

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

---



Université M'Hamed Bougara Boumerdès  
Faculté des sciences  
Département Mathématique

Mémoire présenté pour l'obtention du diplôme de Master en  
Mathématiques

Spécialité : Modélisation Stochastique et Statistique

---

# L'impact des variations des prix du brut sur l'activité d'exploration

---

Réalisé par :  
Melle BAHLOUL Nour el houda  
Melle DIGUER Lamia

Encadré par :  
Mr TAZROUTI Moussa(UMBB)  
Mme SLIMANI Leila(EXPLO)

*Soutenu le 06 juillet 2022, Devant le jury composé de :*

Mme BEN MANSOUR : Présidente  
Mme CHAMRIKH : Examinatrice

Promotion : 2021/2022

# Remerciements

Tout d'abord, on remercie Allah le tout puissant de m'avoir donné le courage et la patience nécessaire à mener ce travail à son terme.

On tiens à remercier tout particulièrement **Monsieur M.TAZROUTI** , pour l'aide compétente qu'il nous avons apporté, pour sa patience et son encouragement. Son œil critique m'a été très précieux pour structurer le travail et pour améliorer la qualité des différentes sections.

On tiens à remercier également notre promotrice **Mme.Slimani Laila** pour son aide immense, la qualité de son suivie ainsi que pour tous les conseils et les informations qu'il m'a prodigués avec un degré de patience et de professionnalisme sans égal et sans oublier **Mr.ATTAR Mohamed** pour son aide .

Que les membres de jury trouvent, ici **Mme Ben Mansour Mme Chamrikh** l'expression de mes sincères remerciements pour l'honneur qu'ils me font en prenant le temps de lire et d'évaluer ce travail.

On souhaite aussi remercier l'équipe pédagogique et administrative de Département Mathématique pour leurs efforts dans le but de nos offrir une excellente formation. Pour finir, je souhaite remercier toute personne ayant contribué de près ou de loin à la réalisation de ce travail.

Dédicaces :

*Je dédie ce modeste travail aux personnes qui ont toujours été là pour me soutenir et qui ont fait tout ce qui été en leur pouvoir pour que je serai là aujourd'hui :mes chères père et mère.*

*À mes très chers frères et sœurs :  
Chérifa, Abdelhak, Sihem, Yacine, Abdellah, Asma, Assia et Aicha avec qui j'ai  
partagé chaque instant de ma vie. Dieu vous garde pour moi.*

*À mon oncle Rabah, ma tante Aicha, ma cousine Zineb et à toute ma  
famille .*

*À mes très chères amies Imene, Asma, Fahima et Melissa avec eux j'ai  
passé des moments inoubliables.*

*À mon binôme DIGUER LAMIA .*

*À tous ceux qui me sont chers, à vous tous*

*Merci.*

***Bahloul Nour el houda***

Dédicaces :

*Je dédie ce modeste travail à ....*

*À Mes très chers parents*

*À Ma Chère Maman Aucune dédicace très chère maman, ne pourrait exprimer la profondeur des sentiments que j'éprouve pour vous, vos sacrifices innombrables et votre dévouement firent pour moi un encouragement.*

*À mon père qui peut être fier et trouver ici le résultat de longues années de sacrifices. Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.*

*À Mon cher fiancé RABEH Toi qui n'as cessé de croire en moi et de m'encourager et qui as toujours été à mes cotés.*

*À Mes très chers frères et sœurs (Hichem, Khalil, Mohamed Islam, Amir Chaouki, Rania, Djawed et Abd Allah) Qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité. Dieu vous garde pour moi*

*À mon binôme Bahloul Nour el houda*

*À tous ceux qui me sont chers, à vous tous*

*Merci.*

**DIGUER LAMIA**

# Table des matières

Introduction générale	8
Problématique	10
<b>I</b>	<b>12</b>
<b>1 Exploration-Production</b>	<b>13</b>
1.1 Prospection / exploration gazière et pétrolière . . . . .	14
1.1.1 Définition . . . . .	14
1.1.2 La découverte du champ . . . . .	14
1.1.3 L'évaluation du champ . . . . .	18
1.2 LA PRODUCTION DE PÉTROLE ET DE GAZ . . . . .	18
1.2.1 Le développement du champ . . . . .	18
1.2.2 La production du champ . . . . .	19
1.2.3 L'abandon du champ . . . . .	19
1.3 les variations du prix du pétrole sur l'activité d'exploration . . . . .	19
1.4 L'importance de l'industrie pétrolière en Algérie . . . . .	20
1.5 L'évolution des investissements en exploration compte tenu de l'évolution des prix du pétrole . . . . .	21
1.6 L'évolution des découvertes pétrolières à la lumière de l'évolution des prix du pétrole . . . . .	22
1.7 Organigramme de la macrostructure de Sonatrach . . . . .	24
1.8 Organigramme de la division exploration . . . . .	25
<b>2 Processus stochastique et séries chronologiques :</b>	<b>26</b>
2.1 Définitions et concept de base . . . . .	26
2.1.1 Processus aléatoire . . . . .	26
2.1.2 Processus aléatoire du second ordre . . . . .	26
2.1.3 Les processus stationnaires . . . . .	27
2.2 Fonction d'autocovariance . . . . .	28
2.3 Fonction d'autocorrelation . . . . .	28
2.4 Fonction d'autocorrelation partielle . . . . .	29

## Table des matières

---

2.5	SERIES CHRONOLOGIQUES . . . . .	30
2.5.1	Définition . . . . .	30
2.5.2	Opérateurs linéaires . . . . .	31
2.5.3	Analyse des séries chronologiques . . . . .	32
2.5.4	Modélisation des séries chronologique . . . . .	33
2.5.5	Séries non stationnaires : . . . . .	43
2.5.6	Extension des modèles ARMA : . . . . .	49
<b>3</b>	<b>La méthodologie de BOX et JENKINS</b>	<b>51</b>
3.1	Introduction . . . . .	51
3.2	Test des composantes saisonnières et tendancielle . . . . .	51
3.2.1	Test de Fisher . . . . .	51
3.2.2	Test de Dickey-Fuller . . . . .	52
3.3	Identification . . . . .	53
3.4	Estimation . . . . .	53
3.5	Tests de validation . . . . .	54
3.6	Test sur les paramètres . . . . .	54
3.6.1	Tests sur les résidus . . . . .	55
3.7	Prévision : . . . . .	57
<b>4</b>	<b>La modélisation multivariée</b>	<b>59</b>
4.1	Processus multivariés . . . . .	59
4.1.1	Fonction d'autocorrélation . . . . .	61
4.2	La représentation générale d'un modèle VAR . . . . .	61
4.3	Estimation des paramètres du modèle VAR : . . . . .	62
4.3.1	Méthode d'estimation : . . . . .	63
4.3.2	Détermination du nombre de retards p : . . . . .	63
4.4	Prévision . . . . .	63
4.5	Causalité au sens de Granger : . . . . .	64
<b>5</b>	<b>Régression simple et multiple</b>	<b>65</b>
5.1	Modèle de régression simple . . . . .	65
5.1.1	Présentation du modèle . . . . .	65
5.1.2	Rôle du terme aléatoire . . . . .	65
5.1.3	Estimation des paramètres . . . . .	66
5.1.4	Equation et tableau de la variance . . . . .	66
5.1.5	Tableau d'analyse de la variance . . . . .	68
5.1.6	Test de spécification de l'ajustement . . . . .	68
5.2	Modèle de régression multiple . . . . .	69
5.2.1	Présentation du modèle . . . . .	69
5.2.2	Estimation des coefficients de régression . . . . .	70

## Table des matières

---

5.2.3	Hypothèses des estimateurs . . . . .	70
5.2.4	Equation d'analyse de la variance . . . . .	71

## II 73

### 6 Application 74

	Application de la méthode de BOX ET JENKINS . . . . .	74
6.1	Étude de la série du Prix <sub>t</sub> . . . . .	74
6.1.1	Identification . . . . .	74
6.1.2	Test de la stationnarité de la série $dPRIX_t$ . . . . .	79
6.1.3	Identification du modèle . . . . .	81
6.1.4	Estimation des paramètre du modèle . . . . .	81
6.1.5	Validation du modèle . . . . .	82
6.1.6	La prévision . . . . .	85
6.2	Étude de la série du DCOVT <sub>t</sub> . . . . .	87
6.2.1	Identification : . . . . .	87
6.2.2	Test de la stationnarité de la série $dDCOVT_t$ . . . . .	93
6.2.3	Identification du modèle : . . . . .	95
6.2.4	Estimation des paramètre du modèle : . . . . .	95
6.2.5	Validation du modèle : . . . . .	96
6.2.6	La prévision . . . . .	100
6.3	Étude de la série du INVES <sub>t</sub> . . . . .	102
6.3.1	Identification : . . . . .	102
6.3.2	Test de la stationnarité de la série $dINVES_t$ . . . . .	106
6.3.3	Identification du modèle : . . . . .	108
6.3.4	Estimation des paramètre du modèle . . . . .	109
6.3.5	Validation du modèle : . . . . .	109
6.3.6	La prévision . . . . .	113
	Application du Modèle VAR . . . . .	114
6.4	Etude multivariée des séries( $dPRIX_t, dDCOVT_t$ ) . . . . .	114
6.4.1	Identification du modèle VAR(P) . . . . .	115
6.4.2	Estimation des paramètres du modèle VAR(6) . . . . .	115
6.4.3	Validation . . . . .	117
6.4.4	Test de Granger de Causalité . . . . .	119
6.4.5	Test sur l'exogénéité des variables ( $dPRIX_t$ , $dDCOVT_t$ ) . . . . .	120
6.4.6	Prévision . . . . .	121
6.4.7	Conclusion . . . . .	123
6.4.8	Comparaison . . . . .	123
	Application de la Régression . . . . .	125
6.5	Etude de l'influence des forages,et du sismiques sur les découvertes . . . . .	125

## Table des matières

---

6.5.1	Statistique du modèle . . . . .	125
6.5.2	Estimation des coefficients . . . . .	125
6.5.3	Analyse des résidus . . . . .	126
6.5.4	Identification du modèle . . . . .	129
6.5.5	Estimation des paramètre du modèle . . . . .	129
6.6	Conclusion : . . . . .	133
	<b>Conclusion générale</b>	<b>134</b>
	<b>Bibliographie</b>	<b>136</b>

# Introduction générale

En matière d'énergie, les pays producteurs d'hydrocarbures et particulièrement ceux qui exportent la majorité de leur production, doivent veiller à l'équilibre du potentiel des réserves d'hydrocarbures découverts, par rapport au volume de production annuelle.

La traduction de cet équilibre consiste à ce que chaque année, les nombres de découvertes couvrent suffisamment les volumes de la production, afin de maintenir le ratio stratégique du secteur des hydrocarbures.

L'exploration systématique du sous-sol algérien, en vue d'approfondir les connaissances dans le domaine minier et dans la recherche de nouveaux gisements, constitue depuis longtemps un objectif majeur de la politique énergétique nationale.

Devant l'augmentation de la demande annuelle d'hydrocarbures, SONATRACH s'est empressée de s'inscrire dans une nouvelle politique de partenariat. Il s'agira pour elle de rechercher l'amélioration du taux de récupération par l'association du capital étranger aux gisements et aux contrats de prospection et d'exploration des nouveaux gisements par des compagnies pétrolières internationales.

la stratégie de la Sonatrach à l'horizon 2030 (SH 2030) d'aller vers les régions pour expliquer les objectifs à court, à moyen et à long termes.

la présentation de la stratégie s'est focalisée sur les aspects de développement en relation avec l'activité Aval, et permettre au groupe pétrolier national de consolider sa place parmi les cinq grandes compagnies pétrolières mondiales.

La prévision est une activité préalable à toute prise de décision stratégique. D'une manière générale, une entreprise doit prendre en considération l'outil statistique qui est considéré aujourd'hui comme fiable, il peut fournir une représentation et une interprétation des données économiques ou autres. Elle doit établir des prévisions régulièrement pour pouvoir contrôler les paramètres jugés importants par l'entreprise.

Notre étude consiste à analyser les séries chronologiques des découvertes en hydrocarbures (pétrole et gaz) de l'Algérie depuis l'origine, réalisées par SONATRACH seule ou en partenariat, ainsi que celles des efforts d'exploration en terme physique et financier. La deuxième partie de notre travail sera consacré à trouver un modèle mathématique qui permettra de mettre en évidence une corrélation entre le nombre des découvertes et l'effort financier de l'exploration

### Notre travail s'articulera donc autour de sept chapitres :

1. • Dans le premier chapitre, nous exposons l'activité de l'Exploration Pétrolière, nous présentons ses caractéristiques et ses techniques, et les plus importantes études antérieures sur l'impact des fluctuations des prix du pétrole sur les activités d'exploration et de production pétrolière de Sonatrach.
2. • Quand au deuxième chapitre, nous parlons du Pétrole et les marchés pétroliers.
3. • Dans le troisième chapitre, nous donnons des notions de base des processus stochastique, nous étudions les séries chronologiques.
4. • Dans le quatrième chapitre, est consacré à la méthodologie de Box-Jenkins. L'approche adoptée par cette méthode consiste en une représentation d'une classe de modèles qui représenteront le mieux les données observées, en passant par les étapes suivantes : l'identification, l'estimation et la validation du modèle.
5. • Dans le cinquième chapitre, nous élargissons notre étude à une analyse multivariée ; ainsi nous élaborons un modèle dont le comportement de sa variable endogène n'est plus expliquée que par ses observations passées, mais aussi par d'autre variables pouvant avoir une influence sur la variable endogène en question, en utilisant la modélisation VAR qui permet d'interpréter des relations dynamiques existant entre des variables stationnaires.
6. • Dans le sixième chapitre nous introduisons le concept de la régression simple et multiple, nous donnons des définitions et concept de base, et les différents estimateurs.
7. • Nous réservons le septième chapitre pour l'application des différentes méthodes. Nous colorerons notre étude par une comparaison entre les différentes techniques utilisées via le critère des écarts prévisionnels moyens (RMSE).
8. • Enfin, la clôture de la présente étude est faite par une conclusion.

# Problématique

Dans le monde actuel, le pétrole et le gaz naturel occupent une place essentielle parmi les sources d'énergie exploitées. La consommation annuelle d'hydrocarbures a augmenté, elle doit, selon les prévisions de certaines agences spécialisées, croître encore de manière importante pendant les prochaines décennies.

Pour satisfaire ces besoins, on recherche et on exploite actuellement des gisements situés à des profondeurs de plus en plus grandes, dont les caractéristiques posent des problèmes techniques difficiles. La prospection et l'exploitation ont du aussi être étendues, souvent au prix de la mise au point de techniques difficiles et des aires géographiques nouvelles.

L'exploration pétrolière est une opération fortement capitaliste, entourée de risques et d'aléas ; par ce fait le manager est confronté à de lourdes responsabilités se résumant à la prise de décision de forer ou non les prospects proposés.

En effet, l'analyse des incertitudes et des risques sera une composante fondamentale des études préalables à une décision d'investissement ; pour cela les décideurs ont tenté de trouver les outils leur permettant de s'orienter vers les choix les plus cohérents en vue de s'asseoir une stratégie d'exploration et d'optimiser la politique d'investissement dans cette activité.

Une découverte d'hydrocarbures est considérée comme étant le résultat final du processus d'exploration, les questions qui se pose sont :

1. De quelle manière l'évolution des découvertes suit l'évolution de l'effort d'exploration ?
2. Est-il possible de trouver une relation entre l'effort d'exploration et les résultats qui en découlent : les découvertes, et est ce que ce modèle représente assez bien le phénomène afin de faire des prévisions et jusqu'à quel horizon ?

Face à cette préoccupation la SONATRACH, nous a chargé de réaliser une étude dont l'objectif est :

- a. Analyser les séries chronologiques des découvertes en hydrocarbures (pétrole et gaz) de l'Algérie depuis l'origine réalisées par SONATRACH seule ou en partenariat ainsi que celle des efforts d'exploration en terme physique et financier. Pour cela nous disposons d'informations obtenues en 36 ans d'observation (annuelles 1986-2021), elles contiennent :
  - Les nombres de découvertes.
  - Le prix du pétrole brut à chaque année en Dollars (en \$ U.S).

## Problématique

---

- b. Faire des prévisions sur les séries étudiées durant les 04 prochaines années.
- c. Trouver un modèle mathématique qui permettra de mettre en évidence une corrélation entre les nombres de découvertes, les efforts physique et financier d'exploration et le niveau des prix du pétrole brut.

# Première partie

# Chapitre 1

## Exploration-Production

Depuis la création de SONATRACH, l'Activité Exploration-Production en hydrocarbures est montée en puissance grâce à l'expertise technologique et le savoir-faire des équipes dans la recherche et la découverte de nouveaux gisements sur le territoire national et à travers le monde.

L'activité exploration-production, est la division la plus stratégique de la compagnie nationale des hydrocarbures. C'est le cœur de Sonatrach, soit 70 % de son budget. Pour le personnel, cette activité compte 27000 permanents et 4000 temporaires. C'est cette activité qui assure les ressources pétrolières et gazières pour l'entreprise Sonatrach et pour le pays. Et pour ce faire, des processus sont mis en œuvre dont chacune de ses six divisions techniques :

Division Exploration

Division Petroleum Engineering et Développement

Division Forage

Division Production

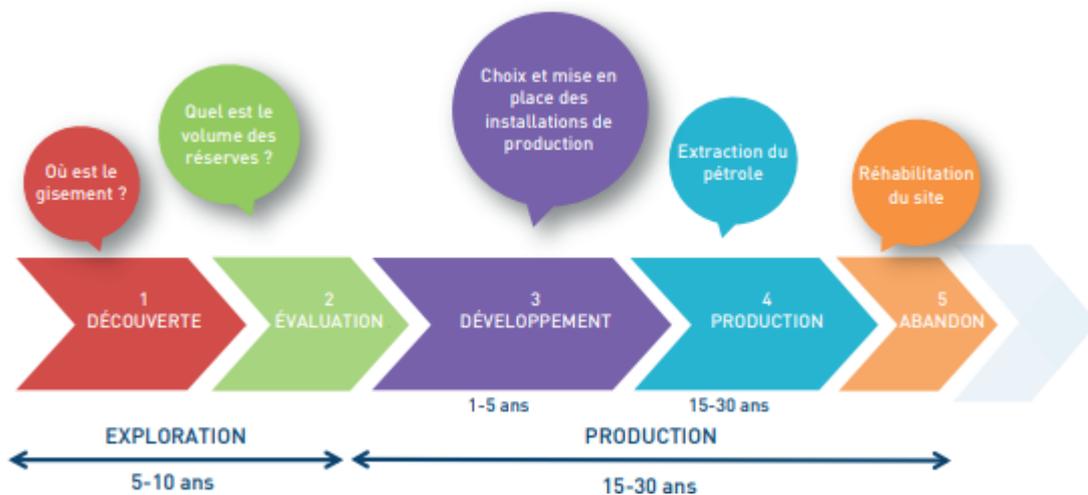
Division Associations

Division Laboratoire

Par ailleurs, l'activité exploration-production (EP) de Sonatrach a pour mission la recherche, le développement, l'exploitation et la production des hydrocarbures. Elle s'articule autour de trois axes :

1. le développement et l'exploitation des gisements pour une valorisation optimale des ressources , de développement et d'exploitation des gisements.
2. la gestion des activités en partenariat dans les phases d'exploration.
3. la recherche, la négociation et le développement de nouveaux projets sur le territoire national.

### 1.1 Prospection / exploration gazière et pétrolière



#### 1.1.1 Définition

Prospection et exploration gazières et/ou pétrolières visent à découvrir de nouveaux gisements de gaz naturel ou de pétrole.

Ces deux ressources fossiles résultent de la transformation des déchets organiques animaux et végétaux contenus dans les boues gorgées d'eau se sédimentant au fond des mers. Leur enfouissement progressif pendant des millions d'années les soumet à des températures (géothermie) et des pressions (gravité) croissantes.

En l'absence d'oxygène, ces matières organiques se pyrolysent en hydrocarbures au sein des roches-mères, soit sous forme d'huile (pétrole) entre 1,5 et 3 km de profondeur, soit sous forme de gaz (méthane) au-delà. Sous l'effet de la pression intense, ces hydrocarbures de densité inférieure à l'eau, sont alors poussés à migrer hors de leur roche-mère et à remonter vers la surface au travers de couches plus perméables. Si leur remontée est bloquée par des strates imperméables (argile) formant voûte, ils peuvent s'accumuler dans des roches perméables sous-jacentes (roches-réservoirs) en formant un gisement de pétrole et/ou de gaz naturel.

L'objectif de la prospection est de localiser un gisement. L'exploration doit en vérifier l'existence et en évaluer l'importance et la qualité grâce à des forages dont l'emplacement est déterminé en associant géologie et géophysique. En cas de succès, ces deux phases en amont sont suivies par les phases d'exploitation, de transport et de commercialisation.

#### 1.1.2 La découverte du champ

1. La prospection géologique en surface :

## Chapitre 1. Exploration-Production

---

Les caractéristiques géologiques des gisements pétroliers diffèrent en fonction de leur âge (de 5 à 400 millions d'années), de leur profondeur (de 1 à 10 km) et de leur thermique (la formation de l'huile se situant entre 60 et 150 °C).

Pour identifier les régions potentiellement pétrolifères, les géologues s'interrogent sur les points suivants :

Quelle est la nature des roches ?

Ont-elles été soumises à des conditions favorables à la création d'hydrocarbures ?

Ces hydrocarbures ont-ils pu migrer et être piégés par des couches imperméables ?

Les géologues dressent une carte du sous-sol à partir des informations obtenues en surface par examen des affleurements et dans les airs par photogéologie(1).

Lorsqu'une zone favorable (prospect) est repérée par les géologues depuis la surface, c'est au tour des géophysiciens d'explorer le sous-sol.

### 2. La prospection géophysique en profondeur :

La sismique réflexion est la méthode principale des géophysiciens pour repérer des gisements potentiels :

- Sur terre (onshore), à partir d'un choc ou de vibrations sonores ébranlant le sol, on détecte par un réseau de géophones les échos réfléchis partiellement par les couches géologiques. On obtient ainsi une échographie 2D de la structure des couches prospectées.

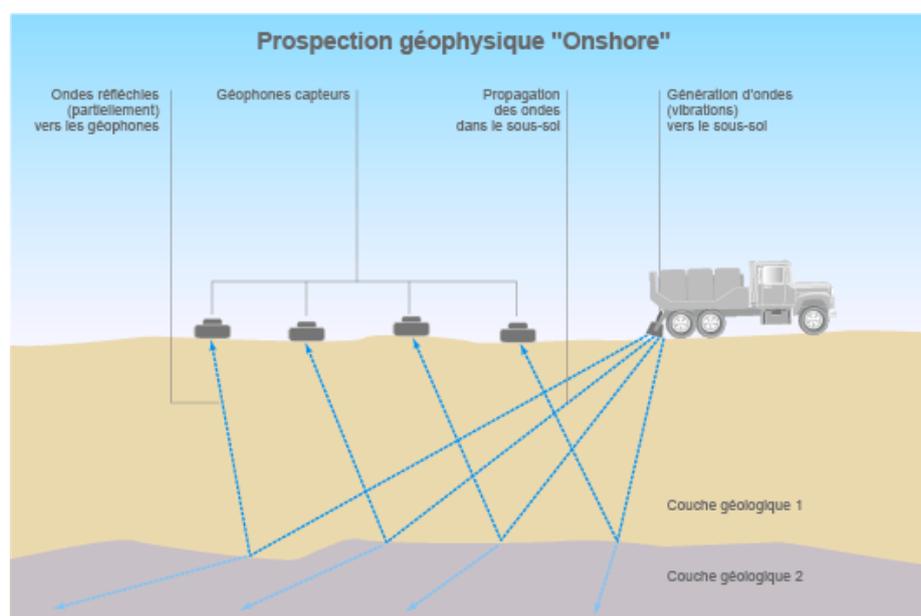


FIGURE 1.1 – Prospection géophysique sur terre ((©)Connaissance des Énergies)

- En mer (offshore), on produit l'onde sismique par air comprimé à haute pression et on recueille les échos sur des hydrophones flottants (flûtes), la couche d'eau étant

considérée comme homogène.

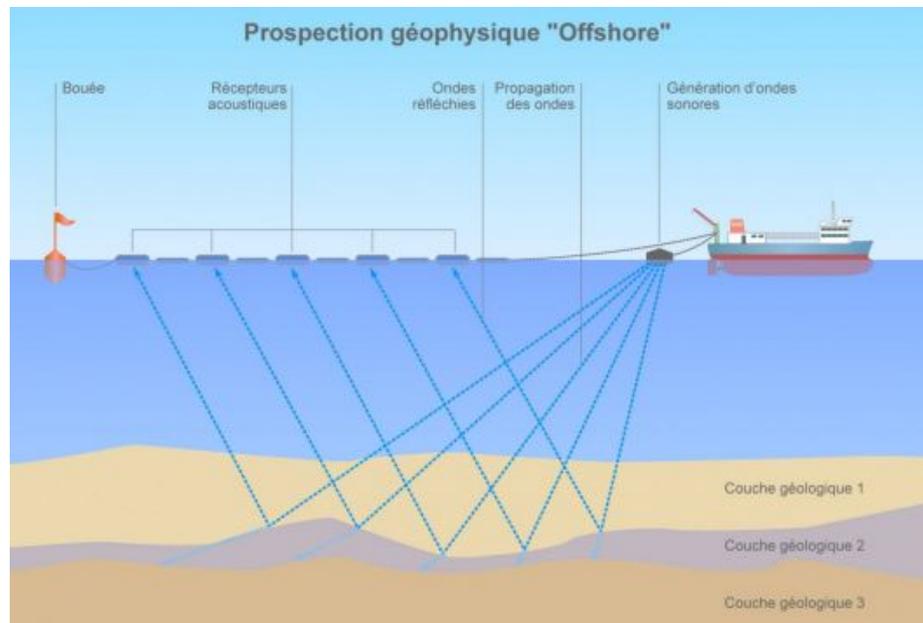


FIGURE 1.2 – Prospection géophysique offshore (©Connaissance des Énergies)

Pour passer en sismique 3D, on multiplie les flûtes et les angles de production des ondes sismiques pour permettre de construire des images du sous-sol en volume. Beaucoup plus onéreuse(2), l'imagerie sismique 3D est aussi beaucoup plus précise et permet de visualiser les volumes des gisements. En intégrant le facteur temps, on peut analyser l'évolution des gisements en cours d'exploitation en 4D.

### 3. Les forages d'exploration :

Après la prospection, le forage est la seule méthode pour confirmer la présence d'hydrocarbures et pour définir :

- la qualité de l'effluent du puits (huile saturée de gaz ,eau) .
- la perméabilité du réservoir .
- la production potentielle et la quantité d'huile.

Forer consiste à percer l'écorce terrestre pour atteindre les zones pétrolifères, au-delà de deux kilomètres. Pour les gisements conventionnels terrestres, on fore généralement à la verticale mais des forages horizontaux sont pratiqués pour les gisements de grande étendue et de faible épaisseur. En mer, pour des raisons économiques, des forages orientés multiples sont effectués à partir d'une plateforme unique.

Dans un forage vertical classique, la tête de forage est un trépan doté de dents en acier très dur, parfois diamanté, mis en rotation rapide par un train de tiges

creuses reliées à une tour verticale d'une trentaine de mètres de haut dans laquelle sont regroupés la table de rotation et les pompes d'aspiration et d'injection. Au fur et à mesure de la descente du trépan, on visse en surface des tiges supplémentaires. Simultanément, on procède au tubage externe du forage par des cylindres creux en acier de diamètre supérieur au trépan que l'on gaine de ciment. Pour débarrasser en permanence le fond du forage des débris de roche arrachés par le trépan, on injecte sous haute pression dans le train de tiges en rotation une boue fluide qui traverse le trépan et remonte par le tubage externe en entraînant les débris. Cette boue est filtrée en surface, analysée et réinjectée dans le train de tiges. Au-delà de l'évacuation des débris, ce fluide équilibre la pression sur les parois du puits, lubrifie et refroidit le trépan et peut empêcher d'éventuelles éruptions.

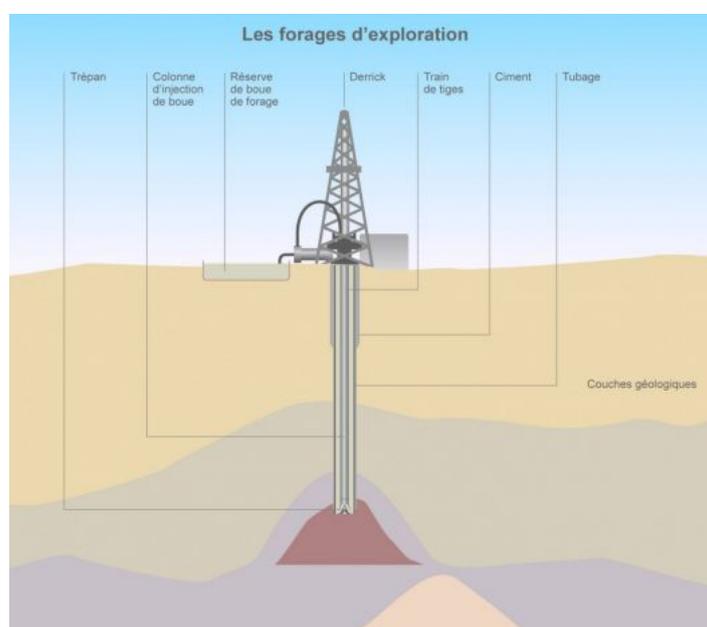


FIGURE 1.3 – Les forages d'exploration ((©)Connaissance des Énergies)

La profondeur des trous de forage est habituellement comprise entre 2 000 et 4 000 m et peut atteindre 6 000 m. Lorsque des traces d'hydrocarbures sont détectées dans le fluide remontant en surface, on procède à un carottage avec un trépan spécial qui découpe un cylindre dans la roche. Une fois remontée, cette carotte fournit des informations clés sur la teneur en hydrocarbures de la roche traversée. Si un gisement est atteint, le forage est arrêté. Des explosifs sont descendus pour percer le tubage et laisser le pétrole pénétrer dans le puits et remonter à la surface si la pression est forte. Une tête de puits est alors installée pour mesurer le débit et évaluer la productivité du gisement. En cas de succès, d'autres forages sont réalisés pour en confirmer le potentiel. Puis viennent les multiples études économiques pour

en estimer la rentabilité avant une décision de mise en exploitation.

### 1.1.3 L'évaluation du champ

Une fois la découverte confirmée, des modèles numériques du gisement (souvent en 3D) sont réalisés sur ordinateur. Ils permettent :

- D'estimer le volume de pétrole et de gaz en place dans le réservoir avant la mise en production du gisement.
- De simuler l'extraction des fluides contenus dans la roche réservoir.
- De créer des scénarii potentiels de mise en production du champ, en fonction de divers facteurs tels que techniques d'extraction, nombre de puits, taille et type des installations de production, etc.

Des forages d'appréciation du gisement sont implantés pour obtenir de nouvelles données sur les caractéristiques du réservoir.

**Si la balance entre la quantité des « ressources en place » et le coût estimé du développement est jugée économiquement satisfaisante, le gisement est développé puis mis en production.**

## 1.2 LA PRODUCTION DE PÉTROLE ET DE GAZ

### 1.2.1 Le développement du champ

C'est la phase d'exploitation du gisement qui demande la mise en place de tout l'équipement nécessaire.

Le plan de développement du champ définit :

- le nombre de puits à forer pour pouvoir produire.
- les techniques de récupération et d'extraction du pétrole emprisonné dans la roche réservoir.
- le type et le coût des installations comme les plates-formes, en fonction des aléas du milieu marin (marées, tempêtes, courants, vents, corrosion...),
- les dispositifs de séparation des gaz et des fluides.
- les sites de traitements pour préserver l'environnement.

**Remarque sur le forage :**

**La technique de forage la plus répandue est celle du forage Rotary qui s'est beaucoup renouvelée, en particulier avec les forages déviés - permettant de contourner un obstacle souterrain - ou horizontaux – permettant de traverser le réservoir sur toute sa longueur. Les puits multidrains, quant à eux, permettent**

de limiter le nombre de forages, en produisant à partir d'une tête de puits unique

### 1.2.2 La production du champ

La période pendant laquelle on extrait les hydrocarbures varie généralement de 15 à 30 ans et peut se prolonger jusqu'à 50 ans et plus pour les « champs géants ».

La durée de vie du gisement se compose de différentes phases successives :

- Une période de croissance.
- Une phase de stabilisation ou « plateau ».
- Des phases d'injection d'eau, de gaz ou de produits chimiques pour aider ou « assister » la récupération du pétrole et maintenir un volume satisfaisant de ressources produites.
- Une phase d'épuisement : la production de pétrole décline progressivement.

**Remarque sur les grandes profondeurs d'eau (« Deep Offshore ») :**

Depuis quelques années, l'exploration mais surtout la production par grande profondeur d'eau (supérieure à 1 000 m d'eau) ont connu des avancées technologiques majeures. Au total, plus de 1 300 puits ont été forés à ces profondeurs. Cette production reste cependant particulièrement complexe et coûteuse ; elle représente, encore aujourd'hui, un challenge technologique.

### 1.2.3 L'abandon du champ

Quand le débit d'hydrocarbures devient minimum, le gisement est épuisé ; c'est la phase de fermeture du gisement.

Avant d'abandonner le champ, les compagnies pétrolières :

- Démantèlent les plates-formes.
- Mettent en sécurité les puits.
- Préservent les réserves de pétrole résiduelles.
- Nettoient, dépolluent et réhabilitent le site si nécessaire, conformément aux législations nationales et internationales.

## 1.3 les variations du prix du pétrole sur l'activité d'exploration

les fluctuations des prix du pétrole est considéré comme l'un des événements les plus importants en termes d'impact sur les changements macroéconomiques et les indicateurs au niveau des pays exportateurs ainsi que des partiels au niveau des compagnies pétrolières. Par conséquent, des prévisions liées à la taille des réserves de pétrole équitables sont utilisées, la capacité de production, les taux de consommation et les niveaux de prix afin de

développer les programmes d'investissement des compagnies pétrolières, et sur la base de ces prévisions sont pris décisions d'investissement dans l'industrie pétrolière, en particulier dans la phase amont, qui se caractérise par un degré de risque élevé ainsi qu'un coût en capital élevé.

La baisse des prix du pétrole entraîne une réduction des revenus pétroliers de la société, ce qui peut entraîner une réduction des dépenses d'investissement des projets L'industrie pétrolière, en particulier les projets d'exploration et de production, qui sont les plus pris en compte dans les dépenses d'investissement dans les étapes de l'industrie pétrolière.

Pour tout cela, il valait la peine de se poser la question suivante :

**: Quel est l'impact des variation du prix du pétrole sur les investissements pétroliers de Sonatrach en étape amont ?**

### **l'importance de l'étude :**

L'autorisation de l'étude réside dans le fait qu'il existe une relative rareté d'études économiques sur l'impact des fluctuations des prix du pétrole sur les compagnies pétrolières, car la plupart des études ont examiné l'impact de ces fluctuations sur l'économie globale et sur les pays exportateurs et importateurs, alors que nous avons essayé dans cette étude de clarifier l'impact de Les fluctuations du prix du pétrole au niveau micro, notamment au niveau de Sonatrach, qui est considéré comme une valeur ajoutée pour la recherche et les études dans ce domaine.

### **Objectifs de l'étude :**

Les objectifs de cette étude sont les suivants :

- Se familiariser avec les investissements pétroliers de Sonatrach ;
- Clarifier et suivre l'impact des fluctuations des prix du pétrole sur les activités d'exploration et de production pétrolière de Sonatrach.

## **1.4 L'importance de l'industrie pétrolière en Algérie**

L'industrie pétrolière en Algérie est considérée comme l'une des composantes les plus importantes du secteur de l'énergie, car l'économie algérienne est complètement dépendante du secteur des hydrocarbures, où tous Les activités économiques et sociales et les politiques de développement dépendent des revenus d'exportation des hydrocarbures, car le secteur des hydrocarbures contribue de manière significative aux revenus de l'État. Et cela grâce à la fiscalité pétrolière, qui représente environ 97% des revenus, puisque Sonatrach obtient environ 80% de son budget d'investissement à l'Exploration et production pétrolières, et Sonatrach a décidé d'investir 53 milliards de dollars dans la production d'exploration au

cours de la période (2017-2021) Cela représente 9 milliards de dollars pour l'exploration, avec l'objectif de pouvoir forer environ 100 puits par an dans le cadre de sa stratégie de croissance SH2030. Sonatrach ambitionne de doubler le volume annuel des découvertes et d'augmenter les réserves prouvées de 50 à 100 millions de tonnes par an.

### 1.5 L'évolution des investissements en exploration compte tenu de l'évolution des prix du pétrole

- Le prix du pétrole est resté élevé depuis 2009, atteignant son pic en 2011, atteignant 112 dollars le baril, en raison de la reprise économique après la crise financière, ainsi que la réduction de la production pétrolière à l'OPEP, à partir de janvier 2009 Cette hausse des prix a fortement encouragé l'investissement de la société dans l'exploration.
- Sonatrach a relevé le volume de ses investissements orientés vers l'exploration à partir de 2012, avec un volume de 1897 millions de dollars à 2724 millions de dollars en 2014, soit une augmentation estimée jusqu'en 2014, à un rythme estimé à 43 %.  
- Malgré la forte baisse des prix du pétrole sur la période (2013-2016), l'entreprise est restée ferme dans ses dépenses d'investissement dans l'activité d'exploration pétrolière, car la baisse à court terme n'affecte pas les investissements pétroliers de l'entreprise, mais à moyen et long terme, de fortes fluctuations conduisent à une baisse de l'investissement, par exemple le prix du pétrole est passé de 109 dollars le baril en 2013 à 52 dollars en 2015 le volume des investissements était de 2526 millions de dollars et 2559 millions de dollars sur Cela signifie que la volatilité des investissements n'était pas grande, contrairement à la volatilité des prix
- Les investissements dans le secteur de l'exploration ont enregistré une baisse notable au cours des dernières années, passant de 1 409 millions de dollars en 2014 à 2784 millions de dollars en 2018, avec une baisse estimée à 45%, due à la baisse des prix du pétrole, 60 dollars par baril.
- Les prix du pétrole ont commencé à se redresser depuis 2017, enregistrant un prix moyen de 74 dollars le baril, atteignant 71 dollars le baril sur l'année En 2018, toutefois, les investissements dans le secteur de l'exploration sont demeurés stagnants, atteignant un niveau record de 1 611 M\$ en 2017 Il est reparti à la baisse à 1 409 millions de dollars en 2018, et cette baisse est due à une forte diminution de l'absentéisme du partenaire étranger, ainsi qu'à la stratégie d'investissement suivie par l'entreprise.

## 1.6 L'évolution des découvertes pétrolières à la lumière de l'évolution des prix du pétrole

- Le nombre de puits découverts par an pour la société est passé de 9 découvertes pétrolières en 2000 à 30 découvertes pétrolières en 2018, de 230 % Le baril de pétrole brut est également passé de 29 dollars le baril à 71 dollars, soit environ 250 %.
- Une évolution remarquable du nombre de puits découverts Au cours de la période 2000-2009, le nombre de puits découverts est passé de 9 puits en 2000 à 19 puits en 2009, en raison de l'amélioration des prix du pétrole au cours de cette période, qui a conduit à une augmentation des investissements dans le processus d'exploration d'une part, ainsi qu'à une augmentation Ouverture au partenariat étranger dans l'exploration des puits de pétrole et de gaz

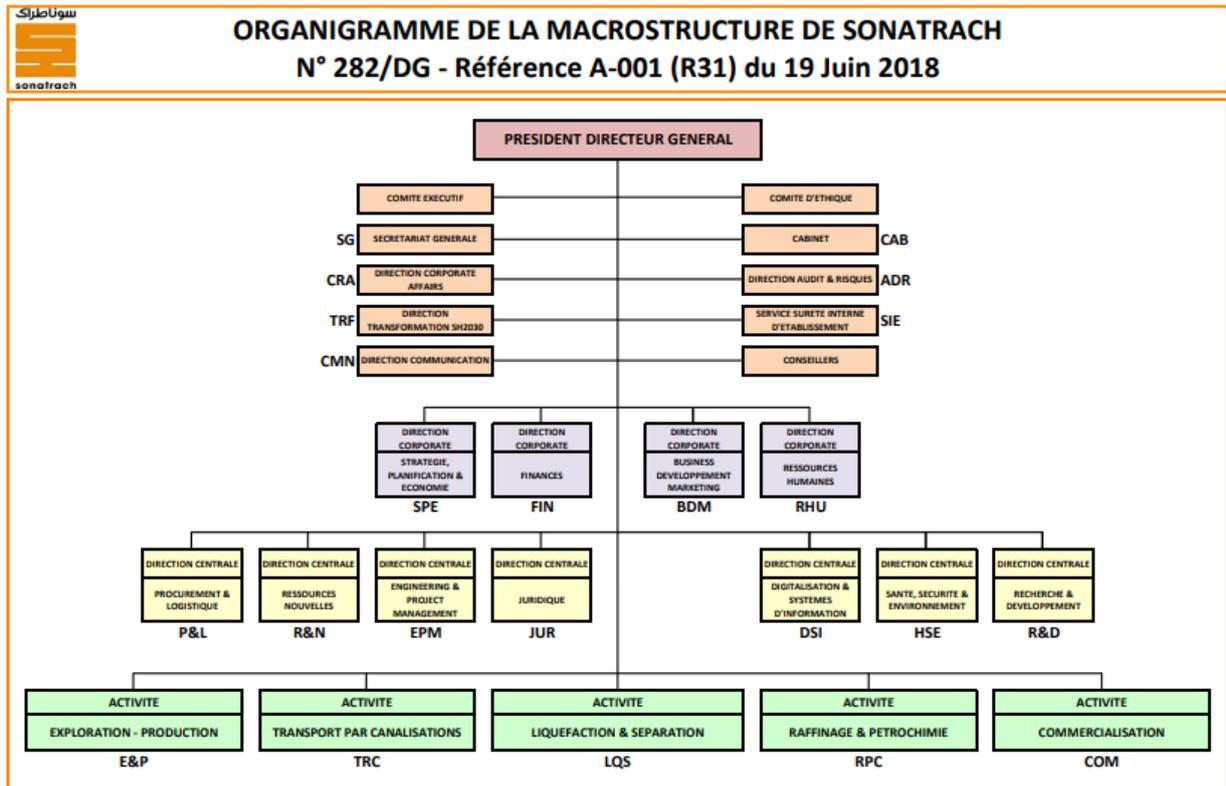
## Chapitre 1. Exploration-Production

---

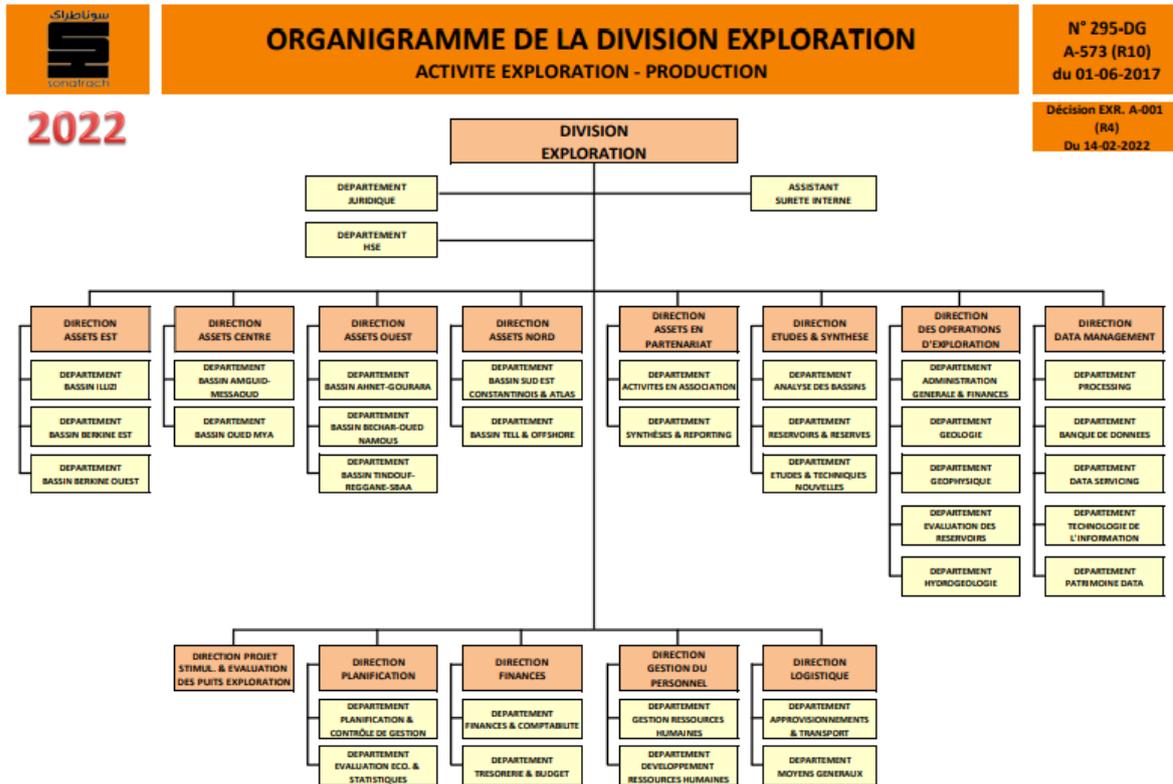
L ' année	Découverte de pétrole	Découverte de gaz	total
2000	6	3	9
2001	3	3	6
2002	3	3	6
2003	4	3	7
2004	4	8	12
2005	6	2	8
2006	11	6	17
2007	5	15	20
2008	3	13	16
2009	4	12	16
2010	14	15	29
2011	10	10	20
2012	8	23	31
2013	12	20	32
2014	15	17	32
2015	12	18	30
2016	17	16	33
2017	20	13	33
2018	14	16	30

TABLE 1.1 – : L'évolution des découvertes pétrolières et gazières de Sonatrach (2000-2018)

## 1.7 Organigramme de la macrostructure de Sonatrach



## 1.8 Organigramme de la division exploration



# Chapitre 2

## Processus stochastique et séries chronologiques :

### Introduction :

L'étude des séries chronologiques ou temporelles correspond à l'analyse statistique d'observations régulièrement espacées dans le temps, dans le but de représenter des phénomènes aléatoires qui évoluent dans le temps, le modèle obtenu sera par la suite utilisé selon les objectifs désirés, tels que la prévision ou le contrôle.

Une démarche générale de prévision a été popularisée par Box et Jenkins (1970), c'est une méthode extrapolative du fait que seul le passé de la variable étudiée est utilisé. Cette méthode est fondée sur la notion des modèles ARMA, et utilise des concepts et résultats de la théorie des processus aléatoires.

### 2.1 Définitions et concept de base

Dans ce qui suit  $(\Omega, A, P)$  désigne un espace probabilisé et  $(T, \tau)$  un espace probabilisable.

#### 2.1.1 Processus aléatoire

On appelle processus stochastique ou processus aléatoire toute famille de variables aléatoires  $X_t$ . Cela signifie qu'à tout  $t \in T$  est associée une variable aléatoire prenant ses valeurs dans un ensemble numérique  $E$ . On note le processus  $X_t$ . Si  $T$  est dénombrable, on dit que le processus est discret ; si  $T$  est un intervalle, on dit que le processus est permanent.

#### 2.1.2 Processus aléatoire du second ordre

Un processus aléatoire  $X_t, t \in Z$  est dit du second ordre si et seulement si :

$E(X_t^2) < \infty \forall t \in T$ , c'est - à - dire que  $X_t$  admet un moment d'ordre deux fini.

### 2.1.3 Les processus stationnaires

La notion de stationnarité joue un rôle central dans la théorie des processus. Deux types de stationnarité sont généralement considérés. Et dans ce qui suit nous passons en revue les définitions de bases liées à ces type de processus stationnaires.

#### Processus strictement stationnaire (la stationnarité fort) :

Soit un processus stochastique  $X_t \in Z$ , le processus est dit strictement (ou fortement)stationnaire si :

$\forall (t_1, t_2, \dots, t_n) \in Z^n$  alors la suite  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$  a la même loi de probabilité que la suite  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  autrement dit :

$$P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) = P(X_{t_1+h} \leq x_1, \dots, X_{t_n+h} \leq x_n)$$

$$\forall (t_1, t_2, \dots, t_n) \in Z^n; \forall (x_1, x_2, \dots, x_n) \in R^n, \forall h \in Z$$

#### Processus faiblement stationnaire (second ordre) :

Le processus  $X_t, t \in Z$  est dit faiblement stationnaire si :

- $E(X_t) = \mu, \forall t \in Z$

- $Var(X_t) = \sigma_x^2 = \gamma_0, \forall t \in Z$

- $Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)] = \gamma_x(h), \forall t, h \in Z$

$\gamma_x(h)$  est la fonction d'autocovariance du processus  $X_t, t \in Z$

#### Remarque :

La covariance d'un processus faiblement stationnaire dépend seulement de la différence entre les instants.

Désormais, le terme stationnaire renverra au concept de stationnarité du second ordre, sauf mention contraire.

#### Processus bruit blanc :

Un bruit blanc  $\epsilon_t, t \in Z$  est une suite de variables aléatoires non corrélées de moyenne nulle et de variance finie constante.

Un processus bruit blanc vérifie les propriétés suivantes :

- $E(\epsilon_t) = 0, \forall t \in Z$

- $Var(\epsilon_t) = E(\epsilon_t^2) = \sigma_\epsilon^2, \forall t \in Z$

Et on conséquence sa fonction d'autocovariance est donnée par :

$$\gamma_h = \text{Cov}(X_t, X_{t+h}) = E(\epsilon_t, \epsilon_{t+h}) = \begin{cases} 0 & h \neq 0 \\ \sigma_\epsilon^2 & h = 0 \end{cases}$$

Si de plus  $\epsilon_t \rightarrow N(0, \sigma^2)$ , alors  $\epsilon_t$  est un processus bruit blanc gaussien.

## 2.2 Fonction d'autocovariance

La fonction d'autocovariance du processus stationnaire (du second ordre)  $X_t, t \in \mathbb{Z}$  notée  $\gamma(h)$  est définie par :

$$\gamma(h) = \text{Cov}(X_t, X_{t-h}) = E[(X_t - E(X_t))(X_{t-h} - E(X_{t-h}))]; \forall (t, h) \in \mathbb{Z}^2$$

$$h = 0, \gamma(0) = \text{Var}(X_t)$$

### Propriétés :

$\gamma(-h) = \gamma(h); \forall h \in \mathbb{Z}$  (la fonction d'autocovariance est symétrique)

$|\gamma(h)| \leq \gamma(0) \leq \text{Var}(X_t); \forall h, t \in \mathbb{Z}$

### Estimateur :

Considérons  $(X_t, \dots, X_T)$ , l'estimateur de la fonction d'autocovariance est donnée par :

$$\hat{\gamma}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T)$$

$$\text{avec } \bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$$

$\hat{\gamma}(h)$  est appelée l'autocovariance empirique.

### Remarque :

La fonction  $\gamma(h)$  mesure la covariance pour un couple de valeurs séparées par un intervalle de longueur  $h$ , appelé retard .

Elle fournit une information sur la variabilité de la série ,

et sur les liaisons temporelles qui existent entre les diverses composantes de la série.

## 2.3 Fonction d'autocorrelation

La fonction d'autocorrelation de  $h$  ( $h \in \mathbb{Z}$ ) , d'un processus stationnaire du second ordre de moyenne  $E(X_t) = \mu$ , notée  $\rho(h)$ , est défini par :

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

### Propriétés :

- $-1 \leq \rho(h) \leq 1$

- $\rho(0) = 1$

**Estimateur :**

$$\rho(\hat{h}) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{T}{T-h} \frac{\sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T)}{\sum_{t=1}^T (X_t - \bar{X}_T)^2}$$

**Remarque :**

- La représentation graphique de  $\rho(h)$  est appelée corrélogramme.
- Si la fonction d'autocorrélation  $\rho(h)$  décroît rapidement quand le nombre de retards augmente, cela signifie que la série est stationnaire
- $\rho(h)$  représente le coefficient de corrélation pour un couple de valeurs séparées par un intervalle de longueur  $h$ .
- L'un des concepts les plus importants dans l'analyse des séries chronologiques est l'autocorrélation.

## 2.4 Fonction d'autocorrélation partielle

Elle mesure la corrélation entre  $X_t$  et  $X_{t-h}$ , l'influence des variables (pour  $i < h$ ) ayant été retirée

$\phi_{tt}$  : Fonction d'autocorrélation partielle de  $X_t$ .

Soit  $\underline{\rho}(h)$  la matrice symétrique formée des  $(h-1)$  premières autocorrélations de  $X_t$ .

$$\underline{\rho}(h) = \begin{bmatrix} 1 & \rho_1 & \rho_1 & \cdots & \rho_{h-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{h-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{h-2} \\ \vdots & & & \ddots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \rho_1 & 1 \end{bmatrix} \quad \underline{\rho}(h)^* = \begin{bmatrix} 1 & \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_2 \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_3 \\ \vdots & & & \ddots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \rho_1 & \rho_h \end{bmatrix}$$

$\underline{\rho}(h)^*$  est la matrice  $\underline{\rho}(h)$  dans laquelle on a remplacé la dernière colonne par le vecteur  $[\rho_1, \rho_2, \dots, \rho_n]^t$

$$\phi_{tt} = \frac{|\underline{\rho}(h)^*|}{|\underline{\rho}(h)|}$$

La fonction d'autocorrélation partielle d'un processus stochastique  $X_t$   $t \in \mathbb{Z}$  stationnaire du second ordre est définie par le coefficient de corrélation partielle  $\pi_h$ . Ces coefficients s'obtiennent à partir d'un système d'équation (algorithme de Durbin), avec :

$$\begin{cases} \pi(1) = \phi_{11} = \rho(1) & \text{initialisation} \\ \pi(h) = \phi_{hh} = \frac{\rho_h - \sum_{j=1}^{h-1} \phi_{h-1,j} \rho^{(h-j)}}{1 - \sum_{j=1}^{h-1} \phi_{h-1,j} \rho^{(j)}} & h = 2, 3, \dots, k \\ \phi_{hj} = \phi_{h-1,j} - \phi_{hh} \phi_{h-1,h-j} & j = 1, 2, \dots, h-1 \text{ et } h = 2, 3, \dots, k \end{cases}$$

**Remarque :** l'algorithme est basé sur les équation de Yule-Walker.

## 2.5 SERIES CHRONOLOGIQUES

L'approche de modélisation par les séries chronologiques est utilisée pour faire des prévisions, elle consiste à exploiter l'information contenue dans les valeurs passées d'une variable et des perturbations aléatoires (on aura besoin de collecter des informations sur une assez longue période pour avoir des prévisions fiables), de façon à déterminer les caractéristiques intrinsèques et la nature de l'évolution dans le temps de la série, nous pourrons alors prévoir les valeurs futures de la variable.

A titre d'exemple, nous citerons quelques domaines d'application :

- L'économétrie (prédiction de quantités économiques, les prix de ventes et d'achats, ...).
- La finance (évaluation des cours de la bourse au cours d'une séance ...).
- La météorologie (analyse des données climatiques, prévision, ...).

### 2.5.1 Définition

Une série chronologique dite aussi chronique ou série temporelle (time series en terminologie anglaise) est une suite d'observations  $X_t, t \in \mathbb{Z}$  indexées par un ensemble ordonné  $T$ . Autrement dit, une série chronologique est une réalisation du processus stochastique  $X_t, t \in \mathbb{Z}$ . les dates d'observations sont en général équidistantes : c'est le cas des séries journalières, hebdomadaire, mensuelles, annuelles... Suivant l'ensemble des indices  $T$  nous distinguons deux type de chroniques, à savoir :

#### Série continue :

Une série est continue lorsque l'ensemble des valeurs possibles de  $t$  est non dénombrable. Nous pourrons rencontrer ce genre de séries en physique quantique.

#### Séries discrètes :

Une série est discrète lorsque l'ensemble des valeurs possible de  $t$  est un ensemble dénombrable.

Nous distinguons deux types de variables constituant une série discrète :

**Les variables de flux :** elles représentent le mouvement intervenu durant un certain intervalle de temps (le nombre d'accidents durant l'année en cours, le trafic aérien quotidien...).

**Les variables de niveau :** elles représentent un état à un moment donnée (taux de chômage, température à un lieu fixe, ...). Une série chronologique peut être représentée graphiquement en plaçant les instants ( $t_i, 1 \leq i \leq n$ ) en abscisses et les observations ( $y_i, 1 \leq i \leq n$ ) en ordonnées (n le nombre d'observations).

### 2.5.2 Opérateurs linéaires

Nous considérons l'ensemble des entiers  $\mathbb{Z}$ , sur cet ensemble, nous définirons les opérateurs de retard  $B$ , d'avance  $F$ , de différence ordinaire  $\nabla$  et de différence saisonnière  $\nabla_s$ .

#### Opérateur de retard (Backward) :

Pour formaliser le déplacement dans le temps de la série temporelle, nous définissons une application, qu'à partir d'une observation prise à une date donnée nous permet d'exprimer les observations passées.

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

De manière générale :

$$B^n X_t = X_{t-n} \text{ et } (\sum_{i=0}^n a_i B^i) X_t = \sum_{i=0}^n a_i X_{t-i}$$

#### Opérateur d'avance (Forward) :

- $FX_t = X_{t+1}$
- $F^n X_t = X_{t+n}$

#### Propriétés :

- $B^0 X_t = X_t$
- $B^i B^j X_t = B^{i+j} X_t = X_{t-i-j}$
- $B^{-i} X_t = X_{t+i}; \forall i \in \mathbb{Z}$
- $B^{-1} = F$
- $(B^i + B^j) X_t = B^i X_t + B^j X_t = X_{t-i} + X_{t-j}, \forall (i, j) \in \mathbb{Z}^2$

Notons que ces propriétés s'appliquent également à l'opérateur  $F$ .

### Opérateur de différence ordinaire :

L'opérateur de différence ordinaire noté  $\nabla$ , associé au processus  $X_t, t \in \mathbb{Z}$  est telle que :

$$\forall t \in \mathbb{Z}, \nabla X_t = X_t - X_{t-1} = X_t - BX_t = (1 - B)X_t$$

Et par construction, nous obtiendrons l'opérateur de la  $d$ ème différence noté  $\nabla^d$  tel que :

$$\forall t \in \mathbb{Z}, \nabla^d X_t = (1 - B)^d X_t$$

### Opérateur de différence saisonnier $\nabla^s$ :

L'opérateur de différence saisonnier d'ordre noté  $\nabla^s$  associer au processus  $X_t, t \in \mathbb{Z}$  est tel que :  $\forall t \in \mathbb{Z}, \nabla_s X_t = (1 - B^s)X_t$

Et par construction, nous obtiendrons l'opérateur de la  $d$ ème différence d'ordre  $s$ , notée  $\nabla_s^d$  tel que :

$$\forall t \in \mathbb{Z}, \nabla_s^d X_t = (1 - B^s)^d X_t$$

### Remarque :

$\ll s \gg$  est la périodicité des données (  $s=12$  pour les données mensuelles ,  $s=4$  pour les données trimestrielles ,  $\dots$  ).

## 2.5.3 Analyse des séries chronologiques

L'analyse des séries chronologiques a pour objectif de d'écrire les principales caractéristiques du processus générateur de la série, l'ajustement du modèle adéquat, la prévision et le contrôle. Autrement dit, la première étape dans cette étude consiste à faire une analyse graphique des donnée, qui permet d'entrevoir les composantes fondamentales du mouvement d'ensemble, et donc d'avoir une idée générale sur le comportement de la série chronologique. -Composantes d'une série chronologique Les premières études sur les séries chronologiques ont amené à considérer que la chronique peut se mettre sous la forme fonctionnelle suivante :

$$Y_t = f(T_t, S_t, C_t, \epsilon_t)$$

- $T_t$  :représente la tendance de la chronique.
- $S_t$  :représente la saisonnalité.
- $C_t$  :représente le cycle conjoncturel.
- $\epsilon_t$  : représente les fluctuations irréguliers (erreurs) Donnons pour chacune de ces composantes, quelques définitions :

### La tendance (trend en terminologie anglaise) Notée $T$ :

La tendance décrit le mouvement à long terme de la série, ce mouvement est traditionnellement représenté par la formes : polynomial, logarithmique, exponentielle,  $\dots$  . Elle est en fonction du temps et marque l'allure générale du problème, elle peut être croissante ou décroissante.

### Le cycle conjoncturel :

Cette composante regroupe les variations autour de la tendance avec des alternances d'époques ou des phases d'expansion et de contradiction. Il décrit un mouvement ascendant ou descendant successif.

### La saisonnalité :

Cette composante décrit un mouvement plus ou moins régulier d'amplitude presque constante et qui se répète à des intervalles égaux. Elle représente les fluctuations de nature périodique et regroupe les effets météorologiques, les activités économiques et sociales (vacances, élection,  $\dots$ ).

### L'erreur (variation accidentelle ou aléa) :

Cette composante rassemble tout ce que les autres composantes n'ont pas pu expliquer du phénomène observé, elle résulte de multiples causes de nature aléatoire qu'on ne peut ni mesurer ni prévoir (panne, grève, catastrophes naturelles...), des erreurs de mesures, des aléa climatiques (inondation, sécheresses,  $\dots$ ), économiques  $\dots$ .

- Ces différentes composantes peuvent être combinées selon un des trois modèles suivants :

**Modèle additif :** Qui est sous la forme suivante :

$$Y_t = T_t + S_t + C_t + \epsilon_t ; t \in \mathbb{Z}$$

**Modèle multiplicatif :** Qui est sous la forme suivante :

$$Y_t = T_t * S_t * C_t * \epsilon_t ; t \in \mathbb{Z}$$

**Modèle mixte :** Il s'agit là de modèle, où addition et multiplication sont utilisés. - toutes les autres combinaisons sont également possibles.

## 2.5.4 Modélisation des séries chronologique

L'objectif de la modélisation est de construire des modèles permettant de décrire le comportement d'une chronique, et de ce fait résoudre les problèmes liés à la prévision.

### Décomposition de wold (1938) :

Le théorème de wold(1938) est un théorème fondamental dans la modélisation des séries chronologiques stationnaires, ou qui peuvent être stationnarisées par des transformations adéquates.

### Théorème :

Tous processus stochastique du second ordre  $X_t \in \mathbb{Z}$ , possède une décomposition unique donnée par :

$$X_t = U_t + V_t, \text{ tel que}$$

- Ces deux processus  $(U_t, V_t)$  sont orthogonaux de plus  $U_t$  est purement déterminable et  $V_t$  est purement interminable (aléatoire).
- Le processus  $V_t \in \mathbb{Z}$  peut être représenté sous forme d'une combinaison linéaire infinie du présent et du passé du processus bruit blanc  $\epsilon_t \in \mathbb{Z}$

Cette expression devait être convergente en moyenne quadratique cela veut dire que  $Var(v_t)$  doit être finie c'est-à-dire :

$$V_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j} \text{ avec } a_0 = 1 \ a_j \in \mathbb{R}, j \geq 1$$

$$Var(v_t) < \infty \text{ i.e } \sum_{j=0}^{\infty} |a_j| < \infty, a_0 = 1$$

### Modèle autorégressif moyenne mobile ARMA(p,q) :

Dans le but de trouver une méthode de prévision celle qui exploite toute l'information contenue dans la série, on essaie de modéliser la série chronologique, en utilisant une classe de modèle probabiliste. Ceci nous amène à réintroduire le processus stochastiques, en nous limitant à une classe de processus aléatoire stationnaires appelés les processus ARMA (Auto Regressive Moving Average). Les modèles ARMA résultent de la combinaison de processus autorégressifs (partie AR) et moyenne mobile (partie MA). Ils ont été popularisés par BOX and JENKINS (1970).

L'objectif est de modéliser une série temporelle en fonction de ses valeurs passées, mais aussi en fonction des valeurs présentes et passées d'un bruit blanc.

### Modèle autorégressif AR(p) :

#### Définition :

Le processus stationnaire  $X_t \in \mathbb{Z}$  satisfait une représentation AR d'ordre p, noté AR(p) , s'il est solution de l'équation aux différences stochastique suivante :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t; \text{ ou encore}$$

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

## Chapitre 2. Processus stochastique et séries chronologiques :

---

En utilisant l'opérateur de retard  $B$ , l'équation (1) devient :

$$X_t - \phi_1 B X_t - \phi_2 B^2 X_t - \dots - \phi_p B^p X_t = \epsilon_t$$

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

$$\Phi(B) X_t = \epsilon_t$$

$$\text{Où } \Phi(B) = - \sum_{j=0}^p \phi_j B^j; \text{ ou } \phi_0 = -1; \phi_j \in \mathbb{R}; \forall j < p; \phi_p \in \mathbb{R}^*$$

Où  $\Phi(B)$  représente le polynôme de retard et  $\epsilon_t$  est un bruit blanc de moyenne nulle et de variance  $\sigma_\epsilon^2$  (polynôme auto régressif)

$$\text{AR}(1) : X_t = \phi_1 X_{t-1} + \epsilon_t$$

$$\text{AR}(2) : X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$$

⋮

$$\text{AR}(p) : X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t$$

**Remarque :** Le modèle autorégressif d'ordre (p) explique la valeur de la chronique à l'instant t comme une combinaison linéaire de p observations antérieures. Il apparaît aussi comme une régression multiple où l'on explique les valeurs de la série Chronologique aux instants t-1, t-2, ..., t-p, c'est pour cela que nous l'appelons autorégressif d'ordre (p) (p représente le plus grand retard utilisé dans l'autorégression).

### Notion de causalité (stationnarité) :

#### Définition :

Un modèle de série chronologique (linéaire ou non linéaire) de la forme :

$$X_t = g(X_{t-1} X_{t-2} \dots X_{t-p}; \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q})$$

Où  $\epsilon_t$  est un bruit blanc, est dit « causal », si et seulement si, on peut exprimer le processus stochastique  $X_t$  sous forme d'une combinaison linéaire (finie ou infinie) convergente, en moyenne quadratique, du présent et du passé d'un bruit blanc  $\epsilon_t$ .

**N.B** La causalité n'est pas une propriété du processus  $X_t, t \in \mathbb{Z}$  à lui seul, mais plutôt de la relation avec  $\epsilon_t, t \in \mathbb{Z}$ . Par suite, si le modèle est causal, alors le processus  $X_t, t \in \mathbb{Z}$  est stationnaire, puisque ce dernier s'écrit en fonction de bruit blanc stationnaire.

#### Théorème :

Le modèle autorégressif  $X_t = \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t$  est causal si et seulement si : les racines de la fonction caractéristique  $\Phi(B) = 0$  sont en valeur absolue supérieures à 1, c'est-à-dire  $|Z| > 1$

### Exemple :

$$\text{AR}(1) : X_t = \phi_1 X_{t-1} + \epsilon_t \Leftrightarrow X_t = \phi_1 B X_t + \epsilon_t \Leftrightarrow (1 - \phi_1 B) X_t = \epsilon_t$$
$$1 - \phi_1 Z = 0 \Rightarrow Z = \frac{1}{\phi_1}, |Z| > 1 \Rightarrow \phi_1 < 1$$

### Notion d'inversibilité :

#### Définition :

Un modèle de série chronologique (linéaire ou non linéaire) de la forme :

$$X_t = g(X_{t-1} X_{t-2} \cdots X_{t-p}; \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-q})$$

Où  $\epsilon_t$  est un bruit blanc, est dit « *inversible* », si est seulement si, on peut exprimer le processus  $\epsilon_t$  sous forme d'une combinaison linéaire (finie ou infinie) convergente, en moyenne quadratique, du présent et du passé du processus stochastique  $X_t$ .

• D'après la définition d'inversibilité, un modèle autorégressif d'ordre fini (AR(p)) est toujours.

#### Fonction d'autocorrélation :

Si les racines de  $\Phi(B)$  sont situées à l'extérieur du disque unité (pour avoir un processus causal). Considérons la relation :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t; \forall t \in \mathbb{Z}$$

avec :

- $E(X_t) = 0$  (processus centré),
- $E(X_{t-h} \epsilon_t) = 0$  (comme  $\epsilon_t$  est indépendant de  $\epsilon_{t-1} \epsilon_{t-2} \epsilon_{t-3} \cdots$  alors  $\epsilon_t$  est indépendant du passé constitué par les variables  $X_{t-1}, X_{t-2}, \cdots, X_{t-h}$  pour  $h \geq 0$ ).

#### Equation de Yule-Walker :

Multiplions l'équation (1) par  $X_{t-h}$  et prenons l'espérance des deux cotés, on obtient :

$$E(X_{t-h}) = \phi_1 E(X_{t-1} X_{t-h}) + \phi_2 E(X_{t-2} X_{t-h}) + \cdots + \phi_p E(X_{t-p} X_{t-h}) + E(\epsilon_t X_{t-h})$$

- pour  $h = 0$  on obtient :

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \cdots + \phi_p \gamma(p) + \sigma_\epsilon^2$$

Pour  $h > 0$  on obtient :

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) + \cdots + \phi_p \gamma(h-p) \cdots \cdots (2)$$

En divisons (2) par  $\gamma(0)$  on obtient :

## Chapitre 2. Processus stochastique et séries chronologiques :

$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2) + \dots + \phi_p\rho(h-p) \dots\dots\dots (3)$$

En développant la relation (3) de proche en proche pour  $h = 1, 2, \dots, p$  on obtient le système suivant dit de Yule-Walker :

$$\begin{cases} \rho(1) = \phi_1 + \phi_2\rho(1) + \dots + \phi_p\rho(p-1) \\ \rho(2) = \phi_1\rho(1) + \phi_2 + \dots + \phi_p\rho(p-2) \\ \vdots \\ \rho(p) = \phi_1\rho(1) + \phi_2\rho(2) + \dots + \phi_p \end{cases}$$

D'où l'écriture matricielle suivante :

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(p-2) \\ \vdots & & & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \phi(p) \end{bmatrix}$$

Donc estimer les paramètres du modèle AR (p) revient à résoudre le système linéaire (ou matricielle) des p équations de Yule-Walker à p inconnus  $\phi_1, \phi_2, \dots, \phi_p$ , et les valeurs estimées de  $\rho(h)$  sont  $\rho(\hat{h})$ .

Sous cette forme, ces équations permettent d'obtenir par inversion, les fonctions  $\phi_i$  en fonction de  $\phi_1, \phi_2, \dots, \phi_p$ . En pratique  $\rho_i$  sera remplacée par son estimation :  $\rho(\hat{h}) = \frac{\gamma(\hat{h})}{\gamma(0)}$  tel que  $\rho = \Re\Phi \Leftrightarrow \Phi = \Re^{-1}\rho$

### Remarque :

- Le corrélogramme d'un modèle AR est un corrélogramme dont les valeurs absolues diminuent, jusqu'à qu'elles deviennent nulles.
- Il n'est pas toujours facile d'identifier un modèle AR par sa fonction d'autocorrélation, sauf dans le cas AR(1), c'est la raison pour laquelle nous avons recours aux autocorrélation partielles.

### Fonction d'autocorrélation partielle :

On appelle autocorrélation partielle du retard h le coefficient de corrélation entre  $X_t$  et  $X_{t-h}$  dans lequel l'influence de  $X_{t-1}X_{t-1+h}$  est éliminée, on le note ici  $\pi(h)$  et La fonction d'autocorrélation partielle de retard h est définie par :

$$\pi(h) = \phi_{hh} = \frac{\varrho_1}{\varrho_2}$$

avec

$$\varrho_1 = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(2) \\ \vdots & & & \ddots & \vdots \\ \rho(h-1) & \rho(h-2) & \dots & \dots & \rho(h) \end{pmatrix} \text{ et } \varrho_2 = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(h-2) \\ \vdots & & & \ddots & \vdots \\ \rho(h-1) & \rho(h-2) & \dots & \dots & 1 \end{pmatrix}$$

et  $|\underline{\rho}_h| \neq 0$

## Chapitre 2. Processus stochastique et séries chronologiques :

---

La fonction d'autocorrélation partielle est construite en reportant pour chaque valeur positive de  $h$ , la valeur du coefficient d'autocorrélation partielle  $\pi(h)$ . Elle est souvent utilisée pour déterminer l'ordre d'autocorrélation d'un modèle autorégressif.

- La fonction d'autocorrélation partielle  $AR(p)$  d'un satisfait à :

$$\phi_{11} = \rho(1) \text{ et } \phi_{hh} = 0, \forall h > 0, \forall h > p$$

on dit que cette fonction est «tronquée» au delà de l'ordre  $p$ . On peut donc lire l'ordre  $p$  d'un modèle autorégressif sur le corrélogramme  $\phi$  des l'autocorrélations partielles, ce dernier s'annule à l'ordre  $p+1$ . Ainsi, si la fonction d'autocorrélation partielle d'une série est calculée et si elle paraît tronquée, on peut modéliser la série par un modèle autorégressif.

- On ne manquera pas de noter que la résolution matricielle est très lourde. Néanmoins, la méthode récursive de Durbin peut fournir des autocorrélations partielles  $\phi_h$ , ainsi que les autres coefficients très rapidement, à partir de coefficient de corrélation partielle  $\pi_h$ , dont la définition suivante :

### Définition :

On appelle algorithme de Durbin (1960), le système des équations de récurrence suivant :

$$\left\{ \begin{array}{ll} \pi(1) = \phi_{11} = \rho(1) & \textit{initialisation} \\ \pi(h) = \phi_{hh} = \frac{\rho(h) - \sum_{j=1}^{h-1} \phi_{h-1,j} \rho(h-j)}{1 - \sum_{j=1}^{h-1} \phi_{h-1,j} \rho(j)} & \text{pour } h=2,3,\dots,k \\ \phi_{hj} = \phi_{h-1,j} - \phi_{hh} \phi_{h-1,h-j} & \text{pour } h=2,3,\dots,k ; j=1,2,\dots,h-1 \end{array} \right.$$

On obtient ainsi, une matrice triangulaire celle de Yule- Walker.

### Cas particulier AR(1) : (processus de Markov)

Un processus autorégressive d'ordre 1 est défini par :

$$X_t = \phi X_{t-1} + \epsilon_t \forall t \in \mathbb{Z}$$

Un modèle AR(1) est toujours inversible. Il est causal, si et seulement si :  $|\phi| < 1$

-La moyenne :  $E(X_t) = \mu = 0$

-La variance :  $Var(X_t) = \frac{(\sigma_\epsilon)^2}{(1-\phi^2)}, |\phi| < 1$

-Fonction :  $\rho(h) = (\phi)^{|h|}, h \in \mathbb{Z}$

-Fonction d'autocorrélation partielle :  $\phi_1 = \rho(1)$  et  $\phi_{hh} = 0, h \geq 2$

### Modèle moyenne mobile MA(q) (Moving Average)

#### Définition :

## Chapitre 2. Processus stochastique et séries chronologiques :

---

Un processus stationnaire  $\{X_t, t \in Z\}$  satisfait une représentation moyenne mobile d'ordre  $q$ ,

noté MA ( $q$ ), s'il est solution de l'équation aux différences stockastique suivante :

$$X_t = \epsilon_t - \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

En introduisant le polynôme de retard on obtient :

$$X_t = \Theta(B)\epsilon_t \text{ ou } \Theta(B) = - \sum_{j=1}^q \theta_j B^j ; \theta_0 = -1; \theta_j \in R, \forall j \in R, \theta_q \in R^*$$

$\Theta(B)$  : Le polynôme moyenne mobile de degré  $q$  ( $q$  représente le plus grand retard utilisé).

### Remarque :

- Le modèle moyenne mobile d'ordre  $q$ , MA ( $q$ ) explique la valeur de la série à l'instant  $t$  par une moyenne pondérée d'aléas  $\epsilon_t$  jusqu'à la  $q^{eme}$  période qui sont supposés générés par un processus de type bruit blanc.
- Nous pouvons interpréter le modèle MA comme étant une représentation d'une série chronologique fluctuant autour de sa moyenne de manière aléatoire, d'où le terme de moyenne mobile.

### - Condition de causalité :

D'après la définition de causalité, un modèle moyenne mobile d'ordre fini est toujours causal, car c'est une combinaison linéaire finie de processus stationnaire  $\{\epsilon_t, t \in Z\}$

### - Condition d'inversibilité :

Le modèle moyenne mobile MA ( $q$ ) :  $X_t = \epsilon_t - \sum_{j=1}^q \theta_j \epsilon_{t-j}$  est inversible si et seulement si les racines de sont en valeur absolues supérieurs à 1, c'est-à-dire  $|Z| > 1$  .

### - Fonction d'autocovariance :

-La moyenne :  $E(X_t) = E(\epsilon_t) - \theta_1 E(\epsilon_{t-1}) - \dots - \theta_q E(\epsilon_{t-q}) = 0$

car  $E(\epsilon_t) = 0$  donc  $E(X_t) = 0$  .

-La variance :  $Var(X_t) = \gamma(0) = (\sigma_\epsilon)^2 \sum_{i=0}^q (\theta_i)^2; \theta_0 = 1$ .

-La covariance :  $\gamma(h) = \begin{cases} (1 + (\theta_1)^2 + (\theta_2)^2 + \dots + (\theta_q)^2)(\sigma_\epsilon)^2 & \text{si } h=0 \\ (-\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h})(\sigma_\epsilon)^2 & \text{si } 1 \leq h \leq q \\ 0 & \text{si } h \geq q+1 \end{cases}$

### - Fonction d'autocorrélation :

La fonction d'autocorrélation d'un processus MA (q) est donné par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1 & \text{si } h=0 \\ \frac{(-\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q\theta_{q-h})}{(1+\theta_1^2 + \theta_2^2 + \dots + \theta_q^2)} & \text{si } 1 \leq h \leq q \\ 0 & \text{si } h \geq q+1 \end{cases}$$

### Remarque :

La fonction d'autocorrélation s'annule à partir d'un décalage supérieur à q, on dit qu'elle est tronquée au-delà du retard q. Donc on peut identifier un MA (q) à partir du corrélogramme qui s'annule à partir d'un retard supérieur à q.

### - Cas particulier MA (1) :

Un processus moyenne mobile d'ordre 1 est défini par :

$$X_t = \epsilon_t - \theta\epsilon_{t-1}; \forall t \in Z$$

- Un modèle MA (1) est toujours causal.
- Il est inversible, si et seulement si :  $|\theta| < 1$
- La moyenne :  $E(X_t) = E(\epsilon_t) - \theta E(\epsilon_{t-1}) = 0$
- La variance :

$$Var(X_t) = \gamma(0) = Var(\epsilon_t - \theta\epsilon_{t-1}) = Var(\epsilon_t) + (-\theta^2)Var(\epsilon_{t-1}) + 2(-\theta)Cov(\epsilon_t, \epsilon_{t-1}) = \sigma^2 + \sigma^2\theta^2$$

$$Var(X_t) = \sigma^2(\theta^2 + 1)$$

### - Fonction d'autocovariance :

$$\begin{aligned} Cov(X_t, X_{t-1}) &= Cov(\epsilon_t - \theta\epsilon_{t-1}, \epsilon_{t-1} - \theta\epsilon_{t-2}) \\ &= Cov(\epsilon_t, \epsilon_{t-1}) - \theta Cov(\epsilon_t, \epsilon_{t-2}) - \theta Cov(\epsilon_{t-1}, \epsilon_{t-1}) + \theta^2 Cov(\epsilon_{t-1}, \epsilon_{t-2}) \\ &\implies Cov(X_t, X_{t-1}) = -\theta\sigma^2 \\ Cov(\epsilon_t, \epsilon_{t-2}) &= Cov(\epsilon_t - \theta\epsilon_{t-1}, \epsilon_{t-2} - \theta\epsilon_{t-3}) = 0 \end{aligned}$$

### - Fonction d'autocorrélation :

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{-\theta\sigma^2}{\sigma^2(\theta^2+1)} = \frac{-\theta}{(\theta^2+1)}$$

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

L'autocorrélation de retard 1 est différente de 0.

Donc la fonction d'autocorrélation est tronquée au delà du retard 1.

### - Fonction d'autocorrélation partielle :

Afin de calculer les autocorrélations partielles d'un modèle MA, nous utilisons l'algorithme de Durbin.

Contrairement au modèle AR(p) la fonction d'autocorrélation partielle d'un modèle MA n'a pas d'expression explicite.

Les autocorrélations partielles sont donc données récursivement par l'algorithme de Durbin.

$$\begin{cases} \phi_{11} = \rho_1 = \frac{-\theta}{1+\theta^2} \\ \phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{-\rho_1^2}{1 - \rho_1^2} \\ \phi_{33} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_2 - \phi_{22}\rho_1} = \frac{-\phi_{22}\rho_1}{1 - \phi_{21}\rho_1} \end{cases}$$

Comme nous avons :

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \phi_{11}(1 - \phi_{22}) = \rho_1\left(1 + \frac{\rho_1^2}{1 - \rho_1^2}\right)$$

$$\text{Donc : } \phi_{33} = \frac{\rho_1^3}{1 - 2\rho_1^2}$$

Nous pourrions par la suite poursuivre les calculs pour déterminer les autocorrélations partielles d'ordre supérieur, en exprimant les autocorrélations en fonction de  $\theta$  pour obtenir une suite récurrente on a :

$$\phi_{22} = \frac{-\rho_1^2}{1 - \rho_1^2} \text{ et } \rho_1 = \frac{-\theta}{1 + \theta^2} \implies \phi_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4}$$

$$\text{On remarque que : } (1 + \theta^2 + \theta^4) = \frac{1 - \theta^6}{1 - \theta^2}$$

$$\text{Donc } \phi_{22} = \frac{-\theta^2(1 - \theta^2)}{1 - \theta^6}$$

$$\text{- En raisonnant de la même manière pour } \phi_{33} \text{ on trouve : } \phi_{33} = \frac{-\theta^3(1 - \theta^2)}{1 - \theta^8}$$

- La formule de récurrence pour les autocorrélations partielles d'un modèle MA(q) est alors donnée par :

$$\phi_{kk} = \frac{-\theta^k(1 - \theta^2)}{1 - \theta^{2(k+1)}}$$

### Modèle mixte ARMA(p,q) :

#### Définition :

Le processus stationnaire  $\{X_t, t \in Z\}$  satisfait une représentation ARMA d'ordre p et q, noté ARMA(p,q), s'il est solution de l'équation aux différences stochastiques suivante :

$$X_t - \sum_{j=1}^p \phi_j X_{t-j} = \epsilon_t \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

$$\text{Ou encore : } \Phi(B)X_t = \Theta(B)\epsilon_t$$

avec :  $\phi_p \neq 0, \theta_q \neq 0$

- Les polynômes  $\Phi$  et  $\Theta$  d'ordre respectif p et q n'ont pas de racines communes.

-  $\epsilon_t$  bruit blanc de variance  $\delta_\epsilon^2 \neq 0$

Donc on peut noter AR(p) par ARMA(p,0), et MA(q) par ARMA(0,q).

### - Condition de causalité :

## Chapitre 2. Processus stochastique et séries chronologiques :

---

Le modèle ARMA( $p, q$ ) défini par  $\Phi(B)X_t = \Theta(B)\epsilon_t$  est causal, si et seulement si les racines de  $\Phi$  sont de module strictement supérieure à 1, c'est-à-dire  $|z| > 1$  ;  
ou encore  $\Phi(z) \neq 0, \forall z \in Z, |z| \leq 1$ .

### - Condition d'inversibilité :

Le modèle ARMA( $p, q$ ) défini par  $\Phi(B)X_t = \Theta(B)\epsilon_t$  est inversible, si et seulement si :  
Les racine de  $\Theta$  sont on valeur absolues supérieurs à 1, c'est-à-dire  $|z| > 1$ .

### Remarque :

En réalité, la stationnarité traduit la condition sous laquelle le processus  $\{X_t\}$  admet une écriture du type moyenne mobile.

Si l'on cherche la condition permettant une écriture autorégressive du processus, alors on parlera de condition d'inversibilité.

Evidement tout processus autorégressif est inversible, mais ce n'est pas forcément le cas des processus moyenne mobile.

### - Fonction d'autocovariance :

Si les conditions d'inversibilité et de causalité sont respectées, la fonction d'autocovariance,  $\gamma(h)$  d'un processus satisfait une représentation ARMA est donnée par :

$$\gamma(h) = \begin{cases} \phi_1\gamma(1) + \dots + \phi_p\gamma(p) + \sigma^2 - \theta_1\gamma(-1) - \dots - \theta_q\gamma(-q) & \text{si } h=0 \\ \phi_1\gamma(h-1) + \dots + \phi_p\gamma(h-p) + \sigma^2(-\theta_h\theta_{h+1}\theta_1 + \dots\theta_q\theta_{q-h}) & \text{si } 1 \leq h \leq q \\ \phi_1\gamma(h-1) + \dots + \phi_p\gamma(h-p) & \text{si } h \geq q+1 \end{cases}$$

$$\text{Avec : } E(\epsilon_t X_{t-h}) = \begin{cases} \sigma_t^2 & \text{pour } h=0 \\ 0 & \text{pour } h \neq 0 \end{cases}$$

### - Fonction d'autocorrélation :

Pour calculer les autocorrélation d'un modèle ARMA, on procède comme dans le cas des modèles AR. A partir de l'équation :

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

En multipliant les deux membres par  $X_{t-h}$  et en introduisant l'espérance, on obtient l'équation suivante :

$$\frac{1}{\gamma(0)} [E(X_t X_{t-h}) - \phi_1 E(X_{t-1} X_{t-h}) - \dots - \phi_p E(X_{t-p} X_{t-h})] = \frac{1}{\gamma(0)} [E(\epsilon_t X_{t-h}) - \phi_1 E(\epsilon_{t-1} X_{t-h}) - \dots - \phi_q E(\epsilon_{t-q} X_{t-h})]$$

$\epsilon_t$  est un bruit blanc, et par conséquent non corrélé avec le passé du processus  $X_t$ ,

$$\text{donc : } E(\epsilon_t X_{t-h}) = 0$$

$$E(\epsilon_{t-q} X_{t-h}) = 0 \text{ si } h - q > 0 \Rightarrow h > q$$

On obtient :

$$\rho(h) = \phi_1\rho(h-1) - \dots - \phi_p\rho(h-p) = 0, \forall h > q$$

$$\rho(h) = \sum_{i=1}^p \phi_i\rho(h-i), \forall h > q$$

Donc :

$$\rho(h) = \begin{cases} \phi_1\rho_{h-1} + \dots + \phi_p\rho(h-p) + \frac{\sigma^2}{\gamma(0)}(-\theta_h + \dots + \theta_q\theta_{q-h}) & \text{si } 1 \leq h \leq q \\ \phi_1\rho_{h-1} + \dots + \phi_p\rho(h-p) & \text{si } h \geq q+1 \end{cases}$$

### Remarque :

La fonction d'autocorrelation décroît exponentiellement, ou sinusoidalement après un grand nombre de retard.

### - Fonction d'autocorrélation partielle :

La forme de cette fonction est la résultante d'un processus mettant en présence une partie autorégressive et une partie moyenne mobile, elle dépend de chaque partie et de la valeur des paramètres  $(p, q)$ .

### Remarque :

La fonction d'autocorrélation partielle possède  $p$  premières valeurs significatives, puis elle décroît exponentiellement vers zéro ou sous forme d'un sinusöide amortie.

## 2.5.5 Séries non stationnaires :

Les séries économetriques présentent souvent des réalisations non stationnaires. Elle admettent une moyenne et une variance qui varient au cours de temps, or les moyennes statistiques ne s'applique pas à des séries dont la stationnarité n'est pas vérifiée.

La non stationnarité peut être détectée graphiquement par la présence d'une tendance, d'une saisonnalité, ou d'une modification dans la structure de la série ; pour ce fait, on essaie généralement d'analyser les graphes de la fonction d'autocorrélation (fonction d'autocorrélation lentement décroissante).

Mais cela reste insuffisant ; et pour avoir une certitude, il est nécessaire d'appliquer les tests pour confirmer ou infirmer les déductions tirées par les graphes de la série et par les corrélogrammes. La plupart des résultats et des méthodes utilisées dans l'analyse des séries temporelles repose sur la notion de stationnarité du second ordre, ce qui nous mène à appliquer à la chronique non stationnaire certaines transformations (différence ordinaire, différence saisonnière, la formule de BOX-COX...).

### Analyse de la tendance :

La non stationnarité d'un processus aléatoire dépend du moment du premier ordre (espérance mathématique) et celui du second ordre (variance et covariance du processus),

## Chapitre 2. Processus stochastique et séries chronologiques :

---

qui doivent être indépendants du temps. Dans le cas contraire, le processus aléatoire est dit non stationnaire. Suite aux travaux de NELSON et PLOSSER (1982), on analyse les cas de non stationnarité à partir de deux types de processus :

- Les processus de type TS qui représentent la non stationnarité de type déterministe.
- Les processus de type DS qui représentent la non stationnarité de type aléatoire (stochastique).

### - Les processus TS (Trend stationary) :

#### Définition :

Les processus TS sont des processus non stationnaires de type déterministe qui s'écrivent sous la forme :  $X_t = f_t + \epsilon_t$

Où :

$f_t$  : fonction polynomial qui dépend du temps qui peut être linéaire ou non linéaire.

$\epsilon_t$  : processus de type ARMA (*stationnaire*).

Le processus TS le plus simple est représenté par une fonction polynomiale de degré 1. Le processus s'écrit :

$$X_t = a_0 + a_1t + \epsilon_t$$

si  $\epsilon_t$  est un bruit blanc, les caractéristiques de ce processus sont alors :

$$E(X_t) = a_0 + a_1t + E(\epsilon_t) = a_0 + a_1t$$

$$Var(X_t) = 0 + Var(\epsilon_t) = \sigma_\epsilon^2$$

$$Cov(X_t, X'_t) = 0 \text{ pour } t \neq t'$$

Nous constatons que le processus TS est caractérisé par une espérance mathématique à tendance déterministe, une variance constante au cours du temps et par des covariances nulles, dans un tel modèle la réalisation des prévisions n'est pas une tâche facile.

#### Remarque :

La non stationnarité de ce processus est due au fait que sa moyenne dépend du temps.

\*\* Les étapes de la méthode utilisée pour le rendre stationnaire sont :

- Estimer les coefficients  $a_0$  et  $a_1$  par la méthode des moindres carrés ordinaires (*MCO*).
- Retrancher de  $X_t$ , la valeur estimée  $\hat{a}_0 + \hat{a}_1t$ .

### - Les processus DS (Differency Stationary)

Les processus DS sont des processus non stationnaires de type aléatoire qui s'écrivent sous la forme :  $X_t = \rho X_{t-1} + \beta + \epsilon_t$

Où :

$\beta$  : une constante réelle.

$\epsilon_t$  : un processus stationnaire de type ARMA ou un bruit blanc.

## Chapitre 2. Processus stochastique et séries chronologiques :

---

\*\* On peut le rendre stationnaire on utilisant un filtre aux différences à l'ordre  $d$ ,  $d \geq 1$  :

$$(1 - B)^d X_t = \beta + \epsilon_t$$

Où :

B : l'opérateur de décalage.

$d$  : l'ordre de filtre aux différences.

Par récurrence, on obtient :

$$X_t = \rho^t X_{t-\tau} + \beta \sum_{j=0}^{\tau-1} \rho^j + \sum_{j=0}^{\tau-1} \rho^j \epsilon_{t-j}$$

Nous supposons que  $|\rho| = 1$  et que  $\tau = t$  nous aurons donc  $X_t = X_0 + \beta t \sum_{j=1}^t \epsilon_j$

Où  $X_0$  désigne le premier terme de la série  $X_t$ .

### Remarque

Dans la pratique, le processus DS le plus utilisé à un filtre aux différences d'ordre 1 ( $d = 1$ ), appelé processus de marche aléatoire, qui s'écrit sous la forme :

$$(1 - B)X_t = \beta + \epsilon_t \Leftrightarrow X_t = X_{t-1} + \beta + \epsilon_t$$

Où  $\epsilon_t$  : un processus bruit blanc de variance  $\sigma_\epsilon^2$  et  $\rho = 1$ .

\*\* Passons maintenant à l'étude des caractéristiques de ce processus :

- **Processus DS sans dérive :** (Modèle de marche aléatoire sans dérive)

Dans le cas de processus DS sans dérive,  $\beta$  est égale à zéro ( $\beta = 0$ ). Il s'écrit sous la forme :  $X_t = X_{t-1} + \epsilon_t$

Par conséquent :

$$E(X_t) = X_0$$

$$Var(X_t) = t\sigma_\epsilon^2$$

$$Cov(X_t, X_s) = Min(t, s)\sigma_\epsilon^2 \quad \forall t \neq s$$

- La non stationnarité de ce processus est due au fait que sa variance dépend du temps.

### Test de racine unitaire :

Ces tests permettent d'examiner l'existence d'une racine unitaire dans les processus qui ont générés la série chronologique en permettant de déterminer le type de la non stationnarité qui peut être déterministe (de type TS) que l'on peut le rendre stationnaire par régression sur une tendance déterministe, ou stochastique (de type DS) que l'on peut les rendre stationnaires en utilisant un filtre au différence à l'ordre  $d$ .

Il existe un grand nombre de tests de racine unitaire, les tests les plus utilisés sont ceux de Dickey-Fuller (1979).

### - Test de Dickey-Fuller simple (DF) :

Les modèles suivant de base à la construction de ces tests sont au nombre de trois, et dans ce qui suit est un processus bruit blanc (supposition).

[1] : Modèle sans constante ni tendance déterministe

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

[2] : Modèle avec constante et sans tendance déterministe

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

[3] : Modèle avec constante et avec tendance déterministe

$$X_t = c + bt + \rho X_{t-1} + \epsilon_t$$

On teste l'hypothèse nulle  $H_0$  de présence de racine unitaire ( $X_t$  est intégré d'ordre 1,  $I(1)$ , donc non stationnaire) contre l'hypothèse alternative  $H_1$  en l'absence de racine unitaire ( $X_t$  est intégré d'ordre 0, c'est-à-dire que  $X_t$  est stationnaire).

L'hypothèse du test comme suit :  $\begin{cases} H_0 : \rho = 1 \\ H_1 : |\rho| < 1 \end{cases}$

En s'inspirant du modèle [1]

$$X_t = \rho X_{t-1} + \epsilon_t \dots (1)$$

Retranchons  $X_{t-1}$  de chaque coté de l'équation (1)

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

$$\Delta X_t = (\rho - 1)X_{t-1} + \epsilon_t$$

En pratique et en posant  $\phi = (\rho - 1)$  on estime les modèles suivant :

Modèle [4] :

$$\Delta X_t = \phi X_{t-1} + \epsilon$$

Modèle [5] :

$$\Delta X_t = c + \phi X_{t-1} + \epsilon$$

Modèle [6] :

$$\Delta X_t = c + bt + \phi X_{t-1} + \epsilon$$

Ce qui revient à dire que le test de racine unitaire repose sur le test de l'hypothèse nulle  $\phi = 0$  (non stationnaire) contre l'hypothèse alternative  $\phi \neq 0$  (stationnaire), et donc le

système d'hypothèse devient :  $\begin{cases} H_0 : \phi = 0 \\ H_1 : |\phi| \neq 0 \end{cases}$

### Principe des Tests de Dickey-Fuller :

Sous l'hypothèse  $H_0$ , le processus  $X_t$  n'est pas stationnaire quelque soit le modèle retenu. Les règles habituelles de l'inférence statistique ne peuvent donc pas être appliquées pour tester cette hypothèse, en particulier la distribution de student du paramètre  $\rho$ .

Dickey et Fuller ont étudiés la distribution asymptotique de l'estimateur du paramètre  $\rho$  sous l'hypothèse  $H_0$  à l'aide des simulations de Monte-Carlo, ils ont tabulé les valeurs critiques pour des échantillons de tailles différentes.

Soit la t-statistique notée ( $t_{\hat{\phi}}$ ) tel que  $t_{\hat{\phi}} = \frac{\hat{\phi}-1}{\sigma_{\hat{\phi}}^2}$ , on compare alors la  $t_{\hat{\phi}}$  avec la valeur critique  $t_{tabulé}$  :

-Si  $t_{\hat{\phi}} \geq t_{tabulé}$  alors on accepte l'hypothèse  $H_0$  il existe une racine unitaire.

-Sinon on rejette l'hypothèse  $H_0$  .

### Remarque :

- Ces tests révèlent l'existence d'une racine unitaire mais restent insuffisants pour discriminer entre les processus TS et DS, c'est ainsi qu'on adopte un algorithme en trois étapes.

- On dit que la tendance est significativement différente de 0, ssi  $t_{\hat{\phi}} \geq t_{tabulé}$  alors la tendance existe, sinon elle est dite non significativement différente de 0.

- On dit que la constante est non significativement différente de 0, ssi t-statistique < valeur critique, sinon elle est dite significativement différente de 0.

### Enoncé de l'algorithme :

#### Etape 1 :

Dans cette étape on estime le modèle [3] et on teste la signification de la tendance déterministe.

- Si la tendance n'est pas significativement différente de zéro (t-statistique de la tendance est inférieure aux valeurs critiques de la tendance tabulées par Dickey-Fuller) .

On passe à l'étape 2.

- Si la tendance est significativement différente de zéro, on teste l'hypothèse nulle de racine unitaire en comparant  $t_{\hat{\phi}}$  aux valeurs tabulées par Dickey-Fuller.

Si on accepte  $H_0$ ,  $X_t$  est non stationnaire de type DS.

Si on rejette  $H_0$ ,  $X_t$  est de type TS.

#### Etape 2 :

Cette étape n'est applicable que si la tendance n'est pas significative, donc on estime le modèle[2] et on teste la signification de la constante.

- Si la constante n'est pas significative on passe à l'étape 3.

- Si la constante est significative, on teste l'hypothèse de racine unitaire.

Si  $H_0$  est acceptée, donc  $X_t$  est non stationnaire de type DS.

Si  $H_0$  n'est acceptée,  $X_t$  est stationnaire.

#### Etape 3 :

Cette étape n'est applicable que si la constante dans le modèle n'est pas significative. On estime dans ce cas le modèle [1] et on teste l'hypothèse nulle de racine unitaire.

Si  $H_0$  est acceptée,  $X_t$  est non stationnaire de type DS.

Si  $H_0$  est rejetée,  $X_t$  est stationnaire.

### - Test de Dickey-Fuller augmenté :

#### Transformation de modèle de base :

Dans les modèles précédents, utilisés par les tests de Dickey-Fuller simple, le processus  $\epsilon_t$  est par hypothèse un bruit blanc. Or il n'y a aucune raison pour que à priori, l'erreur soit non corrélée ; on appelle tests de Dickey-Fuller augmentés (ADF, 1981) la prise en compte de cette hypothèse. Les tests ADF s'effectuent exactement comme les tests DF sur les modèles suivants :

Modèle [4] :

$$\Delta X_t = \phi X_{t-1} + \sum_{j=1}^p \phi_j \Delta X_{t-j} + \epsilon_t$$

Modèle [5] :

$$\Delta X_t = \phi X_{t-1} + \sum_{j=1}^p \phi_j \Delta X_{t-j} + c + \epsilon$$

Modèle [6] :

$$\Delta X_t = \phi X_{t-1} + \sum_{j=1}^p \phi_j \Delta X_{t-j} + c + bt + \epsilon$$

#### Remarque :

- On pratique nous allons utiliser les tests de ADF.
- Avant d'appliquer le test ADF il faut préciser l'ordre de décalage p en utilisant le critère d'Akaike.
- Les principaux logiciels d'analyse de séries temporelles calculent automatiquement les valeurs critiques à l'instar de EVIEWS 10.

#### Analyse de la saisonnalité :

Une série chronologique saisonnière est une série dont les données relatives à une même période (la période est plus courte qu'une année) de différentes années ont tendance à ce situer de façon analogue par rapport à la moyenne annuelle. Elle peut se relater à des observations trimestrielles et mensuelles aussi bien que d'heure en heure ou aux observations quotidiennes. Il est possible de détecter cette saisonnalité par un examen graphique de la série, qui se manifeste par la répétition d'un certain phénomène dans chaque période. Ou par un examen fait sur le corrélogramme de la série étudiée, qui laisse apparaître des pics très marqués aux retards 1, S, 2S,.... On en déduit une saisonnalité de périodicité S (S=3, 6, 12, ...).

### 2.5.6 Extension des modèles ARMA :

L'hypothèse de stationnarité, présente sous certaines conditions dans les modèles ARMA, n'est que rarement vérifiée pour les séries économiques ; Elles comportent également une tendance, une saisonnalité ou même une structure plus complexe. Par conséquent, l'intérêt des modèles ARMA semble assez limité.

#### Modèles ARIMA (p, d, q) :

Si on considère par exemple les différences premières (ou en général les différences d'ordre d) de telles séries l'hypothèse de stationnarité devient souvent vraisemblable.

Il est donc naturel de considérer la classe des processus dont la différence d'un certain ordre satisfait une représentation ARMA.

Si on note  $\Delta^d X_t$  la différence d'ordre d de  $X_t$ , c'est-à-dire :  $\Delta^d X_t = (1B)^d X_t$

On va s'intéresser aux processus satisfaisant :

$$\Phi(B)\Delta^d X_t = \Theta(B)\epsilon_t$$

#### Définition :

Un processus  $X_t$  est un modèle « autoregressif moyenne mobile intégré » d'ordre (p, d, q) ARIMA(p, d, q) s'il vérifie une équation de type :

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\epsilon_t, \text{ pour tout } t \geq 0 \dots\dots(1)$$

Où

$\Phi(B)$  : Polynôme autoregressif dont les racines sont de module supérieures à 1 .

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \text{ où } \phi_p \neq 0$$

$\Theta(B)$  : Polynôme moyenne mobile dont les racines sont de module supérieures à 1 .

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \text{ où } \theta_q \neq 0$$

Aucune des racines de  $\Phi(B)$  n'est égale à une racine de  $\Theta(B)$ .

Les coefficients réels  $\phi_i, i=1, \dots, p$  et  $\theta_j, j=1, \dots, q$ , sont fixés et  $\{\epsilon_t\}$  est un bruit blanc.

#### Remarque :

- La famille ARIMA désigne parfois la classe de tous les modèles, stationnaires et non stationnaires, en convenant que les ARIMA (p, 0, q) sont les ARMA(p, q).

- L'équation (1) peut s'écrire comme suite :

$$\Psi(L)X_t = \Theta_q(L)\epsilon_t, \forall t \in Z$$

$$\text{avec } \Psi(L) = \Phi_q(L)(1 - L)^d = (1 - \psi_1 L - \dots - \psi_{p+d} L^{p+d})$$

Donc, cette équation est la même que celle de la définition d'un modèle ARMA(p+d, q), tel que le polynôme  $\Psi(Z)$  admet 1 comme racine multiple d'ordre (dite racine unitaire).

## Chapitre 2. Processus stochastique et séries chronologiques :

---

### Modèles SARIMA (ARIMA saisonnier) :

Il est possible de trouver que certaines séries chronologique peuvent être caractérisées par une allure graphique périodique, pour cela, il est important de les analyser en tenant compte de l'effet saisonnier. Box et Jenkins (1970) ont proposés une classe de modèles particulière appelées : classe de modèles ARIMA saisonniers.

# Chapitre 3

## La méthodologie de BOX et JENKINS

### 3.1 Introduction

George Box et Gwilym Jenkins ont développé en 1976 une véritable méthodologie de recherche et d'étude systématique en fonction de l'étude des corrélogrammes. C'est une approche itérative qui consiste à identifier un modèle susceptible de représenter le phénomène étudié. Elle se réfère à deux types de modèles, autorégressifs et moyenne mobile ou à une combinaison des deux. Cependant, la modélisation d'une chronique nécessite au préalable une stationnarisation. le diagramme qui suit illustre la démarche générale de Box Jenkins :

### 3.2 Test des composantes saisonnières et tendancielle

#### 3.2.1 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, trimestriel...).

On considère :

- $n$  : le nombre d'années.
- $p$  : Le nombre d'observation dans l'année.
- $X_{ij}$  : La valeur de la série pour la  $i^{\text{ème}}$  année et la  $j^{\text{ème}}$  période.
- $\bar{X}_{..}$  : La moyenne générale.

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

- $var_A = \frac{p \sum_{i=1}^n (\bar{X}_i - \bar{X}_{..})^2}{n-1}$

- $var_P = \frac{n \sum_{j=1}^p (\bar{X}_j - \bar{X}_{..})^2}{p-1}$

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 générale

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

l'hypothèse est :

$$\begin{cases} H_0 : \ll \text{pas de saisonnalité} \gg \\ H_1 : \ll \text{il existe une saisonnalité} \gg \end{cases}$$

La valeur calculée  $F_0 = \frac{var_P}{var_R}$  ; que l'on compare à la valeur tabulée de Fisher au seuil de signification  $\alpha F_{v_1 v_2}^\alpha$

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

- si  $F_0 > F_{v_1 v_2}^\alpha$

on rejette  $H_0$  la série est saisonnière.

Le test de Fisher permet également de tester l'effet de la tendance.

Soient les hypothèses :

$$\begin{cases} H_0 : \ll \text{La série n'est pas affectée d'une tendance} \gg \\ H_1 : \ll \text{La série est affectée d'une tendance} \gg \end{cases}$$

On calcule :

$F_1 = \frac{var_A}{var_R}$  ; que l'on compare avec la valeur tabulée de Fisher  $F_{v_3 v_2}^\alpha$  à seuil  $\alpha$

Avec  $v_3 = (n - 1)$  ,  $v_2 = (p - 1)(n - 1)$  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, le plus connu est de Dickey-Fuller.

### 3.2.2 Test de Dickey-Fuller

Il permet de détecter non seulement l'existence d'une tendance mais aussi de quel type elle est pour la construction de ce test, Dickey et Fuller ont proposés trois modèles de base :

$(1 - \phi_1 B)X_t = \epsilon_t$  : modèle autorégressif d'ordre 1 ..... Modèle (1)

$(1 - \phi_1 B)(X_t - c) = \epsilon_t$  : modèle autorégressif d'ordre 1 avec constante ..... Modèle (2)

$(1 - \phi_1 B)(X_t - c - bt) = \epsilon_t$  : modèle autorégressif d'ordre 1 avec tendance ..... Modèle (3)

les hypothèses à tester sont :

$$\begin{cases} H_0 : \ll \phi_1 = 1 \gg \\ H_1 : \ll |\phi_1| < 1 \gg \end{cases}$$

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

- Si dans l'un des modèles la statistique relative à  $\phi_1$  est supérieure aux valeurs tabulées  $t_{0,05}$ , on déduit qu'il existe une racine unitaire.

On estime en premier le modèle (3) .

On commence par tester le significativité de la tendance :

- Si la tendance n'est pas significative alors on teste 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 marche aléatoire dit DS sans dérive, on différencie la série une fois et on recommence.
- 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 non stationnaire de type TS
- Sinon la série est stationnaire.

### 3.3 Identification

La phase d'identification est la plus importante et la plus difficile, elle consiste à déterminer le modèle adéquat dans la famille des modèles ARMA. elle est fondée sur l'étude des corrélogramme simple et partiel. Une fois la série stationnarisée, nous identifions les paramètres  $p$  et  $q$  de la manière suivante : -Si le corrélogramme simple n'a que ses  $q$  premiers termes différents de zéro et que les termes du corrélogramme partiel diminuent lentement, nous pouvons pronostiquer un MA( $q$ ) (moyenne mobile d'ordre  $q$ ). -Si le corrélogramme partiel n'a que ses  $p$  premiers termes différents de zéro et que les termes du corrélogramme simple diminuent lentement, cela caractérise un AR( $p$ ) (autorégressif d'ordre  $p$ ). -Si les fonctions d'autocorrelation simple et partiel ne paraissent pas tronquées, il s'agit alors d'un processus ARMA( $p, q$ ).

### 3.4 Estimation

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

### 3.5 Tests de 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.

### 3.6 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 :

$$\begin{cases} H_0 : \ll \text{le coefficient } \hat{B}_p = 0 \gg \\ H_1 : \ll \hat{B}_p \neq 0 \text{ est significative} \gg \end{cases}$$

statistique de Student représente le ratio du coefficient à estimer sur son écart type .

- si  $\frac{\hat{B}_p}{\hat{\sigma}_{\hat{B}_p}} > 1.96$  ( $\alpha = 5\%$ ) on accepte  $H_1$

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.

#### Critère d'informations

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

**Critère d'information 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.

$$AIC(p,q) = n \log \sigma_\epsilon^2 + 2(p + q)$$

ou  $n$  : est le nombre d'observation de la série

**Critère d'information Bayésien (BIC)** Il a les memes caractéristiques que l'AIC, néanmoins ; jugé plus intéressant dans la mesure où il pénalise les paramètres en nombre excessif plus fortement que l'AIC.

$$BIC(p,q) = n \log \hat{\sigma}_\epsilon^2 + (p+q) \log \left( \frac{\hat{\sigma}_\epsilon^2}{n} \right) + (p+q) \log(p+q)$$

**Critère de Schwartz 1979** Ce critère est défini comme suit :

$$SC(p,q) = n \log \hat{\sigma}_\epsilon^2 + (p+q) \log n$$

**Critère de Hannan-Quin 1979** Ce critère de Hannan-Quin s'écrit sous la forme suivante :

$$HQ(p,q) = \log \hat{\sigma}_\epsilon^2 + (p+q)c \log \left[ \frac{\log n}{n} \right]$$

$c$  est une constante à spécifier.

Les critères les plus utilisés sont AIC, BIC, SC. Le modèle choisi (optimal) est celui qui minimise ces critères. Ils existent d'autres critères appelés "critères de pouvoir prédictif" Utilisés pour vérifier les performances prévisionnelles.

**Principe de Parcimonie** Dans le cas où les critères d'informations AIC et BIC de deux ou plusieurs modèles retenus seraient très proches ou contradictoires, on fait intervenir ce principe puisque le but de la modélisation d'une chronique est de minimiser le nombre de paramètres à estimer.

### 3.6.1 Tests 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  $\epsilon_t, t \in \mathbf{Z}$

**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 :

$$\begin{cases} H_0 : \ll \rho_0 = 0, \rho_1 = 0, \dots, \rho_k = 0 \gg \\ H_1 : \ll \text{il existe au moins un } \rho_i \text{ significativement différent de zéro} \gg \end{cases}$$

Sa statistique est :

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

- $K$  : nombre de retards choisis
- $n$  : nombre d'observations

En l'absence d'autocorrélation la statistique  $Q$  obéit à une  $\chi^2(K - p - q - P - Q)$  degré de libertés L'hypothèse  $H_0$  est rejetée au seuil de 5% si  $Q$  est supérieure au quantile 0.95 de 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 (asymétrie) et de Kurtosis (queue de distribution), par ailleurs il

## Chapitre 3. La méthodologie de BOX et JENKINS

---

existe un autre test celui de Kolmogorov-Smirnov. Ces deux tests permettent par la suite de calculer les intervalles de prévisions.

### Test de Jarque et Bera :

Soit  $\mu_k = \frac{1}{n} \sum_{t=1}^n (\epsilon_t - \bar{\epsilon})^k$  le moment centré d'ordre k du processus  $\{\epsilon_t, t \in Z\}$

\* Le coefficient de Skewness est défini par :  $B_1^{\frac{1}{2}} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}$

\* Le coefficient de Kurtosis est défini par :  $B_2 = \frac{\mu_4}{\mu_2^2}$

- Si la distribution est normal et le nombre d'observations grand ( $n > 30$ ) :

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

On construit alors les statistiques centrées réduites correspondantes à  $B_1^{\frac{1}{2}}$  données par :

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

Que l'on compare à 1.96 (valeur de la loi normale au seuil de 5%)

Soient les hypothèses :

$$\begin{cases} H_0 : V_1 = 0 & \text{Symétrie} \\ H_0 : V_2 = 0 & \text{Aplatissement normal} \end{cases}$$

Nous acceptons l'hypothèse  $H_0$  si  $V_1 \leq 1.96$  et  $V_2 \leq 1.96$ ; dans le cas contraire, l'hypothèse de normalité est rejetée.

Le test de Jarque et Bera synthétise les résultats précédents. En effet,

• Si  $B_1^{\frac{1}{2}}$  obéissent à des lois normales alors la statistique :

$$JB = \frac{6}{n} B_1^{\frac{1}{2}} + \frac{n}{24} (B_2 - 3)^2 \text{ suit une } \chi^2 \text{ à deux degrés de liberté.}$$

• si  $JB > \chi_{1-\alpha}^2(2)$  on rejette l'hypothèse 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 à t-1 a une influence sur l'erreur produite à l'instant t. Le test de Durbin et Watson (1951) permet de détecter l'autocorrélation des résidus pour un ordre (corrélation entre  $\epsilon_t$  et  $\epsilon_{t-1}$ ) sous la forme :

$$\epsilon_t = \rho \epsilon_{t-1} + V_t \text{ où } V_t \rightarrow N(0, \sigma_v^2)$$

On teste :

$$\begin{cases} H_0 : \ll \rho = 0 \gg (\text{absence d'autocorrélation à l'ordre 1 des résidus}). \\ H_1 : \ll \rho \neq 0 \gg (\text{Présence d'autocorrélation à l'ordre 1 des résidus}). \end{cases}$$

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

$$DW = \frac{\sum_{t=2}^n (\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=2}^n \epsilon_t^2}$$

avec  $\epsilon_t$  : sont les résidus de l'estimation du modèle.

Cette statistique varie entre 0 et 4 et nous avons DW égale à 2 lorsque  $\hat{\rho} = 0$  ( $\hat{\rho}$  est l'estimateur de  $\rho$ ).

D'où :

- Il existe une autocorrélation positive, si  $DW = 0$ .
- Il existe une autocorrélation négative, si  $DW = 4$ .
- Il y a l'absence d'autocorrélation, si  $DW \simeq 2$ .

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

Pour ce faire, on utilise les corrélogrammes des carrés des résidus, si un ou plusieurs termes sont significativement différents de zéro, on déduit 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  $\epsilon_t$  tel que :

$$\epsilon_t = u_t h_t \text{ avec } u_t \sim N(0, 1) \text{ et } h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2$$

$$\text{Soit l'hypothèse : } \begin{cases} H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0 \\ H_1 : \text{contre.} \end{cases}$$

- $LM < \chi^2(1)$  on accepte  $H_0$ . La variance de l'erreur est constante  $h_t^2 = \alpha(0)$
- Dans le cas contraire  $LM > \chi^2(1)$  à  $p$  degrés de liberté compris entre 1 et 3, on rejette  $H_0$  et le processus est justifiable d'un modèle ARCH( $p$ ).
- Si  $p \geq 3$  le modèle sera justifié d'un modèle de type GARCH.

## 3.7 Prévision :

C'est l'ultime étape de cette méthode, pour 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 :

$$\hat{X}_{t+1} = E(X_{t+1}/I_t) \text{ ou } I_t \text{ est l'information disponible jusqu'à l'instant } t.$$

$$I_t = (X_1, X_2, \dots, X_t, \epsilon_1, \epsilon_2, \dots, \epsilon_t)$$

Considérons un ARMA(1,1) et un horizon de prévision  $h=1$

$$X_t = \phi_1 X_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1}$$

$$X_{t+1} = \phi_1 X_t + \epsilon_{t+1} - \theta_1 \epsilon_t$$

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

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

$$\hat{X}_{t+2} = E(X_{t+2}/I_t) = \theta_1 E(X_{t+1}/I_t)$$

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

### Chapitre 3. La méthodologie de BOX et JENKINS

---

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

En généralisant pour un horizon  $h \geq 1$  on trouve :

$$\Phi(B)X_t = \Theta(B)\epsilon_t \leftrightarrow X_t = \frac{\Theta(B)}{\Phi(B)}\epsilon_t = \Psi(B)\epsilon_t$$

$$X_t = \sum_{i=0}^{\infty} \Psi_i \epsilon_{t-i}$$

$$\text{et } \hat{X}_{t+h} = E(X_{t+h}/I_t) = E\left(\sum_{i=0}^{\infty} \Psi_i \epsilon_{t+h-i}/I_t\right)$$

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

$$\hat{X}_{t+h} = \sum_{i=0}^{\infty} \Psi_i \epsilon_{t+h-i}$$

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

$$\hat{\epsilon}_{t+1} = X_{t+1} - \hat{X}_{t+1} = \epsilon_{t+1} \quad \hat{\epsilon}_{t+2} = X_{t+2} - \hat{X}_{t+2} = \epsilon_{t+2} + \Psi_1 \epsilon_{t+1}$$

$$\hat{\epsilon}_{t+h} = X_{t+h} - \hat{X}_{t+h} = \sum_{i=0}^{h-1} \Psi_i \epsilon_{t+h-i} \quad \text{avec } \Psi_0 = 1$$

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

$$\text{Var}(\hat{\epsilon}_{t+h}) = \text{Var}\left[\sum_{i=0}^{h-1} \Psi_i \epsilon_{t+h-i}\right] = E\left(\sum_{i=0}^{h-1} \Psi_i \epsilon_{t+h-i}\right)^2 = \sigma_\epsilon^2 \sum_{i=0}^{h-1} \Psi_i^2$$

D'où l'intervalle de prévision :

$$\hat{X}_{t+h} \pm 1.96\sigma_\epsilon \left(\sum_{i=0}^{h-1} \Psi_i^2\right)^{\frac{1}{2}}$$

Noté par :  $[\hat{X}_{t+h} \pm 1.96\hat{h}_{t+k}]$

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 oeuvre.

Néanmoins, elle permet non seulement de traiter et d'analyser les processus stationnaires mais également les processus qui ne le sont pas.

# Chapitre 4

## La modélisation multivariée

### Introduction :

La modélisation économétrique classique à plusieurs équations a connu beaucoup de critiques (*GRANGER*1969) et de défaillances face à un environnement économique très perturbé. Les prévisions élaborées à l'aide de ces modèles se sont révélées très médiocres. Les critiques principales formulées à l'encontre de ces modèles concernent la simultanéité des relations et la notion de variable exogène. La représentation VAR (Vector AutoRegressive) - généralisation des modèles autorégressifs (AR) au cas multivarié- apporte une réponse statistique à l'ensemble de ces critiques. Dans cette représentation les variables sélectionnées en fonction du problème étudié ont toutes, a priori, le même statut et on s'intéresse alors à des relations purement statistiques.

La modélisation vectorielle autorégressive (*VAR*) a pour motivation de décrire les interdépendances entre un ensemble de variables, sans faire d'hypothèses a priori sur la valeur des coefficients des variables.

### 4.1 Processus multivariés

#### Définition

Un processus  $X_t, t \in \mathbb{Z}$  multivarié est une famille de variables aléatoires vectorielles définies sur  $\mathbb{R}^n$ . Comme dans le cas univarié, la stationnarité joue un rôle important dans la théorie des processus, c'est ainsi qu'on s'étalera à étudier les processus multivariés stationnaires dans ses différents champs.

#### Fonction d'autocovariance d'un processus multivarié

Considérons un processus multivarié  $X_t, t \in \mathbb{Z}$  de moyenne  $\mu$ , la covariance entre  $X_t$  et  $X_s$  est donnée par :

$$\gamma(t, s) = \text{Cov}(X_t, X_s) = E[(X_t - \mu)(X_s - \mu)'] ; \forall t, s \in \mathbb{Z}$$

### Processus multivarié fortement stationnaire :

Soit un processus multivarié  $X_t, t \in \mathbb{Z}$  le processus est dit fortement (ou strictement) stationnaire si :

$$\forall n \in \mathbb{N}^* \quad \forall (t_1, t_2, \dots, t_n) \in \mathbb{Z}^n \text{ et } \forall h \in \mathbb{Z}$$

le vecteur  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$  a la même loi de probabilité que la suite  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$

autrement dit :

$$P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) = P(X_{t_1+h} \leq x_1, \dots, X_{t_n+h} \leq x_n)$$

$$\forall (t_1, t_2, \dots, t_n) \in \mathbb{Z}^n ; \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n, \forall h \in \mathbb{Z}$$

Ainsi tous les moments d'ordre, d'un processus multivarié strictement stationnaire sont invariants pour toute translation dans le temps, or cette définition est rarement vérifiée en pratique, c'est ainsi que nous nous intéressons à un second type de stationnarité des processus multivariés, dit du second ordre.

### Processus multivarié faiblement stationnaire :

Le processus multivarié  $X_t, t \in \mathbb{Z}$  est dit faiblement stationnaire (du second ordre) si sa moyenne est finie indépendante du temps et de plus le processus est stationnaire en sa covariance.

$$\bullet E(X_t) = E(X_{t+h}) = \mu, \forall t \in \mathbb{Z}$$

$$\bullet Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)] = \Gamma(h) \forall t, h \in \mathbb{Z}$$

Où  $\Gamma(h)$  est la fonction d'autocovariance matricielle du processus.

### Proposition :

Si  $X_t, t \in \mathbb{Z}$  est stationnaire alors  $\forall i \in 1, \dots, n$   $X_{i,t}$  stationnaire. La réciproque est fausse.

### Remarque :

Dans ce qui suit le terme stationnaire, sauf mention contraire signifiera la stationnarité du second ordre

### Processus bruit blanc multivariée :

Un vecteur bruit blanc multivariée  $\epsilon_t, t \in \mathbb{Z}$  est une suite de variables aléatoires non corrélées de moyenne nulle et de matrice de covariance  $\Sigma_e$

$$\begin{cases} E(\epsilon_t) = 0 & \forall t \in \mathbb{Z} \\ E(\epsilon_t \epsilon_t') = \Sigma_e & \forall t \in \mathbb{Z} \end{cases}$$

Et on conséquence sa fonction d'autocovariance est donnée par :

$$\Gamma(h) = E(\epsilon_t, \epsilon_{t+h}) = \begin{cases} 0 & h \neq 0 \\ \Sigma_e & h = 0 \end{cases}$$

### Estimateur :

En pratique la fonction d'autocovariance est estimée à l'aide de l'estimateur suivant :

$$\Gamma(\hat{h}) = \frac{1}{t-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T)'$$

avec  $\bar{X}_t = \frac{1}{T} \sum_{t=1}^T X_t$

### 4.1.1 Fonction d'autocorrélation

La fonction d'autocorrélation d'un processus stationnaire multivarié faiblement stationnaire de moyenne  $\mu$  et de matrice de covariance  $\Gamma(h)$  notée  $A_h = \rho_{ij}$  est définie par

$$A_h = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}$$

### Décomposition de Wold :

Le théorème de Wold, est également valable dans le cas multivarié

### Théorème de Wold :

Tout processus stationnaire  $X_t, t \in \mathbb{Z}$  peut se décomposer en la somme d'une composante régulière prévisible (déterministe) et d'une expression linéaire stochastique tel que :

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

Où  $C_j$  est une suite de matrices carrées de taille  $n$  avec  $C_0 = I_n$  et  $\epsilon_t \in \mathbb{R}^n$  avec  $\epsilon_t$  bruit blanc de matrice de covariance  $\Sigma_e$

## 4.2 La représentation générale d'un modèle VAR

La généralisation de la représentation VAR à kvariables et p décalages (noté VAR(p)) s'écrit sous forme matricielle :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \epsilon_t \tag{4.1}$$

avec :

$$Y_t = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ \vdots \\ Y_{k,t} \end{bmatrix};$$

$$A_p = \begin{bmatrix} a_{1,p}^1 & a_{1,p}^2 & \cdots & \cdots & a_{1,p}^k \\ a_{2,p}^1 & a_{2,p}^2 & \cdots & \cdots & a_{2,p}^k \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{k,p}^1 & a_{k,p}^2 & \cdots & \cdots & a_{k,p}^k \end{bmatrix} ;$$

$$A_0 = \begin{bmatrix} a_1^0 \\ a_2^0 \\ \vdots \\ \vdots \\ a_k^0 \end{bmatrix} ; \epsilon_t = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \vdots \\ \epsilon_{k,t} \end{bmatrix}$$

où :

$Y_t = (Y_{1,t}, Y_{2,t}, \dots, Y_{k,t})$  est un vecteur colonne de dimension  $k$ .

$A_p$  est une matrice carrée d'ordre  $k$  tel que  $k = 1, \dots, p$ .

$A_0$  est un vecteur colonne de dimension  $k$ .

$\epsilon_t$  est un vecteur colonne de dimension  $k$  représentant le bruit blanc.

On note :  $\Sigma = E(\epsilon_t \epsilon_t')$  la matrice de dimension  $(k, k)$  des variances covariances des erreurs. Cette matrice est bien sur inconnue.

La représentation VAR ( $p$ ) sous forme matricielle peut s'écrire à l'aide de l'opérateur de retard  $B$  comme suit :

$$(1 - A_1 B - A_2 B^2 - \dots - A_p B^p) Y_t = A_0 + \epsilon_t ;$$

$$\text{ou encore } A(B) Y_t = A_0 + \epsilon_t$$

### Condition de stationnarité :

Un modèle VAR est stationnaire, s'il satisfait les trois conditions classiques :

- $E(Y_t) = \mu, \forall t \in Z$  .
- $Var(Y_t) < \infty$ .
- $Cov(Y_t, Y_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)'] = \Gamma_k, \forall t \in Z$

On démontre qu'un processus VAR( $p$ ) est stationnaire si le polynôme défini à partir du déterminant :

$det(I - A_1 z - A_2 z^2 - \dots - A_p z^p) = 0$  a ses racines à l'extérieur du cercle unité du plan complexe.

## 4.3 Estimation des paramètres du modèle VAR :

Les paramètres du modèle VAR ne peuvent être estimés que sur des séries chronologiques stationnaires.

### 4.3.1 Méthode d'estimation :

Dans le cas d'un processus VAR, chacune des équations [1] peut être estimée par la méthode des moindres carrés ordinaires (MCO), indépendamment les unes des autres (ou par une méthode de maximum de vraisemblance).

Soit le modèle VAR(p) estimé :  $Y_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \dots + \hat{A}_p Y_{t-p} + e_t$

avec :  $e_t$  étant le vecteur de dimension (k,1) des résidus d'estimation  $e_{1t}, e_{2t}, \dots, e_{kt}$ .

Et on note :  $\sum$  e la matrice des variances-covariances estimées des résidus du modèle.

### 4.3.2 Détermination du nombre de retards p :

La détermination de l'ordre d'un modèle VAR se fait à l'aide des deux critères d'informations : Akaike Information Criterion (AIC) et Schwarz Criterion (SC).

Dans le cas d'un modèle VAR, ces deux critères peuvent être utilisés pour déterminer l'ordre p du modèle.

La procédure de sélection de l'ordre de la représentation consiste à estimer tous les modèles VAR pour un ordre allant de 0 à ( h étant le retard maximum admissible par la théorie économique ou par les données disponibles).

Les fonctions AIC(p) et SC(p) sont calculées de la manière suivante :

$$AIC(p) = \ln(\det|\sum e|) + \frac{2k^2p}{n}$$

$$SC(p) = \ln(\det|\sum e|) + \frac{k^2 p \ln(n)}{n}$$

Avec :

k : nombre de variable du système ;

n : nombre d'observations ;

p : nombre de retards ;

$\sum e$  : matrice des variances covariances des résidus du modèle.

Notons que le retard p qui minimise les critères AIC ou SC est retenu.

## 4.4 Prévision

Les coefficients du modèle étant estimés, la prévision peut être calculée en n à l'horizon d'une période, par exemple pour un VAR(1), de la manière suivante :

$$\hat{Y}_n(1) = \hat{A}_0 + \hat{A}_1 Y_n$$

A l'horizon de 2 périodes, la prévision est :

$$\hat{Y}_n(2) = \hat{A}_0 + \hat{A}_1 \hat{Y}_n(1) = \hat{A}_0 + \hat{A}_1 \hat{A}_0 + \hat{A}_1^2 Y_n$$

A l'horizon de 3 périodes , la prévision s'écrit :

$$\hat{Y}_n(3) = \hat{A}_0 + \hat{A}_1 \hat{Y}_n(2) = (I + \hat{A}_1 + \hat{A}_1^2) \hat{A}_0 + \hat{A}_1^3 Y_n.ect$$

L'espérance de l'erreur de prévision est nulle, sa variance est donnée par :  $\sum e(h) = \sum e + M_1 \sum e M'_{h-1} + \dots + M_{h-1} \sum e M'_{h-1}$

Où  $M_i$  est calculé par la formule de récurrence suivante :

$$M_i = \sum_{j=1}^{\min(p,i)} \hat{A}_j M_{i-j} \quad i = 1, 2, 3, \dots \text{ et } M_0 = I$$

Ainsi il vient :

$$M_1 = \hat{A}_1;$$

$$M_2 = \hat{A}_1 M_1 + \hat{A}_2 M_0 = \hat{A}_1^2 + \hat{A}_2;$$

$$M_3 = \hat{A}_1 M_2 + \hat{A}_2 M_1 + \hat{A}_3 M_0 = \hat{A}_1^3 + \hat{A}_1 \hat{A}_2 + \hat{A}_2 \hat{A}_1 + \hat{A}_3 \text{ ect}$$

La variance de l'erreur de prévision pour chacune des prévisions des  $k$  variables ( $\hat{\sigma}_n^2(h)$ ) se lit sur la première diagonale de la matrice  $\sum e(h)$ .

L'intervalle de prévision au seuil  $(1 - \alpha/2)$  est donnée par :  $\hat{Y}_n(h) \pm t^{\alpha/2} \hat{\sigma}_n(h)$  telle que  $t^{\alpha/2}$  Valeur de la loi normale.

### 4.5 Causalité au sens de Granger :

Granger (1969) a proposé les concepts de causalité et d'exogénéité : la variable  $y_{2t}$  est la cause de  $y_{1t}$  si notre pouvoir de prédiction de  $y_{1t}$  est meilleur lorsque l'information relative à  $y_{2t}$  est induite dans l'analyse.

Soit le modèle VAR(p) pour lequel les variables  $y_{1t}$  et  $y_{2t}$  sont stationnaires :

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} a_1^1 & b_1^1 \\ a_1^2 & b_1^2 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \dots + \begin{bmatrix} a_p^1 & b_p^1 \\ a_p^2 & b_p^2 \end{bmatrix} \begin{bmatrix} Y_{1t-p} \\ Y_{2t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

Le bloc de variables  $(y_{2t-1}, \dots, y_{2t-p})$  est considéré comme exogène par rapport au bloc de variables  $(y_{1t-1}, \dots, y_{1t-p})$  si le fait de rajouter le bloc  $y_{2t}$  n'améliore pas significativement la détermination des variables  $y_{1t}$ .

Ceci consiste à effectuer un test de restrictions sur les coefficients des variables  $y_{2t}$  de la représentation VAR.

$y_{2t}$  ne cause pas  $y_{1t}$  si l'hypothèse suivante est acceptée :

$$H_0 : \ll b_1^1 = b_2^1 = \dots = b_p^1 = 0 \gg$$

$y_{1t}$  ne cause pas  $y_{2t}$  si l'hypothèse suivante est acceptée :

$$H_0 : \ll a_1^1 = a_2^1 = \dots = a_p^1 = 0 \gg$$

Ces tests peuvent être menés à l'aide d'un test de Fisher classique de nullité des coefficients, équation par équation.

# Chapitre 5

## Régression simple et multiple

### 5.1 Modèle de régression simple

#### 5.1.1 Présentation du modèle

Le modèle de régression décrit entre une variable  $Y$ , dont on cherche à expliquer le comportement, dite variable endogène, et une variable explicative  $X$ , dite variable exogène, cette liaison se formule par :

$$Y_t = \alpha_0 + \alpha_1 X_t + \epsilon_t; \forall t = 1, \dots, n$$

- $Y_t$  = variable à expliquer au temps  $t$  ;
- $X_t$  = variable explicative au temps  $t$  ;
- $\alpha_0, \alpha_1$  = paramètres du modèle ;
- $\epsilon_t$  = erreur de spécification (différence entre le modèle vrai et le modèle spécifié), cette erreur est inconnue et restera inconnue ;
- $n$  = nombre d'observations.

#### 5.1.2 Rôle du terme aléatoire

Le terme  $\epsilon_t$  mesure la différence entre les valeurs réellement observées  $Y_t$  et les valeurs qui auraient été observées si la relation spécifiée avait été rigoureusement exacte. Le terme donc regroupe trois erreurs : une erreur de spécification, une erreur de fluctuation d'échantillonnage, et une erreur de mesure.

Dans la réalité, nous connaissons les deux séries d'observations  $X_t$  et  $Y_t$ , mais pas les valeurs de  $\alpha_0$  et  $\alpha_1$ , ou les estimateurs notés  $\hat{\alpha}_0, \hat{\alpha}_1$  et respectivement sont des variables aléatoires, qui suivent les mêmes lois de probabilité, celle de  $\epsilon_t$ .

Les caractéristiques de moyenne et d'écart-type de ces coefficients permettent de construire des tests de validité du modèle estimé.

### 5.1.3 Estimation des paramètres

**Hypothèses :**

- $H_1$  : Le modèle est linéaire en  $X_t$ ;
- $H_2$  : Les valeurs sont observées sans erreur  $X_t$ ;
- $H_3$  :  $E(\epsilon_t) = 0$ ;
- $H_4$  :  $E(\epsilon_t^2) = \sigma_{\epsilon_t}^2 = Var(\epsilon_t)$ ;  $H_4$  s'appelle hypothèse d'homoscédastique
- $H_5$  :  $E(\epsilon_t, \epsilon_s) = 0$ ; si  $t \neq s$
- $H_6$  :  $Cov(X_t, X_{ct}) = 0$  l'erreur est indépendante de la variable explicative.

**Formulation des estimateurs :**

**Graphiquement :**

Nous pouvons faire un ajustement à l'aide d'une droite ,par le graphe des nuages de points, tel que l'estimateur des coefficients  $\alpha_0$  et  $\alpha_1$  est obtenu en minimisant la distance au carrée entre chaque observation et la droite, d'où le nom d'estimateur des moindres carrés ordinaires :MCO.

**Analytiquement :**

$$Min \sum_{t=1}^n \epsilon_t^2 = Min \sum_{t=1}^n (Y_t - \alpha_0 - \alpha_1 X_t)^2 = Min S$$

La solution du problème de minimisation est donnée par :

$$\begin{cases} \frac{\partial S}{\partial \alpha_0} = 0 \\ \frac{\partial S}{\partial \alpha_1} = 0 \end{cases}$$

$$\iff \begin{cases} \sum_{t=1}^n -2(Y_t - \hat{\alpha}_0 - \hat{\alpha}_1 X_t) = 0 \\ \sum_{t=1}^n -2X_t(Y_t - \hat{\alpha}_0 - \hat{\alpha}_1 X_t) = 0 \end{cases}$$

Et par suite

$$\iff \begin{cases} \hat{\alpha}_1 = \frac{\sum_{t=1}^n (X_t - \bar{X}_t)(Y_t - \bar{Y}_t)}{\sum_{t=1}^n (X_t - \bar{X}_t)^2} \\ \hat{\alpha}_0 = \bar{Y}_t - \hat{\alpha}_1 \bar{X}_t \end{cases}$$

Le modèle estimé peut s'écrire sous la forme suivante :

$$Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_t + e_t$$

Avec  $e_t = Y_t - \hat{Y}_t$  le résidu et  $\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_t$

### 5.1.4 Equation et tableau de la variance

**Equation d'analyse de la variance :**

Démontrons la relation suivante :

$$\sum e_t = 0$$

## Chapitre 5. Régression simple et multiple

---

$$Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_t + e_t \Rightarrow \sum_t Y_t = \sum_t \hat{\alpha}_0 + \sum_t \hat{\alpha}_1 X_t + \sum_t e_t$$

$$\Rightarrow \sum_t e_t = \sum_t Y_t - n\hat{\alpha}_0 - \hat{\alpha}_1 \sum_t X_t$$

Et comme  $\hat{\alpha}_0 = \bar{Y}_t - \hat{\alpha}_1 \bar{X}_t$ , alors  $\sum_t e_t = 0$  ;

$$\text{aussi } \sum_t Y_t = \sum_t \hat{Y}_t$$

$$\text{car } e_t = Y_t - \hat{Y}_t \Rightarrow \sum_t e_t = \sum_t Y_t - \sum_t \hat{Y}_t = 0$$

Donc, l'équation fondamentale d'analyse de la variance est :

$$\sum_t (Y_t - \bar{Y}_t)^2 = \sum_t (\hat{Y}_t - \bar{Y}_t)^2 + \sum_t (e_t)^2$$

$$SCT = SCE + SCR$$

avec :

- SCT : la somme des carrées totale  $\sum_t (Y_t - \bar{Y}_t)^2$
- SCE : la somme des carrées expliquée  $\sum_t (\hat{Y}_t - \bar{Y}_t)^2$
- SCR : la somme des carrées résiduelle  $\sum_t (e_t)^2$

Il est important de calculer :

**Le coefficient de corrélation**  $\rho_{x,y}$

Il mesure le degré de liaison entre deux phénomènes représentés par les variables X et Y , et qui est donné par :

$$\rho_{x,y} = \frac{\sum_{t=1}^n (X_t - \bar{X}_t)(Y_t - \bar{Y}_t)}{[\sum_{t=1}^n (X_t - \bar{X}_t)^2 \sum_{t=1}^n (Y_t - \bar{Y}_t)^2]^{\frac{1}{2}}} = \frac{Cov(X_t, Y_t)}{\sigma_x \sigma_y}$$

avec  $\rho_{x,y}$  entre -1 et 1

- lorsque  $\rho_{x,y}$  est proche e 1 (respectivement proche de -1), il existe une liaison apparente forte entre X et Y et donc les variables sont corrélées positivement (respectivement corrélées négativement) .
- lorsque  $\rho_{x,y}$  est faible, donc soit la liaison réelle est non linéaire, soit il n'existe pas de liaison entre X et Y.

**Le coefficient de détermination**  $R^2$  :

Il indique la qualité de l'ajustement réalisé, avec :

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y}_t)^2}{\sum_t (y_t - \bar{y}_t)^2} = 1 - \frac{\sum_t (e_t)^2}{\sum_t (y_t - \bar{y}_t)^2}$$

- une valeur de  $R^2$  proche de 1 montre que l'ajustement est bon.
- une valeur de  $R^2$  proche de 0 montre que l'ajustement est mauvais.

### 5.1.5 Tableau d'analyse de la variance

Source de variation	Somme de carées	Degrés de liberté	Carrés moyens
X	$SCE = \sum_t (\hat{Y}_t - \bar{Y}_t)^2$	1	SCE/1
Résidu	$SCR = \sum_t (e_t)^2$	n-2	SCR/n-2
Total;	$SCT = \sum_t (Y_t - \bar{Y}_t)^2$	n-1	•

Le degré de liberté correspondant au nombre de valeurs que nous pouvons choisir arbitrairement.

### 5.1.6 Test de spécification de l'ajustement

La détermination des paramètres  $\alpha_0, \alpha_1, \rho$  et  $R^2$  n'est pas suffisante pour affirmer que la liaison entre X et Y est significative pour cela, nous présenterons les différents tests d'hypothèses.

#### Test de signification des paramètres :

Pour une droite de régression :

$$Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_t + e_t$$

Les hypothèses sont formulées par :

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

La statistique utilisée est celle de student. Pour un seuil  $\alpha = 0.05$ , la règle de décision est alors la suivante :

$$\text{si } t_{\hat{\alpha}_1}^* = \frac{\hat{\alpha}_1}{\hat{\sigma}_{\hat{\alpha}_1}} > t_{n-2}^{\alpha/2}$$

Alors l'hypothèse  $H_0$  est rejetée, et donc la variable est contributive à l'explication de la variable  $Y_t$ .

Sinon, l'hypothèse  $H_0$  est acceptée, et donc la variable  $X_t$  n'est pas explicative de la variable  $Y_t$ .

#### Remarque

Si le nombre d'observation  $> 30$ , alors  $t_{n-2}^{\alpha/2} = 1.96$  avec  $\alpha = 0.05$ .

#### - Test d'analyse de la variance

La relation suivante :

$$SCT = SCE + SCR$$

Permet de juger la qualité de l'ajustement d'un modèle. En effet, plus la variance expliquée

## Chapitre 5. Régression simple et multiple

---

est proche de la variance totale, meilleur est l'ajustement du nuage de points par la droite des moindres carrées.

### - Test de signification globale (ou test de Fisher)

Le coefficient de corrélation  $\rho$  peut servir à tester les hypothèses suivantes

$$\begin{cases} H_0 : \rho = 0 : \text{la relation entre X et Y n'est pas significative.} \\ H_1 : \rho \neq 0 : \text{la relation entre X et Y est significative.} \end{cases}$$

La statistique utilisée est celle de Fisher, avec :

$$F^* = \frac{SCE/1}{SCR/(N-2)} = \frac{R^2}{(1-R^2)(N-2)}$$

Donc F empirique suit la loi de Fisher à 1 et (N-2) degrés de liberté, où :

$$F^* = (t^*)^2;$$

$t^*$  est la Student empirique ;

$R^2$  coefficient de détermination.

Si  $F^* > F_{(1,n-2)}^0 .05$  alors l'hypothèse  $H_0$  est rejetée ; donc  $X_t$  est une variable réellement explicative ;

Ainsi, si la variance expliquée est significativement supérieure à la variance résiduelle, alors  $X_t$  est une variable réellement explicative.

Dans le cas contraire, la variable  $X_t$  n'est pas explicative de la variable .

## 5.2 Modèle de régression multiple

### 5.2.1 Présentation du modèle

Le modèle linéaire général est une généralisation du modèle de régression simple dans lequel figurent plusieurs variables explicatives :

$$y_t = a_0 + a_1x_{1t} + a_2x_{2t} + \dots + a_kx_{kt} + \epsilon_t; t = 1, \dots, n$$

Avec :

$y_t$  : la variable à expliquer à la date t ;

$x_{1t}$  : La variable explicative 1 à la date t ;

$x_{2t}$  : La variable explicative 2 à la date t ;

:

:

$x_{kt}$  : La variable explicative k à la date t ;

$a_0, a_1, \dots, a_k$  : Les paramètres du modèle ;

$\epsilon_t$  : L'erreur de spécification (différence entre le modèle vrai et le modèle spécifié), cette erreur est inconnue et restera inconnue ;

n : Le nombre d'observation. En prenant la formulation matricielle du modèle linéaire général :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad Y = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}; \quad Y = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & \cdots & x_{k2} \\ \vdots & & & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

Donc  $Y = Xa + \epsilon$

### 5.2.2 Estimation des coefficients de régression

Afin d'estimer  $a = (a_0, a_1, \dots, a_k)'$ , nous appliquons la méthode des Moindres Carrés Ordinaires **MCO**, qui consiste à minimiser  $\sum_{t=1}^n \epsilon_t^2$ , soit :

$$\text{Min} \sum_{t=1}^n \epsilon_t^2 = \text{Min} \epsilon' \epsilon = \text{Min} (Y - Xa)' (Y - Xa) = \text{Min} S$$

avec  $\epsilon'$  : transposé du vecteur  $\epsilon$ .

Pour minimiser cette fonction par rapport à  $a$ , nous différencions  $S$  par rapport à  $a$ .

$$\frac{\partial S}{\partial a} = -2X'X\hat{a} = 0 \Rightarrow \hat{a} = (X'X)^{-1}X'Y$$

Cette solution est réalisable si la matrice carrée  $X'X$  de dimension  $(k+1, k+1)$  est inversible.

Le modèle estimé s'écrit :

$$y_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \dots + \hat{a}_k x_{kt} + e_t$$

avec  $e_t = y_t - \hat{y}_t$  où  $e_t$  est le résidu, c'est-à-dire l'écart entre la valeur observée de la variable à expliquer et sa valeur estimée (ajustée).

#### Attention

Il convient de bien distinguer entre l'erreur de spécification du modèle (noté  $\epsilon_t$ ) qui est et restera inconnue et le résidu ( $e_t$ ) qui lui est connu.

### 5.2.3 Hypothèses des estimateurs

$H_1$  : les valeurs sont observées sans erreur .

$H_2$  :  $E(\epsilon_t) = 0$ .

$H_3$  :  $E(\epsilon_t) = \sigma_\epsilon^2$ , la variance de l'erreur est constante ( $\forall t$ ) (homoscédasticité).

$H_4$  :  $E(\epsilon_t, \epsilon_{t'}) = 0$  si  $t \neq t'$ , les erreurs sont non corrélées (indépendantes).

$H_5$  :  $Cov(x_{i,t}, \epsilon_t) = 0$ , l'erreur est indépendante des variables explicatives.

### 5.2.4 Equation d'analyse de la variance

Nous avons les relations suivantes :

$$\sum_t y_t = \sum \hat{y}_t \Rightarrow \bar{y} = \bar{\hat{y}}$$

$$\sum_t \epsilon_t = 0$$

De ces deux relations, nous en déduisons l'équation fondamentale d'analyse de la variance :

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t \epsilon_t^2$$

$$SCT = SCE + SCR$$

La variabilité totale ( SCT ) est égale à la variabilité expliquée ( SCE ) + la variabilité des résidus ( SCR ).

Plus la variance expliquée est " proche " de la variance totale, meilleur est l'ajustement global du modèle.

#### - Coefficient de corrélation $\rho$

$\rho$  est le coefficient de corrélation entre la série  $y_t$  et  $\hat{y}_t$ , il représente la valeur maximale du coefficient de corrélation linéaire simple entre les coordonnées de Y et celle de tout vecteur de la forme  $X\hat{a}$ . Il est donné par :

$$\rho = \frac{Cov(Y, \hat{y}_t)}{\Sigma_y \Sigma_{\hat{y}_t}}$$

#### Coefficient de détermination $R^2$

Il indique la quantité de l'ajustement y par  $\hat{y}$ , avec :

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{\sum_t \epsilon_t^2}{\sum_t (y_t - \bar{y})^2} = \frac{SCE}{SCT}$$

$R^2$  : Variance expliquée par la regression / Variance de Y .

- Une valeur de  $R^2$  proche de 1, montre que l'ajustement est bon.
- Une valeur de  $R^2$  proche de 0, montre que l'ajustement est mauvais.
- Si  $R^2=1$ , alors  $y_t = \hat{y}_t$ , donc l'ajustement est parfait.

#### - Test globale de Fisher

Ce test permet de voir la relation entre Y et les variables  $x_1, x_2, \dots, x_k$  dans leur ensemble, ce qui est équivalent à l'hypothèse selon laquelle tous les coefficients sont nuls :

## Chapitre 5. Régression simple et multiple

---

$$\begin{cases} H_0 : & a_0 = a_1 = \dots = a_k = 0 \\ H_1 : & \text{Il existe au moins un des coefficients non nul.} \end{cases}$$

La statistique utilisée est :

$$F^* = \frac{SCE/k}{SCR/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Si  $F^* > F_{(k, n-k-1)}^0 \cdot 05$  alors l'hypothèse  $H_0$  est rejetée, le modèle est globalement explicatif. Sinon l'hypothèse  $H_0$  est acceptée.

## Deuxième partie

# Chapitre 6

## Application

### Application de la méthode de BOX ET JENKINS

Nous nous proposons dans ce présent chapitre de modéliser et de prévoir trois séries temporelles représentant les variations des prix du brut du pétrole et les nombres des découvertes , a fin de déterminer , parmi la classe des modèles ARIMA ( ils permettent de représenter la plupart des processus stationnaires ), le modèle adéquat qui représente au mieux la réalité .

#### 6.1 Étude de la série du Prix<sub>t</sub>

##### 6.1.1 Identification

Nous disposons d'une série annuel évoluant de 1986 à 2021 , ou toutes les variables sont en une série de 36 observations . Elles représente les variations des prix du brut du pétrole en Algérie ou l'unité est en dollars .

Notation :

Nous notons la série prix  $PRIX_t$

##### Représentation graphique de la série $PRIX_t$

Pour avoir une idée globale mais non décisive sur la nature et les caractéristiques du processus (tendance, saisonnalité,..), il est nécessaire d'analyser le graphe représentant l'évolution de la série en question.

D'après la représentation graphique de la série brute ( $PRIX_t$ ) (Figure 7.1) nous remarquons : une non stationnarité moyenne, qui se traduit par une légère tendance par d'autre descendants autour des valeurs 13 et 113 dollars .

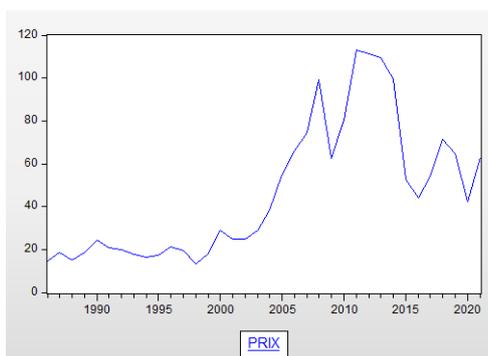


FIGURE 6.1 – le graphe de la série  $PRIX_t$

### Examen du corrélogramme de la série $PRIX_t$ :

Le logiciel économique Eviews fournit les résultats des fonctions d'autocorrélation simple (colonne AC) et partielle (colonne PAC), ainsi que leurs corrélogrammes associés. Les bornes de l'intervalle de confiance sont stylisées par des pointillés horizontaux ; tous les termes qui sont à l'intérieur de cet intervalle sont significativement nuls et ceux qui sortent de ce dernier sont significativement différent de zéro au seuil de 5%

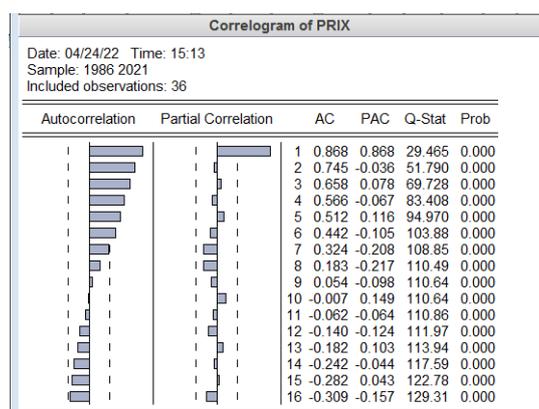


FIGURE 6.2 – le corrélogramme  $PRIX_t$

L'analyse du corrélogramme simple et partielle de la série brute (figure 7.2) nous indique une non stationnarité de la série. En effet, la fonction d'autocorrélation simple et partielle diminue lentement vers zéro, on remarque aussi de nombreux pics sortant de la bande de confiance .

L'hypothèse du test est :

$$\begin{cases} H_0 : \text{la série stationnaire} \\ H_1 : \text{la série n'est pas stationnaire} \end{cases}$$

Si les probabilités  $< 5\%$  on  $H_0$  Sinon on accepte  $H_0$

Nous constatons que la série n'est pas stationnaire puisque tous les probabilités  $<$

5%, On rejette  $H_0$ .

La série  $PRIX_t$  semble être générée par un processus non stationnaire. Afin de confirmer ou infirmer la non stationnarité de la série, nous proposons d'appliquer le test de DICKEY-FULLER Augmenté sur la série  $PRIX_t$ .

### Test de DICKEY-FULLER Augmenté

L'application du test de DICKEY-FULLER Augmenté nécessite la sélection de nombre du retard  $p$ , ainsi nous avons choisi le retard qui minimise le critère d'Akaike égale à 0.

Les résultats sont regroupés dans le tableau ci-dessous où la constante est notée C et la tendance @TREND. Au début, nous avons effectué un test sur la signification de la tendance, en se référant aux tables de Dickey-Fuller.

Les hypothèses à tester sont :

$$\begin{cases} H_0 : \text{la tendance n'est pas significative.} \\ H'_0 : \text{la constante n'est pas significative.} \\ H''_0 : \text{l'absence de la racine unitaire.} \end{cases}$$

On commence par l'estimation du modèle [3].

**Modèle [3] :** "Test sur la tendance"

$\Delta PRIX_t = \phi PRIX_{t-1} + bt + c + \epsilon_t$ ; Ou  $\epsilon_t$  est un processus stationnaire.

Nous utilisons le test programmé sous le logiciel Eviews.

Le résultat de l'affichage pour la série  $PRIX_t$  est donnée dans le tableau suivant :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRIX(-1)	-0.228758	0.116292	-1.967106	0.0579
C	3.834377	5.170042	0.741653	0.4637
@TREND("1986")	0.445169	0.363073	1.226114	0.2291
R-squared	0.110104	Mean dependent var		1.387429
Adjusted R-squared	0.054485	S.D. dependent var		15.30080
S.E. of regression	14.87813	Akaike info criterion		8.319478
Sum squared resid	7083.479	Schwarz criterion		8.452794
Log likelihood	-142.5909	Hannan-Quinn criter.		8.365499
F-statistic	1.979625	Durbin-Watson stat		1.747426
Prob(F-statistic)	0.154678			

FIGURE 6.3 – ADF modèle 3 de la série  $PRIX_t$

À la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (t-statistic = 1.226114) est inférieure à toutes les valeurs critiques 3.53, 2.79, 2.38 lues dans la table de Dickey-Fuller respectivement aux seuils 1%, 5%, 10%.

- La probabilité de nullité du coefficient de la tendance (@trend) qui est égale à 0.2291 est supérieure au seuil de 5%.

Donc on accepte  $H_0$

## Chapitre 6. Application

la tendance n'est pas significative.

Nous passons alors au test du modèle [2]

**Modèle [2] :** "Test sur la constante"

$\Delta PRIX_t = \phi PRIX_{t-1} + c + \epsilon_t$ ; Ou  $\epsilon_t$  est un processus stationnaire.

Augmented Dickey-Fuller Unit Root Test				
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(PRIX)				
Method: Least Squares				
Date: 04/18/22 Time: 11:11				
Sample (adjusted): 1987 2021				
Included observations: 35 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRIX(-1)	-0.124990	0.080363	-1.555312	0.1294
C	7.102592	4.463603	1.591224	0.1211
R-squared	0.068297	Mean dependent var		1.387429
Adjusted R-squared	0.040063	S.D. dependent var		15.30080
S.E. of regression	14.99117	Akaike info criterion		8.308245
Sum squared resid	7416.260	Schwarz criterion		8.397122
Log likelihood	-143.3943	Hannan-Quinn criter.		8.338925
F-statistic	2.418996	Durbin-Watson stat		1.843270
Prob(F-statistic)	0.129411			

FIGURE 6.4 – ADF modèle 2 de la série  $PRIX_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (tstatistic = 1.591224) est inférieure à toutes les valeurs critique 3.22, 2.54, 2.17 lues dans la table de Dickey–Fuller respectivement aux seuils 1%, 5%, 10% .

- La probabilité de nullité du coefficient de la constante ( $c$ ) qui est égale à 0.1211 est supérieure au seuil de 5%.

Donc on accepte  $H_0'$

la constante n'est pas significative.

Nous passons alors au test du modèle [1]

**Modèle [1] :** "Test sur la racine unitaire"

$\Delta PRIX_t = \phi PRIX_{t-1} + \epsilon_t$ , Ou  $\epsilon_t$  est un processus stationnaire.

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.422776	0.5234
Test critical values:		
1% level	-2.632688	
5% level	-1.950687	
10% level	-1.611059	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(PRIX)  
 Method: Least Squares  
 Date: 04/18/22 Time: 11:12  
 Sample (adjusted): 1987 2021  
 Included observations: 35 after adjustments

FIGURE 6.5 – ADF modèle 1 de la série  $PRIX_t$

## Chapitre 6. Application

---

La valeur de la t-statistic de test ADF qu'est égale à  $-0,422776$  est supérieur à la différente valeur critique aux différent seuils :  $-2,632688$  ;  $-1,950687$  ;  $-1,611059$  ( 1% ,5% ,10%) respectivement (donner par le test ADF )

Donc on rejette  $H_0''$ .

La série  $PRIX_t$  possède une racine unitaire et le processus qui génère cette série est non stationnaire de type DS sans dérivé .

### Stationarisation de la série :

On procède à la différentiation ordinaire d'ordre 1 . Sous EVIEWS on insere la nouvelle série engendrée par la différentiation première :

$$dprix_t = prix_t - prix_{t-1}$$

### Le graphe de la série dPRIX<sub>t</sub> :

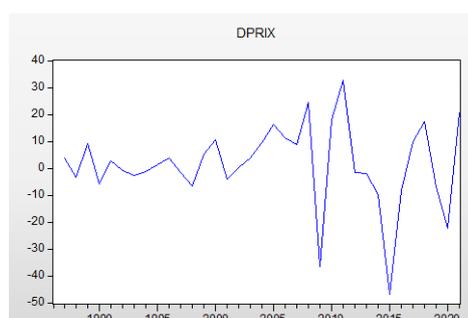


FIGURE 6.6 – graphe de la série dPRIX<sub>t</sub> :

Nous remarquons que l'effet de la tendance a été absorbé.

### Analyse du corrélogramme de la série dPRIX<sub>t</sub>

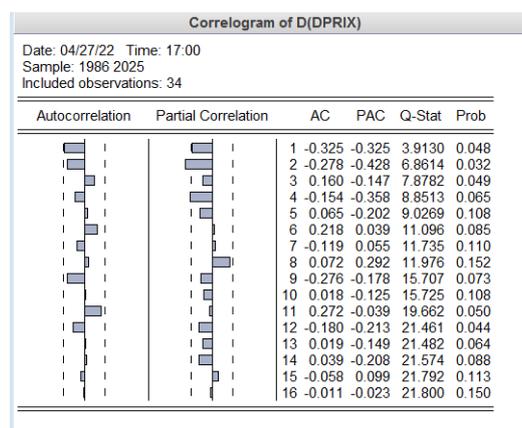


FIGURE 6.7 – correlogramme de la série  $dPRIX_t$

- L'examen du correlogramme de la série (figure 6.7) montre que la série représente un bruit blanc, car tous les termes ne sont pas significativement différents de zéro, on peut en conclure que le processus étudié est sans mémoire.

la série peut être stationnaire, pour confirmer ou infirmer cette hypothèse, il convient d'appliquer à la série  $dPRIX_t$  le test de la racine unitaire DICKEY-FULLER Augmenté.

### 6.1.2 Test de la stationnarité de la série $dPRIX_t$

**Test de DICKEY-FULLER Augmenté sur la série  $dPRIX_t$  :**

On applique le test de DFA pour infirmer ou confirmer notre hypothèse concernant la stationnarité du processus qui génère la série  $dprix_t$ , on a besoin de déterminer le retard

qui minimise le critère d'Akaike et Schwarz qui égale à 0. Par la suite on estime par la méthode des moindres carrés

les trois modèles.

Les hypothèses à tester sont :

$$\begin{cases} H_0 : \text{la tendance n'est pas significative.} \\ H'_0 : \text{la constante n'est pas significative.} \\ H''_0 : \text{l'absence de la racine unitaire.} \end{cases}$$

**On commence par l'estimation du modèle [3].**

**Modèle [3] :** "Test sur la tendance"

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (t-statistic = -0,241600) est inférieure à toutes les valeurs critiques 3.53, 2.79, 2.38 lues dans la table de Dickey-Fuller respectivement aux seuils 1%, 5%, 10%.

## Chapitre 6. Application

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(DPRIX)  
 Method: Least Squares  
 Date: 04/18/22 Time: 11:19  
 Sample (adjusted): 1988 2021  
 Included observations: 34 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DPRIX(-1)	-0.998816	0.185377	-5.388016	0.0000
C	2.574404	5.910623	0.435555	0.6662
@TREND("1986")	-0.068096	0.281855	-0.241600	0.8107

R-squared	0.485118	Mean dependent var	0.504118
Adjusted R-squared	0.451900	S.D. dependent var	21.61512
S.E. of regression	16.00249	Akaike info criterion	8.467463
Sum squared resid	7938.469	Schwarz criterion	8.602142
Log likelihood	-140.9469	Hannan-Quinn criter.	8.513392
F-statistic	14.60400	Durbin-Watson stat	1.939026
Prob(F-statistic)	0.000034		

FIGURE 6.8 – ADF modèle 3 de la série  $dPRIX_t$

- La probabilité de nullité du coefficient de la tendance (@trend) qui est égale à 0.8107 est supérieure au seuil de 5%.

Donc on accepte  $H_0$

la tendance n'est pas significative.

**Nous passons alors au test du modèle [2]**

**Modèle [2] :** "Test sur la constante"

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(DPRIX)  
 Method: Least Squares  
 Date: 04/18/22 Time: 11:20  
 Sample (adjusted): 1988 2021  
 Included observations: 34 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DPRIX(-1)	-0.993335	0.181256	-5.480272	0.0000
C	1.310179	2.707722	0.483868	0.6318

R-squared	0.484149	Mean dependent var	0.504118
Adjusted R-squared	0.489028	S.D. dependent var	21.61512
S.E. of regression	15.76529	Akaike info criterion	8.410520
Sum squared resid	7953.416	Schwarz criterion	8.500306
Log likelihood	-140.9788	Hannan-Quinn criter.	8.441140
F-statistic	30.03338	Durbin-Watson stat	1.942269
Prob(F-statistic)	0.000005		

FIGURE 6.9 – ADF modèle 2 de la série  $dPRIX_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (t-statistic = 0,483868) est inférieure à toutes les valeurs critique 3.22, 2.54, 2.17 lues dans la table de Dickey-Fuller respectivement aux seuils 1%, 5%, 10%.

- La probabilité de nullité du coefficient de la constante (c) qui est égale à 0.6318 est supérieure au seuil de 5%.

Donc on accepte  $H'_0$

la constante n'est pas significative.

Nous passons alors au test du modèle [1]

Modèle [1] : "Test sur la racine unitaire"

Null Hypothesis: DPRIX has a unit root				
Exogenous: None				
Lag Length: 0 (Automatic - based on SIC, maxlag=8)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-5.526560	0.0000
Test critical values:				
	1% level		-2.634731	
	5% level		-1.951000	
	10% level		-1.610907	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DPRIX)				
Method: Least Squares				
Date: 04/18/22 Time: 11:21				
Sample (adjusted): 1988 2021				
Included observations: 34 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DPRIX(-1)	-0.988570	0.178876	-5.526560	0.0000
R-squared	0.480374	Mean dependent var	0.504118	

FIGURE 6.10 – ADF modèle 1 de la série  $dPRIX_t$

La valeur de la t-statistic de test ADF qu'est égale à -5,526560 est inférieure à la différente valeur critique aux différent seuils : -2,632688 ; -1,950687 ; -1,611059 1% ,5% ,10% respectivement (donner par le test ADF) , alors on accepte  $H_0''$  .

La série  $dPRIX_t$  ne possède pas une racine unitaire .

En conclusion , notre série est stationnaire .Et on peut passer à l'autre étape à savoir l'identification .

### 6.1.3 Identification du modèle

L'analyse du corrélogramme simple de la série stationnaire ( $dprix_t$ ) montre qu'aux retards ( $q = 1$ ) le termes sont à l'extérieure de l'intervalle de confiance et ce qui concerne le corrélogramme partiel , on remarque que les valeurs de la fonction d'autocorrélation partielles sont élevées aux différents retard ( $p = 1, 2$ ) ,ce qui amène à estimer plusieurs modèles parmi eux on a choisit : $ARMA(1,1)$  D'ou le modèle  $ARIMA(p = 1, d = 1, q = 1)$  est celui le plus adéquant pour modéliser notre série .

### 6.1.4 Estimation des paramètre du modèle

Le tableau suivant contient l'estimation du modèle  $ARMA(1,1)$ , qui a été choisi parmi d'autre modèles estimés on se basant sur les critères de pouvoir prédictif à savoir :  $R^2$  statistique de Fisher (*maximum*) , AIC , SC (*minimum*) et sur la statistique de Durbin Watson qui égale 1,84 présage un bon ajustement .

## Chapitre 6. Application

Dependent Variable: DPRIX  
 Method: ARMA Maximum Likelihood (OPG - BHHH)  
 Date: 04/19/22 Time: 10:44  
 Sample: 1987 2021  
 Included observations: 35  
 Convergence achieved after 30 iterations  
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.873538	0.284462	3.070840	0.0043
MA(1)	-0.938001	0.229498	-4.087185	0.0003
SIGMASQ	224.7405	45.96267	4.889630	0.0000

R-squared	0.011806	Mean dependent var	1.387429
Adjusted R-squared	-0.049956	S.D. dependent var	15.30080
S.E. of regression	15.67833	Akaike info criterion	8.428103
Sum squared resid	7865.916	Schwarz criterion	8.561419
Log likelihood	-144.4918	Hannan-Quinn criter.	8.474124
Durbin-Watson stat	1.842896		

Inverted AR Roots	.87
Inverted MA Roots	.94

FIGURE 6.11 – le modèle

### 6.1.5 Validation du modèle

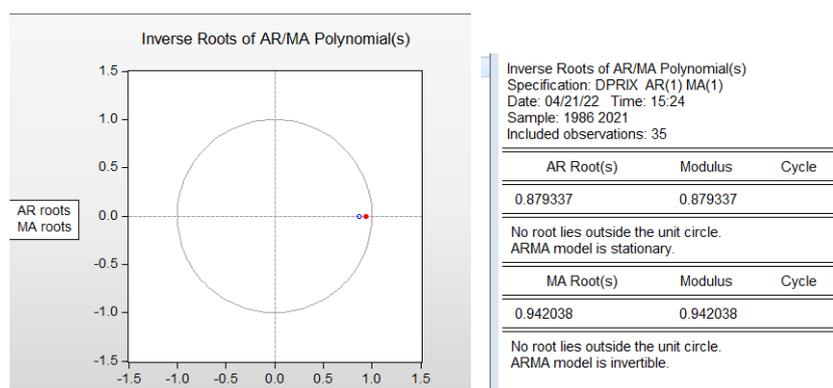


FIGURE 6.12

On remarque que :

- Tous les coefficients du modèle sont significative et différent de 0 .
  - La probabilité de nullité des composantes AR et MA est inférieur à 5% .
  - La statistique de Student des composantes AR et MA en valeurs absolue est supérieur à 1,96 .
  - Ces composantes n'ont pas de racine commune .
  - Les inverses des racines sont des modules inférieur à 1 .
- Et donc ce modèle est retenu , pour cela on va s'assurer avec des tests a l'étape suivante .

Test sur les résidus :

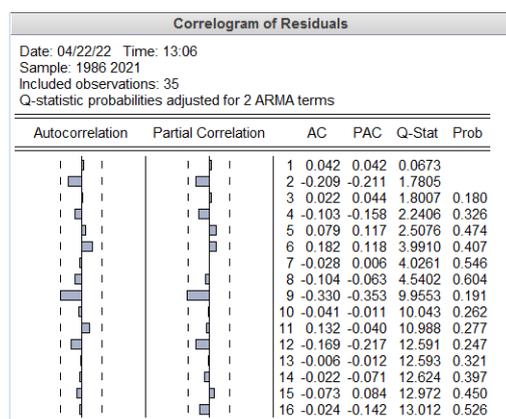


FIGURE 6.13

**Examen du correlogramme des résidus :** Le correlogramme des résidus forment un bruit blanc ,car il n'apparait aucun terme en dehors l'intervalle de confiance au seuil 5 % . donc, l'estimation du modèle  $ARIMA(1,1,1)$  est validée, la série  $dprix_t$  peut etre valablement représentée par un processus ARIMA de type  $ARIMA(1,1)$  .

**Ljung-Box (test d'autocorrélation :** Le correlogramme des résidus montre bien que les résidus forment un bruit blanc .Ce qu' est confirmé par la statistique de Ljung- Box ( $Q - stat$ ) (la dernière valeur calculée sur le correlogramme) qu'est inférieure à la valeur théorique de  $khi2(N - p - q)$  :

- N :le nombre d'oservation.
- p :l'ordre d'autorégressif.
- q :l'ordre de moyenne mobile.

En effet ,

$Q - stat = 13,01 < \chi^2(34) = 49,765$  au seuil 5% d'où la validation du modle

**Durbin Watson :** on test :

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

si DW est proche de 2 on accepte  $H_0$  la statistique

$DW = 1.84 \approx 2$ , on accepte l'hypotèse de non corrélation des résidus.

**test de normalité des résidus :**

les tests sont effectués à partir des valeurs empiriques des coefficients de SKewness, Kurtosis et la statistique de Jarque- Berra par le logiciel EVIEWS 10.

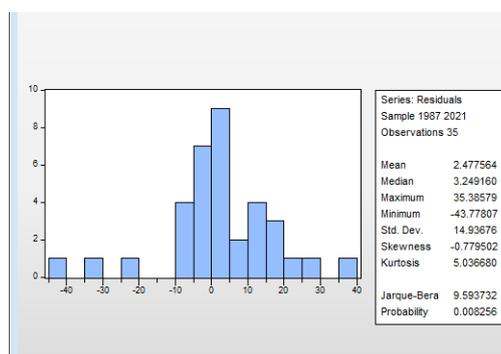


FIGURE 6.14

### Test de Skewness :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_1 = 0 \text{ (la distribution des résidus est Symétrique )} \\ H_1 : \gamma_1 \neq 0 \text{ (la distribution des résidus est Asymétrique)} \end{cases}$$

$$\gamma_1 = \frac{|\sqrt{\beta_1}|}{\sqrt{\frac{6}{n}}} = \frac{\sqrt{0.779502}}{\sqrt{\frac{6}{35}}} = 2.13 > 1.96$$

on rejette  $H_0$

### Test de Kurtosis :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_2 = 0 \text{ (la distribution des résidus est relativement Aplatie )} \\ H_1 : \gamma_2 \neq 0 \text{ (la distribution des résidus est n'est pas Aplatie )} \end{cases}$$

$$\gamma_2 = \frac{|\beta_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|5.036680 - 3|}{\sqrt{\frac{24}{35}}} = 2.45 > 1.96$$

On rejette  $H_0$  .

nous confirmons par la statistique de Jarque-Berra

### Test des résidus de Jarque-Berra :

L'hypothèse de test :

$$\begin{cases} H_0 : \text{Les résidus forment un bruit blanc gaussien} \\ H_1 : \text{Les résidus forment un bruit blanc non gaussien} \end{cases}$$

La statistique de Jarque-Berra :  $JB = \frac{n}{6}\beta_1 + \frac{n}{24}(\beta_2 - 3)^2$   $JB = 9.593732 > \chi_{0.05}^2(2) = 5.99$   
les résidus forment un bruit blanc non gaussien.

Le processus ARIMA(1, 1, 1) est donc un bruit blanc non gaussien .

### corrélogramme résiduel au carré

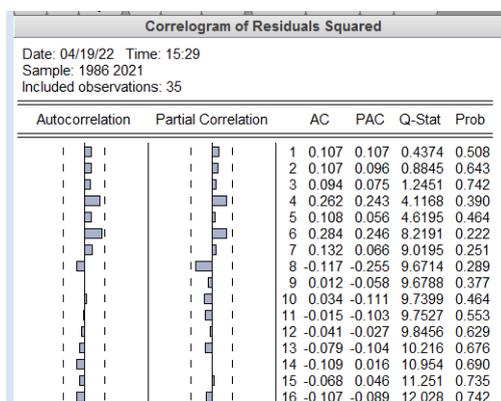


FIGURE 6.15 – corrélograme de résiduel au carré

on rejette l'hypothèse nulle d'hétéroscédasticité,  
il n'existe pas un effet ARCH

### Graphes des séries résiduelle, actuelle et estimée :

La représentation graphique des séries résiduelle, actuelle et estimée montre que le modèle a bien expliqué la série

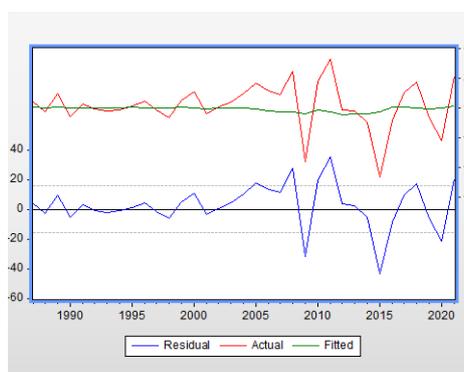


FIGURE 6.16 – Le graphe des séries résiduelle, actuelle et estimée

### 6.1.6 La prévision

Rappelons que la série sous étude "prix" comprend 36 observations de 1986 à 2021 donc l'instant est 2021.

on souhaitera prédire 2022 2023 2024 2025 sachant que

## Chapitre 6. Application

---

$$dprix_t = prix_t - prix_{t-1}$$

en remplaçant t par t+h on obtient :

$$prix_{t+h} = prix_{t-1+h} + 0.87prix_{t+h-1} + \epsilon_{t+h} - 0.93\epsilon_{t+h-1}$$

avec h=1,2,3,4

on obtient les prévisions suivants :

années	prix en dollars
2022	84.06
2023	85.44
2024	86.82
2025	88.2

Le graphe des prévision est donné par :

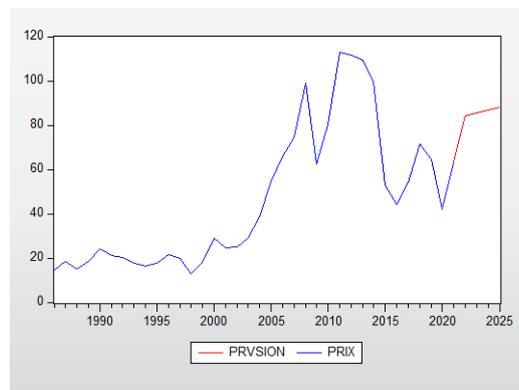


FIGURE 6.17 – la prévision

### 6.2 Étude de la série du $DCOVT_t$

#### 6.2.1 Identification :

Nous disposons d'une série annuel évoluant de 1986 à 2021 , ou toutes les variables sont en une série de 36 observations . Elles représente la découverte annuel des sismique et forage du brut du pétrole en Algérie .

Notation :

Nous notons la série Découverte  $DCOVT_t$

#### Représentation graphique de la série $DCOVT_t$

Pour avoir une idée globale mais non décisive sur la nature et les caractéristiques du processus (tendance, saisonnalité,...), il est nécessaire d'analyser le graphe représentant l'évolution de la série en question. (Figure 7.18)

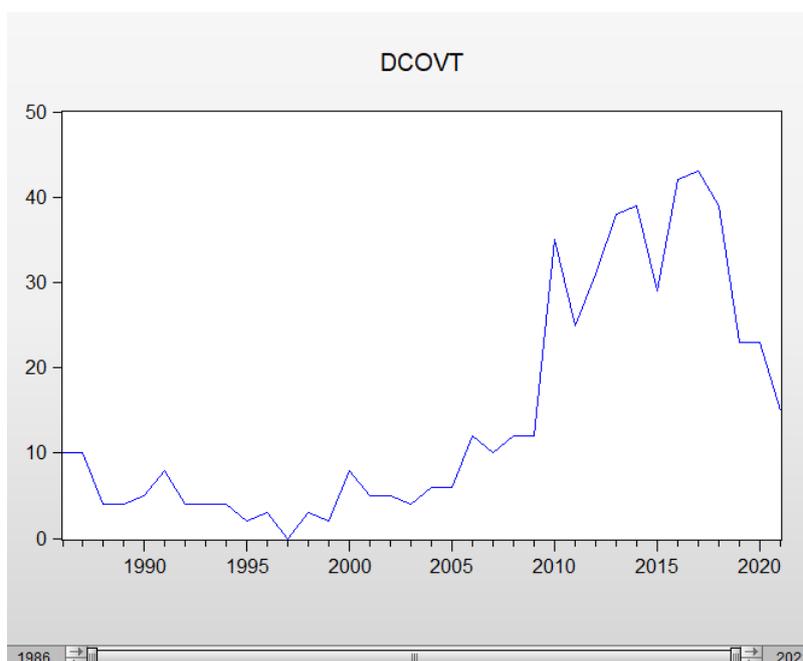


FIGURE 6.18 – le graphe de la série  $DCOVT_t$

D'après la représentation graphique de la série brute ( $DCOVT_t$ ) (Figure 7.18) nous remarquons : une non stationnarité moyenne, qui se traduit par une légère tendance par d'autre descendants autour des valeurs 0 et 43 découverte .

### Examen du corrélogramme de la série $DCOVT_t$ :

Le logiciel économique Eviews fournit les résultats des fonctions d'autocorrélation simple (colonne AC) et partielle (colonne PAC), ainsi que leurs corrélogrammes associés. Les bornes de l'intervalle de confiance sont stylisées par des pointillés horizontaux; tous les termes qui sont à l'intérieur de cet intervalle sont significativement nuls et ceux qui sortent de ce dernier sont significativement différent de zéro au seuil de 5% .

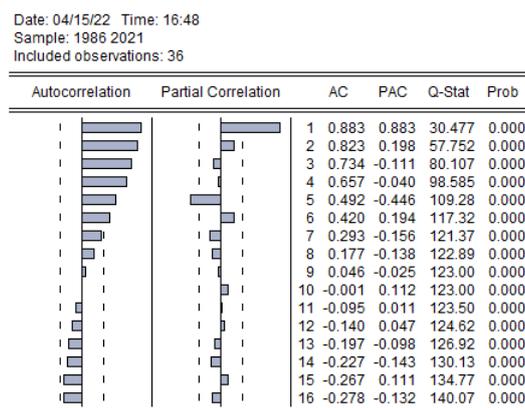


FIGURE 6.19 – le corrélogramme  $DCOVT_t$

L'analyse du corrélogramme simple et partielle de la série brute (figure 7.19) nous indique une non stationnarité de la série. En effet, la fonction d'autocorrélation simple et partielle diminue lentement vers zéro, on remarque aussi de nombreux pics sortant de la bande de confiance .

L'hypothèse du test est :

$$\begin{cases} H_0 : & \text{la série stationnaire} \\ H_1 : & \text{la série n'est pas stationnaire} \end{cases}$$

Si les probabilités  $< 5\%$  on rejette  $H_0$  Sinon on accepte  $H_0$

Nous constatons que la série n'est pas stationnaire puisque tous les probabilités  $< 5\%$  .

On rejette  $H_0$  .

la série  $DCOVT_t$  semble etre générée par un procesus non stationnaire .

Afin de confirmer ou infermer la non stationnarité de la série , nous nous proposons d'appliquer le test de DICKEY-FULLER Augmenté sur la série  $DCOVT_t$  .

### Test de DICKEY-FULLER Augmenté

l'application du test de DICKEY-FULLER Augmenté nécessite la selection de nombre du retard  $p$  .

Les résultats sont regroupés dans le tableau ci-dessous où la constante est notée  $C$  et la

## Chapitre 6. Application

tendance @TREND. Au début, nous avons effectué un test sur la signification de la tendance, en se référant aux tables de Dickey-Fuller.

Les hypothèses à tester sont :

$$\begin{cases} H_0 : \text{la tendance n'est pas significative .} \\ H'_0 : \text{la constante n'est pas significative .} \\ H''_0 : \text{l'absence de la racine unitaire .} \end{cases}$$

On commence par l'estimation du modèle [3] .

**Modèle [3] :** "Test sur la tendance"

$\Delta DCOVT_t = \phi DCOVT_{t-1} + Bt + C + \epsilon_t$ ; Ou  $\epsilon_t$  est un processus stationnaire.

Nous utilisons le test programmé sous le logiciel Eviews.

Le résultat de l'affichage pour la série  $DCOVT_t$  est donnée dans le tableau suivant :

Augmented Dickey-Fuller Unit Root Test on DCOVT				
Null Hypothesis: DCOVT has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 0 (Automatic - based on SIC, maxlag=9)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-2.119035	0.5178
Test critical values:	1% level		-4.243644	
	5% level		-3.544284	
	10% level		-3.204699	
*Mackinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DCOVT)				
Method: Least Squares				
Date: 04/15/22 Time: 22:12				
Sample (adjusted): 1987 2021				
Included observations: 35 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DCOVT(-1)	-0.270578	0.127689	-2.119035	0.0419
C	-0.683630	2.285637	-0.299098	0.7668
@TREND("1986")	0.264955	0.172383	1.537009	0.1341
R-squared	0.123976	Mean dependent var		0.142857
Adjusted R-squared	0.069224	S.D. dependent var		6.642744
S.E. of regression	6.408701	Akaike info criterion		6.635007
Sum squared resid	1314.286	Schwarz criterion		6.768322
Log likelihood	-113.1126	Hannan-Quinn criter.		6.681027
F-statistic	2.264337	Durbin-Watson stat		2.144779
Prob(F-statistic)	0.120298			

FIGURE 6.20 – ADF modèle 3 de la série  $DCOVT_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (t-statistic = 1.537009) est inférieure à toutes les valeurs critiques 3.53, 2.79, 2.38 lues dans la table de Dickey-Fuller respectivement aux seuils 1%, 5%, 10% .
- La probabilité de nullité du coefficient de la tendance (@trend) qui est égale à 0.1341 est supérieure au seuil de 5% .

## Chapitre 6. Application

Donc on accepte  $H_0$   
la tendance n'est pas significative.

Nous passons alors au test du modèle [2]

Modèle [2] : "Test sur la constante"

$\Delta DCOVT_t = \phi DCOVT_{t-1} + C + \epsilon_t$ ; Ou  $\epsilon_t$  est un processus stationnaire.

Augmented Dickey-Fuller Unit Root Test on DCOVT				
Null Hypothesis: DCOVT has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=9)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.442355	0.5504
Test critical values:	1% level		-3.632900	
	5% level		-2.948404	
	10% level		-2.612874	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DCOVT)				
Method: Least Squares				
Date: 04/15/22 Time: 22:42				
Sample (adjusted): 1987 2021				
Included observations: 35 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DCOVT(-1)	-0.116937	0.081073	-1.442355	0.1586
C	1.846792	1.617875	1.141493	0.2619
R-squared	0.059303	Mean dependent var		0.142857
Adjusted R-squared	0.030798	S.D. dependent var		6.642744
S.E. of regression	6.539654	Akaike info criterion		6.649091
Sum squared resid	1411.314	Schwarz criterion		6.737968
Log likelihood	-114.3591	Hannan-Quinn criter.		6.679771
F-statistic	2.080389	Durbin-Watson stat		2.328070
Prob(F-statistic)	0.158627			

FIGURE 6.21 – ADF modèle 2 de la série  $DCOVT_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (tstatistic = 1.141493) est inférieure à toutes les valeurs critique 3.22, 2.54, 2.17 lues dans la table de Dickey-Fuller respectivement aux seuils 1%, 5%, 10%

- La probabilité de nullité du coefficient de la constante ( $c$ ) qui est égale à 0.2619 est supérieure au seuil de 5% .

Donc on accepte  $H'_0$

la constante n'est pas significative.

Nous passons alors au test du modèle [1]

## Chapitre 6. Application

**Modèle [1] :** "Test sur la racine unitaire"

$\Delta DCOVT_t = \phi DCOVT_{t-1} + \epsilon_t$ ; Ou  $\epsilon_t$  est un processus stationnaire.

Augmented Dickey-Fuller Unit Root Test on DCOVT				
Null Hypothesis: DCOVT has a unit root				
Exogenous: None				
Lag Length: 0 (Automatic - based on SIC, maxlag=9)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-0.887170	0.3247
Test critical values:				
	1% level		-2.632688	
	5% level		-1.950687	
	10% level		-1.611059	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DCOVT)				
Method: Least Squares				
Date: 04/15/22 Time: 22:56				
Sample (adjusted): 1987 2021				
Included observations: 35 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DCOVT(-1)	-0.049361	0.055639	-0.887170	0.3812
R-squared	0.022160	Mean dependent var		0.142857
Adjusted R-squared	0.022160	S.D. dependent var		6.642744
S.E. of regression	6.568730	Akaike info criterion		6.630673
Sum squared resid	1467.039	Schwarz criterion		6.675112
Log likelihood	-115.0368	Hannan-Quinn criter.		6.646014
Durbin-Watson stat	2.395653			

FIGURE 6.22 – ADF modèle 1 de la série  $DCOVT_t$

la valeur de la t-statistic de test ADF qu'est égale à -0,887 est supérieure à la différence valeur critique aux différents seuils : -2,632688 ; -1,950687 ; -1,611059 1%, 5%, 10% respectivement (donner par le test ADF )

Donc on rejette  $H_0''$  .

La série  $DCOVT_t$  possède une racine unitaire et le processus qui génère cette série est non stationnaire de type DS sans dérivé .

**Stationnarisation de la série  $DCOVT_t$  :**

On procède à la différentiation ordinaire d'ordre 1 . Sous EVIEWS on insère la nouvelle série engendrée par la différentiation première :

$$dDCOVT_t = DCOVT_t - DCOVT_{t-1}$$

**Le graphe de la série  $dDCOVT_t$  :**

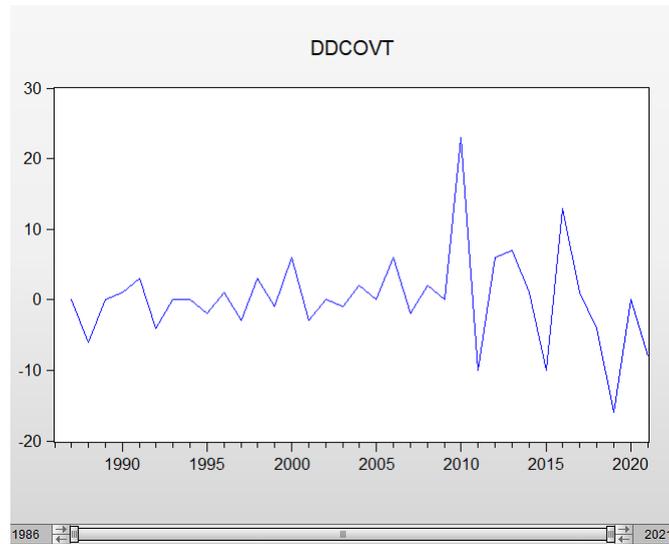


FIGURE 6.23 – graphe de la série  $dDCOVT_t$

Nous remarquons que l'effet de la tendance a été absorbé.

Analyse du corrélogramme de la série  $dDCOVT_t$  :

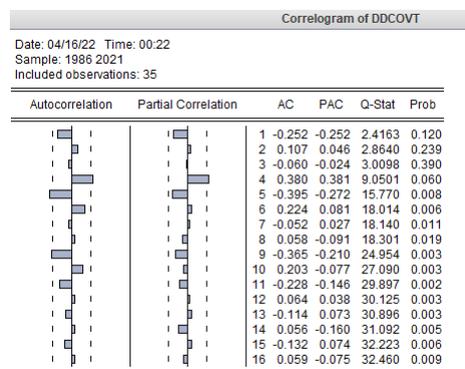


FIGURE 6.24 – corrélogramme de la série  $dDCOVT_t$

- L'examen du corrélogramme de la série (figure 7.24) montre que la série représente un bruit blanc, car tous les termes ne sont pas significativement différents de zéro, on peut en conclure que le processus étudié est sans mémoire.

la série peut être stationnaire, pour confirmer ou infirmer cette hypothèse, il convient d'appliquer à la série  $dDCOVT_t$  le test de la racine unitaire DICKEY-FULLER Augmenté.

### 6.2.2 Test de la stationnarité de la série $dDCOVT_t$

Test de DICKEY-FULLER Augmenté sur la série  $dDCOVT_t$  :

On applique le test de DFA pour infirmer ou confirmé notre hypothèse concernant la stationnarité du processus qui génère la série  $dDCOVT_t$ , on a besoin de déterminer le retard qui minimise le critère d'Akaïke et Schwarz. Par la suite on estime par la méthode des moindres carrés les trois modèles.

Les hypothèses à tester sont :

$$\begin{cases} H_0 : \text{la tendance n'est pas significative.} \\ H'_0 : \text{la constante n'est pas significative.} \\ H''_0 : \text{l'absence de la racine unitaire.} \end{cases}$$

On commence par l'estimation du modèle [3].

Modèle [3] : "Test sur la tendance"

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.110244	0.0000
Test critical values:		
1% level	-4.252879	
5% level	-3.548490	
10% level	-3.207094	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(DDCOVT)  
 Method: Least Squares  
 Date: 05/14/22 Time: 12:31  
 Sample (adjusted): 1988 2021  
 Included observations: 34 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DDCOVT(-1)	-1.263039	0.177637	-7.110244	0.0000
C	0.585380	2.459472	0.238011	0.8134
@TREND("1986")	-0.018257	0.117508	-0.155365	0.8775

R-squared	0.620553	Mean dependent var	-0.235294
Adjusted R-squared	0.596073	S.D. dependent var	10.57171
S.E. of regression	6.718881	Akaike info criterion	6.731818
Sum squared resid	1399.444	Schwarz criterion	6.866497
Log likelihood	-111.4409	Hannan-Quinn criter.	6.777747
F-statistic	25.34894	Durbin-Watson stat	1.906102
Prob(F-statistic)	0.000000		

FIGURE 6.25 – ADF modèle 3 de la série  $dDCOVT_t$

À la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (t-statistic = -0,155365) est inférieure à toutes les valeurs critique 3.53, 2.79, 2.38 lues dans la table de Dickey-Fuller respectivement aux seuils 1%, 5%, 10%.

## Chapitre 6. Application

---

- La probabilité de nullité du coefficient de la tendance ( $@trend$ ) qui est égale à 0.8775 est supérieure au seuil de 5%.

Donc on accepte  $H_0$

la tendance n'est pas significative.

**Nous passons alors au test du modèle [2]**

**Modèle [2] :** "Test sur la constante"

		t-Statistic	Prob.*	
<hr/>				
Augmented Dickey-Fuller test statistic		-7.229640	0.0000	
Test critical values:	1% level	-3.639407		
	5% level	-2.951125		
	10% level	-2.614300		
<hr/>				
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DDCOVT)				
Method: Least Squares				
Date: 05/14/22 Time: 12:33				
Sample (adjusted): 1988 2021				
Included observations: 34 after adjustments				
<hr/>				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DDCOVT(-1)	-1.263899	0.174822	-7.229640	0.0000
C	0.247962	1.136540	0.218172	0.8287
<hr/>				
R-squared	0.620258	Mean dependent var	-0.235294	
Adjusted R-squared	0.608391	S.D. dependent var	10.57171	
S.E. of regression	6.615640	Akaike info criterion	6.673772	
Sum squared resid	1400.534	Schwarz criterion	6.763558	
Log likelihood	-111.4541	Hannan-Quinn criter.	6.704392	
F-statistic	52.26769	Durbin-Watson stat	1.902947	
Prob(F-statistic)	0.000000			
<hr/>				

FIGURE 6.26 – ADF modèle 2 de la série  $dDCOVT_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student ( $tstatistic = 0,218172$ ) est inférieure à toutes les valeurs critique 3.22, 2.54, 2.17 lues dans la table de Dickey–Fuller respectivement aux seuils 1%, 5%, 10%.

- La probabilité de nullité du coefficient de la constante (c) qui est égale à 0.8287 est supérieure au seuil de 5%.

Donc on accepte  $H_0$

la constante n'est pas significative.

**Nous passons alors au test du modèle [1]**

**Modèle [1] :** "Test sur la racine unitaire"

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.335957	0.0000
Test critical values:		
1% level	-2.634731	
5% level	-1.951000	
10% level	-1.610907	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(DDCOVT)  
 Method: Least Squares  
 Date: 05/14/22 Time: 12:33  
 Sample (adjusted): 1988 2021  
 Included observations: 34 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DDCOVT(-1)	-1.261656	0.171982	-7.335957	0.0000

R-squared	0.619693	Mean dependent var	-0.235294
Adjusted R-squared	0.619693	S.D. dependent var	10.57171
S.E. of regression	6.519475	Akaike info criterion	6.616435
Sum squared resid	1402.617	Schwarz criterion	6.661328
Log likelihood	-111.4794	Hannan-Quinn criter.	6.631745
Durbin-Watson stat	1.904596		

FIGURE 6.27 – ADF modèle 1 de la série  $dDCOVT_t$

la valeur de la t-statistic de test ADF qu'est égale à  $-7,335957$  est inférieure à la différente valeur critique aux différent seuils :  $-2,632688$  ;  $-1,950687$  ;  $-1,611059$  1% ,5% ,10% respectivement (donner par le test ADF ) .

Alors on accepte  $H_0''$  .

La série  $dDCOVT_t$  ne possède pas une racine unitaire .

En conclusion , notre série est stationnaire .Et on peut passer à l'autre étape à savoir l'identification .

### 6.2.3 Identification du modèle :

l'analyse du corrélogramme simple de la série stationnaire ( $dDCOVT_t$ ) (figure 7.24) montre qu'aux retards ( $q=4,5$ ) le termes sont à l'extérieure de l'intervalle de confiance et ce qui concerne le corrélogramme partiel , on remarque que les valeurs de la fonction d'autocorrélation partielles sont élevées aux différents retard ( $p=4$ ) ,ce qui amène à estimer plusieurs modèles parmi eux on a choisit :ARMA(4,0)

D'ou le modèle ARIMA ( $p=4,d=1,q=0$ ) est celui le plus adéquant pour modéliser notre série .

### 6.2.4 Estimation des paramètre du modèle :

le tableau suivant contient l'estimatio du modèle ARMA(4,0) , qui a été choisi parmi d'autre modèles estimés on se basant sur les critères de pouvoir prédictif à savoir :  $R^2$  statistique de Fisher : maximum ; AIC , SC : minimum et sur la statistique de Durbin

## Chapitre 6. Application

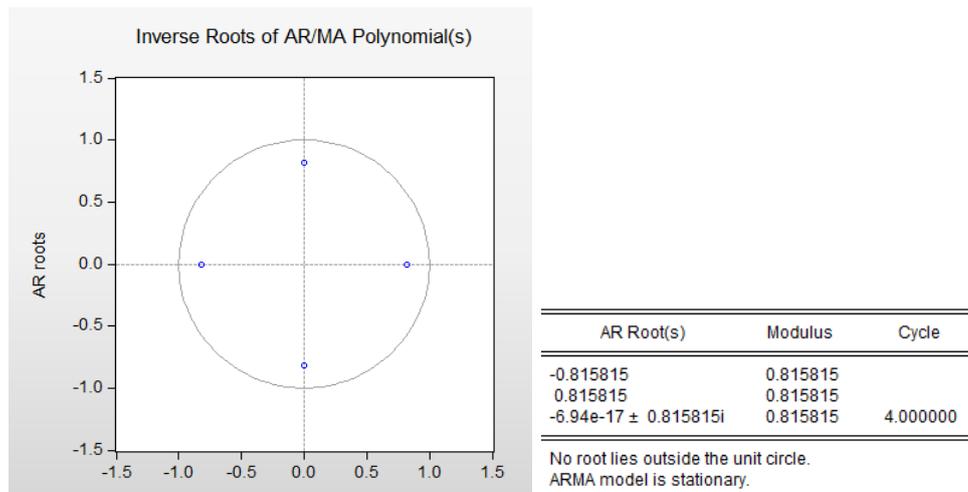
Watson qui égale 2,18 présage un bon ajustement .

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(4)	0.442963	0.129449	3.421923	0.0017
SIGMASQ	34.62967	6.143686	5.636628	0.0000
R-squared	0.192128	Mean dependent var		0.142857
Adjusted R-squared	0.167647	S.D. dependent var		6.642744
S.E. of regression	6.060399	Akaike info criterion		6.521837
Sum squared resid	1212.038	Schwarz criterion		6.610714
Log likelihood	-112.1321	Hannan-Quinn criter.		6.552517
Durbin-Watson stat	2.188728			
Inverted AR Roots	.82	.00-.82i	-.00+.82i	-.82

FIGURE 6.28 – le modèle

### 6.2.5 Validation du modèle :

Test sur les paramètres :



On remarque que :

- Tous les coefficients du modèle sont significative et différent de 0 .
- La probabilité de nullité des composantes AR et MA est inférieur à 5% .
- La statistique de Student des composantes AR et MA en valeurs absolue est supérieur à 1,96 .
- Ces composantes n'ont pas de racine commune .
- Les inverses des racines sont des modules inférieur à 1 .

Et donc ce modèle est retenu , pour cela on va s'assurer avec des tests a l'étape suivante .

Test sur les résidus :

Examen du corrélogramme des résidus :

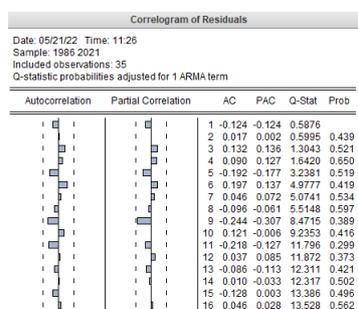


FIGURE 6.29 – corrélogramme des résidus

le corrélogramme des résidus forment un bruit blanc ,car il n'apparait aucun terme en dehors l'intervalle de confiance au seuil 5 % .

Donc,l'estimation du modèle ARIMA(4,1,0) est validée,la série  $dDCOVT_t$  peut etre valablement représentée par un processus ARIMA de type ARIMA(4,0) .

**Ljung-Box test d'autocorrélation :**

le corrélogramme des résidus montre bien que les résidus forment un bruit blanc .Ce qu' est confirmé par la statistique de Ljung-Box (Q-stat) qu'est inférieure à la valeur valeur théorique de  $\chi^2(N-p-q)$  :

- N :le nombre d'oservation.
- p :l'ordre d'autorégressif.
- q :l'ordre de moyenne mobile.

En effet ,

$$Q - stat = 13,528 < \chi^2(34) = 49,765$$

au seuil 5% d'ou la validation du modèle.

**Durbin Watson :**

on test :

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

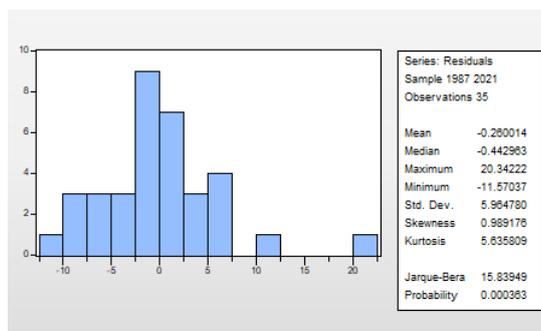
Si DW est proche de 2 on accepte  $H_0$  la statistique  $DW = 2.18 \approx 2$ , on accepte l'hypotèse de non corrélation des résidus

## Chapitre 6. Application

---

### Test de normalité des résidus :

les tests sont effectués à partir des valeurs empiriques des coefficients de SKewness, Kurtosis et la statistique de Jarque- Berra par le logiciel EVIEWS 10.



### Test de Skewness :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_1 = 0 (\text{la distribution des résidus est Symétrique}) \\ H_1 : \gamma_1 \neq 0 (\text{la distribution des résidus est Asymétrique}) \end{cases}$$

$$\gamma_1 = \frac{|\sqrt{\beta_1}|}{\sqrt{\frac{6}{n}}} = \frac{\sqrt{0.989178}}{\sqrt{\frac{6}{35}}} = 2.4 > 1.96$$

On rejette  $H_0$

### Test de Kurtosis :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_2 = 0 (\text{la distribution des résidus est relativement Aplatie}) \\ H_1 : \gamma_2 \neq 0 (\text{la distribution des résidus est n'est pas Aplatie}) \end{cases}$$

$$\gamma_2 = \frac{|\beta_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|5.635809 - 3|}{\sqrt{\frac{24}{35}}} = 3.18 > 1.96$$

On rejette  $H_0$ .

nous confirmons par la statistique de Jarque-Berra

### Test des résidus de Jarque-Berra :

L'hypothèse de test :

$$\begin{cases} H_0 : \text{Les résidus forment un bruit blanc gaussien} \\ H_1 : \text{Les résidus forment un bruit blanc non gaussien} \end{cases}$$

## Chapitre 6. Application

La statistique de Jarque-Berra :

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

$$JB = 15.83949 > \chi_0^2.05(2) = 5.99$$

les residus forment un bruit blanc non gaussien.

Le processus ARIMA(4,1,0) est donc un bruit blanc non gaussien .

**Correlograme de résidus au carré :**

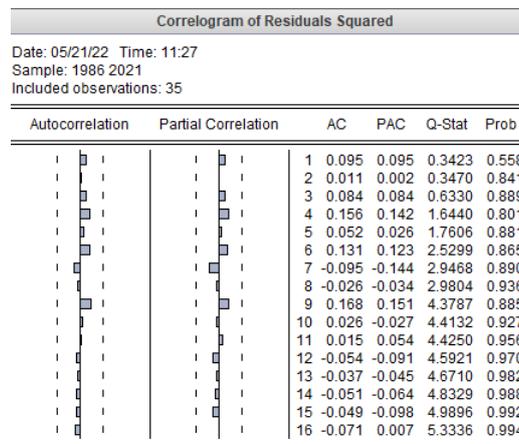


FIGURE 6.30 – corrélograme de résidus au carré

D'après le corrélogramme , nous remarquons que tous les pics sont a l'interieure de la bande de confiance . on rejette l'hypothèse nulle d'hémoscédasticité,il n'existe pas un effet ARCH .

**graphe des séries résiduelle,actuelle et estimée :**

la représentation graphique des séries résiduelle,actuelle et estimée montre que le modèle a bien expliqué la série .

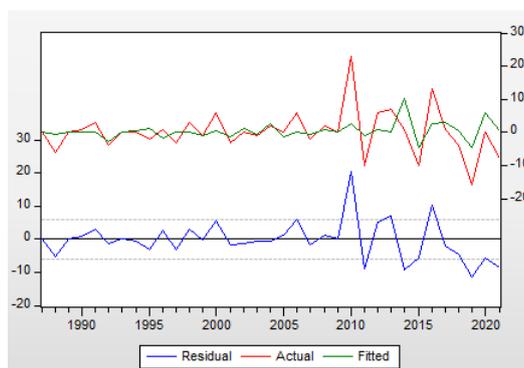


FIGURE 6.31 – Le graphe des séries résiduelle,actuelle et estimée

### 6.2.6 La prévision

Rappelons que la série sous étude "Découverte" comprend 36 observations de 1986 à 2021 donc l'instant est 2021.

On souhaitera prédire 2022 2023 2024 2025 sachant que :

$$dDCOVT_t = DCOVT_t - DCOVT_{t-1}$$

en remplaçant t par t+h on obtient :

$$DCOVT_{t+h} = DCOVT_{t+h-1} + 0.442DCOVT_{t+h-4} + \epsilon_{t+h}$$

avec h=1,2,3,4

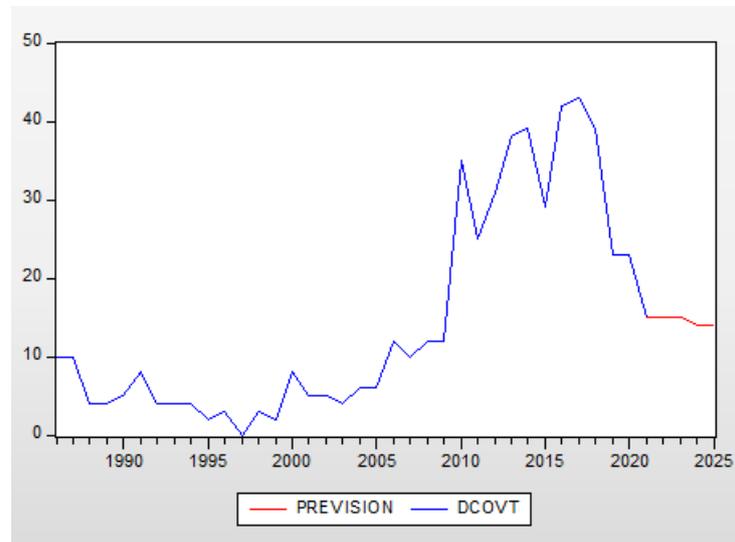
On obtient les prévisions suivants :

années	Découverte
2022	15
2023	15
2024	14
2025	14

## Chapitre 6. Application

---

Le graphe des prévision est donné par :



### 6.3 Étude de la série du $INVES_t$

#### 6.3.1 Identification :

Nous disposons d'une série annuel évoluant de 1986 à 2021 , ou toutes les variables sont en une série de 36 observations . Elles représente l'évolution annuelle des investissements financiers pour l'exploration pétrolière .

Notation :

Nous notons la série Investissement  $INVES_t$

#### Représentation graphique de la série $INVES_t$

Pour avoir une idée globale mais non décisive sur la nature et les caractéristiques du processus (tendance, saisonnalité,..), il est nécessaire d'analyser le graphe représentant l'évolution de la série en question. (Figure 7.32)

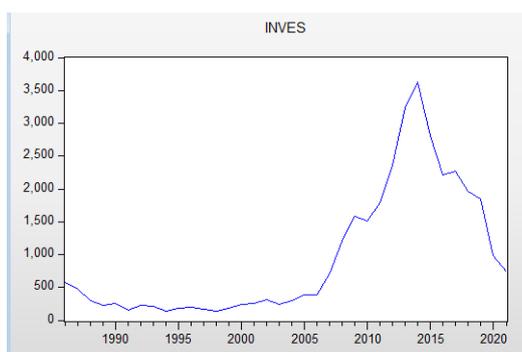


FIGURE 6.32 – le graphe de la série  $INVES_t$

D'après la représentation graphique de la série brute ( $INVES_t$ ) (Figure 7.32) nous remarquons : une non stationnarité moyenne, qui se traduit par une légère tendance par d'autre descendants .

#### Examen du corrélogramme de la série $INVES_t$ :

Le logiciel économique Eviews fournit les résultats des fonctions d'autocorrelation simple (colonne AC) et partielle (colonne PAC), ainsi que leurs corrélogrammes associées. Les bornes de l'intervalle de confiance sont stylisées par des pointillés horizontaux ; tous les termes qui sont à l'intérieure de cet intervalle sont significativement nuls et ceux qui sortent de ce dernier sont significativement différent de zéro au seuil de 5% .

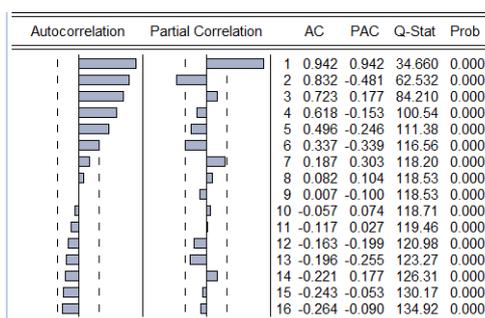


FIGURE 6.33 – le corrélogramme  $INVES_t$

L'analyse du corrélogramme simple et partielle de la série brute (figure 7.33) nous indique une non stationnarité de la série. En effet, la fonction d'autocorrélation simple et partielle diminue lentement vers zéro, on remarque aussi de nombreux pics sortant de la bande de confiance .

L'hypothèse du test est :

$$\begin{cases} H_0 : \text{la série stationnaire} \\ H_1 : \text{la série n'est pas stationnaire} \end{cases}$$

Si les probabilités  $< 5\%$  on rejette  $H_0$  Sinon on accepte  $H_0$

Nous constatons que la série n'est pas stationnaire puisque tous les probabilités  $< 5\%$  .

On rejette  $H_0$  .

la série  $INVES_t$  semble être générée par un processus non stationnaire .

Afin de confirmer ou infirmer la non stationnarité de la série , nous nous proposons d'appliquer le test de DICKEY-FULLER Augmenté sur la série  $INVES_t$  .

### Test de DICKEY-FULLER Augmenté

l'application du test de DICKEY-FULLER Augmenté nécessite la sélection de nombre du retard  $p$  .

Les résultats sont regroupés dans le tableau ci-dessous où la constante est notée  $C$  et la tendance @TREND. Au début, nous avons effectué un test sur la signification de la tendance, en se référant aux tables de Dickey-Fuller.

Les hypothèses à tester sont :

$$\begin{cases} H_0 : \text{la tendance n'est pas significative .} \\ H'_0 : \text{la constante n'est pas significative .} \\ H''_0 : \text{l'absence de la racine unitaire .} \end{cases}$$

On commence par l'estimation du modèle [3] .

Modèle [3] : "Test sur la tendance"

$$\Delta INVES_t = \phi INVES_{t-1} + Bt + C + \epsilon_t; \text{Ou } \epsilon_t \text{ est un processus stationnaire.}$$

## Chapitre 6. Application

Nous utilisons le test programmé sous le logiciel Eviews.

Le résultat de l'affichage pour la série  $INVES_t$  est donnée dans le tableau suivant :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INVES(-1)	-0.376805	0.114617	-3.287505	0.0034
D(INVES(-1))	0.596827	0.176732	3.377023	0.0027
D(INVES(-2))	0.052464	0.237239	0.221144	0.8270
D(INVES(-3))	0.244314	0.262773	0.929753	0.3626
D(INVES(-4))	0.163225	0.249665	0.653774	0.5200
D(INVES(-5))	0.591435	0.236456	2.501242	0.0203
C	-166.3476	165.5278	-1.004953	0.3258
@TREND("1986")	25.02415	11.85773	2.110366	0.0464
R-squared	0.577717	Mean dependent var		19.40000
Adjusted R-squared	0.443355	S.D. dependent var		361.1877
S.E. of regression	269.4773	Akaike info criterion		14.25402
Sum squared resid	1597597.	Schwarz criterion		14.62768
Log likelihood	-205.8104	Hannan-Quinn criter.		14.37356
F-statistic	4.299688	Durbin-Watson stat		1.665859
Prob(F-statistic)	0.003919			

FIGURE 6.34 – ADF modèle 3 de la série  $INVES_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (t-statistic = 2.110366) est inférieure à toutes les valeurs critique 3.53, 2.79, 2.38 lues dans la table de Dickey–Fuller respectivement aux seuils 1%, 5%, 10%.

Donc on accepte  $H_0$

la tendance n'est pas significative.

**Nous passons alors au test du modèle [2]**

**Modèle [2] :** "Test sur la constante"

$$\Delta INVES_t = \phi INVES_{t-1} + C + \epsilon_t; \text{Ou } \epsilon_t \text{ est un processus stationnaire.}$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INVES(-1)	-0.085515	0.051445	-1.662248	0.1065
D(INVES(-1))	0.511659	0.155172	3.297375	0.0025
C	84.87547	71.40900	1.188582	0.2436
R-squared	0.282398	Mean dependent var		7.588235
Adjusted R-squared	0.236101	S.D. dependent var		341.1067
S.E. of regression	298.1318	Akaike info criterion		14.31705
Sum squared resid	2755360.	Schwarz criterion		14.45172
Log likelihood	-240.3898	Hannan-Quinn criter.		14.36298
F-statistic	6.099707	Durbin-Watson stat		1.844212
Prob(F-statistic)	0.005837			

FIGURE 6.35 – ADF modèle 2 de la série  $INVES_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (tstatistic = 1.188582) est inférieure à toutes les valeurs critique 3.22, 2.54, 2.17 lues dans la table de Dickey–Fuller respectivement aux seuils 1%, 5%, 10%.

- La probabilité de nullité du coefficient de la constante (c) qui est égale à 0.2436 est supérieure au seuil de 5%.

## Chapitre 6. Application

Donc on accepte  $H'_0$

la constante n'est pas significative.

Nous passons alors au test du modèle [1]

Modèle [1] : "Test sur la racine unitaire"

$$\Delta INVES_t = \phi INVES_{t-1} + \epsilon_t; \text{Ou } \epsilon_t \text{ est un processus stationnaire.}$$

	t-Statistic	Prob.*		
Augmented Dickey-Fuller test statistic	-1.155316	0.2211		
Test critical values:				
1% level	-2.634731			
5% level	-1.951000			
10% level	-1.610907			
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(INVES)				
Method: Least Squares				
Date: 06/01/22 Time: 12:21				
Sample (adjusted): 1988 2021				
Included observations: 34 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
INVES(-1)	-0.042856	0.037095	-1.155316	0.2565
D(INVES(-1))	0.495752	0.155587	3.186331	0.0032
R-squared	0.249695	Mean dependent var	7.588235	
Adjusted R-squared	0.226248	S.D. dependent var	341.1067	
S.E. of regression	300.0483	Akaike info criterion	14.30279	
Sum squared resid	2880927.	Schwarz criterion	14.39257	

FIGURE 6.36 – ADF modèle 1 de la série  $INVES_t$

la valeur de la t-statistic de test ADF qu'est égale à -1.155316 est supérieur à la différente valeur critique aux différent seuils : -2,634731 ; -1,951000 ; -1,6108907 1% ,5% ,10% respectivement (donner par le test ADF )

Donc on rejette  $H''_0$  .

La série  $INVES_t$  possède une racine unitaire et le processus qui génère cette série est non stationnaire de type DS sans dérivé .

**Stationarisation de la série  $INVES_t$  :**

On procède à la différentiation ordinaire d'ordre 1 . Sous EVIEWS on insere la nouvelle série engendrée par la différentiation première :

$$dINVES_t = INVES_t - INVES_{t-1}$$

**Le graphe de la série  $dINVES_t$**

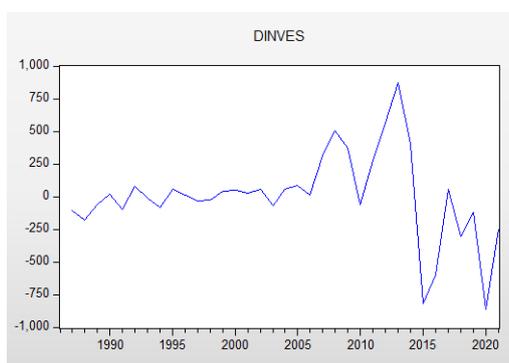


FIGURE 6.37 – graphe de la série  $dINVES_t$

Nous remarquons que l'effet de la tendance a été absorbé.

**Analyse du corrélogramme de la série  $dINVES_t$**

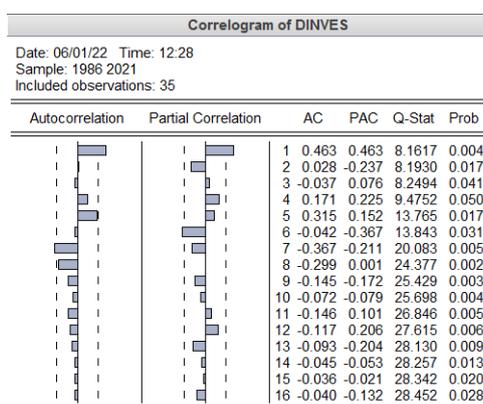


FIGURE 6.38 – corrélogramme de la série  $dINVES_t$

- L'examen du corrélogramme de la série (figure 7.38) montre que la série représente un bruit blanc, car tous les termes ne sont pas significativement différents de zéro, on peut en conclure que le processus étudié est sans mémoire.

la série peut être stationnaire, pour confirmer ou infirmer cette hypothèse, il convient d'appliquer à la série  $dINVES_t$  le test de la racine unitaire DICKEY-FULLER Augmenté.

### 6.3.2 Test de la stationnarité de la série $dINVES_t$

**Test de DICKEY-FULLER Augmenté sur la série  $dINVES_t$  :**

On applique le test de DFA pour infirmer ou confirmé notre hypothèse concernant la stationnarité du processus qui génère la série  $dINVES_t$ , on a besoin de déterminer le retard qui minimise le critère d'Akaike et Schwarz.

Par la suite on estime par la méthode des moindres carrés les trois modèles.

## Chapitre 6. Application

Les hypothèses à tester sont :

$$\begin{cases} H_0 : \text{la tendance n'est pas significative .} \\ H'_0 : \text{la constante n'est pas significative .} \\ H''_0 : \text{l'absence de la racine unitaire .} \end{cases}$$

On commence par l'estimation du modèle [3] .

Modèle [3] : "Test sur la tendance"

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DINVES(-1)	-2.418296	0.627071	-3.856493	0.0011
D(DINVES(-1))	2.037735	0.608646	3.347983	0.0034
D(DINVES(-2))	1.692727	0.626633	2.701306	0.0141
D(DINVES(-3))	1.978303	0.607991	3.253837	0.0042
D(DINVES(-4))	1.291700	0.538689	2.397860	0.0269
D(DINVES(-5))	1.887413	0.424164	4.449721	0.0003
D(DINVES(-6))	1.190613	0.490984	2.424954	0.0254
C	-409.8685	202.5748	-2.023294	0.0573
@TREND("1986")	28.46843	11.30150	2.518994	0.0209
R-squared	0.711668	Mean dependent var	-8.303571	
Adjusted R-squared	0.590264	S.D. dependent var	384.0086	
S.E. of regression	245.8062	Akaike info criterion	14.10206	
Sum squared resid	1147993.	Schwarz criterion	14.53026	
Log likelihood	-188.4288	Hannan-Quinn criter.	14.23296	
F-statistic	5.862019	Durbin-Watson stat	2.149738	
Prob(F-statistic)	0.000754			

FIGURE 6.39 – ADF modèle 3 de la série  $\text{DINVES}_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student (t-statistic = 2.518994) est inférieure à toutes les valeurs critiques 3.53, 2.79 lues dans la table de Dickey-Fuller respectivement aux seuils 1%, 5%.

Donc on accepte  $H_0$

la tendance n'est pas significative.

Nous passons alors au test du modèle [2]

Modèle [2] : "Test sur la constante"

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DINVES(-1)	-0.945391	0.324792	-2.910763	0.0081
D(DINVES(-1))	0.666152	0.367313	1.813580	0.0834
D(DINVES(-2))	0.190924	0.294193	0.648976	0.5231
D(DINVES(-3))	0.572825	0.342437	1.672789	0.1085
D(DINVES(-4))	-0.009015	0.215670	-0.041799	0.9670
D(DINVES(-5))	0.999744	0.282786	3.535336	0.0019
C	56.16223	55.62348	1.009686	0.3236
R-squared	0.600661	Mean dependent var	-11.06897	
Adjusted R-squared	0.491751	S.D. dependent var	377.3829	
S.E. of regression	269.0423	Akaike info criterion	14.23412	
Sum squared resid	1592442.	Schwarz criterion	14.56416	
Log likelihood	-199.3947	Hannan-Quinn criter.	14.33748	
F-statistic	5.515184	Durbin-Watson stat	2.131040	
Prob(F-statistic)	0.001294			

FIGURE 6.40 – ADF modèle 2 de la série  $\text{DINVES}_t$

## Chapitre 6. Application

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student ( $t_{\text{statistic}} = 1.009686$ ) est inférieure à toutes les valeurs critiques 3.22, 2.54 lues dans la table de Dickey–Fuller respectivement aux seuils 1%, 5%.
- La probabilité de nullité du coefficient de la constante ( $c$ ) qui est égale à 0.3236 est supérieure au seuil de 5%.

Donc on accepte  $H_0'$

la constante n'est pas significative.

Nous passons alors au test du modèle [1]

Modèle [1] : "Test sur la racine unitaire"

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.417268	0.0012
Test critical values:		
1% level	-2.634731	
5% level	-1.951000	
10% level	-1.610907	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(DINVES)  
 Method: Least Squares  
 Date: 06/01/22 Time: 12:33  
 Sample (adjusted): 1988 2021  
 Included observations: 34 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DINVES(-1)	-0.529200	0.154860	-3.417268	0.0017

R-squared	0.261271	Mean dependent var	-4.147059
Adjusted R-squared	0.261271	S.D. dependent var	350.8850
S.E. of regression	301.5663	Akaike info criterion	14.28483
Sum squared resid	3001093.	Schwarz criterion	14.32972
Log likelihood	-241.8421	Hannan-Quinn criter.	14.30014

FIGURE 6.41 – ADF modèle 1 de la série  $dINVES_t$

la valeur de la t-statistic de test ADF qu'est égale à -3.417268 est inférieure à la différente valeur critique aux différents seuils : -2,634731 ; -1,951000 ; -1,610907 1% ,5% ,10% respectivement (donner par le test ADF) .

Alors on accepte  $H_0''$ .

La série  $dINVES_t$  ne possède pas une racine unitaire .

En conclusion , notre série est stationnaire .

Et on peut passer à l'autre étape à savoir l'identification .

### 6.3.3 Identification du modèle :

l'analyse du corrélogramme simple de la série stationnaire ( $dINVES_t$ )(figure 7.38) montre qu'aux retards ( $q=1,7$ ) les termes sont à l'extérieur de l'intervalle de confiance et ce qui concerne le corrélogramme partiel, on remarque que les valeurs de la fonction d'autocorrélation partielles sont élevées aux différents retards ( $p=1,6$ ) ,ce qui amène à estimer plusieurs modèles parmi eux on a choisit :ARMA(1,7)

D'où le modèle ARIMA ( $p=1,d=1,q=7$ ) est celui le plus adéquat pour modéliser notre série .

### 6.3.4 Estimation des paramètre du modèle

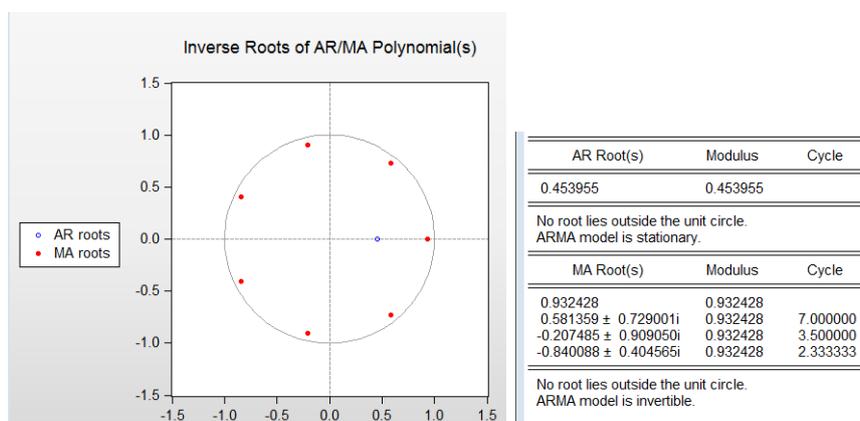
le tableau suivant contient l'estimatio du modèle ARMA(1,7) , qui a été choisi parmi d'autre modèles estimés on se basant sur les critères de pouvoir prédictif à savoir :  $R^2$  statistique de Fisher : maximum ; AIC , SC : minimum et sur la statistique de Durbin Watson qui égale 1.79 présage un bon ajustement .

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.453955	0.105337	4.309558	0.0001
MA(7)	-0.612781	0.213027	-2.876538	0.0071
SIGMASQ	62219.06	12682.82	4.905774	0.0000
R-squared	0.434664	Mean dependent var	4.371429	
Adjusted R-squared	0.399331	S.D. dependent var	336.5914	
S.E. of regression	260.8680	Akaike info criterion	14.14776	
Sum squared resid	2177667.	Schwarz criterion	14.28107	
Log likelihood	-244.5858	Hannan-Quinn criter.	14.19378	
Durbin-Watson stat	1.796294			
Inverted AR Roots	.45			
Inverted MA Roots	.93	.58+ .73i	.58- .73i	-.21-.91i
	-.21+ .91i	-.84-.40i	-.84+.40i	

FIGURE 6.42 – le modèle

### 6.3.5 Validation du modèle :

Test sur les paramètres :



On remarque que :

- Tous les coefficients du modèle sont significative et différent de 0 .
- La probabilité de nullité des composantes AR et MA est inférieur à 5% .
- La statistique de Student des composantes AR et MA en valeurs absolue est supérieur à 1,96 .
- Ces composantes n'ont pas de racine commune .
- Les inverses des racines sont des modules inférieur à 1 .

Et donc ce modèle est retenu , pour cela on va s'assurer avec des tests a l'étape suivante .

Test sur les résidus :

Examen du corrélogramme des résidu :

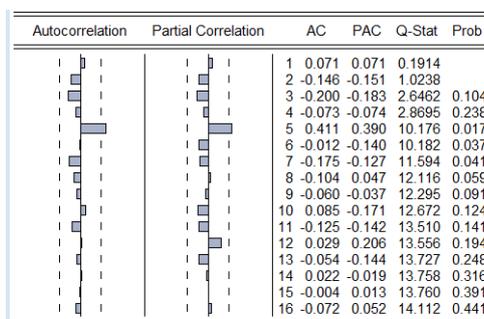


FIGURE 6.43 – corrélogramme des résidus

le corrélogramme des résidus forment un bruit blanc ,car il n'apparaît aucun terme en dehors l'intervalle de confiance au seuil 5 % .

Donc,l'estimation du modèle ARIMA(1,1,7) est validée, la série  $dINVES_t$  peut être valablement représentée par un processus ARIMA de type ARIMA(1,7) .

**Ljung-Box test d'autocorrélation :**

le corrélogramme des résidus montre bien que les résidus forment un bruit blanc .Ce qu' est confirmé par la statistique de Ljung-Box • N :le nombre d'observation.

- p :l'ordre d'autorégressif.
- q :l'ordre de moyenne mobile.

En effet ,

$$Q - stat = 14.112 < \chi^2_{(34)} = 49,765$$

au seuil 5% d'ou la validation du modèle.

**Durbin Watson :**

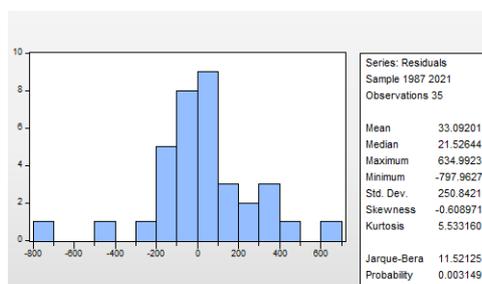
on test :

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

Si DW est proche de 2 on accepte  $H_0$  la statistique  $DW = 1.79 \approx 2$ , on accepte l'hypotèse de non corrélation des résidus

### Test de normalité des résidus :

les tests sont effectués à partir des valeurs empiriques des coefficients de SKewness, Kurtosis et la statistique de Jarque- Berra par le logiciel EVIEWS 10.



### Test de Skewness :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_1 = 0 \text{ (la distribution des résidus est symétrique )} \\ H_1 : \gamma_1 \neq 0 \text{ (la distribution des résidus est n'est pas symétrique)} \end{cases}$$

$$\gamma_1 = \frac{|\sqrt{\beta_1}|}{\sqrt{\frac{6}{n}}} = \frac{\sqrt{0.608971}}{\sqrt{\frac{6}{35}}} = 1.88 < 1.96$$

On accepte  $H_0$

### Test de Kurtosis :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_2 = 0 \text{ (la distribution des résidus est relativement Aplatie )} \\ H_1 : \gamma_2 \neq 0 \text{ (la distribution des résidus est n'est pas Aplatie)} \end{cases}$$

$$\gamma_2 = \frac{|\beta_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|5.53316 - 3|}{\sqrt{\frac{24}{34}}} = 3.01 > 1.96$$

On rejette  $H_0$  .

nous confirmons par la statistique de Jarque-Berra

### Test des résidus de Jarque-Berra :

L'hypothèse de test :

$$\begin{cases} H_0 : \text{Les résidus forment un bruit blanc gaussien} \\ H_1 : \text{Les résidus forment un bruit blanc non gaussien} \end{cases}$$

La statistique de Jarque-Berra :

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

$$JB = 11.52125 > \chi_0.05^2(2) = 5.99$$

## Chapitre 6. Application

les résidus forment un bruit blanc non gaussien.

Le processus ARIMA(1, 1, 7) est donc un bruit blanc non gaussien .

**correlogramme de résidus au carré :**

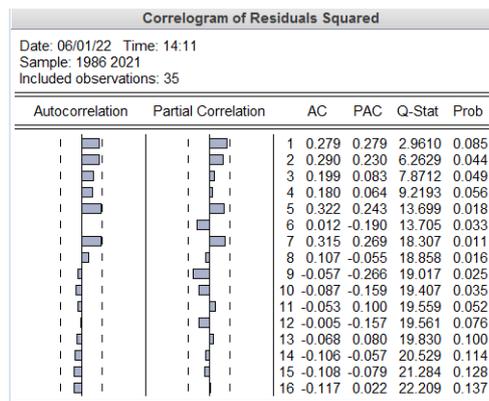


FIGURE 6.44 – correlogramme de résidus au carré

D'après le correlogramme , nous remarquons que tous les pics sont a l'interieure de la bande de confiance .

donc on jettère l'hypothèse nulle d'hémoscédasticité,il n'existe pas un effet ARCH

**graphe des séries résiduelle,actuelle et estimée :**

la représentation graphique des séries résiduelle,actuelle et estimée montre que le modèle a bien expliqué la série .

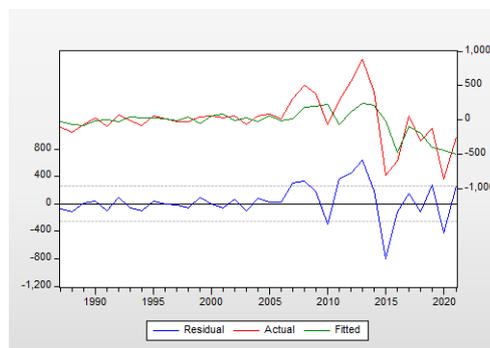


FIGURE 6.45 – Le graphe des séries résiduelle,actuelle et estimée

### 6.3.6 La prévision

Rappelons que la série sous étude "Investissement" comprend 36 observations de 1986 à 2021 donc l'instant est 2021.

on souhaitera prédire 2022 2023 2024 2025 sachant que :

$$dINVES_t = INVES_t - INVES_{t-1}$$

en remplaçant t par t+h on obtient :

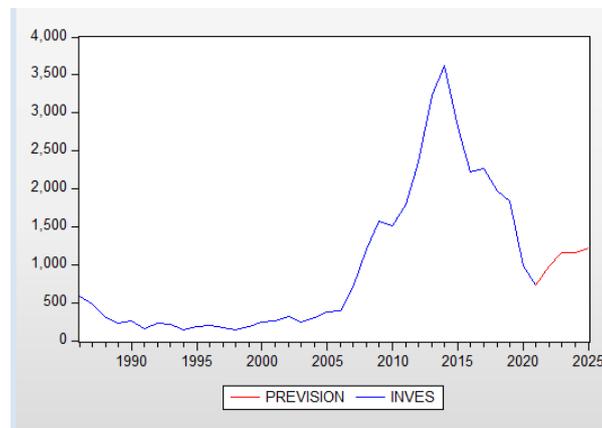
$$INVES_{t+h} = INVES_{t+h-1} + 0.453955INVES_{t+h-1} + \epsilon_{t+h} - 0.612781\epsilon_{t+h-7}$$

avec  $h=1, \dots, 7$

on obtient les prévisions suivants :

années	Investissement
2022	983.722
2023	1165.6188
2024	1157.1391
2025	1222.5659

Le graphe des prévision est donné par :



### Application du Modèle VAR

Suite a l'application de la méthodologie de Box Jenkins sur les séries Prix de pétrole et Découverte,celle-ci sont rendues stationnaires, nous pouvons donc appliquer la modélisation VAR .

La méthodologie des vecteurs autorégressifs VAR consiste à modéliser un ensemble de séries chronologiques à l'aide d'un processus AR(p) vectoriel .

En pratique on dispose souvent d'observations simultanées de plusieurs séries chronologiques qui nous permettent d'étudier les liaisons existant entre elles et de voir comment celles-ci évoluent dans le temps.

#### 6.4 Etude multivariée des séries( $dPRIX_t, dDCOVT_t$ )

Dans cette partie nous analysons la série chronologique multivariée  $X_t$  qui couvre la période allant de 1986 à 2021. Elle est composée de deux séries .

$PRIX_t$  : le prix du brut du pétrole en Algérie .

$DCOVT_t$  : le nombres de découverte (sismique et forage) en Algérie .

Après avoir stationnarisé les variables,nous construisons un modèle VAR à deux variables et à p décalages noté :

$X_t = ( dPRIX_t , dDCOVT_t )$  ;  $X_t$  est un vecteur de dimension (2,1) .

$\epsilon_t = ( \epsilon_{1t} ; \epsilon_{2t} )'$  ;est un vecteur résiduel associée à  $X_t$  de matrice de variance-covariance  $\sigma$  .

**Représentation graphique :**

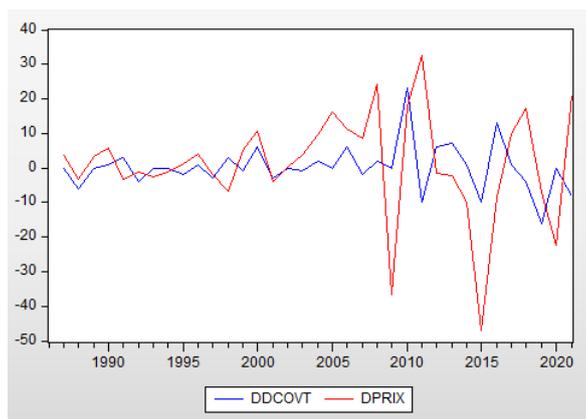


FIGURE 6.46 – Graphe de dPRIX et dDCOVT

### 6.4.1 Identification du modèle VAR(P)

Avant tout traitement, il convient de déterminer le retard  $p$  optimal : il s'agit de celui qui minimise les critères d'Akaike (AIC) et Schwartz (SC).

A cette fin, nous avons estimé divers processus VAR pour des ordres de retards  $p$  allant de 0 à 9. Pour chaque modèle, nous avons calculé les critères d'informations précédemment cités.

Le tableau ci-dessous reporte les résultats obtenus.

Lag	LogL	LR	FPE	AIC	SC	HQ
1	-194.1763	NA	14316.78	15.24433	15.43789	15.30007
2	-190.3920	6.404225	14619.25	15.26092	15.64803	15.37240
3	-188.4380	3.006193	17317.66	15.41831	15.99897	15.58552
4	-181.7480	9.262989	14426.79	15.21139	15.98560	15.43433
5	-170.9727	13.26201	8936.052	14.69020	15.65797	14.96889
6	-158.8019	13.10697*	5099.176*	14.06168*	15.22300*	14.39610*
7	-155.0232	3.488016	5750.569	14.07871	15.43358	14.46886
8	-153.1578	1.434918	7909.230	14.24291	15.79134	14.68880
9	-152.8106	0.213678	13205.69	14.52389	16.26587	15.02552

\* indicates lag order selected by the criterion  
 LR: sequential modified LR test statistic (each test at 5% level)  
 FPE: Final prediction error  
 AIC: Akaike information criterion  
 SC: Schwarz information criterion  
 HQ: Hannan-Quinn information criterion

La comparaison des modèles suivant les critères AIC et SC montre que le modèle VAR(6) pourrait peut être convenir.

Le nombre de décalages retenu correspond à la valeur la plus faible des critères soit :  $p=6$

### 6.4.2 Estimation des paramètres du modèle VAR(6)

**Estimation du modèle VAR(6) avec constante :**

Le modèle VAR(6) avec constante s'écrit sous la forme suivante :

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varphi_3 X_{t-3} + \varphi_4 X_{t-4} + \varphi_5 X_{t-5} + \varphi_6 X_{t-6}$$

ou  $\varphi_0$  représente l'estimation de la constante et  $\varphi_p$  ( $p = 6$ ) sont des matrices carrées d'ordre 2. On effectue l'estimation en utilisant le logiciel Eviews 10.0.

Le tableau contient deux colonnes représentant le nombre de variables du modèle VAR

-Ce tableau peut être décomposé en deux blocs, chaque bloc est associé à une série

-Chaque bloc contient  $p = 1$  lignes

-La ligne ( $i=1$ ) d'un bloc précis correspond à la série associée à l'instant  $t - i$

-Chaque ligne contient les coefficients au retard  $i$  (donnés en haut), ainsi que les t-statistiques associées (donnée en bas bas entre crochets).

Nous obtenons le tableau présenté ci-dessous :

A la lecture du tableau on constate que la constante est non significative aux seuils 1%, 5%, 10% puisque les t-statistiques (données par les valeurs entre crochets) sont inférieures aux différentes valeurs critiques.

## Chapitre 6. Application

	DDCOVT	DPRX
DDCOVT(-1)	-0.175420 (0.20591) [-0.85193]	0.757500 (0.96366) [0.78606]
DDCOVT(-2)	-0.279826 (0.20124) [-1.39054]	0.289444 (0.94180) [0.30733]
DDCOVT(-3)	-0.277125 (0.17415) [-1.59127]	-2.153976 (0.81505) [-2.64276]
DDCOVT(-4)	0.048065 (0.20150) [0.23854]	-0.127760 (0.94303) [-0.13548]
DDCOVT(-5)	-0.143296 (0.19268) [-0.74371]	-0.971196 (0.90174) [-1.07702]
DDCOVT(-6)	-0.169662 (0.18236) [-0.93038]	-0.757171 (0.85345) [-0.88719]
DPRX(-1)	-0.174151 (0.06491) [-2.68294]	0.014662 (0.30378) [0.04826]
DPRX(-2)	0.273091 (0.06635) [4.11599]	-0.088239 (0.31052) [-0.28417]
DPRX(-3)	-0.073986 (0.07073) [-1.04464]	0.204031 (0.33101) [0.61880]
DPRX(-4)	0.162453 (0.07295) [2.22698]	-0.648197 (0.34140) [-1.89279]
DPRX(-5)	0.169197 (0.08019) [2.10982]	0.603377 (0.37532) [1.60765]
DPRX(-6)	0.320630 (0.08560) [3.74577]	-0.147377 (0.40060) [-0.36789]
C	0.273534 (0.72607) [0.37673]	5.514062 (3.39806) [1.62271]
R-squared	0.884339	0.534739
Adj. R-squared	0.797593	0.185793
Sum sq. resid	165.9541	3634.889
S.E. equation	3.220579	15.07251
F-statistic	10.19457	1.532440
Log likelihood	-56.44324	-111.1993
Akaike AIC	5.478844	8.565467
Schwarz SC	6.091770	9.178393
Mean dependent	0.379310	1.484626
S.D. dependent	7.158480	16.70391
Determinant resid covariance (dof adj.)		1926.984
Determinant resid covariance		596.5731
Log likelihood		-174.7257
Akaike information criterion		13.84315
Schwarz criterion		15.06901
Number of coefficients		26

Donc on réestime le modèle sans constante

Estimation du modèle VAR(6) sans constante :

	DDCOVT	DPRIX
DDCOVT(-1)	-0.190138 (0.20027) [-0.89947]	0.682372 (1.00702) [0.65775]
DDCOVT(-2)	-0.279527 (0.19609) [-1.42550]	0.295471 (0.98599) [0.29967]
DDCOVT(-3)	-0.262929 (0.16568) [-1.58696]	-1.987817 (0.83309) [-2.24204]
DDCOVT(-4)	0.075672 (0.18290) [0.41373]	0.428761 (0.91969) [0.46620]
DDCOVT(-5)	-0.122113 (0.17958) [-0.67999]	-0.544181 (0.90297) [-0.60266]
DDCOVT(-6)	-0.159949 (0.17591) [-0.90926]	-0.581371 (0.88453) [-0.63465]
DPRIX(-1)	-0.169439 (0.06207) [-2.72999]	0.109638 (0.31208) [0.35131]
DPRIX(-2)	0.272947 (0.06465) [4.22183]	-0.091131 (0.32508) [-0.28033]
DPRIX(-3)	-0.072796 (0.06886) [-1.05712]	0.226805 (0.34626) [0.65502]
DPRIX(-4)	0.162133 (0.07108) [2.28106]	-0.652651 (0.35740) [-1.82612]
DPRIX(-5)	0.171540 (0.07791) [2.20179]	0.650611 (0.39175) [1.66079]
DPRIX(-6)	0.320128 (0.08340) [3.83848]	-0.157503 (0.41936) [-0.37558]
R-squared	0.883313	0.459169
Adj. R-squared	0.807809	0.107572
Sum sq. resids	167.4262	4233.097
S.E. equation	3.138248	15.77991
F-statistic	11.93995	1.306627
Log likelihood	-66.57129	-113.4084
Akaike AIC	5.418710	8.648857
Schwarz SC	5.984488	9.214634
Mean dependent	0.379310	1.484626
S.D. dependent	7.158480	16.70391
Determinant resid covariance (dof adj.)	2000.890	587.5826
Determinant resid covariance		-177.0296
Log likelihood		13.86411
Akaike information criterion		14.99566
Schwarz criterion		24
Number of coefficients		

FIGURE 6.47

A partir de la table d'estimation on obtient l'équation suivante :

$$X_t = \begin{pmatrix} -0.18 & -0.17 \\ 0.66 & 0.11 \end{pmatrix} X_{t-1} + \begin{pmatrix} -0.28 & 0.27 \\ 0.30 & -0.09 \end{pmatrix} X_{t-2} + \begin{pmatrix} -0.26 & -0.07 \\ -1.87 & 0.22 \end{pmatrix} X_{t-3} + \\ \begin{pmatrix} 0.08 & 0.16 \\ 0.4 & -0.65 \end{pmatrix} X_{t-4} + \begin{pmatrix} -0.12 & 0.17 \\ -0.54 & 0.65 \end{pmatrix} X_{t-5} + \begin{pmatrix} -0.16 & 0.32 \\ -0.56 & -0.16 \end{pmatrix} X_{t-6} + \epsilon_t$$

Bien entendu cette écriture du modèle subira un remaniement après l'épreuve des tests suite à l'étape de validation.

Nous obtenons également la matrice suivante de variance covariance :

$$= \begin{pmatrix} 9.84 & 21.24 \\ 21.24 & 249.005 \end{pmatrix}$$

### 6.4.3 Validation

**Test sur les racines :**

Les racines des polynômes autorégressifs des séries sont supérieures en module à 1, car leurs inverses calculés par Eviews sont tous inférieures à 1, ainsi les conditions de stationnarité et d'inversibilité sont vérifiées.

## Chapitre 6. Application

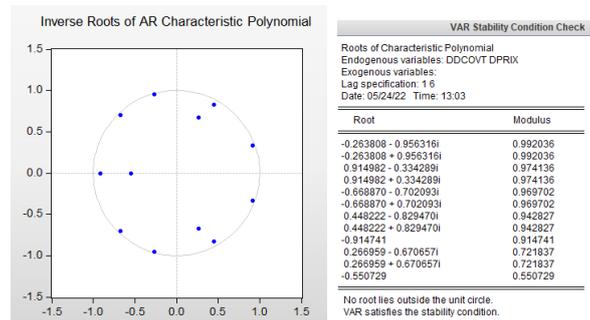


FIGURE 6.48

### Test sur les résidus :

De la même façon que la méthodologie de Box Jenkins, il convient de vérifier si les résidus forment un bruit blanc, une observation des corrélogrammes des résidus des deux séries s'impose.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.235	-0.235	1.7676	0.184
		2	0.001	-0.057	1.7676	0.413
		3	0.012	-0.002	1.7723	0.621
		4	-0.056	-0.057	1.8843	0.757
		5	-0.208	-0.250	3.5111	0.622
		6	0.038	-0.089	3.5676	0.735
		7	0.169	0.164	4.7337	0.692
		8	-0.010	0.081	4.7377	0.785
		9	-0.249	-0.314	7.5139	0.584
		10	0.207	0.015	9.5325	0.482
		11	-0.141	-0.041	10.522	0.484
		12	0.078	0.151	10.843	0.542

FIGURE 6.49

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.127	-0.127	0.5156	0.473
		2	0.054	0.038	0.6124	0.736
		3	-0.142	-0.133	1.3090	0.727
		4	0.016	-0.019	1.3183	0.858
		5	0.059	0.072	1.4506	0.919
		6	-0.187	-0.198	2.8211	0.831
		7	-0.147	-0.211	3.7089	0.813
		8	-0.188	-0.225	5.2208	0.734
		9	0.277	0.201	8.6833	0.469
		10	-0.324	-0.373	13.639	0.190
		11	0.125	-0.002	14.422	0.211
		12	-0.227	-0.240	17.142	0.144

FIGURE 6.50

### Corrélogramme des résidus croisés entre ( $dPRIX_t$ et $dDCOVT_t$ ) :

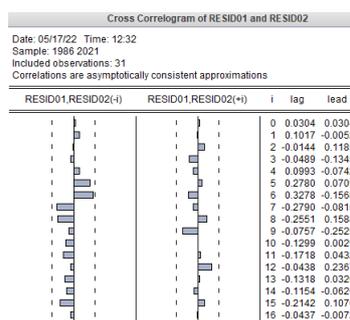


FIGURE 6.51

Sur la base de l'étude des corrélogrammes simple et croisé des résidus, on déduit que les résidus associés au modèle VAR(6) forment un vecteur bruit blanc.

### Graphe des résidus :

La représentation graphique des résidus est donnée comme suit :

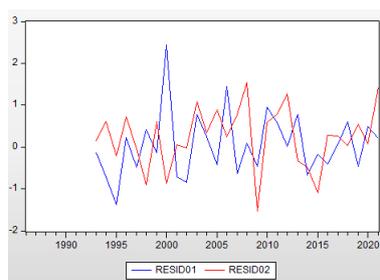


FIGURE 6.52

### Test sur les estimations :

En tenant compte du tableau des estimations et le fait qu'un coefficient est significativement différent de zéro au seuil 5% , si la t-stat en valeur absolue associée est supérieure à 1.96, et en vertu des résultats des tests précédents, le modèle VAR(6) est validé et il s'écrit de la façon suivante :

$$dDCOVT_t = -0.169dPRIX_{t-1} + 0.272dPRIX_{t-2} + 0.162dPRIX_{t-4} + 0.171dPRIX_{t-5} + 0.32dPRIX_{t-6} + \epsilon_t$$

$$dPRIX_t = -1.867dDCOVT_{t-3} + \epsilon_t$$

### 6.4.4 Test de Granger de Causalité

Dans notre étude nous retenons le concept de causalité définie par Granger ( 1969 ), car il a l'avantage d'être très opérationnel dans les travaux appliqué et s'inscrit parfaitement dans le cadre des modèles vectoriels autorégressifs : il est en effet facilement testable,

## Chapitre 6. Application

puisque il s'agit ici seulement de tester la nullité jointe de certains coefficients. Nous allons donc tester la causalité entre les deux séries.

Null Hypothesis:	Obs	F-Statistic	Prob.
DDCOVT does not Granger Cause DPRIX	29	2.38343	0.0773
DPRIX does not Granger Cause DDCOVT		13.1914	2.E-05

FIGURE 6.53

### Interprétation :

L'hypothèse de test :

$$\begin{cases} H_0 : \text{L'absence de causalité.} \\ H_1 : \text{La présence de causalité.} \end{cases}$$

- D'après le tableau ci-dessus :

#### causalité au sens de Granger de dDCOVT vers dPRIX :

nous acceptons l'hypothèse  $H_0$  car la probabilité associée est de 0.0773 ; elle est supérieure au seuil statistique usuel de 5% ,donc il n'y a pas de causalité .

#### causalité au sens de Granger de dPRIX vers dDCOVT :

nous rejettons l'hypothèse  $H_0$  car la probabilité associée est de  $2.10^{-5}$  ; elle est inférieure au seuil statistique usuel de 5% ,donc la présence de causalité .

### 6.4.5 Test sur l'exogénéité des variables ( $dPRIX_t$ , $dDCOVT_t$ )

Afin de confirmer l'hypothèse précédente, nous avons étudié les fonctions de réponse des deux séries  $dPRIX_t$  et  $dDCOVT_t$  ou leurs représentations graphique est donnée comme suit :

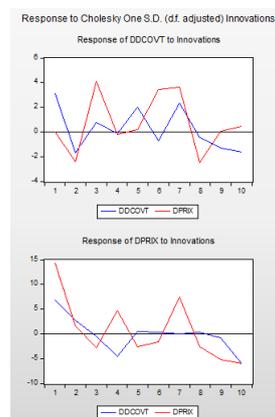


FIGURE 6.54

## Chapitre 6. Application

Le premier graphe ci-dessus décrit les réponses de la série dDCOVT à un choc sur elle-même .

Le deuxième graphe nous contatons ont particulier qu'un choqe positive sur la série dPRIX a un impact positive de la série dDCOVT .

### Analyse de la décomposition de la variance :

Variance Decomposition using Cholesky (d.f. adjusted) Factors			
Variance Decomposition of DDCOVT:			
Period	S.E.	DDCOVT	DPRIX
1	3.138248	100.0000	0.000000
2	4.314408	68.66471	31.33529
3	5.978782	37.55232	62.44768
4	5.983644	37.53598	62.46402
5	6.318141	43.89197	56.10803
6	7.218541	34.65694	65.34306
7	8.402135	33.40381	66.59619
8	8.775602	30.89349	69.10651
9	8.875429	32.43527	67.56473
10	9.036082	34.58713	65.41287
Variance Decomposition of DPRIX:			
Period	S.E.	DDCOVT	DPRIX
1	15.77991	18.40954	81.59046
2	16.10608	20.73926	79.26074
3	16.34348	20.24036	79.75964
4	17.63901	24.22120	75.77880
5	17.83583	23.75588	76.24412
6	17.91373	23.57418	76.42582
7	19.39880	20.10513	79.89487
8	19.58115	19.76233	80.23767
9	20.26895	18.62652	81.37348
10	21.97394	23.42861	76.57139
Cholesky Ordering: DDCOVT DPRIX			

FIGURE 6.55

D'après le tableau ci – dessus, nous remarquons que la variance de l'erreur de prévision de la série dDCOVT est due 35% à ses propres innovations et à 65% à celle de la série dPRIX .

Parcontre, la variance de l'erreur de prévision de la série dPRIX est due à 77% à ses propres innovations et 33% à celle de la série dDCOVT.

Par conséquent, un choc sur les prix du brut du pétrole à un impact plus important sur les nombres des découverte (sismique + forage) qu'un choc sur les nombres des découverte (sismique + forage) sur les prix du brut du pétrole .

### Conclusion :

Il y'a une seule variable exogène (dPRIX).

### 6.4.6 Prévision

Les graphe et les valeurs prédites relatifs aux prévisions précédente des deux séries ( $dPRIX_t, dDCOVT_t$ ) sont donnés par :

## Chapitre 6. Application

---

l'année	découverte
2022	11
2023	23
2024	30
2025	17

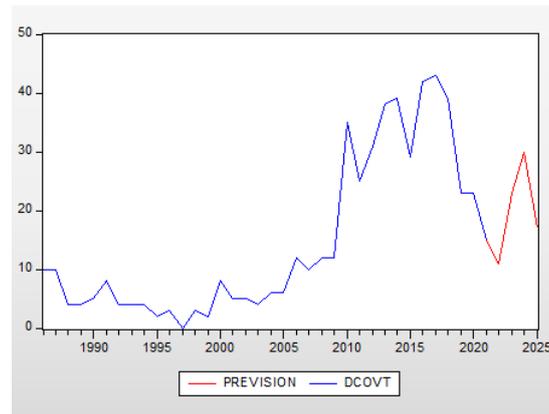


FIGURE 6.56

l'année	prix en dollars
2022	77.32
2023	75.73
2024	118.99
2025	121.67

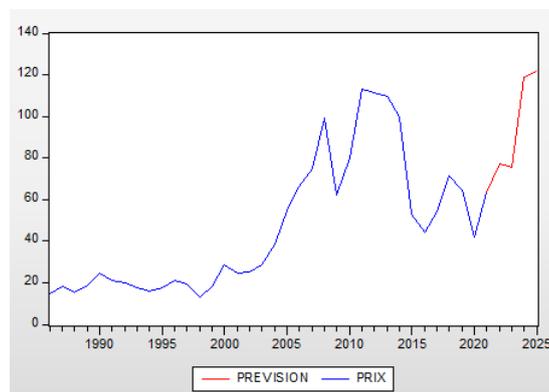


FIGURE 6.57

### 6.4.7 Conclusion

Comme l'étude multivariée est achevée, nous pouvons tirer les résultats suivants : Les deux modèles VAR(6) sont retenus pour modéliser les deux vecteurs des séries ( $dPRIX_t$ ,  $dDCOVT_t$ )

Le modèle VAR(6) est validé, par la suite exploité pour effectuer des prévisions, en examinant les équations liées aux prévisions.

A cette étape bien précise nous sommes en possession de deux groupes de résultats associés à deux approches différentes pour un même objectif à savoir effectuer des prévisions.

Pour choisir les meilleurs résultats une comparaison sera entreprise par la suite pour ne pas perdre de vue le but de notre étude.

### 6.4.8 Comparaison

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

Le calcul des écarts prévisionnels moyens, relativement à chacune des deux méthodes traitées dans ce présent mémoire nous a permis de dresser le tableau suivant :

Séries	Type de modélisation	$RMSE = \sqrt{\frac{\sum_{i=1}^N \varepsilon_t^2}{N}}$
PRIX <sub>t</sub>	Box-Jenkins	<b>14,7816</b>
	VAR	<b>10,8435</b>
DCOVT <sub>t</sub>	Box-Jenkins	<b>5,8023</b>
	VAR	<b>2,1565</b>

FIGURE 6.58

## Chapitre 6. Application

---

**Remarque :** Une méthode est jugée plus appropriée qu'une autre (c'est à dire que ses résultats prévisionnels sont plus fiables) si la valeur du critère RMSE donnée par cette méthode est petite par rapport à celle données par les autres méthodes.

**Conclusion :** Nous concluons que les prévisions obtenues par la modélisation VAR sont plus fiables pour les deux série .

## Application de la Régression

### 6.5 Etude de l'influence des forages, et du sismiques sur les découvertes

$$DCOVT_t = \alpha_0 + \alpha_1 FORAGE_t + \alpha_2 SISMIQUE_t + \epsilon_t$$

Où  $\epsilon_t$  erreur de spécification du modèle, elle restera inconnue

#### 6.5.1 Statistique du modèle

- Somme des carrés des résidus (SSR)=728.165
- Erreur standard de l'estimation =4.69740
- R= 0.942
- $R^2 = 0.888$
- $\bar{R}^2=0.881$

La table d'ANOVA : F=130.941

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	5778,585	2	2889,292	130,941	,000 <sup>b</sup>
	Résidu	728,165	33	22,066		
	Total	6506,750	35			

a. Variable dépendante : découverte

b. Prédicteurs : (Constante), sismique, forage

FIGURE 6.59

On remarque que la signification de F est égale à 0.00 inférieur au seuil choisi 0.05, donc la régression est significative .

#### 6.5.2 Estimation des coefficients

Modèle		Coefficients non standardisés		Coefficients standardisés		t	Sig.	Intervalle de confiance à 95,0% pour B		Corrélations			Statistiques de colinéarité		
		B	Erreur standard	Bêta				Borne inférieure	Borne supérieure	Corrélation simple	Partielle	Partielle	Tolérance	VIF	
1	(Constante)	-,142	2,134			-,067	,947	-4,483	4,199						
	forage	,287	,022	,961		13,127	,000	-,243	,332	,942	,916	,764	,633	1,579	
	sismique	-,9,381E-5	,000	-,031		-,419	,678	-,001	,000	,551	-,073	-,024	,633	1,579	

a. Variable dépendante : découverte

Test d'hypothèse :

$$\begin{cases} H_0 : \alpha_i = 0 \text{ (le coeff n'est pas significatif); } tq : i = 0, 1, 2 \\ H_1 : \alpha_i \neq 0 \text{ (le coeff significatif); } tq : i = 0, 1, 2 \end{cases}$$

test sur  $\alpha_0$  :

## Chapitre 6. Application

---

$\alpha_0 = -1.42$ ; t-statistic = -0.067; signification de t = 0.947

On remarque que t-statistic(-0.067) la est inferieur à 1.96 et la probabilité critique est superieur à 0.05.

Donc on accepete  $H_0$  .

Alors  $\alpha_0$  n'est pas significatif .

**test sur  $\alpha_1$  :**

$\alpha_1 = 0.287$ ; t - statistic = 13.127; signification de t = 0.00

On remarque que t-statistic(13.127) la est superieur à 1.96 et la probabilité critique est inferieur à 0.05.

Donc on rejette  $H_0$  .

Alors  $\alpha_1$  est significatif .

Alors la variable explicative  $FORAGE_t$  est contributive à l'explication de la variable  $DCOVT_t$  .

**test sur  $\alpha_2$  :**

$\alpha_2 = -9.381 \times 10^{-5}$ ; t-statistic = -0.419; signification de t = 0.678

On remarque que t-statistic(-0.419) la est inferieur à 1.96 et la probabilité critique est superieur à 0.05.

Donc on accepete  $H_0$  .

Alors  $\alpha_2$  n'est pas significatif .

Alors la variable explicative  $SISMIQUE_t$  n'est pas contributive à l'explication de la variable  $DCOVT$  .

**On retire le terme constant et SISMIQUE; on obtient :**

$$DCOVT_t = \alpha_1 FORAGE_t + \epsilon_t$$

**D'où le modèle :**

$$DCOVT_t = 0.287 FORAGE_t + e_t$$

où  $e_t$  :le résidus qui est connue.

L'intervalle de confiance est égal à [0.250 ; 0.294]

Nous avons donc un risque de 5% que le véritable coefficient se trouve à l'extérieur de l'intervalle de confiance, ce qui est bien entendu cohérent avec le rejet de  $\alpha_1 = 0$

### 6.5.3 Analyse des résidus

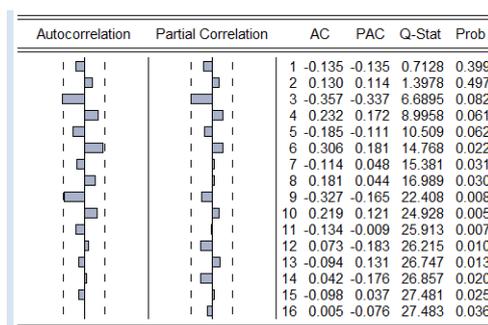


FIGURE 6.60

L'analyse du corrélogramme des résidus (Figure 7.60) permet de voir que le troisième terme est en dehors de l'intervalle de confiance.

### Test de Durbin-Watson :

Test d'hypothèse :

$$\begin{cases} H_0 : \rho = 0 \text{ (les résidus ne sont pas corrélés)} \\ H_1 : \rho \neq 0 \text{ (les résidus sont corrélés)} \end{cases}$$

la statistique DW = 2.249 et  $2.249 \approx 2$

Donc on accepte  $H_0$ , les résidus ne sont pas corrélés.

Nous constatons que les résidus forment un bruit blanc.

### Etude de la série des résidus $e_t$ :

Pour l'étude de la série, nous allons appliquer la méthode de Box et Jenkins qui permet de déterminer un modèle adéquat dans la classe des modèles ARMA (ils permettent de représenter la plupart des processus stationnaires), autrement dit on cherche le modèle le plus adapté pour représenter le phénomène.

### Test de DICKEY-FULLER Augmenté sur la série $e_t$ :

On applique le test de DFA pour infirmer ou confirmer notre hypothèse concernant la stationnarité du processus qui génère la série  $e_t$ , on a besoin de déterminer le retard

qui minimise le critère d'Akaike et Schwarz qui égale à 0. Par la suite on estime par la méthode des moindres carrés les trois modèles.

Les hypothèses à tester sont :

$$\begin{cases} H_0 : \text{la tendance n'est pas significative.} \\ H'_0 : \text{la constante n'est pas significative.} \\ H''_0 : \text{l'absence de la racine unitaire.} \end{cases}$$

On commence par l'estimation du modèle [3].

**Modèle [3] :** "Test sur la tendance"

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES1(-1)	-1.200669	0.173622	-6.915403	0.0000
C	-2.875833	1.616077	-1.779515	0.0847
@TREND("1986")	0.132988	0.077812	1.709080	0.0971
R-squared	0.599466	Mean dependent var	0.127311	
Adjusted R-squared	0.574433	S.D. dependent var	6.968589	
S.E. of regression	4.545995	Akaike info criterion	5.948187	
Sum squared resid	661.3142	Schwarz criterion	6.081502	
Log likelihood	-101.0933	Hannan-Quinn criter.	5.994207	
F-statistic	23.94668	Durbin-Watson stat	1.962160	
Prob(F-statistic)	0.000000			

FIGURE 6.61 – ADF modèle 3 de la série  $e_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student ( $t$ -statistic = 1.709080) est inférieure à toutes les valeurs critiques 3.53, 2.79 lues dans la table de Dickey–Fuller respectivement aux seuils 1%, 5%.  
Donc on accepte  $H_0$

la tendance n'est pas significative.

**Nous passons alors au test du modèle [2]**

**Modèle [2] :** "Test sur la constante"

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES1(-1)	-1.138478	0.174638	-6.519090	0.0000
C	-0.450491	0.795416	-0.566360	0.5750
R-squared	0.562905	Mean dependent var	0.127311	
Adjusted R-squared	0.549660	S.D. dependent var	6.968589	
S.E. of regression	4.676436	Akaike info criterion	5.978395	
Sum squared resid	721.6789	Schwarz criterion	6.067272	
Log likelihood	-102.6219	Hannan-Quinn criter.	6.009075	
F-statistic	42.49853	Durbin-Watson stat	1.931129	
Prob(F-statistic)	0.000000			

FIGURE 6.62 – ADF modèle 2 de la série  $e_t$

A la lecture du tableau ci-dessus, on remarque que :

- La statistique de student ( $t$ -statistic = -0.566360) est inférieure à toutes les valeurs critiques 3.22, 2.54 lues dans la table de Dickey–Fuller respectivement aux seuils 1%, 5%.

- La probabilité de nullité du coefficient de la constante ( $c$ ) qui est égale à 0.3236 est supérieure au seuil de 5%.

Donc on accepte  $H'_0$

la constante n'est pas significative.

**Nous passons alors au test du modèle [1]**

**Modèle [1] :** "Test sur la racine unitaire"

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.562318	0.0000
Test critical values:		
1% level	-2.632688	
5% level	-1.950687	
10% level	-1.611059	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(RES1)  
 Method: Least Squares  
 Date: 06/04/22 Time: 23:51  
 Sample (adjusted): 1987 2021  
 Included observations: 35 after adjustments

FIGURE 6.63 – ADF modèle 1 de la série  $e_t$

la valeur de la t-statistic de test ADF qu'est égale à -3.417268 est inférieure à la différente valeur critique aux différent seuils : -6.562318 ; -2.632688 ; -1,950687,-1.611059 1% ,5% ,10% respectivement (donner par le test ADF ) , alors on accepte  $H_0''$  .

La série  $e_t$  ne possède pas une racine unitaire .

En conclusion , notre série est stationnaire .Et on peut passer à l'autre étape à savoir l'identification .

### 6.5.4 Identification du modèle

l'analyse du corrélogramme simple de la série stationnaire ( $e_t$ ) montre qu'aux retards ( $q = 3$ ) le termes sont à l'extérieure de l'intervalle de confiance et ce qui concerne le corrélogramme partiel , on remarque que les valeurs de la fonction d'autocorrélation partielles sont élevées aux différents retard ( $p=3$ ) ,ce qui amène à estimer plusieurs modèles parmi eux on a choisit :ARMA(3.0) D'ou le modèle ARIMA ( $p=3,q=0$ ) est celui le plus adéquant pour modéliser notre série .

### 6.5.5 Estimation des paramètre du modèle

le tableau suivant contient l'estimatio du modèle ARMA(3.0) , qui a été choisi parmi d'autre modèles estimés on se basant sur les critères de pouvoir prédictif à savoir :  $R^2$  statistique de Fisher : maximum ; AIC , SC : minimum et sur la statistique de Durbin Watson qui égale 1.877576 présage un bon ajustement .

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(3)	-0.370531	0.124373	-2.979193	0.0053
SIGMASQ	17.69526	4.861005	3.640247	0.0009
R-squared	0.133860	Mean dependent var	-0.386605	
Adjusted R-squared	0.108385	S.D. dependent var	4.584074	
S.E. of regression	4.328528	Akaike info criterion	5.834592	
Sum squared resid	637.0293	Schwarz criterion	5.922655	
Log likelihood	-103.0226	Hannan-Quinn criter.	5.865297	
Durbin-Watson stat	1.877576			
Inverted AR Roots	.36+ .62i	.36- .62i	-.72	

FIGURE 6.64

Tests sur les résidus du modèle ARMA(3.0) :

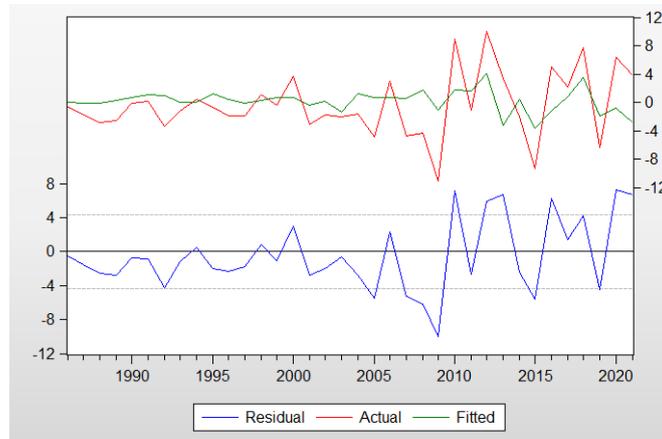


FIGURE 6.65

le modèle explique 13.36% de la variation total de la variable dépendante ( $R^2=13.36\%$ )

Corrélogramme des résidus du modèle ARMA(3.0) :

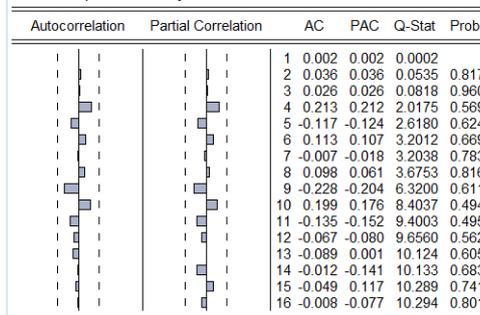


FIGURE 6.66

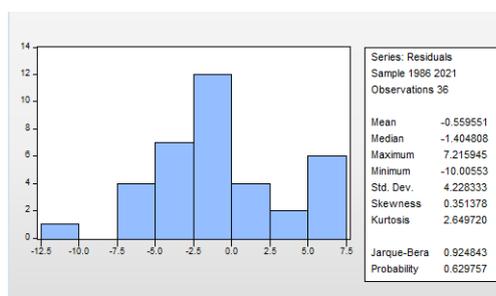
L'analyse du corrélogramme des résidus, illustré par la figure montre que tous les termes sont à l'intérieur de l'intervalle de confiance, donc les résidus forment un bruit blanc, chose confirmée par la statistique de LJUNG-BOX (Q-stat) qui est inférieure à la valeur théorique de  $\chi^2(h - 1)$  quelque soit le retard h, on particulier pour h=16 on a :

$$Q - stat(16) = 10.294 < 25$$

au seuil 0.05, donc les résidus forment un bruit blanc.

### Test de normalité des résidus :

les tests sont effectués à partir des valeurs empiriques des coefficients de SKewness, Kurtosis et la statistique de Jarque- Berra par le logiciel EVIEWS 10.



### Test de Skewness :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_1 = 0 \text{ (la distribution des résidus est symétrique)} \\ H_1 : \gamma_1 \neq 0 \text{ (la distribution des résidus n'est pas symétrique)} \end{cases}$$

$$\gamma_1 = \frac{|\sqrt{\beta_1}|}{\sqrt{\frac{6}{n}}} = \frac{\sqrt{0.351378}}{\sqrt{\frac{6}{36}}} = 1.45 < 1.96$$

on accepte  $H_0$

### Test de Kurtosis :

L'hypothèse de test :

$$\begin{cases} H_0 : \gamma_2 = 0 \text{ (la distribution des résidus est relativement Aplatie)} \\ H_1 : \gamma_2 \neq 0 \text{ (la distribution des résidus n'est pas Aplatie)} \end{cases}$$

$$\gamma_2 = \frac{|\beta_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|2.649720 - 3|}{\sqrt{\frac{24}{36}}} = 0.42 < 1.96$$

On accepte  $H_0$ .

nous confirmons par la statistique de Jarque-Berra

### Test des résidus de Jarque-Berra :

L'hypothèse de test est :

$$\begin{cases} H_0 : \text{Les résidus forment un bruit blanc gaussien} \\ H_1 : \text{Les résidus forment un bruit blanc non gaussien} \end{cases}$$

La statistique de Jarque - Berra :

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

$$JB = 0.924843 < \chi_{0.05}^2(2) = 5.99$$

## Chapitre 6. Application

---

les résidus forment un bruit blanc gaussien.

Le processus ARMA(3,0) est donc un bruit blanc gaussien .

**Le modèle est représenté par l'équation suivante :**

$$e_t = w_t - 0.37e_{t-3}$$

**Le modèle de la régression est donné par :**

$$DCOVT_t = 0.287FORAGE_t - 0.37e_{t-3} + w_t$$

ou  $w_t$  est un bruit blanc

### 6.6 Conclusion :

Le nombre de découvertes est logiquement une récompense des efforts d'exploration (Sismique;Forage), nous avons donc pensé à étudier l'influence des efforts d'exploration sur le nombre de découvertes; pour cela, nous avons utilisé la méthode de Régression : Celle-ci révèle en effet- l'existence d'un retard des efforts d'exploration par rapport au nombre découvert .

Concernant l'explication de l'investissement (comme variable endogène), nous remarquons que les plus importants paramètres qui influent sur l'investissement sont : le forage et le prix du pétrole brut. La fiabilité et la véracité résident dans l'accès direct au sous-sol par le forage pour se prononcer sur la réalité de l'existence du gisement. Quand au coût du forage, il consomme à lui seul la plus importante dépense d'exploration (65 à 80%) du coût total.

Les décisions d'investissements étant orientées par une vision long terme des prix assez liée aux variations de la conjoncture. La capacité de financement des compagnies pétrolières est cependant sensible aux variations de prix du pétrole brut, ce qui a conduit les acteurs privés à réduire leurs investissements dans l'arrière en 1998 , 1999,2020 et 2021. Donc le forage et le prix du pétrole sont deux variantes importantes et incontournables de l'investissement .

# Conclusion générale

L'Algérie est l'un des pays les plus riches en hydrocarbures, elle possède d'importantes réserves de pétrole et de gaz naturel dont son activité économique dépend fortement. La hausse des prix du pétrole durant ces dernières années incite les entreprises à investir encore plus dans le domaine de l'exploration ; les découvertes permettront ainsi de fortifier le volume des réserves et d'augmenter le rythme de production pour maximiser le profit. Il est donc primordial pour la SONATRACH de bien comprendre les mécanismes qui régissent le domaine de l'exploration pétrolière et les paramètres qui l'influencent. Pour cela, l'analyse et l'évolution de certaines grandeurs se révèlent de première importance ; Ainsi deux séries ont été mise à notre disposition, à savoir : les nombres de Découvertes, le Prix du pétrole brut

La première partie de notre étude a porté sur l'analyse des séries chronologiques représentant l'évolution des grandeurs citées ci dessus.

Il existe plusieurs méthodes pour aborder l'étude d'une série chronologique. Cependant le nombre d'observation dont on dispose peut restreindre ce choix de manière importante.

Une série annuelle observée sur 36 ans ne constitue pas un grand champ d'investigation, néanmoins nous avons proposé dans un premier temps l'étude individuelle des séries en utilisant l'approche univariée la méthodologie de Box Jenkins.

- L'application de la méthode de Box Jenkins nous a permis de modéliser les phénomènes étudiés.

En effet nous avons pu déterminer dans la famille des modèles ARIMA, celui qui est le plus adapté à représenter le comportement de nos séries. Par la suite nous avons calculé les prévisions de chacune d'entre elles.

L'examen du corrélogramme des séries différenciées Découverte ,Prix nous à pas permis de choisir un modèle pour les représenter car tous les termes ne sont pas significativement différents de zéro. Ces deux séries représentent 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 ente elles. Afin d'améliorer les prévisions fournies par cette méthode nous avons appliqué la modélisation multivariée (VAR) sur les séries nombres de Découvertes,et Prix du pétrole brut. Le modèle choisi est un VAR(6), il nous a permis de calculer les prévisions pour chacune des séries.

## Conclusion générale

---

En se basant sur les critères de comparaison en terme de performance prévisionnelle, Nous concluons que les prévisions obtenues par la modélisation VAR sont plus fiables pour les deux série puisqu'elle donne la plus petite valeur des écarts prévisionnels moyens (RMSE). Dans la deuxième partie, notre objectif a été de trouver un modèle mathématique qui permettra de mettre en évidence une corrélation entre les nombres de découvertes et l'effort d'exploration (Forage ;Sismique) Pour cela nous avons utilisé la méthode de la régression simple et multiple. Celle-ci permet de bien aborder les problèmes économiques et de tester les hypothèses portant sur la relation entre une variable dépendante et au moins deux variables indépendantes.

nous remarquons que le plus important paramètre qui influence sur les nombres de Découverte est Forage .

Finalement, cette étude nous a permis non seulement de manipuler les logiciels utilisés, mais aussi d'adopter autant de notions et des méthodes 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] AILI. A, et KERSANI.F, *Etude Prévisionnelle des ventes des produits laitiers au sein du LFB (Laiterie Fromagerie de Boudouaou)*, Université M'hamed BOUGARA Boumerdès, Faculté des Sciences, (2006-2007).
- [2] Arthur CHARPENTIER, *Séries temporeles : Théorie et applications*.
- [3] BOUAZA. M , *La méthodologie de Box Jenkins et applications*, Université de de Saida.
- [4] BOUBCHIR. L, *Modélisation statistique multivariée des images dans le domaine des transformées multi-échelles orientées et non-orientées*.
- [5] Christian GOURIEROUX Alain MONFORT *Statistique et modèles économétriques* ,Edition Economica 1996.
- [6] Didier Delignières *Série temporelles –modèles ARIMA* Mars 2020
- [7] GATT. F et KEDOUI. S, *Modélisation et Prévision Journalières des Courbes de Charges*, Université des sciences et de la technologie Houari Boumediene, Faculté de Mathématiques.
- [8] Guy MELARD, *Méthodes de prévision à court terme*, Edition ELLIPSE ;1990
- [9] HAMRITA. M, *Séries temporelles*, Institut Supérieur de Mathématiques Appliqués Informatique - Kairouan, (Décembre 2013).
- [10] Jacques FLANDRIN, Jean CHAPELLE, *Le pétrole* 1961 ; Edition TECHNIP
- [11] Jean MASSRON, *Economie des hydrocarbures* 1975 ; Edition TECHNIP.
- [12] KESRAOUI Mahrez NSISABIRA Cléophas *Etude et modéfication d'une évolution d'une pathologie* 2016/2017
- [13] MOKHTARI Loubna - ZADDEM Nawal *étude de l'impact des importations de lait et de blé en Algérie* Université M'hamed BOUGARA Boumerdes 2017/2018.
- [14] M. Sofiane LADJOUZE- M. Yahia ZAIR *Modélisation et prévision de l'activité d'exploration pétrolière en Algérie* 2002/2003
- [15] PARDOUX. C, et GOLDFARB.B, *Prévision à court terme : méthodes de lissage exponentiel* , Université Paris-Dauphine.

## Bibliographie

---

- [16] TAZEROUTI. M, et HANECHÉ .M, *Modélisation des ventes des produits Pharmaceutiques Laboratoire Pharmaceutique Algérien*, Université M'hamed BOUGARA Boumerdes