

Prévisions de la production annuelle de pétrole non raffiné pour 2022 à 2025 au Canada

Econométrie des séries temporelles



Alexandra Millot

Contexte Historique : La production de pétrole au Canada



Au cours de ces trente dernières années, la production de pétrole brut au Canada a fortement augmenté, passant de 1,7 million de barils par jour (« Mb/j ») en 1990 à 4,7 Mb/j en 2019. Cette croissance a propulsé le Canada au rang de quatrième producteur mondial de pétrole. En plus de cette production significative, le Canada détient d'importantes réserves de sables bitumineux. L'industrie pétrolière canadienne est principalement concentrée en Alberta, où les sables bitumineux sont extraits et transformés en pétrole brut.

Actuellement, le Canada produit environ 4,625 millions de barils par jour de pétrole. Des projections suggèrent que la production du pays pourrait augmenter de plus de 120 % d'ici 2050, ce qui représente un taux de croissance supérieur à celui de tout autre pays non membre de l'OPEP (Organisation des pays exportateurs de pétrole). Cependant, l'exploitation des sables bitumineux est également sujette à des controverses en raison de ses implications environnementales, notamment en ce qui concerne les émissions de gaz à effet de serre et la déforestation. Le Canada est donc confronté à des défis complexes en matière de développement énergétique, cherchant à concilier la croissance économique avec la protection de l'environnement et la transition vers des sources d'énergie plus durables.

1858 : Découverte du premier gisement de pétrole commercial

1947 : Découverte du gisement de pétrole de Leduc No.1 en Alberta

1973 : Crise pétrolière (prix du pétrole qui augmente)

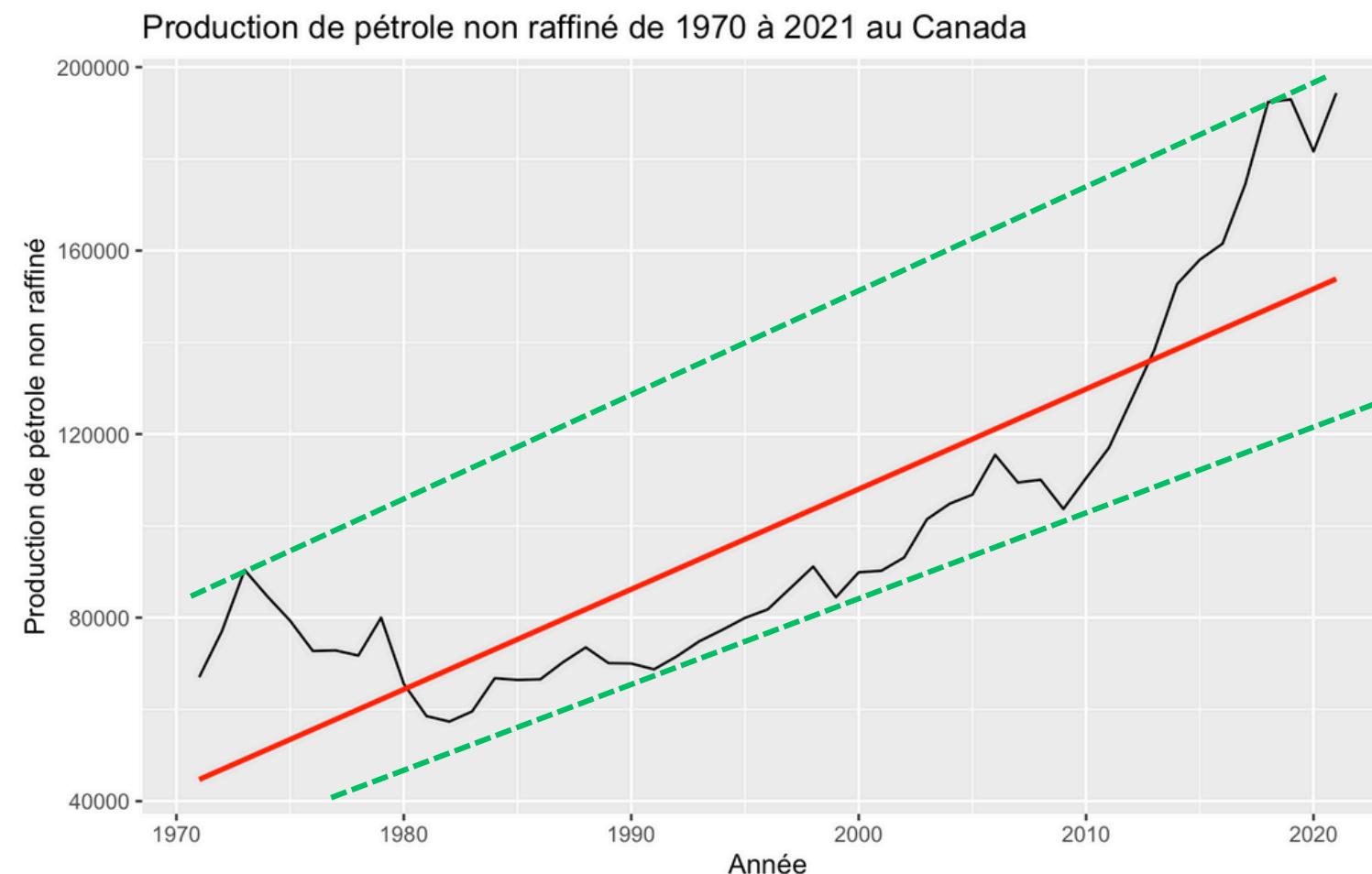
2003 : Début de la production pétrolière dans les sables bitumineux

2015 : Baisse des prix du pétrole

2020 : Pandémie de COVID-19

Chronogramme de la production annuelle de pétrole non raffiné de 1971 à 2021

Nous traçons tout d'abord le chronogramme de nos données représentant notre variable quantitative étant la production annuelle de pétrole non raffiné (sur l'axe des Y) en fonction du temps, de 1970 à 2021 (sur l'axe des X).



Interprétation

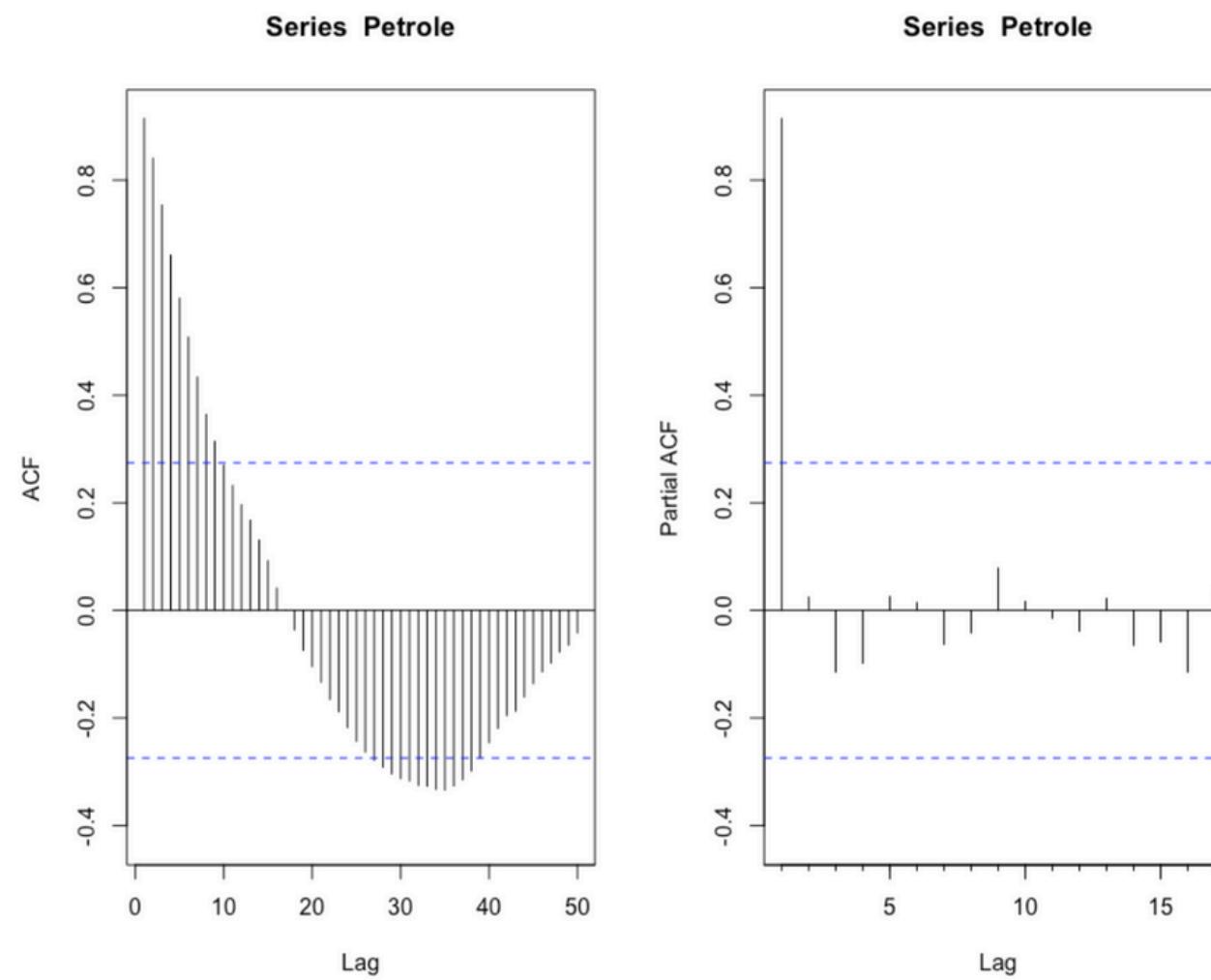
La courbe rouge de la série temporelle met en évidence une tendance à la hausse.

La série temporelle n'a pas d'effets saisonniers étant donné qu'on possède qu'une seule valeur par an.

À l'aide des deux droites vertes nous observons qu'elles ne sont pas parallèles entre elles. On en déduit une hétéroscédasticité, la variance des données augmente avec le temps.

Fonction d'autocorrélation totale et partielle de la production annuelle de la série d'origine

Nous traçons tout d'abord le chronogramme de nos données représentant notre variable quantitative étant la production annuelle de pétrole non raffiné (sur l'axe des Y) en fonction du temps, de 1970 à 2021 (sur l'axe des X).



Interprétation

Nous observons pour l'ACF une variation des valeurs des coefficients d'autocorrélation, le premier décalage comporte un pic au-dessus de 0.8, suivi d'une vague décroissante qui alterne entre les corrélations positives et négatives. Cela indique une structure d'autocorrélation.

Pour poursuivre l'analyse, la première valeur du PACF est hors de l'intervalle de confiance avec une forte corrélation de 0.9, suivie de l'absence d'autocorrélation ultérieure, cela peut indiquer que nous sommes dans un modèle AR(1) de la forme :

$$\text{Petrole}_t = \phi_0 + \phi_1 \text{Petrole}_{t-1} + \varepsilon_t$$

Avec ε_t un bruit blanc gaussien.

Racines unitaires

Détermination du PGD de la série, afin de la rendre stationnaire

- Dickey Fuller
- Dickey Fuller Augmenté
- Zivot et Andrews
- Lee et Strazicich
- Lee et Strazicich (Bootstrap)

Dickey-Fuller : “Trend”

Consiste à estimer par les MCO :

$$\Delta \text{Petrole}_t = (\rho - 1) \text{Petrole}_{t-1} + \beta_0 + \beta_1 \text{tendance}_t + \varepsilon_t$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \text{ et } \beta_1 = 0 \text{ versus } H_a : |\rho| < 1 \text{ et } \beta_1 \neq 0$$

Tout d'abord nous regardons la p-value associé à β_1 qui est de 0.101
 > 0.05 (seuil de significativité)

On en déduit que β_1 n'est pas significatif. Alors la spécification « trend » n'est pas la bonne et nous passons à la spécification « drift »

```
> summary(ur.df(x,type="trend",lag=0))

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend
```

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt)

Residuals:

Min	1Q	Median	3Q	Max
-16630.9	-3256.4	-27.6	2323.2	14945.2

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-536.91325	2769.71748	-0.194	0.847
z.lag.1	-0.01896	0.04639	-0.409	0.685
tt	193.33788	115.61438	1.672	0.101

Residual standard error: 6477 on 47 degrees of freedom
Multiple R-squared: 0.1139, Adjusted R-squared: 0.0762
F-statistic: 3.021 on 2 and 47 DF, p-value: 0.05831

Value of test-statistic is: -0.4087 4.5916 3.0209

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-4.04	-3.45	-3.15
phi2	6.50	4.88	4.16
phi3	8.73	6.49	5.47

Dickey-Fuller : "Drift"

Consiste à estimer par les MCO :

$$\Delta \text{Petrole}_t = (\rho - 1) \text{Petrole}_{t-1} + \beta_0 + \varepsilon_t$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \text{ et } \beta_0 = 0 \text{ versus } H_a : |\rho| < 1 \text{ et } \beta_0 \neq 0$$

Tout d'abord nous regardons la p-value associé à β_0 qui est de $0.4792 > 0.05$ (seuil de significativité)

On en déduit que β_0 n'est pas significatif. Alors la spécification « drift » n'est pas la bonne et nous passons à la spécification « none »

```
> summary(ur.df(x,type="drift",lag=0))

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift
```

Call:
lm(formula = z.diff ~ z.lag.1 + 1)

Residuals:
Min 1Q Median 3Q Max
-18285.7 -2524.2 664.8 3800.0 11800.4

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.920e+03 2.692e+03 -0.713 0.4792
z.lag.1 4.588e-02 2.594e-02 1.769 0.0833 .

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 ' ' 1

Residual standard error: 6597 on 48 degrees of freedom
Multiple R-squared: 0.06119, Adjusted R-squared: 0.04163
F-statistic: 3.128 on 1 and 48 DF, p-value: 0.0833

Value of test-statistic is: 1.7687 5.2911

Critical values for test statistics:
1pct 5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1 6.70 4.71 3.86

Dickey-Fuller : "None"

Consiste à estimer par les MCO :

$$\Delta \text{Petrole}_t = (\rho - 1) \text{Petrole}_{t-1} + \varepsilon_t$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \quad \text{versus} \quad H_a : |\rho| < 1$$

Pour savoir si on accepte H_0 ou non, la règle de décision est si la valeur de la statistique t associée à $(\rho - 1)$ > à la valeur critique donnée à l'intersection de la ligne tau1 et de la colonne 5%, on accepte H_0 .

Ici la valeur de la statistique t associée à $(\rho - 1)$ est 3.1899 et la valeur critique est de -1.95. On a bien $3.1899 > -1.95$

On accepte alors $H_0 : \rho - 1 = 0$

Le processus est DS d'après Dickey-Fuller

```
> summary(ur.df(x,type="none",lag=0))

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-16857.5 -3105.5   367.8  3697.4 12908.3 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
z.lag.1  0.028525  0.008942   3.19  0.00248 **  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6564 on 49 degrees of freedom
Multiple R-squared:  0.172,    Adjusted R-squared:  0.1551 
F-statistic: 10.18 on 1 and 49 DF,  p-value: 0.002483

Value of test-statistic is: 3.1899

Critical values for test statistics:
      1pct   5pct 10pct 
tau1 -2.6 -1.95 -1.61
```

Test de racine unitaire

Les conclusions du test DF ne sont valables que si les 23 ne sont pas autocorrélés. Testons cela avec l'ACF et le PACF des résidus des régressions DF.

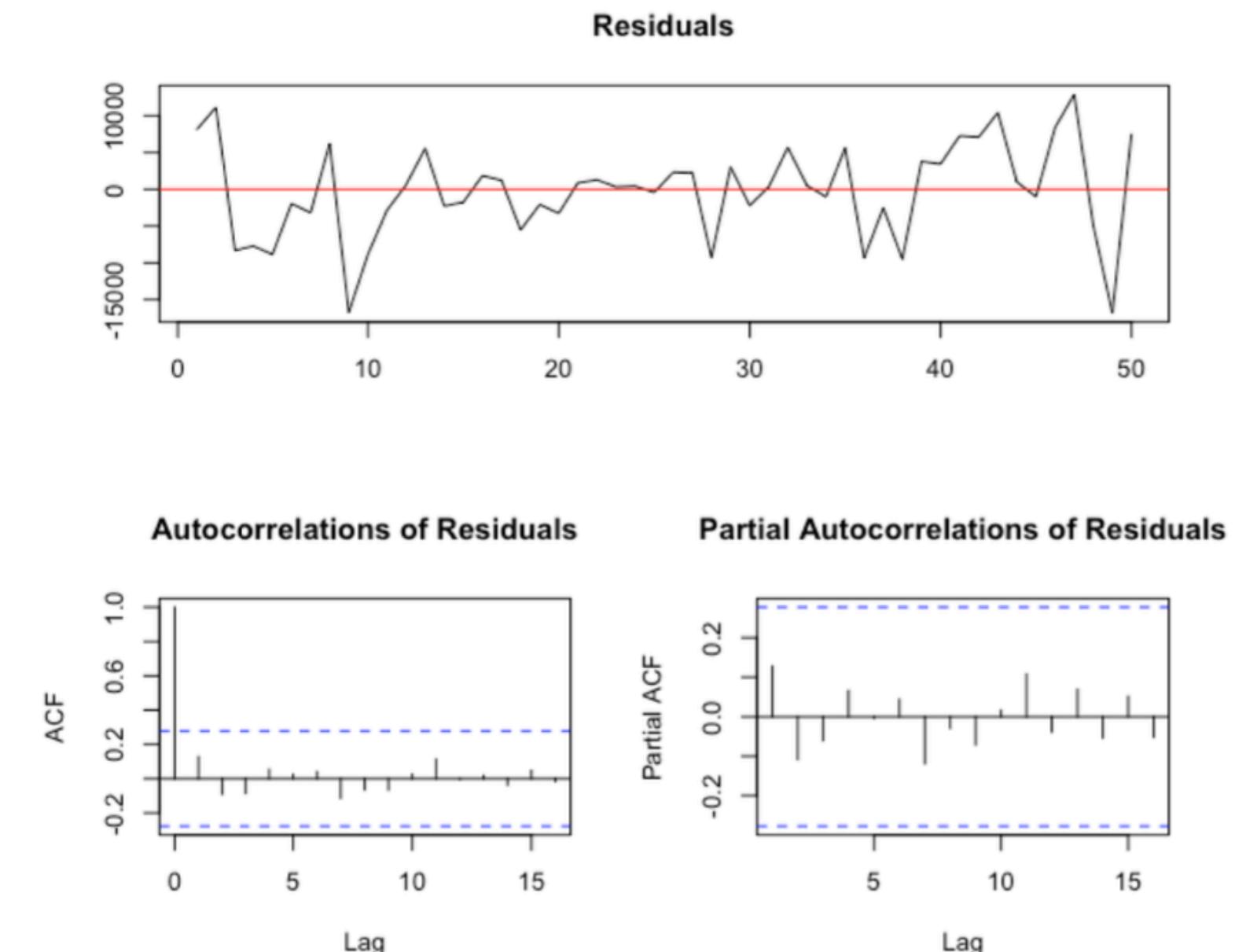
Interprétation

D'après l'ACF tous les valeurs des coefficients d'autocorrélation semblent se situer dans l'intervalle de confiance mis à part celle à l'ordre 0. Il ne semble donc pas y avoir d'autocorrélation.

Dans le PACF, cela semble se confirmer car l'ensemble des coefficients d'autocorrélation partiels se situent dans l'intervalle de confiance.

Cela confirme la conclusion du test de Dickey-Fuller : Le PGD est DS.

Remarque : Il n'est pas nécessaire dans ce cas là d'effectuer Dickey-Fuller Augmenté.



Zivot et Andrews : "Both"

Pour connaître le lag optimal dans Zivot et Andrews on part de la formule de Schwert qui nous donne un pmax de 10. On part donc de lag = 10 puis on diminue le nombre de γ de tel sorte que le dernier ait une statistique t en valeur absolue < 1.6 et l'avant dernier une statistique t en valeur absolue > 1.6 . Pour le Petrole, cela nous donne $p = 8$

Le coefficient δ_1 n'est pas significatif mais comme δ_2 est significatifs nous gardons la spécification « Both ». En effet, nous ne pouvons pas enlever le coefficient associé à DU (δ_1) sans enlever le coefficient associé à DT (δ_2).

la statistique calculée est de -4.11 supérieur à la valeur critique à 5% étant de -5.08 donc on accepte H_0 .

Interprétation :

Sous H_0 , on a DS sans changement structurel. Avec un pic à la 41^{ème} position. Le pic s'est donc produit en 2012. Cette année là, des changements dans la réglementation environnementale ont été effectué, notamment en ce qui concerne l'exploitation des sables bitumineux en Alberta. Des problèmes liés aux infrastructures, tels que des pannes dans les pipelines ont aussi pu avoir un impact

```
> summary(ur.za(Petrole, model="both", lag=8))
```

```
#####
# Zivot-Andrews Unit Root Test #
#####
```

Call:
lm(formula = testmat)

Residuals:

	Min	1Q	Median	3Q	Max
	-10149.4	-2472.5	-688.1	2069.2	9921.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.669e+04	7.035e+03	2.373	0.02448 *
y.l1	4.278e-01	1.392e-01	3.073	0.00458 **
trend	1.270e+03	2.535e+02	5.010	2.47e-05 ***
y.dl1	2.287e-01	1.613e-01	1.418	0.16689
y.dl2	-1.903e-01	1.569e-01	-1.213	0.23494
y.dl3	-1.748e-01	1.775e-01	-0.985	0.33283
y.dl4	-2.098e-02	1.813e-01	-0.116	0.90869
y.dl5	2.180e-02	1.878e-01	0.116	0.90838
y.dl6	-2.426e-01	1.731e-01	-1.402	0.17165
y.dl7	-2.759e-01	1.616e-01	-1.708	0.09841 .
y.dl8	5.349e-02	1.609e-01	0.332	0.74198
du	6.636e+03	5.151e+03	1.288	0.20782
dt	3.754e+03	1.252e+03	2.997	0.00554 **

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

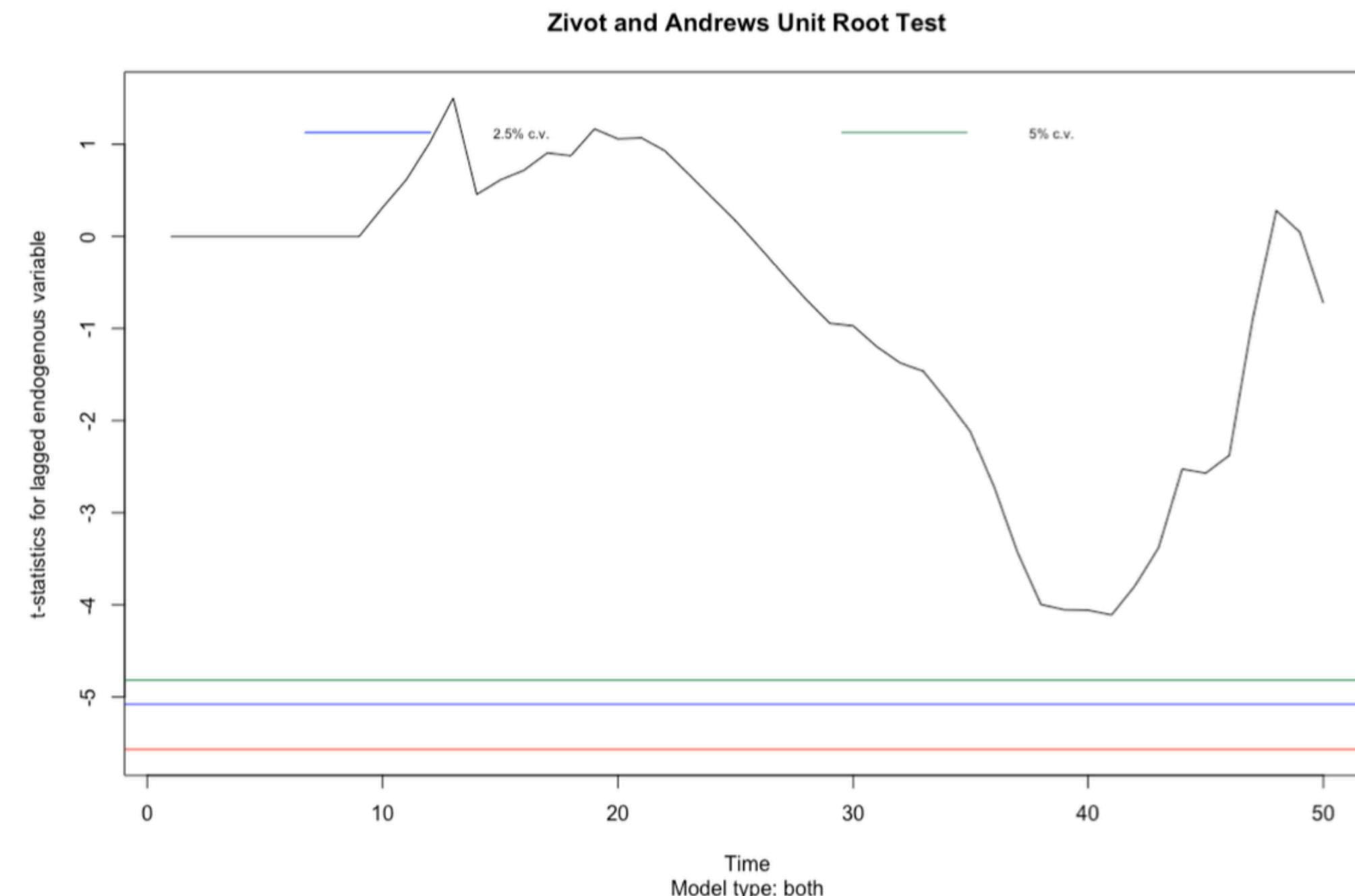
Residual standard error: 4594 on 29 degrees of freedom
(9 observations deleted due to missingness)

Multiple R-squared: 0.991, Adjusted R-squared: 0.9873
F-statistic: 267 on 12 and 29 DF, p-value: < 2.2e-16

Teststatistic: -4.1111
Critical values: 0.01= -5.57 0.05= -5.08 0.1= -4.82

Potential break point at position: 41

Zivot et Andrews : "Both"



Lee et Strazicich : 1 date de rupture

La valeur de la statistique de test est de -8.91. λ est estimé à 0.6 qui n'apparaît pas dans le tableau des valeurs critiques on prend donc la valeur la plus proche 0.5. La valeur critique au seuil de risque de 5% vaut -4.51 :

$$-8.91 < -4.51$$

Donc on rejette H_0 donc le PGD qui a généré Petrole est TS avec un changement structurel dans la constante et la pente de la partie déterministe de la série.

pour finir un seul λ a été introduit dans le modèle pour tenir compte de l'autocorrélation.

```
> y <- Petrole
> myBreaks <- 1
> myModel <- "break"
> myLags <- 4 # car 5 renvoie une erreur
> myLS_test <- ur.ls(y=y, model = myModel, breaks = myBreaks, lags = myLags,
  method = "GTOS", pn = 0.1, print.results = "print" )
[1] -8.917015
[1] "First possible structural break at position: 32"
[1] "The location of the first break - lambda_1: 0.6 , with the number of total observations: 51"
Critical values - Break model:
      lambda    1%     5%    10%
[1,] 0.1 -5.11 -4.50 -4.21
[2,] 0.2 -5.07 -4.47 -4.20
[3,] 0.3 -5.15 -4.45 -4.18
[4,] 0.4 -5.05 -4.50 -4.18
[5,] 0.5 -5.11 -4.51 -4.17
[1] "Number of lags determined by general-to-specific lag selection: 1"
Runtime:
Time difference of 0.0004862507 mins
```

Lee et Strazicich : 1 date de rupture

La première rupture structurelle possible est à la position 32 autrement dit en 2003.

Cette année là, les réglementations environnementales ont évolué, telles que les normes d'émissions ou les exigences de réhabilitation des sites, auraient pu avoir un impact sur la production de pétrole. Ainsi que les événements géopolitiques tels que la guerre en Irak ou des tensions au Moyen-Orient auraient pu entraîner des variations importantes des prix du pétrole influençant la production pétrolières au Canada.

On teste maintenant *Lee et Strazicich* avec 2 dates de rupture.

Lee et Strazicich : 2 dates de rupture

```
> myBreaks <- Z
> myLS_test <- ur.ls(y=y , model = myModel, breaks = myBreaks, lags = myLags, method = "GTOS",pn = 0.1,
  print.results = "print" )
[1] -11.57808
[1] "First possible structural break at position: 20"
[1] "The location of the first break - lambda_1: 0.4 , with the number of total observations: 51"
[1] "Second possible structural break at position: 30"
[1] "The location of the second break - lambda_2: 0.6 , with the number of total observations: 51"
Critical values:
      Break 2 - 0.4 - 1% Break 2 - 0.4 - 5%
Break 1 - 0.2      -6.16      -5.59
Break 1 - 0.4      NA        NA
Break 1 - 0.6      NA        NA
      Break 2 - 0.4 - 10% Break 2 - 0.6 - 1%
Break 1 - 0.2      -5.27      -6.41
Break 1 - 0.4      NA        -6.45
Break 1 - 0.6      NA        NA
      Break 2 - 0.6 - 5% Break 2 - 0.6 - 10%
Break 1 - 0.2      -5.74      -5.32
Break 1 - 0.4      -5.67      -5.31
Break 1 - 0.6      NA        NA
      Break 2 - 0.8 - 1% Break 2 - 0.8 - 5%
Break 1 - 0.2      -6.33      -5.71
Break 1 - 0.4      -6.42      -5.65
Break 1 - 0.6      -6.32      -5.73
      Break 2 - 0.8 - 10%
Break 1 - 0.2      -5.33
Break 1 - 0.4      -5.32
Break 1 - 0.6      -5.32
[1] "Number of lags determined by general-to-specific lag selection: 1"
Runtime:
Time difference of 0.007990551 mins
```

La valeur de la statistique de test est de -11.578.

λ_1 est estimé à 0.4 ce qui donne $T_{B1} = 20$ ce qui correspond à 1991 donc la valeur critique dans break 1 avec valeur la plus proche 0.4 donc entre 0.2 et 0.4 (entre 1ière et seconde ligne).

λ_2 est estimé à 0.6 ce qui donne $T_{B2} = 30$ ce qui correspond à 2001 donc la valeur critique dans break 2 à 5% avec la valeur = 0.6

La valeur critique est à l'intersection de break1 = 0.4 et break 2 = 0.6 à 5% soit -5.67.

$$-11.57 < -5.67$$

Donc on rejette H_0 donc le PGD qui a générée Petrole est TS avec changements structurels.

Lee et Strazicich : 2 dates de rupture

Remarque : Comme le test avec 2 dates est moins puissant que celui avec 1 seule date de rupture on garde la conclusion de LS avec 1 seule date.

Les deux dates de rupture structurelle possible sont 1991 et 2001.

En 1991, il y a eu la guerre du Golfe de janvier à février qui a entraîné des perturbation significative de la production de pétrole dans la région du Golfe Persique. Cet événement conduit à des fluctuations importantes des prix du pétrole sur le marché mondial, influençant les décisions de production au Canada. De plus, les investissements dans les infrastructures pétrolières, tels que les pipelines et les installations de traitement, ainsi que les retards ou les problèmes rencontrés dans ces projets, auraient influencé la capacité de production du pays.

En 2001, des événements tels que les attentats du 11 septembre 2001 aux États-Unis ont provoqué des perturbations économiques mondiales, ce qui a pu influencer les prix du pétrole. Les politiques énergétiques du gouvernement canadien, y compris les réglementations environnementales et les incitations fiscales pour l'industrie pétrolière, auraient influencé la production. Les décisions politiques concernant l'exploration et l'exploitation des ressources pétrolières et gazières auraient également joué un rôle.

Lee et Strazicich : 2 dates de rupture

Jusque là, nous avons trouvé le **PGD DS** avec les tests de *Dickey-Fuller* et *Zivot et Andrews*. Ainsi que le **PGD TS** avec *Lee et Strazicich*.

Le nombre de donnée qu'on étudie est restreint, nous en avons 50. Afin d'être sûr de la valeur de notre PGD, nous allons effectuer le test de Lee et Strazicich avec Bootstrap (méthode avec rééchantillonnage) qui sera le plus approprié et qui nous permettra de trancher entre les conclusions jusque là effectué.

Lee et Strazicich avec Bootstrap : 1 date de rupture

La valeur de la statistique de test est de -3.28. λ est estimé à 0.7 qui n'apparaît pas dans le tableau des valeurs critiques on prend donc la valeur la plus proche 0.5. La valeur critique au seuil de risque de 5% vaut -4.51 :

$$-3.288 > -4.51$$

Donc on accepte H_0 donc le PGD qui a générée Petrole est DS.

pour finir quatre λ ont été introduit dans le modèle pour tenir compte de l'autocorrélation.

```
> myParallel_LS <- ur.ls.bootstrap(y=Petrole , model = myModel, breaks = myBreaks, lags = myLags, method = "Fixed",pn = 0.1, critval = "bootstrap", print.results = "print")
[[1]]
[1] -3.288709

[1] "First possible structural break at position: 35"
[1] "The location of the first break - lambda_1: 0.7 , with the number of total observations: 51"
Critical values - Break model:
  lambda    1%    5%   10%
[1,] 0.1 -5.11 -4.50 -4.21
[2,] 0.2 -5.07 -4.47 -4.20
[3,] 0.3 -5.15 -4.45 -4.18
[4,] 0.4 -5.05 -4.50 -4.18
[5,] 0.5 -5.11 -4.51 -4.17
[1] "Number of lags used: 4"
Runtime:
Time difference of 0.002265803 mins
```

Lee et Strazicich avec Bootstrap : 1 date de rupture

La première rupture structurelle possible est à la position 35 autrement dit en 2006.

Cette année là, l'exploitation des sables bitumineux en Alberta était déjà une composante importante de la production de pétrole au Canada. Les progrès technologiques dans l'extraction des sables bitumineux ainsi que les investissements dans cette industrie ont probablement influencé la production de pétrole à cette époque. En plus, des projets d'expansion ou de construction de nouvelles infrastructures pourraient avoir été entrepris pour répondre à la demande croissante de transport de pétrole.

On teste maintenant *Lee et Strazicich Bootstrap* avec 2 dates de rupture.

Lee et Strazicich avec Bootstrap : 2 dates de rupture

```
[[1]]  
[1] -4.512272  
  
[1] "First possible structural break at position: 9"  
[1] "The location of the first break - lambda_1: 0.2 , with the number of total observations: 51"  
[1] "Second possible structural break at position: 37"  
[1] "The location of the second break - lambda_2: 0.7 , with the number of total observations: 51"  
Critical values:  
Break 2 - 0.4 - 1% Break 2 - 0.4 - 5% Break 2 - 0.4 - 10% Break 2 - 0.6 - 1%  
Break 1 - 0.2      -6.16        -5.59        -5.27        -6.41  
Break 1 - 0.4      NA          NA          NA          -6.45  
Break 1 - 0.6      NA          NA          NA          NA  
Break 2 - 0.6 - 5% Break 2 - 0.6 - 10% Break 2 - 0.8 - 1% Break 2 - 0.8 - 5%  
Break 1 - 0.2      -5.74        -5.32        -6.33        -5.71  
Break 1 - 0.4      -5.67        -5.31        -6.42        -5.65  
Break 1 - 0.6      NA          NA          -6.32        -5.73  
Break 2 - 0.8 - 10%  
Break 1 - 0.2      -5.33  
Break 1 - 0.4      -5.32  
Break 1 - 0.6      -5.32  
[1] "Number of lags used: 4"  
Runtime:  
Time difference of 0.004104368 mins
```

La valeur de la statistique de test est -4.512.

λ_1 est estimé à 0.4 ce qui donne $T_{B1} = 9$ ce qui correspond à 1980 donc la valeur critique dans break 1 avec valeur de 0.2.

λ_2 est estimé à 0.7 ce qui donne $T_{B2} = 37$ ce qui correspond à 2008 (crise économique) donc la valeur critique dans break 2 à 5% avec la valeur entre 0.6 et 0.8.

La valeur critique est à l'intersection de break1 = 0.2 et break 2 entre 0.6 et 0.8 à 5% soit -5.72.

$$-4.512 > -5.72$$

Donc on accepte H_0 donc le PGD qui a généré Petrole est DS.

Lee et Strazicich avec Bootstrap : 2 dates de rupture

Les deux dates de rupture structurelle possible sont 1980 et 2008.

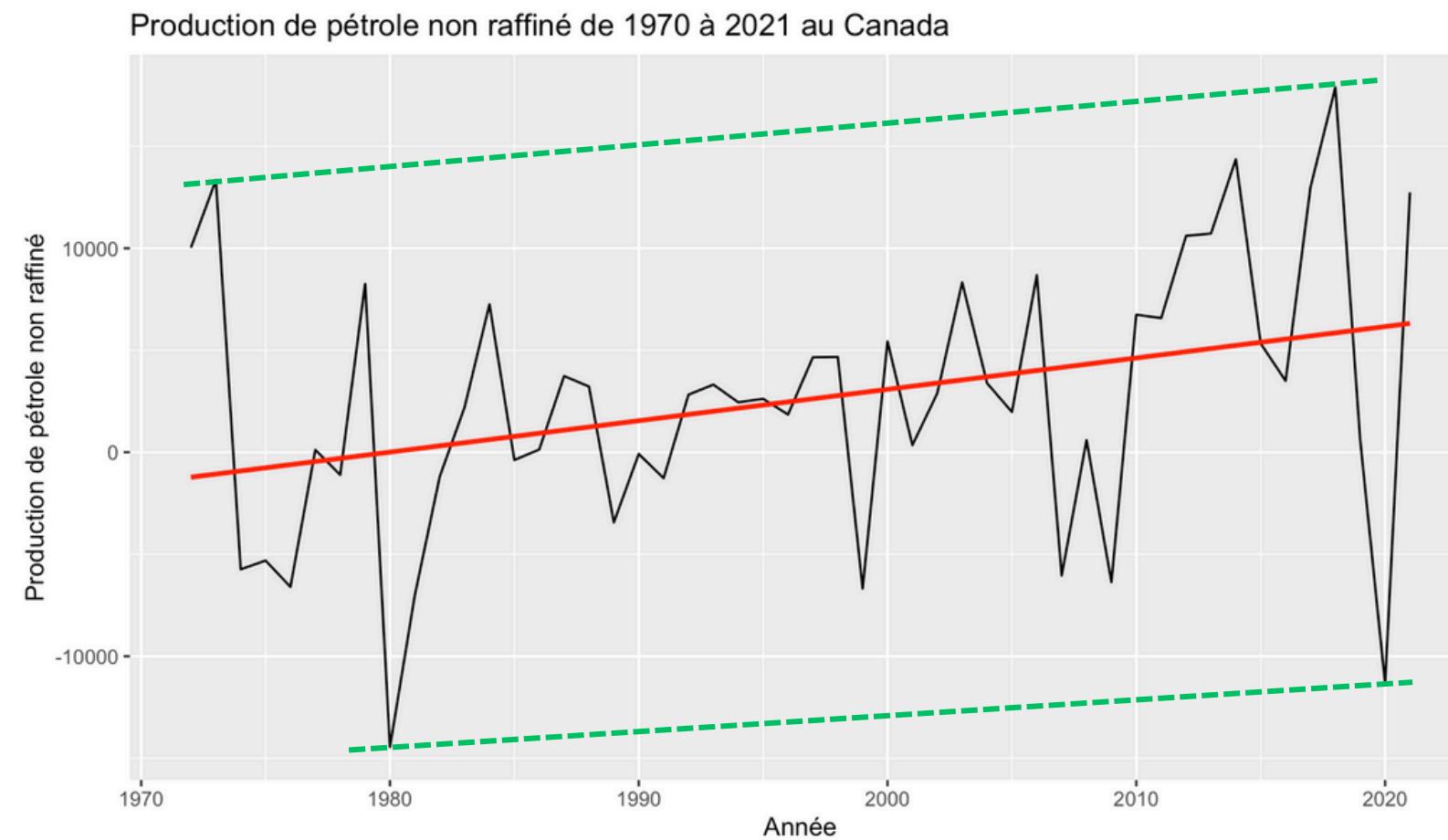
En 1980, la crise pétrolière des années 1970, notamment le choc pétrolier de 1979, a eu un impact significatif sur le marché mondial du pétrole. Les prix élevés du pétrole et les incertitudes géopolitiques associées à cette crise ont pu affecter la production de pétrole au Canada.

En 2008, La crise financière qui a débuté en 2008 a entraîné une récession mondiale, ce qui a conduit à une baisse de la demande mondiale de pétrole. Cette diminution de la demande a pu avoir un impact sur la production de pétrole au Canada, car les entreprises ont peut-être réduit leur production pour s'adapter à la baisse de la demande. De plus, les prix du pétrole ont atteint des niveaux historiquement élevés, dépassant même les 140 dollars le baril avant de chuter brutalement vers la fin de l'année en raison de la crise financière. Ces fluctuations des prix ont pu influencer les décisions de production des entreprises pétrolières canadiennes.

Remarque : Comme nous travaillons sur des échantillons assez petits nous allons privilégier les résultats du test Lee et Strazicich avec Bootstrap donc c'est cette conclusion que je garde. Pour conclure ces tests le PGD est DS on va donc procéder à une différentiation.

Chronogramme de la production annuelle de pétrole non raffiné de 1971 à 2021 différencié à l'ordre 1

Nous traçons tout d'abord le chronogramme de notre série différencié à l'ordre 1 représentant notre variable quantitative étant la production annuelle de pétrole non raffiné (sur l'axe des Y) en fonction du temps, de 1970 à 2021 (sur l'axe des X).



Interprétation

La courbe rouge de la série temporelle met en évidence une tendance à la hausse.

La série temporelle n'a pas d'effets saisonniers étant donné qu'on possède qu'une seule valeur par an.

À l'aide des deux droites vertes nous observons qu'elles sont quasiment parallèles entre elles. On en déduit une légère homoscédasticité, la variance des données augmente peu avec le temps.

Dickey-Fuller : “Trend”

Consiste à estimer par les MCO :

$$\Delta dPetrole_t = (\rho - 1)dPetrole_{t-1} + \beta_0 + \beta_1 \text{tendance}_t + \varepsilon_t$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \text{ et } \beta_1 = 0 \text{ versus } H_a : |\rho| < 1 \text{ et } \beta_1 \neq 0$$

Tout d'abord nous regardons la p-value associé à β_1 qui est de 0.014 < 0.05 (seuil de significativité).

On en déduit que β_1 est significatif. Testons maintenant $H_0 : \rho - 1 = 0$ versus $H_a : |\rho| < 1$. La règle de décision est si la statistique t calculé pour $\rho - 1$ > à la valeur critique donnée à l'intersection de la ligne tau3 et de la colonne 5%, vous acceptez H_0 .

La statistique t est de $-6.3535 < -3.50$ (la valeur critique). On rejette donc H_0

Le processus est TS d'après Dickey-Fuller.

```
> summary(ur.df(dPetrole,type="trend",lag=0))

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt)

Residuals:
    Min      1Q  Median      3Q     Max 
-17502.9 -2920.7   -69.3  3244.1 14334.5 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2070.6127  1834.6252 -1.129   0.265  
z.lag.1       -0.9102    0.1433 -6.354 8.53e-08 *** 
tt            170.1985   66.6155  2.555   0.014 *  
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 6300 on 46 degrees of freedom
Multiple R-squared:  0.4705,    Adjusted R-squared:  0.4474 
F-statistic: 20.43 on 2 and 46 DF,  p-value: 4.462e-07

Value of test-statistic is: -6.3535 13.6244 20.4347

Critical values for test statistics:
          1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Test de racine unitaire

Les conclusions du test DF ne sont valables que si les Et ne sont pas autocorrélés. Testons cela avec l'ACF et le PACF des résidus des régressions DF.

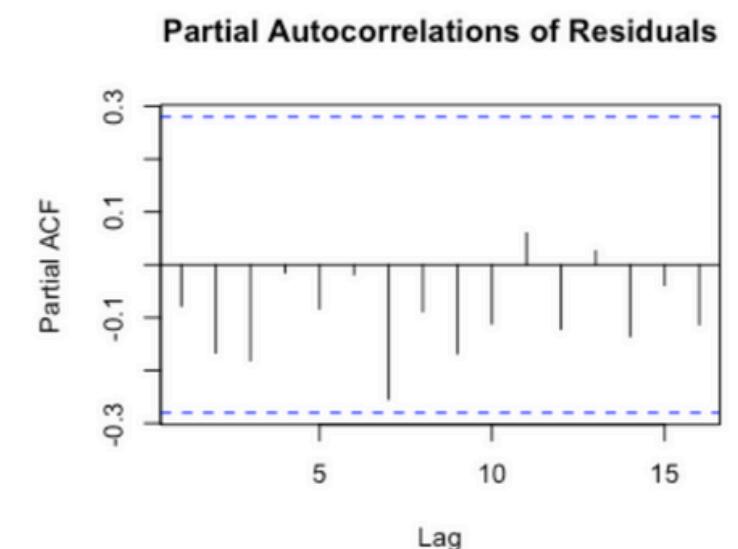
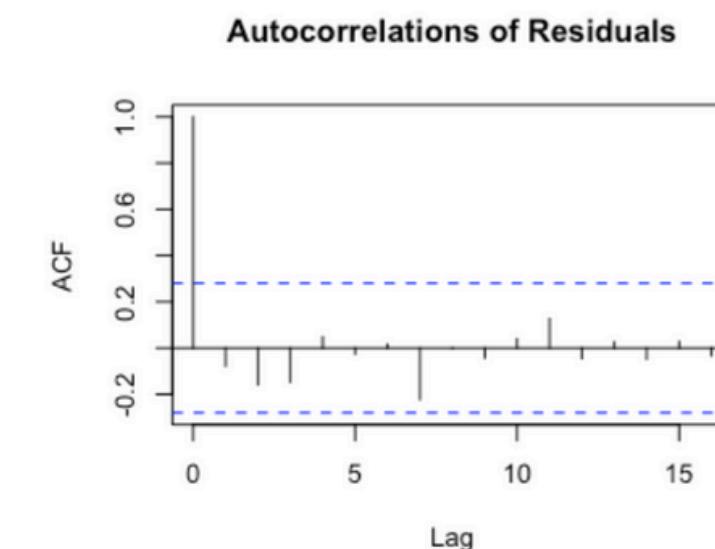
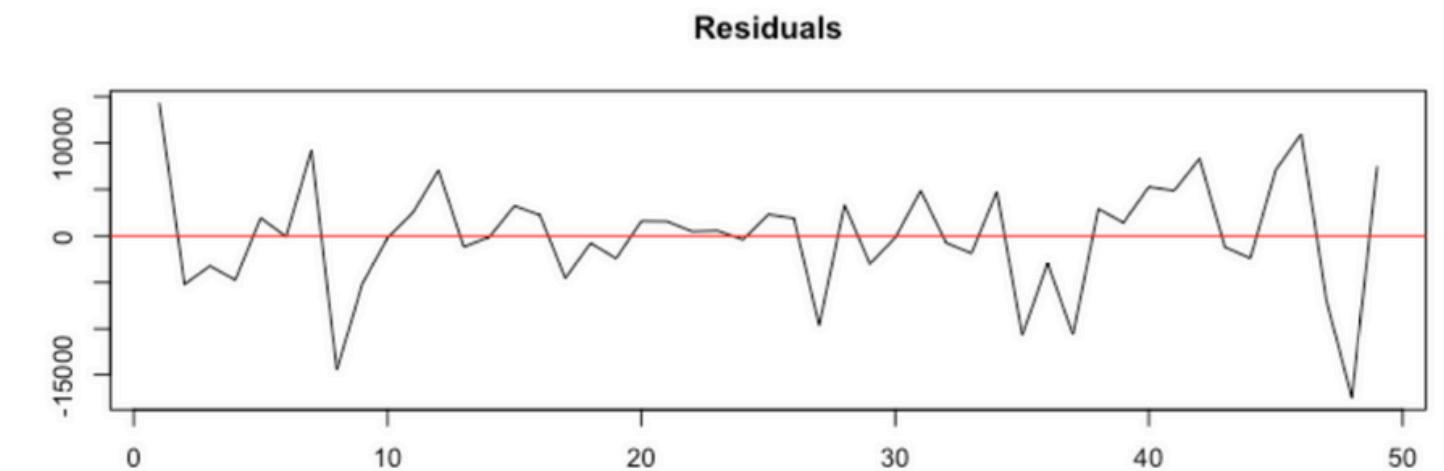
Interprétation

D'après l'ACF tous les valeurs des coefficients d'autocorrélation semblent se situer dans l'intervalle de confiance mis à part celle à l'ordre 0. Il ne semble donc pas y avoir d'autocorrélation.

Dans le PACF, cela semble se confirmer car l'ensemble des coefficients d'autocorrélation partiels se situent dans l'intervalle de confiance.

Cela confirme la conclusion du test de Dickey-Fuller :
Le PGD est DS.

Il n'est donc pas nécessaire d'effectuer Dickey-Fuller Augmenté.



Zivot et Andrews : "Both"

Pour connaître le lag optimal dans Zivot et Andrews on part de la formule de Schwert qui nous donne un pmax de 10. On part donc de lag = 10 puis on diminue le nombre de γ de tel sorte que le dernier ait une statistique t en valeur absolue < 1.6 et l'avant dernier une statistique t en valeur absolue > 1.6 . Pour le Petrole, cela nous donne $p = 7$

Les coefficients δ_1 et δ_2 sont significatifs nous gardons la spécification « Both ».

la statistique calculée est de -5.5161 est inférieur à la valeur critique à 5% étant de -5.08 donc on rejette H_0 .

Interprétation :

Sous H_a , on a TS avec un changement structurel. Avec un pic à la 35^{ème} position. Le pic s'est donc produit en 2006.

```
> summary(ur.za(dPetrole, model="both",lag=7))
```

```
#####
# Zivot-Andrews Unit Root Test #
#####
```

```
Call:  
lm(formula = testmat)
```

Residuals:

Min	1Q	Median	3Q	Max
-12827	-2943	1242	2284	11668

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.505e+04	3.837e+03	-3.922	0.000473 ***
y.l1	-2.689e+00	6.687e-01	-4.021	0.000361 ***
trend	8.235e+02	1.817e+02	4.532	8.69e-05 ***
y.dl1	2.547e+00	6.037e-01	4.219	0.000208 ***
y.dl2	2.098e+00	5.491e-01	3.820	0.000625 ***
y.dl3	1.607e+00	4.694e-01	3.424	0.001804 **
y.dl4	1.368e+00	3.801e-01	3.599	0.001134 **
y.dl5	1.129e+00	2.976e-01	3.792	0.000673 ***
y.dl6	7.195e-01	2.238e-01	3.215	0.003117 **
y.dl7	2.487e-01	1.684e-01	1.477	0.150103
du	-1.210e+04	3.908e+03	-3.096	0.004229 **
dt	1.317e+03	4.247e+02	3.102	0.004167 **

Signif. codes:	0 ****	0.001 **	0.01 *	0.05 .
	0.1 ' '	1		

Residual standard error: 5168 on 30 degrees of freedom
(8 observations deleted due to missingness)

Multiple R-squared: 0.5513, Adjusted R-squared: 0.3867
F-statistic: 3.35 on 11 and 30 DF, p-value: 0.00415

Teststatistic: -5.5161

Critical values: 0.01= -5.57 0.05= -5.08 0.1= -4.82

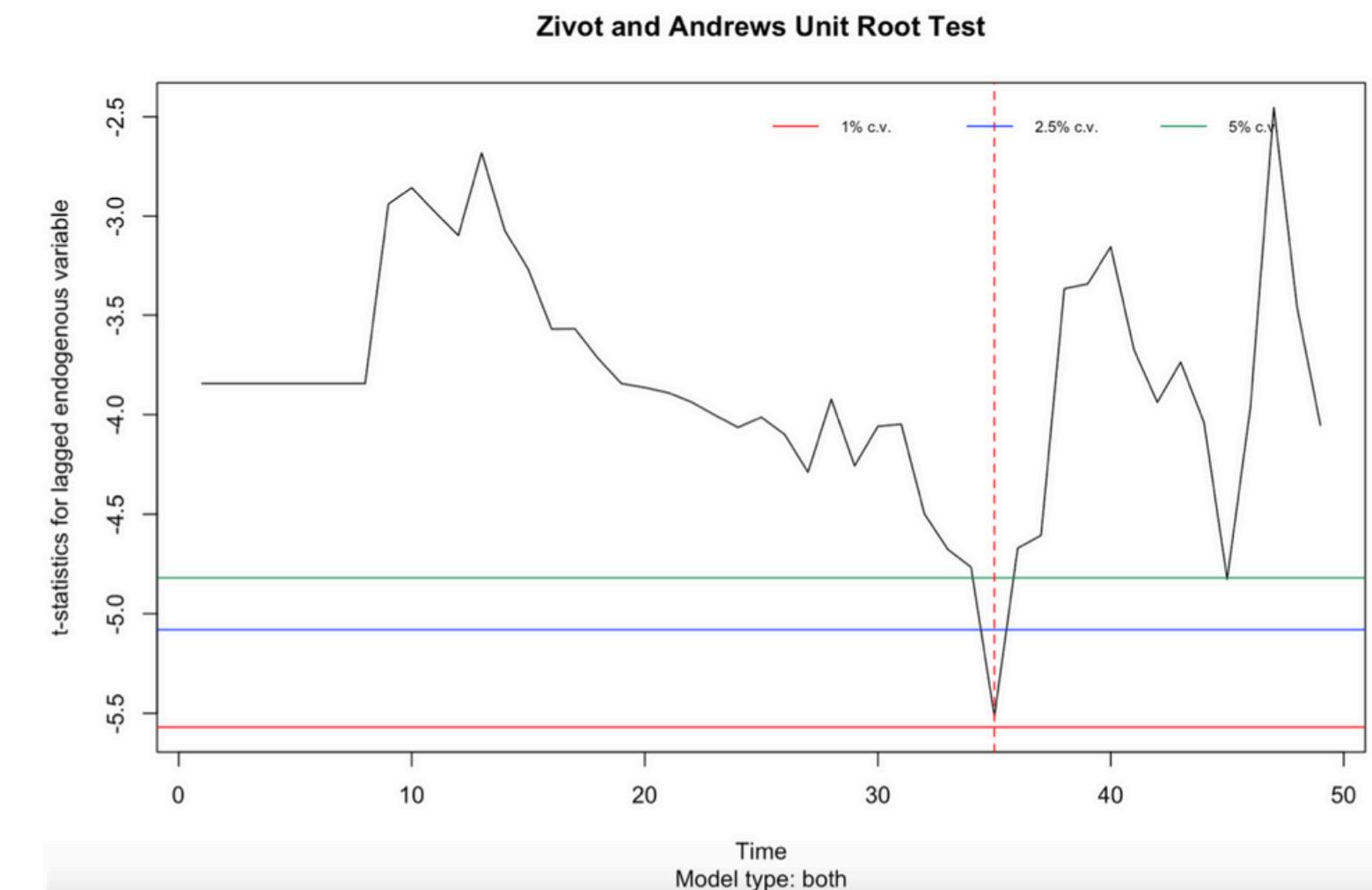
Potential break point at position: 35

Zivot et Andrews : “Both”

En 2006, les prix du pétrole ont connu une période de forte hausse en 2006 en raison de la demande mondiale croissante et des préoccupations concernant l'offre. Il y a probablement eu des investissements importants dans les infrastructures pétrolières au Canada, tels que de nouveaux pipelines et des installations de traitement.

Ces investissements visent à accroître la capacité de production et à faciliter le transport du pétrole extrait des régions productrices vers les marchés.

On effectue maintenant les tests de *Lee et Strazicich avec 1 et 2 dates de rupture*.



Lee et Strazicich : 1 date de rupture

```
[1] -6.521028
[1] "First possible structural break at position: 12"
[1] "The location of the first break - lambda_1: 0.2 , with the number of total observations: 50"
Critical values - Break model:
      lambda    1%     5%    10%
[1,] 0.1 -5.11 -4.50 -4.21
[2,] 0.2 -5.07 -4.47 -4.20
[3,] 0.3 -5.15 -4.45 -4.18
[4,] 0.4 -5.05 -4.50 -4.18
[5,] 0.5 -5.11 -4.51 -4.17
[1] "Number of lags determined by general-to-specific lag selection: NA"
Runtime:
Time difference of 0.001035217 mins
```

La valeur de la statistique de test est de -6.52. λ est estimé à 0.2. La valeur critique au seuil de risque de 5% vaut -4.47 :

$$-8.91 < -4.47$$

Donc on rejette H_0 donc le PGD qui a générée d'Pétrole est TS avec un changement structurel dans la constante et la pente de la partie déterministe de la série.

Lee et Strazicich : 1 date de rupture

La première rupture structurelle possible est à la position 12 autrement dit en 1983.

Cette année là, le marché pétrolier mondial était caractérisé par une période de volatilité des prix. Une surproduction mondiale de pétrole, combinée à une diminution de la demande, a entraîné une chute significative des prix du pétrole tout au long de l'année.

On teste maintenant *Lee et Strazicich* avec 2 dates de rupture.

Lee et Strazicich : 2 dates de rupture

```
> myBreaks <- 2
> myLS_test <- ur.ls(y=y , model = myModel, breaks = myBreaks, lags = myLags, method = "GTOS",pn = 0.1,
+ print.results = "print" )
[1] -11.57808
[1] "First possible structural break at position: 20"
[1] "The location of the first break - lambda_1: 0.4 , with the number of total observations: 51"
[1] "Second possible structural break at position: 30"
[1] "The location of the second break - lambda_2: 0.6 , with the number of total observations: 51"
Critical values:
      Break 2 - 0.4 - 1% Break 2 - 0.4 - 5%
Break 1 - 0.2          -6.16          -5.59
Break 1 - 0.4          NA            NA
Break 1 - 0.6          NA            NA
      Break 2 - 0.4 - 10% Break 2 - 0.6 - 1%
Break 1 - 0.2          -5.27          -6.41
Break 1 - 0.4          NA            -6.45
Break 1 - 0.6          NA            NA
      Break 2 - 0.6 - 5% Break 2 - 0.6 - 10%
Break 1 - 0.2          -5.74          -5.32
Break 1 - 0.4          -5.67          -5.31
Break 1 - 0.6          NA            NA
      Break 2 - 0.8 - 1% Break 2 - 0.8 - 5%
Break 1 - 0.2          -6.33          -5.71
Break 1 - 0.4          -6.42          -5.65
Break 1 - 0.6          -6.32          -5.73
      Break 2 - 0.8 - 10%
Break 1 - 0.2          -5.33
Break 1 - 0.4          -5.32
Break 1 - 0.6          -5.32
[1] "Number of lags determined by general-to-specific lag selection: 1"
Runtime:
Time difference of 0.007990551 mins
```

La valeur de la statistique de test est -11.578.

λ_1 est estimé à 0.4 ce qui donne $T_{B1} = 20$ ce qui correspond à 1991 donc la valeur critique dans break 1 avec valeur la plus proche 0.4 donc entre 0.2 et 0.4 (entre 1ière et seconde ligne).

λ_2 est estimé à 0.6 ce qui donne $T_{B2} = 30$ ce qui correspond à 2001 donc la valeur critique dans break 2 à 5% avec la valeur = 0.6

La valeur critique est à l'intersection de break1 = 0.4 et break 2 = 0.6 à 5% soit - 5.67.

$$-11.57 < -5.67$$

Donc on rejette H_0 donc le PGD qui a générée d'Petrole est TS avec changements structurels.

Comme le test avec 2 dates est moins puissant que celui avec 1 seule date de rupture on garde la conclusion de LS avec 1 seule date

Les deux dates de rupture structurelle possible sont 1991 et 2001 (cf. page 15).

Lee et Strazicich : 2 dates de rupture

Jusque là, nous avons trouvé le **PGD TS** avec les tests de *Dickey-Fuller*, *Zivot et Andrews* et *Lee et Strazicich*.

Le nombre de donnée qu'on étudie est restreint, nous en avons 50. Afin d'être sûr de la valeur de notre PGD, nous allons effectuer le test de Lee et Strazicich avec Bootstrap (méthode avec rééchantillonnage) qui sera le plus approprié et qui nous permettra de trancher entre les conclusions jusque là effectué.

Lee et Strazicich avec Bootstrap : 1 date de rupture

La valeur de la statistique de test est de -3.99. λ est estimé à 0.1. La valeur critique au seuil de risque de 5% vaut -4.50 :

$$-3.99 > -4.50$$

Donc on accepte H_0 donc le PGD qui a généré d'Petrole est DS.

pour finir quatre λ ont été introduit dans le modèle pour tenir compte de l'autocorrélation.

```
[[1]]  
[1] -3.991365  
  
[1] "First possible structural break at position: 7"  
[1] "The location of the first break - lambda_1: 0.1 , with the number of total observations: 50"  
Critical values - Break model:  
lambda 1% 5% 10%  
[1,] 0.1 -5.11 -4.50 -4.21  
[2,] 0.2 -5.07 -4.47 -4.20  
[3,] 0.3 -5.15 -4.45 -4.18  
[4,] 0.4 -5.05 -4.50 -4.18  
[5,] 0.5 -5.11 -4.51 -4.17  
[1] "Number of lags used: 4"  
Runtime:  
Time difference of 0.0033379 mins
```

Lee et Strazicich avec Bootstrap : 1 date de rupture

La première rupture structurelle possible est à la position 7 autrement dit en 1978.

Cette année là, le Canada était confronté à des décisions politiques telles que la réglementation de l'exploration et de l'exploitation pétrolières, les taxes et les incitations fiscales, ainsi que les politiques d'exportation de pétrole brut. Au cours des années 1970, le monde a connu une pénurie de pétrole due à des événements tels que le premier choc pétrolier de 1973 et le deuxième choc pétrolier de 1979. Ces événements ont entraîné une augmentation significative des prix du pétrole, ce qui a probablement incité le Canada à augmenter sa production pour répondre à la demande croissante.

On teste maintenant *Lee et Strazicich Bootstrap avec 2 dates de rupture*.

Lee et Strazicich avec Bootstrap : 2 dates de rupture

```
> myBreaks <- 2
> myParallel_LS <- ur.ls.bootstrap(y=dPetrole , model = myModel, breaks = myBreaks, lags = myLags, method = "Fixed",pn = 0.
1, critval = "bootstrap", print.results = "print")
[[1]]
[1] -5.491944

[1] "First possible structural break at position: 7"
[1] "The location of the first break - lambda_1: 0.1 , with the number of total observations: 50"
[1] "Second possible structural break at position: 41"
[1] "The location of the second break - lambda_2: 0.8 , with the number of total observations: 50"
Critical values:
      Break 2 - 0.4 - 1% Break 2 - 0.4 - 5% Break 2 - 0.4 - 10% Break 2 - 0.6 - 1% Break 2 - 0.6 - 5%
Break 1 - 0.2          -6.16        -5.59        -5.27        -6.41        -5.74
Break 1 - 0.4            NA          NA          NA        -6.45        -5.67
Break 1 - 0.6            NA          NA          NA          NA          NA
      Break 2 - 0.6 - 10% Break 2 - 0.8 - 1% Break 2 - 0.8 - 5% Break 2 - 0.8 - 10%
Break 1 - 0.2          -5.32        -6.33        -5.71        -5.33
Break 1 - 0.4          -5.31        -6.42        -5.65        -5.32
Break 1 - 0.6            NA        -6.32        -5.73        -5.32
[1] "Number of lags used: 4"
Runtime:
Time difference of 0.004052035 mins
```

La valeur de la statistique de test est -5.49.

λ_1 est estimé à 0.1 ce qui donne $T_{B1} = 7$ ce qui correspond à 1978 (cf. page 30) donc la valeur critique dans break 1 avec valeur de 0.2.

λ_2 est estimé à 0.8 ce qui donne $T_{B2} = 41$ ce qui correspond à 2012 (cf. page 10) donc la valeur critique dans break 2 à 5% avec la valeur 0.8.

La valeur critique est à l'intersection de break1 = 0.2 et break 2 = 0.8 à 5% soit -5.71.

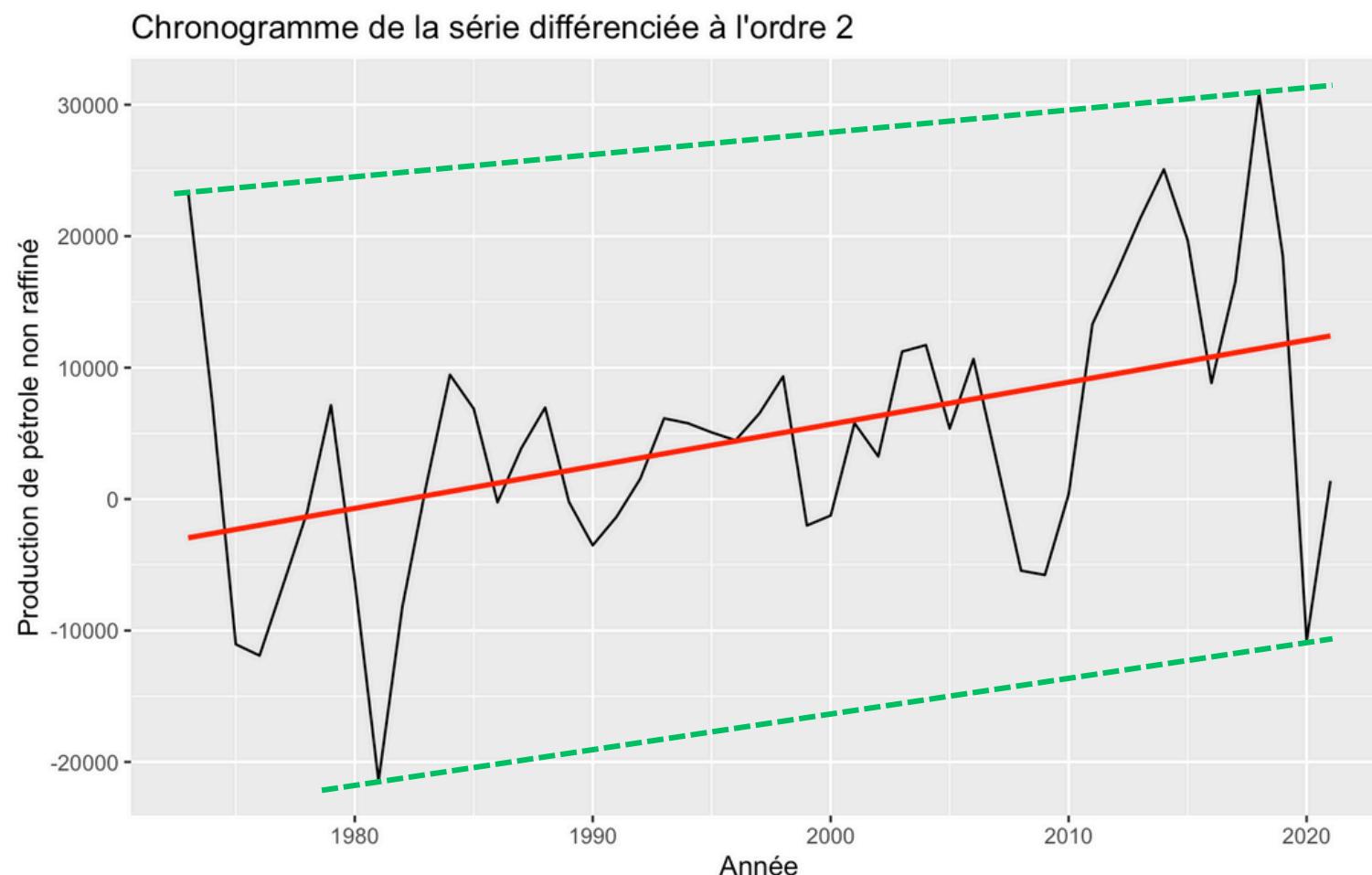
$$-5.49 > -5.71$$

Donc on accepte H_0 donc le PGD qui a générée dPetrole est DS.

Comme nous travaillons sur des échantillons assez petits nous allons privilégier les résultats du test Lee et Strazicich avec Bootstrap donc c'est cette conclusion que je garde. Pour conclure ces tests le PGD est DS on va donc procéder à une nouvelle différentiation.

Chronogramme de la production annuelle de pétrole non raffiné de 1971 à 2021 différenciée à l'ordre 2

Nous traçons tout d'abord le chronogramme de notre série différenciée à l'ordre 1 représentant notre variable quantitative étant la production annuelle de pétrole non raffiné (sur l'axe des Y) en fonction du temps, de 1970 à 2021 (sur l'axe des X).



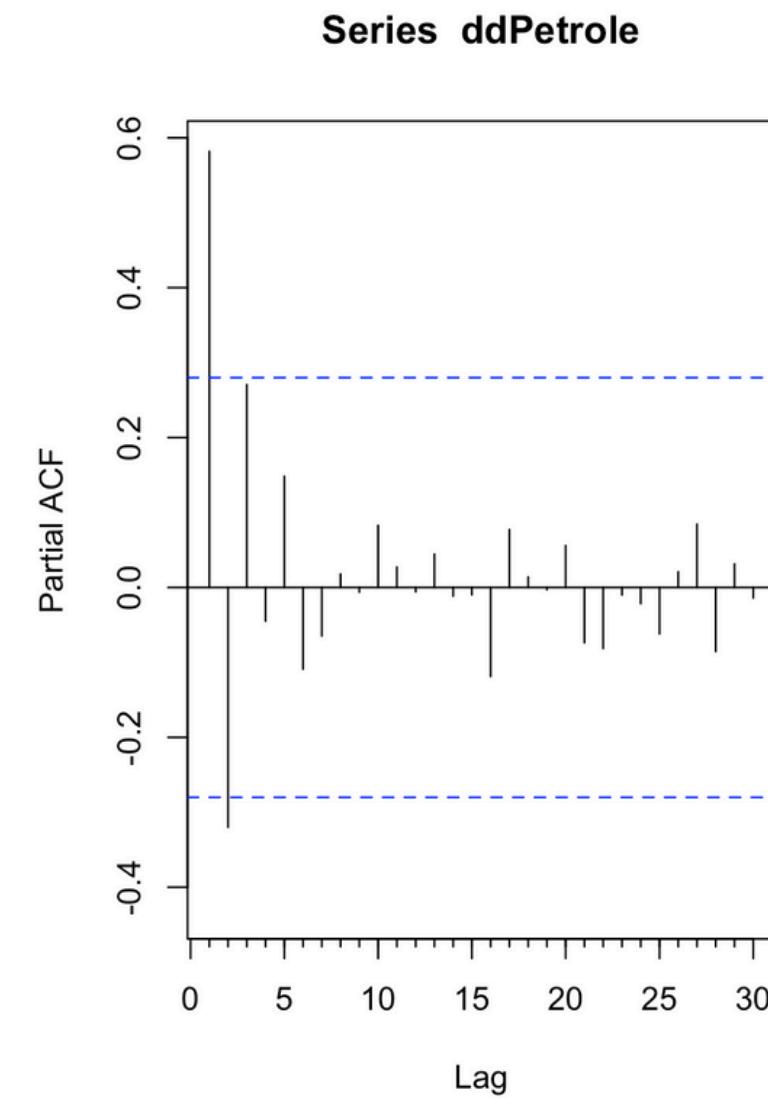
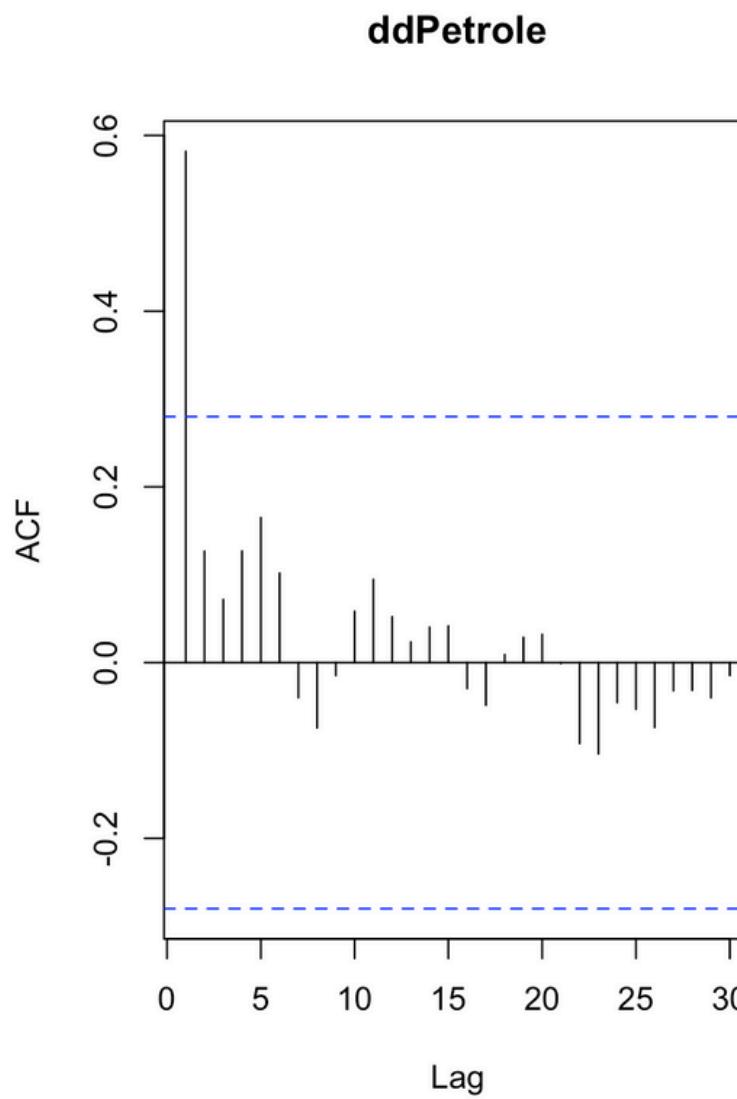
Interprétation

La courbe rouge de la série temporelle met en évidence une tendance à la hausse.
La série temporelle n'a pas d'effets saisonniers étant donné qu'on possède qu'une seule valeur par an.

À l'aide des deux droites vertes nous observons qu'elles sont quasiment parallèles entre elles. On en déduit une légère homoscédasticité, la variance des données augmente peu avec le temps.

Corrélogramme de la série différenciée à l'ordre 2

Détermination de présence de corrélation dans la série



ACF

Pas d'autocorrélation détecter dans l'ACF de la série différenciée à l'ordre 2

PACF

Pour poursuivre l'analyse, la première et la deuxième valeur du PACF sont hors de l'intervalle de confiance avec une forte corrélation de 0.6 et -0.3, suivie de l'absence d'autocorrélation ultérieure

Dickey-Fuller : “Trend”

Consiste à estimer par les MCO :

$$\Delta ddPetrole_t = (\rho - 1)ddPetrole_{t-1} + \beta_0 + \beta_1 \text{tendance}_t + \varepsilon_t$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \text{ et } \beta_1 = 0 \text{ versus } H_a : |\rho| < 1 \text{ et } \beta_1 \neq 0$$

Tout d'abord nous regardons la p-value associé à β_1 qui est de 0.554
 > 0.05 (seuil de significativité)

On en déduit que β_1 n'est pas significatif. Alors la spécification « trend » n'est pas la bonne et nous passons à la spécification « drift »

```
> summary(ur.df(ddPetrole,type="trend",lag=0))
```

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
```

Test regression trend

Call:

```
lm(formula = z.diff ~ z.lag.1 + 1 + tt)
```

Residuals:

Min	1Q	Median	3Q	Max
-20008.7	-3872.2	945.8	5054.7	18186.2

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)		
(Intercept)	-1421.7765	2361.9892	-0.602	0.550		
z.lag.1	-1.4134	0.1507	-9.377	3.75e-12 ***		
tt	50.0102	83.9548	0.596	0.554		

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’	0.1 ‘ ’	1

Residual standard error: 8055 on 45 degrees of freedom
Multiple R-squared: 0.6634, Adjusted R-squared: 0.6484
F-statistic: 44.34 on 2 and 45 DF, p-value: 2.297e-11

Value of test-statistic is: -9.3769 29.6046 44.3375

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-4.15	-3.50	-3.18
phi2	7.02	5.13	4.31
phi3	9.31	6.73	5.61

Dickey-Fuller : "Drift"

Consiste à estimer par les MCO :

$$\Delta ddPetrole_t = (\rho - 1)ddPetrole_{t-1} + \beta_0 + \varepsilon_t$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \text{ et } \beta_0 = 0 \text{ versus } H_a : |\rho| < 1 \text{ et } \beta_0 \neq 0$$

Tout d'abord nous regardons la p-value associé à β_0 qui est de 0.865
 > 0.05 (seuil de significativité)

On en déduit que β_0 n'est pas significatif. Alors la spécification « drift » n'est pas la bonne et nous passons à la spécification « none »

```
> summary(ur.df(ddPetrole,type="drift",lag=0))  
#####  
# Augmented Dickey-Fuller Test Unit Root Test #  
#####
```

Test regression drift

Call:

```
lm(formula = z.diff ~ z.lag.1 + 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-18926.9	-3451.2	743.3	5618.7	19332.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-197.6719	1156.3217	-0.171	0.865
z.lag.1	-1.4160	0.1496	-9.464	2.27e-12 ***

Signif. codes:	0 ***	0.001 **	0.01 *	0.05 .
	1			

Residual standard error: 7998 on 46 degrees of freedom
Multiple R-squared: 0.6607, Adjusted R-squared: 0.6533
F-statistic: 89.58 on 1 and 46 DF, p-value: 2.274e-12

Value of test-statistic is: -9.4645 44.8586

Critical values for test statistics:

	1pct	5pct	10pct
tau2	-3.58	-2.93	-2.60
phi1	7.06	4.86	3.94

Dickey-Fuller : “None”

Consiste à estimer par les MCO :

$$\Delta Petrole_t = (\rho - 1) Petrole_{t-1} + \varepsilon_t$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \quad \text{versus} \quad H_a : |\rho| < 1$$

Pour savoir si on accepte H_0 ou non, la règle de décision est si la valeur de la statistique t associée à $(\rho - 1)$ > à la valeur critique donnée à l'intersection de la ligne tau1 et de la colonne 5%, on accepte H_0 .

Ici la valeur de la statistique t associée à $(\rho - 1)$ est 3.1899 et la valeur critique est de -1.95. On a bien $-9.569 < -1.95$

On rejette alors $H_0 : \rho - 1 = 0$

Le processus est AR(1) d'après Dickey-Fuller

```
> summary(ur.df(ddPetrole, type="none", lag=0))
```

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
```

Test regression none

Call:

```
lm(formula = z.diff ~ z.lag.1 - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-19099.0	-3645.9	561.4	5404.4	19151.9

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
z.lag.1	-1.4145	0.1478	-9.57	1.3e-12 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7915 on 47 degrees of freedom

Multiple R-squared: 0.6608, Adjusted R-squared: 0.6536

F-statistic: 91.58 on 1 and 47 DF, p-value: 1.299e-12

Value of test-statistic is: -9.5697

Critical values for test statistics:

	1pct	5pct	10pct
tau1	-2.62	-1.95	-1.61

Test de racine unitaire

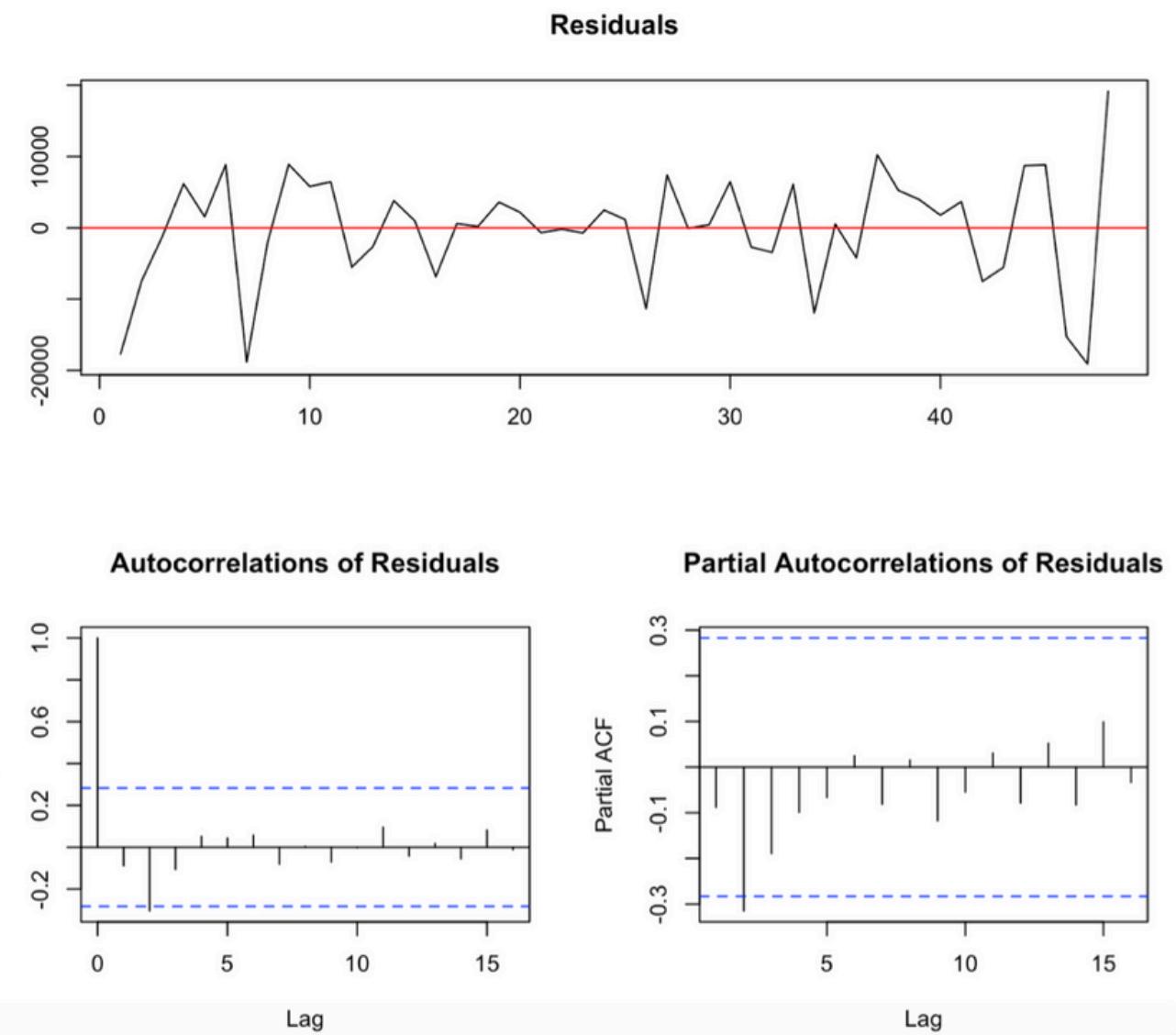
Les conclusions du test DF ne sont valables que si les *epsilons t* et ne sont pas autocorrélés.
Testons cela avec l'ACF et le PACF des résidus des régressions DF.

Interprétation

D'après l'ACF la valeur du lag 2 ne se situe pas dans l'intervalle de confiance. Il semble donc y avoir d'autocorrélation.

Dans le PACF, cela semble se confirmer car le coefficient d'autocorrélation partiels associé au lag 2 se situent hors l'intervalle de confiance.

Cela confirme la nécessité de passer par Dickey-Fuller Augmenté, la conclusion n'est pas fiable avec seulement Dickey-Fuller.



Nombre de retard P

```
> summary(CADFtest(ddPetrole,criterion="MAIC",type="none",max.lag.y=pmax))
Augmented DF test

t-test statistic:          ADF test
p-value:                  -8.230284e+00
Max lag of the diff. dependent variable: 4.106429e-15
                                         0.000000e+00

Call:
dynlm(formula = formula(model), start = obs.1, end = obs.T)

Residuals:
    Min      1Q  Median      3Q     Max 
-19530.0 -3394.8   369.1  3925.1 18855.2 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
L(y, 1) -1.4394    0.1749   -8.23 4.11e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7320 on 37 degrees of freedom
Multiple R-squared:  0.6467, Adjusted R-squared:  0.6372 
F-statistic: NA on NA and NA DF, p-value: NA
```

Le nombre de retards, P , qu'on ajoute dans les régressions pour tenir compte de l'autocorrélation dans les aléas a un impact sur la taille et la puissance des tests.

Tout d'abord il faut calculer la valeur maximale de retard à introduire, p_{\max} , donné par la formule de Schwert (1989) :

$$p_{\max} = \left[12 \times \left(\frac{T}{100} \right)^{0.25} \right] = 10$$

Ensuite on réalise le MAIC(p) dont la sortie R est ci-contre. On remarque que le max lag est égale à 0. On emploie alors le BIC.

Nombre de retard P, Dickey-Fuller Augmenté

On regarde ici la valeur absolue de la statistique t égale à 2.53 étant supérieur à 1.6, le coefficient gamma 1 est significatif. **On garde donc ce test de Dickey-Fuller Augmenté.**

Consiste à estimer par les MCO :

$$\Delta ddPetrole_t = (\rho - 1)ddPetrole_{t-1} + \varepsilon_t + \sum_{p=1}^P \gamma_p \Delta ddPetrole_{t-p}$$

Puis à tester :

$$H_0 : \rho - 1 = 0 \quad \text{versus} \quad H_a : |\rho| < 1$$

Pour savoir si on accepte H_0 ou non, la règle de décision est si la valeur de la statistique t associée à $(\rho - 1)$ > à la valeur critique donnée à l'intersection de la ligne tau1 et de la colonne 5%, on accepte H_0 .

Ici la valeur de la statistique t associée à $(\rho - 1)$ est -7.26 et la valeur critique est de -1.95. On a bien $-7.26 > -1.95$

On rejette alors $H_0 : \rho - 1 = 0$

Le processus est stationnaire d'après Dickey-Fuller Augmenté

```
> summary(ur.df(ddPetrole, type="none", lag=pmax, selectlag="BIC"))
```

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
```

Test regression none

Call:

```
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
```

Residuals:

Min	1Q	Median	3Q	Max
-19388	-3052	1229	4560	12180

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
z.lag.1	-2.0005	0.2753	-7.267	1.47e-08 ***
z.diff.lag	0.4439	0.1753	2.532	0.0158 *

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
	0.1 ‘ ’			1

Residual standard error: 6837 on 36 degrees of freedom

Multiple R-squared: 0.7002, Adjusted R-squared: 0.6835

F-statistic: 42.03 on 2 and 36 DF, p-value: 3.838e-10

Value of test-statistic is: -7.2671

Critical values for test statistics:

1pct	5pct	10pct	
tau1	-2.62	-1.95	-1.61

Choix du modèle

Rappel des différents processus

La fonction d'autocorrélation étendue (EACF)

Choix du meilleur modèle :

- Tests post prévision
- Critère d'Information Bayésien

Rappel des différents processus

Pour estimer un modèle nous pouvons procéder de deux manières :

- **L'estimateur du Maximum de vraisemblance** pour estimer les modèles *MA* et *ARMA* :
Il cherche à maximiser la vraisemblance des observations données les paramètres du modèle.
- **Les MCO (conditionnel)** pour les modèles *AR* :
Les MCO conditionnels consistent à minimiser la somme des carrés des résidus entre les valeurs observées et les valeurs prédites par le modèle.

Revoyons plus en détail les caractéristiques distinctes des processus AR (Auto-Régressifs), MA (Moyenne Mobile) et ARMA (Auto-Régressifs Moyenne Mobile).

Processus auto-régressif (AR)

Définition on appelle processus autorégressif d'ordre p usuellement noté $AR(p)$, un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ vérifiant une relation du type :

$$X_t + \sum_{i=1}^p \phi_i X_{\{t-i\}} = \varepsilon_t \quad \forall t \in \mathbb{Z}$$

Avec $\phi_i \in \mathbb{R}$ et ε_t un bruit blanc de variance σ^2

Cette relation s'écrit également :

$$\Phi(L)X_t = \varepsilon_t$$

$$\text{où } \Phi(L) = 1 + \phi_1 L + \cdots + \phi_p L^p$$

Processus moyennes mobiles (MA)

définition on appelle processus moyenne mobile d'ordre q , usuellement noté MA(q) (moving average), un processus $(X_t)_{t \in \mathbf{Z}}$ défini par:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \forall t \in \mathbf{Z}$$

où les θ_i sont des réels et ε_t est un bruit blanc de variance σ^2 .

De même que pour les processus autorégressifs cette relation s'écrit:

$$X_t = (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t = \Theta(L) \varepsilon_t$$

Contrairement à l'AR(p) la définition du MA(q) est explicite et le processus X_t est stationnaire.

Processus autorégressifs moyennes mobiles (ARMA)

Les processus ARMA(p,q) généralise les modèles autorégressifs et moyennes mobiles. Ces modèles sont très utiles en pratique pour modéliser des séries réelles en nécessitant moins de paramètres que les modèles AR ou MA simples.

définition un processus stationnaire X_t admet une représentation ARMA(p,q) minimale s'il satisfait

$$X_t + \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \forall t \in \mathbf{Z}$$

soit, avec les notations précédentes

$$\Phi(L)X_t = \Theta(L)\varepsilon_t, \quad \forall t \in \mathbf{Z}$$

avec les conditions suivantes:

1. $\phi_p \neq 0$ et $\theta_q \neq 0$
2. Φ et Θ n'ont pas de racines communes et leurs racines sont de modules > 1
3. ε_t est un BB de variance σ^2

Processus AutoRegressive Integrated Moving Average (ARIMA)

On dit que y_t suit un processus ARIMA(p,d,q) si $\Delta^d y_t$ suit un processus ARMA(p,q) c'est-à-dire si on a besoin de différencier d fois y_t pour le rendre stationnaire.

- Un ARIMA(p,d,q) s'écrira

$$\Phi_p(L)(1 - L)^d y_t = c + \Theta_q(L)\epsilon_t$$

Fonction d'autocorrélation étendue (EACF)

L'**eacf** "Extended AutoCorrelation Function" est un tableau à double entrée avec les valeurs de p en ligne et les valeurs de q en colonne. Les valeurs de p et de q à choisir sont celles qui correspondent au coin supérieur gauche d'un triangle qui ne contient que des 'o' qui est rectangle en haut à droite et ouvert à droite et qui est le plus en haut et à gauche possible (pour minimiser la somme de p + q)

```
> eacf(ddPetrole) # p = 0 et q = 1  
AR/MA  
 0 1 2 3 4 5 6 7 8 9 10 11 12 13  
0 x o o o o o o o o o o o o o  
1 x x o o o o o o o o o o o o  
2 x o o o o o o o o o o o o o  
3 o x x o o o o o o o o o o o  
4 x o o o o o o o o o o o o o  
5 x o o o o o o o o o o o o o  
6 x o o o o o o o o o o o o o  
7 x o o o o o o o o o o o o o
```

Ici nous obtenons **p = 0** et **q = 1**

Pour modéliser la série différenciée à l'ordre 1 on peut utiliser un MA(1) donc on utilisera un ARIMA(0,2,1) pour la série originale.

Afin d'être sûr d'obtenir le meilleur modèle nous allons comparer plusieurs modèles entre eux. Tel que ARIMA(0,2,1), ARIMA(1, 2, 3), ARIMA(2,2,3) et bien plus encore (voir code R) pour la série originale.

Test sur les coefficients

`coeftest()` est utilisé pour tester des hypothèses sur les coefficients estimés dans les modèles de régression, ce qui peut être utile pour évaluer la signification statistique de ces coefficients. Cette étape s'effectue avant de faire la série de tests post prévision.

```
> reg1 = Arima(Petrole, order = c(0, 2, 1), include.constant = T)
> coeftest(reg1) # Le coefficient ma1 est significatif.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ma1	-0.893912	0.068065	-13.133	< 2.2e-16 ***

Signif. codes:	0	***	0.001	**
			0.01	*
			0.05	.
			0.1	'
			1	'

ARIMA(0,2,1)

La p-value associé au coefficient ma1 est inférieur à 0.05. Le coefficient ma1 est significatif.

Test sur les coefficients

ARIMA(1,2,1)

```
> reg2 = Arima(Petrole, order = c(1, 2, 1), include.constant = T)
> coeftest(reg2)
```

73

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)		
ar1	0.141597	0.159412	0.8882	0.3744		
ma1	-0.915566	0.065223	-14.0374	<2e-16 ***		

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’	0.1 ‘ ’	1

Le coefficient ar1 n'est pas significatif. Si on estime pas ce coefficient, cela revient à estimer le modèle ARIMA(0,2,1) déjà abordé.

Test sur les coefficients

```
> reg3 = Arima(Petrole, order = c(1, 2, 2), include.constant = T)
> coeftest(reg3) # Le coefficient ma1 n'est pas significatif.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.953485	0.056062	-17.0078	<2e-16 ***
ma1	0.103836	0.085099	1.2202	0.2224
ma2	-0.896163	0.081870	-10.9462	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> reg3 = Arima(Petrole, order = c(1, 2, 2), include.constant = T,
+                 fixed = c(NA, 0, NA))
> coeftest(reg3) # Tous les coefficients sont significatifs
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.870921	0.096161	-9.0569	< 2.2e-16 ***
ma2	-0.806279	0.124050	-6.4996	8.053e-11 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ARIMA(1,2,2)

Le coefficient ma1 n'est pas significatif, donc on réestime le modèle sans.

Sans le coefficient ma1, tous les coefficients sont significatifs.

Test sur les coefficients

ARIMA(2,2,3)

```
> reg5 = Arima(Petrole, order = c(2, 2, 3), include.constant = T)
> coeftest(reg5) # Tous les coefficients sont significatifs.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.63031	0.21603	2.9177	0.003526 **
ar2	-0.88483	0.20318	-4.3550	1.331e-05 ***
ma1	-1.42708	0.26574	-5.3702	7.867e-08 ***
ma2	1.18162	0.37554	3.1465	0.001653 **
ma3	-0.59927	0.30400	-1.9713	0.048693 *

Signif. codes:	0	***	0.001	**
			0.01	*
			0.05	.
			0.1	'
			1	'

Tous les coefficients sont significatifs.

Test sur les coefficients

ARIMA(1,2,0)

```
> reg8 = Arima(Petrole, order = c(1, 2, 0), include.constant = T)
> coefest(reg8) # Le coefficient ar1 est significatif.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.40582	0.14285	-2.8409	0.004498 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

La p-value associé au coefficient ar1 est inférieur à 0.05. Le coefficient est significatif.

Tests post



Lorsque nous estimons des modèles statistiques pour prédire des valeurs futures, que le modèle soit du type AR, MA ou ARMA. Nous posons comme hypothèse que les aléas (= erreurs) de notre modèle sont des bruits blancs. Comme on ne connaîtra jamais les aléas, nous les approcherons par les résidus de notre régression. L'hypothèse de BB implique que :

- L'espérance des aléas est nulle ;
- Les aléas ne sont pas autocorrélés ;
- Les aléas ne sont pas conditionnellement hétéroscléastiques

De plus, il est essentiel de vérifier que les erreurs suivent une distribution gaussienne pour garantir la validité de nos inférences, c'est-à-dire les conclusions que nous tirons sur la significativité des coefficients. Afin de confirmer ces hypothèses, nous allons entreprendre différents tests pour évaluer la conformité de nos données aux critères susmentionnés.

Test d'espérance nulle pour les aléas

Ce test à comme hypothèses :

$$H_0 : E(\varepsilon_t) = 0 \quad \text{et} \quad H_a : E(\varepsilon_t) \neq 0$$

On va utiliser la moyenne des résidus, notée $m(e)$ pour approcher l'espérance des aléas et on va calculer un t de Student. Si la statistique t, représenté par la formule ci-dessous :

$$\frac{|m(e)|}{\sigma_e} \times \sqrt{T}$$

est inférieur au quantile de la loi normale centrée réduite soit 1.96 quand on prend le seuil de risque de 5% alors nous acceptons l'hypothèse nulle. Sous H_0 cette statistique est distribuée comme une normale centrée réduite.

Test d'autocorrélation de Ljung-Box

- La statistique de Ljung-Box (fonction **Box.test**) s'écrit :

$$Q_K = T(T + 2) \sum_{k=1}^K \frac{\hat{\rho}(k)^2}{T - k}$$

- Elle permet de tester H_0 : indépendance contre H_a : présence d'autocorrélation
 $H_0 : \rho(1) = \dots = \rho(K) = 0$ versus H_a : un des K coefficients ρ est NON nul
- Sous H_0 , $Q_K \xrightarrow[T \rightarrow \infty]{L} \chi^2(K)$
- Si la valeur calculée de $Q_K >$ à la valeur tabulée de $\chi^2(K) \rightarrow$ rejet de H_0

Test d'hétéroscédasticité conditionnelle d'Engle

L'idée de l'hétéroscédasticité conditionnelle (Auto-Regressive Conditionnal Heteroscedasticity, ARCH) est que la variance de l'aléa en t dépend de l'importance des aléas au carré des périodes précédentes. Si vous notez ϵ_t le t ème aléa de votre modèle de régression, le modèle ARCH(p) s'écrit

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_p \epsilon_{t-p}^2 \quad (3)$$

- Le test ARCH(p) d'Engle que vous appliquez aux résidus (e_t) de votre modèle de régression (AR, MA ou ARMA) consiste à estimer par MCO l'équation (4) suivante

$$e_t^2 = a + c_1 e_{t-1}^2 + c_2 e_{t-2}^2 + \dots + c_p e_{t-p}^2 + \text{erreur}_t \quad (4)$$

puis à tester la nullité de tous les coefficients c de cette régression. Remarquez que si tous les c sont nuls, alors la variance est égale à a c'est-à-dire une constante (i.e homoscédasticité conditionnelle).

- Le test d'Engle permet de tester H_0 : homoscédasticité conditionnelle versus H_a : hétéroscédasticité conditionnelle.
- La statistique de test est $LM = T \times R^2$ avec le coefficient de détermination associé à l'estimation de (4).
- Sous H_0 , $LM \sim \chi^2(p)$. Donc si $LM > \chi^2(p)$ ou si la p-value < 0.05, on rejette H_0 et on considère que le processus est justifiable d'un ARCH(p).

Test de normalité de Jarque-Bera

- il permet de tester H_0 : normalité versus H_a : la non normalité
- la statistique de test s'écrit

$$JB = \frac{T - k}{6} \left(S^2 + \frac{(K - 3)}{4} \right) \quad (5)$$

avec

- T , le nb d'observations
- k , le nb de variables explicatives si les données proviennent des résidus d'une régression linéaire. Sinon, $k = 0$.
- S , le coefficient d'asymétrie i.e. le moment d'ordre 3 d'une variable centrée-réduite
- K , le Kurtosis i.e. le moment d'ordre 4 d'une variable centrée-réduite
- Sous H_0 , la statistique JB suit asymptotiquement une χ^2 à 2 ddl.
- Si $JB > \chi^2(2)$, vous rejetez H_0 ; de même si la p-value < 0.05.

Tableau comparatif des tests post

	ARIMA(0, 2, 1)	ARIMA(1,2,2)	ARIMA(2,2,3)	ARIMA(1,2,0)
t.test	0.7686	0.7454	0.6282	0.9093
Box.test	0.3332 0.6066 0.7693 0.8358 0.9160 0.9485 0.9559 0.9786 0.9860 0.9935 0.9942 0.9969	0.4380 0.7383 0.8392 0.8551 0.9309 0.9503 0.9504 0.9745 0.9817 0.9913 0.9931 0.9964	0.8099 0.8527 0.9387 0.9655 0.9343 0.9709 0.9672 0.9657 0.9805 0.9898 0.9919 0.9958	0.4127 0.0517 0.0861 0.1471 0.2275 0.3048 0.3483 0.4506 0.5264 0.6212 0.6396 0.7058
ArchTest	0.2442 0.5036 0.6205 0.3943 0.5890 0.1475 0.2175 0.2213 0.2970 0.8478 0.8852 0.2890	0.2780 0.5374 0.6363 0.4202 0.5967 0.1409 0.2048 0.2293 0.3007 0.8833 0.9010 0.3029	0.2438 0.4819 0.6927 0.4430 0.5015 0.0961 0.1611 0.2378 0.2554 0.7631 0.5347 0.3301	0.0040 0.0168 0.0451 0.0259 0.0230 0.0235 0.0219 0.0425 0.0636 0.0527 0.0096 0.0190
jarque.bera.test	0.0594	0.0722	0.2964	0.1235

Chaque p-value est supérieur à 0.05 on accepte donc chaque hypothèse nulle de chaque modèle ce qui indique les aléas sont des bruits blancs et ils suivent une distribution gaussienne. Sauf pour le modèle ARIMA(1,2,0), le ArchTest indique que le modèle possède des effets ARCH. On ne prendra donc pas en compte ce modèle

BIC

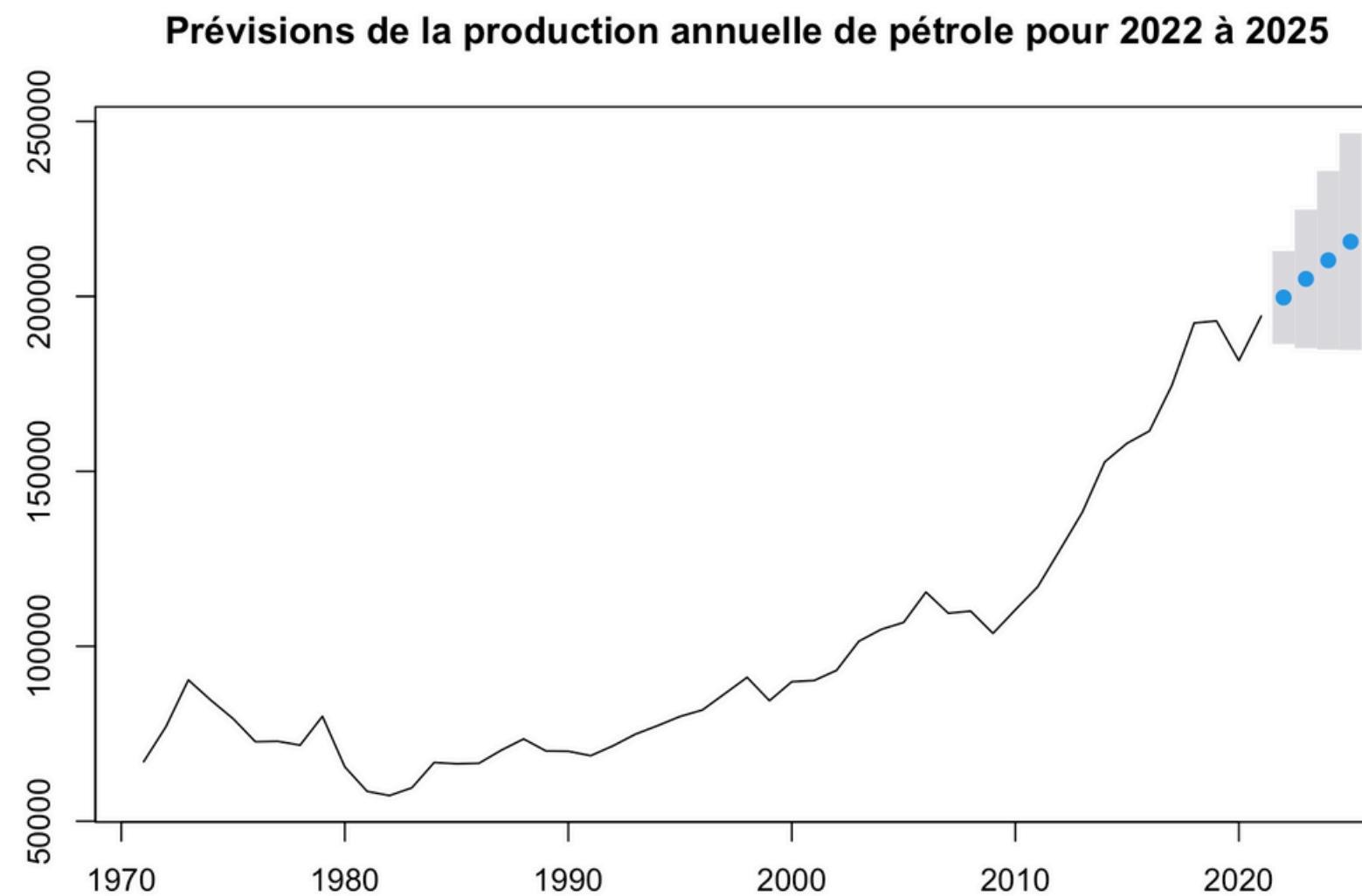
Le Critère d'Information Bayésien (BIC) est un outil statistique utilisé pour comparer différents modèles en prenant en compte à la fois leur ajustement aux données et leur complexité.

Dans cette situation, le modèle **ARIMA(0,2,1)** se distingue comme étant le meilleur choix, étant donné qu'il affiche le BIC le plus faible.

BIC
ARIMA(0,2,1)
ARIMA(1,2,2)
ARIMA(2,2,3)

Prévision de la production annuelle de pétrole non raffiné pour 2022 à 2025 au Canada

Résultat des prévisions avec le modèle ARIMA(0,2,1)



	Prévision	Lo 95	Hi 95
2022	199679.1	186400.7	212957.6
2023	205000.2	185200.5	224799.9
2024	210321.3	184804.5	235838.1
2025	215642.4	184698.6	246586.2

Différences entre les valeurs réelles et les prévisions

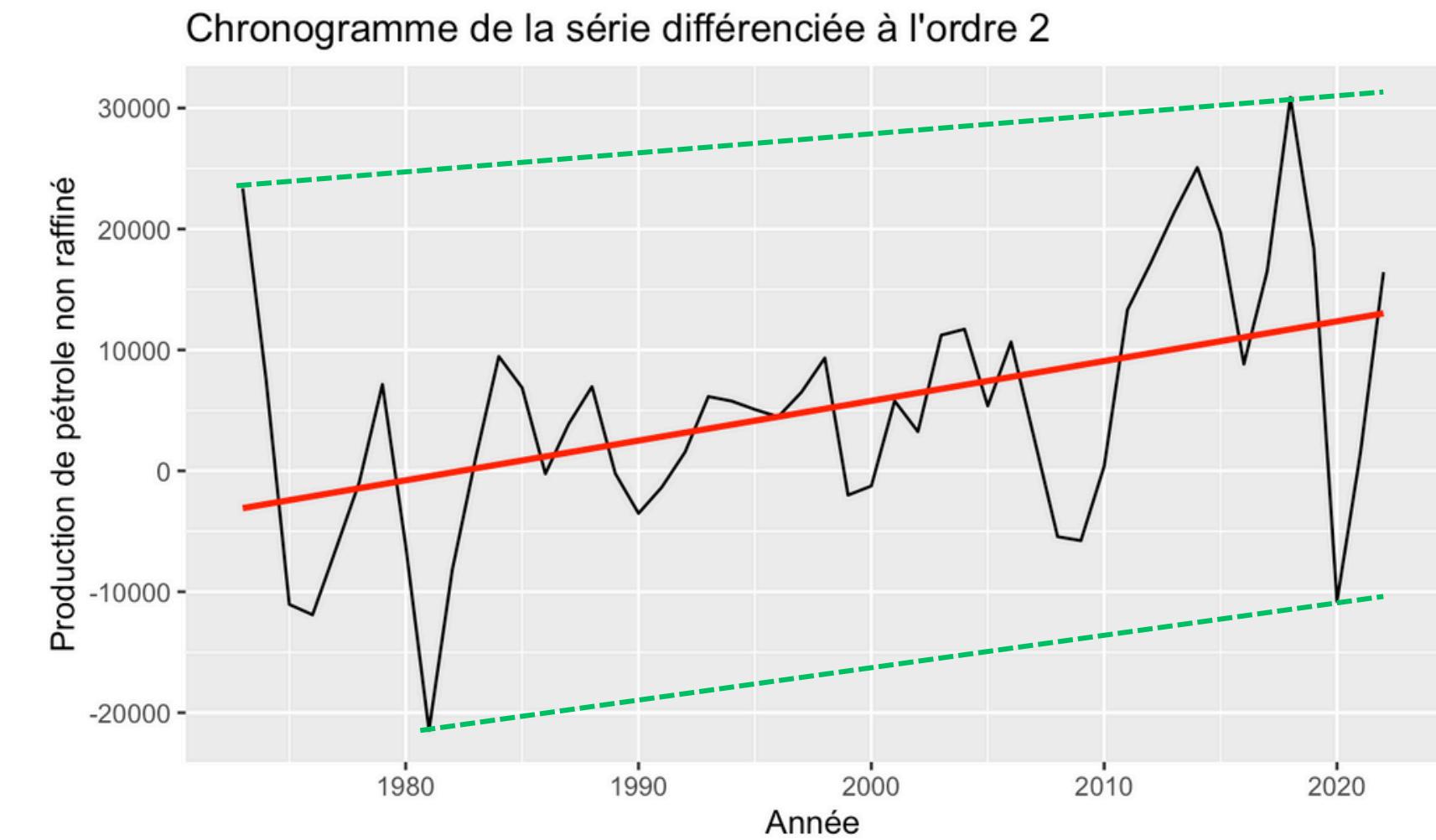
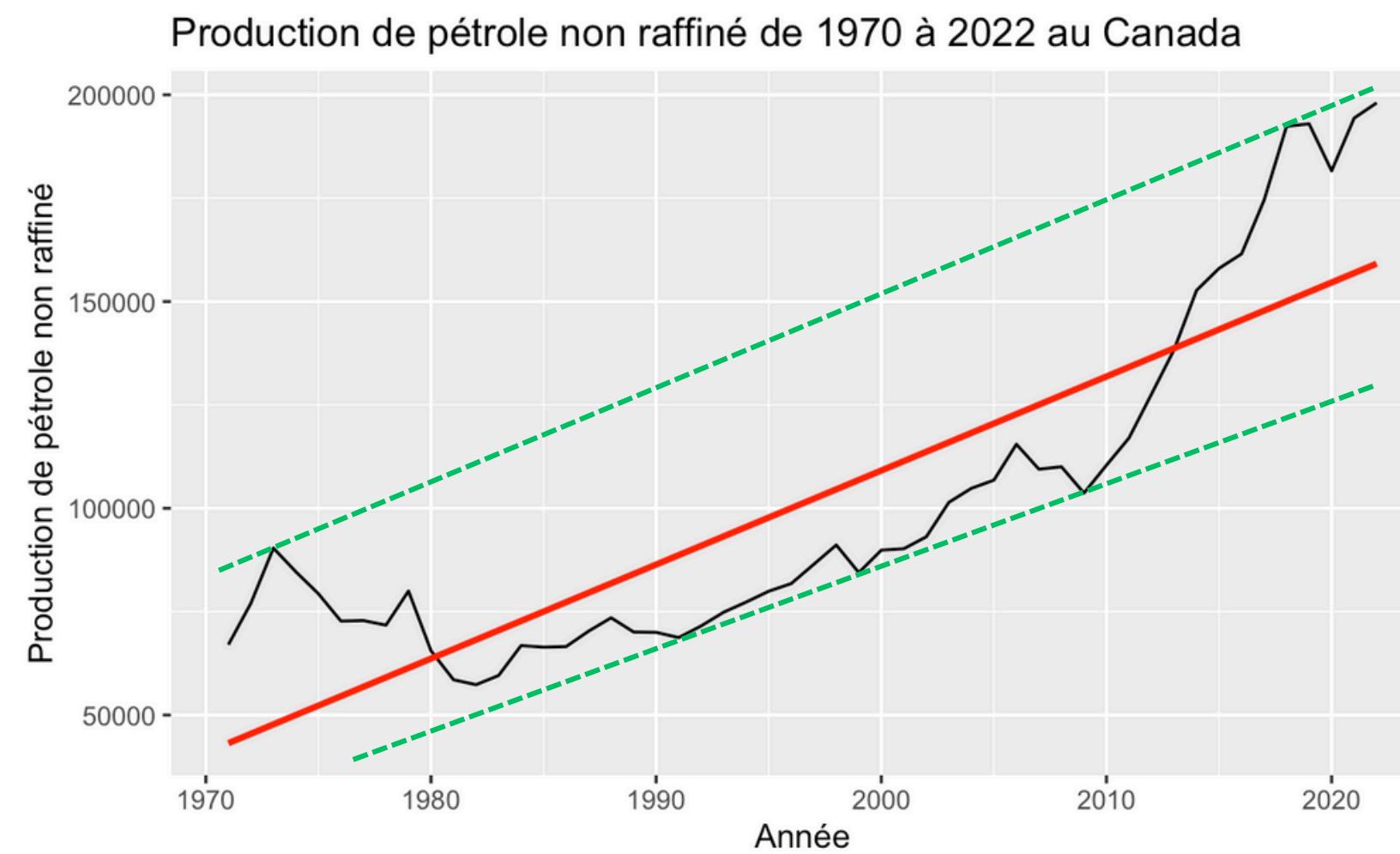
En l'absence des données réelles exactes pour les années 2022 et 2023, nous avons obtenu le pourcentage d'augmentation de la production de pétrole au Canada pour ces années afin de pouvoir comparer les valeurs réelles aux prédictions que nous venons de générer.

	Prévision	Lo 95	Hi 95	%	Valeur réelle	Erreur	Erreur en %
2022	199679.1	186400.7	212957.6	+ 1,9 %	198050,802	-1628.298	0.82
2023	205000.2	185200.5	224799.9	+ 2,1 %	202209,8688	-2790.331	1.38

Nous remarquons que les valeurs réelles se situent à l'intérieur des intervalles de confiance et que les taux d'erreur sont inférieurs à 2%.

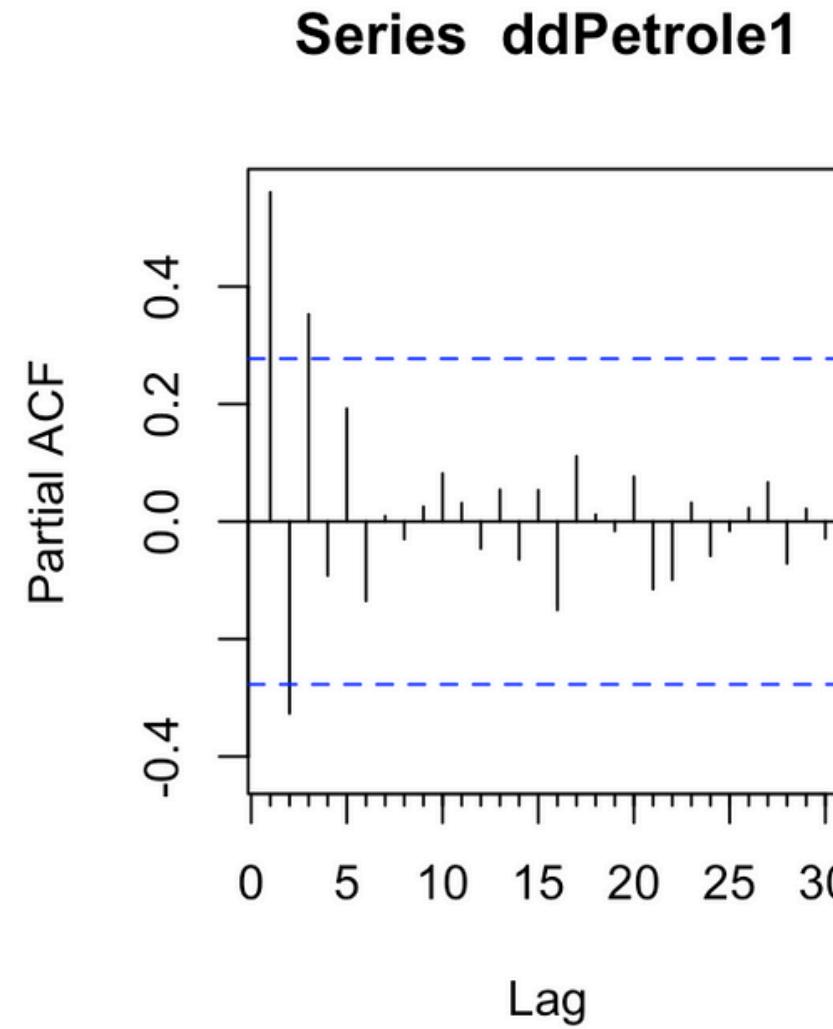
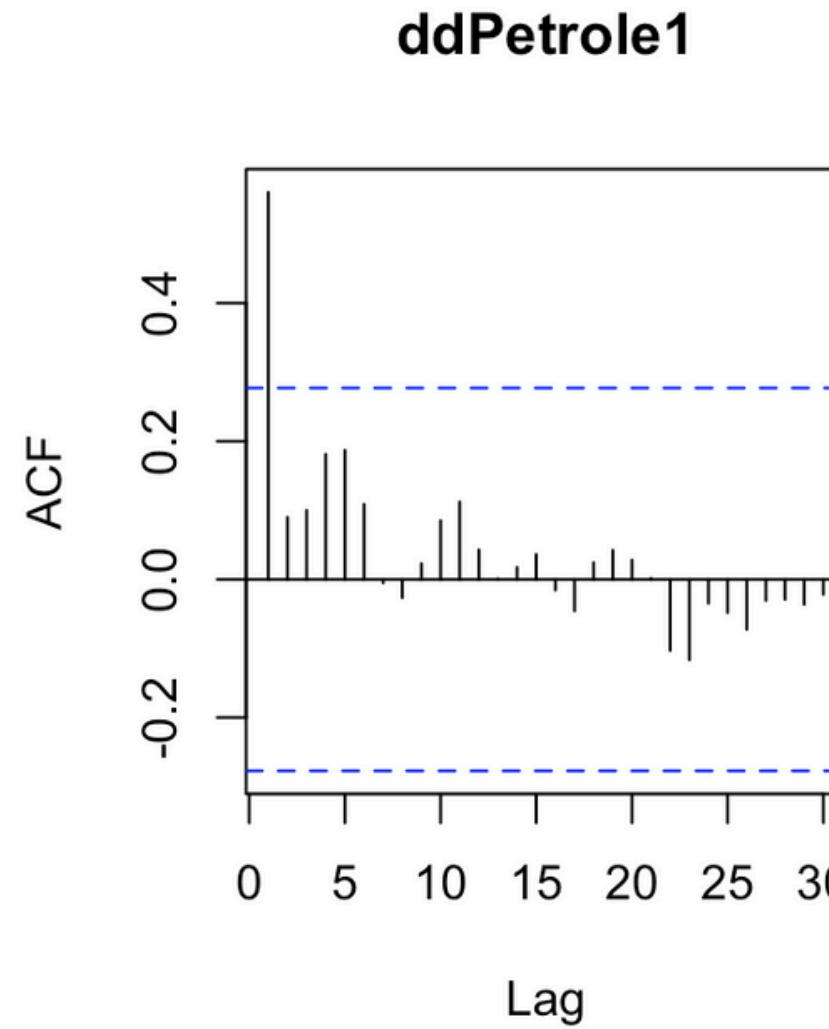
L'estimation et la
prévision avec
l'année 2022 en plus

Chronogramme



Corrélogramme de la série différenciée

Détermination de présence de corrélation dans la série



ACF

Pas d'autocorrélation détecter dans l'ACF de la série différenciée à l'ordre 2

PACF

Pour poursuivre l'analyse, les trois premières valeurs du PACF sont hors de l'intervalle de confiance avec une forte corrélation de 0.5, -0.3 et 0.35, suivie de l'absence d'autocorrélation ultérieure

Fonction d'autocorrélation étendue (EACF)

```
> eacf(ddPetrole1) # p = 0 et q = 1  
AR/MA  
 0 1 2 3 4 5 6 7 8 9 10 11 12 13  
0 x o o o o o o o o o o o o o  
1 x x o o o o o o o o o o o o  
2 x o o o o o o o o o o o o o  
3 x x x o o o o o o o o o o o  
4 x o x o o o o o o o o o o o  
5 x o o o o o o o o o o o o o  
6 x x o o o o o o o o o o o o  
7 x o o o o o o o o o o o o o
```

Ici nous obtenons **p = 0** et **q = 1**

Pour modéliser la série différenciée à l'ordre 1 on peut utiliser un MA(1) donc on utilisera un ARIMA(0,2,1) pour la série originale.

Afin d'être sûr d'obtenir le meilleur modèle nous allons comparer plusieurs modèles entre eux. Tel que ARIMA(1,2,1), ARIMA(1, 2, 2), ARIMA(2,2,3) et bien plus encore (voir code R) pour la série originale.

Test sur les coefficients

ARIMA(0,2,1)

```
> reg1 = Arima(Petrole1, order = c(0, 2, 1), include.constant = T)
> coeftest(reg1) # Le coefficient ma1 est significatif.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ma1	-0.896570	0.066527	-13.477	< 2.2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

La p-value associé au coefficient ma1 est inférieur à 0.05. Le coefficient ma1 est significatif.

Test sur les coefficients

```
> reg3 = Arima(Petrole1, order = c(1, 2, 2), include.constant = T)
> coeftest(reg3) # Le coefficient ma1 n'est pas significatif.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.952246	0.055223	-17.244	<2e-16 ***
ma1	0.101736	0.082916	1.227	0.2198
ma2	-0.898263	0.079848	-11.250	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> reg3 = Arima(Petrole, order = c(1, 2, 2), include.constant = T,
+               fixed = c(NA, 0, NA))
> coeftest(reg3) # Tous les coefficients sont significatifs
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.870921	0.096161	-9.0569	< 2.2e-16 ***
ma2	-0.806279	0.124050	-6.4996	8.053e-11 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ARIMA(1,2,2)

Le coefficient ma1 n'est pas significatif, donc on réestime le modèle sans.

Sans le coefficient ma1, tous les coefficients sont significatifs.

Test sur les coefficients

ARIMA(2,2,3)

```
> # modèle de la série Petrole  
> reg5 = Arima(Petrole1, order = c(2, 2, 3), include.constant = T)  
> coefest(reg5) # Tous les coefficients sont significatifs.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.599714	0.047107	12.7310	< 2.2e-16 ***
ar2	-0.992506	0.023516	-42.2063	< 2.2e-16 ***
ma1	-1.448597	0.091475	-15.8359	< 2.2e-16 ***
ma2	1.496348	0.136830	10.9358	< 2.2e-16 ***
ma3	-0.892257	0.107847	-8.2734	< 2.2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Tous les coefficients sont significatifs.

Test sur les coefficients

ARIMA(1,2,0)

```
> reg8 = Arima(Petrole1, order = c(1, 2, 0), include.constant = T)
> coefest(reg8) # Le coefficient ar1 est significatif.
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.40100	0.12936	-3.0999	0.001936 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

La p-value associé au coefficient ar1 est inférieur à 0.05. Le coefficient est significatif.

Tableau comparatif des tests post

	ARIMA(0, 2, 1)	ARIMA(1,2,2)	ARIMA(2,2,3)	ARIMA(1,2,0)
t.test	0.7814	0.7539	0.7149	0.9197
Box.test	0.3490 0.6399 0.8042 0.8717 0.9396 0.9623 0.9672 0.9850 0.9894 0.9952 0.9958 0.9977	0.4525 0.7529 0.8632 0.8866 0.9498 0.9617 0.9613 0.9815 0.9856 0.9934 0.9949 0.9973	0.8072 0.9144 0.9588 0.9707 0.9712 0.9891 0.9925 0.9941 0.9954 0.9959 0.9976 0.9989	0.4169 0.0476 0.0780 0.1336 0.2085 0.2830 0.3220 0.4215 0.4968 0.5921 0.6093 0.6786
ArchTest	0.2543 0.5578 0.6878 0.5880 0.8213 0.1319 0.2021 0.2421 0.3155 0.9391 0.9573 0.5891	0.2898 0.6082 0.7068 0.6105 0.8123 0.1271 0.1939 0.2482 0.3199 0.9584 0.9632 0.6265	0.3500 0.6844 0.7900 0.6134 0.7762 0.1353 0.2235 0.2531 0.2678 0.7737 0.6487 0.5231	0.0252 0.0903 0.2032 0.1805 0.1874 0.1967 0.1657 0.2568 0.3333 0.3978 0.2517 0.2297
jarque.bera.test	0.0605	0.0757	0.2473	0.09825

Chaque p-value est supérieur à 0.05 on accepte donc chaque hypothèse nulle de chaque modèle ce qui indique les aléas sont des bruits blancs et ils suivent une distribution gaussienne. Sauf pour le modèle ARIMA(1,2,0), le ArchTest indique que le modèle possède des effets ARCH. On ne prendra donc pas en compte ce modèle

BIC

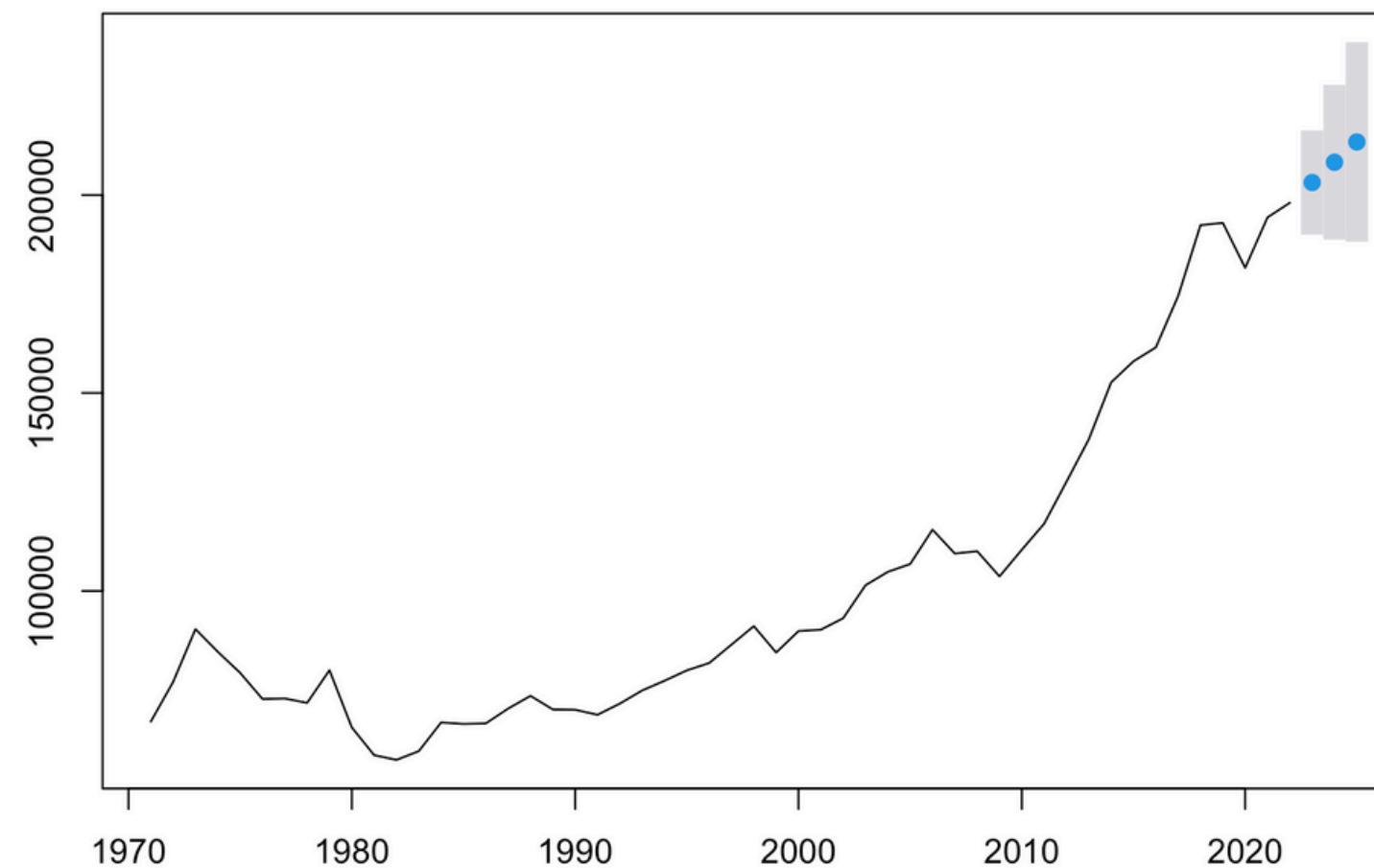
Le Critère d'Information Bayésien (BIC) est un outil statistique utilisé pour comparer différents modèles en prenant en compte à la fois leur ajustement aux données et leur complexité.

Dans cette situation, le modèle **ARIMA(0,2,1)** se distingue comme étant le meilleur choix, étant donné qu'il affiche le BIC le plus faible.

BIC	
ARIMA(0,2,1)	1031.437
ARIMA(1,2,2)	1035.21
ARIMA(2,2,3)	1043.14

Résultat des prévisions avec le modèle ARIMA(0,2,1)

Prévisions de la production annuelle de pétrole pour 2023 à 2025



	Prévision	Lo 95	Hi 95
2023	203170.5	190023.7	216317.3
2024	208290.2	188712.6	227867.7
2025	213409.9	188210.2	238609.5