

*République Algérienne Démocratique et populaire*  
*Ministère de l'enseignement Supérieur et de la Recherche Scientifique*

**Université M'hamed Bougara de BOUMERDES**  
*Faculté des Mathématiques*

**Mémoire de Master Mathématique Financière**



Perspective de l'activité exploration a  
l'horizon 2030 en fonction des prix du  
pétrole

**Réalisé par :**  
M<sup>r</sup> Sfina Sidali  
M<sup>r</sup> Chennah Anouar

**Soutenu à UMBB, le 30/06/2019, devant le jurys composé de :**

<i>Encadreur</i>	<i>M<sup>me</sup></i>	<i>BENMANSOUR.M</i>	<i>M.A.A</i>	<i>UMBB</i>
<i>Prsidente</i>	<i>M<sup>me</sup></i>	<i>LARABI.R</i>	<i>M.A.A</i>	<i>UMBB</i>
<i>Examinatrice</i>	<i>M<sup>me</sup></i>	<i>GATT.F</i>	<i>M.A.A</i>	<i>UMBB</i>

**Année Universitaire 2018/2019**

---

## *Remerciement*

Nos remerciements au dieu le tout puissant qui nous a donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de M.BENMANSOUR enseignante de statistique à l'université de Mehmd Bougera Boumerdes, on le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire.

Nos remerciement s'adresse à Ms AZIZ HEOUA et Ms BenAhmed Mourad pour sont aides pratique et son soutien moral et ses encouragements. Nous sommes conscientes de l'honneur que nous a fait M. A.Arabî en étant président du jury et M. Gatt d'avoir accepté d'examiner ce travail.

On remercie aussi M. Flici cheffe département pour leurs aides.

Nos remerciement s'adresse également à tout nos professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leurs charges académiques et professionnelles.

Nos profonds remerciements vont également à toutes les amis et collègues qui m'ont apporté leur soutien moral et intellectuel tout au long de ma démarche.

---

## *Dédicace*

Je dédie ce travail qui n'aura jamais pu voir le jour sans les soutiens indéfectibles et sans limite de mes chers parents, Sfina Bouelam et Oualed-Amri Noura qui ne cessent de me donner avec amour le nécessaire pour que je puisse arriver à ce que je suis aujourd'hui.  
Que dieux vous protège et que la réussite soit toujours à ma portée pour que je puisse vous combler de bonheur.

Je dédie aussi ce travail à :

- Mes frères, Mourad et Houssef.
- Mes grandes-mères.
- Mes oncles, mes tantes et leur famille.f
- Tous mes cousins et cousines.
- Tous mes amis, mes collègues et tous ceux qui m'estiment.

*SFINA SIDALI*

---

## *Dédicace*

- Toutes les lettres ne sauraient trouver les mots qu'il faut ... Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance ... Aussi, c'est tous simplement que Je dédie ce modeste travail à :

- Ma mère, qui a œuvré pour ma réussite, par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

- Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

- Mes frères et sœurs qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

Mon collègue de MF.

- A ceux qui m'ont encouragé et soutenu dans mes moments les plus difficiles.

*ChENAH ANOUAR.*

---

## *Introduction générale*

Le développement économique et social est conditionnée par l'utilisation optimale des ressources disponibles, naturelles et humaines. Par ailleurs, la consommation de pétrole est à la base de toutes les activités économiques.

Le secteur de l'énergie et notamment le secteur pétrolière joue le rôle de locomotive de développement dans la sphère économique nationale dans le sens ou son essor conditionne l'essor d'une croissance économique dont notre pays. Son rôle est plus que vital du moment où elle est présente dans l'industrie, le transport et de nombreux autres secteurs.

Les perspectives pour 2018 De nombreux paramètres seront, comme par le passé, susceptibles d'influencer avec plus ou moins d'ampleur et sur des délais plus ou moins longs, les évolutions du prix du pétrole. On peut, en particulier, citer la croissance économique des pays émergents, la stabilité des pays producteurs, le taux du dollar ou les évolutions des marchés financiers.

Le Changement de la tendance de l'investissement dans l'amont pétrolier après un cycle haussier de plus d'une décennie en raison de la persistance de très bas prix du pétrole et du gaz.. En 2017 au niveau mondial , l'investissement dans l'amont est estimé à 394 G\$, le chiffre le plus bas depuis 2007, en baisse de 124 G\$ par rapport à 2015.

Face à cette changement qui cesse de décroître et afin d'assurer l'équilibre entre la demande et l'offre (la production), la Société Nationale Sonatrach est obligé de prendre des mesures pour satisfaire la demande potentielle.

Pour cette raison et afin de satisfaire la demande au moindre coût, il est utile pour la Sonatrach d'étudier et d'analyser l'évolution de la demande nationale et internationale en pétrole, afin de tirer des informations et des conclusions pour la phase de prise des décisions liées à l'adaptation du parc de production (programmes de production, l'entretien, l'approvisionnement, l'équipement et calcul des coûts marginaux de production).

Ainsi, l'objectif attendu de notre modeste travail de fin d'études consiste à analyser les investissements sur les projets, les réserves pétrolière récupérable et les prix moyennes de pétrole, identifier les éléments qui influent et caractérisent l'exploration et la production de lui et élaborer un modèle mathématique prévisionnelle qui permettra aux décideurs de

---

prendre les décisions adéquates pour renouvellement des investissements moins coût et faire un prix de vente acceptable, afin de satisfaire le besoin des consommateurs .

Notre travail est composé de deux sections ; la première forme le premier chapitre, elle comporte des définitions des hydrocarbure et ses sources, et fera la présentation de l'organisme d'accueil ou s'est déroulé notre lieu de stage, notamment département d'exploration ; une filiale du groupe SONATRACH ;

La deuxième section contient trois chapitres : dans le premier nous essayerons de définir les séries chronologiques, étudier les typologies des modèles, la non stationnarité des séries chronologiques, le deuxième chapitre sera consacré a l'étalement théorique des méthodes de prévisions utilisées dans notre étude afin de mieux comprendre le mécanisme de prévision ; en fin le dernier chapitre sera consacré à la modélisation des séries tout en exposant le modèle le plus adéquat pour notre prévision. Nous clôturons cette étude par une conclusion générale qui consolidera les connaissances acquises et exposera les résultats auxquels nous avons abouti

# Table des matières

<b>1</b>	<b>Aspect descriptif de SONATRACH et de son environnement</b>	<b>9</b>
1.1	Généralisation sur les hydrocarbures : . . . . .	9
1.1.1	Définition du pétrole : . . . . .	9
1.1.2	Histoire du pétrole : . . . . .	10
1.1.3	Les activités de l'industrie pétrolière : . . . . .	10
1.1.4	Etapes de la recherche des hydrocarbures : . . . . .	12
1.1.5	Techniques de L'exploration : . . . . .	12
1.2	Les gisements et les réserves : . . . . .	14
1.2.1	Les gisements : . . . . .	14
1.2.2	Les réserves : . . . . .	17
1.3	Les risque et couts dans le secteur pétrolier : . . . . .	19
1.3.1	Type de couts en exploitation / production : . . . . .	19
1.3.2	Les risque dans le secteur pétrolière : . . . . .	21
<b>2</b>	<b>Aspect théorique des séries chronologiques</b>	<b>27</b>
2.1	Définition d'une série chronologique . . . . .	27
2.1.1	Objectifs de l'analyse d'une chronique . . . . .	28
2.1.2	Les composantes d'une série chronologique . . . . .	28
2.1.3	Les schémas de composition . . . . .	29
2.1.4	Procédure de choix d'un schéma de décomposition . . . . .	31
2.2	Caractéristiques d'une série chronologique . . . . .	32
2.2.1	Moyenne et Variance : . . . . .	32

2.2.2	Fonction d'auto-covariance : . . . . .	32
2.2.3	Fonction d'auto-corrélation (AC) . . . . .	33
2.2.4	Fonction d'auto-corrélation partielle (PAC) . . . . .	33
2.3	Analyse d'une série chronologique . . . . .	34
2.3.1	Les opérateurs retard, avance et différenciation . . . . .	34
2.3.2	Les effets des opérateur . . . . .	35
2.4	Processus aléatoire stationnaire . . . . .	35
2.4.1	Définition d'un processus aléatoire . . . . .	35
2.4.2	Processus aléatoire stationnaire . . . . .	35
2.5	Les processus aléatoire non stationnaire . . . . .	36
2.5.1	Les processus ( $TS$ ) ( trend stationary) . . . . .	36
2.5.2	Les processus ( $DS$ ) (différence stationary) . . . . .	37
2.6	Modélisation des séries temporelles . . . . .	38
2.6.1	Les processus linéaire stationnaire . . . . .	38
2.6.2	Les processus non stationnaires . . . . .	40
2.6.3	Processus à effet ARCH : . . . . .	40
2.7	Étude de la stationnarité d'une série temporelle . . . . .	41
2.8	Conclusion . . . . .	43
<b>3</b>	<b>Méthodes de prévision</b> . . . . .	<b>44</b>
3.1	Une prévision ? . . . . .	44
3.1.1	Les méthodes de prévision . . . . .	45
3.2	La méthodologie de Box et Jenkins : . . . . .	45
3.2.1	L'identification : . . . . .	45
3.2.2	L'estimation des paramètres : . . . . .	46
3.2.3	La validation du modèle : . . . . .	47
3.2.4	La prévision : . . . . .	48
3.2.5	L'interprétation des résultats : . . . . .	49
3.3	Prévision par le modèle ARFIMA . . . . .	50
3.3.1	Les modèles ARFIMA( $p,d,q$ ) . . . . .	50

---

3.3.2	Stationnarité des modèles ARFIMA(p,d,q) . . . . .	50
3.3.3	Les modèles longue mémoire . . . . .	52
3.3.4	Estimation par la méthode de Geweke Porter-Hudak (1983) . . . . .	53
3.4	Prévision par La régression linéaire . . . . .	54
3.4.1	La régression linéaire simple : . . . . .	55
3.4.2	La régression linéaire multiple : . . . . .	55
3.4.3	Estimation de $\beta$ par les moindres carrées : . . . . .	57
3.4.4	Test d'hypothèses : . . . . .	58
3.4.5	Prévision d'une valeur ultérieure : . . . . .	59
3.5	Conclusion . . . . .	60
<b>4</b>	<b>APPLICATION :</b>	<b>62</b>
4.1	Présentation des données : . . . . .	62
4.2	Prévision par la méthodologie de Box et Jenkins . . . . .	64
4.2.1	Études de la série d'investissement sur les projets d'exploration . . . . .	64
4.2.2	Étudie la série des réserves pétrolière récupérable . . . . .	74
4.2.3	Etudie la série du prix moyenne de pétrole . . . . .	83
4.3	Application ARFIMA . . . . .	91

# 1

## Aspect descriptif de SONATRACH et de son environnement

### 1.1 Généralisation sur les hydrocarbures :

#### 1.1.1 Définition du pétrole :

Le pétrole (petroleum, du mot grec *petra*, roche et du mot latin *oleum*, huile) est un liquide d'origine naturelle, une huile minérale composée d'une multitude de composés organiques essentiellement des hydrocarbures piégés dans des formations géologiques particulières. Le pétrole dans son gisement est fréquemment associé à des fractions légères qui se séparent spontanément du liquide à la pression atmosphérique, ainsi que diverses impuretés comme le CO<sub>2</sub>, H<sub>2</sub>S, l'eau de formation et des traces métalliques. Chaque gisement pétrolier recèle une qualité particulière de pétrole, déterminée par la proportion relative en molécules lourdes et légères, mais aussi par la quantité d'impuretés. L'industrie pétrolière caractérise la qualité d'un pétrole à l'aide de sa densité API, correspondant à sa « légèreté » : un brut de moins de 10<sup>deg</sup> API est plus dense que l'eau et correspond à un bitume, tandis qu'une huile de plus de 31.1<sup>deg</sup> API correspond à un brut léger. Les pétroles compris entre 10 et 45<sup>deg</sup> API étaient dits conventionnels, tandis qu'en dehors de cet intervalle les pétroles étaient dits non conventionnels ; cette définition est néanmoins évolutive car les technologies actuelles

permettent de traiter par des procédés standards des pétroles jusqu'alors considérés comme exotiques : les condensats, situés au-delà des 45<sup>deg</sup>API, en sont une bonne illustration.

### 1.1.2 Histoire du pétrole :

En l'Algérie :

#### *De l'indice des hydrocarbures en surfaces aux découvertes de gisements*

L'histoire du pétrole en Algérie a commencé suite aux observations d'indices d'hydrocarbures observés en surface au Nord de l'Algérie. Les explorateurs se sont vite intéressés à cette région dès la fin des années 1880 ou des premiers travaux d'exploration furent entamés. Des puits peu profonds ont montré la présence de pétrole dans la région de Ain Zeft vers 1885 ou encore ceux dans le bas de Chellif à Tiliouanet au sud de Rélizane en 1915.

Vu l'intérêt vital porté au pétrole après la seconde guerre mondiale, la prospection s'est intensifiée et en 1948 le gisement de Oued Guetérini fut découvert au Sud de Sour el Ghozlane à 150 Km au sud d'Alger, découverte modeste mais commerciale, sa production atteignait son maximum en 1955.

Le tournant de l'histoire pétrolière de l'Algérie a commencé lorsque l'explorateur C.Kilian lors de sa longue traversée du Sahara a révélé par ses observations l'existence des couches géologiques schisteuses carbonées et bitumineuses témoignant de l'existence possible d'accumulation de pétrole au Sahara. C'est à ce moment-là que des études géologiques de surface furent entreprises dans cette région.

L'introduction des techniques géologiques (techniques qui permettent de montrer l'architecture du sous-sol ) dans la prospection pétrolière notamment de la sismique réfraction a permis la réalisation de forage en juin 1956 de Hassi-Messaoud produisant d'importantes quantités d'hydrocarbures (pétrole essentiellement ), d'autres gisements aussi importants ont suivi, il s'agit de Hassi-Rmel (GAZ), In salah(GAZ) etc. Avec ces gisements géants, l'Algérie faisait ainsi son entrée sur la scène énergétique mondiale. D'autres gisements ont été découverts par la suite, Rhourde nous, Nezla-hassichergui , In Amenas, Zarzatine, Alrar, Ohanet Berkin et bien d'autres encore.

### 1.1.3 Les activités de l'industrie pétrolière :

L'industrie pétrolière traite de la chaîne industrielle du pétrole et du gaz naturel, du gisement jusqu'au consommateur. Le système pétrolier désigne la combinaison de facteurs géologiques majeurs qui ont permis d'obtenir l'accumulation d'hydrocarbures. Tout d'abord, la présence d'une roche mère est fondamentale pour la génération des hydrocarbures. Ensuite une roche- réservoir poreuse et perméable est nécessaire pour contenir des hydrocarbures

et leur permettre de s'accumuler. Ce réservoir doit être surmonté d'une couverture qui agit comme une barrière contre les mouvements ascendants de fluides. Un piège doit sceller le tout, afin de permettre des accumulations. Enfin, la succession des événements géologiques, que l'on appelle le timing, doit avoir été favorable, et en particulier il est indispensable que le piège se soit formé avant que les hydrocarbures n'aient migré.

C'est l'occurrence de chacun de ces événements individuels que les explorateurs tentent d'évaluer, au moyen de probabilités, lors de la prospection pétrolière, afin d'évaluer les chances de rencontrer une accumulation d'hydrocarbures dans un endroit précis du sous-sol.

### **1. L'exploration :**

Pour avoir du pétrole, il faut d'abord faire de l'exploration, c'est-à-dire rechercher les endroits où la topographie du sous-sol permet de « piéger » l'or noir et donc de rechercher les gisements d'hydrocarbures qui seront ensuite exploités. L'exploration constitue une phase particulière où règne un contexte de grande incertitude, en effet l'exploration du pétrole consiste à étudier la géologie pétrolière. Cette exploration commence par l'établissement des cartes à l'aide de photos aériennes. L'étude géologique proprement dite, s'intéressant à la formation des gisements et autres caractéristiques des roches en tant que réservoirs (ou couvertures).

L'étude des structures internes tendant à définir l'existence des « pièges » à partir des méthodes de surface ; c'est la géophysique que pratiquent des équipes parcourant les terrains à prospecter (à explorer) et dessinant des cartes structurales. Les moyens les plus sophistiqués sont mis en œuvre, la prospection étant affinée par un maillage sismique.

Sur terre, la génération des ondes se fait soit à l'aide d'explosifs, soit avec des camions vibreurs. Les données sont enregistrées à l'aide de sismographes.

En mer, un bateau remorque un dispositif de génération des ondes à air comprimé (canon) ainsi qu'un réseau de capteurs de pression (hydrophones) répartis en lignes (streamers) pouvant atteindre 10km de long.

L'ensemble des techniques ci-dessus aboutit au « prospect » c'est-à-dire au gisement « putatif » qu'il faut finalement vérifier sur le terrain par forage.

**2. La production ;**

**3. Le transport ;**

**4. Le raffinage ;**

**5. La distribution.**

### **1.1.4 Etapes de la recherche des hydrocarbures :**

#### **La prospection :**

La prospection est la première étape de la filière du pétrole. De plus en plus de compagnies pétrolières font appel à des entreprises spécialisées qui réalisent des études géologiques. Puis vient la décision de pratiquer ou non des forages. De la précision et de la justesse des méthodes d'analyse géologique découlent les possibilités de rencontrer effectivement du pétrole lors des forages. Plus ces prédictions seront justes, plus le coût global de la phase de prospection sera réduit compte tenu du nombre de forages limités à entreprendre.

En cas de découverte de pétrole, il reste à construire l'ensemble des infrastructures nécessaires à l'extraction du pétrole, à son stockage et à son transport.

### **1.1.5 Techniques de L'exploration :**

#### **A. Prospection géologique :**

La recherche géologique consiste en l'observation des affleurements de terrain et a quelques travaux en laboratoire afin d'aboutir à une meilleur connaissance du terrain. Cette activité se réalise en surface.

##### **- La géologie de terrain :**

Consiste en des prélèvements d'échantillon et des observations : cette discipline fait appel à des géologues spécialisés équipés de véhicules tout terrain et de matériel (marteau, boussole, acide, etc...).

##### **- La géologie de synthèse :**

Consiste à faire la synthèse des données issues de terrain, des puits et du laboratoire, en établissant des coupes géologiques des corrélations et des cartes.

#### **B. Méthode Géophysique :**

La recherche pétrolière doit faire appel à la géophysique pour obtenir des renseignements sur la structure profonde du sous-sol ; on distingue différentes méthodes :

**La gravimétrie :**

Elle permet de :

- Rechercher les grandes structures dans la couverture sédimentaire de la plate forme.
- Etudier la constitution géologique régionale.
- Tracer des failles.
- Rechercher les zones récifales et dômes de sel.

**La magnétométrie :**

Elle consiste à détecter les anomalies du champ magnétique terrestre et à connaître la formation de la plate forme.

**La sismique :**

Technique permettant de visualiser les différentes couches sédimentaires du sous-sol en envoyant des ondes sonores qui sont récupérées par des géophones. L'image peut être donnée en 2D ou en 3D.

**- La sismique 2D :**

Elle permet d'évaluer le potentiel en hydrocarbures, la morphologie du sous-sol, la mise en relief des zones favorables au piégeage des hydrocarbures, en interprétant les data sismiques.

**- La sismique 3D :**

Elle réduit considérablement les risques sur le positionnement des futures implantations, en donnant une image -en trois (03) dimensions du gisement- qui permet de voir avec plus de détails, la surface complète du gisement ainsi que sa profondeur, elle est utilisée surtout en phase de développement.

**- La sismique 4D :**

Appelée également de réservoir, il s'agit de conjuguer la sismique 3D avec la dimension temps. Elle est utilisée pour le suivi d'un gisement en production dans le temps, son application est à ses débuts.

Ces méthodes débouchent sur le tracé de cartes en sub-surface après avoir traité les données acquises. Ces cartes peuvent être considérées comme une échographie du sous-sol. Elles permettent de dégager une image structurale du

sous-sol et parfois de localiser les réservoirs et d'identifier les différents contacts (gaz -eau, huile -eau).

### C. Le forage d'exploration :

Pour l'explorateur, il s'agit plutôt de sondage car ce qu'il recherche c'est l'information géologique et pétrolière. Le forage d'un puits apporte la certitude de l'imprégnation des couches, il décide de l'existence ou de l'absence du pétrole. Par ailleurs, en l'absence d'accumulation, le puits sec extrêmement précieux, car il permet par l'intermédiaire de la géologie de subsurface d'examiner la nature et les propriétés des couches traversées.

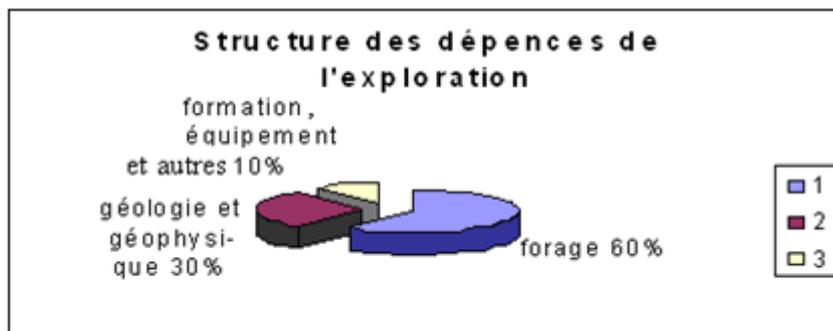


FIGURE 1.1: Structure des dépenses de forage

## 1.2 Les gisements et les réserves :

### 1.2.1 Les gisements :

#### 1. Définition d'un gisement :

Un gisement est constitué d'une accumulation de pétrole ou de gaz dans les pores d'une roche sédimentaire, le réservoir. Les gisements d'hydrocarbures sont donc situés dans les bassins sédimentaires, il y a de cela des millions d'années.

Ces sédiments proviennent soit de l'érosion de reliefs émergés bordant ces dépressions (argile, sable), soit d'une activité biologique (calcaire), ou encore de l'évaporation de l'eau dans des lagunes.

En d'autres termes, le gisement est l'aire géographique dans le sous-sol, qui est constitué par un ou plusieurs autres réservoirs.

## **2. L'étude de gisement :**

L'étude d'un gisement à partir de la découverte d'un réservoir productif a pour but d'établir un projet de développement qui cherchera à optimiser la récupération des hydrocarbures dans le cadre d'une politique économique donnée. Les spécialistes des gisements continueront aussi à étudier le gisement pendant le cycle de production afin d'en tirer les informations nécessaires à l'exploitation optimale du gisement.

Tout ceci nécessite notamment l'estimation des :

- Valeurs d'hydrocarbures (quantité en place)
- Quantité récupérable (estimées à partir de plusieurs modes d'exploitation possible).
- Potentiels de production de rentabilité pour un projet donné.

Pour qu'un gisement soit réputé commercialement exploitable , le produit des ventes des quantités hydrocarbures susceptibles d'être produites dans ce dernier permet de couvrir les investissements de recherche, les investissements de développement et les investissements d'exploitation, ainsi que d'autres charges et dépenses, et d'obtenir un bénéfice net qui justifie pour l'exploitant l'intérêt de la mise en exploitation du dit gisement.

## **3. La vie d'un gisement :**

La durée de vie d'un gisement pétrolier ou gazier correspond à la période durant laquelle on extrait les hydrocarbures qu'il renferme. Cette durée varie généralement de 15 à 30 ans. Pour les gisements abritant de très grandes quantités d'hydrocarbures, elle peut se prolonger jusqu'à 50 ans et plus.

La durée de vie d'exploitation dépend de :

- La conjoncture favorable,
- Le volume et la nature des quantités d'hydrocarbures à extraire,
- Le rythme de production,
- Les techniques d'exploitation.

Le cycle de vie du gisement comprend trois étapes :

- la phase de l'exploration : 2 à 5 ans
- phase de développement : 3 à 4 ans
- la phase de production : suivant la taille de gisement et le rythme de la production.

**L'abandon** : On cesse d'exploiter un gisement à partir du moment où, pour diverses raisons, la production de pétrole et de gaz coûte plus d'argent qu'elle n'en rapporte ou si les recettes générées par le projet n'arrivent pas à couvrir les dépenses nécessaires par l'investissement et aussi pour payer la fiscalité.

On arrête la production lorsque le cumul des cash-flows devient stable ou négatif.

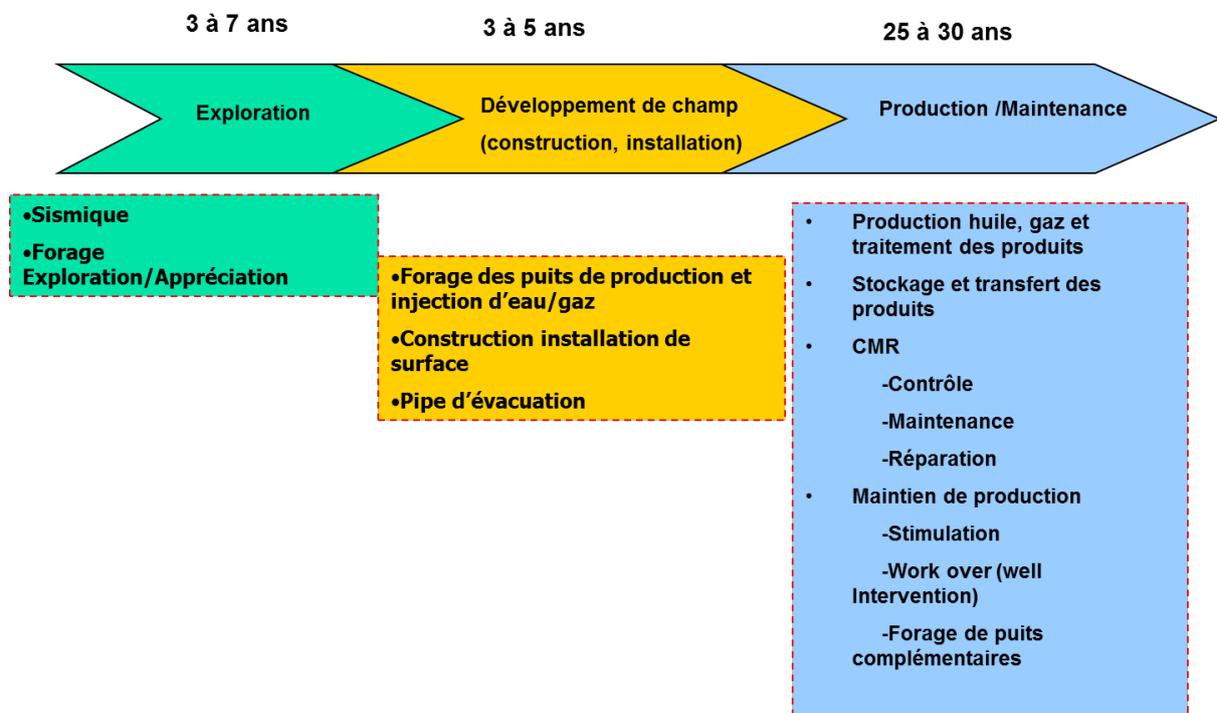


FIGURE 1.2: La vie d'un gisement

## 1.2.2 Les réserves :

### 1. Définition des réserves :

Les réserves pétrolières désignent le volume de pétrole récupérable, à partir du champ de pétrole découvert, sur la base des contraintes économique et technologique actuelles. La part récupérable du pétrole présent dans le sol est appelée « réserve » lorsqu'elle est exploitable techniquement et économiquement ; pour un champ donné, on tient compte de ces contraintes pour évaluer la part récupérable, sachant qu'une partie importante du pétrole ne sera pas extraite du sol. Le taux de récupération est le pourcentage de pétrole récupérable par rapport à la quantité totale initialement en place. Ce taux est de l'ordre de 35

### 2. Types de réserves :

Les réserves pétrolières sont appelées possibles, probables ou prouvées selon le degré croissant de certitude que l'on a de leur existence, en fonction des données et des interprétations géologiques et techniques, pour chaque localisation. La mesure de ces valeurs est faite par la ou les compagnies pétrolières procédant à l'exploration/production.

Les réserves pétrolières sont essentiellement une mesure de risque géologique et de la probabilité de l'existence du pétrole et de son exploitation dans les conditions économiques et les technologies actuelles.

- Les réserves sont largement utilisées également les trois valeurs  $1P$ ,  $2P$  et  $3P$  dérivées des précédentes qui évaluent également de manière probabiliste le potentiel de réserves d'un champ<sup>1</sup> :

- Le  $P1$  est généralement égale au  $P90$  ou au  $P95$ .
- Le  $P2$  est toujours égale au  $P50$ .
- Le  $P3$  est généralement égale au  $P10$  ou au  $P5$ .

- Les trois termes correspondent le plus souvent mais les exceptions sont légions aux valeurs respectives  $1P$ ,  $(2P \sim 1P)$  et  $(3P - 2P - 1P)$  ce qui signifie, en inversant le système, que l'on a :

- $1P$  = prouvées.
- $2P$  = prouvées + probables.
- $3P$  = prouvées + probables + possibles.

---

1. Denis BABUSIAUX « Recherche et production du pétrole et du gaz » Editions TECHNIP.P96-99.

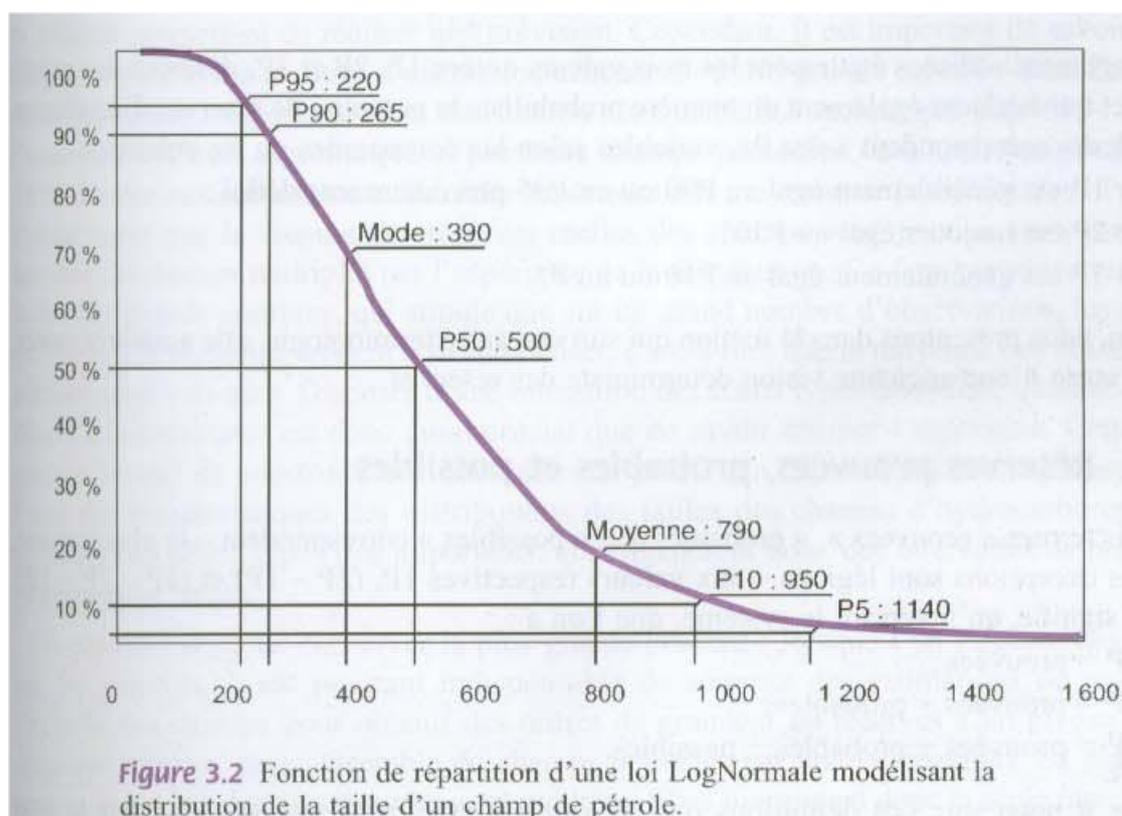


FIGURE 1.3: Fonction de répartition d'une loi log Normale modélisant la distribution de la taille d'un champ de pétrole.

- **Les réserves prouvées ou  $F_{95}$  :**

C'est la quantité de pétrole qui sera exploitée avec les moyens actuels, avec une probabilité de 95%. Les réserves prouvées correspondent aux quantités d'hydrocarbures récupérables avec une quasi-certitude, à partir de réservoirs connus, forés aux conditions économiques et technologiques du moment.

- **Les réserves probables ou  $F_{50}$  :**

Sont les quantités de pétrole qui seront produites, mais avec une probabilité de 50%, ce sont les quantités d'hydrocarbures potentiellement récupérables des réservoirs connus.

- **La réserve possible ou  $F5$  :**

Sont les quantités de pétrole produites avec une probabilité de 5% ce sont les quantités d'hydrocarbures susceptibles d'être découvertes à partir de réservoirs encore inconnus, et extraites à des conditions techniques et économiques envisageables pour les années à venir

- **Les réserves ultimes :**

Les réserves ultimes sont constituées par l'addition des réserves prouvées et possibles et probables sont à l'ordre de 215 milliards de T de pétrole brut, et environ 278000 milliards de m<sup>3</sup> de gaz naturel.

- **Les réserves non conventionnelles :**

Les réserves non conventionnelles correspondant aux schistes bitumineux, et au pétrole extra lourd. Le potentiel théorique de ces réserves est de l'ordre de grandeur des réserves précédentes.

## **1.3 Les risque et couts dans le secteur pétrolier :**

### **1.3.1 Type de couts en exploitation / production :**

#### **1. Le coût géologique :**

Le prix de revient d'une équipe géologique dépend beaucoup des conditions de vie et des facilités de déplacement.on compte par compagnie et non par mois . Les dépenses géologiques couvrent environ de 10 à 30% du coût total d'exploration.

#### **2. Le coût de la géophysique :**

Quant à la géophysique,elle représente en moyenne 25 à 35% des dépenses de l'exploration. Leur distribution est comme suit : 1 à 2% magnétométrie, 1 à 3% gravimétrie et 95 à 98 p.100 sismique.

### 3. Le coût d'un forage :

Les facteurs déterminants dans le coût d'un forage sont :

► **Les données techniques** : type d'appareils et la durée du puits, elles sont à leur tour fonction de :

- La zone géographique (terre/mer).
- Les conditions climatiques.
- L'accès et la nature de la zone (vierge ou mature) .
- Les puits : sont de type (d'exploration ou de développement), leur profondeur et leurs caractéristiques techniques (puits horizontal ...).
- Disponibilité des appareils de forage et des supports en mer.

► **Les données économiques** : (internationales ou régionales).

Les coûts d'un forage représentent la part la plus importante des dépenses d'exploration avec 65 à 80%. Ils comportent tous les frais depuis la préparation jusqu'à la mise en production ou l'abandon du puits. Ils sont distribués comme suit :

Génie civil :9%, forage : 50 à 60%, consommables : 14 à 17% autres opérations : 10 à 15% et les charges indirectes : 9%.a En moyenne, le coûts d'un appareil de forage est de 180 000\$/mois.

### 4. Le coût de développement :

La mise en production d'un puits nécessite le développement du gisement. Cette phase couvre, en termes de dépenses, de 40 à 60% du coût technique du pétrole. Celle-ci s'étale sur plusieurs années (2 à 4 années généralement).

Après une étude d'optimisation, on commence à mettre en place : des forages de développement, qui représente de 60 à 85 % du cout de développement, ainsi que les installations de surface, de récupération et de transport .

### 5. Le coût d'exploitation :

Les coûts d'exploitation, appelés également coûts opératoires, ou OPEX (Operating Expenditures). A distinguer des CAPEX (Capital Expenditures) qui sont les investissements. Les OPEX sont alors tous les décaissements propres au fonctionnement d'une installation. Il sont distribués comme suit : les supports généraux des filiales :20%, les opérations puits/surface : 15%, la logistique : 15%, d'autres ( l'inspection, la sécurité... etc) : 35%

### **1.3.2 Les risques dans le secteur pétrolière :**

#### **1. Définition :**

L'exploitation pétrolière est une activité à forte intensité en capital dont le niveau de risque est élevé. Le risque est défini comme étant la possibilité qu'un fait ayant des conséquences non souhaitables se produise.

Le risque est défini aussi comme étant la possibilité qu'un projet ne se déroule pas conformément aux prévisions.

Les écarts qui peuvent survenir dans les coûts, par exemple, ou les dates d'achèvement par rapport aux prévisions étant considérés comme inacceptables.

#### **2. Types de risques :**

##### **2.1. Le risque géologique :**

C'est le risque inhérent à la nature du sous-sol, c'est-à-dire à la disposition des couches souterraines.

C'est possible que les études géologiques et géophysiques s'entendent sur l'existence des hydrocarbures et qu'après le forage on ne trouve rien, on évoque dans ce cadre, l'aspect aléatoire caractérisant l'existence des hydrocarbures dans le site prospecté.

Dans le cas d'un puits positif, il reste à savoir si le volume d'accumulation d'hydrocarbures est suffisant pour que son exploitation soit économiquement rentable. Un autre facteur qui s'inscrit dans la même logique du risque géologique est la nature des hydrocarbures en place (huile, gaz ou condensat) et leur qualité (léger ou lourd ; gaz humide ou sec) ; c'est la nature qui détermine le prix de ces produits sur le marché.

##### **2.2. Le risque économique :**

L'évaluation de la rentabilité des projets d'investissement repose essentiellement sur le prix de vente du pétrole, c'est le facteur déterminant les recettes procurées par l'exploitation du gisement.

Comme le prix est sujet à plusieurs facteurs, il n'est pas déterminé seulement par la simple superposition de l'offre et de la demande mais également par la prise en considération des paramètres politico-économiques (choc pétrolier...). D'où la difficulté de l'estimer d'une manière déterministe.

A cela s'ajoute le risque d'évaluation des investissements et des frais opératoires. Cette évaluation se trouve biaisée par l'inflation.

### **2.3. Le risque pays :**

Ce risque est d'ordre politique. Il englobe les éléments suivant :

- Risque de souveraineté : l'existence d'une loi claire et bien déterminée. Stabilité politique et fiscale (changement de régime, décision de nationalisation...etc.) ;
- Nature des relations internationales.

### **2.4. Le risque associé :**

C'est le risque relatif au partenaire choisi en fonction de ses compétences pour l'accomplissement des tâches. Parfois, on se trouve face à des situations où les envergures technologiques et financières sont limitées. Ce qui joue à la défaveur de la rentabilité du projet.

### **2.5. Le risque du marché :**

En ce qui concerne le pétrole, le risque de marché est pratiquement inexistant, car le pétrole est un produit qui se vend relativement très bien sur le marché et quel que soit son prix.

Mais ce n'est pas le cas pour le gaz, où le risque du marché est tout à fait réel, du fait de la nature du produit qui pose des problèmes de transport.

### **2.6. Le risque de charge :**

Si la valeur du dollar US augmente par rapport aux devises de référence, cela favorise le secteur pétrolier (une hausse de 5% du dollar US a le même effet qu'une hausse du brut de 1 dollar le baril).

### **Conclusion :**

Le pétrole a cette particularité d'être à la fois une matière première (la pétrochimie et la fabrication des matières plastiques par exemple) et une source d'énergie dont l'usage est aussi bien industriel que domestique. L'ensemble de ses atouts lui a conféré un poids considérable dans l'économie mondiale et le commerce international si bien qu'aujourd'hui le pétrole se présente comme une marchandise stratégique (sens militaire) et spéculative (sens économique). Ainsi le pétrole représente un enjeu économique impliquant l'intervention réfléchie des acteurs (États, compagnies) dans le contexte planétaire.

**Organigramme de l'activité Exploration – Production :** Notre stage pratique est effectué à Boumerdes au niveau du département Finance et Planification de la division Exploration de SONATRACH. Créée en 1965, cette activité est l'une des importantes étapes par laquelle passe l'industrie pétrolière.

La Division Exploration est organisée comme suit :

- Une Direction Assets Est.
- Une Direction Assets Centre.
- Une Direction Assets Ouest.
- Une Direction Assets Nord.
- Une Direction Assets en Partenariat.
- Une Direction Etudes et Synthèses.
- Une Direction des Opérations d'Exploration.
- Une Direction Data Management.
- Une Direction Projet Simulation et Evaluation des Puits d'Exploration.
- Une Direction Projet Ressources Non Conventionnelles.
- Une Direction Planification.
- Une Direction Finances.
- Une Direction Gestion du Personnel.
- Une Direction Logistique.
- Un Département Finances.
- Un Département Juridique.

Organigramme de la Division Exploration est donné dans la figure 1.2 :

### **La Division Exploration (EXP)**

La Division Exploration a pour missions essentielles :

- La conduite et le développement des activités de prospection et de recherche des Hydrocarbures ;

-La participation avec les autres Divisions aux appels d'offres d'exploration en Algérie et à L'étranger ;

- La participation à l'évaluation des offres de partenariat sur des projets d'exploration en Algérie et à l'étranger ;

- La mise en œuvre de la stratégie de la Société en matière d'exploration ;

- La préparation, l'établissement et la recommandation des programmes techniques D'exploration et leur suivi ;

Le développement et la conduite des travaux d'analyse en matière de géologie et de Géophysique ;

- La gestion et le suivi des contrats en effort propre et en association ;
- Le développement d'expertise dans le domaine de l'exploration.

Direction Planification

La Direction Planification a pour missions :

-La mise en œuvre des orientations stratégiques et directives arrêtées par le management en Matière d'élaboration des plans et budgets ;

- L'élaboration des projets de plans à court et moyen termes en collaboration avec les structures concernées et leur diffusion au sein de la Division ;

- L'analyse et l'évaluation des programmes d'activités et des budgets proposés ;

- Le suivi de la réalisation des plans et budgets arrêtés ;

- L'établissement des bilans d'exécution des plans ;

- La réalisation de toute évaluation technico-économique des projets d'exploration et Asset, en effort propre de SONATRACH ;

- La consolidation et le reporting requis par le management et par les autorités compétentes conformément au nouveau cadre législatif.

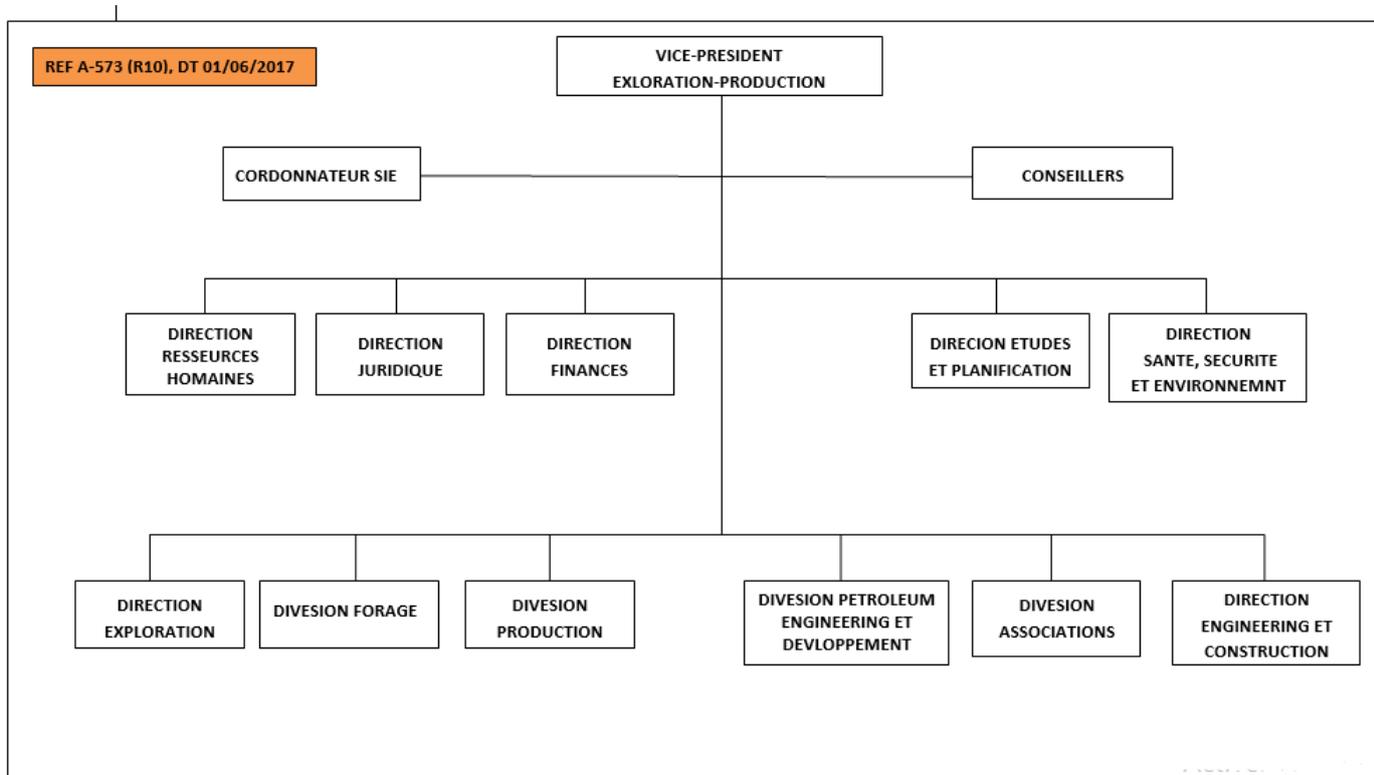


FIGURE 1.4: Organigramme de la SONATRACH

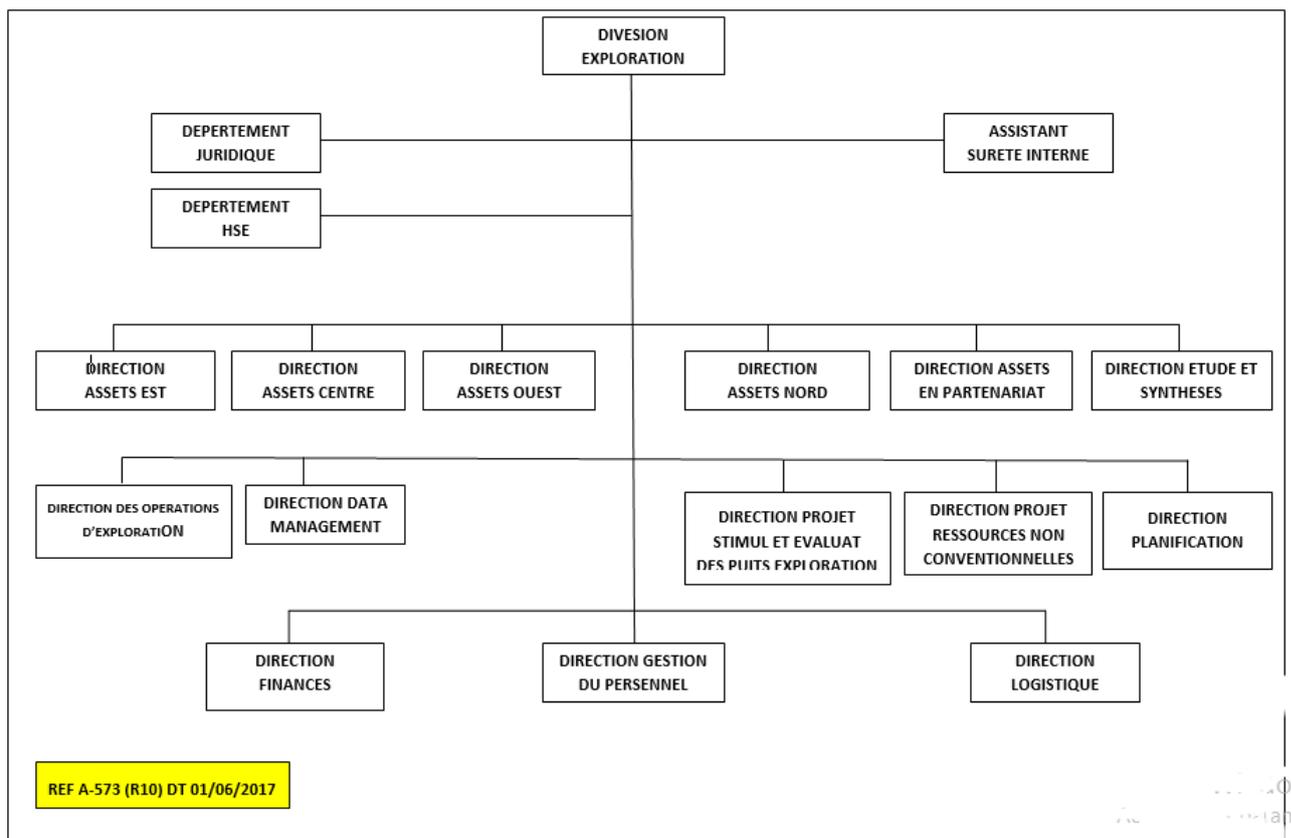


FIGURE 1.5: Organigramme de la Division Exploration.

# 2

## Aspect théorique des séries chronologiques

### Introduction

La théorie des séries chronologiques (ou temporelles) est appliquée de nos jours dans des domaines aussi variés que l'économétrie, la médecine ou la démographie. Dans ce chapitre, plusieurs concepts importants liés à l'analyse des séries chronologiques seront abordés. Parmi ceux-ci, on retrouve les notions de bruit blanc, d'auto-corrélation, de stationnarité et opérateurs définis sur une série chronologique.

### 2.1 Définition d'une série chronologique

Une série statistique est désignée comme étant chronologique quand les données qui la constituent sont les valeurs d'une variable enregistrée en fonction de la date pendant une certaine période.

**Définition :** Une série chronologique, chronique ou série temporelle, est une suite finie de données indexée par le temps.

Si  $(t_1, t_2, \dots, t_n)$  sont les  $n$  instants successifs d'observations et si  $X_{t_i}$  est la valeur

mesurée à l'instant  $t_i$ , on notera la série chronologique  $(X_{t_i}), t_i \in T$  où  $T$  est un ensemble ordonné  $T = (t_1, t_2, \dots, t_n)$

### 2.1.1 Objectifs de l'analyse d'une chronique

Les objectifs de l'analyse d'une chronique sont :

#### 1. La modélisation : elle consiste à

- développer des modèles permettant de décrire le comportement d'une ou plusieurs séries chronologiques.
- de mettre au point une méthodologie pour spécifier, estimer, valider (juger) un modèle approprié pour des données particulières.

#### 2. La prévision :

Étant donnée des observations  $X_1, \dots, X_T$ , la prévision consiste à évaluer une valeur non observée,  $X_{T+h}$ . La prévision peut être ponctuelle, ou par intervalle.

### 2.1.2 Les composantes d'une série chronologique

La première étape dans l'étude des séries chronologiques consiste à représenter graphiquement l'événement qui permet d'entrevoir les quatre composantes fondamentales d'une chronique :

1. **Tendance**  $Z_t$  : Représente l'évolution à long terme de la grandeur étudiée, et traduit l'aspect général de la série.
2. **Les variations saisonnières**  $S_t$  : Sont liées au rythme imposé par les saisons météorologiques (production agricole, consommation de gaz,...), ou encore par des activités économiques et sociales (fêtes, vacances, solde, etc). Elles sont de nature périodique, c'est-à-dire qu'il existe un entier  $p$ , appelé période, tel que  $S_t = S_{t+p}$  pour tout  $t \geq 1$ .
3. **Cycle**  $C_t$  : Regroupe les variations autour de la tendance avec des alternances de phases d'expansion et de récession. Ces phases durent généralement plusieurs années, mais n'ont pas de durée fixe.
4. **Les variations accidentelles ou résiduelles**  $\varepsilon_t$  : les variations accidentelles sont des fluctuations irrégulières et imprévisibles. Elles sont supposées en général de faible amplitude. Elles proviennent de circonstances non prévisibles : catastrophes naturelles, crise boursière, grèves,...

Le graphique suivant présente l'ensemble des composantes précitées

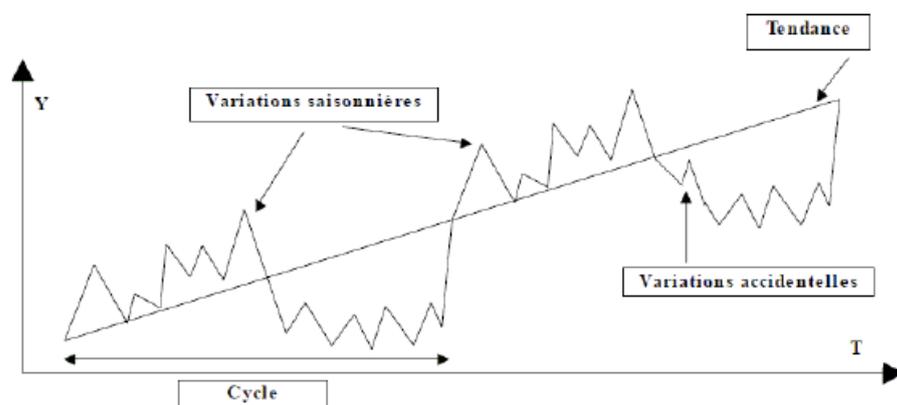


FIGURE 2.1: les composantes d'une série chronologique

### 2.1.3 Les schémas de composition

La décomposition d'une série temporelle en composantes déterministes ( $Z_t$  et  $S_t$ ) et aléatoire  $\varepsilon_t$  est un phénomène important car l'étude d'une série brute ne peut donner d'indications précisées que dans le cas où, à chaque instant on peut faire la part des influences respectives de la tendance du cycle, de la saisonnalité et de résidu, pour cela, les composantes du mouvement peuvent se combiner en trois schéma de décomposition :

#### 1. Schémas additif :

Dans ce modèle, la série chronologique, s'écrit de la façon suivante :

$$X_t = Z_t + S_t + C_t + \varepsilon_t$$

La tendance, les variations saisonnières, les variations accidentelles et cyclique, ont un effet additif dans ce modèle de série chronologique.

## 2. Schéma multiplicatif :

Dans ce modèle, la série chronologique, s'écrit de la façon suivante :

$$X_t = Z_t * S_t * C_t * \varepsilon_t$$

La tendance, les variations saisonnières, les variations accidentelles et cyclique, ont un effet multiplicatif sur la taille de la série chronologique. Ce modèle peut être ramené à un modèle additif par transformation logarithmique :

$$\log(X_t) = \log(Z_t) + \log(S_t) + \log(C_t) + \log(\varepsilon_t)$$

## 3. Schéma mixte :

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

$$X_t = Z_t * S_t * C_t + \varepsilon_t$$

- La tendance  $Z_t$  prend des différentes forme :

- a. **Linéaire** :  $Z_t = a + bt$ .
- b. **Quadratique** :  $Z_t = a + bt + ct^2$ .
- c. **Exponentielle** :  $Z_t = ae^{bt}$ .

### 2.1.4 Procédure de choix d'un schéma de décomposition

Afin d'identifier le type de schéma correspondant à la chronique étudiée, on fait référence à certains tests parmi lesquels :

- **Le test graphique de la bande** : permet de détecter une relation entre la volatilité et la moyenne. Il consiste à relier toutes les valeurs hautes de la chronique par une ligne plafond, puis toutes les valeurs basses par une ligne plancher. Si les deux lignes sont parallèles, l'écart-type ne varie pas avec la moyenne et le schéma est additif. Au contraire, si les deux lignes se rapprochent ou s'écartent, le schéma est multiplicatif.

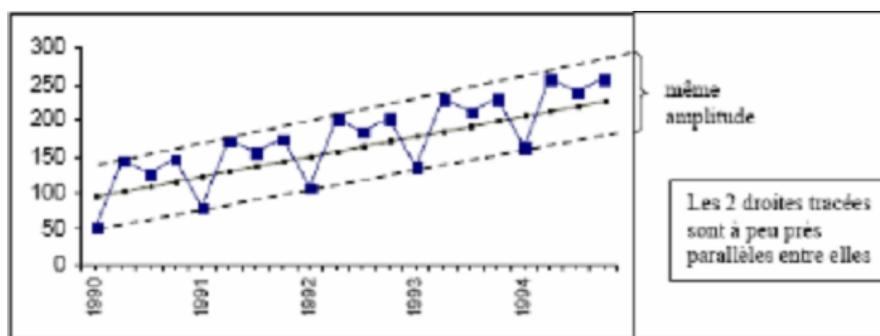


FIGURE 2.2: Exemple d'un schéma additif

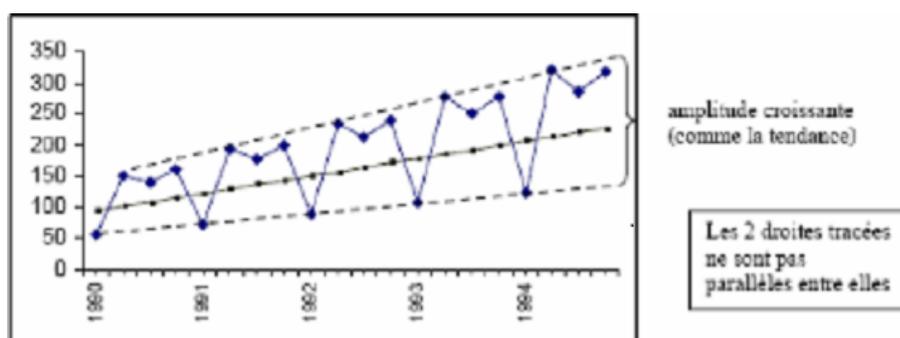


FIGURE 2.3: Exemple d'un schéma multiplicatif

- **Le test de Buys-Ballot** : le test de Buys-Ballot est fondé sur les méthodes analytiques prenant en compte le calcul des moyennes  $\bar{X}_i$  et les écarts types  $\sigma_i$  pour chaque période (année).

pour ces paires d'observations  $(\bar{X}_i, \sigma_i)$ , nous estimons par la méthode des Moindres Carrés Ordinaires les paramètres  $a$  et  $b$  de l'équation :  $\sigma_i = a + b\bar{X}_i + \varepsilon_i$ .

si le coefficient  $b$  n'est pas significativement différent de 0 ( test de Student) alors on accepte l'hypothèse d'un schéma additif; dans le cas contraire, nous retenons un schéma multiplicatif.

## 2.2 Caractéristiques d'une série chronologique

Parmi les caractéristiques les plus importantes d'une série chronologique  $(X_t)_{t \in T}$  :

### 2.2.1 Moyenne et Variance :

Soit une série temporelle  $(X_t)_{t \in Z}$ ,  $t = 1, \dots, T$ . Les expressions de la moyenne et de la variance  $\forall t \in T$  sont :

- **Moyenne** : qui est un moment d'ordre un, est donnée par :

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

- **Variance** : la variance mesure la dispersion de  $X_t$  autour de sa moyenne, donnée par :

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

### 2.2.2 Fonction d'auto-covariance :

Soit  $(X_t)_{t \in T}$  un processus aléatoire de variance finie. on définit la fonction d'auto-covariance de  $(X_t)_{t \in T}$  pour tout  $t$ , par :

$$\gamma(h) = cov(X_t, X_{t+h}) = E[(X_t - E(X_t))(X_{t+h} - E(X_{t+h}))] \in Z$$

Cette fonction mesure la covariance entre deux valeurs de  $(X_t)_{t \in T}$  séparées par un certain délai. Elle fournit des informations sur la variabilité de la série et sur les liaisons temporelles qui existent entre les différentes composantes de la série  $(X_t)_{t \in T}$ .

### 2.2.3 Fonction d'auto-corrélation (AC)

La fonction d'auto-corrélation d'un processus  $(X_t)_{t \in T}$  de moyenne  $E(X_t) = \mu$ , notée  $\rho(h)$ , est défini par :

$$\rho(h) = \rho_h = \frac{\text{cov}(X_t, X_{t+h})}{\sqrt{V(X_t)}\sqrt{V(X_{t+h})}}$$

Cette fonction mesure la corrélation des séries avec elle-même décalée de k période.

La représentation graphique de  $\rho(h)$  est appelée corrélogramme.

### 2.2.4 Fonction d'auto-corrélation partielle (PAC)

La fonction d'auto-corrélation partielle mesure la corrélation entre  $X_t$  et  $X_{t-h}$  l'influence des variables  $X_{t-h+i}$  (pour  $i < h$ ) ayant été retirée. Notons  $\rho_h$  et  $\rho_{hh}$  les fonctions respectivement d'auto-corrélation et d'auto-corrélation partielle de  $X_t$ . Soit  $P_h$  la matrice symétrique formée des  $(h - 1)$  premières auto-corrélations de  $X_t$  :

$$P_h = \begin{pmatrix} 1 & \rho_1 & \cdot & \cdot & \cdot & \rho_{h-1} \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{h-1} & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

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

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

Ou  $|P_h^*|$  est le déterminant de la matrice  $P_h^*$  et  $P_h^*$  est donné par :

$$P_h^* = \begin{pmatrix} 1 & \rho_1 & \cdot & \cdot & \rho_{h-2} & \rho_1 \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \rho_{h-1} & \cdot & \cdot & \cdot & \cdot & \rho_h \end{pmatrix}$$

$P_h^*$  est ainsi la matrice  $P_h$  dans laquelle on a remplacé la dernière colonne par le vecteur  $[\rho_1 \dots \rho_h]$ .

## 2.3 Analyse d'une série chronologique

Les modèles de prévision traditionnels sont basés sur l'extrapolation des tendances passées ie : utilisent le passé de la variable elle-même . seul le passé de la variable est utilisé en vue de la prévoir sans apport d'information extérieur.

### 2.3.1 Les opérateurs retard, avance et différenciation

- **L'opérateur Retard :**

On aura souvent à considérer une variable en fonction de son passé. Il est donc commode de définir un opérateur qui transforme une variable  $X_t$  en fonction de sa valeur passée. C'est l'opérateur retard désigné par la lettre B et tel que :

$$BX_t = X_{t-1} \text{ et } B^k(X_t) = X_{t-k}$$

- **L'opérateur d'avance (forward) :**

L'opérateur d'avance est l'inverse de l'opérateur de retard, noté **F** (Forward) il transforme une observation en sa valeur futur, il est défini par :

$$F(X_t) = X_{t+1} \text{ et } F^n(X_t) = X_{t+n}$$

- **L'opérateur de différence ordinaire :**

L'opérateur de différence ordinaire noté  $\nabla$  associé au processus  $(X_t)_{t \in T}$  tel que :

$$\nabla X_t = X_t - X_{t-1} = X_t - (BX_t) = (1 - B)X_t \quad \forall t \in T$$

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

$$\nabla^d X_t = (1 - B)^d X_t$$

- **L'opérateur de différence saisonnière  $\nabla_s$  :**

L'opérateur est définit comme suit :

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

En d'autre termes :

$$\nabla_s = (1 - B^s)$$

## 2.3.2 Les effets des opérateur

L'opérateur  $\nabla$  :

permet d'éliminer la tendance de la série .

peut être répété plusieurs fois si la tendance n'est pas linéaire

L'opérateur  $\nabla_s$  :

permet d'éliminer la saisonnalité de période  $s$

on peut également l'appliquer plusieurs fois

## 2.4 Processus aléatoire stationnaire

### 2.4.1 Définition d'un processus aléatoire

Les processus aléatoires ou stochastiques notés  $\{X_t, t \in T\}$  décrivent l'évolution d'une grandeur aléatoire en fonction du temps, pour chaque instant du temps, la valeur de la quantité étudiée  $X_t$  est appelée variable aléatoire et l'ensemble des valeurs  $X_t$  quand  $t$  varie appelé processus aléatoire.

La chronique est dite échantillon ou réalisation du processus aléatoire et ce dernier est appelé le processus générateur de la chronique.

### 2.4.2 Processus aléatoire stationnaire

On dit qu'un processus est stationnaire si ses caractéristiques ne varient pas avec la définition de l'origine du temps, ou encore, si ses caractéristiques statistiques ne varient pas au cours du temps. On définit :

#### 1. La stationnarité au sens strict (ou stationnarité forte) :

Le processus  $\{X_t, t \in T\}$  est stationnaire au sens strict si pour tout  $(t_1, t_2, \dots, t_n)$  avec  $t_i \in T, i = 1, 2, \dots, n$  et si pour tout  $h \in T$  avec  $t_{i+h} \in T, (X_{t_1}, \dots, X_{t_n})$  a la même distribution de probabilité jointe que  $(X_{t_1+h}, \dots, X_{t_n+h})$ .

Un processus strictement stationnaire a toutes ses caractéristiques (tous ses moments) invariantes dans le temps.

#### 2. La stationnarité au second ordre (ou stationnarité faible) :

Un processus  $\{X_t, t \in T\}$  est faiblement stationnaire si :

- $E(X_t) = m$ , constante et indépendante de  $t$ ,  $t \in T$
- $Var(X_t) < \infty$ ,  $t \in T$  (finie et indépendante de temps).
- $Cov(X_t, X_{t+h}) = E[(X_t - m)(X_{t+h} - m)] = \gamma(h)$  indépendante de  $t$ .

### 3. le processus Bruit Blanc (white noise) :

Parmi la classe des processus stationnaires, il existe des processus particuliers dont la structure est la plus simple, dits bruit blanc.

Un processus bruit blanc  $(\varepsilon_t)_{t \in T}$  est une suite de variables aléatoires non corrélées de moyenne nulle et de variance constante  $\sigma^2$  est donc caractérisé par la fonction d'autocovariance suivante :

$$\mathbf{E}(\varepsilon_t \varepsilon_{t+h}) = \begin{cases} \sigma^2 & \text{si } h = 0, \forall t \in T \\ 0 & \text{si } h \neq 0, \forall t \in T \end{cases}$$

$$0 \rightarrow \text{si } h \neq 0, \forall t \in T$$

On parle souvent de bruit blanc gaussien, il s'agit d'un bruit blanc dont la distribution marginale suit une loi normale.

## 2.5 Les processus aléatoire non stationnaire

Les processus stochastiques non stationnaire sont caractérisés par des propriétés stochastiques qui évoluent en fonction du temps. Il existe deux sortes de non stationnarité : la non stationnarité de nature déterministe ( $TS$ ) et la non stationnarité de nature stochastique ( $DS$ )

### 2.5.1 Les processus ( $TS$ ) ( trend stationary)

Un processus  $TS$  s'écrit  $x_t = f_t + \varepsilon_t$  ou  $f_t$  est une fonction polynômiale du temps, linéaire ou non linéaire, et  $\varepsilon_t$  un processus stationnaire.

Le processus ( $TS$ ) s'écrit comme :  $x_t = a_0 + a_1 t + \varepsilon_t$

ce processus ( $TS$ ) est non stationnaire car  $E(x_t)$  dépend du temps.

connaissant  $\hat{a}_0$  et  $\hat{a}_1$ , le processus  $x_t$  peut être fonctionnariser en retranchant de la valeur de  $x_t$ , en  $t$ , la valeur estimée  $\hat{a}_0 + \hat{a}_1 t$

## 2.5.2 Les processus ( $DS$ ) (différence stationary)

Les processus  $DS$  sont des processus que l'on peut rendre stationnaire par l'utilisation d'un filtre aux différences :

$$(1 - D)^d x_t = \beta + \varepsilon_t$$

Où :

$\varepsilon_t$  est un processus stationnaire.

$\beta$  une constante réelle,  $B$  est l'opérateur décalage (opérateur de retard), et  $d$  l'ordre du filtre aux différences.

Ces processus sont souvent représentés en utilisant le filtre aux différences premières ( $d = 1$ ), le processus est dit alors processus du premier ordre .il s'écrit :

$$(1 - D)^1 x_t = \beta + \varepsilon_t \iff x_t = x_{t-1} + \beta + \varepsilon_t$$

L'introduction de la constante dans le processus  $DS$  permet de définir deux processus différents :

$\beta = 0$  : le processus  $DS$  est dit sans dérive il s'écrit :  $x_t = x_{t-1} + \varepsilon_t$  comme  $\varepsilon_t$  est un bruit blanc, ce processus porte le nom de modèle de marche au hasard ou de marche aléatoire (*Random Walk Model*). Il est très fréquemment utilisé pour analyser l'efficience des marchés financiers.

Pour fonctionnariser la marche aléatoire, il suffit d'appliquer le filtre aux différences premières :

$$x_t = x_{t-1} + \varepsilon_t \iff (1 - D)x_t = \varepsilon_t$$

$\beta \neq 0$  le processus porte alors le nom de processus  $DS$  avec dérive, il s'écrit :

$$x_t = x_{t-1} + \beta + \varepsilon_t$$

La stationnarisation de ce processus est réalisée en utilisant le filtre aux différences premières :

$$x_t = x_{t-1} + \beta + \epsilon_t \iff (1 - D)x_t = \beta + \epsilon_t$$

## 2.6 Modélisation des séries temporelles

L'objet 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.

### 2.6.1 Les processus linéaire stationnaire

#### 1. Processus autorégressif $AR(p)$ :

On appelle processus autorégressif d'ordre  $p$ , noté  $AR(p)$ , un processus stationnaire  $(X_t)_{t \in T}$  vérifiant une relation de type :

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

ou encore :

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

- Les  $\phi_i$  sont des réels positifs ou négatifs à déterminer.
- $\epsilon_t$  est un bruit blanc de variance  $\sigma^2$ .

#### Notion stationnarité :

Le modèle autorégressif est stationnaire si et seulement si les racines  $Z$  de la fonction caractéristique  $\phi_B = 0$  sont en module supérieur à 1, c'est-à-dire :  $\|Z\| > 1$ . Ou  $\phi_B$  est le polynôme de retard de degré  $p$  donnée par :

$$\phi_B = 1 - \sum_{j=1}^p \phi_j B^j, \phi_j \in R \text{ et } \phi_p \neq 0$$

Un processus  $(X_t, t \in T)$  satisfaisant une représentation  $AR(p)$  est toujours inversible.

#### 2. Processus moyenne mobile $MA(q)$ :

On appelle processus moyenne mobile d'ordre  $q$ , noté par  $MA(q)$ , un processus  $(X_t)_{t \in T}$  stationnaire vérifiant une relation du type :

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

Où :

les  $\theta_j$  sont des réels positifs ou négatifs à déterminer.

$\varepsilon_t$  est le bruit blanc de variance  $\sigma^2$

en introduisant l'opérateur au retard on obtient :

$$X_t = \Theta(B)\varepsilon_t \text{ où } \Theta(B) = 1 - \sum_{j=0}^q \theta_j B^j, \theta_0 = 1, \theta_j \in \mathbb{R}, \forall j < q : \theta_q \in \mathbb{R}^*$$

Avec :

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

Un processus  $\{X_t, t \in T\}$  satisfaisant une représentation  $MA(q)$  est toujours stationnaire.

Le processus  $MA(q)$  est inversible si les racines  $\vartheta_i, i = 1, \dots, q$  de polynôme  $(1 - \vartheta_1 Z - \vartheta_2 Z^2 - \dots - \vartheta_q Z^q)$  sont en module  $> 1$

### 3. Processus mixtes $ARMA(p, q)$ :

Ces processus constituent une extension naturelle des processus  $AR$  et  $MA$ . Ce sont en effet des processus mixtes au sens où ils incorporent simultanément des composantes  $AR$  et  $MA$  ce qui permet d'obtenir une description plus parcimonieuse des données.

#### Définition :

un processus stationnaire  $\{X_t, t \in T\}$  suit un  $ARMA(p, q)$  s'il vérifie la relation suivante :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)\varepsilon_t$$

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

Les processus  $ARMA(p, q)$  peuvent donc se mettre :

- (a) Sous la forme  $MA(\infty)$  en écrivant :  $X_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t$ , si toutes les racines de  $\Phi$  sont à l'extérieur de l'unité.

- (b) Ou sous la forme  $AR(\infty)$  en écrivant :  $\Theta^{-1}(B)\Phi(B)X_t = \varepsilon_t$  , si toutes les racines de  $\Theta$  sont à l'extérieure du l'unité.

## 2.6.2 Les processus non stationnaires

### 1. Les processus $ARIMA(p, d, q)$ et sa version saisonnière $SARIMA$ :

Les processus  $ARIMA$  (AutoRegressive Integrated Moving Average) sont la généralisation des modèles  $ARMA$  pour des processus non stationnaires, admettant une tendance stochastique . Un modèle  $ARIMA$  est d'ordre  $(p,d,q)$ ,  $p$  correspondant au nombre de termes autorégressifs,  $d$  au nombre de différenciations, et  $q$  au nombre de termes de moyenne mobile, il vérifie la relation suivante :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

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

### 2. Processus $SARMA$ ( $ARMA$ saisonniers) :

Les processus  $SARMA$  ou  $ARMA$  saisonniers représentent les mêmes caractéristiques qu'un modèle  $ARMA(p, q)$  mais il contient l'effet saisonnier, on les notes  $ARMA_s(p, q)$  et il s'écrit :

$$(1 - \phi_{1s} B^s - \dots - \phi_{ps} B^{ps}) X_t = (1 - \theta_{1s} B^s - \dots - \theta_{qs} B^{qs}) \varepsilon_t$$

$$\Phi_s(B^s)(1 - B)^d X_t = \Theta_{s'}(B^s) \varepsilon_t$$

Avec :

$\Theta_s(B)$  représente le polynôme moyen mobile saisonnier.

$\Phi_s(B^s)$  représente le polynôme autorégressif saisonnier.

$s$  et représente la période de la saisonnalité.

### 2.6.3 Processus à effet ARCH :

On considère un processus ARCH pour modélisé les résidu d'un linéaire. Prenons l'exemple d'un modèle linéaire ARMA avec résidu de type ARCH(P) soit :

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

On dit que  $X_t$  est un processus ARMA avec effet ARCH d'ordre P si :

$$\varepsilon_t = Z_t * \sqrt{h_t} \text{ avec}$$

$$h_t = V(\varepsilon_t/\varepsilon_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-1}^2$$

L'exemple le plus simple c'est  $X_t$  est un AR(1) avec effet ARCH(1) :

$$X_t = \Phi X_{t-1} + \varepsilon_t \text{ avec } \varepsilon_t = Z_t * \sqrt{h_t}$$

$$\text{Et } h_t = V(\varepsilon_t/\varepsilon_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-1}^2$$

-  $\Phi(B)X_t = \Theta(B)\varepsilon_t$  s'appelle équation de la moyenne.

-  $h_t = V(\varepsilon_t/\varepsilon_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-1}^2$  s'appelle équation de la variance.

## 2.7 Étude de la stationnarité d'une série temporelle

### • Méthode graphique :

La représentation graphique permet de détecter la présence d'une saisonnalité, tendance ou d'une rupture (modification), ainsi l'observation du correlogramme d'après l'étude de la fonction d'autocorrélation.

✓ Dans le correlogramme de la fonction d'autocorrélation partiel des pics important apparaissent due aux retards  $S, 2S, \dots$ , ce qui fait signe de la présence d'une saisonnalité de période  $S$ .

✓ La manière dans la quelle la fonction d'autocorrélation décroît montre la présence de période  $S$ .

La représentation graphique de la série peut (dans certains cas) nous indiquer que la série n'est pas stationnaire, mais elle ne nous permet pas de discriminer entre les deux types de non stationnarité (la présence de non stationnarité de type stochastique peut produire une série avec une allure croissante, de la même manière que la non stationnarité de type déterministe).

### • Méthode analytique :

✓ **Analyse de la tendance** : Les tests de stationnarité permettent de vérifier si une série est stationnaire ou non. Il y a deux types de test différents, les tests de stationnarité, comme le test *Kwiatkowski Phillips SchmidtShin (KPSS)*, pour lesquels l'hypothèse nulle  $H_0$  est que la série est stationnaire, et les tests de racine unitaire comme

le test de *Dickey–Fuller* et le test augmenté de *Dickey–Fuller* (*ADF*), pour lesquels l’hypothèse nulle est que la série a été générée par un processus présentant une racine unitaire, et donc, qu’elle n’est pas stationnaire. **Tests de racine unitaire** : Il existe plusieurs tests de racine unitaire, parmi lesquels nous citons les tests de *Dickey–Fuller* simple et augmenté

▷ **Test de *Dickey – Fuller* simple (*DF*) :**

Dicky et Fuller considèrent trois modèles de base pour la série  $X_t$ ,  $t = 1, \dots, T$ . Les trois modèles de base sont :

**Modèle [1]** : modèle sans constante ni tendance déterministe :

$$(1 - \phi_1 B)X_t = \varepsilon_t \Rightarrow X_t = \phi_1 X_{t-1} + \varepsilon_t, \phi_1 \in \mathfrak{R}$$

**Modèle [2]** : modèle avec constante sans tendance déterministe :

$$(1 - \phi_1 B)(X_t - \mu) = \varepsilon_t \Rightarrow X_t = \phi_1 X_{t-1} + c + \varepsilon_t, \\ \text{tel que } c = (1 - \phi_1 \mu)$$

**Modèle [3]** : modèle avec constante et tendance déterministe :

$$(1 - \phi_1 B)(X_t - \alpha - \beta t) = \varepsilon_t \Rightarrow X_t = \phi_1 X_{t-1} + c + b_t \varepsilon_t,$$

tel que :

$$c = \alpha(1 - \phi_1) + \beta \phi_1$$

$$b = (1 - \phi_1)\beta$$

$B$  : opérateur de retard et  $\varepsilon_t$  un bruit blanc.

▷ **Test de *Dickey – Fuller* augmenté (*ADF*) :**

Dans les modèles précédents utilisés pour les tests de *Dickey–Fuller* simple, le processus  $\varepsilon_t$  est par hypothèse un bruit blanc. Or les tests de *Dickey–Fuller* augmentés prennent en compte l’hypothèse d’existence d’autocorrélation des erreurs (ce test est utilisé pour blanchir les résidus).

Les différents modèles retenus pour construire les statistiques de tests sont les suivants :

$$\text{Modèle [4]} : \Delta X_t = \rho X_{t-1} + \sum_{i=1}^p \lambda_i \Delta X_{t-1} + \varepsilon_t$$

$$\text{Modèle [5]} : \Delta X_t = \rho X_{t-1} + c + \sum_{i=1}^p \lambda_i \Delta X_{t-1} + \varepsilon_t$$

$$\text{Modèle [6]} : \Delta X_t = \rho X_{t-1} + c + b_t + \sum_{i=1}^p \lambda_i \Delta X_{t-1} + \varepsilon_t$$

$$\text{Où } \rho = (1 - \phi_1), c = (1 - \phi_1)\mu, b = (1 - \phi_1)\beta$$

## 2.8 Conclusion

L'étude d'une série temporelle à pour but de : modéliser le mécanisme de production de cette série ; et prévoir l'évolution au futur.

L'intérêt de notre travail est bien l'étude prévisionnelle d'un ensemble des séries temporelles, dans ce qui va suivre on s'intéresse au traitement de quelques méthodes de prévision afin de pouvoir faire une étude comparative.

# 3

## Méthodes de prévision

### Introduction

Dans ce chapitre nous examinerons la notion de la prévision et son importance dans les ressources pétrolière et le prix moyenne du pétrole ainsi les différentes approches de prévisions pour les investissements qui à engagé tel que la méthodologie de Box et Jenkins, la prévisions par le modèle ARFIMA, et la régression linéaire.

### 3.1 Une prévision ?

Prévoir, c'est observer un ensemble de données qui permet d'envisager une situation future et d'entreprendre des actions pour y parer concrètement autrement dit c'est porter un jugement sur les évènements ou évolutions possibles à venir en utilisant comme outils le passé et le présent.

Il en résulte que les prévisions sont toujours entachées d'erreur et qu'il est possible d'en établir plusieurs pour un même évènement à venir (qui constitue l'objet de la prévision). Des outils d'analyse doivent être donc développés afin de comparer et de hiérarchiser les prévisions pour discerner ce qui fait qu'on puisse, ou non, en qualifier certaines de "bonnes".

### 3.1.1 Les méthodes de prévision

- **Les méthodes extrapolatives** : Ces méthodes utilisent le passé de la variable elle-même nous citons à titre d'exemple : modélisation ARMA, les modèles ARFIMA.
- **Les méthodes explicatives** : Celles-ci utilisent les valeurs passées et présentes d'une ou de plusieurs variables pour prévoir  $y$ . Parmi ces méthodes on peut citer : la régression linéaire.

## 3.2 La méthodologie de Box et Jenkins :

La technique de Box-Jenkins consiste à identifier le modèle approprié susceptible de représenter la série, elle se base sur la forme des fonctions d'autocorrélation simple et d'autocorrélation partielle de la série étudiée, afin de choisir un modèle ou éventuellement plusieurs modèles qui seront examinés à tour de rôle, on aura donc besoin de critères pour le choix. La modélisation uni-variée de Box et Jenkins concerne les processus ARMA, ARIMA ou SARIMA. Le choix du modèle résulte d'une procédure itérative qui comporte trois étapes : identification, estimation et validation du modèle. Une fois le modèle déterminé nous pouvons faire des prévisions.

### 3.2.1 L'identification :

- **La familiarisation avec les données :**

Cette étape consiste à identifier le modèle ARMA susceptible de représenter la série, c'est pour cela qu'il est important de se familiariser avec les données en examinant le graphe de la série chronologique (présence de saisonnalité, stationnarité,...) qui permet de faire une analyse préliminaire qui consiste par exemple à corriger les données aberrantes, transformer les données (transformation logarithmique, inverse, racine carrée,...) puisqu'il faut se ramener à une série stationnaire, le recours aux différences premières ordinaires, différences premières saisonnières, différences ordinaires et saisonnières. Le choix de la transformation est dicté par l'allure graphique de la série. D'ailleurs le choix de la transformation des données est plus facile après avoir appliqué les opérateurs de différence adéquats. Il est conseillé de comparer les variances des différentes séries. Comme l'inspection des autocorrélations partielles (PAC) donne une idée sur l'ordre du modèle autoregressif et celle des autocorrélations simples (AC) donne une idée sur l'ordre du modèle moyenne mobile.

Plus précisément si la fonction d'autocorrélations simple décroît rapidement vers zéro et la fonction d'autocorrélations partielle présente un cut-off après  $p$  retard, on peut conclure que la série provient d'un processus autorégressif d'ordre  $p$  ( $AR(p)$ ). Par contre si la fonction d'autocorrélation simple présente un cut-off après un  $q$  retards et que la fonction d'autocorrélation partielle décroît rapidement vers zéro, alors on peut déduire que la série est générée à partir d'un modèle moyenne mobile d'ordre  $q$  ( $MA(q)$ ). Evidement il se peut aussi que les fonctions d'autocorrélations simple et d'autocorrélations partielle présentent une forme exponentielle ou sinusoidal, on constate dans ce cas qu'on est en présence d'un processus autorégressif moyenne mobile ARMA ( $p, q$ ).

### •Critère d'information :

Ils existent des critères d'informations qui sont utilisés, comme guide, dans le choix du modèle, ce qui nous permet d'éviter la sélection arbitraire des paramètres  $p$  et  $q$  du modèle.

Parmi ces critères, ils existent les critères d'information qui mesurent l'écart entre la vraie loi inconnu et celle du modèle proposé. Les estimateurs de la quantité d'information qui ont été proposés sont :

#### 1. Critère d'Akaike (1970) : appelé aussi AIC

Akaike a démontré que le meilleur des modèles ARMA non filtré est celui qui minimise la statistique :

$$AIC = \log(\sigma^2) + \frac{2(p+q)}{N}$$

#### 2. Critère de Schwartz (1978) : appelé aussi SCH tel que :

$$SCH = \log(\sigma^2) + (p+q) \frac{\log(N)}{N}$$

Le critère de choix consiste à minimiser AIC et/ou SCH.

### 3.2.2 L'estimation des paramètres :

Une fois l'étape de l'identification terminée, il faut estimer les paramètres qui sont les coefficients des polynômes AR et MA ainsi que les polynômes saisonniers SAR et SMA, et la variance des innovations .

La méthode d'estimation la plus utilisée est celle des moindres carrés ou bien la

méthode du maximum de vraisemblance. Cette dernière a donnée beaucoup de satisfaction aussi bien pour la précision des résultats obtenus que pour la stabilité et la rapidité des calculs. Quand l'étape d'estimation est achevée, l'étape suivante va nous permettre de valider le(s) modèle(s) estimé(s).

### 3.2.3 La validation du modèle :

A l'étape de l'identification, les incertitudes liées aux méthodes employées font que plusieurs modèles en général sont estimés et c'est l'ensemble de ces modèles qui subit alors l'épreuve des tests, il existe de très nombreux tests permettant de comparer les performances entre modèles ; nous pouvons citer les tests sur le modèle, les tests sur les paramètres et les tests sur les résidus.

#### 3.2.1 Tests des paramètres :

Tous les coefficients du modèle retenu doivent être significativement différents de zéro, il convient donc d'utiliser le test de *Student* classique. Il s'agit dans cette étape de tester la significativité des paramètres  $\phi_i$  et  $\theta_j$  (avec  $i = 1; \dots; p$  et  $j = 1; \dots; q$ ) dans la formulation obtenue. Nous rejeterons avec un risque  $\alpha\%$  l'hypothèse que le paramètre est nul si :

$$\frac{|\hat{\phi}_i|}{\sqrt{\text{var}(\hat{\phi}_i)}} > t_\alpha \quad (t_\alpha = 1.96 \text{ Si } \alpha = 5\%)$$

#### 3.2.2 Tests sur les résidus :

Pour que les modèles obtenus soient valides, il convient de vérifier que les résidus estimés suivent bien un bruit blanc et ceci à l'aide des tests suivants :

- **Test "Portmanteau" (Box-Pierce 1970) :**

Le test de Box-Pierce, a pour objet de tester le caractère non autocorrélé des résidus. Ce test est fondé sur la statistique  $Q = \sum_{h=1}^K \hat{\zeta}_h^2(\hat{\epsilon})$ , ou  $\hat{\zeta}_h(\hat{\epsilon})$  est le coefficient d'autocorrélation d'ordre  $h$  des résidus estimé, et  $K$  c'est le nombre maximal de retards. les hypothèses de ce test sont les suivants :

$H_0 : \zeta_1 = \dots = \zeta_h = 0$  non significativement différent de zéro.

$H_1 : \exists j$  tel que  $\zeta_j \neq 0$  significativement différent de zéro.

Les règles de test sont :

— Si :  $Q \geq \chi_{1-\alpha}^2 (K - p - q)$  on accepte  $H_0$

— Si :  $Q > \chi_{1-\alpha}^2 (K - p - q)$  on rejette  $H_0$

Avec  $\chi_{1-\alpha}^2 (K - p - q)$  est le quantile d'ordre  $(1 - \alpha)$  de loi  $\chi_{1-\alpha}^2$  de  $(K - p - q)$  degrés de liberté.

#### • Test de Lyung et Box :

Lyung et Box ont proposé une modification qui améliore l'approximation Q.

On appelle cette statistique, de Lyung et Box au retard K :

$$Q_k^k = n(n+2) \sum k = 1^k \frac{\theta_k^2}{n-k} < \chi_{0.05(k)^2}$$

- Si  $Q < \chi_{1-\alpha}^2 (K - p - q)$  alors la série comporte comme un bruit blanc.
- Si  $Q > \chi_{1-\alpha}^2 (K - p - q)$  alors la série ne comporte pas comme un bruit blanc.

#### • Test de normalité :

Il convient de vérifier la normalité des résidus, ceci est possible par un test de Jarque et Bera (1984), qui est basé sur le skewness (coefficient d'asymétrie de la distribution) et la kurtosis (aplatissement-épaisseur des queues).

En notant  $\mu_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$  le moment d'ordre k de la distribution, on appelle skewness le coefficient  $\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}$  et kurtosis  $\beta_2 = \frac{\mu_4}{\mu_2^2}$ .

Sous des hypothèses de normalité ; on a :

$$\beta_1 \rightarrow N(0, \sqrt{\frac{6}{n}}) \text{ et } \beta_2 \rightarrow N(0, \sqrt{\frac{24}{n}})$$

Le test de Jarque-Bera sur le fait que, si la distribution suit une loi normale ; alors la quantité  $S = \frac{n}{6} \beta_1^2 + \frac{n}{24} (\beta_2 - 3)^2$  suit asymptotiquement une loi du  $\chi_{1-\alpha}^2$  à 2 degré de liberté.

Ainsi si  $S \geq \chi_{1-\alpha}^2$  on rejette l'hypothèse  $H_0$  de normalité des résidus au seuil  $\alpha$

### 3.2.4 La prévision :

Pour identifier le processus étudié à un processus ARMA, on a appliqué différentes transformations (exemple différenciation dans le cas d'une série I(1)), il nécessaire

lors de la phase de prévision de prendre en compte la transformation retenue et de recolorer la prévision. Plusieurs cas sont possibles :

Si le processus contient une tendance déterministe, on extrait cette dernière par régression afin d'obtenir une série stationnaire lors de la phase d'estimation. Ensuite, lors de la phase de prévision, on adjoint aux prévisions réalisées sur la composante ARMA stationnaire, la projection de la tendance.

Si la transformation résulte de l'application d'un filtre linéaire (de type par exemple différences premières), on réalise les prévisions sur les séries filtrées stationnaires et l'on reconstruit ensuite par inversion du filtre les prévisions sur la série initiale.

Prenons l'exemple d'un processus ARMA(1,1) :

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

Avec  $\phi_1 < 1$  et  $\theta_1 < 1$  ( stationnarité et inversibilité ).

Calculons les prévisions pour divers horizons.

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

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

On en déduit la relation suivante donnant la série des prévision récursives :

$$\hat{X}_{t+h} = \phi_1 X_{t+h-1} \quad \forall h > 1$$

L'erreur de prévision est donnée par :

$$\hat{e}_{t+h} = X_{t+h} - \hat{X}_{t+h}$$

### 3.2.5 L'interprétation des résultats :

Elle n'est pas toujours aisée. Les éléments les plus importants, pour lesquels il faudra trouver une explication, sont les opérations de différence utilisés et éventuellement les constantes. Ceux-ci déterminent en effet le comportement de la prévision à long terme. Le polynôme autorégressif joue également un rôle essentiel sur la fonction de prévision. En effet, pour celui-ci les racines réelles induisent une composante amortie dans la fonction de prévision. Mais, les racines complexes donnent lieu à une composante pseudo-périodique amortie.

### 3.3 Prévision par le modèle ARFIMA

Dans le domaine des méthodes d'analyse des séries temporelles, il arrive souvent qu'on modélise des séries à mémoire longue au moyen de processus à mémoire courte tels que les modèles ARMA. Ceci revient alors à approximer la fonction d'autocorrélation qui décroît hyperboliquement au moyen d'une somme d'exponentielles. Même si une telle procédure est toujours possible, elle ne conduit pas à un modèle parcimonieux puisqu'il est nécessaire de considérer des retards très élevés dans la modélisation ARMA. Cette difficulté peut être résolue grâce à l'introduction des processus ARFIMA (Autoregressive Fractionnally Integrated Moving Average) dont la caractéristique essentielle est la présence d'un paramètre d'intégration fractionnaire prenant explicitement en compte le comportement de long terme de la série.

#### 3.3.1 Les modèles ARFIMA(p,d,q)

Le modèle général pour représenter des séries chronologiques stationnaires ou non stationnaires comportement homogène peut être présenté sous la forme suivante :

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t \quad (3.1)$$

Ce modèle est alors appelé modèle autoregressif moyenne mobile fractionnaire intégré (AutoRegressive Fractionally Integrated Moving Average) ou modèle ARFIMA(p,d,q). Un modèle fractionnaire intégré a pour caractéristique une dépendance entre des observations éloignées.

**Remarque** L'expression (3.1) peut s'écrire sous la forme suivante :

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t \quad (3.2)$$

Où

$$Y_t = \nabla_d X_t$$

. Ainsi dans l'expression (3.2),  $\Theta(B)$  est un opérateur autoregressif stationnaire, donc le processus  $Y_t$  est un processus stationnaire.

#### 3.3.2 Stationnarité des modèles ARFIMA(p,d,q)

Il s'agit d'obtenir les conditions sur "d" pour qu'un processus ARFIMA(p,d,q) soit stationnaire, et pour cela il suffit d'étudier le comportement d'un processus ARFIMA

d'ordre (0,d,0) qui s'appelle "bruit fractionnaire intégré" et qui satisfait la relation :

$$(1 - B)^d X_t = \varepsilon_t \quad (3.3)$$

L'expression (3.3) est équivalente à :

$$X_t = (1 - B)^{-d} \varepsilon_t \quad (3.4)$$

Le développement en série de  $(1 - B)^{-d}$  est donné par :

$$(1 - B)^{-d} = \frac{(-d)(-d-1)\dots(-d-j+1)}{j!} (-B)^j$$

$$(1 - B)^{-d} = \frac{(d)(d+1)\dots(d+j-1)}{j!} (-B)^j$$

Notons :

$$\frac{\Gamma(d+j)}{\Gamma(d)} = (d)(d+1)\dots(d+j-1)$$

avec  $\Gamma(n) = (n-1)!$

L'expression (3.4) peut être représentée sous la forme :

$$X_t = \sum_{j=0}^{+\infty} b_j \varepsilon_{t-j} \quad (3.5)$$

Et

$$b_j = \frac{\Gamma(d-j)}{\Gamma(d)\Gamma(j+1)}$$

Le comportement asymptotique des coefficients ( $\alpha_j$ ) se déduit directement de la formule de Sheppard :

$$\frac{\Gamma(j-\alpha)}{\Gamma(j-b)} \sim j^{\alpha-b} \text{ si } j \rightarrow \infty$$

Alors

$$b_j \sim \frac{j^{d-1}}{\Gamma(d)} \text{ si } j \rightarrow \infty$$

D'où

$$b_j^2 \sim \frac{j^{2d-2}}{\Gamma(d)^2} \text{ si } j \rightarrow \infty$$

La suite des  $b_j$  est de carré sommable si et seulement si :

$$2d - 2 < -1 \Leftrightarrow d < \frac{1}{2}$$

Donc on déduit qu'un processus ARFIMA(p,d,q) est stationnaire si et seulement si  $0 < d < 1$

On peut alors établir une classification des séries temporelles d'après les résultats du théorème précédent en fonction des valeurs du paramètre d'intégration fractionnaire d :

- ) Lorsque  $d \geq \frac{1}{2}$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  est un processus non stationnaire.
- Si  $d = 0$ , le processus ARFIMA (p,0,q) se réduit au processus ARMA standard et exhibe uniquement une mémoire de court terme (ne présente aucune structure de dépendance à long terme).
- Si  $\frac{-1}{2} < d < 0$  le processus est anti-persistant, les autocorrélations alternent de signe et la densité spectrale est dominée par des composantes de haute fréquence (la densité spectrale tend vers zéro lorsque la fréquence tend vers zéro).

### 3.3.3 Les modèles longue mémoire

Les processus ARMA s'appellent souvent les processus courte mémoire parce que leurs covariances décroissent rapidement. D'autre part, un processus longue mémoire est caractérisé par la propriété suivante : la fonction d'autocovariance décroît plus lentement c'est-à-dire : n'est pas absolument sommable

$$\sum_{k=-\infty}^{+\infty} |\gamma(k)| = \infty \quad (3.6)$$

Ils sont appelés ainsi en raison de la corrélation forte entre les observations largement séparées dans le temps. Ils permettent donc d'identifier les phénomènes de persistance. Les modèles de séries chronologiques longue mémoire ont attirés beaucoup d'attention récemment car ces processus modélisent une grande variété de phénomènes et il y a maintenant une réalisation croissante que les caractéristiques de possession de longue mémoire de séries chronologiques surgissent dans des domaines aussi divers que des sciences économiques, géophysique ... etc.

### 3.3.4 Estimation par la méthode de Geweke Porter-Hudak (1983)

Cette méthode est l'une des méthodes d'estimation semi-paramétrique du paramètre d'intégration fractionnaire d'un processus ARFIMA (p,d,q). Dans la suite, nous présentons cette méthode dite aussi log-périodogramme proposée par Geweke et Porter-Hudak (1983).

#### Principe de la méthode :

D'une manière générale, on considère un processus scalaire,  $\{X_t\}_{t \in Z}$  stationnaire, dont la densité spectrale est de la forme suivante sur l'intervalle  $[0, 2\pi[$  :

$$f(\lambda) = |2\sin(\frac{\lambda}{2})|^{-2d} f^*(\lambda) \quad (3.7)$$

Où

$$f^*(\lambda) = \frac{\sigma^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2} \quad (3.8)$$

Où d est le paramètre de mémoire compris dans l'intervalle  $(\frac{-1}{2}, \frac{1}{2})$  et  $f^*$  est une fonction continue bornée sur tout l'intervalle  $[0, 2\pi[$ . le paramètre d contrôle le comportement de la densité spectrale dans un voisinage de zéro alors que  $f^*$  contrôle le comportement de courte mémoire.

En calculant l'équation (3.7) aux fréquences de Fourier :  $\lambda_j = 2\pi_j/N$

pour  $j = 0, \dots, N-1$ , où N est la taille de l'échantillon, et par passage aux logarithmes, on obtient alors :

$$\log I_N(\lambda_j) = -2d \log |2\sin(\frac{\lambda_j}{2})| + \log f^*(0) + \log \frac{f^*(\lambda_j)}{f^*(0)} + \log \frac{I_N(\lambda_j)}{f^*(\lambda_j)} \quad (3.9)$$

Où  $I_N(\lambda_j)$  est le périodogramme (L'estimateur asymptotiquement sans biais de la densité spectrale ( $\lambda$ )) calculé à la fréquence  $\lambda_j$ , défini par l'expression :

$$I_N(\lambda_j) = \frac{1}{2\pi N} |\sum e^{i\lambda_j t} (X_t - \bar{X}_N)|^2 \quad (3.10)$$

On considère alors l'équation (2.3) lorsque T tend vers l'infini, j étant fixé.

L'estimateur GHP nécessite à ce niveau deux hypothèses cruciales, relatives au comportement asymptotique des éléments de l'équation (2.3)

(H1) : pour des fréquences suffisamment basses ( $\lambda \rightarrow 0$ ) le terme  $\log \frac{f^*(\lambda_j)}{f^*(0)}$

est négligeable.

(H2) : la suite des terme  $\log \frac{I_N(\lambda_j)}{f(\lambda_j)}$  pour  $j = 1, \dots, m$  est asymptotiquement indépendante distribuée (i.i.d), le nombre  $m$  de fréquence considère est alors appelé la largeur de bande.

Sous les hypothèses (H1) et (H2), l'estimateur GPH du paramètre de mémoire  $d$ , noté  $\hat{d}_{GPH}$ , s'obtient alors en considérant la régression linéaire simple, pour  $j = 1, \dots, m$  :

$$\log I_N(\lambda_j) = \alpha + \beta Z_j + \varepsilon_i \quad (3.11)$$

Où

- $\alpha = \log f^*(0) - \gamma$
- $Z_j = -2d \log |2 \sin \left( \frac{\lambda_j}{2} \right)|$
- $\varepsilon_i = \log \frac{I_N(\lambda_j)}{f(\lambda_j)}$

l'estimateur GPH est alors explicitement défini par l'égalité suivante :

$$\hat{d}_{GPH} = \frac{\sum_{j=1}^m (Z_j - \bar{Z}) \log I_N(\lambda_j)}{\sum_{j=1}^m (Z_j - \bar{Z})^2} \quad (3.12)$$

Geweke et Porter-Hudak (1983) montrent que, quand  $\frac{-1}{2} < d < \frac{1}{2}$ , la loi de l'estimateur  $\hat{d}_{GPH}$  de  $d$  tend vers une loi normale lorsque  $N \rightarrow \infty$  :

$$\hat{d}_{GPH} \sim N \left( d, \pi^2 \left[ 6 \sum_{j=1}^m (Z_j - \bar{Z})^2 \right]^{-1} \right)$$

### 3.4 Prévision par La régression linéaire

#### Introduction :

La régression linéaire est une modélisation linéaire qui permet d'établir des estimations dans le futur à partir d'informations provenant du passé. Dans ce modèle de régression linéaire, on a plusieurs variables dont une qui est une variable explicative et les autres qui

sont des variables expliquées. Cet outil est utilisé pour les analyses techniques boursières mais aussi pour la gestion de budgets. Elle est souvent calculée avec la méthode des moindres carrés qui permet de réduire les erreurs en ajoutant de l'information.

### 3.4.1 La régression linéaire simple :

Un modèle de régression linéaire simple est de la forme :

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (3.13)$$

Où

- ▶ Y est la variable dépendante (une v.a.).
- ▶  $\beta_0$  et  $\beta_1$  sont les coefficients (ordonnée à l'origine et pente).
- ▶ X est la variable indépendante (variable explicative).
- ▶  $\varepsilon$  est une erreur aléatoire

Les hypothèses relatives à ce modèle sont les suivantes :

1. la distribution de l'erreur  $\varepsilon$  est indépendante de X ou X est fixe,
2. l'erreur est centrée et de variance constante (homoscédasticité) :

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \forall i = 1 \dots n$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = \delta_{i,j} \sigma^2 \text{ pour tout couple } (i,j), (\delta_{i,j} \text{ est le symbole de Kronecker, i.e. } \delta_{i,j} = 1 \text{ si } i = j, \delta_{i,j} = 0 \text{ si } i \neq j)$$

3.  $\beta_0$  et  $\beta_1$  sont constants, pas de rupture du modèle.

### 3.4.2 La régression linéaire multiple :

Le modèle de régression linéaire multiple est l'outil statistique le plus habituellement mis en œuvre pour l'étude de données multidimensionnelles. Cas particulier de modèle linéaire, il constitue la généralisation naturelle de la régression simple.

Un modèle de régression linéaire multiple est de la forme suivante :

$$y = \beta_0 + \sum_{j=1}^p \beta_j x_j + \varepsilon \quad (3.14)$$

Où :

y est la variable à expliquer (à valeurs dans R);  $x_1, \dots, x_p$  sont les variables explicatives (à valeurs dans R);  $\varepsilon$  est le terme d'erreur aléatoire du modèle;  $\beta_0, \beta_1, \dots, \beta_p$  sont les paramètres à estimer.

La désignation "multiple" fait référence au fait qu'il y a plusieurs variables explicatives  $x_j$  pour expliquer  $y$ .

Pour  $n$  observations, on peut écrire le modèle de régression linéaire multiple sous la forme : pour  $i = 1, \dots, n$

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{i,j} + \varepsilon_i \quad (3.15)$$

On suppose que :  $\varepsilon_i$  est une variable aléatoire, non observée,  $x_{i,j}$  est observé et non aléatoire et  $y_i$  est observé et aléatoire.

On fait les trois hypothèses additionnelles suivantes :

**(A1)**

$$\begin{aligned} \mathbb{E}[\varepsilon_i] &= 0, \forall i = 1, \dots, n \text{ où de manière équivalent :} \\ \mathbb{E}[y_i] &= \beta_0 + \sum_{j=1}^p \beta_j x_{i,j}, \forall i = 1, \dots, n \end{aligned}$$

Indique que les erreurs sont centrées.

**(A2)**

$$\begin{aligned} \mathbb{V}[\varepsilon_i] &= \sigma^2, \forall i = 1, \dots, n \text{ où de manière équivalent :} \\ \mathbb{V}[y_i] &= \sigma^2, \forall i = 1, \dots, n \end{aligned}$$

Indique que On parle d'hypothèse d'homoscédasticité ( $\simeq$  homogénéité des variances). Cette variance  $\sigma^2$  est un paramètre du modèle qu'il faudra estimer.

**(A3)**

$$\begin{aligned} \text{Cov}(\varepsilon_i, \varepsilon_{i'}) &= 0, \forall i \neq i' \text{ où de manière équivalent :} \\ \text{Cov}(y_i, y_{i'}) &= 0, \forall i \neq i' \end{aligned}$$

Sous cette hypothèse, les termes d'erreur  $\varepsilon_i$  sont non corrélés.

On peut écrire matriciellement le modèle (3) de la manière suivante :

$$Y = \beta X + \varepsilon \quad (3.16)$$

Où

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \cdot & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & \cdot & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \cdot & \cdot & \cdot & x_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix}$$

$Y$  désigne le vecteur à expliquer de taille  $n$ ,  $X$  la matrice explicative de taille  $n * (p + 1)$ , et  $\varepsilon$  le vecteur d'erreurs de taille  $n$ .

Les hypothèses peuvent alors s'écrire sous forme matricielle :

(A1')

$$\begin{aligned}\mathbb{E}[\varepsilon] &= 0_n, \text{ où de manière équivalent :} \\ \mathbb{E}[Y] &= X\beta \in \mathbb{R}^n.\end{aligned}$$

(A2')

$$\begin{aligned}\mathbb{V}[\varepsilon] &= \sigma^2 I_n, \text{ où de manière équivalent :} \\ \mathbb{V}[y_i] &= \sigma^2 I_n.\end{aligned}$$

Dans la suite, on suppose que :  $n > (p+1)$  et  $\text{rang}(X) = p+1$

On a donc plus d'observations que de variables et il n'existe pas de liaison linéaire entre les variables explicatives  $x_j$  c'est à dire pas de multicolinéarité.

### 3.4.3 Estimation de $\beta$ par les moindres carrés :

On cherche  $\hat{\beta} \in \mathbb{R}^{p+1}$  qui minimise la somme des erreurs quadratiques

$$\varepsilon_i^2 = (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2$$

On doit donc résoudre le problème d'optimisation suivant :

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^n [y_i - (\beta_0 + \sum_{j=1}^p \beta_j x_{ij})]^2 \quad (3.17)$$

► **Résolution du problème d'optimisation** : Le problème d'optimisation est :

$$\min_{\beta \in \mathbb{R}^{p+1}} F(\beta)$$

avec

$$\begin{aligned}F(\beta) &= \sum_{i=1}^n [y_i - (\beta_0 + \sum_{j=1}^p \beta_j x_{ij})]^2 \\ &= (Y - X\beta)^T (Y - X\beta) \\ &= Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta\end{aligned} \quad (3.18)$$

Le minimum est atteint pour  $\frac{\partial F(\beta)}{\partial \beta} = 0$ .

**Rappels.** Soient  $a$  et  $x$  deux vecteurs de dimension  $K$ , et soit  $A$  une matrice de dimension  $K \times K$ . on a :

$$\frac{\partial a^T x}{\partial x} = \frac{\partial x^T a}{\partial x} = a \text{ et } \frac{\partial a^T A x}{\partial x} = 2Ax \text{ si } A \text{ est symétrique.}$$

► **Solution du problème d'optimisation** : on en déduit après quelque manipulation

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{3.19}$$

Sous réserve que  $X^T X$  soit inversible,

$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j x_{ij}$  est appelé la valeur prédite et  $\hat{\varepsilon}_i = y_i - \hat{y}_i$  est appelé le résidu.

En notant  $x_i^T = (1, x_{i1}, \dots, x_{ip})$ , la valeur prédite  $\hat{y}_i$  s'écrit :  $\hat{y}_i = x_i^T \hat{\beta}$

on dit que le minimum de F est égale à  $\sum_{i=1}^n \hat{\varepsilon}_i^2$ . Ce minimum est appelé **la somme des carrés des résidus (SCR)**. Et la valeur prédite  $\hat{y}_i$  estimé  $\mathbb{E}[y_i] = (\beta_0 + \sum_{j=1}^p \beta_j x_{ij})$  et non pas  $y_i$ . Une meilleur notation serait  $\widehat{\mathbb{E}}[y_i]$ .

Aucune des hypothèses n'a été utilisée ici pour obtenir  $\hat{\beta}$ .

► **Propriétés de  $\hat{\beta}$**  : Sous les hypothèses (A1)' et (A2)', on peut montrer que :

- $\mathbb{E}[\hat{\beta}] = \beta$  , on dit que l'estimateur  $\hat{\beta}$  est sans biais.
- $\mathbb{V}[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$ , on dit que de variance minimale parmi tous les estimateurs linéaires par rapport à Y ) sans biais (propriété dite *Gauss – Markov*)

### 3.4.4 Test d'hypothèses :

On veut maintenant tester la nullité des coefficients  $\beta_j$  du modèle de régression. Pour faire ces tests, il est nécessaire de faire une hypothèse supplémentaire :

(A3)'

$\varepsilon \sim N(0_n, \sigma^2 I_n)$  ou de manière équivalente

$Y \sim N(X\beta, \sigma^2 I_n)$  (L'unique "nouveauté" ici est la normalité).

► **Test de signification du modèle** : typiquement, on commence par tester :

$$H_0 : " \beta_1 = \dots = \beta_p = 0 " \text{ contre } H_1 : " \exists j \in 1, \dots, p, \beta_j \neq 0 "$$

On utilise la statistique suivante :

$$F_n = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 / p}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - p - 1)} = \frac{SCE / p}{SCR / (n - p - 1)}$$

qui est distribuées sous  $H_0$  selon une **loi de Fisher** à p et n-p-1 degrés de libertés. On rejette  $H_0$  avec un risque  $0 < \alpha < 1$  si :

$$F_n \geq f_{1-\alpha}(p, n - p - 1)$$

ou  $f_{1-\alpha}(p, n - p - 1)$  est le fractile d'ordre  $1 - \alpha$  de la loi F(p,n-p-1).

### Significativité d'un coefficient de régression

On veut tester qu'un coefficient  $\beta_i$  vaut certaine valeur de  $\beta_{i0}$  :

$H_0 : \beta_i = \beta_{i0}$  contre  $H_1^A : \beta_i \neq \beta_{i0}$  ou contre  $H_1^B : \beta_i < \beta_{i0}$  ou contre  $H_1^C : \beta_i > \beta_{i0}$

#### Statistique de test

Considérons l'expression de  $\hat{\beta}$ . Un sous vecteur d'un vecteur gaussien est lui-même gaussien, donc  $H_0 : \hat{\beta}_i \sim N(\beta_{i0}, \sigma_0^2)$  où  $\sigma_0^2$  est un élément (i,i) de  $(\sigma_\mu^2(X \cdot X)^{-1})$ .

Sous les quatre présupposées et si  $H_0$  est vérifiée,  $T = (\hat{\beta}_i - \beta_{i0})/s(\hat{\beta}_i) \sim st(n - (p + 1))$ , loi de Student à  $(n-(p+1))$  ddl.

Si l'on choisit  $\beta_{i0} = 0$  et que  $H_0$  est rejetée, on dit que  $\beta_i$  est significatif. En toute rigueur,  $T_0$  suit sous  $H_0$  une loi de student à  $(n-(p+1))$  ddl. La statistique  $\hat{\beta}_i/s(\hat{\beta}_i)$  est appelée *t-statistique*. Des statistique semblent son utilisées dans les modèles de séries temporelles, ou les estimateurs suivant seulement des lois approximativement normales. Nous traiterons la prédictions à la faveur de l'exemple de la section : l'étude exploration.

#### ► Table d'analyse de la variance (ANOVA) :

Source de variance	Somme des carrées	ddl	Carrée moyen	F
Régression (expliquée)	$SCE = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2$	$p$	$\frac{1}{p} \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2$	$\frac{SCE/p}{SCR/(n-p-1)}$
Résiduelle	$SCR = \sum_{i=1}^n (y_i - \hat{y}_n)^2$	$n - (p + 1)$	$\frac{1}{n-p-1} \sum_{i=1}^n (y_i - \hat{y}_n)^2$	/
Total	$\sum_{i=1}^n (y_i - \bar{y}_n)^2$	$n - 1$	$\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2$	/

#### Remarques :

- On retrouve la statistique dite de Fisher  $F_n$  qui permet de tester l'ajustement du modèle.
- On retrouve la propriété fondamentale  $SCT = SCE + SCR$  qui permet de mesurer l'ajustement du modèle par le coefficient de détermination

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

- Le coefficient  $R^2$  donne la proportion de variabilité de y qui est expliquée par le modèle. Plus le  $R^2$  est proche de 1, meilleure est l'adéquation du modèle aux données.

### 3.4.5 Prévision d'une valeur ultérieure :

On désire prévoir à l'aide du modèle la valeur de la variable y pour des observations futures  $(x_{1,0}, \dots, x_{p,0})$  des p variables explicatives. Posons

$$x_0 = (1, x_{1,0}, \dots, x_{p,0})^T \in \mathbb{R}^{p+1}$$

D'après le modèle on a :

$$y_0 = x_0\beta + \varepsilon_0$$

et la prédiction est :

$$\hat{y}_0 = \mathbb{E}[y_0] = x_0^T \hat{\beta}$$

L'erreur de prédiction est défini par  $\hat{y}_0 - y_0$  et on peut montrer que sous les hypothèses du modèle (incluant l'hypothèse de normalité), on a :

$$\hat{y}_0 - y_0 \sim N(0, \sigma^2(1 + x_0^T(X^T X)^{-1}x_0)) \quad (3.20)$$

On en déduit que :

$$\frac{y_0 - \hat{y}_0}{\sigma \sqrt{1 + x_0^T(X^T X)^{-1}x_0}} \sim N(0, 1)$$

. on utilise ce résultat pour construire un **intervalle de prédiction** pour  $y_0$ , c'est à dire l'intervalle  $[A, B]$  tel que :

$$\mathbb{P}(A \leq y_0 \leq B) = 1 - \alpha$$

Ici,  $y_0$  est une variable aléatoire et non pas un paramètre. L'intervalle de prédiction est donc un intervalle dans lequel une future observation  $y_0$  va tomber avec une certaine probabilité (différent d'un intervalle de confiance).

On en déduit l'intervalle de prédiction pour  $y_0$  au niveau de confiance  $1 - \alpha$  suivant :

$$\left[ \hat{y}_0 \pm t_{1-\alpha/2}(n-p-1) s \sqrt{1 + x_0^T(X^T X)^{-1}x_0} \right] \text{ avec } s^2 = \frac{SCR}{n-p-1}$$

## 3.5 Conclusion

Ce chapitre nous a donné une introduction aux méthodes de prévisions qu'on va les appliquer dans le prochain chapitre ou nous allons illustrer ces méthodes afin de faire des prévisions sur différentes séries temporelles.

# 4

## APPLICATION :

### 4.1 Présentation des données :

Dans ce paragraphe, on s'intéresse à étudier la prévision des trois séries chronologiques de taille  $n=50$  observations de 1969-2018, de type pétrolière : série d'investissement sur les projets, série des réserves pétrolières récupérables, et une série des prix moyens de pétrole. Les prévisions seront de 2019-2030.

1. On note la série d'investissement par  $(Inv_t)_t$  dont une représentation graphique est donnée par la figure 1.1.
2. On note la série réserve par  $(Res_t)_t$  dont une représentation graphique est donnée par la figure 1.2.
3. On note la série des prix moyens du pétrole par  $(Prix_t)_t$  dont une représentation graphique est donnée par la figure 1.3.

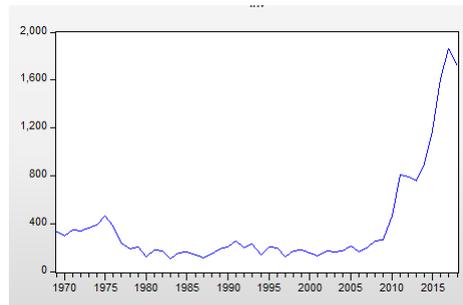


FIGURE 4.1: Évolution de la série d'investissement sur les projets

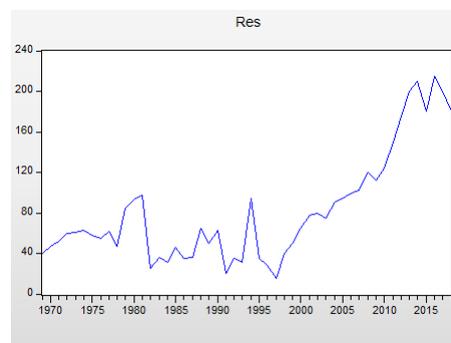


FIGURE 4.2: Évolution de la série réserves pétrolière récupérable

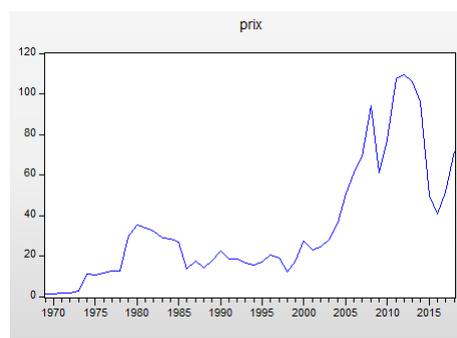


FIGURE 4.3: Évolution de la série des prix moyenne de pétrole

On applique les 3 méthodes de prévision tel que (la méthode de box-jenkins, le modèle ARFIMA(p,d,q) ) sur les trois séries réelles de pétrole. Et faire une modélisations pour chaque série.

Pour chaque séries, on compare les capacités prédictives pour chque méthodes à l'aide du critère de la racine carré de l'erreur quadratique moyenne, noté RMSE ("Root Mean Squared Error"). Ce critère est défini comme suit :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{t+i} - \hat{X}_{t+i})^2}$$

Où h est l'horizon de prévision et  $\hat{X}_{t+i}$  est la valeur prédite de  $(X_{t+i})$ . Pour chaque approche, les prévisions sont calculées pour h=12.

## 4.2 Prévision par la méthodologie de Box et Jenkins

Dans cette section de notre étude on va se focaliser sur les trois séries, et Pour le traitement économétrique de nos séries on utilise le logiciel "Eviews 9"

### 4.2.1 Études de la série d'investissement sur les projets d'exploration

Analyse préliminaire :

Le graphe suivant représente la représentation graphique de La série d'investissement d'exploration en dollar de 1969 jusqu'au 2018, cette série contient 50 observations :

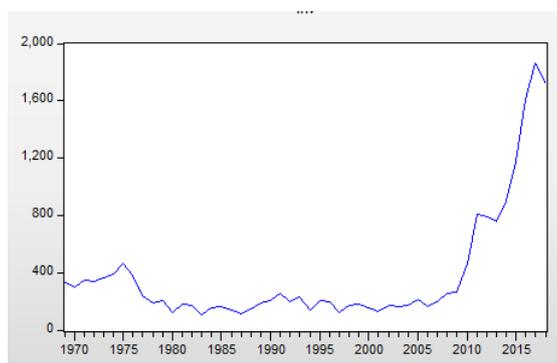


FIGURE 4.4: graphe de la série inv

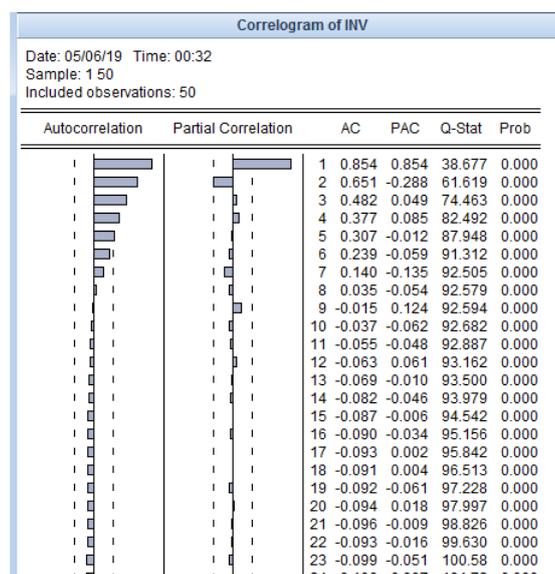


FIGURE 4.5: corrélogramme de la série inv

D'après la représentation graphique ainsi que le correlogramme de la série d'investissement il y a une augmentation visible de la série fluctue d'une allure croissante avec le temps, d'où la série peut être n'est pas stationnaire.

### Dickey-Fuller :

On applique le test de Dickey-Fuller pour infirmer ou confirmer notre hypothèse concernant la stationnarité du processus qui généré la série ( $inv_t$ ), on a besoin de déterminer le retard qui minimise le critère d'Akaike et Schwarz. Par la suite on estimée par la méthode de moindre carrée ordinaire avec le retard égal à 0, donc il suffit d'applique la stratégie de test Dickey-Fuller simple.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INV(-1)	0.044973	0.045378	0.991065	0.3268
C	-55.98250	29.33498	-1.908387	0.0626
@TREND("1969")	2.745958	1.167431	2.352138	0.0230

FIGURE 4.6: MODÈLE [3]

D'après le tableau du modèle [3], nous constatons que la tendance n'est pas significativement différente de zéro puisque sa t-stat est égale à 2.35 est inférieure aux valeurs critiques tubulées par Dickey-Fuller simple qu'est égale à 2.79 aux

seuils 5%. Pour cela on passe au modèle [2] :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INV(-1)	0.096955	0.041500	2.336267	0.0238
C	-5.547841	20.96258	-0.264655	0.7924

FIGURE 4.7: MODÈLE [2]

D'après le modèle [2], nous constatons que la constante n'est pas significativement différente de zéro car sa t-statistique est égale à -0.26 est inférieure aux valeurs critiques tabulé par Dickey-Fuller simple qu'est égale à 2.54 aux seuils 5%. Donc on passe au modèle [1].

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	3.018044	0.9991
Test critical values:		
1% level	-2.613010	
5% level	-1.947665	
10% level	-1.612573	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(INV)  
 Method: Least Squares  
 Date: 06/14/19 Time: 23:19  
 Sample (adjusted): 2 50  
 Included observations: 49 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INV(-1)	0.089336	0.029601	3.018044	0.0041

FIGURE 4.8: MODÈLE [1]

D'après le tableau ci-dessus, nous constatons que la t-statistique qu'est égale à 3.01 est supérieur aux valeurs critiques tabulé de Dickey-Fuller aux seuils 1%, 5% et 10%, donc on accepte l'hypothèse nulle d'existence une racine unitaire, alors la série n'est pas stationnaire de type DS.

Pour rend cette série stationnaire on doit faire un première différenciation :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1))	-0.384855	0.365590	-1.052695	0.2997
D(INV(-1),2)	-0.389952	0.377397	-1.033268	0.3086
D(INV(-2),2)	-0.857776	0.353712	-2.425070	0.0206
D(INV(-3),2)	-0.849280	0.322280	-2.635219	0.0124
D(INV(-4),2)	-0.902947	0.245872	-3.672423	0.0008
D(INV(-5),2)	-0.495018	0.196161	-2.523523	0.0163
C	-85.77064	39.07948	-2.194774	0.0349
@TREND("1969")	3.992965	1.481937	2.694424	0.0108

FIGURE 4.9: MODÈLE [6]

Le modèle [6] indique que la tendance n'est pas significativement différente de zéro puisque sa t-statistique est égale à 2.694424 est inférieure aux valeurs critiques tabulé de Dickey-Fuller Augmenté qu'est égale à 2.79 aux seuils 5%. Donc on passe au modèle [5] :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1))	0.258289	0.300025	0.860891	0.3950
D(INV(-1),2)	-0.897613	0.354292	-2.533538	0.0158
D(INV(-2),2)	-1.283761	0.342820	-3.744706	0.0006
D(INV(-3),2)	-1.139940	0.329036	-3.464483	0.0014
D(INV(-4),2)	-1.064479	0.258353	-4.120253	0.0002
D(INV(-5),2)	-0.547636	0.211477	-2.589575	0.0138
C	13.01565	14.65731	0.887997	0.3804

FIGURE 4.10: MODÈLE [5]

Le modèle [5] indique la que la constante n'est pas significativement différente de zéro car sa t-stat est égale à 0.887997 est inférieure aux valeurs critiques tabulé par Dickey-Fuller Augmenté qu'est égale à 2.54 aux seuils 5%. On passe au modèle [4].

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	1.125761	0.9301
Test critical values:		
1% level	-2.619851	
5% level	-1.948686	
10% level	-1.612036	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(INV,2)  
 Method: Least Squares  
 Date: 06/15/19 Time: 22:19  
 Sample (adjusted): 1976 2018  
 Included observations: 43 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1))	0.325798	0.289402	1.125761	0.2675
D(INV(-1),2)	-0.944406	0.349349	-2.703332	0.0103
D(INV(-2),2)	-1.321478	0.339206	-3.895802	0.0004
D(INV(-3),2)	-1.159939	0.327325	-3.543689	0.0011
D(INV(-4),2)	-1.074996	0.257343	-4.177293	0.0002
D(INV(-5),2)	-0.545343	0.210856	-2.586326	0.0138

FIGURE 4.11: MODÈLE [4]

D'après le tableau ci-dessus, nous constatons que la t-stat est égale à 1.125761 est supérieur aux valeurs critiques tabulé de Dickey-Fuller augmenté aux seuils 1%, 5% et 10%, donc on accepte l'hypothèse nulle de l'existence une racine unitaire, alors la série n'est pas stationnaire.

Donc on passe à la deuxième différenciation :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1),2)	-5.693431	0.717760	-7.932217	0.0000
D(INV(-1),3)	3.941286	0.612088	6.439083	0.0000
D(INV(-2),3)	2.754703	0.499725	5.512437	0.0000
D(INV(-3),3)	1.644533	0.330309	4.978777	0.0000
D(INV(-4),3)	0.574101	0.181478	3.163472	0.0032
C	-63.30507	32.78701	-1.930797	0.0614
@TREND("1969")	2.974421	1.124167	2.645888	0.0120

FIGURE 4.12: MODÈLE [6]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1),2)	-4.960712	0.713867	-6.949068	0.0000
D(INV(-1),3)	3.335306	0.611926	5.450507	0.0000
D(INV(-2),3)	2.309069	0.507206	4.552523	0.0001
D(INV(-3),3)	1.387583	0.340350	4.076932	0.0002
D(INV(-4),3)	0.470827	0.191063	2.464249	0.0185
C	16.21305	14.12926	1.147481	0.2585

FIGURE 4.13: MODÈLE[5]

D'après les deux modèles [6] et [5] nous constatons que la tendance n'est pas significativement différente de zéro car sa t-stat = 2.645888 est inférieure aux valeurs critiques tabulé de Dickey-Fuller Augmenté égale à 2.79 aux seuils 5%.

Et que la constante n'est pas significativement différente de zéro car sa t-stat est égale à 1.147481 est inférieure aux valeurs critiques tabulé qu'est égale à 2.54 aux seuils 5%. On passe au modèle [4].

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.395534	0.0178
Test critical values:		
1% level	-2.622585	
5% level	-1.949097	
10% level	-1.611824	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(INV,3)  
 Method: Least Squares  
 Date: 06/15/19 Time: 22:31  
 Sample (adjusted): 1978 2018  
 Included observations: 41 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INV(-1),2)	-2.954763	1.233446	-2.395534	0.0222
D(INV(-1),3)	1.364580	1.162487	1.173845	0.2486
D(INV(-2),3)	0.520616	1.044543	0.498415	0.6214
D(INV(-3),3)	-0.160489	0.877541	-0.182885	0.8560
D(INV(-4),3)	-0.726578	0.667543	-1.088437	0.2841
D(INV(-5),3)	-0.766632	0.431554	-1.776443	0.0846
D(INV(-6),3)	-0.308975	0.224830	-1.374264	0.1784

FIGURE 4.14: MODÈLE[4]

Le tableau ci-dessus, nous constatons que la t-stat est égale à -2.395534 est inférieur aux valeurs critiques tabulé de Dickey-Fuller augmenté aux seuils 1%, 5% et 10%. Alors la série est stationnaire avec deux différenciation.

## Identification :

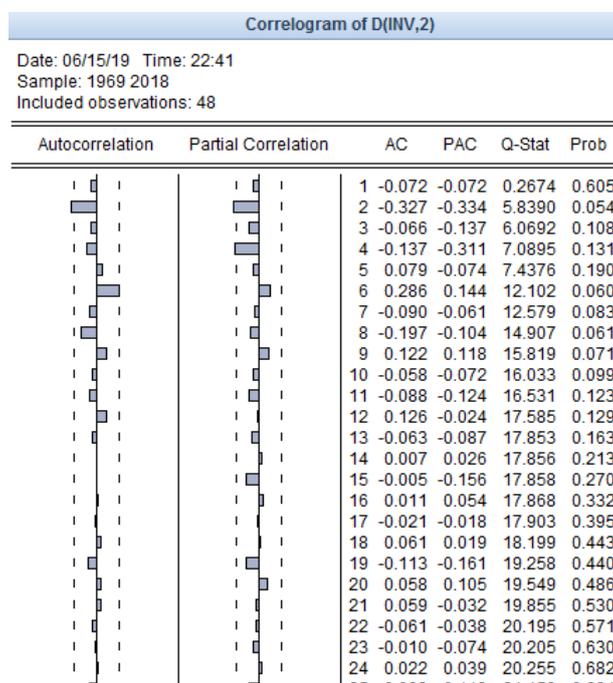


FIGURE 4.15: Corrélogramme de la série

D'après le corrélogramme on remarque que :

La partie autocorrelation simple indique que la 2ème et la 6ème pic qui sort à l'intervalle de confiance, donc cela nous a même à dire qu'il existe un moyenne mobile MA(2) et MA(6). Pour la partie autocorrélation partiel on a la 2ème et 4ème pics qui sort a l'intervalle de confiance donc cela nous a même à dire qu'il existe un autorégressive AR(2) et AR(4). Et on insérant la différenciation appliquée précédemment on arrivant a identifié les modèles suivants :

les modèles	ARIMA(2,2,4)	ARIMA(0,2,4)
$R^2$	0.393	0.189
<b>Akaike info criterion</b>	12.091	12.248
<b>Schwarz criterion</b>	12.247	12.326

D'après le tableau des critères de choix on choisir le modèle ARIMA(2,2,4). Mais lorsque on vérifié le modèle ARIMA(4,2,4), nous l'avons trouvé que le R-squared ( $R^2 = 0.54$ ) est plus grand que celui déduit précédemment.

**Estimation des paramètres :**

Le tableau suivant contient l'estimation du modèle qui a été choisi car il présente des critères de pouvoir prédictif meilleur que ceux des autres modèles estimés (à savoir : R-squared : maximum ; et Akaike Info Criterion, Schwarz Criterion : minimum).

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	-1.223211	0.135253	-9.043865	0.0000
AR(4)	-0.804760	0.133005	-6.050621	0.0000
MA(1)	-0.624351	0.138131	-4.519974	0.0001
MA(2)	0.781049	0.235225	3.320434	0.0019
MA(3)	-0.646124	0.277733	-2.326424	0.0250
MA(4)	0.421865	0.191201	2.206393	0.0330
SIGMASQ	6282.590	1315.587	4.775504	0.0000
R-squared	0.542122	Mean dependent var	-2.214203	
Adjusted R-squared	0.475116	S.D. dependent var	118.3767	
S.E. of regression	85.76262	Akaike info criterion	11.96456	
Sum squared resid	301564.3	Schwarz criterion	12.23745	
Log likelihood	-280.1495	Hannan-Quinn criter.	12.06769	
Durbin-Watson stat	1.870482			

FIGURE 4.16: MODÈLE ARMA(4,4)

**Test sur les paramètres :** On remarque que tous les paramètres du modèle sont significativement différents de zéro. En effet les statistiques de Student associées sont en valeur absolue supérieures à 1,96, ce qu'est confirmé par les probabilités de nullité des coefficients.

Test sur les résidus :

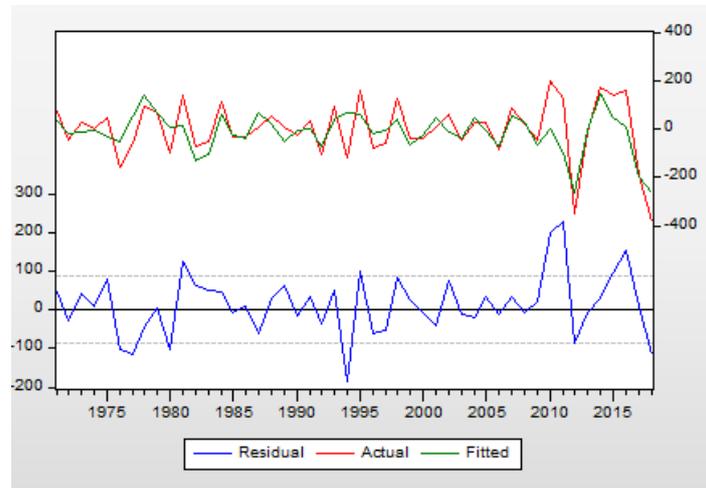


FIGURE 4.17: Actual,Fitted,Residual,graphe

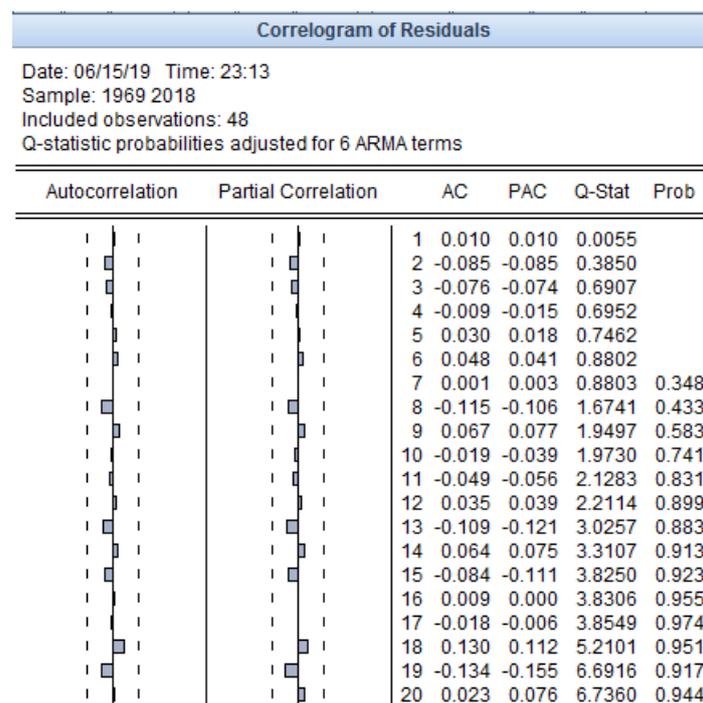


FIGURE 4.18: Corrélogramme des résidus

Après examen du corrélogramme des résidus, on remarque que tout les pics sont à l'intérieur de l'intervalle de confiance on estime que les résidus forme un bruit blanc et non corrélée

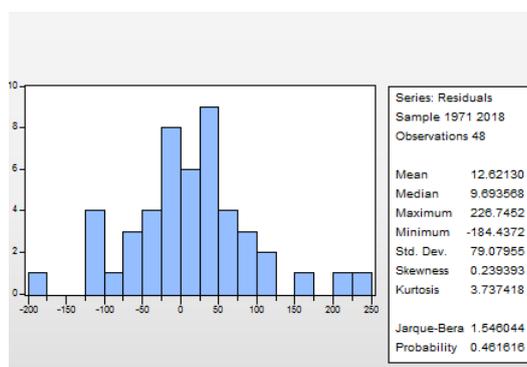


FIGURE 4.19: Histogram-Normality Test

### 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.

Teste de Skewness :

$$|V_1| = \left| \frac{\beta_1^{1/2} - 0}{\sqrt{6/n}} \right| = \left| \frac{0.239393 - 0}{\sqrt{6/50}} \right| = 0.69 < 1.96$$

Test de Kurtosis :

$$|V_2| = \left| \frac{\beta_2 - 3}{\sqrt{24/n}} \right| = \left| \frac{3.737418 - 3}{\sqrt{24/50}} \right| = 1.06 < 1.96$$

En outre **le test de Jarque Bera** :  $JB = 1.546044 < \chi_{0.05}^2(2) = 5.99$  Alors les résidus forment un bruit blanc gaussien

le modèle est validé et peut être représenté sous la forme suivante :

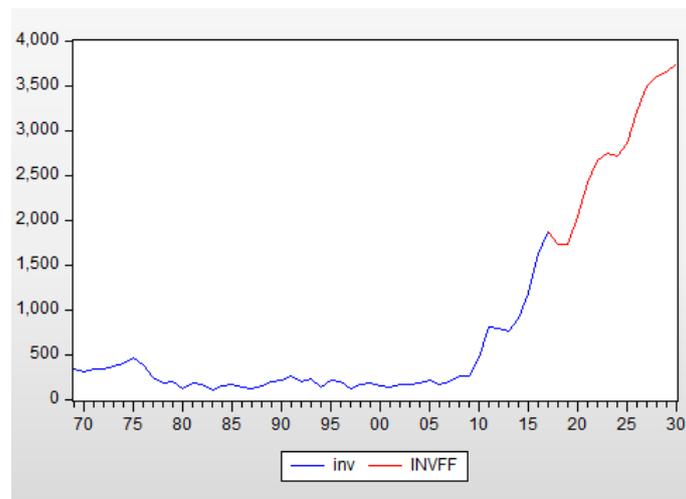
$$(1 + 1.22B^2 + 0.8B^4)(1 - B)^2 inv_t = (1 + 0.62B - 0.78B^2 + 0.64B^3 - 0.42B^4)\epsilon_t$$

### La prévision :

Le tableau suivant représente les valeurs prédites des 12 années depuis 2019 jusqu'au 2030.

La représentation graphique des prévisions (12 valeurs) obtenues par cette méthode sont en harmonie avec la même allure que la série étudiée.

2019	1728.717
2020	2043.687
2021	2416.957
2022	2671.587
2023	2739.645
2024	2704.039
2025	2849.725
2026	3217.696
2027	3487.055
2028	3594.938
2029	3644.522
2030	3745.766



## 4.2.2 Étude la série des réserves pétrolière récupérable

### Analyse préliminaire :

Le graphe suivant représente la représentation graphique de La série des réserves pétrolière de 1969 jusqu'au 2018, cette série contient 50 observations :

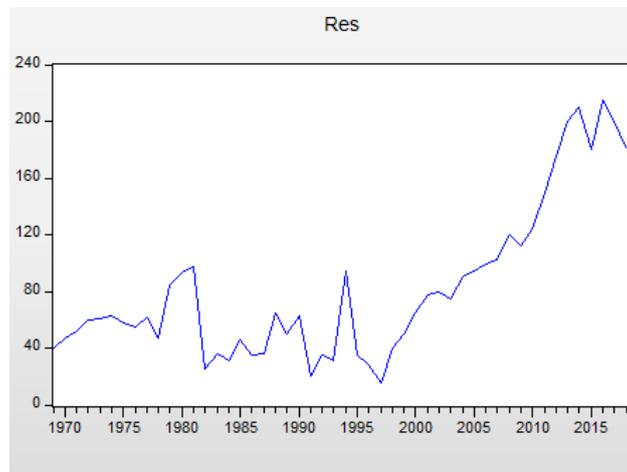


FIGURE 4.20: Graphe de la série res

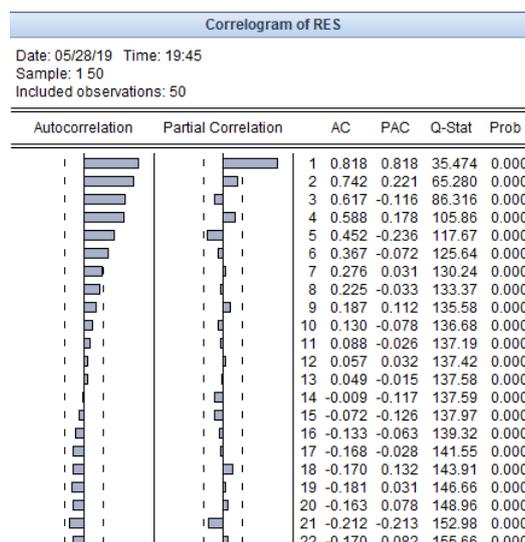


FIGURE 4.21: Corrélrogramme de la série res

La représentation graphique de la série des réserves et le corrélogramme montre une augmentation visible de série fluctue d'une allure croissante avec le temps, d'où la série peut être n'est pas stationnaire.

**Test Dickey-Fuller :** On applique le test de Dickey-Fuller pour infirmer ou confirmer notre hypothèse concernant la stationnarité.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES(-1)	-0.174954	0.084325	-2.074756	0.0436
C	1.954163	6.553841	0.298171	0.7669
@TREND("1969")	0.596330	0.312880	1.905939	0.0629

FIGURE 4.22: MODÈLE [3]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES(-1)	-0.062619	0.061972	-1.010431	0.3175
C	7.869831	5.931739	1.326733	0.1910

FIGURE 4.23: MODÈLE [2]

D'après la table du modèle [3] on remarque que la tendance n'est pas significativement différente de zéro. De même pour le modèle [2] la constante n'est pas significativement différente de zéro puisque sa t-stat est inférieure aux valeurs critiques.

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.752515	0.8734
Test critical values:		
1% level	-2.614029	
5% level	-1.947816	
10% level	-1.612492	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(RES)  
 Method: Least Squares  
 Date: 06/15/19 Time: 00:37  
 Sample (adjusted): 1971 2018  
 Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES(-1)	0.026505	0.035222	0.752515	0.4556
D(RES(-1))	-0.296364	0.149339	-1.984498	0.0532

FIGURE 4.24: MODÈLE [1]

La statistique ADF est égale à 0.752 est supérieur aux trois valeurs critiques pour les seuils 1%, 5% et 10%, on accepte donc l'hypothèse nulle d'existence une racine unitaire. Alors la série n'est pas stationnaire de type DS

pour cette raison on passe a la première différenciation :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(RES(-1))	-1.809758	0.306407	-5.906396	0.0000
D(RES(-1),2)	0.523090	0.245068	2.134472	0.0388
D(RES(-2),2)	0.413190	0.149882	2.756768	0.0087
C	-5.379050	7.116960	-0.755807	0.4541
@TREND("1969")	0.398829	0.250463	1.592367	0.1190

FIGURE 4.25: MODÈLE [6]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(RES(-1))	-1.653872	0.295607	-5.594830	0.0000
D(RES(-1),2)	0.415247	0.239791	1.731703	0.0907
D(RES(-2),2)	0.372799	0.150397	2.478774	0.0173
C	4.666077	3.354610	1.390945	0.1716

FIGURE 4.26: MODÈLE [5]

d'après les deux tableaux on constate que la tendance et la constante ne sont pas significativement différentes de zéro.

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.367737	0.0000
Test critical values:		
1% level	-2.616203	
5% level	-1.948140	
10% level	-1.612320	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(RES,2)  
 Method: Least Squares  
 Date: 06/15/19 Time: 00:44  
 Sample (adjusted): 1973 2018  
 Included observations: 46 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(RES(-1))	-1.532864	0.285570	-5.367737	0.0000
D(RES(-1),2)	0.329837	0.234301	1.407745	0.1664
D(RES(-2),2)	0.338085	0.149915	2.255183	0.0293

FIGURE 4.27: MODÈLE [4]

D'après le tableau du modèle [4] on accepte donc l'hypothèse nulle d'existence une racine unitaire, alors la série est stationnaire avec une différenciation.

Identification :

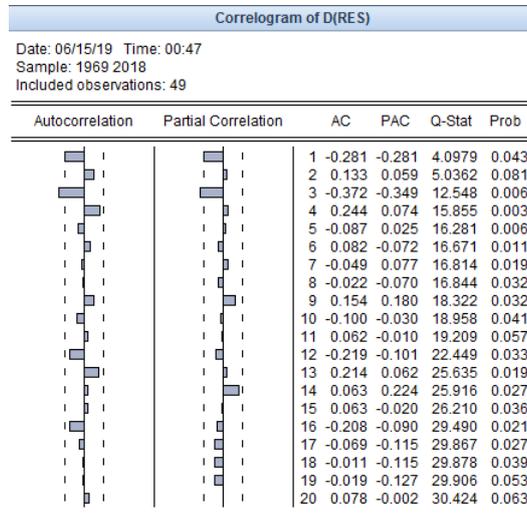


FIGURE 4.28: Correlogramme de la série d(res)

D'après le correlogramme simple et partielle on a identifié les modèles suivants :

les modèles	ARIMA(1,1,3)	ARIMA(0,1,3)
$R^2$	0.196	0.106
<b>Akaike info criterion</b>	9.028	9.182
<b>Schwarz criterion</b>	9.043	9.12

D'après le tableau des critères de choix on a choisir le medéle ARIMA(1,1,3).

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.719485	0.193531	-3.717673	0.0006
MA(1)	0.656582	0.209203	3.138489	0.0030
MA(3)	-0.337040	0.124083	-2.716242	0.0093
SIGMASQ	407.8908	94.93403	4.296572	0.0001
R-squared	0.196754	Mean dependent var		2.857143
Adjusted R-squared	0.143204	S.D. dependent var		22.76801
S.E. of regression	21.07481	Akaike info criterion		9.028478
Sum squared resid	19986.65	Schwarz criterion		9.182912
Log likelihood	-217.1977	Hannan-Quinn criter.		9.087070
Durbin-Watson stat	2.023877			

FIGURE 4.29: Modèle ARMA(1,3)

**Test sur les paramètres :**

D'après le tableaux précédent on remarque que tous les coefficients des paramétrés sont significativement différent de zéro (non null) car : toutes les p-value sont inférieurs à 0.05.

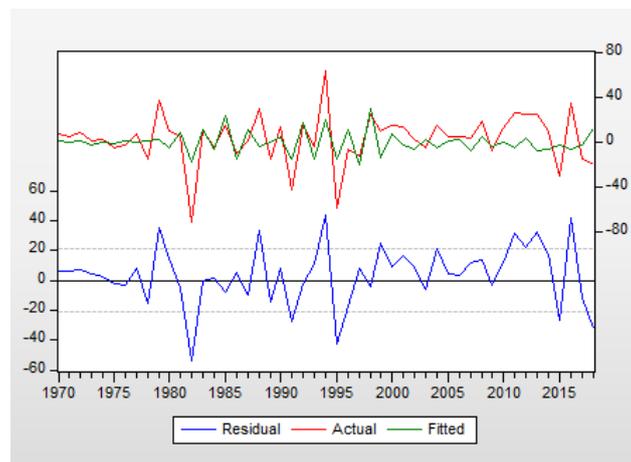


FIGURE 4.30: Actual,Fitted,Residual, graphe

**Test sur les résidus :**

A partir de la représentation graphique des séries résiduelles réelles et estimées nous constatons que le modèle estimé n'ajuste pas la série d'où le coefficient  $R^2$  de cette estimation est proche de 0 que de 1, (égale à 0.196754).

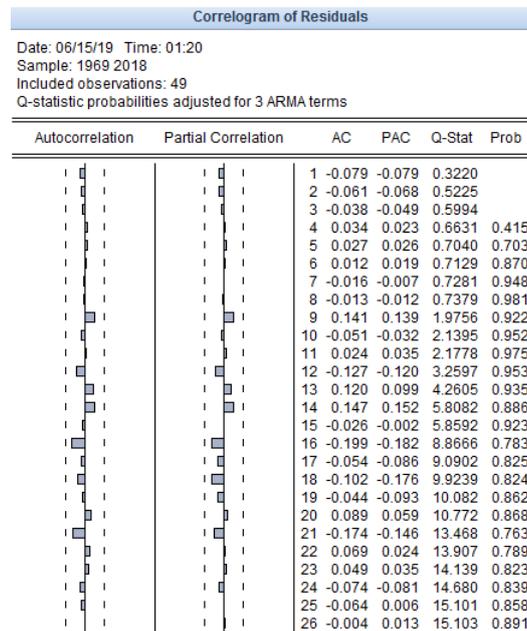


FIGURE 4.31: Le correlogramme des résidus

D'après le correlogramme des résidus, on remarque que tout les pics sont à l'intérieur de l'intervalle de confiance on estime que les résidus forme un bruit blanc et non corrélée

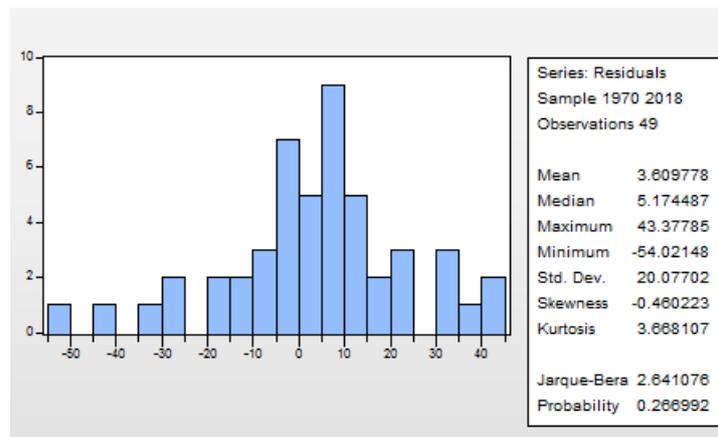


FIGURE 4.32: Histogram-Normality Test

Test de normalité des résidus :

**Teste de Skewness :**

$$|V_1| = \left| \frac{\beta_1^{1/2} - 0}{\sqrt{6/n}} \right| = \left| \frac{-0.460223 - 0}{\sqrt{6/50}} \right| = -1.328 < 1.96$$

**Test de Kurtosis :**

$$|V_2| = \left| \frac{\beta_2 - 3}{\sqrt{24/n}} \right| = \left| \frac{3.668107 - 3}{\sqrt{24/50}} \right| = 0.964 < 1.96$$

**Test de jack-Berra :**  $JB = 2.46 < \chi_{0.05}^2 = 5.99$

Donc les résidus forment un bruit blanc gaussien.

Le modèle est validé et peut être représenté sous la forme suivante :

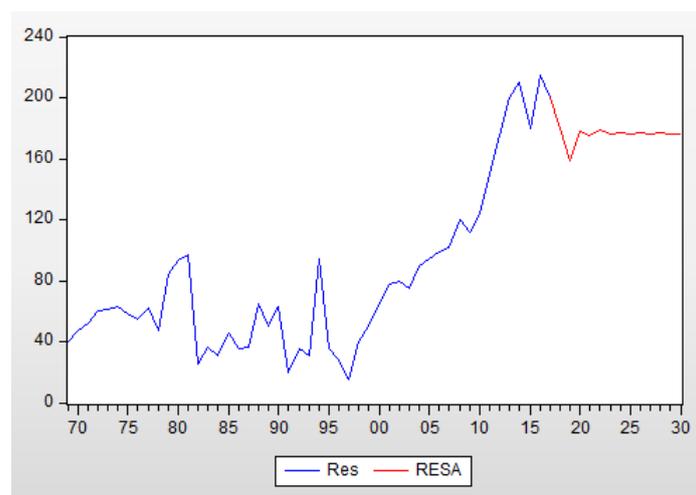
$$(1 + 0.71B)(1 - B)res_t = (1 - 0.65B + 0.33B^3)\epsilon_t$$

**La prévision :**

Le tableau suivant représente les valeurs prédites des 12 années depuis 2019 jusqu'à 2030

2019	159,138
2020	178,030
2021	175,291
2022	179,261
2023	175,844
2024	176,864
2025	176,130
2026	176,658
2027	176,278
2028	176,551
2029	176,355
2030	176,496

La représentation graphique des prévisions (12 valeurs) obtenues par cette méthode ne suit pas vraiment la même allure que la série étudiée puisque le  $R^2$  est proche de 0, ce qui implique que cette méthode donne un mauvais ajustement pour cette série.



### 4.2.3 Etudie la série du prix moyenne de pétrole

Analyse préliminaire :

La série représente les prix moyenne du pétrole annuel de 1969 jusqu'au 2018, elle contient n=50 observations.

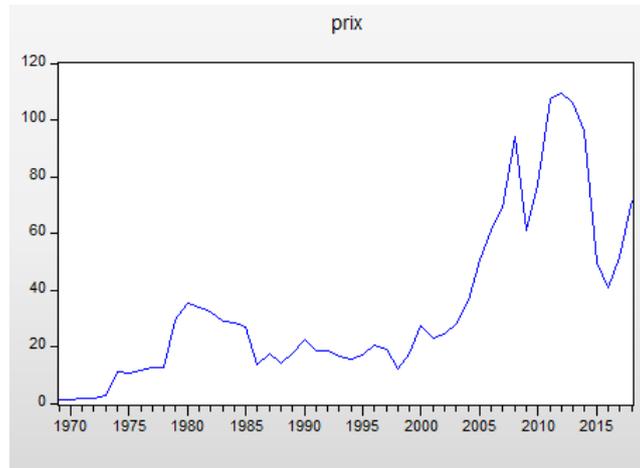


FIGURE 4.33: Graphe de la série de prix

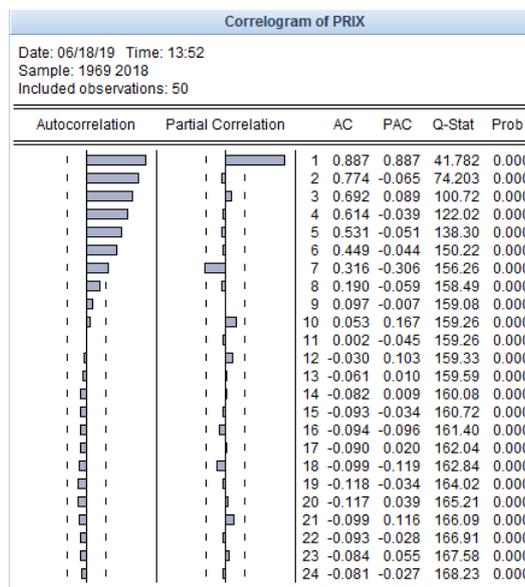


FIGURE 4.34: Corrélogramme de la série de prix

La représentation graphique et le corrélogramme de cette série montrent une augmentation visible de série fluctue d'une allure croissante avec le temps d'où la série peut être n'est pas stationnaire

### Test Dickey-Fuller :

On applique le test de Dickey-Fuller pour infirmer ou confirmer notre hypothèse concernant la stationnarité.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRIX(-1)	-0.194617	0.087223	-2.231248	0.0306
C	0.140329	3.439299	0.040802	0.9676
@TREND("1969")	0.311791	0.179221	1.739701	0.0886

FIGURE 4.35: MODÈLE [3]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRIX(-1)	-0.080952	0.059018	-1.371652	0.1767
C	4.141568	2.611690	1.585781	0.1195

FIGURE 4.36: MODÈLE [2]

La tendance n'est pas significativement différente de zéro dans le modèle [3], et la constante n'est pas significativement différente de zéro dans le modèle [2]. On passe au modèle [1].

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	* -0.263394	0.5861
Test critical values:		
1% level	-2.613010	
5% level	-1.947665	
10% level	-1.612573	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(PRIX)  
 Method: Least Squares  
 Date: 06/18/19 Time: 13:58  
 Sample (adjusted): 1970 2018  
 Included observations: 49 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRIX(-1)	-0.010367	0.039361	-0.263394	0.7934

FIGURE 4.37: MODÈLE [1]

D'après le tableau ci-dessus, nous constatons que la t-stat est égale à -0.263394 est supérieur aux valeurs critiques tabulé par Dickey-Fuller simple aux seuils 1%,5%,10%. Donc on accepte l'hypothèse nulle de existence de racine unitaire alors la série n'est pas stationnaire de type DS.

On passe par la première différenciation :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PRIX(-1))	-0.891409	0.152171	-5.857938	0.0000
C	1.025609	3.773039	0.271826	0.7870
@TREND("1969")	0.012981	0.129718	0.100068	0.9207

FIGURE 4.38: MODÈLE [6]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PRIX(-1))	-0.892056	0.150389	-5.931677	0.0000
C	1.357296	1.783035	0.761228	0.4504

FIGURE 4.39: MODÈLE [5]

La tendance n'est pas significativement différente de zéro puisque sa t-stat est égale a 0.10 est inférieure aux valeurs critiques tubulées par Dickey-Fuller Augmenté .

La constante n'est pas significativement différente de zéro puisque sa t-stat égale à 0.76 est inférieure aux valeurs critiques tabulées par Dickey-Fuller Augmenté On passe au modèle [4].

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.913836	0.0000
Test critical values:		
1% level	-2.614029	
5% level	-1.947816	
10% level	-1.612492	

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

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(PRIX,2)  
 Method: Least Squares  
 Date: 06/18/19 Time: 14:02  
 Sample (adjusted): 1971 2018  
 Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PRIX(-1))	-0.881881	0.149122	-5.913836	0.0000

FIGURE 4.40: MODÈLE[4]

D'après le tableau ci-dessus, nous constatons la t statistique est égale a -5.913836 est inférieure aux valeurs critiques tabulée par Dickey-Fuller aux seuils 1%,5%,10%. Donc la série est stationnaire avec une différenciation.

IDENTIFICATION :

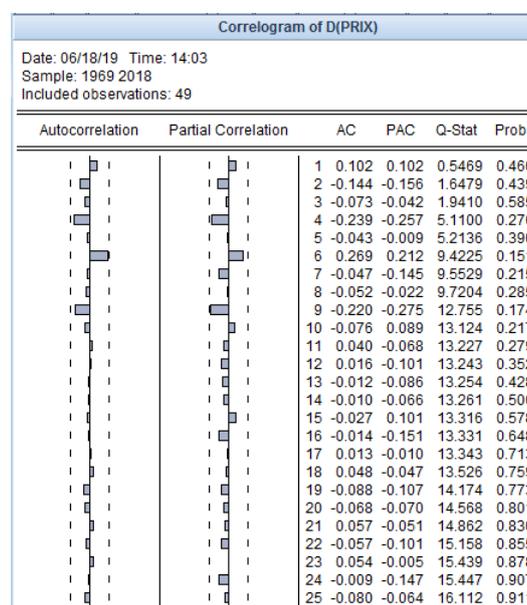


FIGURE 4.41: Correlogramme de la série du prix

D'après le correlogramme on remarque que dans :

La partie simple de corrélogramme nous indique que la 6ème pic sort a l'intervalle de confiance donc cela nous a même a dire qu'il existe un Moyenne Mobile MA(6).

La partie partiel du corrélogramme nous indique que le 9ème pic sort a l'intervalle de confiance donc cela nous a même a dire qu'il existe un Autorégressive AR(9). et on tout compte de la différenciation applique précédant on a identifié les modèle suivante :

les modèles	ARIMA(9,1,6)	ARIMA(0,1,6)
$R^2$	0.26	0.13
<b>Akaike info criterion</b>	7.68	7.77
<b>Schwarz criterion</b>	7.79	7.84

D'après le tableau des critères de choix on a choisir le modèle ARIMA(9,1,6), mais il y a un autre modèle qui voire un R-squard ( $R=0.33$ ) supérieur à cette modèle.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(6)	0.509335	0.129908	3.920743	0.0003
MA(4)	-0.279388	0.121267	-2.303902	0.0259
MA(9)	-0.507783	0.214985	-2.361943	0.0226
SIGMASQ	95.14324	18.13376	5.246745	0.0000
R-squared	0.338209	Mean dependent var		1.439796
Adjusted R-squared	0.294090	S.D. dependent var		12.11451
S.E. of regression	10.17843	Akaike info criterion		7.660144
Sum squared resid	4662.019	Schwarz criterion		7.814579
Log likelihood	-183.6735	Hannan-Quinn criter.		7.718736
Durbin-Watson stat	1.779145			

FIGURE 4.42: ARMA(6,9)

D'après le tableau président tout les coefficients des paramétrés sont significativement différent de zéro (non null )car : tout les t-stat en valeur absolue sont supérieurs à 1,96.

Validation :

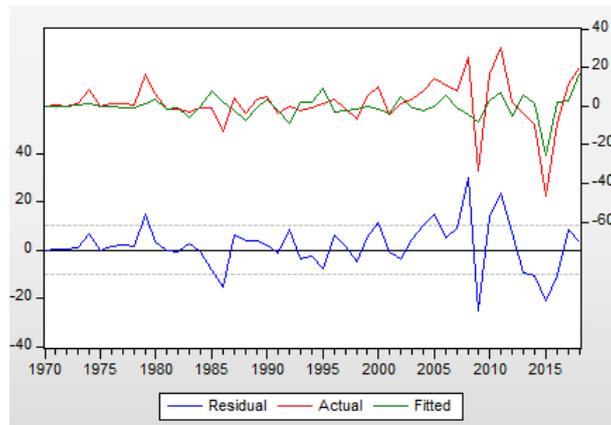


FIGURE 4.43: Actual,fitted,Residual,graphe

Correlogram of Residuals						
Date: 06/18/19 Time: 15:02						
Sample: 1969 2018						
Included observations: 49						
Q-statistic probabilities adjusted for 3 ARMA terms						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.083	0.083	0.3570	
		2	-0.101	-0.108	0.8962	
		3	0.052	0.072	1.0448	
		4	-0.011	-0.034	1.0512	0.305
		5	-0.064	-0.048	1.2843	0.526
		6	-0.016	-0.014	1.2996	0.729
		7	-0.032	-0.040	1.3618	0.851
		8	-0.061	-0.052	1.5879	0.903
		9	-0.092	-0.093	2.1199	0.908
		10	-0.171	-0.173	4.0044	0.779
		11	0.000	0.012	4.0044	0.857
		12	0.029	-0.010	4.0626	0.907
		13	-0.048	-0.046	4.2247	0.937
		14	-0.034	-0.049	4.3055	0.960
		15	-0.069	-0.111	4.6521	0.969
		16	0.005	-0.002	4.6538	0.982
		17	-0.051	-0.103	4.8580	0.988
		18	0.028	0.008	4.9193	0.993
		19	-0.051	-0.127	5.1370	0.995
		20	0.029	0.003	5.2080	0.997
		21	0.049	0.004	5.4195	0.998
		22	-0.121	-0.162	6.7784	0.995
		23	0.022	0.002	6.8232	0.997
		24	0.006	-0.103	6.8263	0.998
		25	-0.047	-0.084	7.0540	0.999
		26	-0.079	-0.136	7.7371	0.999

FIGURE 4.44: Le corrélogramme des résidus

test sur les résidus :

D'après le corrélogramme des résidus, on remarque que tout les pics sont à l'intérieur de l'intervalle de confiance on estime que les résidus forme un bruit blanc et son corrélée

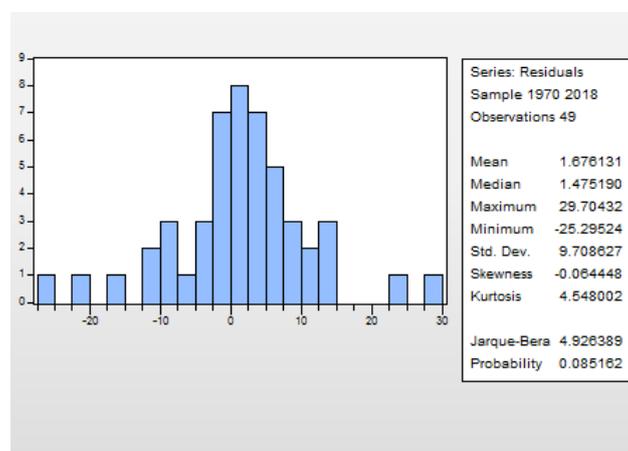


FIGURE 4.45: Histogram-Normality Test

**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.

Teste de Skewness :

$$|V_1| = \left| \frac{\beta_1^{1/2} - 0}{\sqrt{6/n}} \right| = \left| \frac{-0.064448 - 0}{\sqrt{6/50}} \right| = -0.186045 < 1.96$$

test de Kurtosis :

$$|V_2| = \left| \frac{\beta_2 - 3}{\sqrt{24/n}} \right| = \left| \frac{4.548002 - 3}{\sqrt{24/50}} \right| = 2.234349 > 1.96$$

**Test de jack-Berra :**  $JB = 4.92 < \chi_{0.05}^2 = 5.99$  Alors les résidus sous forme bruit blanc gaussien le modèle est validé et peut être représenté sous la forme suivante :

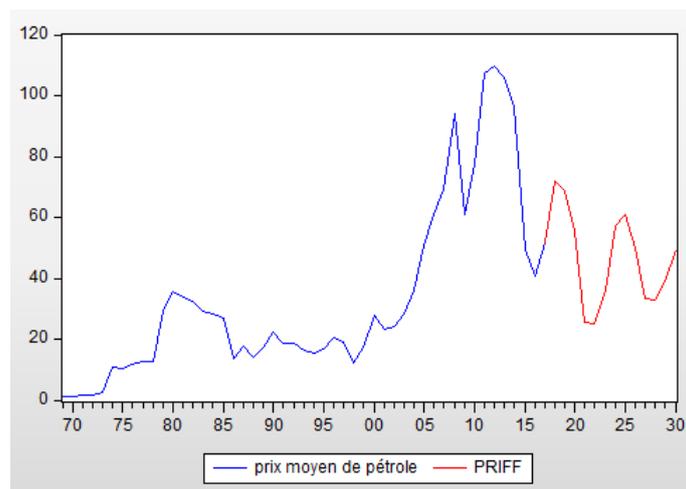
$$(1 - 0.5B^6)(1 - B)prix_t = (1 + 0.27B^4 + 0.5B^9)\epsilon_t$$

**La prévision :**

Le tableau suivant représente les valeurs prédites des (12 valeur ) depuis 2019 jusque a 5 juin 2019 :

2019	68,944
2020	55,278
2021	25,527
2022	25
2023	36,262
2024	57,056
2025	61,051
2026	49,91
2027	33,177
2028	32,91
2029	38,645
2030	49,236

La représentation graphique des prévisions (12 valeurs) obtenues par cette méthode sont pratiquement en harmonie avec la même allure que la série étudiée.



### 4.3 Application ARFIMA

Dans ce paragraphe, nous présentons une application des processus à mémoire longue sur les données réelles.

Notre application est orientée, principalement, dans une optique prévisionnelle. Cette application est réalisée sur les trois séries des données, donnée d'investissement des projets, réserve pétrolière récupérable, et prix moyenne de pétrole.

#### 1. Application sur la série d'investissement des projets :

Nous allons modéliser cette série en utilisant de manière compétitive deux approches différentes. La première approche est basée sur une modélisation linéaire à mémoire courte de type ARIMA, et la seconde approche est basée sur une modélisation à mémoire longue de type ARFIMA. De plus, pour chacune de ces deux approches, nous étudierons l'impact en prévision des différents modèles estimés.

#### Première approche :

Dans cette première approche, nous utilisons la méthode classique de modélisation et de prévision de Box et Jenkins (1970), qu'est déjà faite.

Après le passage par les étapes de la méthodologie de Box et Jenkins, à savoir l'identification, l'estimation, la validation et la prévision, on retient le modèle ARIMA(4,2,4) pour la série  $\Delta^2(Inv_t)_t$  avec le résultat relatif à la qualité des prévisions obtenues à partir de ce modèles.

#### Deuxième approche :

On s'intéresse maintenant à une approche longue mémoire de cette série , à l'aide d'un processus ARFIMA. pour ce la on utilisant le langage statistique R.

Dans un premier temps, on suppose que les ordres p et q des polynômes auto-régressif et moyenne mobile sont nuls. On estime alors le paramètre de mémoire longue par la méthode GPH présentée dans le chapitre 3, et nous prenons en considération la valeur de  $\mu$ , (ici comme la taille de l'échantillon est très petite  $n=50$ , nous prenons  $\mu = 0.4$ ).

Sur R et avec l'instruction "  $fdGPH(Y, bandwidth.exp = 0.4)$  ", avec ( The bandwidth is  $bw = trunc(n^{bandwidth.exp})$ , where  $0 < bandwidth.exp < 1$  and n is the sample size. Default  $bandwidth.exp = \mu = 0.4$ ).

On obtient  $\hat{d} = 0.71$  et le processus ARFIMA(0,d,0) suivant :

$$(1 - B)^{0.71} inv_t = \varepsilon_t$$

Dans une seconde étape, on cherche à spécifier correctement les ordres des parties, autorégressive et moyenne mobile de ces deux processus. Pour cela, on effectue une recherche de  $p$  et  $q$  (étape identification de la méthode Box et Jenkins), puis on passe à l'étape d'estimation des paramètres. On retient un processus de type ARMA(0,4) de la série qui s'écrit de la forme suivante :

$$(1 - B)^{0.71} Inv_t = (1 + 1.33B + 1.17B^2 + 0.86B^3 + 0.31B^4)\varepsilon_t$$

On remarque d'après l'estimation que le paramètre de mémoire longue  $\hat{d} = 0.71$  estimé est supérieur à 0.5, or l'intervalle du paramètre d'intégration fractionnaire d'un processus ARFIMA à mémoire longue qu'est  $[-1/2, 1/2[$ .

Dans ce cas-là, le processus considéré est alors non stationnaire. Donc on peut différencier la série, de manière à ce que le paramètre de mémoire longue soit dans  $[-0.5, 0.5[$ . Cette méthode considère uniquement le problème de l'estimation des paramètres, et elle ne se place pas d'un point de vue prévisionnel.

Laurent FERRARA dans sa thèse de doctorat, a montré de manière empirique que le modèle qui correspond au processus ARFIMA non stationnaire, donne des prévisions qui convergent lentement vers la moyenne non conditionnelle du processus. Par contre les prévisions issues des modèles ARFIMA des séries différencier ne convergent pas. Donc, garder les données brutes, même en cas de non stationnarité, constitue une meilleure approche d'un point de vue prévisionnel.

Donc on conclue que le processus qui concerne la mémoire longue est : ARFIMA(0,0.76,4). On continue, alors, la procédure de prévision du modèle ARFIMA dans ce cas et voir sa qualité prédictive.

#### **Prévision :**

Après la validation du modèle ARFIMA on va prévoir la série d'investissement de 2019 jusqu'au 2030, les valeurs prédictive et le graphe de prévision sont :

2019	1441,01
2020	1163,29
2021	981,48
2022	909,343
2023	870,647
2024	832,021
2025	799,794
2026	772,796
2027	749,868
2028	730,136
2029	712,955
2030	697,841

FIGURE 4.46: les valeurs pridite de l'investissement des projts

### Time Series and Predictions of fitInv

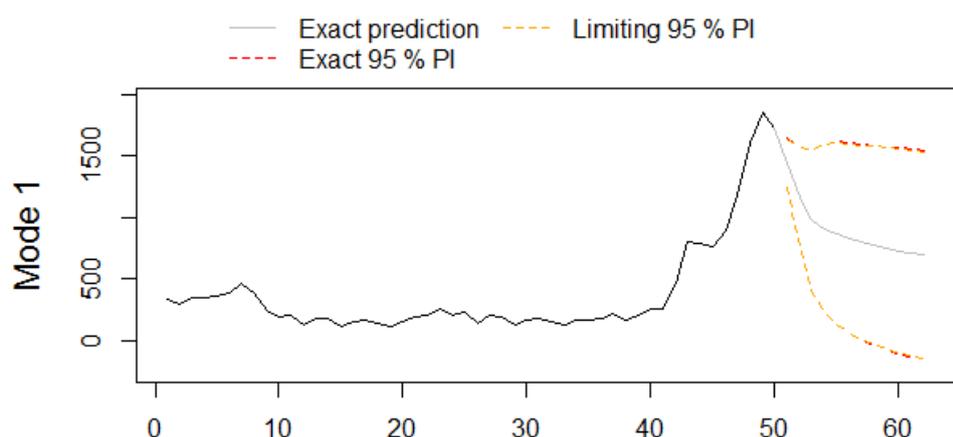


FIGURE 4.47: graphe de prévision de l'investissement des projets

### 2. La série réserve pétrolière récupérable et les prix annuel du pétrole :

On a preuve que les deux séries sont non stationnaire, et de même taille  $n = 50$ , pour cela on peut applique la même structure d'étude, et le tableau suivant résume les résultats de l'approche ARFIMA structure :

séries	ARIMA(p,d,q)	$\mu$	$\hat{d}$	ARFIMA(0, $\hat{d}$ , 0)	ARFIMA(p, $\hat{d}$ , q)
réserve	ARIMA(1,0,3)	0.4	0.66	$(1 - B)^{0.66} X_t = \varepsilon_t$	ARFIMA(1,0.66,0)
prix	ARIMA(6,1,9)	0.4	0.76	$(1 - B)^{0.76} X_t = \varepsilon_t$	ARFIMA(1,0.76,0)

**Pour la série réserve**, On remarque d'après l'estimation que le paramètre de mémoire

longue estimé  $\hat{d} = 0.66$  est supérieur à 0.5, or l'intervalle du paramètre d'intégration fractionnaire d'un processus ARFIMA à mémoire longue est  $[-1/2, 1/2[$ .

On retient un processus de type ARMA(1,0) de la série qui s'écrit de la forme suivante :

$$(1 - 0.97B)(1 - B)^{0.66}Res_t = \varepsilon_t$$

Donc on conclue que le processus qui concerner les mémoire longue est : ARFIMA(1,0.66,0).

Après la validation du modèle ARFIMA on va faire prévoir future des réserves pétrolière récupérable 2019 jusqu'au 2030, les valeur prédictive et le graphe de prévision sont :

2019	174,6798
2020	169,1824
2021	164,2327
2022	159,7987
2023	155,8031
2024	152,1779
2025	148,8683
2026	145,8303
2027	143,0289
2028	140,4353
2029	138,0258
2030	135,7808

FIGURE 4.48: les valeurs pridite des réserves pétrolière récupérable

## Time Series and Predictions of fitResrv

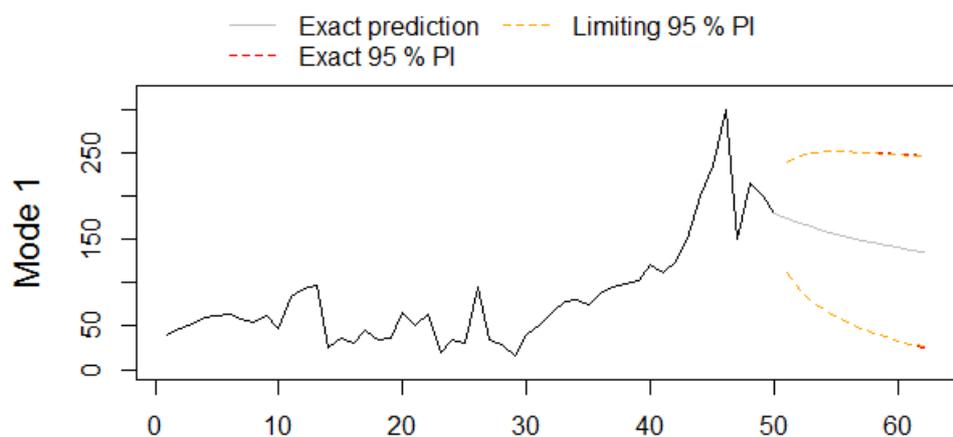


FIGURE 4.49: graphe de prévision des réserves pétrolière récupérable

**Pour la série du prix :** On remarque aussi que le paramètre de mémoire longue  $\hat{d} = 0.76$  estimé est supérieur à 0.5, et on retient un processus de type ARMA(1,0) de la série qui s'écrit de la forme suivante :

$$(1 - 0.99B)(1 - B)^{0.76} \text{Prix}_t = \varepsilon_t$$

Donc on conclure que le processus qui concerner les mémoire longue est : ARFIMA(1,0.76,0).

Après la validation du modèle ARFIMA on va faire prévoir future des prix de pétrole 2019 jusqu'au 2030, les valeur prédictive et le graphe de prévision sont :

2019	70,0438
2020	69,5657
2021	69,2457
2022	69,0023
2023	68,8051
2024	68,6392
2025	68,496
2026	68,3701
2027	68,2578
2028	68,1565
2029	68,0643
2030	67,9798

FIGURE 4.50: les valeurs pridite dse prix moyenne du pétrole

## Time Series and Predictions of fitPrix

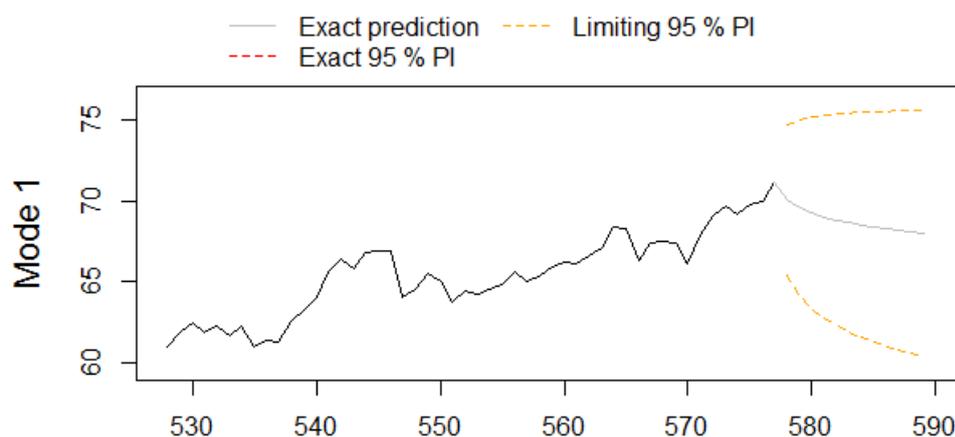


FIGURE 4.51: graphe de prévision des prix moyenne du pétrole

### Comparaison et interprétation des résultats

Après avoir appliqué les deux méthodes de prévision sur les trois séries réelles, les résultats obtenu sont jugés acceptables. On doit faire une comparaison entre les capacités prédictives à l'aide du critère de la racine carré de l'erreur quadratique moyenne de prévision (RMSE) qui on calculé précédemment.

Le tableau suivant représente les RMSE obtenu :

méthodes	box-jenkins	ARFIMA
<b>RMSE(Inv)</b>	124.2886	732.4145
<b>RMSE(Rés)</b>	22.9631	68.6912
<b>RMSE(Prix)</b>	12.8491	24.9406

-  $RMSE(B - J)$  et  $RMSE(ARF)$  sont associée à la méthode de box-Jenkins et le modèle ARFIMA respectivement.

Le tableau de comparaison indique que :

- Pour l'investissement le  $RMSE(B-J) < RMSE(ARF)$  ;
- Pour les réserves le  $RMSE(B-J) < RMSE(ARF)$  ;
- Pour les prix le  $RMSE(B-J) < RMSE(ARF)$ .

Nous concluons que les prévisions obtenues par la méthodologie de Box-Jenkins sont plus faibles que celles obtenues par le modèle ARFIMA pour les trois séries étudiée.

## 4.4 prévision par la régression linéaire :

### 4.4.1 La régression linéaire multiple

Pour applique la régression linéaire multiple sur nos séries on utilise le logiciel Eviews 9.

Le point de départ est l'estimation des paramétrés d'une régression de variable  $Y_t$  (des investissements) en fonction des réserves ( $X_{1,t}$ ) et les prix  $X_{2,t}$ , nous disposons de 50 observation.

L'équation de régression s'écrit sous la forme suivante :

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \epsilon_t \quad (4.1)$$

#### Test de significativité :

Lorsque on traité des données on obtenu le résultat suivant :

- La régression semble une bonne qualité car le  $R^2 = 0.71$  est proche de 1.
- La P-value est inférieur à 0.05 de même la t-statistique de fisher = 57.53 supérieur au valeur tubulée de  $f(2, 47, 0.05)$  alors le modèle est significatif. Donc on peut écrit le modèle sous la forme suivante :

$$Y_t = -122.56 + 8.51X_{1,t} - 5.80X_{2,t} + \epsilon_t \quad (4.2)$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-122.5642	58.48522	-2.095644	0.0415
RES	8.512974	0.939976	9.056583	0.0000
PRIX	-5.801324	1.727904	-3.357433	0.0016
R-squared	0.710002	Mean dependent var	377.8461	
Adjusted R-squared	0.697661	S.D. dependent var	412.3704	
S.E. of regression	226.7432	Akaike info criterion	13.74364	
Sum squared resid	2416386.	Schwarz criterion	13.85836	
Log likelihood	-340.5909	Hannan-Quinn criter.	13.78732	
F-statistic	57.53499	Durbin-Watson stat	0.913775	
Prob(F-statistic)	0.000000			

Test sur les résidu :

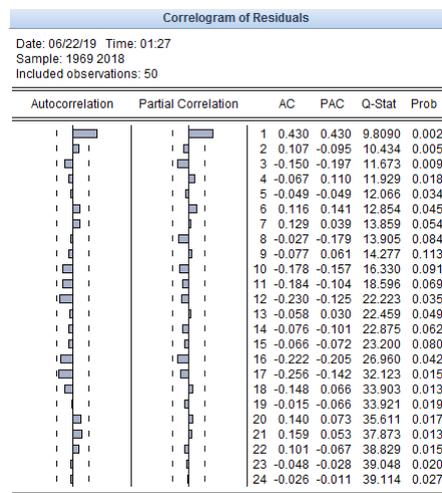


FIGURE 4.52: correlograme des résidus

D'après le correlograme des résidus on remarque que le 1<sup>er</sup> pic dans la partie d'autocorrélation simple et partial sort de l'intervalle de confiance et toutes les proba sont inférieure à 0.05 . Donc les résidus ne forme pas un bruit blanc.

## 4.5 Application la régression linéaire simple :

- On applique la régression linéaire simple de variable  $Y_t$  (l'investissement) en fonction de  $X_t$  (le réserve), nous disposons de 50 observations. L'équation de la régression simple s'écrit sous la forme suivant :

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad (4.3)$$

**Test de significativité d'une coefficient  $\beta_k$  :**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-120.4207	64.43635	-1.868832	0.0678
RES	6.072722	0.656750	9.246620	0.0000
R-squared	0.640449	Mean dependent var		377.8461
Adjusted R-squared	0.632959	S.D. dependent var		412.3704
S.E. of regression	249.8302	Akaike info criterion		13.91862
Sum squared resid	2995925.	Schwarz criterion		13.99510
Log likelihood	-345.9654	Hannan-Quinn criter.		13.94774
F-statistic	85.49999	Durbin-Watson stat		0.486416
Prob(F-statistic)	0.000000			

Lorsque en traité les données on obtient le résultat suivant :

- La régression semble une bonne qualité car le  $R^2 = 0.64$  est proche de 1.
  - La P-value est inférieur à 0.05 de même la t- statistique de fisher = 85.49 supérieur au valeur tabulé de  $f(1, 48, 0.05)$  alors le modèle est significatif .
- Donc en peut écrire le modèle sous la forme suivante :

$$Y_t = 6.07X_{1,t} + \epsilon_t \quad (4.4)$$

Test sur les résidu :

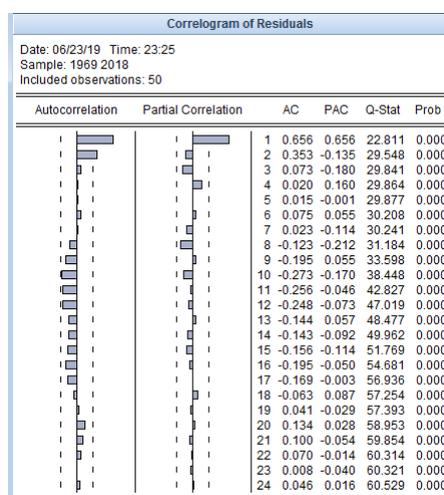


FIGURE 4.53: correlograme des résidus

D'après le correlograme des résidus on remarque que le 1<sup>er</sup> pic dans la partie d'autocorrélation simple et partial sort de l'intervalle de confiance et toutes les proba sont inférieure à 0.05 . Donc les résidus ne forme pas un bruit blanc.

- Pour cela on passe a appliqué la régression linéaire simple de variable Y (l'investissement) en fonction de  $X_t$  (le prix), nous disposons de 50 observation. L'équation de régression s'écrit sous la forme suivante :

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad (4.5)$$

Test de significativité :

On obtenu le résultat suivant :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	162.7766	80.78355	2.014972	0.0495
PRIX	6.298932	1.796401	3.506419	0.0010
R-squared	0.203914	Mean dependent var		377.8461
Adjusted R-squared	0.187329	S.D. dependent var		412.3704
S.E. of regression	371.7448	Akaike info criterion		14.71347
Sum squared resid	6633323.	Schwarz criterion		14.78995
Log likelihood	-365.8368	Hannan-Quinn criter.		14.74259
F-statistic	12.29497	Durbin-Watson stat		0.143319
Prob(F-statistic)	0.000996			

- La régression semble d'un modèle faible car le  $R^2 = 0.20$  est proche de 0.
- La P-value est inférieure à 0.05 de même la t- statistique de fisher =12.29 supérieur au

valeur tabulée de  $f(1, 48, 0.05)$  alors le modèle est significatif .  
 Donc on peut écrire le modèle sous la forme suivant :

$$Y_t = 162.77 + 6.29X_{1,t} + \epsilon_t \tag{4.6}$$

Test sur les résidu :

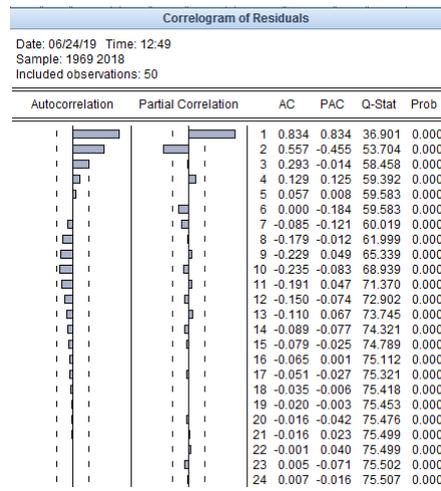
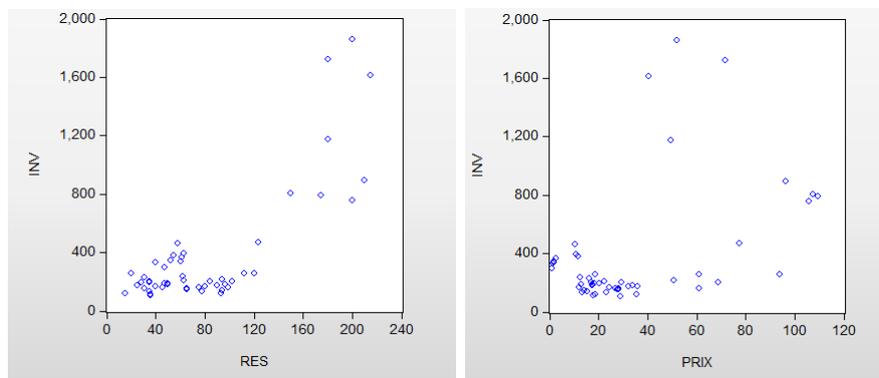


FIGURE 4.54: correlograme des résidus

D'après le correlograme des résidus on remarque que le 1<sup>er</sup> pic dans la partie d'autocorrélation simple et partial sort de l'intervalle de confiance et toutes les proba sont inférieure à 0.05 . Donc les résidus ne forme pas un bruit blanc.

Conclusion :



D'après la forme des deux nuage de point on peut constatons que la série d'investissement est inexplicquée à partir de celles des prix et des réserves.

Et comme le  $R^2$  est proche de 1 pour la régression des investissement en fonction des réserves mais les résidus ne forme pas un bruit blanc pour cela en peut appliqué la modélisation ARMAX.

## 4.6 Conclusion

Dans ce chapitre, on a exposer deux méthodes de prévision, tel que la méthode de Box et Jenkins, la prévision par le modèle ARFIMA.

Ces méthodes présentent des résultats satisfaisants et acceptable.

## *Conclusion Générale*

Dans ce travail nous avons examiné la notion de la prévision et son importance dans les ressources pétrolière et le prix moyenne de pétrole ainsi les investissements qui à engagé.

Prévoir, c'est observer un ensemble de données qui permet d'envisager une situation future et d'entreprendre des actions pour y parer concrètement autrement dit c'est porter un jugement sur les évènements ou évolutions possibles à venir en utilisant comme outils le passé et le présent.

Il existe plusieurs méthodes de prévision permet les quelles on a choisit deux pour la prévision extrapolative et une pour la prévision explicative.

Pour le cas extrapolative on a appliqué la méthodologie de Box-Jenkins qui offre des bonnes prévisions, comme partialement à celle de ARFIMA est sa cause de la taille des séries qu'est on a trité est petite  $n = 50$  obs, ainsi l'horizon de prévision.

Pour le cas explicative on a choisit la méthode de la régression linéaire simple et multiple qui donne des mauvais résultats car les résidus ne forme pas un bruit blanc malgré que le coefficient de détermination  $R^2$  est proche de 1, pour cette résultats on a propose d'applique la modélisation ARMAX.

## Bibliographie :

- [1 ] Charpentier, A. (2004). Cours des séries temporelles. Théorie et Applications. Université de Paris.
- [2 ] Corinne, P. (2005). Séries chronologiques, Quelques éléments du cours. Paris.
- [3 ] Dégerine, S. (2005). Cours de série chronologique. Université Joseph Fourier.
- [4 ] Florenee, N. (2006). Généralités sur les séries chronologiques.
- [5 ] Lévy, C.L. (2013). Introduction l'étude des séries temporelles.
- [6 ] Mechgoug, R. (2013). La Prédiction des Séries Temporelles utilisant les Paradigmes de Soft Computing. Algérie.
- [7 ] Marine, E. (2015). Centre de Formation aux Métier de la Montagne. [15] Monbet, V. (2013). Modélisation de séries temporelles.
- [8 ] Serigne, M. (2007). Analyse et prévision des séries temporelles par la méthode de Box et Jenkins.
- [9 ] Benyammi. Y : "Étude de l'estimateur de l'ordre d'intégration fractionnaire de geweke porter-hudak d'un processus ARFIMA et applications empiriques" thèse de magister, ENSSEA, Alger 2009.
- [10 ] Beran.J, Feng.Y, Ghosh.S et Kulik.R : "Long-Memory Processes : Probabilist Properties and StatisticalMethods", Springer-Verlag Berlin, 2013.
- [11 ] Lardic.S et Mignon.V : "Prévision ARFIMA des taux de change : les modélisateurs doivent-ils encore exhorter à la naïveté des prévision ?" Annales d'économie et de statistique, No 54, 1999.
- [12 ] Palma.W : "long-memory time series :Theory and methodes", Wiley, canada 2007.
- [13 ] Johnston, J. et Dinardo, J. (1999). Méthodes de économétriques, 4ème Edition. Economica
- [14 ] D. C. Hoaglin and R. E. Welsch. The hat matrix in regression and anova. The American Statistician, 32 :17–22, 1978.

- [15 ] A.C. Atkinson. Two graphical displays for outlying and influential observations in regression. *Biometrika*, 68 :13–20, 1981.
- [16 ] Y. Dodge and V. Rousson. *Analyse de régression appliquée*. Dunod, 2004.
- [17 ] A. Sen and M. Srivastava. *Regression Analysis : Theory, Methods, and Applications*. Springer, 1990.