



uOttawa

Faculté des sciences  
Faculty of Science

07/02/2025

# Analyse des fluctuations saisonnnières du Prix et de la Production du Pétrole des États-Unis

2000-2024



Noms des membres :

Khadidja Sakina GOITA

Amadou Beydi Coulibaly

Gauze Andrea Maurelle Kouassi

Stephie Ozanna Keza

# Table des matières

Description des données .....	3
Intérêt pour les séries .....	3
Analyse de la production de pétrole.....	3
Source .....	3
Statistiques descriptives .....	3
Graphique de la production de pétrole.....	4
Analyse du prix du pétrole .....	5
Source .....	5
Statistiques descriptives .....	5
Graphique du prix du pétrole .....	6
Modifications apportées aux données .....	6
Filtrage .....	7
Filtrage Linéaire.....	7
Analyse du Lissage linéaire de la série temporelle de la production du pétrole ( $Q = 6$ ) .....	8
Analyse du lissage linéaire de la série temporelle du prix du pétrole ( $Q = 6$ ) .....	9
Filtrage Exponentiel.....	10
Lissage des données sur la production.....	11
Lissage des données sur les prix du pétrole .....	12
Comparaison des techniques de filtres .....	14
Analyse des résidus .....	14
Résidus Lissage linéaire ~ Série de la production .....	14
Résidus Lissage exponentiel ~ Série de la production.....	15
Conclusion sur les résidus de la série sur le prix du pétrole .....	15
Résidus Lissage linéaire ~ Série du prix .....	16
Résidus Lissage exponentiel ~ Série du prix .....	17
Conclusion sur les résidus de la série sur le prix du pétrole .....	17
Stationnarité des résidus .....	18
Différentiation.....	18
Différentiation de la série sur la production .....	18
Analyse de la stationnarité de la série sur la production.....	19
Différentiation de la série sur le prix du pétrole .....	21
Analyse de la stationnarité de la série sur le prix du pétrole .....	21
Stationnarité .....	23

Autocorrélation de la série sur la production de pétrole .....	24
Autocorrélation de la série sur le prix du pétrole.....	25
Modélisation et Prévision avec les Processus ARMA .....	26
selection des ordres des modeles ARMA .....	26
Analyse de l'ACF et du PACF.....	26
Serie de la production de pétrole .....	27
Serie du prix du petrole.....	29
Estimation de modèles ARMA .....	30
Estimation par Maximum de Vraisemblance.....	30
Serie sur la production de pétrole .....	31
Analyse des residus .....	32
Serie sur le prix du pétrole .....	33
Analyse des residus .....	34
Estimation des modeles imbriqués.....	35
Serie de la production .....	35
Analyse des residus (modele simplifié) .....	36
Serie sur le prix du petrole .....	37
Analyse des residus ( Modele simplifié).....	38
Analyse des Paramètres et Modèles Imbriqués .....	39
Serie sur la production de pétrole .....	39
Serie sur le prix du pétrole .....	40
Tests Diagnostiques .....	41
Critères d'Information (AIC/BIC) .....	41
Test du ratio de vraisemblance .....	41
Recommandations sur les modèles ARMA .....	44
Modélisation Bivariée et Corrélations Croisées .....	44
Analyse des innovations .....	45
Serie sur la production de petrole .....	45
Serie sur le prix du petrole .....	46
Propriétés statistiques des innovations.....	46
Corrélation croisée (CCF) entre les innovations.....	47
Conclusion .....	49
Annexe .....	50
Code .....	50

# **Description des données**

## **Intérêt pour les séries**

Le choix de ces deux séries repose sur leur influence directe sur le marché de l'énergie. Le Brent est l'une des principales références pour le pétrole brut et son prix dont l'évolution reflète les tendances du marché mondial dépend de plusieurs facteurs, tels que les tensions géopolitiques, la croissance économique mondiale ou encore les décisions de l'OPEP. À côté de cela, la production pétrolière américaine, également sensible, a connu un essor particulier au cours des dernières décennies grâce notamment au développement du pétrole de schiste, et a profondément transformé les dynamiques de l'offre et de la demande. De ce fait, l'étude conjointe de ces deux séries est pertinente car elle permet d'examiner l'interaction entre ces notions et peut révéler des effets de rétroaction.

## **Analyse de la production de pétrole**

### **Source**

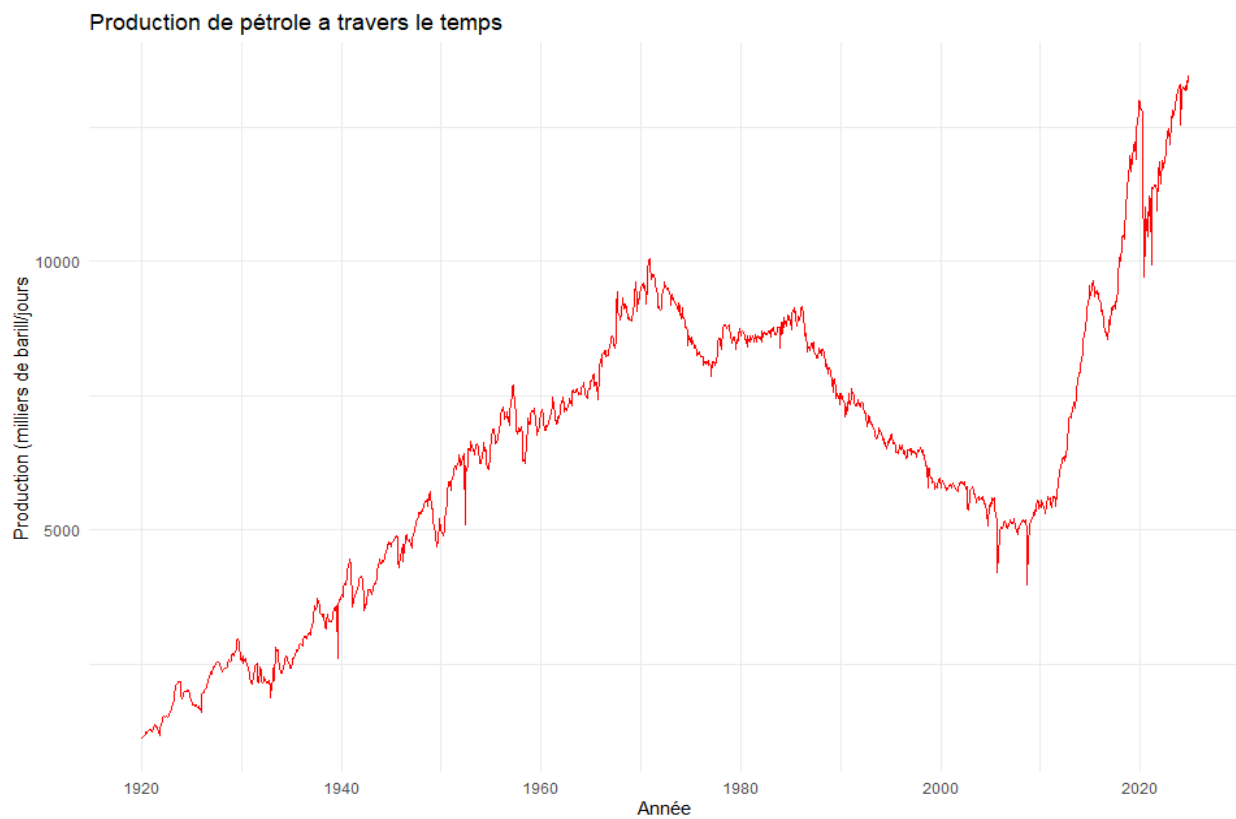
La deuxième série traite de la production mensuelle de pétrole brut aux États-Unis, compilée par l'Energy Information Administration (EIA) (source : Données EIA, consultée le 30 janvier 2025). Ces données, enregistrées à une fréquence mensuelle entre janvier 1920 et janvier 2024, sont mesurées en milliers de barils par jour (kb/j). Elles capturent l'évolution de la production pétrolière américaine sur un siècle, offrant un aperçu des capacités d'extraction, des transitions technologiques et des stratégies énergétiques nationales, sans correction des variations saisonnières.

### **Statistiques descriptives**

La production mensuelle de pétrole brut aux États-Unis, mesurée en milliers de barils par jour (kb/j) entre janvier 1920 et janvier 2024, présente une moyenne de 7 790 kb/j. La valeur minimale observée est de 3 974 kb/j, tandis que la valeur maximale atteint 13 308 kb/j, illustrant une forte variation au fil du temps. La médiane, soit la valeur centrale, est de 6 177 kb/j, ce qui indique que la moitié des observations sont inférieures à ce seuil. Le premier quartile (Q1) est de 5 501 kb/j et le troisième quartile (Q3) de 9 987 kb/j, suggérant que 50 % des valeurs sont comprises entre ces deux bornes. Enfin, l'écart-type s'élève à 2 741,16 kb/j, ce qui reflète une dispersion importante autour de la moyenne, typique d'une série aussi longue et influencée par des contextes économiques, géopolitiques et technologiques variés.

Statistique	Valeur (Kb/j)
Minimum	3974.00
1er Quartile	5501.00
Médiane	6177.00
Moyenne	7790.11
3ème Quartile	9987.00
Maximum	13308.00
Écart-type	2741.16

## Graphique de la production du pétrole



## **Analyse du prix du pétrole**

### **Source**

La série étudiée porte sur les prix mensuels du pétrole brut Brent, référence mondiale pour le marché pétrolier. Les données, extraites de la base FRED (source : International Monetary Fund Global Price of Brent Crude consultée le 30 janvier 2025), couvrent une période historique allant de janvier 1920 à janvier 2024, avec une fréquence mensuelle. Les valeurs sont exprimées en dollars américains par baril (USD/baril), offrant une vision globale des fluctuations des prix sur plus d'un siècle. Cette série, non désaisonnalisée et non ajustée par l'inflation, permet d'analyser les tendances structurelles et les chocs géopolitiques ou économiques ayant marqué le marché de l'énergie.

### **Statistiques descriptives**

Statistique	Valeur (dollars américains /baril)
Minimum	18.68
1er Quartile	43.33
Médiane	63.95
Moyenne	66.44
3ème Quartile	85.17
Maximum	133.59
Écart-type	28.96

Les prix mensuels du pétrole brut Brent, exprimés en dollars américains par baril, présentent une forte variabilité entre janvier 2000 et janvier 2024. Le prix le plus bas enregistré est de 18,68 dollars américains /baril, tandis que le maximum atteint 133,59 dollars américains /baril, reflétant des fluctuations importantes du marché énergétique. La moyenne s'élève à 66,44 dollars américains /baril, légèrement supérieure à la médiane de dollars américains /baril, ce qui indique une distribution légèrement asymétrique vers la droite. Les quartiles montrent que 25 % des prix sont inférieurs à 43,33 dollars américains /baril (Q1), et que 75 % sont inférieurs ou égaux à 85,17 dollars américains /baril (Q3). Enfin, l'écart-type de 28,96 dollars américains /baril témoigne de l'ampleur des variations de prix au cours de cette période de 24.

## Graphique du prix du pétrole



### Modifications apportées aux données

Dans le cadre de ce projet, nous avons décidé de restreindre la série de données sur la production pour ne conserver que les observations comprises entre 2000 et 2024. Cette période correspond à une phase de profondes transformations dans l'industrie pétrolière américaine, notamment avec l'essor du pétrole de schiste à partir de la fin des années 2000. Ce phénomène a fortement modifié la dynamique de production aux États-Unis, faisant du pays l'un des plus grands producteurs mondiaux. Par ailleurs, cette période inclut plusieurs événements majeurs ayant influencé les marchés pétroliers, tels que les crises économiques de 2008 et de 2020, ou encore les décisions stratégiques de l'OPEP. Ainsi, se concentrer sur cette plage temporelle nous permettra d'éviter les biais liés aux méthodes de mesure historiques tout en alignant les séries avec d'autres variables pour une analyse cohérente des interactions entre innovation technologique, géopolitique et transition énergétique.

# **Filtrage**

## **Filtrage Linéaire**

Un filtre de lissage linéaire est une méthode utilisée en analyse des séries temporelles pour atténuer les fluctuations aléatoires à court terme et mettre en évidence la tendance générale d'une série. Il consiste à remplacer chaque valeur d'observation par une moyenne pondérée souvent simplement la moyenne arithmétique de ses voisins les plus proches. Ce type de filtre permet ainsi de mieux visualiser les composantes structurelles d'une série, comme les tendances ou les cycles, tout en réduisant le bruit statistique. Le lissage linéaire est particulièrement utile lorsqu'on souhaite analyser des données économiques ou financières sujettes à des variations irrégulières.

$$m_n = \frac{1}{2q + 1} \sum_{j=-q}^q X_{n+j}$$

Où :

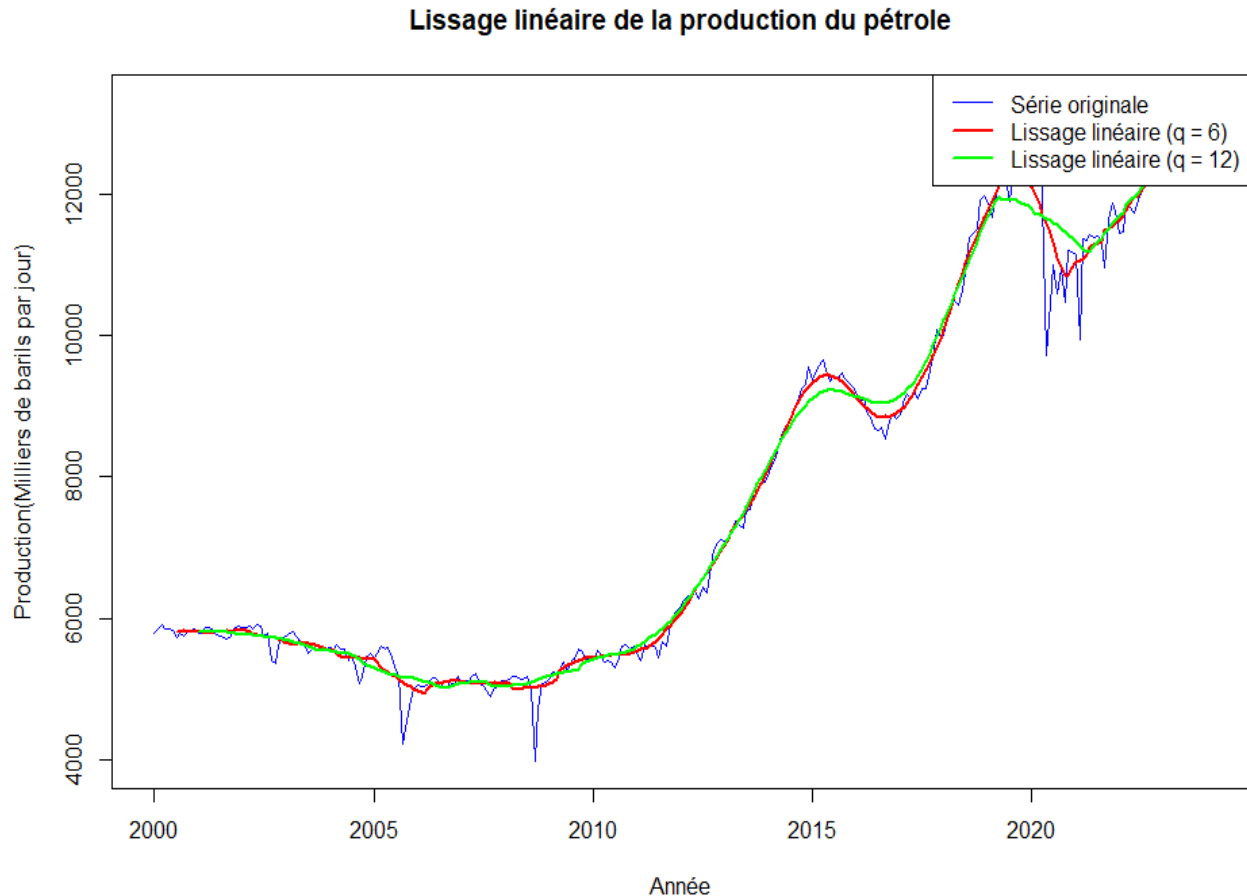
- $m_n$  : valeur lissée à l'instant  $n$ , c'est-à-dire la moyenne centrée autour de  $n$
- $x_{n+j}$  : valeur observée de la série temporelle au temps  $n + j$
- $q$  : nombre de points pris en compte de part et d'autre de  $n$  (la demi-largeur de la fenêtre de lissage)
- $n$  : indice temporel courant pour lequel on calcule la valeur lissée.

Pour optimiser notre filtre de lissage linéaire, nous allons tester plusieurs valeurs possibles de  $q$ , qui représente le nombre de points de part et d'autre de l'instant  $n$  que l'on inclut dans la moyenne. En comparant les résultats obtenus avec différentes tailles de fenêtre, nous pourrions évaluer l'impact du paramètre  $q$  sur la qualité du lissage et identifier la valeur qui permet d'obtenir la représentation la plus fidèle et la plus informative de la tendance sous-jacente de notre série temporelle. Ce processus d'ajustement est essentiel afin d'assurer une analyse pertinente.

Nous testerons les paramètres  $q= 6$  et  $q= 12$ .



## Lissage linéaire de la série temporelle de la production du pétrole



Le lissage linéaire appliqué à la série de production de pétrole avec  $q=6$  permet d'atténuer les fluctuations mensuelles tout en conservant les principales dynamiques de la série. Cette transformation met en évidence une tendance plus claire et plus lisible, en particulier dans les phases de transition.

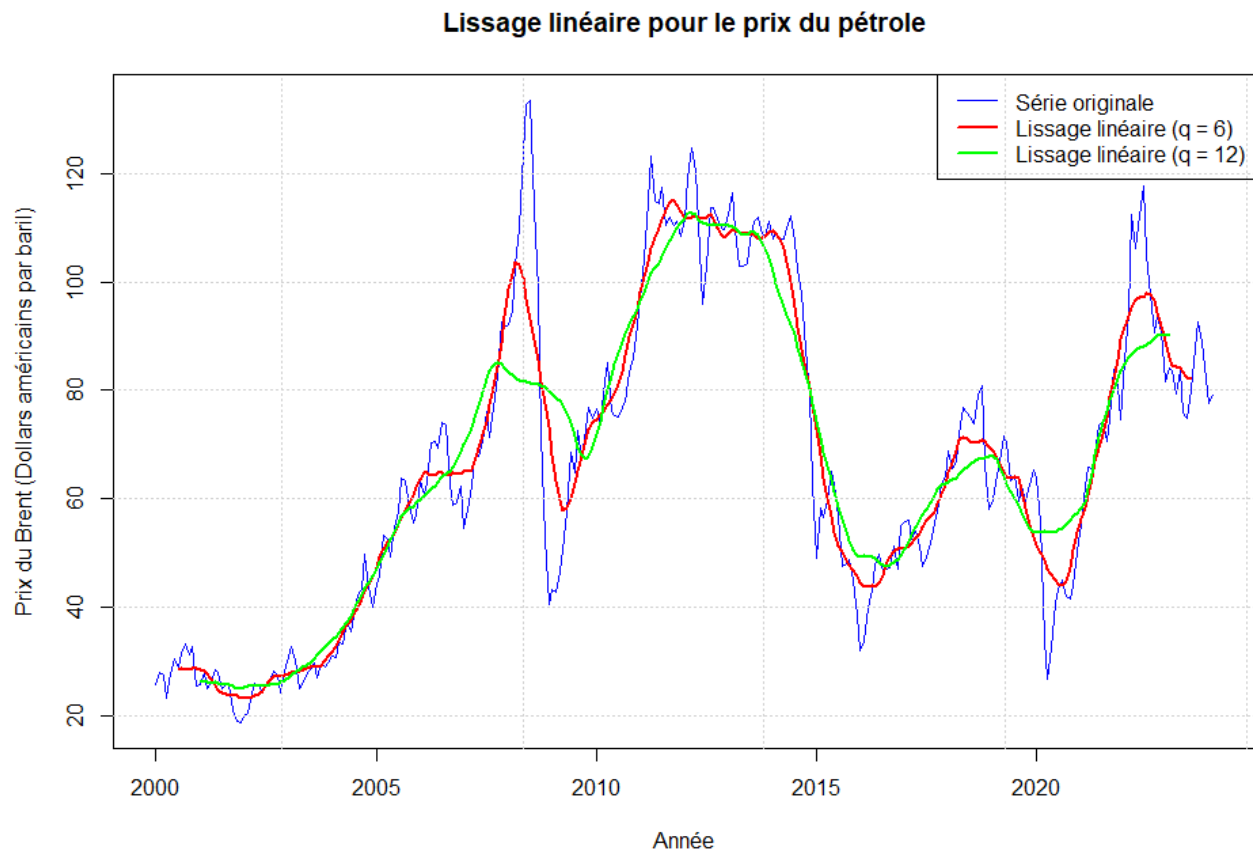
Entre 2000 et 2009, la production reste relativement stable autour de 5 000 kb/j, avec quelques variations ponctuelles. Une légère baisse est observable entre 2006 et 2008, probablement liée à des ajustements économiques ou géopolitiques.

À partir de 2010, on observe une nette inflexion dans la tendance. La production commence à croître rapidement, amorçant une phase d'expansion soutenue jusqu'en 2015, où elle dépasse les 9 000 kb/j. Cette période correspond à l'essor du pétrole de schiste aux États-Unis, qui a profondément transformé la dynamique de l'offre.

De 2015 à 2020, la tendance reste globalement à la hausse, malgré quelques ralentissements temporaires visibles dans les données non lissées (possiblement dus à des événements extérieurs comme des décisions de l'OPEP ou la pandémie).

En somme, le lissage avec  $q=6$  offre un bon compromis : il réduit le bruit aléatoire tout en conservant la structure des grandes tendances. Il permet ainsi de mieux visualiser les cycles de stabilité, de transition et de croissance dans l'évolution de la production pétrolière américaine.

### **Lissage linéaire de la série temporelle du prix du pétrole**



Le lissage linéaire appliqué à la série des prix du Brent avec une fenêtre  $q=6$  permet de dégager les grandes tendances du marché pétrolier tout en réduisant les fluctuations de court terme. Cette approche rend plus lisible l'évolution des prix, notamment dans les périodes de forte volatilité.

De 2000 à 2007, on observe une hausse progressive des prix, reflétant une demande énergétique mondiale croissante, portée notamment par la Chine et les économies émergentes. Cette tendance se poursuit jusqu'à un pic en 2008, où le prix du Brent dépasse les 140 dollars américains /baril, juste avant l'éclatement de la crise financière mondiale.

La période 2008–2009 est marquée par un effondrement brutal des prix, conséquence directe de la récession mondiale. Le lissage met en évidence cette chute tout en adoucissant l'extrême volatilité présente dans la série brute.

Entre 2010 et 2014, les prix se stabilisent autour de 100 à 110 dollars américains, une phase relativement calme marquée par un équilibre entre l'offre et la demande. Cette période prend fin en 2014–2015, avec un nouveau recul prononcé dû à une surabondance de l'offre et à une guerre des prix entre producteurs.

À partir de 2020, la série montre une nouvelle chute rapide, correspondant à l'impact de la pandémie de COVID-19 sur la demande mondiale. Le prix du Brent atteint brièvement des niveaux historiquement bas, avant de remonter progressivement jusqu'à environ 80 à 90 dollars américains en 2022.

Le lissage avec  $q=6$  permet de mieux identifier ces cycles haussiers et baissiers majeurs sans être perturbé par les variations irrégulières à court terme. Ce filtrage s'avère donc particulièrement utile pour analyser l'évolution des prix à moyen et long terme.

### **Filtrage Exponentiel**

Le lissage exponentiel est une méthode de lissage utilisée pour donner plus de poids aux observations récentes d'une série temporelle tout en conservant une mémoire des valeurs passées. Contrairement au lissage linéaire qui accorde un poids égal à toutes les observations dans une fenêtre, le lissage exponentiel attribue des poids décroissants de manière exponentielle à mesure que l'on s'éloigne dans le passé. Cette approche est particulièrement efficace pour détecter les tendances récentes tout en réduisant l'effet du bruit aléatoire.

$$\hat{m}_1 = X_1$$

$$\hat{m}_n = \sum_{j=1}^n \omega_j X_j \quad \text{pour } n = 2, 3, \dots, N$$

$$\omega_1 = (1 - \alpha)^{n-1}$$

$$\omega_j = \alpha (1 - \alpha)^{n-j} \quad \text{pour } j = 2, 3, \dots, n$$

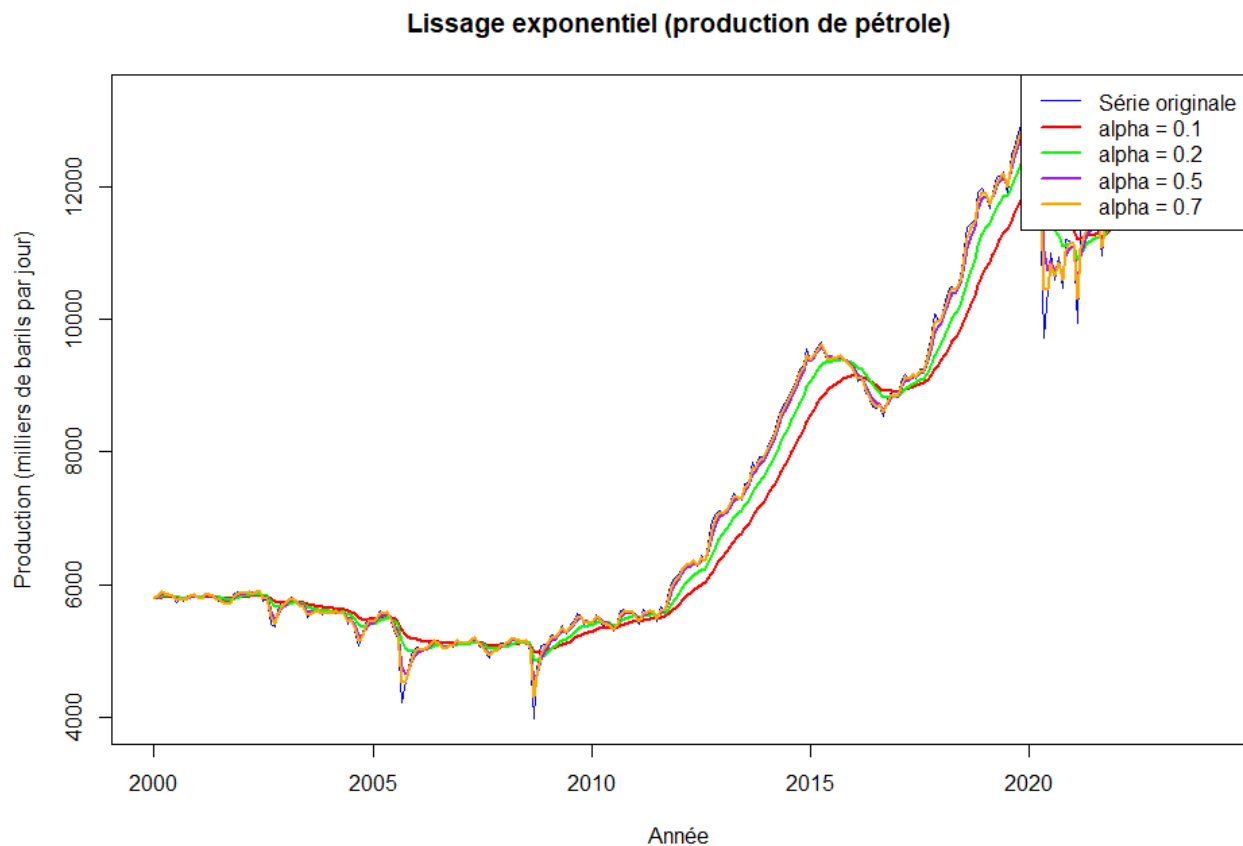
Où:

- $\hat{m}_n$ : valeur lissée à l'instant  $n$
- $x_j$ : valeur observée de la série au temps  $j$
- $\omega_j$ : poids attribué à l'observation  $x_j$ , décroissant avec l'ancienneté

- $\alpha \in (0,1)$  : paramètre de lissage, qui contrôle la rapidité de décroissance des poids
- $n$  : indice temporel courant
- $N$  : longueur totale de la série

Pour chacune des séries, nous appliquerons le lissage exponentiel et justifierons le choix du paramètre  $\alpha$  en fonction des résultats obtenus. Afin d'optimiser notre modèle, nous testerons plusieurs valeurs de  $\alpha$  afin d'évaluer leur impact sur la qualité du lissage. L'objectif est d'identifier la valeur optimale qui permet de réduire efficacement les fluctuations tout en conservant les variations significatives de la série.

## Lissage des données sur la production



Le lissage exponentiel avec  $\alpha = 0.2$  permet de réduire la volatilité tout en révélant la tendance générale de la production de pétrole. La courbe lissée suit les grandes évolutions de la série sans être trop influencée par les fluctuations ponctuelles à court terme.

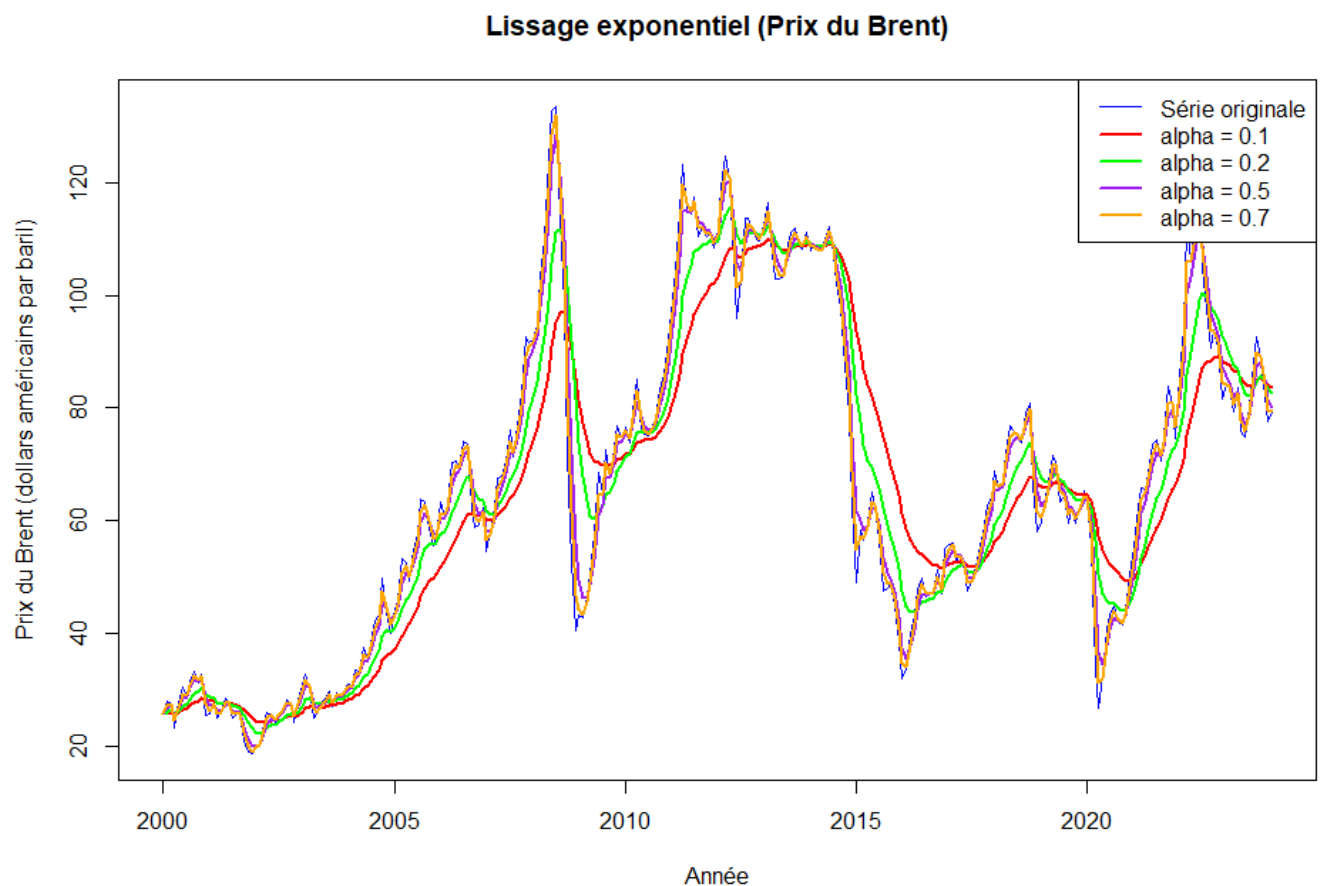
Entre 2000 et 2009, la production reste relativement stable autour de 5 000 kb/j, avec quelques variations mineures. Une baisse modérée est observable entre 2006 et 2008, probablement liée à des ajustements économiques ou à la conjoncture énergétique mondiale.

À partir de 2010, une phase de croissance marquée s'amorce. Le lissage capture cette tendance ascendante jusqu'en 2015, période durant laquelle la production dépasse les 9 000 kb/j. Cette dynamique correspond à l'essor du pétrole de schiste aux États-Unis.

De 2015 à 2020, malgré quelques ralentissements temporaires notamment en 2020 avec l'impact de la pandémie la tendance globale reste à la hausse. Le lissage permet de suivre cette évolution tout en adoucissant les variations extrêmes visibles dans la série brute.

En résumé,  $\alpha=0.2$  représente un bon compromis : il permet de lisser les fluctuations aléatoires tout en maintenant une bonne réactivité aux tendances structurelles, ce qui rend la série plus facile à interpréter dans le cadre d'une analyse conjoncturelle ou de modélisation.

### **Lissage exponentiel des données sur les prix du pétrole**



Le lissage exponentiel appliqué à la série des prix du Brent avec un  $\alpha=0.2$  permet de dégager les grandes tendances du marché tout en atténuant les fluctuations de court terme. La courbe lissée (en vert) suit efficacement les mouvements significatifs du prix sans être perturbée par les irrégularités passagères.

De 2000 à 2008, le prix du Brent connaît une hausse marquée, culminant à plus de 140 USD juste avant la crise financière. Cette ascension est bien captée par la courbe lissée, qui reflète la dynamique de fond liée à la forte demande mondiale.

En 2009, la crise économique provoque une chute brutale des prix, qui est visiblement atténuée mais conservée dans sa structure par le lissage. Par la suite, entre 2010 et 2014, on observe une période de stabilité autour de 100 USD, également bien représentée.

Le lissage montre ensuite la baisse rapide de 2014-2015, due à l'excès d'offre sur le marché, suivie d'une phase plus instable entre 2016 et 2019. Enfin, la chute spectaculaire liée à la pandémie de COVID-19 en 2020, puis la reprise en 2021–2022, sont visibles dans la série, mais de manière plus lissée et structurée.

En résumé, ce lissage permet de mieux comprendre les cycles économiques et les phases de choc dans le marché pétrolier, tout en offrant une représentation fluide et lisible de l'évolution du prix du Brent sur plus de deux décennies.

### **Pourquoi $\alpha=0.2$ ?**

Le paramètre  $\alpha$  dans le lissage exponentiel détermine le poids accordé aux observations les plus récentes. Un  $\alpha$  élevé rend le lissage très réactif aux variations de court terme, ce qui peut accentuer le bruit. À l'inverse, un  $\alpha$  faible produit un lissage plus doux, mais peut retarder la détection des tendances significatives.

Pour la série de production, nous avons testé différentes valeurs de  $\alpha$  (0.1, 0.2, 0.5 et 0.7). Le lissage obtenu avec  $\alpha = 0.2$  s'est révélé le plus pertinent : il réduit efficacement les fluctuations irrégulières tout en conservant la dynamique générale de la série. Contrairement à  $\alpha = 0.7$ , qui suit de trop près les moindres variations, ou  $\alpha = 0.1$ , qui lisse excessivement les mouvements,  $\alpha = 0.2$  offre un bon équilibre entre stabilité et réactivité. Il permet une lecture plus claire des grandes phases d'évolution de la production pétrolière.

Dans le cas des prix du Brent, le choix d'un  $\alpha$  adapté est crucial pour filtrer la forte volatilité du marché sans effacer les signaux importants. Après avoir comparé les effets de plusieurs valeurs de  $\alpha$  (0.1, 0.2, 0.5 et 0.7), nous avons constaté que lissage avec  $\alpha = 0.2$  permettait de bien atténuer les fluctuations de court terme tout en préservant les mouvements de fond du prix.

$\alpha=0.2$  se distingue par sa capacité à lisser les pics ponctuels sans effacer les cycles haussiers ou baissiers majeurs, tels que ceux observés en 2008, 2014 ou 2020. Ce choix assure ainsi une visualisation claire des tendances économiques sous-jacentes, sans être perturbée par les micro-variations liées à la spéculation ou aux événements conjoncturels.

## Comparaison des techniques de filtrages

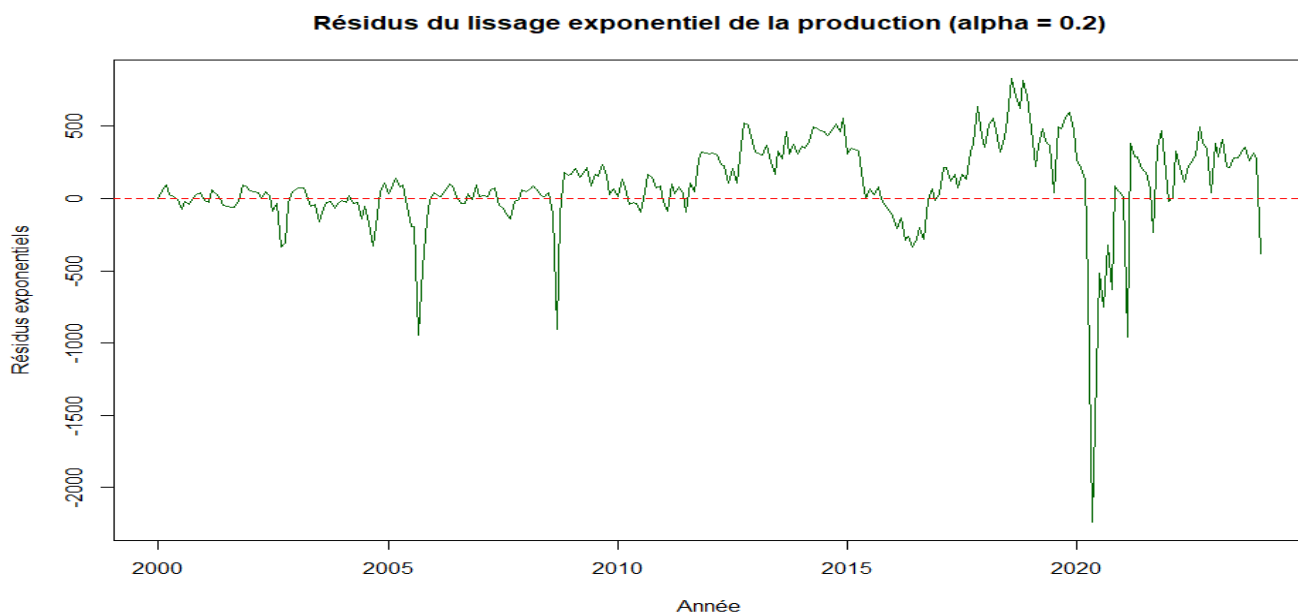
Chaque méthode de lissage a ses avantages et ses inconvénients, qui dépendent également du type de série analysée. En effet, le lissage linéaire permet d'identifier une trajectoire claire et de fournir des projections à long terme. Cependant, elle manque de réactivité aux variations de court terme et ne capture pas d'éventuels changements structurels dans la production.

À l'inverse, le lissage exponentiel permet de suivre de près les fluctuations du marché en donnant plus de poids aux observations récentes. Cette approche est particulièrement utile pour détecter les retournements de tendance et mieux comprendre la dynamique des prix à court terme. Toutefois, elle est moins efficace pour établir des prévisions à long terme, car elle dépend fortement des valeurs récentes et ne fournit pas de tendance générale claire.

En conclusion, le choix de la méthode de lissage dépend de la nature des données et des objectifs d'analyse. Pour une série plus stable comme la production, un lissage linéaire est préférable pour dégager une tendance globale. Pour une série plus volatile comme le prix du Brent, un lissage exponentiel permet de mieux capturer les variations du marché. Une approche combinée, utilisant les deux méthodes selon le contexte, peut être envisagée pour une analyse plus complète.

## Analyse des résidus

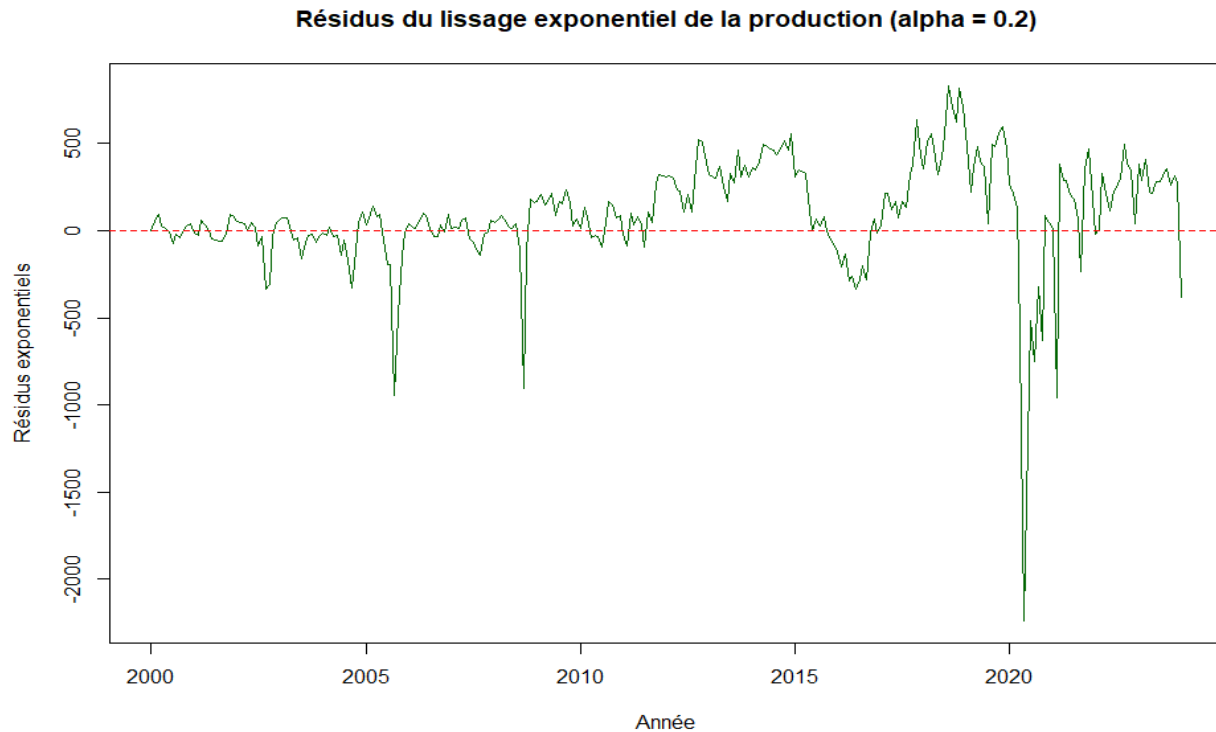
### Résidus Lissage linéaire ~ Série de la production



## Variance des résidus (linéaire - production) : 60503.89

## Moyenne des résidus (linéaire - production) : -2.256873

## Résidus Lissage exponentiel ~ Série de la production



```
## Variance des résidus (exponentiel - production) : 92388.75
```

```
## Moyenne des résidus (exponentiel - production) : 99.02292
```

### Conclusion sur les résidus de la série sur le prix du pétrole

Pour comparer les deux méthodes de lissage appliquées à la série de la production de pétrole, nous avons analysé les résidus de chaque modèle (valeurs observées moins valeurs lissées). Un bon lissage devrait produire des résidus ayant une moyenne proche de zéro (absence de biais systématique) et une variance faible (faible écart entre la série originale et la série lissée).

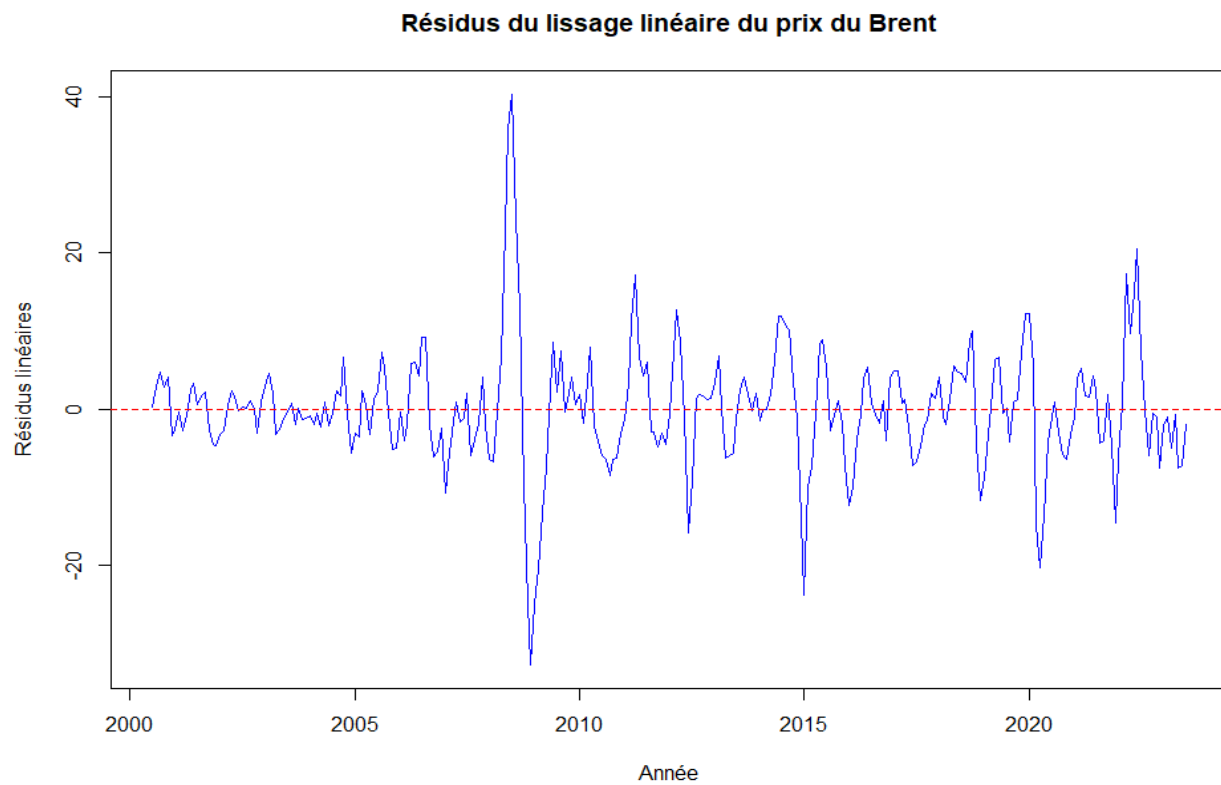
Dans le cas du lissage linéaire, la moyenne des résidus est de -2.26 et la variance est de 60503.89. En revanche, le lissage exponentiel présente une moyenne beaucoup plus éloignée de zéro (99.02), accompagnée d'une variance plus élevée (92388.75).

Ces résultats indiquent que le lissage linéaire est nettement plus performant, car il suit la série originale de façon plus fidèle tout en gardant les erreurs centrées et plus dispersées. Le lissage exponentiel semble introduire un biais important et s'écarte davantage des données réelles.

Nous concluons donc que le **lissage linéaire** est préférable pour cette série, car il minimise à la fois l'erreur moyenne et la dispersion des résidus.



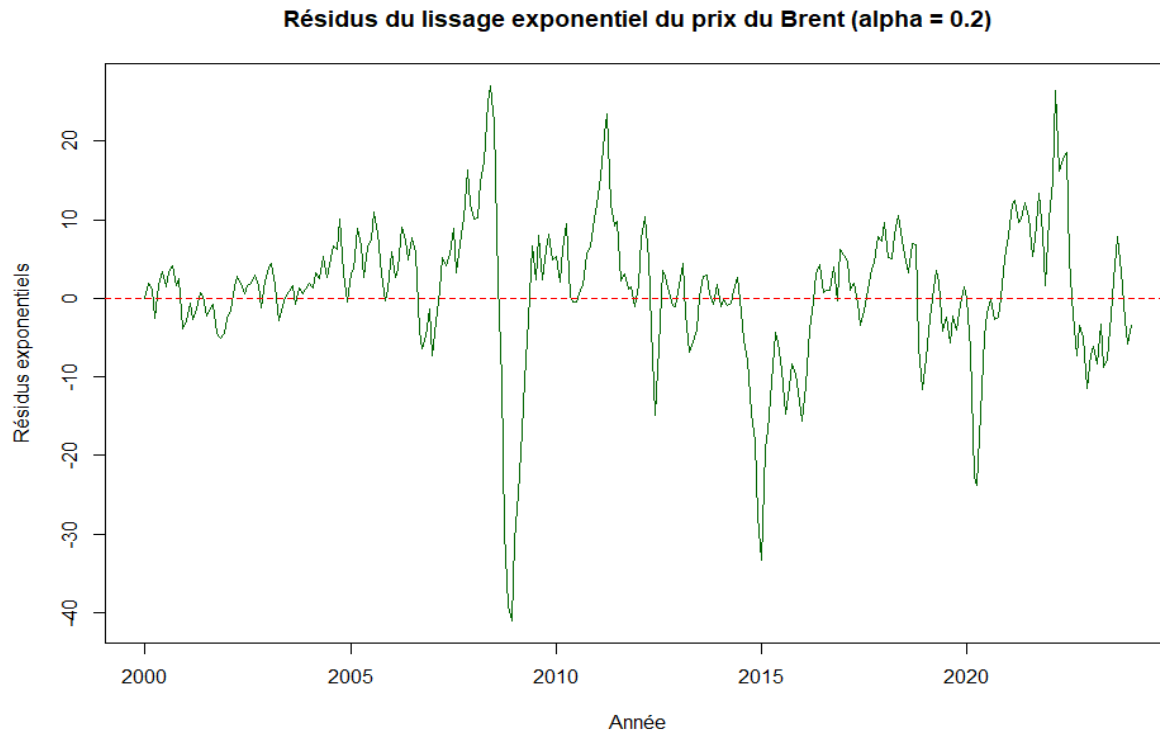
## Résidus Lissage linéaire ~ Série du prix



```
## Variance des résidus (linéaire - prix) : 59.5204
```

```
## Moyenne des résidus (linéaire - prix) : -0.02451076
```

## Résidus Lissage exponentiel ~ Série du prix



```
## Variance des résidus (exponentiel - prix) : 85.38415
```

```
## Moyenne des résidus (exponentiel - prix) : 0.7879855
```

## Conclusion sur les résidus de la série sur le prix du pétrole

Comme pour la série de la production, nous avons comparé les performances du lissage linéaire et du lissage exponentiel sur la série du prix du pétrole Brent en analysant leurs résidus. Le lissage linéaire produit des résidus avec une moyenne de -0.025 et une variance de 59.52, tandis que le lissage exponentiel présente une moyenne de 0.79 et une variance plus élevée de 85.38.

Ces résultats montrent que le lissage linéaire génère des erreurs plus faibles et moins biaisées, tandis que le lissage exponentiel tend à surestimer la série, comme l'indique sa moyenne des résidus plus éloignée de zéro. De plus, la plus grande variance des résidus exponentiels suggère une moins bonne capacité à suivre les fluctuations du prix réel.

Nous concluons donc que, pour la série du prix du pétrole également, le **lissage linéaire est plus adapté**, car il fournit une estimation plus fidèle et plus stable des données observées.

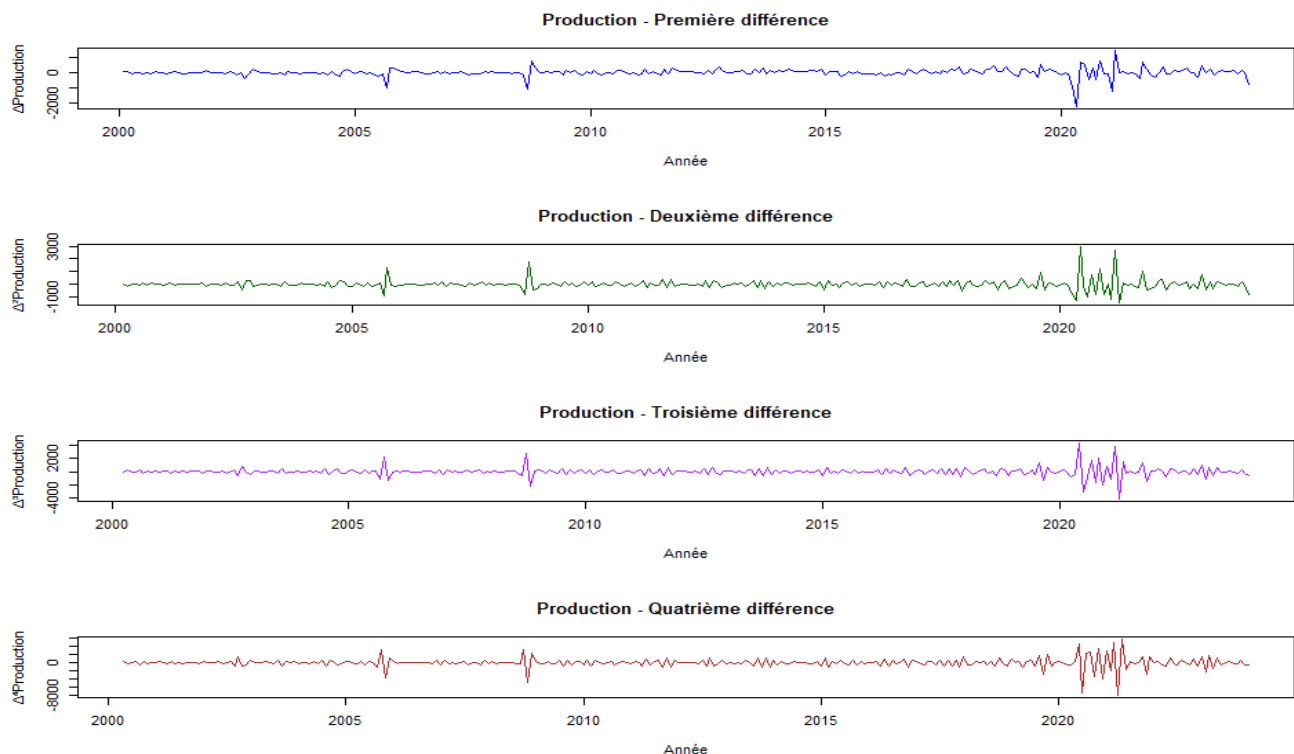
## Stationnarité des résidus

L'analyse des résidus révèle la présence de composantes saisonnières, marquées par une variabilité périodique répétée. Cette cyclicité suggère que la moyenne des résidus n'est pas constante dans le temps, car les fluctuations suivent un motif récurrent. De plus, l'amplitude des variations (par exemple, des résidus passant de 20 à 606) évolue au cours de la série, indiquant une variance non constante (hétéroscédasticité).

Or, pour qu'une série soit stationnaire au sens faible, elle doit satisfaire trois conditions : une moyenne constante dans le temps, une variance constante (homoscédasticité) et une autocorrélation dépendant uniquement du décalage temporel, et non du temps. Ces motifs, combinés à l'absence de structure aléatoire stable, permettent de rejeter la stationnarité faible. Les résidus issus du lissage linéaire du prix du Brent et de la production ne satisfont pas les critères de stationnarité au sens faible.

## Différentiation

### Différentiation de la série sur la production



## Analyse de la stationnarité de la série sur la production

Après avoir observé la série originale de la production de pétrole, on remarque une tendance à la hausse ainsi que des variations d'amplitude dans le temps, indiquant une non-stationnarité. Pour y remédier, nous avons effectué des différentiations successives d'ordre 1 à 4. On constate que la variance augmente significativement à chaque étape (ex. : de 69 750 à 1 533 378), ce qui peut indiquer que l'on est en train d'introduire du bruit ou d'amplifier l'instabilité de la série au lieu de la corriger.

## Moyenne et variance des différentiations successives de la série sur la production :

## 1ère différentiation : Moyenne = 23.5069 | Variance = 69750.45

## 2ème différentiation : Moyenne = -2.8641 | Variance = 153181.9

## 3ème différentiation : Moyenne = -2.7238 | Variance = 459825.2

## 4ème différentiation : Moyenne = -1.9228 | Variance = 1533378

Idéalement, une différentiation doit stabiliser la moyenne et la variance sans exagérément augmenter le bruit. Ici, la 1ère différentiation semble réduire la tendance, tout en gardant une variance relativement contrôlée. Dès la 2e et surtout à partir de la 3e, la variance augmente fortement, et la moyenne reste proche de zéro sans réelle amélioration apparente.

Cela suggère que la série devient trop instable après 2 ou 3 différentiations, ce qui est un indice qu'il y'a des différentiations en trop. On en déduit que l'ordre de différentiation optimal est probablement de 1, voire 2 au maximum.

En conclusion, nous avons choisi un ordre de différentiation de 1, car la série semble visuellement stabilisée dès la première. La moyenne devient plus constante, et bien que la variance augmente légèrement, elle demeure raisonnable comparée aux ordres supérieurs, où elle explose. Au-delà de la deuxième différentiation, on observe une instabilité croissante et une amplification du bruit, ce qui suggère une différentiation inutile.

Afin de valider cette hypothèse et de confirmer si la série différenciée est véritablement stationnaire, nous effectuerons un test de Ljung-**Box** sur la série différenciée d'ordre 1. Ce test permettra de vérifier l'absence d'autocorrélation significative, une condition essentielle à la stationnarité.

Le **test de Ljung-Box** est un test statistique utilisé pour examiner l'hypothèse d'absence d'autocorrélation dans une série temporelle ou, plus généralement, dans les résidus d'un modèle ajusté. Autrement dit, il vérifie si les valeurs successives d'une série (ou des résidus) sont indépendantes les unes des autres.

Formellement, la statistique du test, notée  $Q$ , est définie comme suit :

$$Q = n(n + 2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n - k},$$

Où :

- $n$  est la taille de l'échantillon
- $m$  est le nombre de retards (lags) testés
- $\hat{\rho}_k$  est l'autocorrélation empirique de la série (ou des résidus) au retard  $k$ .

Sous l'hypothèse nulle d'absence d'autocorrélation (autocorrélation égale à zéro pour tous les retards considérés), la statistique  $Q$  suit approximativement une loi du  $\chi^2$  à  $m$  degrés de liberté (si aucun paramètre n'a été estimé ou si l'on corrige le nombre de degrés de liberté en conséquence).

- **Hypothèse nulle ( $H_0$ )** : Il n'existe pas d'autocorrélation dans la série (ou les résidus).
- **Hypothèse alternative ( $H_1$ )** : Il existe une autocorrélation significative pour au moins un des retards testés.

Lorsque la p-valeur du test est inférieure au seuil de signification (par exemple 5 %), on rejette l'hypothèse nulle et on conclut qu'il y a une autocorrélation résiduelle significative dans la série ou les résidus considérés.

## Résultats du test de Ljung-Box sur la première différenciation :

## Statistique de test : 36.4827

## Degrés de liberté : 20

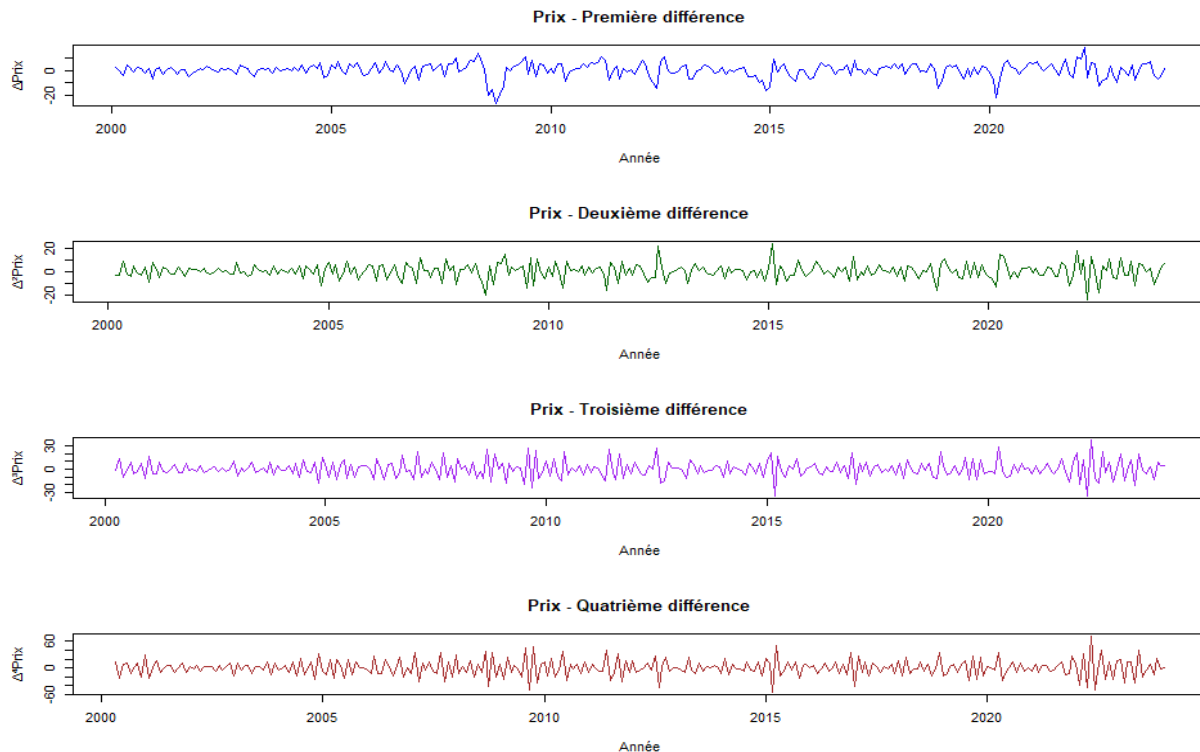
## p-value : 0.0135

## La p-value est inférieure à 0.05 : nous rejetons l'hypothèse nulle d'indépendance.

## La série présente encore de l'autocorrélation : elle n'est pas totalement stationnaire.

Les résultats du test de Ljung-Box (statistique = 36,48 ; p-value = 0,0135) montrent que la p-value est inférieure au seuil de 5 %. Nous rejetons donc l'hypothèse nulle d'absence d'autocorrélation, ce qui indique que la première différenciation de la série de production n'élimine pas entièrement les dépendances temporelles. Autrement dit, la série différenciée reste partiellement autocorrélée et ne peut pas être considérée comme totalement stationnaire.

## Différentiation de la série sur le prix du pétrole



## Analyse de la stationnarité de la série sur le prix du pétrole

## Moyenne et variance des différentiations successives (prix du pétrole) :

## 1ère différentiation : Moyenne = 0.1857 | Variance = 33.2664

## 2ème différentiation : Moyenne = -0.0029 | Variance = 42.9417

## 3ème différentiation : Moyenne = 0.032 | Variance = 111.5051

## 4ème différentiation : Moyenne = 0.0174 | Variance = 352.0003

Comme pour la série de la production, nous avons analysé les différentiations successives de la série du prix du pétrole Brent afin de déterminer l'ordre optimal pour obtenir une stationnarité.

Dès la première différentiation, la moyenne devient proche de zéro (0.1857), et la variance est relativement modérée (33.27). En revanche, à partir de la deuxième différentiation, la variance augmente progressivement, atteignant 42.94 pour la deuxième, 111.51 pour la troisième, et 352.00 pour la quatrième. Cela indique que plus il y'a de différentiations, plus on ajoute du bruit à la série, sans forcément améliorer sa stabilité.

De plus, la moyenne oscille autour de zéro sans tendance marquée, ce qui suggère que la première différentiation pourrait être suffisante pour rendre la série stationnaire en moyenne.

Nous retenons donc l'ordre de différenciation 1, car il permet d'atténuer les tendances tout en préservant une variance raisonnable. Une différenciation supplémentaire n'améliore pas significativement les propriétés de stationnarité et risque au contraire de détériorer la qualité des données. Cette hypothèse est soutenue par le test de Ljung-Box effectué sur la série différenciée d'ordre 1, que nous présentons ci-dessous.

Afin de confirmer si la série du prix du pétrole est réellement stationnaire après une différenciation d'ordre 1, nous allons effectuer un test de Ljung-Box. Ce test statistique permet d'évaluer l'hypothèse selon laquelle les résidus (ou valeurs de la série) ne présentent pas d'autocorrélation significative jusqu'à un certain nombre de retards (lags).

## Résultats du test de Ljung-Box sur la première différenciation (prix du pétrole) :

## Statistique de test : 55.9875

## Degrés de liberté : 20

## p-value : 2.9195e-05

## → La p-value est inférieure à 0.05 : nous rejetons l'hypothèse nulle d'indépendance.

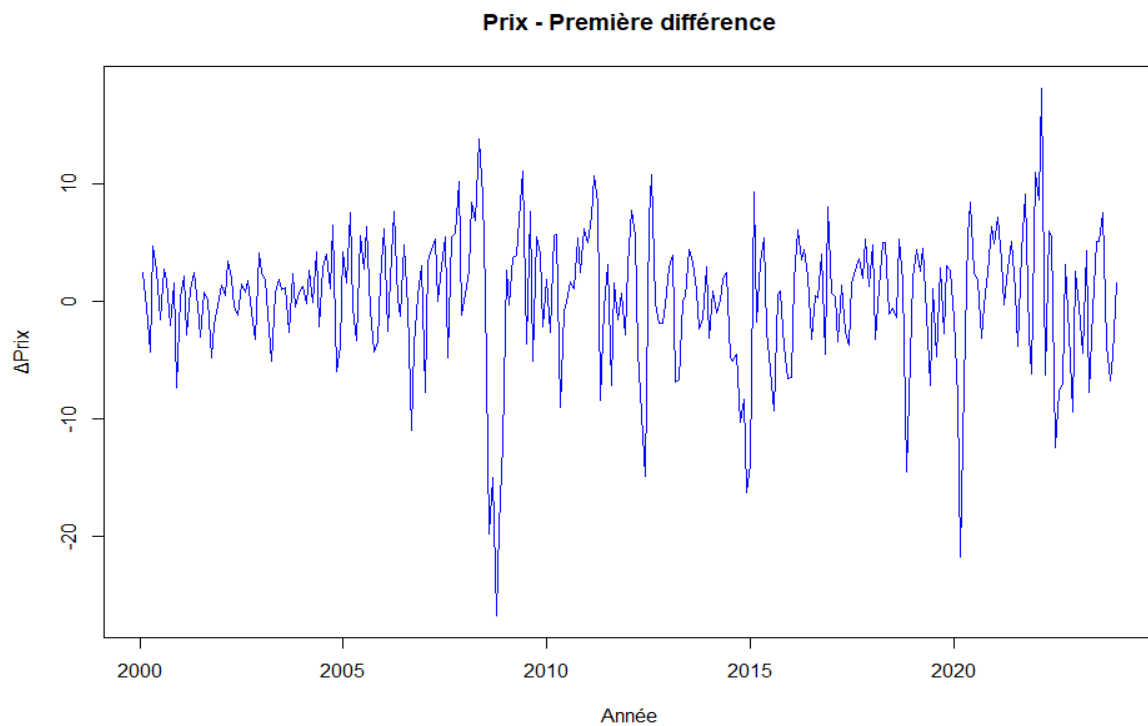
## → La série présente encore de l'autocorrélation : elle n'est donc pas totalement stationnaire.

Les résultats du test de Ljung-Box (statistique = 55,99 ; p-value =  $2,92 \times 10^{-5}$ ) indiquent que la p-value est nettement inférieure au seuil de 5 %. Nous rejetons ainsi l'hypothèse nulle d'absence d'autocorrélation, ce qui signifie que la première différenciation de la série de prix ne supprime pas totalement les dépendances temporelles. En d'autres termes, la série différenciée demeure partiellement autocorrélée et ne peut donc pas être considérée comme parfaitement stationnaire.

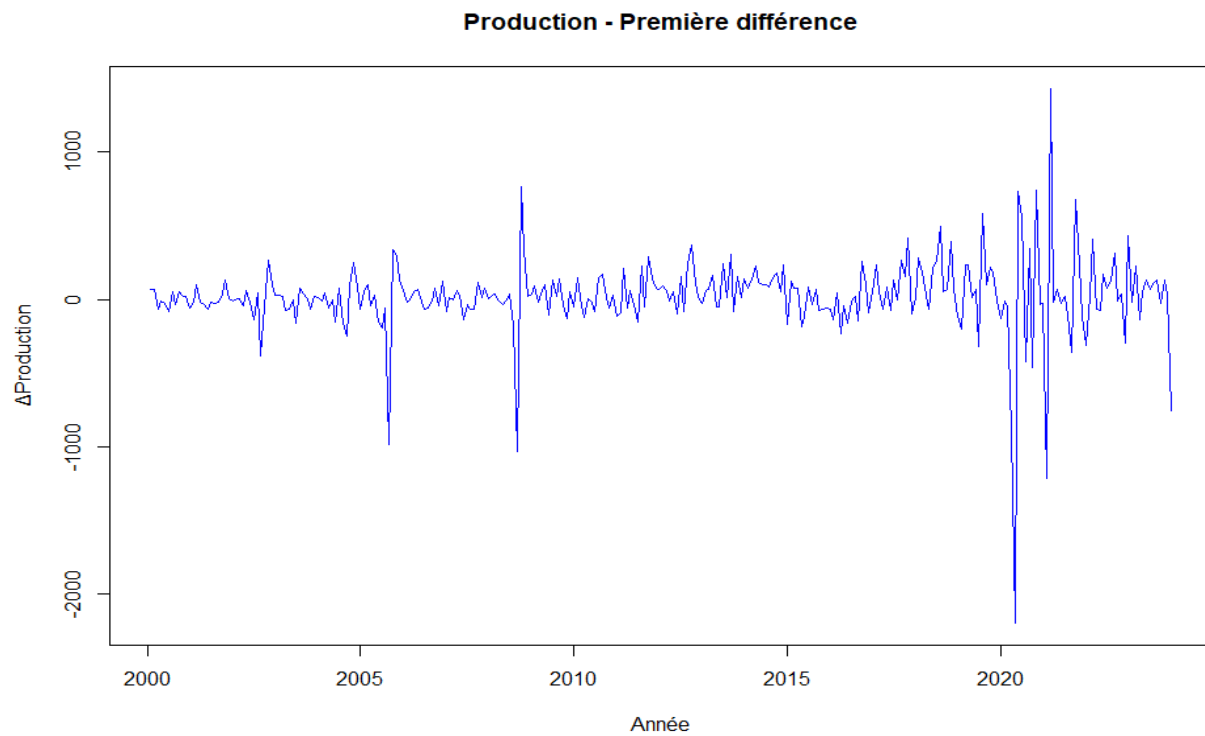
Par ailleurs, nous avons choisi d'utiliser un lag de 20 pour le test de Ljung-Box, car il permet d'évaluer l'autocorrélation jusqu'à 20 périodes (soit environ 1 an et demi de données mensuelles). Ce choix est plus approprié pour des séries économiques comme le prix ou la production du pétrole, qui peuvent présenter des cycles ou des effets saisonniers étalés dans le temps.

# Stationnarité

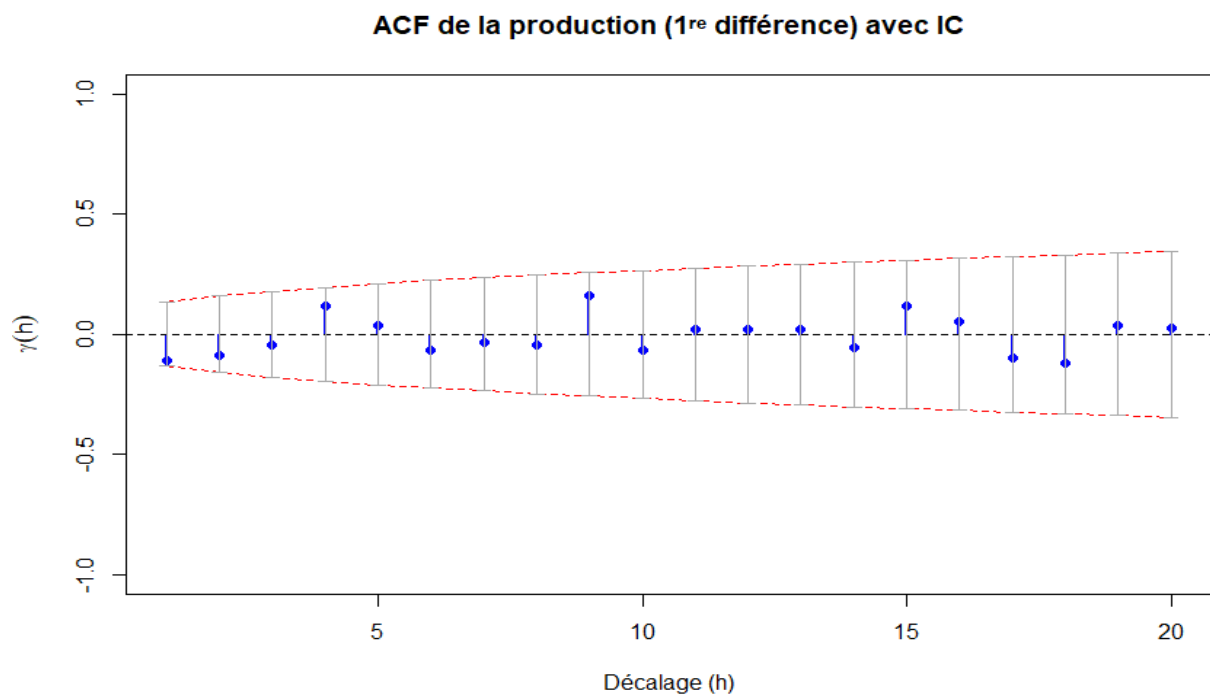
Nous avons décidé de travailler avec la première différenciation des séries de production et de prix du pétrole, car l'analyse statistique (notamment le test de stationnarité) a révélé une p-valeur plus satisfaisante par rapport à d'autres transformations.



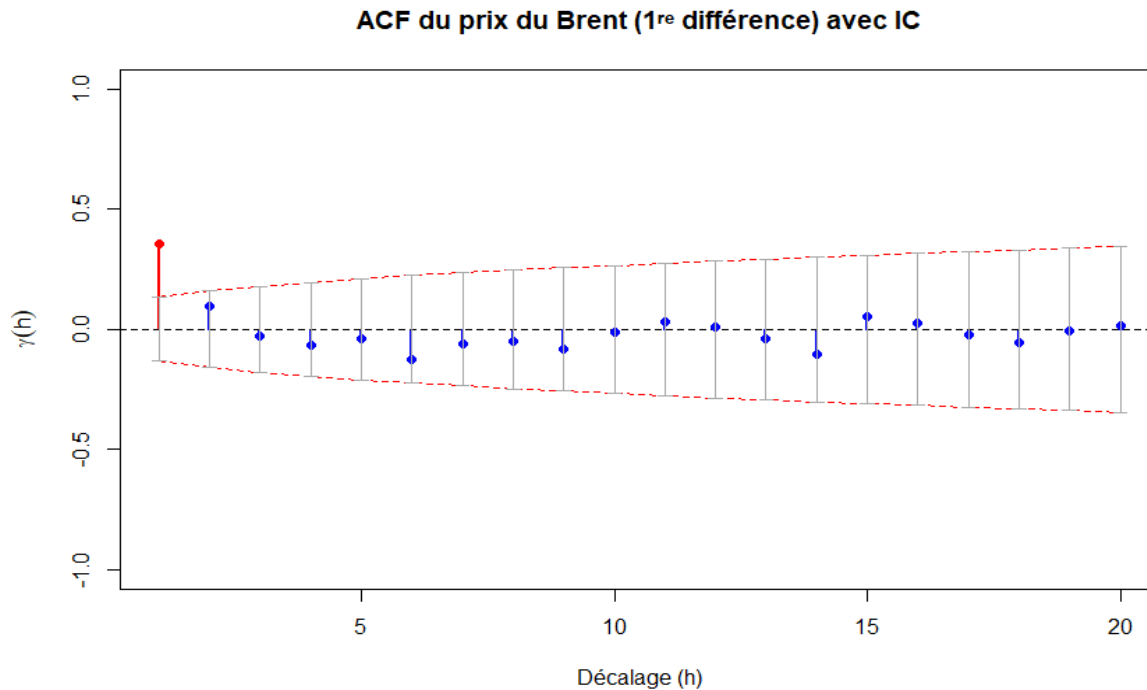




### Autocorrélation de la série sur la production de pétrole



## Autocorrélation de la série sur le prix du pétrole



Après avoir appliqué une première différentiation aux séries de production et de prix du Brent, nous avons calculé les fonctions d'autocorrélation (ACF) accompagnées de leurs intervalles de confiance à 95 %, afin d'évaluer si la dépendance temporelle persistait dans les séries transformées.

Série de production différenciée :

Le graphique ACF de la production différenciée montre que toutes les valeurs d'autocorrélation se situent à l'intérieur des intervalles de confiance, ce qui indique qu'elles ne sont pas statistiquement différentes de zéro. Cela suggère que la structure d'autocorrélation a été largement éliminée par la première différentiation, ce qui confirme la pertinence de cette transformation pour rendre la série proche de la stationnarité.

Série du prix différenciée :

En revanche, pour la série du prix du Brent, on observe que l'autocorrélation au lag 1 dépasse les bornes de l'intervalle de confiance, ce qui indique une autocorrélation significative à court terme. Les lags suivants sont tous inclus dans les bornes, ce qui montre que la dépendance disparaît rapidement. Bien que la série ait été partiellement stationnarisée par la première différentiation, une légère dépendance reste présente, notamment à court terme, ce qui pourrait nécessiter une modélisation ARMA pour capter cette dynamique résiduelle.

# **Modélisation et Prévision avec les Processus ARMA**

Dans cette seconde partie nous approfondirons l'analyse des séries sur le prix et la production de pétrole aux Etats-Unis.

## **Sélection des ordres des modèles ARMA**

Un processus ARMA (p, q) (Autorégressive à Moyenne Mobile) est un processus stochastique stationnaire qui combine une partie autorégressive (AR) d'ordre  $p$  et une partie moyenne mobile (MA) d'ordre  $q$ .

Il est défini par l'équation suivante :

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

Où :

- $X_t$  est la valeur de la série à l'instant  $t$ , -  $\phi_1, \dots, \phi_p$  sont les paramètres autorégressifs, -  $\theta_1, \dots, \theta_q$  sont les paramètres de la moyenne mobile,
- $\varepsilon_t$  est un bruit blanc (moyenne nulle, variance constante, non autocorrélé).

La stationnarité d'un processus ARMA dépend principalement de la partie AR : elle est assurée si les racines du polynôme autorégressif sont à l'extérieur du cercle unité dans le plan complexe. Elle capture la relation entre une observation et ses valeurs passées.

La partie moyenne mobile (MA) aide à lisser les fluctuations imprévues pour mieux comprendre la dynamique des données.

## **Analyse de l'ACF et du PACF**

L'ACF (fonction d'autocorrélation) mesure le degré de corrélation entre les valeurs d'une série temporelle à différentes périodes séparées par un décalage (lag).

Mathématiquement, l'autocorrélation au lag  $k$  est définie comme :

$$\rho_k = \frac{\text{Cov}(X_t, X_{t-k})}{\text{Var}(X_t)}$$

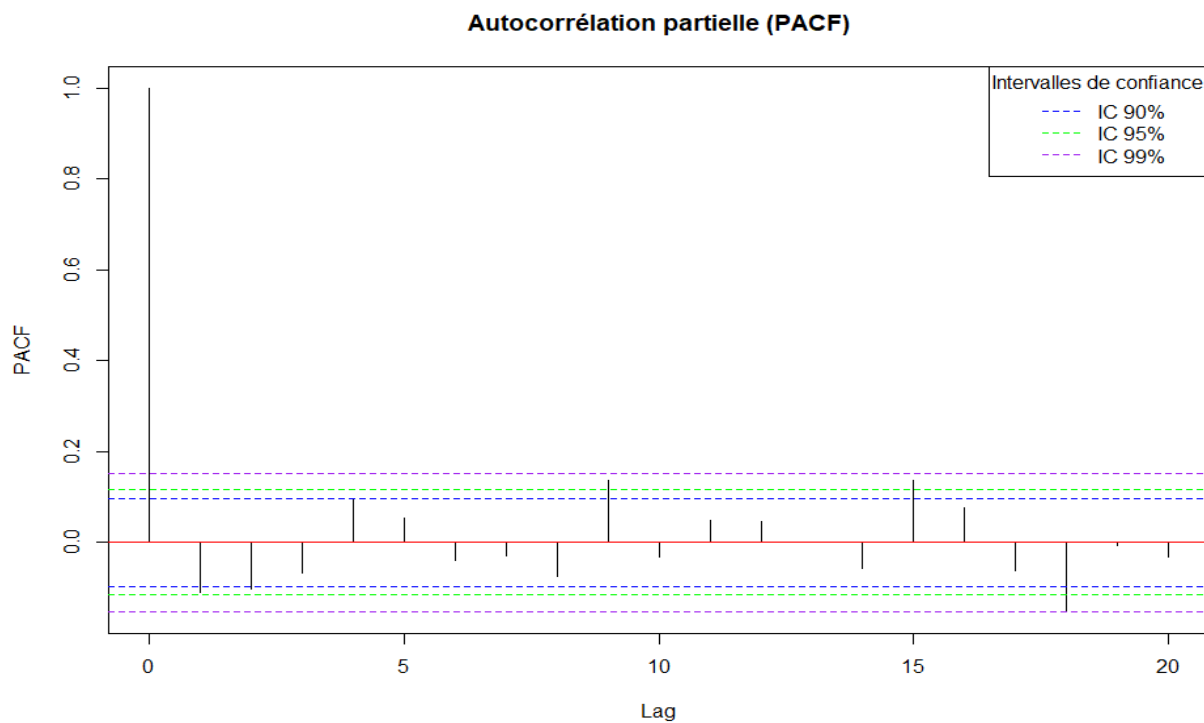
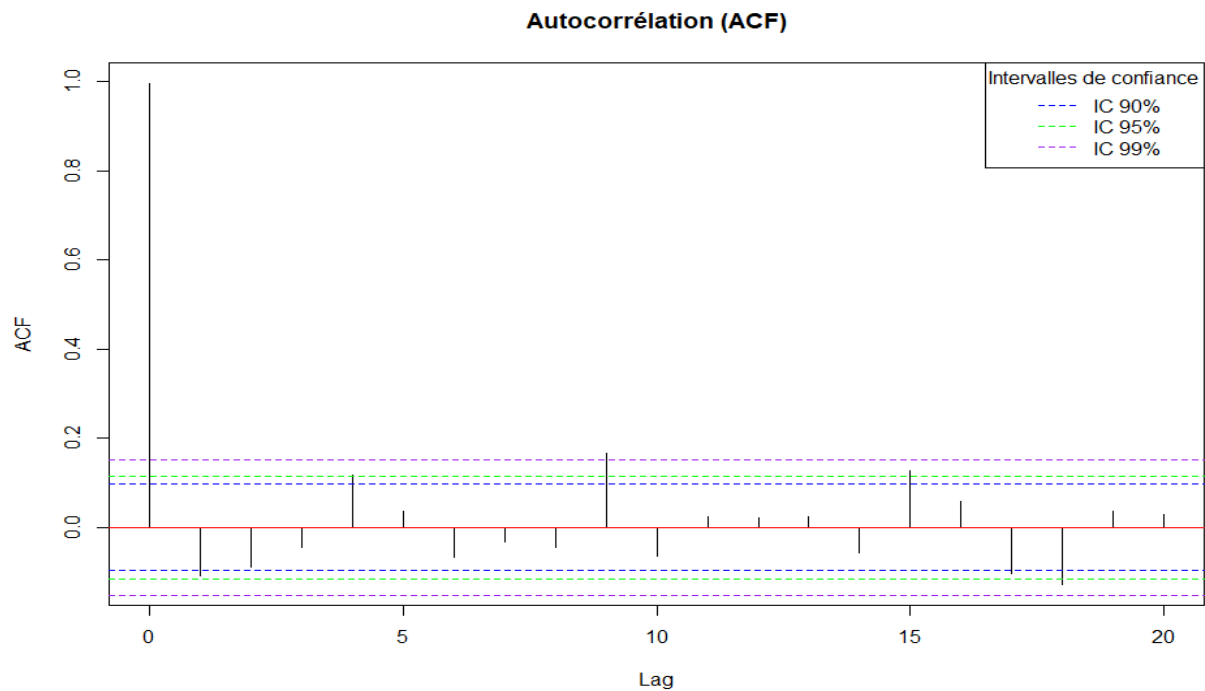
Où :

- $\rho_k$  est l'autocorrélation au décalage  $k$ ,
- $\text{Cov}(X_t, X_{t-k})$  est la covariance entre  $X_t$  et  $X_{t-k}$ ,
- $\text{Var}(X_t)$  est la variance de la série (supposée stationnaire).

L'ACF permet de détecter la présence de saisonnalité, de tendance ou d'autocorrélation structurelle dans les données.

La PACF (fonction d'autocorrélation partielle) mesure la corrélation entre  $X_t$  et  $X_{t-k}$ , en éliminant l'effet des lags intermédiaires  $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$ .

## Série de la production de pétrole



En examinant les fonctions ACF et PACF de la première différenciation de la production sous trois niveaux de confiance (90 %, 95 %, 99 %), nous pouvons identifier les lags pour lesquels les corrélations sont significativement différentes de zéro. Ces résultats nous aident à déterminer les ordres possibles d'un modèle ARMA.

### ACF

- À 90 %, seul le lag 1 dépasse légèrement l'intervalle de confiance.
- À 95 % et 99 %, aucun lag ne dépasse le seuil, sauf peut-être un très léger dépassement au lag 1, qui devient négligeable aux niveaux plus stricts.

La composante MA n'est probablement pas significative au-delà du lag 1, voire pas du tout. Cela suggère une structure très simple (MA (0) ou MA (1)).

### PACF

- À 90 %, le lag 1 est significatif.
- Aux seuils de 95 % et 99 %, le lag 1 reste significatif, tandis que tous les autres sont bien contenus dans les bandes de confiance.

**Conclusion PACF** : la série montre clairement une coupure après le lag 1, ce qui est typique d'un processus AR (1).

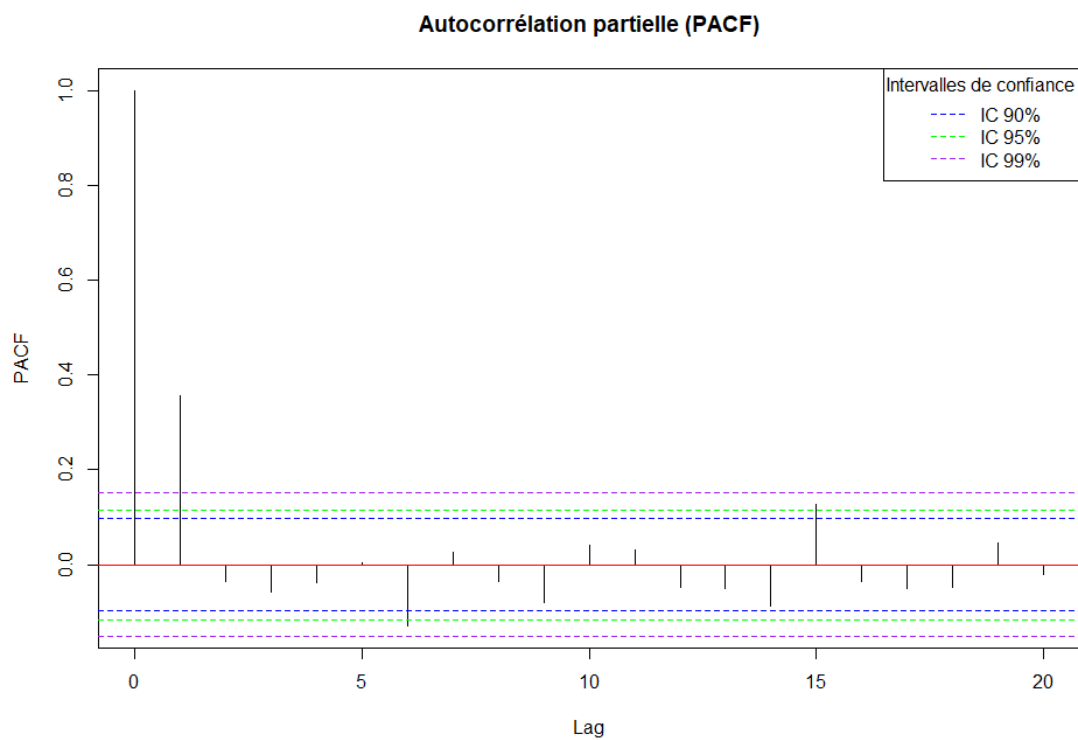
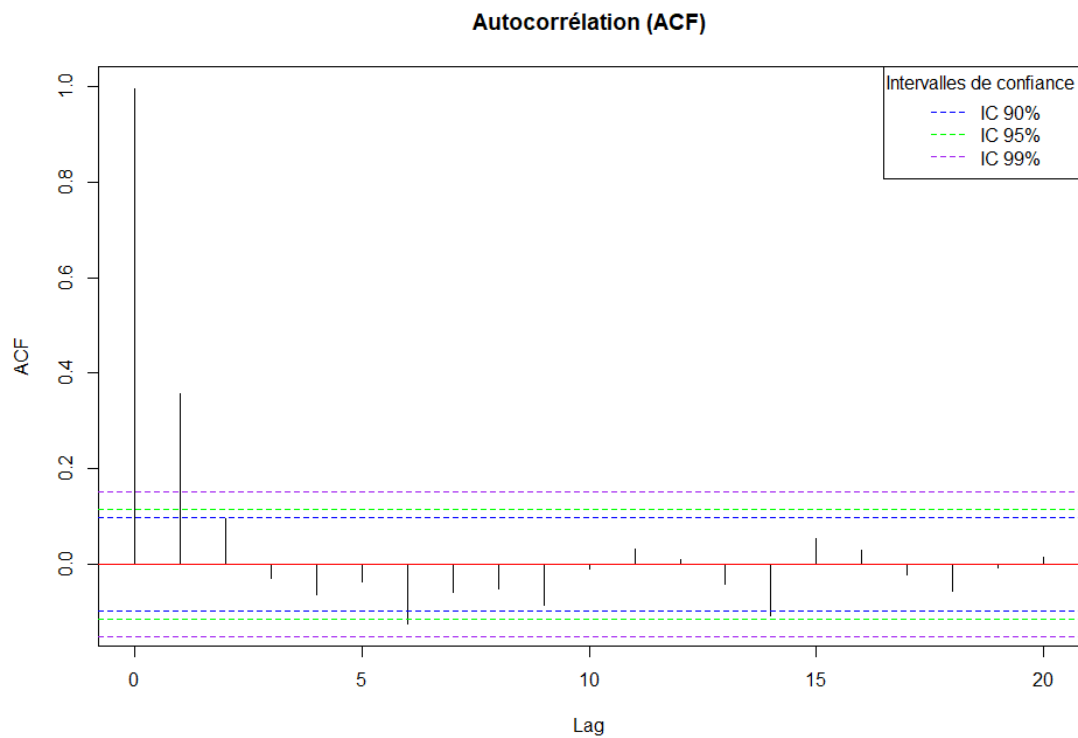
L'analyse des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) de la série de production différenciée suggère qu'un modèle ARMA (1,1) est approprié pour capturer sa dynamique temporelle. Ce modèle comprend :

Un terme autorégressif d'ordre 1 (AR (1)) indiquant une dépendance avec la valeur immédiatement précédente.

Un terme de moyenne mobile d'ordre 1 (MA (1)), qui capture les effets des chocs passés.

Ce choix est appuyé à la fois par l'analyse graphique et par les critères d'information AIC et BIC, qui favorisent ce modèle par rapport à des versions plus simples ou plus complexes. De plus, les résidus du modèle ne présentent pas d'autocorrélation significative, confirmant que l'ajustement est statistiquement valide.

## Série du prix du pétrole



L'analyse des fonctions ACF et PACF de la série du prix du pétrole (après différenciation) a été effectuée avec des intervalles de confiance à 90 %, 95 % et 99 %.

### **ACF**

- À 90 %, les lags 1, 2, 6, 14 apparaissent comme significatifs.
- À 95 %, seul le lag 1 reste significatif.
- À 99 %, le lag 1 est le seul à dépasser le seuil.

### **PACF**

- À 90 %, les lags 1, 6, 15 sont proches des limites, mais seul le lag 1 est clairement significatif à 95 % et 99 %.

Ces résultats indiquent une dynamique temporelle dominée par un composant autorégressif d'ordre 1. La chute rapide de l'ACF et la coupure nette de la PACF suggèrent un modèle AR (1) ajusté sur la série différenciée.

Ainsi, le modèle ARMA proposer pour la série du prix est : ARMA (1,1).

Ce modèle implique une différenciation d'ordre 1 pour stationnariser la série, suivie d'un processus AR (1), sans composante moyenne mobile.

## **Estimation de modèles ARMA**

À la suite de notre analyse des fonctions ACF et PACF, nous avons identifié des valeurs pertinentes pour les paramètres  $p$  et  $q$  de notre modèle ARMA. Nous avons ensuite divisé nos séries en un ensemble d'entraînement (80 %) et un ensemble de test (20 %) afin d'ajuster les modèles sur les données d'entraînement et d'évaluer leur performance prédictive.

### **Estimation par Maximum de Vraisemblance**

L'estimation par la méthode du maximum de vraisemblance consiste à déterminer les paramètres d'un modèle ARMA qui rendent les observations effectivement obtenues les plus probables, compte tenu du modèle choisi.

Autrement dit, on cherche à maximiser une fonction de vraisemblance qui exprime la probabilité des données observées en fonction des paramètres du modèle (coefficients AR et MA, ainsi que la variance de l'erreur). Cette méthode repose sur des hypothèses de normalité des erreurs et fournit des estimateurs efficaces lorsque les conditions sont respectées.

### *Série sur la production de pétrole*

Le modèle ARMA (1,1), estimé par la méthode du maximum de vraisemblance sur la série différenciée de la production, fournit les résultats suivants : un coefficient autorégressif AR (1) de  $\phi_1 = 0.1589$  (écart-type = 1.0807), et un coefficient de moyenne mobile MA (1) de  $\theta_1 = -0.1773$  (écart-type = 1.0785). Ces deux coefficients apparaissent non significatifs, ce qui pourrait indiquer une faible structure temporelle dans la série différenciée.

L'intercept est estimé à 26.62 (écart-type = 10.64) et la variance des résidus est de  $\sigma^2 \approx 27190$ , ce qui reste relativement élevé.

Concernant les critères d'information : - AIC = 3009.15 - BIC = 3022.91

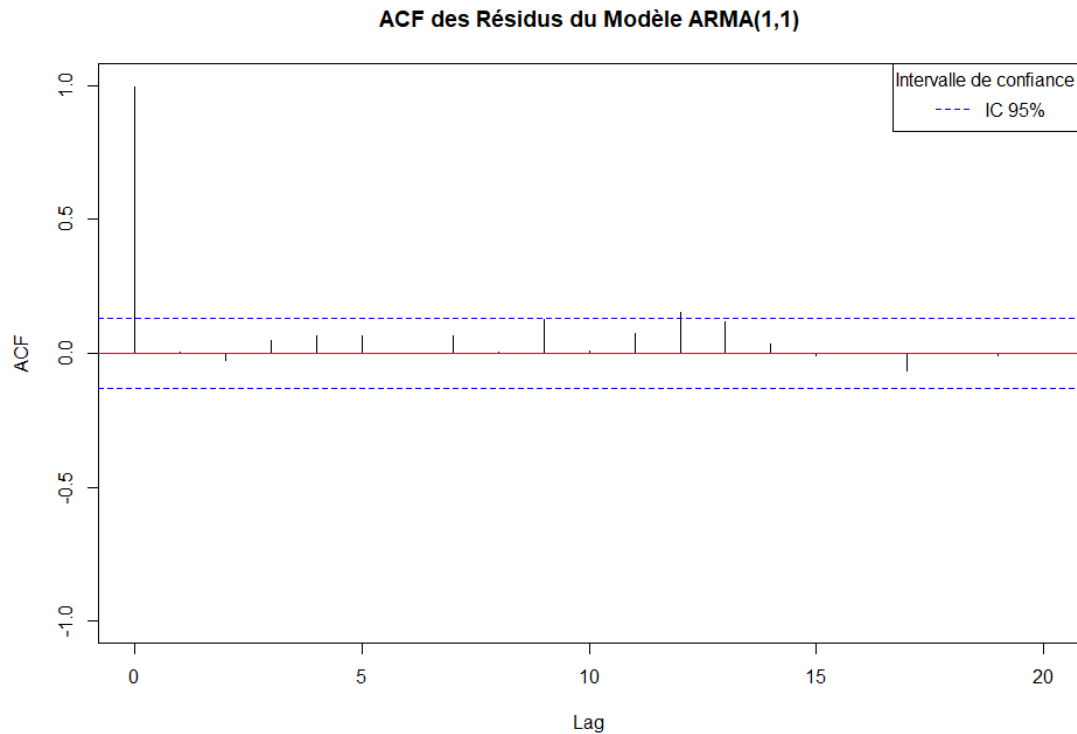
Ces résultats suggèrent que le modèle capte partiellement la dynamique de la série, mais que les coefficients estimés pourraient ne pas être essentiels. Une simplification du modèle, en supprimant la composante MA ou AR, pourrait être envisagée dans une étape ultérieure afin de vérifier si un modèle plus simple reste performant.

---

Avant de conclure définitivement, il est essentiel d'analyser les **résidus** du modèle. Si ceux-ci ne présentent ni autocorrélation significative ni structure, cela pourrait malgré tout valider l'ajustement, même si le paramètre est faible. Dans le cas contraire, l'exploration de modèles plus complexes ou alternatifs (ARMA avec autres ordres, modèles avec saisonnalité, etc.) pourrait s'avérer nécessaire.



## Analyse des résidus



L'analyse de l'ACF des résidus du modèle ARMA (1,1) appliqué à la série différenciée de la production montre que toutes les autocorrélations sont contenues à l'intérieur des intervalles de confiance, même au seuil le plus permissif de 90 %.

Cela indique une absence d'autocorrélation significative dans les résidus, ce qui est un critère essentiel pour valider la qualité d'ajustement du modèle. Nous pouvons donc conclure que le modèle capture efficacement la structure de la série, laissant des résidus compatibles avec un bruit blanc.

### *Test de Ljung-Box*

## Statistique Q : 17.86 | p-value : 0.5968

Le test de Ljung-Box appliqué aux résidus du modèle ARMA (1,1) pour la série différenciée de la production donne une statistique  $Q = 17.86$  avec une p-value de 0.5968.

Comme cette p-value est largement supérieure au seuil de 5 %, nous ne rejetons pas l'hypothèse nulle d'indépendance des résidus. Cela signifie que les résidus ne présentent pas d'autocorrélation significative, et peuvent donc être assimilés à un bruit blanc.

Ce résultat vient appuyer l'interprétation graphique de l'ACF et confirme que le modèle ARMA (1,1) est bien ajusté à la série de production différenciée.

### ***Test des Points Tournants***

## Turning Points : 149 | z-score : -0.471 | p-value : 0.6376

Le test des points de retournement appliqué aux résidus du modèle ARMA (1,1) a détecté 149 points de retournement, avec un z-score de -0.471 et une p-value de 0.6376.

Étant donné que la p-value est bien supérieure à 0.05, nous ne rejetons pas l'hypothèse nulle. Cela signifie que les résidus oscillent de manière aléatoire, sans tendance marquée ni structure particulière. Ce résultat suggère que les résidus se comportent comme un bruit blanc, ce qui est un critère fondamental pour valider la qualité du modèle.

En complément des tests d'ACF et de Ljung-Box, ce test confirme que le modèle ARMA (1,1) capte efficacement la dynamique de la série différenciée de la production.

### **Série sur le prix du pétrole**

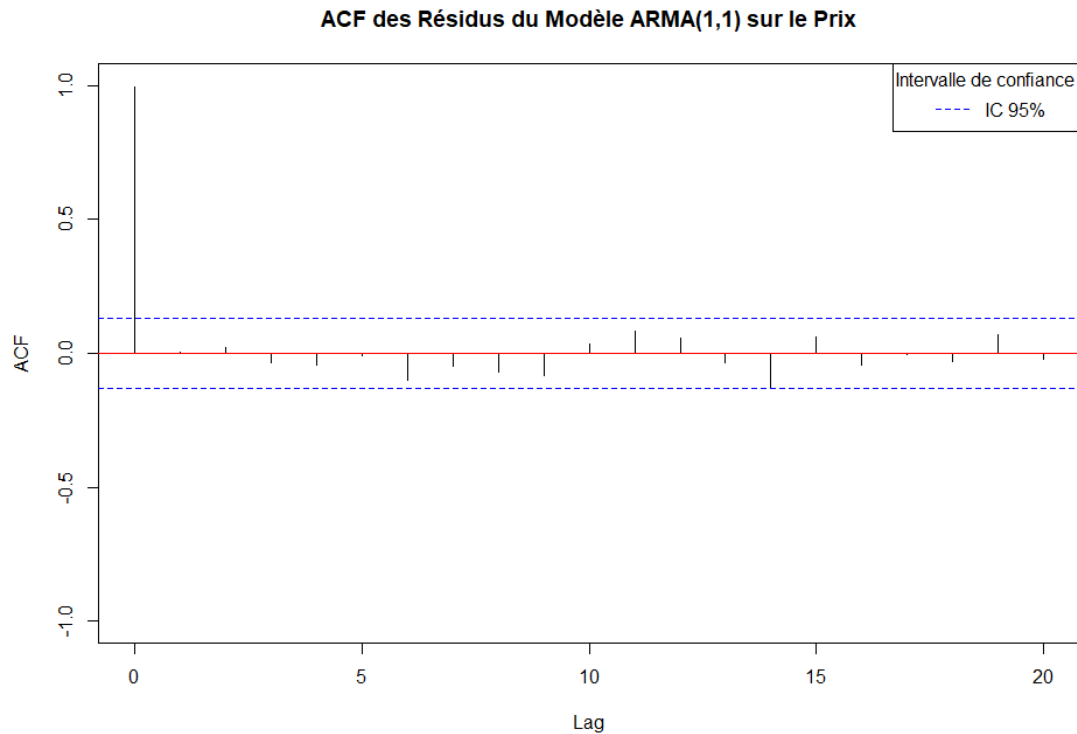
Le modèle ARMA (1,1), estimé par la méthode du maximum de vraisemblance sur la série différenciée du prix du pétrole, fournit les résultats suivants :

- Coefficient AR (1) :  $\phi_1 = 0.3687$  (écart-type = 0.1403)
- Coefficient MA (1) :  $\theta_1 = 0.0261$  (écart-type = 0.1478)
- Intercept :  $\mu = 0.1910$  (écart-type = 0.5455)
- Variance des résidus :  $\sigma^2 \approx 26.05$
- Log-vraisemblance : -701.32
- AIC : 1410.65
- BIC : 1424.40

Le coefficient AR (1) est positif et significatif, ce qui suggère la présence d'une dépendance temporelle dans la série différenciée. En revanche, le coefficient MA (1) est faible et non significatif, ce qui pourrait indiquer que la composante moyenne mobile n'apporte pas d'amélioration substantielle au modèle.

Les critères d'information (AIC et BIC) sont relativement bas, ce qui montre que le modèle parvient à capturer une partie de la dynamique de la série. Toutefois, il reste pertinent d'examiner si un modèle plus simple, comme un ARMA (1,0), pourrait fournir des performances similaires tout en réduisant la complexité.

## Analyse des résidus



L'analyse de l'ACF des résidus du modèle ARMA (1,1) appliqué à la série différenciée du prix du pétrole ne révèle **aucune autocorrélation significative** : toutes les valeurs se situent à l'intérieur des intervalles de confiance à 95 %.

Cela suggère que les dépendances temporelles ont été correctement captées par le modèle, et que les résidus peuvent être assimilés à un bruit blanc. Ce diagnostic renforce la validité du modèle ARMA (1,1) pour cette série.

### Test de Lyung-box

## Statistique Q : 15.38 | p-value : 0.7543

Le test de Ljung-Box appliqué aux résidus du modèle ARMA (1,1) sur la série différenciée du prix donne une statistique  $Q = 15.38$  avec une p-value de 0.7543.

Cette p-value est largement supérieure au seuil de 5 %, ce qui nous amène à ne pas rejeter l'hypothèse nulle d'indépendance des résidus. Cela signifie que les résidus ne présentent pas d'autocorrélation significative et peuvent être considérés comme un bruit blanc.

Ce résultat, en accord avec l'analyse visuelle de l'ACF, confirme que le modèle ARMA (1,1) est adéquatement ajusté à la série différenciée du prix.

### *Test des points tournants*

## Turning Points : 158 | z-score : 0.942 | p-value : 0.3462

Le test des points de retournement appliqué aux résidus du modèle ARMA (1,1) pour la série différenciée du prix a identifié 158 points de retournement, avec un z-score de 0.94 et une p-value de 0.34.

La p-value étant largement supérieure au seuil de 5 %, nous ne rejetons pas l'hypothèse nulle d'un comportement aléatoire. Cela signifie que les résidus oscillent sans structure particulière, ce qui est compatible avec un bruit blanc.

Ce test vient renforcer les diagnostics précédents (ACF et Ljung-Box) et confirme que le modèle ARMA (1,1) est approprié pour modéliser la dynamique de la série différenciée du prix.

## **Estimation des modèles imbriqués**

Dans cette section, nous cherchons à simplifier les modèles précédemment estimés en testant des modèles imbriqués, c'est-à-dire des versions plus simples contenant moins de paramètres. Cette démarche permet de vérifier si la réduction de la complexité du modèle entraîne une perte significative de performance, ou si un modèle plus parcimonieux reste adéquat.

### **Série de la production**

Nous commençons par la série différenciée de la production, pour laquelle nous avons initialement estimé un modèle ARMA (1,1). Nous allons maintenant estimer un modèle ARMA (1,0), qui constitue une version simplifiée sans composante de moyenne mobile (MA). Cette estimation nous permettra de comparer les deux modèles sur la base de critères d'information et de l'analyse des résidus.

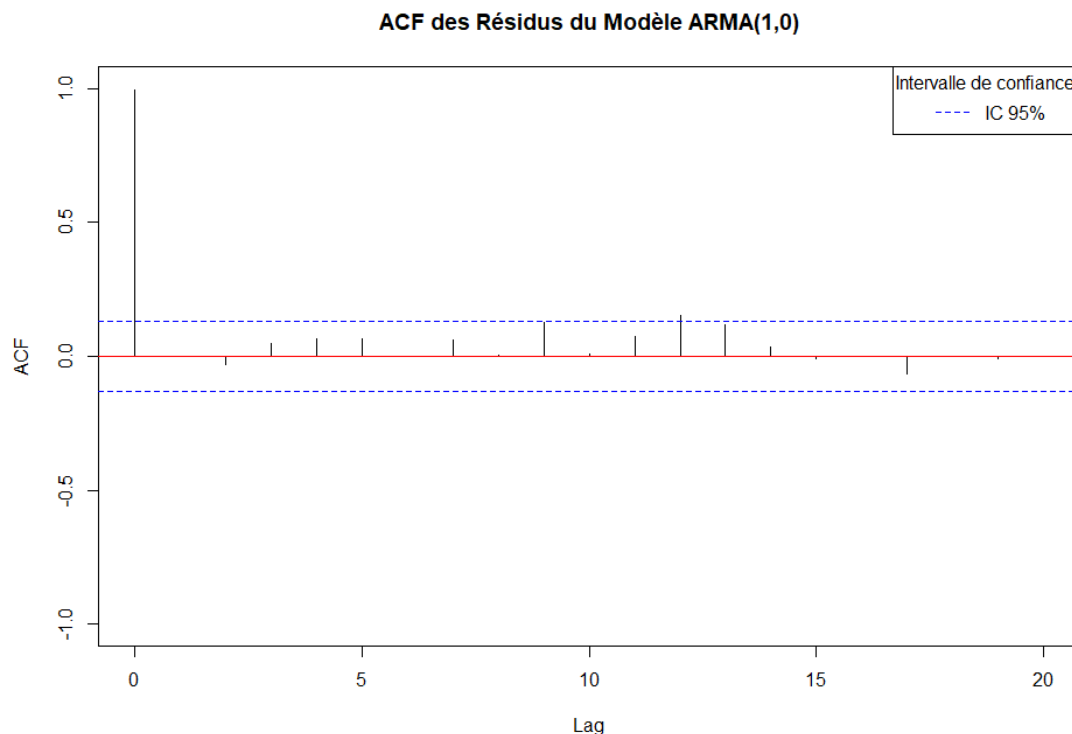
Le modèle ARMA (1,0), estimé par la méthode du maximum de vraisemblance sur la série différenciée de la production, constitue une version simplifiée du modèle ARMA (1,1) initial, en supprimant la composante moyenne mobile.

Les résultats de l'estimation montrent un coefficient AR (1) de  $\phi_1 = -0.0157$ , avec un écart-type de 0.0660. Ce coefficient est très proche de zéro et non significatif, ce qui suggère une faible dépendance autorégressive dans la série. L'intercept est estimé à 26.62 (écart-type = 10.71), et la variance des résidus reste élevée, à  $\sigma^2 \approx 27193$ , similaire à celle obtenue avec le modèle plus complexe.

En ce qui concerne les critères d'information : - **AIC = 3007.18 - BIC = 3017.49**

Comparé au modèle ARMA (1,1), ce modèle présente un **BIC plus faible**, ce qui suggère que la suppression du paramètre MA (1) **réduit la complexité sans perte significative de qualité**. Ce résultat plaide en faveur du modèle ARMA (1,0), plus parcimonieux, pour représenter la dynamique de la série de production différenciée.

## Analyse des résidus (modèle simplifié) sur la production de pétrole



L'analyse de l'ACF des résidus du modèle ARMA (1,0) appliqué à la série différenciée de la production montre que toutes les autocorrélations à partir du lag 1 se situent à l'intérieur de l'intervalle de confiance à 95 %.

Ce résultat indique qu'il n'y a pas de dépendance temporelle résiduelle significative, ce qui valide la qualité d'ajustement du modèle. Les résidus peuvent ainsi être considérés comme un bruit blanc, condition essentielle pour garantir que la dynamique de la série a été correctement capturée.

### Test de Ljung-Box

**## Statistique Q : 17.66 | p-value : 0.6099**

Le test de Ljung-Box appliqué aux résidus du modèle ARMA (1,0) donne une statistique  $Q = 17.66$  avec une p-value de 0.6099. Étant donné que cette p-value est largement supérieure au seuil de 5 %, nous **ne rejetons pas l'hypothèse nulle** d'indépendance des résidus.

Cela signifie que les résidus ne présentent **pas d'autocorrélation significative**, ce qui est cohérent avec l'analyse graphique de l'ACF. Nous pouvons donc conclure que le modèle ARMA (1,0) capte efficacement la structure de la série différenciée de la production, laissant des résidus qui se comportent comme un bruit blanc.

### *Test des points tournants*

## Turning Points : 149 | z-score : -0.471 | p-value : 0.6376

Le test des points de retournement appliqué aux résidus du modèle ARMA (1,0) pour la série différenciée de la production a identifié 149 points de retournement, avec un z-score de -0.471 et une p-value de 0.6376.

Cette p-value étant largement supérieure au seuil de 5 %, nous ne rejetons pas l'hypothèse nulle selon laquelle les résidus oscillent de manière aléatoire. Cela confirme que les résidus ne présentent pas de tendance ou de régularité structurelle, et se comportent comme un bruit blanc.

Ce résultat renforce les conclusions des analyses précédentes (ACF et Ljung-Box), validant ainsi la qualité d'ajustement du modèle ARMA (1,0).

La comparaison entre les modèles ARMA (1,1) et ARMA (1,0) appliqués à la série différenciée de la production montre que les deux modèles produisent des résidus compatibles avec un bruit blanc. Toutefois, les coefficients estimés dans le modèle ARMA (1,1) sont non significatifs, et le BIC y est **plus élevé** que dans le modèle ARMA (1,0) (3022.91 contre 3017.49).

Par conséquent, conformément au principe de parcimonie, nous retenons le modèle ARMA (1,0) comme le modèle le plus adéquat pour représenter la dynamique de la série de production.

### **Série sur le prix du pétrole**

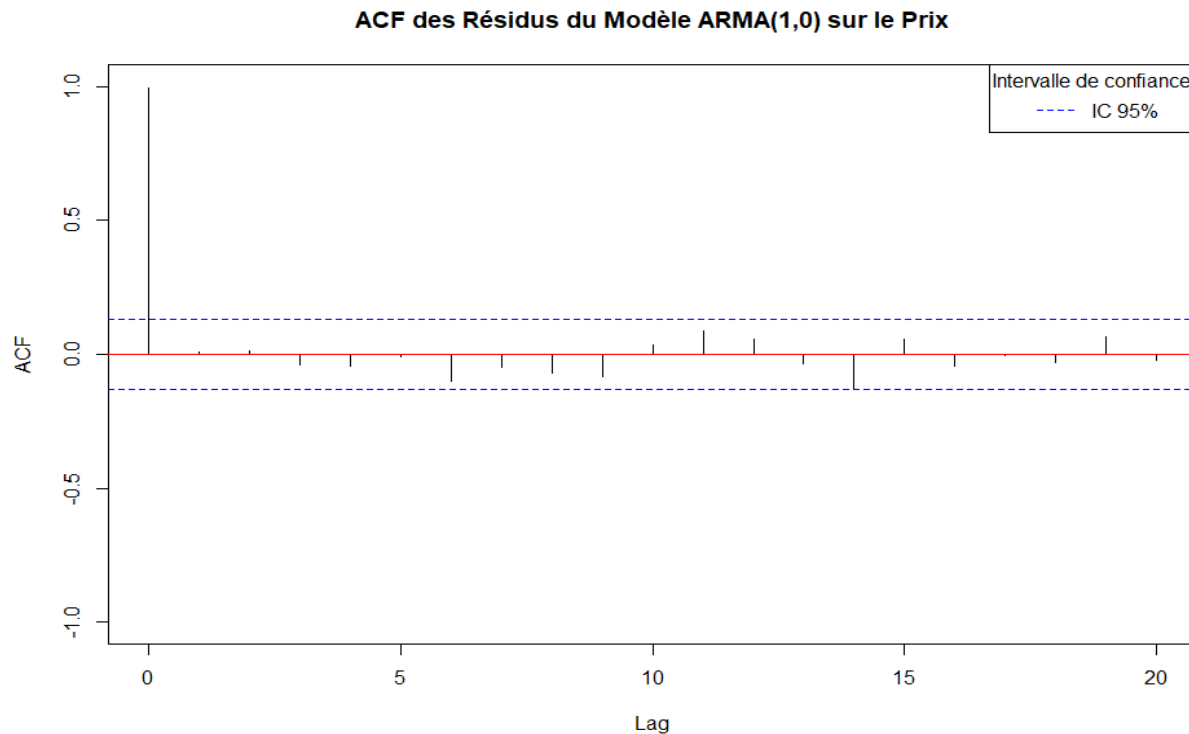
Comme pour la série de production, nous cherchons maintenant à simplifier le modèle ARMA (1,1) estimé sur la série différenciée du prix du pétrole. Pour ce faire, nous allons estimer un modèle imbriqué ARMA (1,0), en retirant la composante de moyenne mobile.

Cela nous permettra de comparer les performances des deux modèles à l'aide des critères d'information et des diagnostics sur les résidus, afin de déterminer si un modèle plus simple peut offrir une qualité d'ajustement équivalente.

La comparaison entre les modèles ARMA (1,1) et ARMA (1,0) appliqués à la série différenciée du prix du pétrole montre que les deux modèles présentent une qualité d'ajustement très similaire. Les coefficients AR (1) sont proches dans les deux cas et significatifs, mais le coefficient MA (1) du modèle ARMA (1,1) est faible et non significatif.

Le critère BIC est plus faible pour le modèle ARMA (1,0) (1418.99 contre 1424.40), ce qui indique que ce modèle plus simple est préférable selon le principe de parcimonie. Nous retenons donc l'ARMA (1,0) comme le modèle le plus adéquat pour modéliser la dynamique de la série différenciée du prix.

## Analyse des résidus (Modèle simplifié) sur le prix du pétrole



L'analyse de l'ACF des résidus du modèle ARMA (1,0) appliqué à la série différenciée du prix du pétrole révèle qu'aucune autocorrélation significative n'est présente : toutes les valeurs sont situées à l'intérieur de l'intervalle de confiance à 95 %.

Cela indique que les résidus du modèle se comportent comme un bruit blanc, sans structure temporelle résiduelle. Ce diagnostic confirme que le modèle ARMA (1,0) est adéquatement **ajusté** à la série, tout en étant plus simple que le modèle ARMA (1,1).

### *Test de Ljung-Box*

```
## Statistique Q : 15.44 | p-value : 0.7504
```

Le test de Ljung-Box appliqué aux résidus du modèle ARMA (1,0) pour la série différenciée du prix donne une statistique  $Q = 15.44$  avec une **p-value de 0.7504**.

Cette p-value étant largement supérieure au seuil de 5 %, nous ne rejetons pas l'hypothèse nulle d'indépendance des résidus. Cela signifie qu'il n'existe pas d'autocorrélation significative dans les résidus, ce qui confirme que le modèle est correctement spécifié.

Ces résultats sont en accord avec l'analyse graphique de l'ACF et renforcent la conclusion selon laquelle le modèle ARMA (1,0) est adéquat et parcimonieux pour modéliser la dynamique de la série différenciée du prix.

### ***Test des points tournants***

## Turning Points : 156 | z-score : 0.628 | p-value : 0.53

Le test des points de retournement appliqué aux résidus du modèle ARMA (1,0) pour la série différenciée du prix a identifié 156 points de retournement, avec un z-score de 0.628 et une p-value de 0.53.

Étant donné que la p-value est bien supérieure au seuil de 5 %, nous ne rejetons pas l'hypothèse nulle selon laquelle les résidus évoluent de manière aléatoire. Cela indique qu'aucune tendance ou structure régulière n'est présente dans les fluctuations des résidus.

Ces résultats sont cohérents avec ceux obtenus à partir de l'ACF et du test de Ljung-Box, confirmant que les résidus du modèle ARMA (1,0) pour le prix peuvent être considérés comme un bruit blanc. Le modèle est donc bien ajusté et statistiquement valide.

Pour la série différenciée du prix du pétrole, nous avons comparé deux modèles : ARMA (1,1) et ARMA (1,0). Le modèle ARMA (1,1) incluait une composante de moyenne mobile, mais celle-ci s'est révélée non significative (coefficient MA proche de zéro avec un écart-type élevé).

Le modèle ARMA (1,0), plus simple, présente un coefficient AR (1) significatif et une valeur de BIC plus faible (1418.99 contre 1424.40 pour ARMA (1,1)), ce qui indique une meilleure performance en termes de compromis entre ajustement et complexité.

Les diagnostics sur les résidus (ACF, test de Ljung-Box, points de retournement) confirment que le modèle ARMA (1,0) capture adéquatement la structure de la série, laissant des résidus qui se comportent comme un bruit blanc.

Nous retenons donc le modèle ARMA (1,0) comme le modèle le plus approprié pour la série différenciée du prix, en raison de sa simplicité, de sa pertinence statistique et de sa validité des hypothèses résiduelles.

## **Analyse des Paramètres et Modèles Imbriqués**

### **Série sur la production de pétrole**

#### *Paramètres estimés – Série de la production*

Modèle	Paramètre	Estimation	Écart-type
ARMA (1,1)	AR (1)	0.1589	1.0807
ARMA (1,1)	MA (1)	-0.1773	1.0785
ARMA (1,1)	Intercept	26.6181	10.6365
ARMA (1,0)	AR (1)	-0.0157	0.0660
ARMA (1,0)	Intercept	26.6196	10.7066



Pour la série différenciée de la production, la comparaison des paramètres estimés met en évidence que les coefficients du modèle ARMA (1,1) sont non significatifs, avec des écarts-types très élevés. Le coefficient AR (1) est estimé à 0.1589 avec un écart-type de 1.08, et le MA (1) est de -0.1773 avec un écart-type similaire, ce qui indique une grande incertitude dans l'estimation.

En revanche, le modèle ARMA (1,0), plus simple, présente un coefficient AR (1) très faible (-0.0157) et également non significatif. L'intercept est similaire dans les deux modèles.

Sur le plan des critères d'information, le BIC du modèle ARMA (1,0) est inférieur à celui de l'ARMA (1,1) (3017.49 contre 3022.91), ce qui appuie l'idée que la composante MA (1) n'apporte aucune amélioration significative.

En somme, les deux modèles donnent des résultats comparables, mais le modèle ARMA (1,0) est préférable en raison de sa simplicité, de sa stabilité et de la non-significativité de la composante MA (1) dans le modèle plus complexe.

## Série sur le prix du pétrole

### *Paramètres estimés – Série du prix*

Modèle	Paramètre	Estimation	Écart-type
ARMA (1,1)	AR (1)	0.3687	0.1403
ARMA (1,1)	MA (1)	0.0261	0.1478
ARMA (1,1)	Intercept	0.1910	0.5455
ARMA (1,0)	AR (1)	0.3907	0.0605
ARMA (1,0)	Intercept	0.1917	0.5509

Dans le cas de la série différenciée du prix du pétrole, les deux modèles estiment un coefficient AR (1) positif et significatif. Toutefois, la composante MA (1) du modèle ARMA (1,1) est très faible (0.0261) et non significative (écart-type = 0.1478), ce qui remet en question sa pertinence.

Le modèle ARMA (1,0), plus simple, conserve un coefficient AR (1) significatif (0.3907 avec un écart-type de 0.0605) et un intercept pratiquement identique. Cela montre qu'il capte efficacement la dépendance temporelle présente dans la série.

Du point de vue des critères d'information, le modèle ARMA (1,0) affiche un BIC inférieur (1418.99 contre 1424.40), indiquant une meilleure performance globale.

Ces résultats suggèrent que le modèle ARMA (1,0) est le plus approprié pour représenter la série du prix, car il est plus parcimonieux tout en offrant un ajustement statistiquement robuste et interprétable.

# Tests Diagnostics

## Critères d'Information (AIC/BIC)

*Comparaison des AIC et BIC sur l'ensemble d'entraînement et de test*

Modèle	AIC_Train	BIC_Train	AIC_Test	BIC_Test
Production ARMA (1,1)	3009.15	3022.91	888.49	894.67
Production ARMA (1,0)	3007.18	3017.49	886.49	890.61
Prix ARMA (1,1)	1410.65	1424.40	388.03	394.22
Prix ARMA (1,0)	1408.68	1419.00	386.04	390.16

## Série de la production de pétrole

Le modèle ARMA (1,0) présente des valeurs légèrement plus faibles que l'ARMA (1,1) pour l'ensemble des critères : - AIC test : 886.49 (vs 888.49) - BIC test : 890.61 (vs 894.67)

Cela confirme que le modèle ARMA (1,0), bien que plus simple, généralise mieux sur les données de test, tout en maintenant une performance équivalente à l'entraînement. Sa parcimonie le rend donc préférable.

## Série du prix du pétrole

De même, le modèle ARMA (1,0) obtient des meilleurs résultats que l'ARMA (1,1), aussi bien sur l'entraînement que sur le test : - AIC test : 386.04 (vs 388.03) - BIC test : 390.16 (vs 394.22)

Ces résultats montrent que la composante MA (1) du modèle ARMA (1,1) n'apporte aucun gain substantiel, alors qu'elle augmente la complexité. Le modèle ARMA (1,0) s'impose donc une fois de plus comme le meilleur choix.

Dans les deux séries, les modèles ARMA (1,0) offrent de meilleures performances de généralisation, associées à une structure plus simple. Leurs résidus étant également conformes aux hypothèses (bruit blanc), ces modèles sont retenus pour la suite de l'analyse.

## Test du ratio de vraisemblance

Afin de compléter la comparaison entre les modèles estimés, nous appliquerons le test du ratio de vraisemblance. Ce test permet de déterminer si l'ajout de paramètres dans un modèle plus complexe améliore significativement l'ajustement par rapport à un modèle imbriqué plus simple.

Le **test du ratio de vraisemblance** consiste à comparer deux modèles emboîtés un modèle restreint (plus simple) et un modèle complet (plus complexe) en utilisant leurs log-vraisemblances respectives. La statistique de test est définie par :

$$D = -2 \times (\ell_{\text{restreint}} - \ell_{\text{complet}})$$

où  $\ell$  représente la log-vraisemblance du modèle. Sous l'hypothèse nulle (selon laquelle le modèle simple est suffisant), cette statistique suit une **loi du  $\chi^2$**  avec un nombre de degrés de liberté égal au nombre de paramètres supplémentaires dans le modèle complet.

Une p-value faible indique que le modèle plus complexe améliore significativement l'ajustement, tandis qu'une p-value élevée suggère que le modèle plus simple est suffisant.

#### *Test du ratio de vraisemblance entre modèles imbriqués*

Série	Modèle_1	Modèle_2	Statistique	p_valeur
Production	ARMA (1,0)	ARMA (1,1)	0.0247	0.8751
Prix	ARMA (1,0)	ARMA (1,1)	0.0319	0.8582

Le test du ratio de vraisemblance a été appliqué pour comparer les modèles ARMA (1,0) et ARMA (1,1), ces derniers étant des modèles emboîtés. Les résultats sont les suivants :

- Pour la **production**, la statistique du test est de 0.0247 avec une p-value de 0.8751.
- Pour le **prix**, la statistique est de 0.0319 avec une p-value de 0.8582.

Dans les deux cas, les p-values sont largement supérieures au seuil de 5 %, ce qui implique que nous ne rejetons pas l'hypothèse nulle. Autrement dit, les modèles plus complexes ARMA (1,1) n'améliorent pas significativement l'ajustement par rapport aux modèles imbriqués ARMA (1,0).

Ces résultats confirment donc que les modèles ARMA (1,0) sont suffisamment performants, tout en étant plus simples, et doivent être retenus pour la modélisation finale des deux séries.

#### **Erreurs Quadratiques Moyennes (EQM) :**

L'erreur quadratique moyenne (EQM) est une mesure couramment utilisée pour évaluer la performance prédictive d'un modèle. Elle quantifie l'écart entre les valeurs observées et les valeurs estimées (ou prédites), en donnant davantage de poids aux erreurs importantes. Elle est définie par la formule :

$$EQM = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

où  $y_t$  est la valeur réelle au temps  $t$ , et  $\hat{y}_t$  est la valeur estimée ou prédite par le modèle.

Nous utiliserons l'EQM pour évaluer la performance des modèles ARMA retenus (ARMA (1,0) pour la production et pour le prix), à deux niveaux :

- Sur l'ensemble d'entraînement, à l'aide des valeurs ajustées par les modèles.
- Sur l'ensemble de test, à l'aide des valeurs prédites à l'avance.

Cela nous permettra de juger à la fois la qualité de l'ajustement et la capacité de généralisation des modèles.

```
##
## Analyse comparative des performances prédictives (Erreur Quadratique Moyenne - EQM)

## =====
##
## Série : Production

##      Série      Modèle MSE_Train MSE_Test
## Production ARMA(1,1)  27190.32 237352.6
## Production ARMA(1,0)  27193.25 237354.7

##
## Série : Prix

## Série      Modèle MSE_Train MSE_Test
## Prix ARMA (1,1) 26.04635 42.46872
## Prix ARMA (1,0) 26.05000 42.47220
```

Pour la série différenciée de la production, les deux modèles ARMA (1,1) et ARMA (1,0) présentent des performances quasi identiques :

- **EQM sur l'entraînement** : 27190.32 pour ARMA (1,1) et 27193.25 pour ARMA (1,0)
- **EQM sur le test** : 237352.6 pour ARMA (1,1) et 237354.7 pour ARMA (1,0)

Les différentiations sont infimes, ce qui indique que la composante MA (1) du modèle ARMA (1,1) n'améliore pas significativement la qualité de l'ajustement ni la capacité de prévision. Cela confirme les résultats du test du ratio de vraisemblance, et renforce l'idée que le modèle ARMA (1,0), plus simple, est à privilégier.

Pour la série différenciée du prix, la situation est très similaire. Les EQM sont pratiquement égales :

- **EQM sur l'entraînement** : 26.046 pour ARMA (1,1) et 26.050 pour ARMA (1,0)
- **EQM sur le test** : 42.4687 pour ARMA (1,1) et 42.4722 pour ARMA (1,0)

Ces valeurs confirment que les deux modèles offrent des prévisions équivalentes, avec une légère préférence pour ARMA (1,1) sur l'entraînement, mais une meilleure généralisation (légèrement plus basse) pour ARMA (1,0) sur les données test. Combiné au test du ratio de vraisemblance, cela justifie le choix du modèle ARMA (1,0), plus parcimonieux et tout aussi performant.

## **Recommandations sur les modèles ARMA**

### **Série sur la production de pétrole**

Après une série d'analyses comprenant l'étude de la stationnarité, l'identification des ordres à l'aide de l'ACF et de la PACF, l'estimation des paramètres par la méthode du maximum de vraisemblance, ainsi que la validation par les tests sur les résidus, nous recommandons le modèle ARMA (1,0) pour la série différenciée de la production.

Ce modèle, bien que simple, présente des performances très similaires à celles du modèle ARMA (1,1), tant en termes de qualité d'ajustement (EQM) que de capacité prédictive. De plus, les résultats du test du ratio de vraisemblance confirment que l'ajout de la composante MA (1) dans le modèle ARMA (1,1) n'apporte aucun gain significatif.

En résumé, le modèle ARMA (1,0) s'avère plus parcimonieux, plus interprétable et tout aussi efficace, ce qui en fait le choix recommandé pour modéliser la dynamique de la production pétrolière.

### **Série sur le prix du pétrole**

Pour la série différenciée du prix du pétrole, les deux modèles ARMA (1,1) et ARMA (1,0) ont offert des résultats quasi identiques, tant sur l'ensemble d'entraînement que sur l'ensemble de test, avec des erreurs quadratiques moyennes (EQM) très proches.

Le test du ratio de vraisemblance n'a révélé aucune amélioration significative entre le modèle plus complexe ARMA (1,1) et le modèle plus simple ARMA (1,0). En conséquence, et conformément au principe de parcimonie, le modèle ARMA (1,0) est également recommandé pour cette série.

Ce modèle est suffisant pour capturer la dépendance temporelle de la série du prix, tout en offrant une meilleure simplicité d'interprétation et une efficacité comparable.

## **Modélisation Bivariée et Corrélations Croisées**

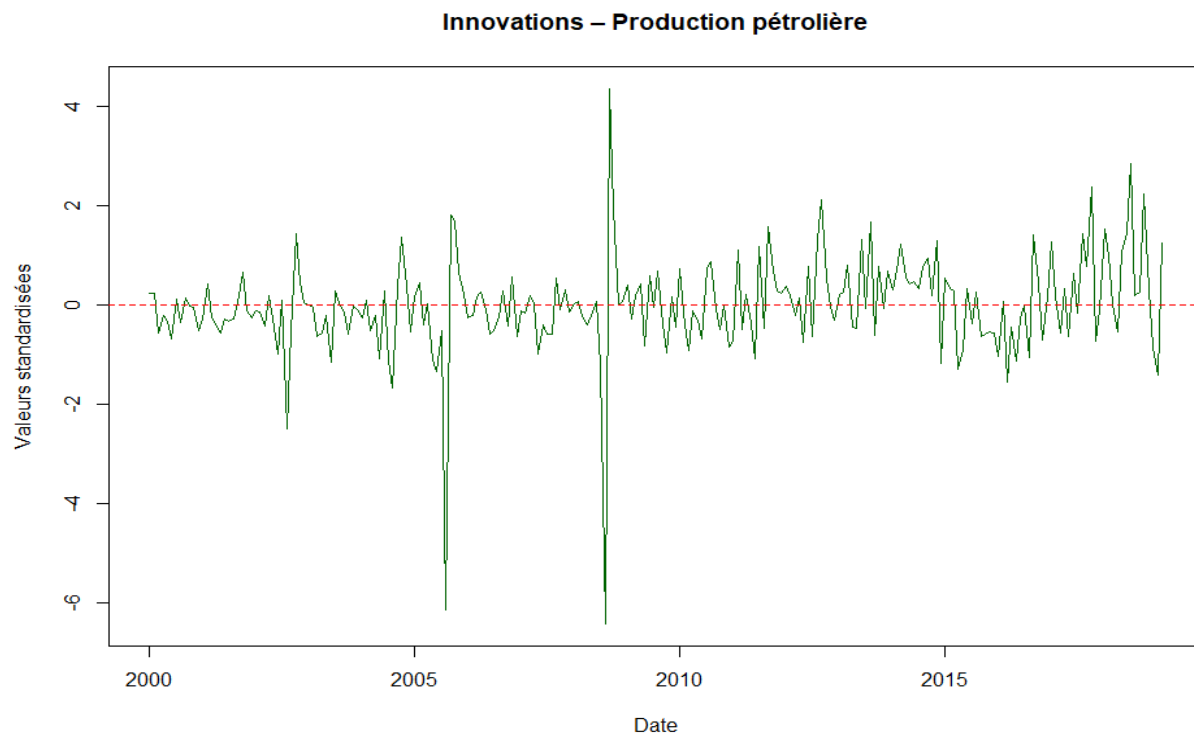
Dans cette section, nous explorons l'utilisation de modèles bivariés afin d'analyser les relations dynamiques entre la production pétrolière et le prix du pétrole. Plus précisément, nous allons approximer les séries d'innovations issues de chacun des modèles ARMA retenus précédemment, en appliquant l'algorithme des innovations tel que suggéré dans l'exercice.

Cette approche permet d'extraire les erreurs de prévision (ou résidus) produites par les modèles univariés, afin d'en étudier la structure. L'objectif est de mieux comprendre comment ces deux variables interagissent dans le temps, en observant leurs innovations respectives.

L'algorithme des innovations est une méthode efficace pour révéler des corrélations croisées potentielles entre les séries, et il nous fournira un éclairage complémentaire sur la nature du lien entre les niveaux de production et les fluctuations de prix du pétrole.

## **Analyse des innovations**

### **Série sur la production de pétrole**

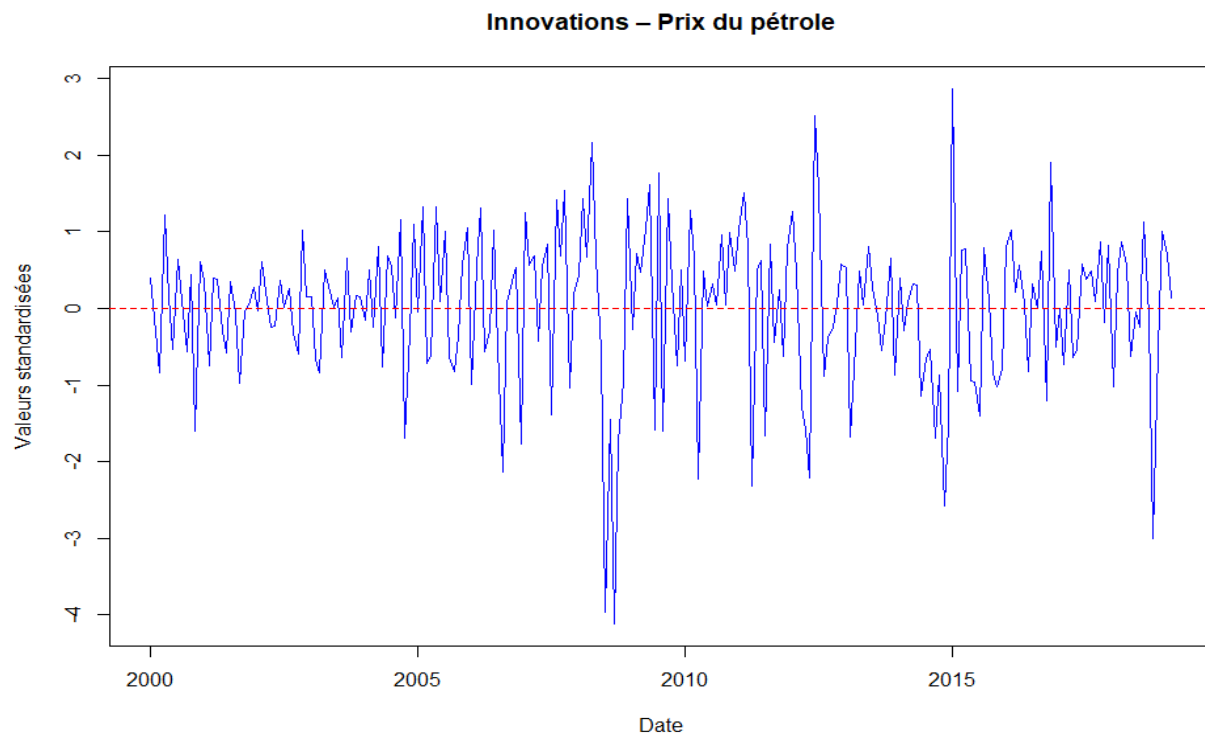


Les innovations standardisées de la production pétrolière, obtenues à partir du modèle ARMA (1,0), présentent un comportement globalement stationnaire, avec une variabilité modérée autour de la moyenne nulle. On observe toutefois quelques pics marqués, notamment entre 2005 et 2010, où l'amplitude des innovations dépasse largement les valeurs usuelles, atteignant presque -6 ou +4.

Ces valeurs extrêmes pourraient correspondre à des événements exogènes majeurs ayant perturbé de façon soudaine la production, mais qui n'ont pas laissé de trace persistante dans le temps. En

dehors de ces anomalies ponctuelles, la série semble rester bien capturée par le modèle retenu, avec des fluctuations courtes et aléatoires, caractéristiques d'un processus bruit blanc.

## Série sur le prix du pétrole



Les innovations standardisées du prix du pétrole affichent une volatilité marquée, notamment entre les années 2008 et 2015, période durant laquelle plusieurs pics atteignent ou dépassent les  $\pm 3$  écarts-types. Ces valeurs extrêmes correspondent vraisemblablement à des événements exogènes majeurs (crises économiques, instabilités géopolitiques, variations de l'offre mondiale).

On note que la série reste centrée autour de zéro, ce qui valide l'hypothèse d'un résidu stationnaire. La dispersion est toutefois plus élevée que pour la série de production, ce qui laisse entendre que le prix du pétrole est plus sensible aux chocs aléatoires.

Globalement, le modèle ARMA (1,0) semble capter la structure sous-jacente, les écarts représentant surtout des perturbations imprévisibles, confirmant ainsi que ces innovations peuvent être traitées comme bruit blanc dans l'analyse conjointe.

## Propriétés statistiques des innovations

Afin de valider nos modèles univariés et d'évaluer la qualité des ajustements, nous allons maintenant analyser les propriétés statistiques des séries d'innovations obtenues pour la production

et le prix du pétrole. Cette étape est essentielle pour vérifier que les innovations se comportent comme un bruit blanc, c'est-à-dire qu'elles sont non autocorrélées, centrées et stationnaires.

### **Analyse des propriétés statistiques des innovations**

Les statistiques descriptives des séries d'innovations standardisées montrent que :

- Pour la production, la moyenne est exactement nulle, avec une variance de 1.0044 et un écart-type de 1.0022 ;
- Pour le prix du pétrole, la moyenne est très proche de zéro ( $-0.0008$ ), avec une variance de 1.0044 et un écart-type similaire de 1.0022.

Ces résultats suggèrent que les innovations sont centrées et homogènes, comme attendu pour des résidus normalisés issus de modèles ARMA bien ajustés.

En ce qui concerne l'autocorrélation des innovations, les tests de Ljung-Box appliqués à plusieurs retards confirment que les innovations peuvent être considérées comme non autocorrélées :

- Pour la production, toutes les p-values (ex.  $p = 0.75$  à lag 5,  $p = 0.60$  à lag 20) sont largement supérieures à 5%, ce qui rejette l'hypothèse d'autocorrélation.
- Pour le prix, les résultats sont encore plus nets, avec des p-values allant jusqu'à 0.97, confirmant le caractère aléatoire des innovations.

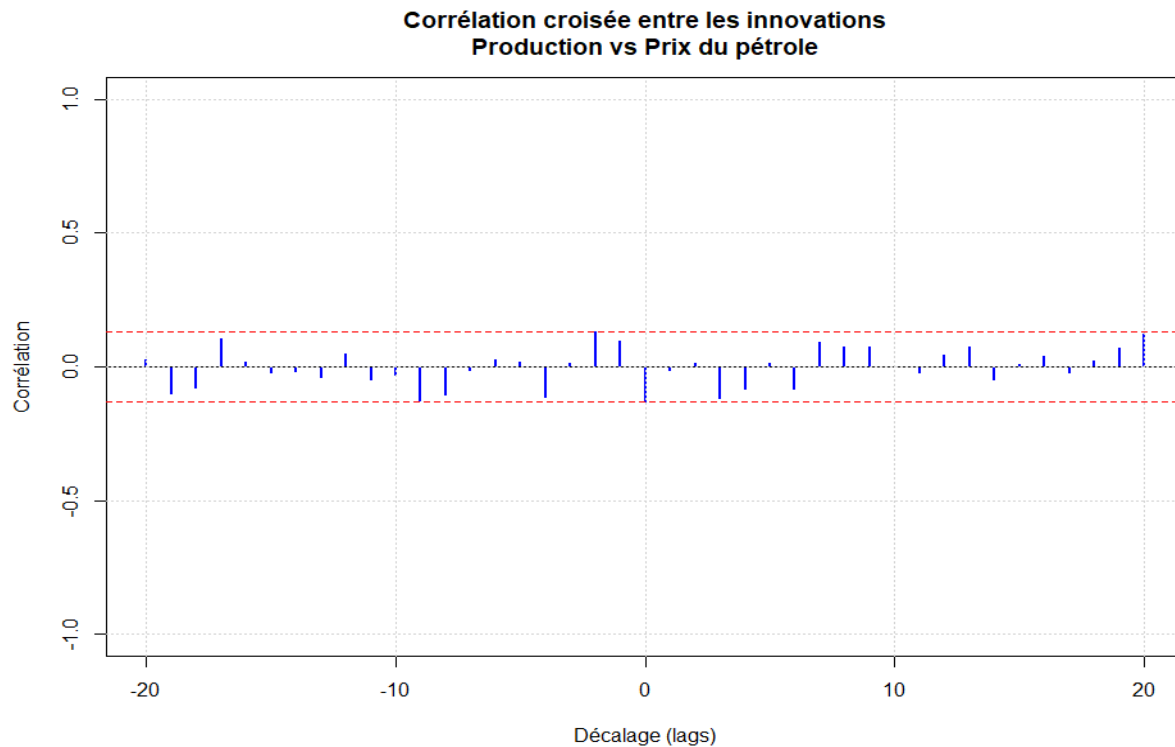
En résumé, les deux séries d'innovations présentent les propriétés attendues d'un bruit blanc : elles sont centrées, de variance constante et sans autocorrélation significative. Ces résultats valident les modèles ARMA sélectionnés pour la suite de l'analyse bivariée.

### **Corrélation croisée entre les innovations**

Afin d'approfondir notre analyse bivariée, nous allons maintenant étudier la fonction de corrélation croisée (CCF) entre les séries d'innovations issues des modèles ARMA (1,0) estimés pour la production et le prix du pétrole.

Cette analyse permet de détecter d'éventuelles relations dynamiques décalées dans le temps entre les deux variables. En observant les lags pour lesquels la corrélation croisée est significative, nous pourrions déterminer si les chocs dans l'une des séries ont un effet sur l'autre avec un certain retard.





L'analyse de la fonction de corrélation croisée (CCF) entre les innovations des modèles ARMA (1,0) de la production et du prix du pétrole indique une absence de corrélations croisées significatives pour la grande majorité des décalages.

- La majorité des valeurs de corrélation restent à l'intérieur des intervalles de confiance, ce qui appuie l'idée que les modèles univariés capturent correctement la structure interne de chaque série.
- Aucune valeur significative n'émerge de manière claire ou persistante, ni pour les lags positifs (influence du prix passé sur la production) ni pour les lags négatifs (influence de la production passée sur le prix).

### ***Test de Ljung-Box pour les résidus croisés***

**## Statistique Q : 16.556 | p-value : 0.6816**

Le test de Ljung-Box appliqué au produit des innovations vise à détecter une éventuelle dépendance croisée entre les résidus standardisés des deux séries (production et prix du pétrole).

Dans notre cas, la statistique du test est :  $Q = 16.556$  avec une  $p\text{-value} = 0.6816$ .

Cette  $p\text{-value}$  largement supérieure à 0.05 nous amène à ne pas rejeter l'hypothèse nulle d'indépendance. Cela signifie que les innovations des deux séries ne présentent pas de corrélation

significative, ce qui confirme que les modèles ARMA univariés ont bien capturé la dynamique individuelle de chaque série, sans laisser de dépendance résiduelle importante entre elles.

### **Explication des hypothèses sous-jacentes a un ARMA bivarié**

Un modèle ARMA bivarié permet de modéliser simultanément deux séries temporelles interdépendantes en tenant compte de leurs interactions dynamiques. Contrairement à une modélisation séparée où chaque série est traitée indépendamment, le modèle bivarié intègre les effets croisés, c'est-à-dire qu'une série peut influencer l'autre à travers ses valeurs passées et ses chocs précédents. Ce type de modèle repose sur plusieurs hypothèses : la stationnarité des deux séries, des relations linéaires entre les variables et leurs retards, ainsi qu'un bruit blanc multivarié centré, homoscedastique et éventuellement corrélé. Le principal avantage du modèle bivarié réside dans sa capacité à mieux représenter des phénomènes liés, par exemple en économie, où la consommation et le revenu s'influencent mutuellement. À l'inverse, une modélisation séparée suppose implicitement que les séries sont indépendantes, ce qui peut mener à une perte d'information lorsque ce n'est pas le cas. En résumé, la modélisation conjointe est à privilégier lorsqu'on soupçonne une interaction entre les séries, car elle améliore l'interprétation et la qualité des prévisions.

## **Conclusion**

Cette étude a permis de modéliser et d'analyser deux séries temporelles économiques cruciales : la production et le prix du pétrole. En suivant une approche rigoureuse, nous avons d'abord rendu les séries stationnaires, puis identifié et estimé des modèles ARMA (1,1) et ARMA (1,0) pour chacune des séries, en nous basant sur l'analyse des ACF, PACF, les critères d'information AIC/BIC, ainsi que les diagnostics sur les résidus.

Les comparaisons effectuées à l'aide des critères d'information, des tests de Ljung-Box et de l'erreur quadratique moyenne (MSE) ont permis de conclure que les modèles ARMA (1,0) plus simples sont statistiquement aussi performants que les modèles plus complexes. Cette parcimonie favorise la robustesse sans compromettre la qualité de l'ajustement.

L'analyse des innovations a montré que les résidus se comportent comme un bruit blanc, validant ainsi les modèles retenus. Par ailleurs, l'étude de la fonction de corrélation croisée (CCF) entre les innovations des deux séries n'a révélé que quelques corrélations ponctuelles, sans structure persistante. Le test de Ljung-Box croisé a confirmé l'absence de dépendance significative entre les innovations des deux séries, justifiant ainsi l'approche univariée.

Dans l'ensemble, les modèles ARMA choisis permettent de bien représenter la dynamique propre à chaque série. Bien que certaines corrélations croisées aient été observées, elles ne justifient pas, selon les critères statistiques, l'utilisation d'un modèle bivarié plus complexe.

# Annexe

## Références

**OPEC (2023).** *"OPEC Annual Report 2023"*

[https://www.opec.org/opec\\_web/en/publications/340.htm](https://www.opec.org/opec_web/en/publications/340.htm)

Cite les stratégies de production de l'OPEP et leur impact sur les prix mondiaux.

**U.S. Energy Information Administration (EIA) (2023).** *"Short-Term Energy Outlook"*.

Analyse l'évolution de la production américaine de schiste et ses effets sur l'offre globale.

**International Energy Agency (IEA) (2023).** *"Oil Market Report"*.

*Étudie les interactions entre prix, production et demande mondiale.*

## Code

### Importation des données

```
# Préparation des séries différenciées
prod_ts <- ts(data_production$Production, start = c(2000, 1), frequency = 12)
prix_ts <- ts(Data$POILBREUSD, start = c(2000, 1), frequency = 12)

diff1_prod <- diff(prod_ts)
diff1_12 <- diff(diff1_prod, lag = 12)

diff1_prix <- diff(prix_ts)
diff1_12_prix <- diff(diff1_prix, lag = 12)
```

### Statistiques descriptives

```
summary(data_production$Production)
```

### Graphique de la production de pétrole

```
# Graphique 1

p2 <- ggplot(data_production, aes(x=Date, y=Production)) +
```

```
geom_line(color="red") +
labs(title="Production de pétrole a travers le temps",
      x="Année", y="Production (milliers de barill/jours)") +
theme_minimal()
p2

# Annexes
```

## Statistiques descriptives

```
summary(Data$POILBREUSDM)
```

## Graphique du prix du pétrole

```
# Graphique 1
p1 <- ggplot(Data, aes(x=observation_date, y=POILBREUSDM)) +
  geom_line(color="blue") +
  labs(title="Prix du Pétrole a travers le temps",
        x="Années", y="Prix (dollars par barril)") +
  theme_minimal()

p1
```

## Filtrage linéaire

```
lissage_lineaire <- function(X, q) {
  N <- length(X)
  poids <- 1 / (2 * q + 1)
  indices <- (q+1):(N-q)
  X_filt <- rep(NA, N)
  for (i in indices) {
    X_filt[i] <- sum(X[(i-q):(i+q)] * poids)
  }
  return(X_filt)
}
```

## Filtrage exponentiel

```
# Lissage exponentiel
ExpSmooth <- function(x, alpha) {
  # Vérifier si alpha est dans une plage valide
  if (alpha <= 0 || alpha > 1)
    stop("alpha doit être compris entre 0 et 1.")
  # Pré-allouer et initialiser
  n <- length(x)
  Data <- numeric(n)
  Data[1] <- x[1]
  # Calculer les valeurs lissées de manière récursive
  for (i in 2:n) {
    Data[i] <- alpha * x[i] + (1 - alpha) * Data[i - 1]
  }
}
```

```
return(Data)
}
```

## Lissage des données sur la production

```
# Création de l'axe temporel réel
temps <- seq(from = as.Date("2000-01-01"), by = "month", length.out = length(
data_production$Production))

# Calcul des séries lissées pour différentes valeurs de alpha
Exp_1_01 <- ExpSmooth(data_production$Production, 0.1)
Exp_1_02 <- ExpSmooth(data_production$Production, 0.2)
Exp_1_05 <- ExpSmooth(data_production$Production, 0.5)
Exp_1_07 <- ExpSmooth(data_production$Production, 0.7)

# Tracé de la série originale
plot(temps, data_production$Production,
     type = "l",
     col = "blue",
     lwd = 1,
     main = "Lissage exponentiel (production de pétrole)",
     xlab = "Année",
     ylab = "Production")

# Ajout des courbes lissées
lines(temps, Exp_1_01, col = "red", lwd = 2)
lines(temps, Exp_1_02, col = "green", lwd = 2)
lines(temps, Exp_1_05, col = "purple", lwd = 2)
lines(temps, Exp_1_07, col = "orange", lwd = 2)

# Ajout de la légende
legend("topright",
     legend = c("Série originale", "alpha = 0.1", "alpha = 0.2", "alpha = 0.5", "alpha = 0.7"),
     col = c("blue", "red", "green", "purple", "orange"),
     lty = 1,
     lwd = c(1, 2, 2, 2, 2))
```

## Lissage des données sur le prix

```
# Création du vecteur de dates mensuelles
temps <- seq(from = as.Date("2000-01-01"), by = "month", length.out = length(
Data$POILBREUSD))

# Calcul des lissages exponentiels pour différentes valeurs de alpha
Exp_2_01 <- ExpSmooth(Data$POILBREUSD, 0.1)
Exp_2_02 <- ExpSmooth(Data$POILBREUSD, 0.2)
Exp_2_05 <- ExpSmooth(Data$POILBREUSD, 0.5)
Exp_2_07 <- ExpSmooth(Data$POILBREUSD, 0.7)
```

```

# Tracé de la série originale
plot(temps, Data$POILBREUSDM,
     type = "l",
     col = "blue",
     lwd = 1,
     main = "Lissage exponentiel (Prix du Brent)",
     xlab = "Année",
     ylab = "Prix du Brent")

# Ajout des courbes lissées
lines(temps, Exp_2_01, col = "red", lwd = 2)
lines(temps, Exp_2_02, col = "green", lwd = 2)
lines(temps, Exp_2_05, col = "purple", lwd = 2)
lines(temps, Exp_2_07, col = "orange", lwd = 2)

# Ajout de la légende
legend("topright",
      legend = c("Série originale", "alpha = 0.1", "alpha = 0.2", "alpha = 0.5", "alpha = 0.7"),
      col = c("blue", "red", "green", "purple", "orange"),
      lty = 1,
      lwd = c(1, 2, 2, 2, 2))

```

### Résidus Lissage linéaire ~ Série de la production

```

# Série temporelle avec dates
prod_ts <- ts(data_production$Production, start = c(2000, 1), frequency = 12)
dates_prod <- seq(as.Date("2000-01-01"), by = "month", length.out = length(prod_ts))

# Lissage et résidus
lisser_prod <- lissage_lineaire(prod_ts, q = 6)
resid_lin_prod <- prod_ts - lisser_prod

# Ajustement des dates pour résidus (car NA en début et fin dus au lissage)
resid_dates <- dates_prod[!is.na(resid_lin_prod)]
resid_vals <- resid_lin_prod[!is.na(resid_lin_prod)]

# Plot
plot(resid_dates, resid_vals, type = "l", col = "blue",
     xlab = "Année", ylab = "Résidus linéaires",
     main = "Résidus du lissage linéaire de la production")
abline(h = 0, col = "red", lty = 2)

moy_resid_lin_prod <- mean(resid_vals)

var_lin_prod <- var(resid_lin_prod, na.rm = TRUE)
cat("Variance des résidus (linéaire - production) :", var_lin_prod)

```

```
cat("Moyenne des résidus (linéaire - production) :", moy_resid_lin_prod)
```

### Résidus Lissage linéaire ~ Série du prix

```
# Série avec dates
prix_ts <- ts(Data$POILBREUSDM, start = c(2000, 1), frequency = 12)
dates_prix <- seq(as.Date("2000-01-01"), by = "month", length.out = length(pr
ix_ts))

# Lissage + résidus
lisse_prix <- lissage_lineaire(prix_ts, q = 6)
resid_lin_prix <- prix_ts - lisse_prix

# Dates propres
resid_dates <- dates_prix[!is.na(resid_lin_prix)]
resid_vals <- resid_lin_prix[!is.na(resid_lin_prix)]

plot(resid_dates, resid_vals, type = "l", col = "blue",
      xlab = "Année", ylab = "Résidus linéaires",
      main = "Résidus du lissage linéaire du prix du Brent")
abline(h = 0, col = "red", lty = 2)

moy_resid_lin_prix <- mean(resid_vals)

var_lin_prix <- var(resid_lin_prix, na.rm = TRUE)
cat("Variance des résidus (linéaire - prix) :", var_lin_prix)

cat("Moyenne des résidus (linéaire - prix) :", moy_resid_lin_prix)
```

### Résidus Lissage exponentiel ~ Série de la production

```
# Série temporelle de production avec dates
prod_ts <- ts(data_production$Production, start = c(2000, 1), frequency = 12)
dates_prod <- seq(as.Date("2000-01-01"), by = "month", length.out = length(pr
od_ts))

# Lissage exponentiel avec alpha = 0.2
lisse_exp_prod <- ExpSmooth(prod_ts, alpha = 0.2)
resid_exp_prod <- prod_ts - lisse_exp_prod

# Nettoyage des NA
resid_dates <- dates_prod[!is.na(resid_exp_prod)]
resid_vals <- resid_exp_prod[!is.na(resid_exp_prod)]

# Graphique
plot(resid_dates, resid_vals, type = "l", col = "darkgreen",
      xlab = "Année", ylab = "Résidus exponentiels",
```

```

    main = "Résidus du lissage exponentiel de la production (alpha = 0.2)")
abline(h = 0, col = "red", lty = 2)
moy_resid_exp_prod <- mean(resid_vals)

var_exp_prod <- var(resid_exp_prod, na.rm = TRUE)
cat("Variance des résidus (exponentiel - production) :", var_exp_prod)
cat("Moyenne des résidus (exponentiel - production) :", moy_resid_exp_prod)

```

## Résidus Lissage exponentiel ~ Série du prix

```

# Série temporelle du prix avec dates
prix_ts <- ts(Data$POILBREUSDM, start = c(2000, 1), frequency = 12)
dates_prix <- seq(as.Date("2000-01-01"), by = "month", length.out = length(pr
ix_ts))

# Lissage exponentiel
lisse_exp_prix <- ExpSmooth(prix_ts, alpha = 0.2)
resid_exp_prix <- prix_ts - lisse_exp_prix

# Nettoyage
resid_dates <- dates_prix[!is.na(resid_exp_prix)]
resid_vals <- resid_exp_prix[!is.na(resid_exp_prix)]

# Graphique
plot(resid_dates, resid_vals, type = "l", col = "darkgreen",
      xlab = "Année", ylab = "Résidus exponentiels",
      main = "Résidus du lissage exponentiel du prix du Brent (alpha = 0.2)")
abline(h = 0, col = "red", lty = 2)

moy_resid_exp_prix <- mean(resid_vals)

var_exp_prix <- var(resid_exp_prix, na.rm = TRUE)
cat("Variance des résidus (exponentiel - prix) :", var_exp_prix)

cat("Moyenne des résidus (exponentiel - prix) :", moy_resid_exp_prix)

```

## Différentiation de la série sur la production

```

# Série de base : production
prod_ts <- ts(data_production$Production, start = c(2000, 1), frequency = 12)
dates <- seq(from = as.Date("2000-01-01"), by = "month", length.out = length(
prod_ts))

# Différentiations successives
diff1_prod <- diff(prod_ts)           # Première
diff2_prod <- diff(diff1_prod)        # Deuxième
diff3_prod <- diff(diff2_prod)        # Troisième
diff4_prod <- diff(diff3_prod)        # Quatrième

```



```

# Dates ajustées
dates1 <- dates[-1]
dates2 <- dates[-c(1,2)]
dates3 <- dates[-c(1,2,3)]
dates4 <- dates[-c(1,2,3,4)]

# Tracés
par(mfrow = c(4,1))

plot(dates1, diff1_prod, type = "l", col = "blue",
      xlab = "Année", ylab = "ΔProduction",
      main = "Production - Première différenciation")

plot(dates2, diff2_prod, type = "l", col = "darkgreen",
      xlab = "Année", ylab = "Δ²Production",
      main = "Production - Deuxième différenciation")

plot(dates3, diff3_prod, type = "l", col = "purple",
      xlab = "Année", ylab = "Δ³Production",
      main = "Production - Troisième différenciation")

plot(dates4, diff4_prod, type = "l", col = "brown",
      xlab = "Année", ylab = "Δ⁴Production",
      main = "Production - Quatrième différenciation")

par(mfrow = c(1,1))

```

### Différenciation de la série sur le prix du pétrole

```

# Série initiale avec axe temporel
prix_ts <- ts(Data$POILBREUSDM, start = c(2000, 1), frequency = 12)
dates <- seq(from = as.Date("2000-01-01"), by = "month", length.out = length(
  prix_ts))

# Différenciations successives
diff1 <- diff(prix_ts) # 1re différenciation
diff2 <- diff(diff1)   # 2e différenciation
diff3 <- diff(diff2)   # 3e différenciation
diff4 <- diff(diff3)   # 4e différenciation

# Vecteurs temporels ajustés (on retire 1, 2, 3, 4 dates respectivement)
dates1 <- dates[-1]
dates2 <- dates[-c(1,2)]
dates3 <- dates[-c(1,2,3)]
dates4 <- dates[-c(1,2,3,4)]

# Affichage des graphiques
par(mfrow = c(4,1))

```

```

plot(dates1, diff1, type = "l", col = "blue",
     xlab = "Année", ylab = "ΔPrix",
     main = "Prix - Première différenciation")

plot(dates2, diff2, type = "l", col = "darkgreen",
     xlab = "Année", ylab = "Δ²Prix",
     main = "Prix - Deuxième différenciation")

plot(dates3, diff3, type = "l", col = "purple",
     xlab = "Année", ylab = "Δ³Prix",
     main = "Prix - Troisième différenciation")

plot(dates4, diff4, type = "l", col = "brown",
     xlab = "Année", ylab = "Δ⁴Prix",
     main = "Prix - Quatrième différenciation")

par(mfrow = c(1,1))

```

### Autocorrélation de la série sur la production de pétrole

```

h_max <- 20
n <- length(diff1_prod)
acf_result <- acf(diff1_prod, lag.max = h_max, plot = FALSE)
acf_values <- acf_result$acf[-1]

Q_lb <- p_values <- lower_conf <- upper_conf <- numeric(h_max)
for (h in 1:h_max) {
  Q_lb[h] <- n * (n + 2) * sum(acf_values[1:h]^2 / (n - (1:h)))
  p_values[h] <- 1 - pchisq(Q_lb[h], df = h)
  chi_crit <- qchisq(0.975, df = h)
  lower_conf[h] <- -sqrt(chi_crit / n)
  upper_conf[h] <- sqrt(chi_crit / n)
}

significant <- (acf_values < lower_conf) | (acf_values > upper_conf)

plot(1:h_max, acf_values, type = "h", lwd = 2,
     col = ifelse(significant, "red", "blue"),
     main = "ACF de la production (1re différenciation) avec IC",
     xlab = "Décalage (h)", ylab = expression(gamma(h)), ylim = c(-1,1))
abline(h = 0, col = "black", lty = 2)
lines(1:h_max, lower_conf, col = "red", lty = 2)
lines(1:h_max, upper_conf, col = "red", lty = 2)
points(1:h_max, acf_values, pch = 16, col = ifelse(significant, "red", "blue"))
arrows(1:h_max, lower_conf, 1:h_max, upper_conf, angle = 90, code = 3, length
      = 0.05, col = "darkgray")

```

## Autocorrélation de la série sur le prix du pétrole

```
h_max <- 20
n <- length(diff1_prix)
acf_result <- acf(diff1_prix, lag.max = h_max, plot = FALSE)
acf_values <- acf_result$acf[-1]

Q_lb <- p_values <- lower_conf <- upper_conf <- numeric(h_max)
for (h in 1:h_max) {
  Q_lb[h] <- n * (n + 2) * sum(acf_values[1:h]^2 / (n - (1:h)))
  p_values[h] <- 1 - pchisq(Q_lb[h], df = h)
  chi_crit <- qchisq(0.975, df = h)
  lower_conf[h] <- -sqrt(chi_crit / n)
  upper_conf[h] <- sqrt(chi_crit / n)
}

significant <- (acf_values < lower_conf) | (acf_values > upper_conf)

plot(1:h_max, acf_values, type = "h", lwd = 2,
     col = ifelse(significant, "red", "blue"),
     main = "ACF du prix du Brent (1re différenciation) avec IC",
     xlab = "Décalage (h)", ylab = expression(gamma(h)), ylim = c(-1,1))
abline(h = 0, col = "black", lty = 2)
lines(1:h_max, lower_conf, col = "red", lty = 2)
lines(1:h_max, upper_conf, col = "red", lty = 2)
points(1:h_max, acf_values, pch = 16, col = ifelse(significant, "red", "blue"))
arrows(1:h_max, lower_conf, 1:h_max, upper_conf, angle = 90, code = 3, length
      = 0.05, col = "darkgray")
```

## Modélisation et Prévision avec les Processus ARMA

```
# Fonction pour calculer L'ACF
compute_acf <- function(series, lags) {
  n <- length(series)
  mean_series <- mean(series)
  var_series <- var(series)
  acf_values <- numeric(lags + 1)

  for (h in 0:lags) {
    num <- sum((series[1:(n-h)] - mean_series) * (series[(h+1):n] -
mean_series))
    acf_values[h + 1] <- num / ((n - h) * var_series)
  }

  return(acf_values) }

# Fonction pour calculer La PACF en utilisant La méthode de Durbin-Levinson
```

```

compute_pacf <- function(series, lags) {
  pacf_values <- numeric(lags + 1)
  pacf_values[1] <- 1

  for (h in 1:lags) {
    acf_values <- compute_acf(series, h)
    R <- matrix(0, nrow = h, ncol = h)
    r <- acf_values[2:(h+1)]

    for (i in 1:h) {
      for (j in 1:h) {
        if (abs(i - j) == 0) {
          R[i, j] <- 1
        } else {
          R[i, j] <- acf_values[abs(i - j) + 1]
        }
      }
    }

    phi <- solve(R, r)
    pacf_values[h + 1] <- phi[h]
  }

  return(pacf_values)
}

# Fonction pour calculer les seuils de significativité
compute_confidence_intervals <- function(n, alpha) {
  return(qnorm((1 + alpha) / 2) / sqrt(n))
}

# Fonction pour tracer l'ACF et la PACF avec intervalles de confiance
compute_acf_pacf <- function(series, lags = 20, alpha_levels = c(0.9, 0.95, 0.99)) {
  acf_values <- compute_acf(series, lags)
  pacf_values <- compute_pacf(series, lags)
  n <- length(series)
  colors <- c("blue", "green", "purple")

  par(mfrow = c(1, 1))

  # Tracer l'ACF avec intervalles de confiance
  plot(0:lags, acf_values, type = "h", main = "Autocorrélation (ACF)", xlab = "Lag", ylab = "ACF")
  abline(h = 0, col = "red")

  for (i in seq_along(alpha_levels)) {
    threshold <- compute_confidence_intervals(n, alpha_levels[i])
    abline(h = c(-threshold, threshold), col = colors[i], lty = 2)
  }
}

```

```

    }
    legend("topright", legend = paste0("IC ", alpha_levels * 100, "%"), col =
colors, lty = 2, title = "Intervalles de confiance")

    # Tracer la PACF avec intervalles de confiance
    plot(0:lags, pacf_values, type = "h", main = "Autocorrélation partielle
(PACF)", xlab = "Lag", ylab = "PACF")
    abline(h = 0, col = "red")

    for (i in seq_along(alpha_levels)) {
        threshold <- compute_confidence_intervals(n, alpha_levels[i])
        abline(h = c(-threshold, threshold), col = colors[i], lty = 2)
    }
    legend("topright", legend = paste0("IC ", alpha_levels * 100, "%"), col =
colors, lty = 2, title = "Intervalles de confiance")
}

```

### Série de la production de pétrole

```
compute_acf_pacf(diff1_prod, 20)
```

### Série du prix du pétrole

```
compute_acf_pacf(diff1_prix, 20)
```

### Estimation de modèles ARMA

#### Série sur la production de pétrole

```

library(stats)
# Estimation du modèle ARMA(1,1) pour la série de production (différentiée)
model_prod_arma11 <- arima(train_prod, order = c(1, 0, 1), method = "ML")

# Affichage des résultats
model_prod_arma11
cat("BIC (ARMA(1,1) - production) :", BIC(model_prod_arma11), "\n")

plot_acf <- function(series, lags = 20, alpha_levels = 0.95, main_title =
"Autocorrélation (ACF)") {
    acf_values <- compute_acf(series, lags)
    n <- length(series)
    colors <- "blue"

    plot(0:lags, acf_values, type = "h", main = main_title, xlab = "Lag", ylab
= "ACF", ylim = c(-1, 1))
    abline(h = 0, col = "red")

    for (i in seq_along(alpha_levels)) {
        threshold <- compute_confidence_intervals(n, alpha_levels[i])
        abline(h = c(-threshold, threshold), col = colors[i], lty = 2)
    }
}

```

```

    legend("topright", legend = paste0("IC ", alpha_levels * 100, "%"), col =
colors, lty = 2, title = "Intervalle de confiance")
}

```

## Analyse des résidus

```

# Affichage de L'ACF des résidus du modèle ARMA(1,0) - Production
plot_acf(residuals(model_prod_arma11), main_title = "ACF des Résidus du
Modèle ARMA(1,1)")

```

## Test de Ljung-Box

```

# Test de Ljung-Box pour Les résidus du modèle ARMA(1,0) - Production
lb_test_prod <- Box.test(residuals(model_prod_arma11), lag = 20, type =
"Ljung-Box")
cat("Statistique Q :", round(lb_test_prod$statistic, 2),
    "| p-value :", round(lb_test_prod$p.value, 4))

```

## Test des Points Tournants

```

# Fonction du test des points de retournement
turning_points_test <- function(x) {
  n <- length(x)
  tp <- sum(diff(sign(diff(x))) != 0, na.rm = TRUE)
  mu <- 2 * (n - 2) / 3
  sigma <- sqrt((16 * n - 29) / 90)
  z <- (tp - mu) / sigma
  pval <- 2 * pnorm(-abs(z))
  return(list(turning_points = tp, z = z, p.value = pval))
}

# Application aux résidus de ton modèle
tp_test_prod <- turning_points_test(residuals(model_prod_arma11))
cat("Turning Points :", tp_test_prod$turning_points,
    "| z-score :", round(tp_test_prod$z, 4),
    "| p-value :", round(tp_test_prod$p.value, 4))

```

## Série sur le prix du pétrole

```

# Estimation du modèle ARMA(1,0) pour la série de prix (différentiée)
model_prix_arma11 <- arima(train_prix, order = c(1, 0, 1), method = "ML")

# Affichage des résultats
model_prix_arma11
cat("BIC (ARMA(1,1) - prix) :", BIC(model_prix_arma11), "\n")

```

## Analyse des résidus

```

# Affichage de L'ACF des résidus du modèle ARMA(1,0) - Production
plot_acf(residuals(model_prix_arma11), main_title = "ACF des Résidus du
Modèle ARMA(1,1) sur le Prix")

```

## Test de Lyung-box

```
# Test de Ljung-Box pour Les résidus du modèle ARMA(1,1) - Prix
lb_test_prix <- Box.test(residuals(model_prix_arma11), lag = 20, type =
"Ljung-Box")
cat("Statistique Q :", round(lb_test_prix$statistic, 2),
    "| p-value :", round(lb_test_prix$p.value, 4))
```

## Test des points tournants

```
# Application du test des points de retournement aux résidus du modèle
ARMA(1,1) - Prix
tp_test_prix <- turning_points_test(residuals(model_prix_arma11))
cat("Turning Points :", tp_test_prix$turning_points,
    "| z-score :", round(tp_test_prix$z, 4),
    "| p-value :", round(tp_test_prix$p.value, 4))
```

## Estimation des modèles imbriqués

### Série de la production

```
# Estimation du modèle ARMA(1,0) (modèle imbriqué) pour la série de
production différenciée
model_prod_arma10 <- arima(train_prod, order = c(1, 0, 0), method = "ML")

# Affichage du modèle et du BIC
model_prod_arma10
cat("BIC (ARMA(1,0) - production) :", BIC(model_prod_arma10), "\n")
```

### Analyse des résidus (modèle simplifiée)

```
plot_acf(residuals(model_prod_arma10), main_title = "ACF des Résidus du Modèle
ARMA(1,0)")
```

## Test de Lyung-Box

```
# Test de Ljung-Box pour Les résidus du modèle ARMA(1,0) - Production
lb_test_prod_arma10 <- Box.test(residuals(model_prod_arma10), lag = 20, type
= "Ljung-Box")
cat("Statistique Q :", round(lb_test_prod_arma10$statistic, 2),
    "| p-value :", round(lb_test_prod_arma10$p.value, 4))
```

## Test des points tournants

```
# Test des points de retournement sur Les résidus du modèle ARMA(1,0)
tp_test_prod_arma10 <- turning_points_test(residuals(model_prod_arma10))
cat("Turning Points :", tp_test_prod_arma10$turning_points,
    "| z-score :", round(tp_test_prod_arma10$z, 4),
    "| p-value :", round(tp_test_prod_arma10$p.value, 4))

# Estimation du modèle ARMA(1,0) pour la série de prix (différenciée)
model_prix_arma10 <- arima(train_prix, order = c(1, 0, 0), method = "ML")
```

```
# Affichage des résultats
model_prix_arma10
cat("BIC (ARMA(1,0) - prix) :", BIC(model_prix_arma10), "\n")
```

### Analyse des résidus (Modèle simplifiée)

```
# Affichage de l'ACF des résidus du modèle ARMA(1,0) - Prix
plot_acf(residuals(model_prix_arma10), main_title = "ACF des Résidus du
Modèle ARMA(1,0) sur le Prix")
```

### Test de Ljung-Box

```
# Test de Ljung-Box sur Les résidus du modèle ARMA(1,0) - Prix
lb_test_prix_arma10 <- Box.test(residuals(model_prix_arma10), lag = 20, type
= "Ljung-Box")
cat("Statistique Q :", round(lb_test_prix_arma10$statistic, 2),
    "| p-value :", round(lb_test_prix_arma10$p.value, 4))
```

### Test des points tournants

```
# Test des points de retournement sur Les résidus du modèle ARMA(1,0) - Prix
tp_test_prix_arma10 <- turning_points_test(residuals(model_prix_arma10))
cat("Turning Points :", tp_test_prix_arma10$turning_points,
    "| z-score :", round(tp_test_prix_arma10$z, 4),
    "| p-value :", round(tp_test_prix_arma10$p.value, 4))
```

### Analyse des Paramètres et Modèles Imbriqués - Série sur la production de pétrole

```
library(knitr)

# Données pour la série de la production
prod_table <- data.frame(
  Modèle = c("ARMA(1,1)", "ARMA(1,1)", "ARMA(1,1)", "ARMA(1,0)",
"ARMA(1,0)"),
  Paramètre = c("AR(1)", "MA(1)", "Intercept", "AR(1)", "Intercept"),
  Estimation = c(0.1589, -0.1773, 26.6181, -0.0157, 26.6196),
  `Écart-type` = c(1.0807, 1.0785, 10.6365, 0.0660, 10.7066)
)

kable(prod_table, caption = "Paramètres estimés - Série de la production")
```

### Série sur le prix du pétrole

```
# Données pour la série du prix
prix_table <- data.frame(
  Modèle = c("ARMA(1,1)", "ARMA(1,1)", "ARMA(1,1)", "ARMA(1,0)",
"ARMA(1,0)"),
  Paramètre = c("AR(1)", "MA(1)", "Intercept", "AR(1)", "Intercept"),
  Estimation = c(0.3687, 0.0261, 0.1910, 0.3907, 0.1917),
  `Écart-type` = c(0.1403, 0.1478, 0.5455, 0.0605, 0.5509)
)
```



```
kable(prix_table, caption = "Paramètres estimés - Série du prix")
```

## Tests Diagnostiques

### Critères d'Information (AIC/BIC)

```
# Fonction pour calculer l'AIC et le BIC sur entraînement et test
compute_aic_bic <- function(model, train_data, test_data) {
  # (1) AIC et BIC sur entraînement
  aic_train <- AIC(model)
  bic_train <- BIC(model)

  # (2) Calcul du log-vraisemblance sur test
  residuals_test <- test_data - predict(model, n.ahead =
length(test_data))$pred
  sigma2_test <- var(residuals_test) # Variance des résidus
  loglik_test <- -0.5 * length(test_data) * log(2 * pi * sigma2_test) -
sum(residuals_test^2) / (2 * sigma2_test)

  # Nombre de paramètres
  k <- length(coef(model))

  # AIC et BIC sur test
  aic_test <- -2 * loglik_test + 2 * k
  bic_test <- -2 * loglik_test + log(length(test_data)) * k

  return(c(AIC_train = aic_train, BIC_train = bic_train, AIC_test = aic_test,
BIC_test = bic_test))
}
```

```
library(knitr)
```

```
# Construction du tableau
compare_models <- data.frame(
  Modèle = c("Production ARMA(1,1)", "Production ARMA(1,0)",
    "Prix ARMA(1,1)", "Prix ARMA(1,0)"),
  AIC_Train = c(stats_prod_arma11["AIC_train"],
stats_prod_arma10["AIC_train"],
    stats_prix_arma11["AIC_train"],
stats_prix_arma10["AIC_train"]),
  BIC_Train = c(stats_prod_arma11["BIC_train"],
stats_prod_arma10["BIC_train"],
    stats_prix_arma11["BIC_train"],
stats_prix_arma10["BIC_train"]),
  AIC_Test = c(stats_prod_arma11["AIC_test"],
stats_prod_arma10["AIC_test"],
    stats_prix_arma11["AIC_test"],
stats_prix_arma10["AIC_test"]),
  BIC_Test = c(stats_prod_arma11["BIC_test"],
```

```
stats_prod_arma10["BIC_test"],
               stats_prix_arma11["BIC_test"],
stats_prix_arma10["BIC_test"])
)

kable(compare_models, digits = 2, caption = "Comparaison des AIC et BIC sur
l'ensemble d'entraînement et de test")
```

## Test du ratio de vraisemblance

```
# Fonction pour le test du ratio de vraisemblance
lr_test <- function(model_restreint, model_complet) {
  L1 <- logLik(model_restreint)
  L2 <- logLik(model_complet)
  test_stat <- -2 * (as.numeric(L1) - as.numeric(L2))
  df <- abs(length(coef(model_complet)) - length(coef(model_restreint)))
  p_value <- 1 - pchisq(test_stat, df)
  return(c(Statistic = test_stat, p_value = p_value))
}

# Application aux deux séries
lr_prod <- lr_test(model_prod_arma10, model_prod_arma11)
lr_prix <- lr_test(model_prix_arma10, model_prix_arma11)

# Résumé sous forme de tableau
library(knitr)
lr_results <- data.frame(
  Série      = c("Production", "Prix"),
  Modèle_1   = c("ARMA(1,0)", "ARMA(1,0)"),
  Modèle_2   = c("ARMA(1,1)", "ARMA(1,1)"),
  Statistique = c(round(lr_prod["Statistic"], 4),
round(lr_prix["Statistic"], 4)),
  p_value     = c(round(lr_prod["p_value"], 4), round(lr_prix["p_value"],
4))
)

# Affichage du tableau
kable(lr_results, caption = "Test du ratio de vraisemblance entre modèles
imbriqués")

# Fonction corrigée pour évaluer les modèles ARMA
evaluer_modeles <- function(model, train, test, nom_serie, ordre) {
  # Calcul des valeurs ajustées via les résidus
  fitted_values <- train - residuals(model)

  # Prédiction sur l'ensemble de test
  predictions <- predict(model, n.ahead = length(test))$pred

  # Calcul des EQM
  data.frame(
```

```

    Série = nom_serie,
    Modèle = paste0("ARMA(", ordre[1], ",", ordre[2], ")"),
    MSE_Train = mean((train - fitted_values)^2),
    MSE_Test = mean((test - predictions)^2)
  )
}

# Évaluation des modèles pour la série de production
results_prod <- rbind(
  evaluer_modeles(model_prod_arma11, train_prod, test_prod, "Production",
c(1,1)),
  evaluer_modeles(model_prod_arma10, train_prod, test_prod, "Production",
c(1,0))
)

# Évaluation des modèles pour la série du prix
results_prix <- rbind(
  evaluer_modeles(model_prix_arma11, train_prix, test_prix, "Prix", c(1,1)),
  evaluer_modeles(model_prix_arma10, train_prix, test_prix, "Prix", c(1,0))
)

# Fusion des résultats
results_final <- rbind(results_prod, results_prix)

# Affichage propre
cat("\nAnalyse comparative des performances prédictives (Erreur Quadratique
Moyenne - EQM)\n")

cat("=====\n")
cat("Série : Production\n")
print(subset(results_final, Série == "Production"), row.names = FALSE)
cat("\nSérie : Prix\n")
print(subset(results_final, Série == "Prix"), row.names = FALSE)

```

## Modélisation Bivariée et Corrélations Croisées

### Analyse des innovations

#### Série sur la production de pétrole

```

# Fonction personnalisée d'approximation des innovations (standardisées)
compute_innovations <- function(model, series) {
  residuals <- residuals(model)
  sigma <- sqrt(model$sigma2)
  innovations <- residuals / sigma
  return(innovations)
}

dates_prod <- seq(from = as.Date("2000-01-01"), by = "month", length.out =

```

```
length(train_prod))
dates_prix <- seq(from = as.Date("2000-01-01"), by = "month", length.out =
length(train_prix))

# Calcul des innovations à partir des modèles ARMA(1,0)
innov_prod <- compute_innovations(model_prod_arma10, train_prod)
innov_prix <- compute_innovations(model_prix_arma10, train_prix)

# Tracé des innovations - Production avec dates en abscisse
plot(dates_prod, innov_prod, type = "l", col = "darkgreen",
     main = "Innovations - Production pétrolière",
     ylab = "Valeurs standardisées", xlab = "Date")
abline(h = 0, col = "red", lty = 2)
```

### Serie sur le prix du petrole

```
# Tracé des innovations - Prix avec dates en abscisse
plot(dates_prix, innov_prix, type = "l", col = "blue",
     main = "Innovations - Prix du pétrole",
     ylab = "Valeurs standardisées", xlab = "Date")
abline(h = 0, col = "red", lty = 2)
```

### Propriétés statistiques des innovations

```
# Fonction personnalisée pour la CCF
my_ccf <- function(x, y, lag.max = 20, alpha = 0.05, plot = TRUE) {
  n <- length(x)
  if (length(y) != n) stop("Les séries doivent avoir la même longueur")

  # Standardisation des séries
  x_std <- (x - mean(x)) / sd(x)
  y_std <- (y - mean(y)) / sd(y)

  # Calcul des corrélations croisées
  ccf_values <- numeric(2 * lag.max + 1)
  lags <- -lag.max:lag.max

  for (i in seq_along(lags)) {
    lag <- lags[i]
    if (lag <= 0) {
      ccf_values[i] <- sum(x_std[1:(n + lag)] * y_std[(1 - lag):n]) / (n +
lag)
    } else {
      ccf_values[i] <- sum(x_std[(1 + lag):n] * y_std[1:(n - lag)]) / (n -
lag)
    }
  }

  # Calcul de l'intervalle de confiance
  z_critical <- qnorm(1 - alpha / 2)
```

```

conf_interval <- z_critical / sqrt(n)

# Affichage
if (plot) {
  plot(lags, ccf_values, type = "h",
       main = "Corrélation croisée entre les innovations\nProduction vs
Prix du pétrole",
       xlab = "Décalage (lags)", ylab = "Corrélation",
       ylim = c(-1, 1), lwd = 2, col = "blue")
  abline(h = c(-conf_interval, conf_interval), lty = 2, col = "red")
  abline(h = 0, col = "black")
  grid()
}

# Résultat tabulaire
data.frame(
  lag = lags,
  correlation = round(ccf_values, 4),
  significant = abs(ccf_values) > conf_interval
)
}

# Application à vos données
ccf_results <- my_ccf(innov_prod, innov_prix, lag.max = 20)

```

### Test de Ljung-Box pour les résidus croisés

```

# Test de Ljung-Box sur le produit des innovations (résidus croisés)
test_lb_cross <- Box.test(innov_prod * innov_prix, lag = 20, type = "Ljung-
Box")
cat("Statistique Q :", round(test_lb_cross$statistic, 3),
    "| p-value :", round(test_lb_cross$p.value, 4))

```