
Mean dependent	14.15385	3.307692	586.5192
S.D. dependent	335.3418	119.2509	5362.601
Determinant resid covariance (dof adj.)		7.01E+15	
Determinant resid covariance		3.69E+15	
Log likelihood		-576.6679	
Akaike information criterion		45.74369	
Schwarz criterion		46.61468	

FIGURE 5.22 – Estimation du VECM

Il est alors possible d'estimer un modèle à correction d'erreur vectoriel (VECM). Dans la mesure où nous avons trois variables, le VECM comportera trois Équations. Les résultats de L'estimation du VECM sont reportés dans le tableau ci-dessus. Dans le tableau précédent, CointEq1 désigne les résidus, retardés d'une période, de la relation de cointégration préalablement trouvée (c'est-à-dire  $\hat{Z}_{t-1}$ )

### Estimation

$$DINV_t = -0.012919\hat{Z}_{t-1} + 0.521674DINV_{t-1} + 1.678965DVD_{t-1} - 0.01082DSIS3D_{t-1} + 11.28207 + \varepsilon_{1t}$$

$$n = 28 \quad R^2 = 0.4213$$

$$DVD_t = -0.010043\hat{Z}_{t-1} + 0.101435DINV_{t-1} + 0.048866DVD_{t-1} + 0.003771DSIS3D_{t-1} + 0.989781 + \varepsilon_{2t}$$

$$n = 28 \quad R^2 = 0.4272$$

$$DSIS3D_t = 0.414989\hat{Z}_{t-1} + 2.377836DINV_{t-1} - 29.3416DVD_{t-1} - 0.405415DSIS3D_{t-1} + 789.7931 + \varepsilon_{3t}$$

$$n = 28 \quad R^2 = 0.5624$$

Le terme à correction d'erreur (-0.012919) est négatif et non significativement différent de 0 (t de Student en valeur absolue est  $< 1.96$  à 5%) dans la relation relative à la variation de l'investissement (DINV).

Dans l'équation de volume de découverte (DVD), le terme à correction d'erreur (- 0.010043) est négatif et significativement différent de 0 (t de Student en valeur absolue est  $> 1.96$  à 5%).

En revanche on remarque que le terme à correction d'erreur (0.414989) du SIS3D est positif et significatif, (t de Student en valeur absolue est  $> 1.96$  à 5%).

### A court terme

1- La variation de D(INV) dépend de la variation de D(INV(-1)) et D(VD(-1)) au seuil statistique de 5% le t de Student en valeur absolue  $> 1.96$ . Ceci reflète l'impact de sa valeur propre décalée d'une période et le volume découvertes.

2- la variation de D(SIS3D) dépend de la variation de D(SIS3D(-1)) et D(VD(-1)) au seuil statistique de 5% le t de Student en valeur absolue  $> 1.96$ . Ceci reflète l'impact de sa valeur propre décalée d'une période et le volume découvertes.

### Analyse de la causalité

On procède donc au test de causalité au sens de Granger à partir de la représentation VAR(1) estimée précédemment. Nous obtenons les résultats suivants :

Pairwise Granger Causality Tests  
Date: 07/04/23 Time: 14:27  
Sample: 1995 2026  
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
DVD does not Granger Cause DINV	26	6.35002	0.0191
DINV does not Granger Cause DVD		0.31280	0.5814
DSIS3D does not Granger Cause DINV	26	0.47494	0.4976
DINV does not Granger Cause DSIS3D		3.37320	0.0792
DSIS3D does not Granger Cause DVD	26	2.55561	0.1236
DVD does not Granger Cause DSIS3D		2.07650	0.1631

FIGURE 5.23 – Test de causalité au sens de Granger

**Test de causalité entre DVD et DINV :** Dans notre cas la probabilité  $< 0.05$ , l'hypothèse nulle est rejetée. On constate que DVD influence DINV. La probabilité  $> 0.05$ , l'hypothèse nulle est acceptée. DINV n'influence pas DVD.

**Test de causalité entre DSIS3D et DINV :** La probabilité  $> 0.05$ , l'hypothèse nulle est acceptée. On constate que DSIS3D n'influence pas DINV, et DINV n'influence pas DSIS3D.

**Test de causalité entre DSIS3D et DVD :** La probabilité  $> 0.05$ , l'hypothèse nulle est acceptée. On constate que DSIS3D n'influence pas DVD, et DVD n'influence pas DSIS3D.

### Prévision

Sachant que nous avons arrêté notre étude en 2022, nous devons faire nos prévisions sur les 4 années qui suivent :

prevision	INV	SIS3D	VD
2023	448.3316	12676.49	104.4983
2024	446.4801	10437.80	92.1101
2025	460.7232	9130.22	90.1872
2026	489.2378	8429.6	90.3166

TABLE 5.5 – Prévision de cointégration multivarié

### 5.3 Comparaison entre la méthodologie de Box-Jenkins et de la méthodologie de cointégration en terme de prévision

La comparaison des résultats prévisionnels obtenus par les deux méthodes à savoir la méthodologie de Box & Jenkins et la modélisation de la cointégration est basée sur l'erreur moyenne quadratique (Root Mean Square) RMSE entre les valeurs prévues et les réalisations des séries de puissance.

Le tableau suivant représente le RMSE calculé à partir de la méthodologie de Box & Jenkins et celle de la modélisation de la cointégration.

	$RMSE_{Box}$	$RMSE_{CB}$	$RMSE_{CM}$
$INV_t$	2216188	7850160	1626889
$VD_t$	316766.6	277957	203616.8

**Remarque :**

$RMSE_{Box}$ ,  $RMSE_{CB}$  et  $RMSE_{CM}$  sont associés à la méthodologie de Box-Jenkins et la méthodologie de la cointégration bivariée et multivariée respectivement.

Pour les séries

- $INV_t : RMSE_{CB} > RMSE_{Box} > RMSE_{CM}$
- $VD_t : RMSE_{Box} > RMSE_{CB} > RMSE_{CM}$

Nous concluons que les prévisions de la méthodologie de cointégration sont plus fiables que celles obtenues par les méthodologies Box & Jenkins et cointégration bivariée.

**Conclusion :** Nous constatons que l'approche de la cointégration multivariée se révèle la meilleure, en terme de prévision.

# Conclusion générale

L'Algérie, en tant que l'un des pays les plus riches en hydrocarbures, détient d'importantes réserves de pétrole et de gaz naturel, qui constituent un pilier essentiel de son activité économique.

La hausse des prix du pétrole au cours des dernières années a incité les entreprises à intensifier leurs investissements dans l'exploration, car de nouvelles découvertes renforcent les réserves et permettent d'augmenter la production pour maximiser les bénéfices.

Comprendre en profondeur les mécanismes et les paramètres régissant le secteur de l'exploration pétrolière est donc d'une importance primordiale pour la SONATRACH.

Afin d'atteindre cet objectif, nous avons analysé l'évolution de quatre indicateurs clés : les investissements financiers pour l'exploration, et le prix du pétrole, les volumes de découvertes et le kilométrage sismique réalisé. Ces données nous ont permis de mieux appréhender le secteur et de prendre des décisions éclairées pour assurer la croissance et la durabilité du secteur de l'exploration pétrolière en Algérie.

Dans la première phase de notre étude, nous nous sommes concentrés sur l'analyse des séries chronologiques univariées représentant l'évolution des indicateurs mentionnés précédemment.

L'analyse d'une série chronologique peut être abordée de différentes manières, mais le nombre limité d'observations dont nous disposons peut restreindre considérablement nos choix. Étant donné que nous disposons d'une série annuelle couvrant 42ans, nous avons opté pour une approche individuelle en utilisant la méthodologie de Box Jenkins.

Grâce à l'application de la méthode de Box Jenkins, nous avons pu déterminer,

parmi les modèles ARIMA, celui qui est le plus adapté pour représenter le comportement de nos séries. Par la suite, nous avons effectué des prévisions pour chaque série.

Cette approche nous a permis de saisir les caractéristiques et les schémas temporels des données, de construire des modèles statistiques appropriés et de générer des prévisions pour évaluer l'évolution future des indicateurs. Ces prévisions nous offrent des informations précieuses pour prendre des décisions éclairées et formuler des recommandations stratégiques.

L'examen du corrélogramme de la série différenciée du prix du pétrole nous a permis de choisir un modèle pour la représenter car tous les termes ne sont pas significativement différents de zéro. Cette série représente un bruit blanc ; nous pensons que le passé de la série n'explique pas vraiment le processus, il y a d'autres paramètres qui influent sur ces séries.

La méthodologie de Box Jenkins permet d'étudier les séries séparément et néglige la corrélation entre elles. Afin d'améliorer les prévisions fournies par cette méthode nous avons fait appel à la théorie de cointégration bivariée (l'investissement, les volumes de découvertes) et (les volumes de découvertes, le kilométrage sismique réalisé) et par la suite à son extension vectorielle (cointégration multivariée) (l'investissement, les volumes de découvertes, le kilométrage sismique). Le modèle choisi est un VAR(1), il nous a permis de calculer les prévisions pour chacune des séries.

Finalement, cette étude nous a permis non seulement de manipuler le logiciel Eviews, mais aussi d'adopter autant de notions et de méthode d'analyse statistique qui est un domaine très vaste. Nous soulignons que les résultats obtenus peuvent être améliorés par une étude sur des séries plus longue (qui disposent plus de données). Par ailleurs il serait intéressant d'analyser les séries pour chaque bassin et de mettre en évidence les corrélations existantes entre elles.

# Bibliographie

- [1] Abbas. Y et Bentarzi. W, Etude des prix spot du gaz naturel, Université des Sciences et de la Technologie Houari Boumedienne, Faculté des mathématiques.
- [2] ADRAR.S et BEN BENNAI.N , Etude prévisionnelle des ventes d'huile par la méthode de BOX et JENKINS(SPA CEVITAL) , Faculté des sciences économiques, commerciales et des sciences de gestion, (2012-2013)
- [3] Bourbonnais Régis et Terraza Michel, Analyse des séries temporelles : Applications à l'économie et à la gestion
- [4] CHARPENTIER Arthur.serie temporelle : théorie et application. soutien de court ;
- [5] Charpentier, Arthur. "Modèles de prévision-Séries temporelles." Paris, France (2012).
- [6] FLORENS J-p, MARIMOUTOU V, PEGUIN-FEISSOLLE, « économétrie ; modélisation et inférence », Armand Colin, 2004, page267.
- [7] Gatt. F et Kedoui. S, Modélisation et Prévision Journalières des Courbes de Charges, Université des Sciences et de la Technologie Houari Boumedienne
- [8] Granger. C.W.J et Newbold. P, Spurious Regressions in Econometrics, Journal of Econometrics, 26, 1045-1066 (1974).
- [9] Johansen. S, Statistical Analysis of Cointegration Vectors, Journal of Economic Dynamic and Control, 12, 231-254 (1988).

- [10] MEDDAHI .S , GAILI.Z , modélisation d'une série financière avec deux approches économétriques et stochastiques, facultés des sciences (2019-2020)
- [11] OUADNI.A, SABER.N, Préviation du taux de change du dinar algérien face aux différentes devises étrangères sur le marché officielles FACULTÉ DES SCIENCES (2019-2020)
- [12] Stock. J, Asymptotic proprieties of Least Squares Estimators of Cointegrating Vectors, *Econometrica*, 55, 1035-1056 (1987).
- [13] ZOGHBI .F et AMIRA.S , Modélisation et prévision de la pluviométrie apport et transfert des eaux (barrage Beni Haroun), FACULTÉ DES SCIENCES EXACTES, (2015-2016).

**SITES INTERNET**

[www.SONATRACH.DZ](http://www.SONATRACH.DZ)

[www.algeriawatch.org/fr/article/eco/hydroc/petrole\\_eserves.htm](http://www.algeriawatch.org/fr/article/eco/hydroc/petrole_eserves.htm).

**Logiciels utilises :**

**EViews 12**