



Faculté des Sciences Exacte et Informatique
Département de Mathématiques

Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

Master

Spécialité : Mathématiques.

Option : Probabilités et Statistique.

Thème

**Détection de rupture dans
les séries temporelles**

Présenté par :
Ghettout Hadjer

Devant le jury composé de :

Sellami Nawel	M.A.A Université Mohammed Seddik Ben Yahia - Jijel	Encadreur
Boudjerda Khawla	M.C.A Université Mohammed Seddik Ben Yahia - Jijel	Président
Madi Meriem	M.A.A Université Mohammed Seddik Ben Yahia - Jijel	Examineur

Promotion **2021/2022**

♡ Remerciements ♡

Avant de présenter ce travail mes remerciements vont tout d'abord à :

*ALLAH tout puissant pour tout nous avoir donné la volonté, la santé, et le courage
d'accomplir ce travail.*

*En second lieu, je remercie ma superviseuse **Mme.Sellami Nawel** pour sa patience, sa
disponibilité et surtout ses judicieux conseils qui ont contribué à alimenter ma réflexion.*

*Mes sincères remerciements s'adressent également aux membres des jurys pour avoir bien voulu
examiner et juger ce travail .*

*Finalement, je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce
mémoire.*

Dédicace

Je dédie ce travail

À mes chères parents

qui m'ont toujours soutenu

À mon mari

source d'espoir et de motivation

À ma soeur

Marwa

À mes chers frères

Nabil et Razki et Mehdi

*À mes meilleurs amies et particulièrement **Bouchra** et **Salima***

Et à tout mes amies de promotion de 2^{ème} année Master probabilités et statistiques

À tout les membre de ma famille

Et toutes personne qui occupe une place dans mon coeur

Merci♥

Table des matières

Résumé	v
Introduction générale	1
1 Concepts de base des séries temporelles	3
Introduction	3
1.1 Présentation d'une série chronologique	3
1.2 Décomposition d'une série temporelle	6
1.3 Modèles de décomposition déterministe	8
1.3.1 Le modèle additif	8
1.3.2 Le modèle multiplicatif	8
1.3.2.1 1 ^{ère} forme de modèle multiplicatif	8
1.3.2.2 2 ^{ème} forme de modèle multiplicatif	9
1.4 Opérateurs définis sur une série chronologique	9
1.5 Élimination de la partie déterministe ($T(t)$ et $S(t)$) [3] [4]	10
1.6 Processus stochastique	10
1.6.1 Processus stationnaire	11
1.6.2 Théorème de Wold (1954)	11
1.7 Autocorrélation d'une série chronologique [5]	12
1.7.1 La fonction d'autocovariance	12
1.7.2 La fonction d'autocorrélation	13
1.7.3 La fonction d'autocorrélation partielle	13
1.8 Modèles d'une série chronologique	14

1.8.1	Processu aléatoire stationnaire	14
1.8.1.1	Processus autorégressifs $AR(p)$ [7] [10]	14
1.8.1.2	Processus moyenne mobile $MA(q)$ [3]	15
1.8.1.3	Processus mixtes $ARMA(p, q)$ [3] [10]	16
1.8.2	Processus aléatoire non stationnaire	18
1.8.2.1	Les processus ARIMA	18
1.8.2.2	Les processus SARIMA	18
1.8.3	Les processus stochastique non linéaire	19
1.8.3.1	Processus ARCH	19
1.8.3.2	Processus GARCH	19
1.9	Tests sur les résidus	20
1.9.1	Test d'autocorrélation des résidus	20
1.9.2	Test de normalité des résidus	20
1.10	Critères d'information	21
1.10.1	AIC (AKaiKe Information Criterion) [1]	21
1.10.2	BIC (Bayesian Information Criterion)	21
1.11	Rappel sur le test ARCH	22
1.12	Modèle AR avec erreurs Hétéroscédastique	23
2	Les tests de rupture	24
	Introduction	24
2.1	Tests de rupture non paramétrique	24
2.1.1	Test de Corrélacion Sur Le Rang	24
2.1.2	Test de Pettitt	25
2.1.3	Test de Lombard (1987)	26
2.1.4	Test de Mann Kendall(Kendall 1938, reprise par Renard 2006) [11] [12]	27
2.1.5	Test de Mann Kendall Modifié [13]	29
2.2	Tests de rupture paramétrique	29
2.2.1	Le test de Dickey-Fuller (DF)	29
2.2.2	Test de Jarušková (1997)	32
2.2.3	Tests de Reeves et al(2007) [16].	34
2.2.4	Lund et Reeves (2007)LR	34
2.2.5	Test de XLW [16]	35

2.3	Package "changeoint"	38
2.3.1	Test de changement de moyenne	38
2.3.2	Test de changement de variance	39
2.3.3	Test de changement de moyenne et variance	39
2.4	Application	40
2.4.1	Analyse préliminaire de la série P	40
2.4.1.1	Représentation graphique de la série P	40
2.4.1.2	Etude du corrélogramme de la série P	42
2.4.1.3	Traitement de la stationnarité	42
2.4.2	Le teste de Mann-Kendall	43
2.4.3	Le test de Pettitt	43
2.4.4	Différenciation de la série P	43
2.4.4.1	La représentation graphique de la série DP	43
2.4.5	La méthodologie de Box-Jenkins	44
2.4.5.1	Identification des modèles possibles :	44
2.4.5.2	Estimation et validation du modèle	45
2.4.5.3	Validation des résidus	46
2.4.5.4	choix de modèle	47
2.4.5.5	Test de normalité pour les résidus du modèle AR(1)	48
2.4.5.6	Application de la modélisation Autorégressif Conditionnellement Hétéroscédastique :	49
2.4.6	Estimation des modèles GARCH sur les résidus du modèles ARMA	51
2.4.7	Validation du modèle	51
	Conclusion	53
	Bibliographie	54

Table des figures

1.1	Nombre de morts accidentelles aux Etats-Unis de 1973 à 1978	5
1.2	Nombre de passagers (en milliers) dans les transports aériens de 1949 à 1960 . .	5
1.3	Nombre annuel de tâches solaires de 1700 à 1980	5
1.4	Exemple d'une tendance linéaire	6
1.5	Exemple d'une saisonnalité	7
1.6	Exemple d'une fluctuation irrégulière	7
2.1	L'exécution de test de changement de moyenne sur R	38
2.2	L'exécution de test de changement de variance sur R	39
2.3	Le graphe du changement de moyenne et la variance sur R	40
2.4	L'évolution de la série (P) entre (2000-2015)	41
2.5	Corrélogramme et corrélogramme partiel de la série P	42
2.6	La représentation graphique de la série DP	44
2.7	Corrélogramme simple et partiel de la série DP	45
2.8	Le graphe, l'ACF et PACF des residus du modèle AR(1)	47
2.9	Le graphe, l'ACF et PACF des residus du modèle MA(1)	48
2.10	l'histogramme des résidus pour le modèle AR(1)	49
2.11	Le corrélogramme des résidus carrés	50

NOTATIONS

$T(t)$: La tendance où trend.
$S(t)$: La variation saisonnière.
ε_t	: La variation résiduelle où bruit blanc.
\mathbb{E}	: Espérance mathématique.
var	: La Variance.
Cov	: La Covariance.
$\gamma(h)$: La fonction d'autocovariance.
ρ_h	: La fonction d'autocorrélation.
ϕ_{hh}	: La fonction d'autocorrélation partielle.
ACF	: Fonction d'auto-corrélation.
$PACF$: Fonction d'auto-corrélation partielle.
$ARMA$: Auto regressive moving average "Auto-régressif moyenne mobile".
AR	: Auto-régressif "Auto-regressive".
MA	: Moyenne mobile "Moving average".
$ARIMA$: Auto-régressif intégré moyenne mobile.
$SARIMA$: Auto-régressif moyenne mobile intégré saisonnier.
$ARCH$: Auto-régressif conditionnellement Hétéroscédastique.
$GARCH$: Auto-régressif conditionnellement Hétéroscédastique Généralisé.
DF	: Le test de Dickey-Fuller.
LR	: Le test de Lund et Reeves.
XLW	: Le test de Wang.
TPR	: Le test de Reeves et al.

Résumé

L'objectif principal de ce travail porte sur les tests de rupture sur les séries chronologiques et ses applications sur les séries des prix de pétrole, nous présentons les tests de rupture paramétrique et non paramétrique. Nous terminons par une application sur des données réelles tel que nous testons la rupture dans la moyenne et la tendance à l'aide de logiciel R.

Mots clés : Série temporelle, Fonction d'autocorrélation, Fonction d'autocorrélation partielle, Processus ARIMA, Processus GARCH, Test Dikey-Fuller, Test de Pettitt, Test de Mann-Kendall.

Introduction générale

L'analyse des séries temporelles est devenue de plus en plus importante dans plusieurs domaines y compris la médecine, l'industrie, la finance, Les séries temporelles sont des séquences de mesures ordonnées chronologiquement qui décrivent des comportements changeant au cours du temps à cause des événements internes ou externes, ce qui entraîne des irrégularités dans les données, dites ruptures ou break points.

La détection des ruptures consiste à identifier les changements dans une série temporelle. En effet, les ruptures divisent les données en plusieurs segments, chacun représente une période du temps notamment : la tendance, la moyenne, et / ou la variance.

L'objectif de ce travail est d'aborder les tests de rupture de la moyenne et la tendance pour les séries temporelles qui font l'objet de nombreuses études dans la littérature voir par exemple : test de Pettitt [15], Lombard [14], Jarušková [8] et Reeves et al [16] [17]. Ces tests s'appuyant sur l'hypothèse de "non rupture" sous l'hypothèse nulle.

Notre mémoire est subdivisé en deux chapitres structure comme suit :

Le premier chapitre intitulé : "Concepts de base des séries temporelles" nous focalisons notre intérêt sur les méthodes que nous utiliserons dans les chapitres suivants. Ce préliminaire débutera par un bref rappel sur les séries temporelles et les différents types de stationnarité. Nous l'enchainons ensuite par la définition de quelques modèles comme les modèles $AR(p)$, $MA(q)$, les modèles $ARMA(p, q)$, et les modèles $ARCH$ [6] et $GARCH$ [2] qui seront notre point d'intérêt plus tard.

Nous avons consacré le deuxième chapitre intitulé : "Les tests de rupture" pour exposer deux types de tests de rupture paramétriques et non paramétriques, nous présenterons quelques

méthodes de détection de rupture souvent utilisées dans les séries temporelles par la suite on effectué une étude pratique tel que on traite la tendance des prix spot du pétrole brut avec un total de données de 190 observations exprimées en dollars US par baril extrait des bulletins statistiques annuels de l'OPEP couvrant la période : Janvier 2000 a Octobre 2015. Les graphes et les calculs sont effectués par le logiciel R.

Concepts de base des séries temporelles

Introduction

Une série chronologique, ou série temporelle (X_t) est une série d'observations ordonnées au cours du temps. La théorie des séries temporelles est une combinaison de deux concepts, probabiliste et statistique, le probabiliste dont on étudie les caractéristiques des variables aléatoires X_t . Le problème statistique est de donner les caractéristiques des distributions de la série temporelle X_t , pour les observations X_1, X_2, \dots, X_T au temps $t = 1, 2, \dots, T$. Le modèle statistique résultant sert à la compréhension du système stochastique d'une part et la prédiction du future (ie. X_{T+1}, X_{T+2}, \dots) d'autre part.

1.1 Présentation d'une série chronologique

La théorie des séries chronologiques (ou temporelles) est appliquée de nos jours dans des domaines aussi variés que l'économétrie, la médecine ou la démographie,... On s'intéresse à l'évolution au cours du temps d'un phénomène, dans le but de **décrire**, **expliquer** puis **prévoir** ce phénomène dans le futur. On dispose ainsi d'observations à des dates différentes c'est à dire d'une suite de valeurs numériques indicées par le temps.

Exemple 1.1.1. *On peut s'intéresser par exemple à l'évolution du nombre de voyageurs utilisant le train, à l'accroissement relatif mensuel de l'indice des prix ou encore à l'occurrence d'un phénomène naturel (comme le nombre de taches solaires).*

Cette suite d'observations est une famille de variables aléatoires réelles notées $(X_t)_{t \in \Theta}$ est appelée **série chronologique** (ou temporelle). Nous la noterons

$$(X_t)_{t \in \Theta} \quad \text{ou} \quad X_t, t \in \Theta$$

où l'ensemble Θ est appelé **espace des temps** qui peut être ;

- **discret** (nombre de voyageurs SNCF quotidien, température maximale...). Dans ce cas, $\Theta \subset \mathbb{Z}$. Les dates d'observations sont le plus souvent équidistantes : par exemple relevés mensuels, trimestriels... Ces dates équidistantes sont alors indexées par des entiers $t = 1, 2, \dots, T$ et T est le nombre d'observations. On dispose donc des observations des variables X_1, X_2, \dots, X_T issues de la famille $(X_t)_{t \in \Theta}$ où $\Theta \subset \mathbb{Z}$ (le plus souvent $\Theta = \mathbb{Z}$). Ainsi si h est l'intervalle de temps séparant deux observations et t_0 l'instant de la première observation, on a le schéma suivant :

$$\begin{array}{cccc}
 t_0 & t_0 + h & \dots & t_0 + (T - 1)h \\
 \downarrow & \downarrow & \dots & \downarrow \\
 X_{t_0} & X_{t_0+h} & \dots & X_{t_0+(T-1)h} \\
 \downarrow & \downarrow & \dots & \downarrow \\
 X_1 & X_2 & \dots & X_T
 \end{array}$$

- **continu** (signal radio, résultat d'un électrocardiogramme...). L'indice de temps est à valeurs dans un intervalle de \mathbb{R} et on dispose (au moins potentiellement) d'une infinité d'observations issues d'un processus $(X_t)_{t \in \Theta}$ où Θ est un intervalle de \mathbb{R} . Un tel processus est dit à temps continu. Les méthodes présentées dans ce cadre sont différentes de celles pour les séries chronologiques à temps discret et présentées dans la suite.

Nous considérons uniquement des **séries chronologiques** $(X_t)_{t \in \Theta}$ à temps discret et **unidimensionnels**.

Voici quelques exemples de séries temporelles :

Exemple 1.1.2. *Nombre de morts accidentelles aux Etats-Unis de 1973 à 1978 Figure (1.1).*

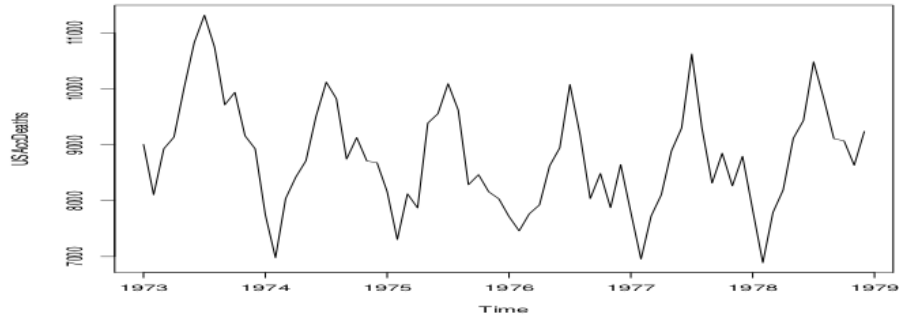


FIGURE 1.1 – Nombre de morts accidentelles aux Etats-Unis de 1973 à 1978

Exemple 1.1.3. *Nombre de passagers par mois (en milliers) dans les transports aériens, de 1949 à 1960 Figure (1.2).*

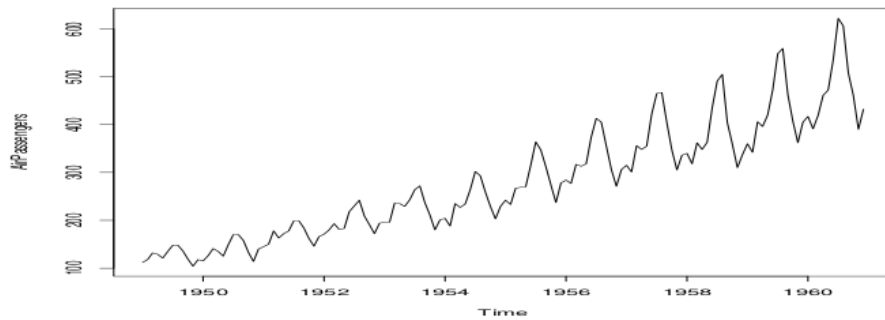


FIGURE 1.2 – Nombre de passagers (en milliers) dans les transports aériens de 1949 à 1960

Exemple 1.1.4. *Nombre annuel de tâches solaires observées à la surface du soleil de 1700 à 1980 Figure (1.3).*

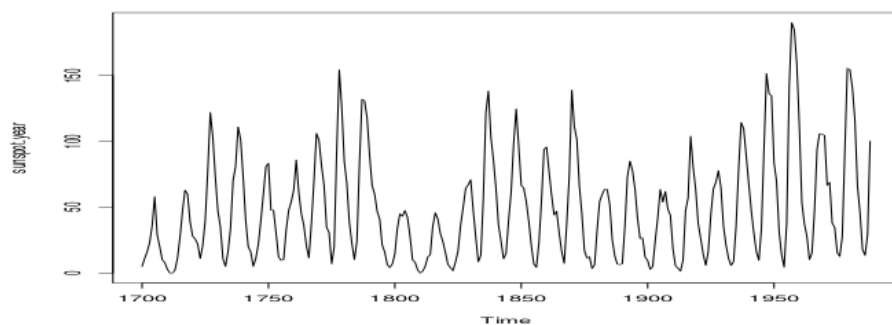


FIGURE 1.3 – Nombre annuel de tâches solaires de 1700 à 1980

Représentation graphique

On représente graphiquement la série chronologique $(X_t)_{t \in T}$ on deux étape :

- 1- En dessinant le nuage formée par les points $(X_t)_{1 \leq t \leq T}$.
- 2- En reliant les points entre eux par des segments de droite, pour indiquer la chronologique.

1.2 Décomposition d'une série temporelle

La première étape dans l'étude d'une série chronologique consiste à représenter graphiquement cette série. On distingue trois composantes principales d'une série chronologique :

- La tendance $T(t)$.
- La variation saisonnière $S(t)$.
- La variation résiduelle où bruit blanc ε_t .

La tendance $T(t)$ où trend

Définition 1.2.1. *La tendance représente l'évolution à long terme de la série étudiée, elle traduit le comportement moyen de la série, cette tendance peut être linéaires, polynomiale, exponentielle, ...*

La série admet une tendance si on peut écrire :

$$X_t = T(t) + \varepsilon_t$$

telle que ε_t est le bruit blanc.

La figure(1.4) suivante présente la tendance qui augmente d'une façon linéaire.

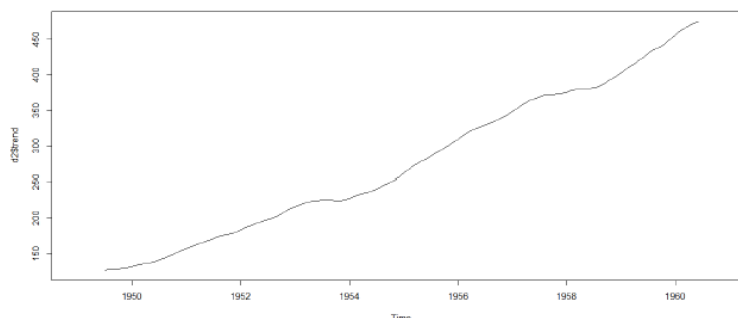


FIGURE 1.4 – Exemple d'une tendance linéaire

La variation saisonnière $S(t)$

Définition 1.2.2. *La variation saisonnière correspond à un phénomène qui se répète à des intervalles de temps réguliers (périodiques). Elle est notée par $S(t), t = 1, \dots, T$ telle que :*

$$S(t + ks) = S(t), \forall k \in \mathbb{Z}$$

s : la période de $S(t)$.

La figure (1.5) représente des cycles réguliers au cours du temps et de même amplitude.

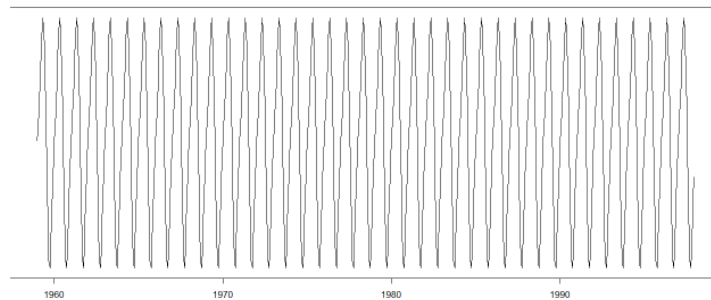


FIGURE 1.5 – Exemple d'une saisonnalité

La variation résiduelle où bruit blanc ε_t

Définition 1.2.3. *Les variations résiduelles sont des fluctuations irrégulières. Elle sont supposées en général de faibles amplitude.*

Par exemple la figure (1.6) représente des fluctuations irrégulières :

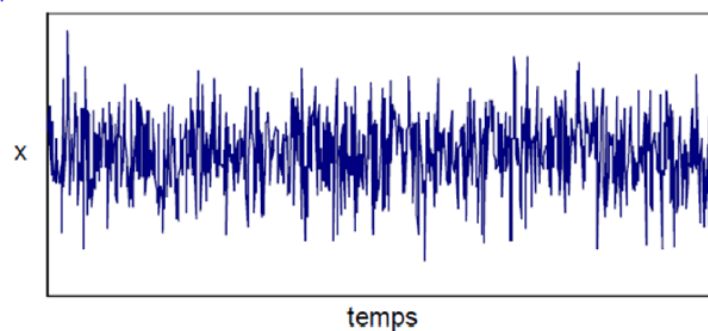


FIGURE 1.6 – Exemple d'une fluctuation irrégulière

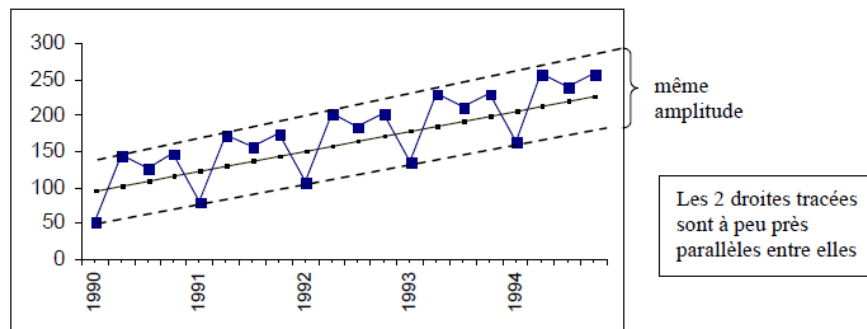
1.3 Modèles de décomposition déterministe

1.3.1 Le modèle additif

Dans un modèle additif, on suppose que les trois composantes : tendance $T(t)$, variations saisonnières $S(t)$ et variations accidentelles ε_t sont indépendantes les unes des autres, tel que :

$$X_t = T(t) + S(t) + \varepsilon_t, \forall t \in T$$

Graphiquement, l'amplitude des variations est constante autour de la tendance.



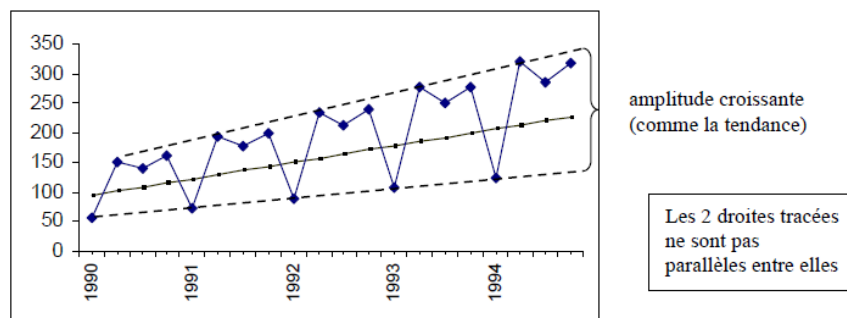
1.3.2 Le modèle multiplicatif

1.3.2.1 1^{ère} forme de modèle multiplicatif

On suppose que les variations saisonnières dépendent de la tendance. Et on considère que X_t s'écrit de la manière suivante :

$$X_t = T(t) \times S(t) + \varepsilon(t), \forall t \in T$$

Graphiquement, l'amplitude des variations saisonnières varie.



1.3.2.2 2^{ème} forme de modèle multiplicatif

On suppose que les variations saisonnières et les variations accidentelles dépendent de la tendance. Et on considère que X_t s'écrit de la manière suivante :

$$X_t = T(t) \times S(t) \times \varepsilon_t, \forall t \in T$$

Remarque 1.3.1. 1-Dans le cas de 2^{ème} forme de modèle multiplicatif à valeurs positives, se ramène à un modèle additif en considérant la série $(\ln(X_t))$:

$$\ln(X_t) = \ln(T(t)) + \ln(S(t)) + \ln(\varepsilon_t), \forall t \in T$$

2-La partie $T(t) + S(t)$ est dite partie déterministe.

1.4 Opérateurs définis sur une série chronologique

Opérateur de retard \mathbf{B}

L'opérateur de retard \mathbf{B} (Backward) se définit de la manière suivante :

$$\mathbf{B} X_t = X_{t-1}$$

Si on applique k fois cet opérateur on décale le processus k unité de temps vers le passé

$$\mathbf{B}^k X_t = X_{t-k}$$

Les propriétés de l'opérateur \mathbf{B}

L'opérateur \mathbf{B} possède les propriétés suivantes :

- 1- $\mathbf{B}^j X_t = X_{t-j} \quad \forall j \in \mathbb{Z}$
- 2- Si $X_t = c \in \mathbb{R}$ pour tout $t \in \mathbb{Z}$; alors :
 - $\mathbf{B}^j X_t = \mathbf{B}^j c = c$ pour tout $j \in \mathbb{Z}$
 - $\mathbf{B}^j(\mathbf{B}^k X_t) = \mathbf{B}^{j+k} X_t = X_{t-j-k}$
 - $\mathbf{B}^{-j} X_t = X_{t+j}$
 - $(\mathbf{B}^j + \mathbf{B}^k) X_t = \mathbf{B}^j X_t + \mathbf{B}^k X_t = X_{t-j} + X_{t-k}$

Opérateur de différence Δ^d

L'opérateur de la d -ième différence noté Δ^d est définie par :

$$\Delta^d X_t = (1 - \mathbf{B})^d X_t, \quad \forall t \in T, \forall d \in \mathbb{N}$$

tel que : $\Delta^d = (1 - \mathbf{B})^d$

Opérateur de différence Δ_s

L'opérateur Δ_s est défini par : $\Delta_s X_t = X_t - X_{t-s}$ tel que : s est la période de la saisonnalité, en d'autres termes : $\Delta_s X_t = (1 - \mathbf{B}^s) X_t$ et $\Delta_s = 1 - \mathbf{B}^s$

Δ sous forme d'un polynôme $\Delta = 1 - \mathbf{B}$

1.5 Elimination de la partie déterministe ($T(t)$ et $S(t)$) [3] [4]

Il y a plusieurs méthodes pour éliminer $T(t)$ et $S(t)$, on présente deux méthodes :

- La méthode de la moyenne mobile pour éliminer la tendance $T(t)$.
- La méthode de différence pour éliminer la tendance $T(t)$ et la saisonnalité $S(t)$. Cette élimination peut s'opérer au moyen de différentiation de différents types :
 - $\Delta^d = (1 - \mathbf{B})^d$ pour éliminer une tendance de degré d .
 - $\Delta_s = (1 - \mathbf{B}^s)$ pour éliminer une saisonnalité de période s .

Dans ce travail, on utilise la méthode de différence pour éliminer la tendance $T(t)$ et $S(t)$.

1.6 Processus stochastique

Définition 1.6.1. *Un processus stochastique est une famille de variables aléatoires $X(t)$, $t \in T$ définie sur un espace de probabilité ou l'ensemble T représente le temps ou bien le domaine d'évolution.*

- $T = \mathbb{N}$: on parlera alors de processus stochastique de temps discret.
- $T = \mathbb{R}$: on dira que $(X(t))_{t \geq 0}$ est un processus stochastique à temps continu.

La stationnarité joue également un rôle important dans la prédiction des séries temporelles, l'intervalle de prédiction étant différent selon que la série est stationnaire ou non. Pour cela on définit les processus stationnaires :

1.6.1 Processus stationnaire

Définition 1.6.2. On dit que le processus $(X_t)_{t \in T}$ est stationnaire au sens strict (ou fortement stationnaire) si la loi de $\{X_{t_1}, \dots, X_{t_n}\}$ est la même que la loi de $\{X_{t_1+h}, \dots, X_{t_n+h}\}$ pour tout (t_1, t_2, \dots, t_n) avec, $t_i \in T$; pour $i = 1, \dots, n$, et pour tout $h \in T$ avec, $t_i+h \in T$. Ainsi, la stationnarité dite forte exprime qu'il y a invariance dans le temps de toutes les caractéristiques du processus.

La stationnarité au sens strict est trop restrictive et on assouplit cette condition en définissant la stationnarité faible ou la stationnarité du second ordre.

Définition 1.6.3. Un processus aléatoire $(X_t)_{t \in T}$ est dit stationnaire du second ordre (ou faiblement stationnaire) s'il remplit les conditions suivantes :

- 1- $\mathbf{E}(X_t) = \mu = Cte; \forall t \in T$
- 2- $\text{var}(X_t) = \sigma^2; \forall t \in T$
- 3- $\text{Cov}(X_t, X_{t-h}) = \gamma(h); \forall t, \forall h \in T$

Les propriétés (1) et (2) signifient que la moyenne et la variance du processus sont indépendants du temps (ie des constantes). La condition(3) traduit le fait que la covariance entre deux période t et $t+h$ est uniquement fonction de la différence des temps h .

1.6.2 Théorème de Wold (1954)

Théorème 1.6.1. Considérons un processus stationnaire X_t . Il est toujours possible de décomposer X_t en une composante déterministe d_t et une composante stochastique u_t tell que :

$$X_t = d_t + u_t,$$

Avec :

$$u_t = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i},$$

où ε_t est un bruit blanc, c'est-à-dire un processus de moyenne nulle et variance constante et non autocorrélé.

1.7 Autocorrélation d'une série chronologique [5]

Moyenne et variance d'une série temporelle

Définition 1.7.1. Soit une série temporelle stationnaire $(X_t, t = 1, \dots, T)$, les expressions de la moyenne et la variance sont :

$$\begin{aligned}\mathbf{E}(X_t) &= \frac{1}{T} \sum_{t=1}^T X_t. \\ \text{var}(X_t) &= \frac{1}{T} \sum_{t=1}^T (X_t - E(X_t))^2.\end{aligned}$$

1.7.1 La fonction d'autocovariance

Soit $(X_t)_{t \in T}$ un processus stochastique, la fonction d'autocovariance est la fonction qui mesure la covariance pour un couple de valeurs associées à des dates différentes, et notée $\gamma(h)$, cette fonction est :

$$\begin{aligned}\gamma(h) &= \text{Cov}(X_t, X_{t+h}) \\ &= \mathbb{E}[(X_t - \mathbb{E}[X_t]) (X_{t+h} - \mathbb{E}[X_{t+h}])]\end{aligned}$$

La fonction d'autocovariance sera noté acvf

Les propriétés de la fonction d'autocovariance

Soit $(X_t)_{t \in T}$ un processus stochastique et $\gamma(h)$ la fonction d'autocovariance.

- 1- pour $h = 0$, $\gamma(0) = \text{Var}(X_t) = \mathbb{E}(X_t)^2 - \mathbb{E}(X_t)^2 = \sigma_X^2 \geq 0, \forall t \in T$
- 2- $\gamma(h) = \gamma(-h)$ (γ est une fonction paire)
- 3- $|\gamma(h)| < \gamma(0)$

1.7.2 La fonction d'autocorrélation

La fonction d'autocorrélation $\rho(h)$ d'un processus $(X_t)_{t \in T}$ est donnée par :

$$\begin{aligned}\rho(h) &= \frac{\text{Cov}(X_t, X_{t+h})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+h})}} \\ &= \frac{\gamma(h)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} \\ &= \frac{\gamma(h)}{\gamma(0)}\end{aligned}$$

Les propriétés de la fonction d'autocorrélation

- 1- $\rho(0) = 1$
- 2- $|\rho(h)| \leq 1$
- 3- $\rho(h) = \rho(-h)$, ρ est une fonction paire.

1.7.3 La fonction d'autocorrélation partielle

La fonction d'autocorrélation partielle notée par ϕ_{hh} est donnée par :

$$\phi_{hh} = \frac{|R^*(h)|}{|R(h)|}$$

Où : $|R^*(h)|$ (resp $|R(h)|$) est le déterminant de la matrice $R^*(h)$ (resp $R(h)$) tels que :

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

$R(h)$ est la matrice symétrique formés des $(h-1)$ premières autocorrélations de X_t .

$$R^*(h) = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ \vdots & & \ddots & \vdots \\ \rho(h-1) & \rho(h-2) & \dots & \rho(h) \end{bmatrix}$$

$R^*(h)$ est obtenue en remplaçant la dernière colonne de $R(h)$ par le vecteur : $(\rho(1) \dots \rho(h))'$

$$\phi_{00} = \rho(0) = 1$$

$$\begin{aligned}\phi_{11} &= \rho(1) \\ \phi_{22} &= \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}\end{aligned}$$

1.8 Modèles d'une série chronologique

1.8.1 Processu aléatoire stationnaire

On explique dans cette partie les processus autorégressifs et les processus moyenne mobile, ainsi que les processus mixtes *ARMA*.

1.8.1.1 Processus autorégressifs $AR(p)$ [7] [10]

Les premiers processus autorégressifs ont été introduits par *George Udny Yule*[7]. Dans cet article, *Yule* utilisait le modèle AR pour modéliser la série chronologique.

Définition 1.8.1. *On dira que $(X_t)_{t \in \mathbb{Z}}$ est un processus autorégressif d'ordre p , noté $AR(p)$ s'il peut être écrit sous la forme suivante :*

$$X_t = \varepsilon_t + \sum_{k=1}^p \varphi_k X_{t-k}$$

tel que : $\varphi_1, \varphi_2, \dots, \varphi_p$ sont des réels et $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

Soit encore ;

$$\Phi(\mathbf{B})X_t = \varepsilon_t.$$

avec : $\Phi(\mathbf{B}) = 1 - \sum_{k=1}^p \varphi_k \mathbf{B}^k$ est le polynôme caractéristique du processus X_t .

propriété 1.8.1. • *Un processus $AR(p)$ est toujours inversible.*

• *Un processus $AR(p)$ est causal et stationnaire si et seulement si $\Phi(z) \neq 0, \forall z \in \mathbb{C}, \text{ tel que } |z| > 1$.*

où $\Phi(z) = 1 - \sum_{k=1}^p \varphi_k z^k$

• **La fonction d'autocovariance**

$$\gamma(h) = \begin{cases} \varphi_1 \gamma(h-1) + \varphi_2 \gamma(h-2) + \dots + \varphi_p \gamma(h-p) + \sigma_\varepsilon^2 & \text{si } h = 0 \\ \varphi_1 \gamma(h-1) + \varphi_2 \gamma(h-2) + \dots + \varphi_p \gamma(h-p) & \text{si } h = 1, \dots, p \end{cases}$$

• **La fonction d'autocorrélation**

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} \varphi_1 \rho(h-1) + \varphi_2 \rho(h-2) + \cdots + \varphi_p \rho(h-p) + \frac{\sigma_\varepsilon^2}{\gamma(0)} & \text{si } h = 0 \\ \varphi_1 \rho(h-1) + \varphi_2 \rho(h-2) + \cdots + \varphi_p \rho(h-p) & \text{si } h = 1, \dots, p \end{cases}$$

• **La fonction d'autocorrélation partielle** Les autocorrélations partielles d'un processus $AR(p)$ sont nulles à partir du rang $p+1$.

Remarque 1.8.1. Si le polynôme Φ a toutes ses racines à l'extérieur du disque unité, alors on peut inverser ce polynôme et écrire $AR(p)$ sous la forme $MA(\infty)$

$$X_t = \Phi^{-1}(\mathbf{B})\varepsilon_t = \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i}$$

Avec : $\varphi_0 = 1$ et $\sum_{i=0}^{+\infty} |\psi_i| < \infty$

1.8.1.2 Processus moyenne mobile $MA(q)$ [3]

Définition 1.8.2. On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus MA ou encore moyenne mobile d'ordre q , noté $MA(q)$ si :

$$X_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \forall t \in \mathbb{Z}$$

tel que : $\theta_1, \theta_2, \dots, \theta_q$ sont des réels ; $\theta_q \neq 0$ et $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

On utilise généralement la notation $X_t = \Theta(\mathbf{B})\varepsilon_t$ et l'on appelle écriture polynomiale, telle que :

$$\Theta(\mathbf{B}) = 1 + \sum_{i=1}^q \theta_i \mathbf{B}^i \quad (\theta_0 = 1)$$

propriété 1.8.2. • Un processus $MA(q)$ est toujours stationnaire et causal.

• Un processus $MA(q)$ est inversible si et seulement si son polynôme $\Theta(z)$ est tel que $\Theta(z) \neq 0, \forall z \in \mathbb{C}, \text{ tel que } : |z| > 1$.

• **La fonction d'autocovariance**

$$\gamma(h) = \begin{cases} \sum_{j=0}^q \theta_j^2 \sigma_\varepsilon^2 & \text{si } h = 0. \\ \sum_{j=0}^{q-|h|} \theta_j \theta_{j+h} \sigma_\varepsilon^2 & \text{si } |h| = 1, \dots, q. \\ 0 & \text{si } |h| > q. \end{cases}$$

• **La fonction d'autocorrélation**

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0. \\ \frac{\sum_{j=0}^{q-|h|} \theta_j \theta_{j+h}}{\sum_{j=0}^q \theta_j^2} & \text{si } |h| = 1, \dots, q. \\ 0 & \text{si } |h| > q. \end{cases}$$

• **La fonction d'autocorrélation partielle** Les autocorrélations partielles sont solution d'une équation de récurrence linéaire simple d'ordre q . Elle décroissant vers 0 de manière exponentielle.

Remarque 1.8.2. Si le polynôme Θ a toutes ses racines à l'extérieur du cercle unité, alors on peut inverser ce polynôme et écrire un $MA(q)$ sous la forme d'un $AR(\infty)$

$$\varepsilon_t = \sum_{i=0}^{+\infty} v_i X_{t-i}$$

Avec : $v_0 = 0$ et $\sum_{i=0}^{+\infty} |v_i| < \infty$

1.8.1.3 Processus mixtes $ARMA(p, q)$ [3] [10]

Définition 1.8.3. Le processus $(X_t)_{t \in \mathbb{Z}}$ satisfait une représentation d'un processus autorégressif à moyenne mobile d'ordre p et q , noté $ARMA(p, q)$ s'il vérifie la relation suivante :

$$\Phi(\mathbf{B})X_t = \Theta(\mathbf{B})\varepsilon_t$$

telle que : $\Phi(\mathbf{B}) = 1 - \sum_{k=1}^p \varphi_k \mathbf{B}^k$ est un polynôme de degré p en \mathbf{B}

$\Theta(\mathbf{B}) = 1 + \sum_{i=1}^q \theta_i \mathbf{B}^i$ est un polynôme de degré q en \mathbf{B}

Les coefficients $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ sont des réels, avec $\varphi \neq 0, \theta_q \neq 0, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

propriété 1.8.3. Soit $(X_t)_{t \in \mathbb{Z}}$ est un processus $ARMA(p, q)$, supposons que les polynômes $\Phi(z) = 1 - \sum_{k=1}^p \varphi_k z^k$ et $\Theta(z) = 1 + \sum_{i=1}^q \theta_i z^i$

• $(X_t)_{t \in \mathbb{Z}}$ est stationnaire et causale si et seulement si $\Phi(z) \neq 0, \forall z \in \mathbb{C}$ tel que $|z| > 1$ Ou $\sum_{k=0}^{\infty} \varphi_k z^k = \frac{\Theta(z)}{\Phi(z)}$ si $|z| < 1$.

• $(X_t)_{t \in \mathbb{Z}}$ est inversible si et seulement si $\Theta(z) \neq 0, \forall z \in \mathbb{C}$ tel que $|z| > 1$ Ou $\sum_{i=0}^{\infty} \varphi_i z^i = \frac{\Phi(z)}{\Theta(z)}$ si $|z| < 1$.

• **La fonctions d'autocovariance**

La fonction d'autocovariance $\rho(h)$ d'un processus stationnaire $ARMA(p, q)$ satisfait la relation de récurrence de la forme :

$$\gamma(h) = \begin{cases} \varphi_1 \gamma(h-1) + \varphi_2 \gamma(h-2) + \cdots + \varphi_p \gamma(h-p) & h > q. \\ \sum_{i=1}^p \varphi_i \gamma(h-i) + \sigma_\varepsilon^2 \sum_{j=h}^q \theta_j \psi_{j-h} & 0 \leq h \leq q. \end{cases}$$

• **La fonctions d'autocorrélation**

La fonction d'autocorrélation $\rho(h)$ d'un processus $ARMA(p, q)$ satisfait une relation de la forme :

$$\rho(h) = \begin{cases} \varphi_1 \rho(h-1) + \varphi_2 \rho(h-2) + \cdots + \varphi_p \rho(h-p), & h > q. \\ \sum_{i=1}^p \varphi_i \rho(h-i) + \frac{\sigma_\varepsilon^2}{\gamma(0)} \sum_{j=h}^q \theta_j \psi_{j-h} & 0 \leq h \leq q. \end{cases}$$

Remarque 1.8.3. L'écriture $MA(\infty)$ et $AR(\infty)$ d'un processus $ARMA(p, q)$

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $ARMA(p, q)$ canonique (ie : inversible et causal) :

$$\Phi(\mathbf{B})X_t = \Theta(\mathbf{B})\varepsilon_t$$

Alors :

1. Il admet une écriture $MA(\infty)$:

$$\begin{aligned} X_t &= \Phi^{-1}(\mathbf{B})\Theta(\mathbf{B})\varepsilon_t \\ &= \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i} \end{aligned} \tag{1.1}$$

Où $(\psi_i)_{i \in \mathbb{N}}$ est une réelle.

En posant $\psi_i = 0$ pour $i < 0$, $\theta_0 = 1$, on a :

$$\forall i \in \mathbb{N} : \psi_i - \sum_{j=1}^{+\infty} \varphi_j \psi_{i-j} = \theta_i.$$

2. Il admet une écriture $AR(\infty)$:

$$\begin{aligned} \varepsilon_t &= \Phi^{-1}(\mathbf{B})\Theta(\mathbf{B})X_t \\ &= X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i} \Leftrightarrow X_t = \varepsilon_t - \sum_{i=1}^{+\infty} \pi_i X_{t-i} \end{aligned} \tag{1.2}$$

Où $(\pi_i)_{i \in \mathbb{N}}$ est une réelle.

En posant $\pi_i = 0$ pour $i < 0$, on a :

$$\forall i \in \mathbb{N} : \pi_i + \sum_{j=1}^q \theta_j \pi_{i-j} = -\varphi_i$$

1.8.2 Processus aléatoire non stationnaire

Les processus ARIMA et SARIMA sont la généralisation des modèles ARMA pour des processus aléatoire non stationnaires admettant une tendance (ARIMA) ou encore une tendance et une saisonnalité (SARIMA)

1.8.2.1 Les processus ARIMA

Définition 1.8.4. On dit qu'un processus $(X_t)_{t \in \mathbb{Z}}$ admet une représentation ARIMA(p, d, q) s'il satisfait

$$(1 - \mathbf{B})^d \Phi(\mathbf{B})X_t = \Theta(\mathbf{B})\varepsilon_t$$

Pour tout $t \in \mathbb{Z}$

$$\Phi(\mathbf{B}) = 1 - \sum_{k=1}^p \varphi_k \mathbf{B}^k$$

$$\Theta(\mathbf{B}) = 1 + \sum_{i=1}^q \theta_i \mathbf{B}^i$$

Les coefficients $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ sont des réels, avec $\varphi \neq 0, \theta_q \neq 0$, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. Les processus ARIMA(p, d, q) sont donc bien adaptés aux séries temporelles présentant une tendance polynomiale de degré $d - 1$

1.8.2.2 Les processus SARIMA

Définition 1.8.5. Soient p, d, q et $s \geq 0$, un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus SARIMA(p, d, q)(P, D, Q) s'il vérifie l'équation :

$$\Delta^d \Phi_p(\mathbf{B}) \Delta_s^D \Phi_P(\mathbf{B}^s) X_t = \Theta_q(\mathbf{B}) \Phi_Q(\mathbf{B}^s) \varepsilon_t$$

$$(1 - \mathbf{B})^d \Phi_p(\mathbf{B}) (1 - \mathbf{B}^s)^D \Phi_P(\mathbf{B}^s) X_t = \Theta_q(\mathbf{B}) \Theta_Q(\mathbf{B}^s) \varepsilon_t$$

Où :

$$\Phi_P(\mathbf{B}^s) = 1 + \Phi_1 \mathbf{B}^s + \dots + \Phi_P \mathbf{B}^{Ps}$$

$$\Theta_Q(\mathbf{B}^s) = 1 + \Theta_1 \mathbf{B}^s + \dots + \Theta_Q \mathbf{B}^{ps}$$

les polynômes $\Phi_p(\mathbf{B}), \Phi_P(\mathbf{B}), \Theta_q(\mathbf{B}), \Theta_Q(\mathbf{B})$ sont à coefficient inconnus.

- s correspond à la période du processus SARIMA qu'on peut identifier en regardant l'autocorrélogramme.
- Les entiers d et D sont choisis de sorte que la série différenciée $= (1 - \mathbf{B})^d (1 - \mathbf{B}^s)^D X_t$ soit stationnaire.
- Les ordres p et q s'obtiennent comme pour les modèles ARMA(p, q) (autocorrélation partielle et simple).
- Les ordres P et Q en regardant les ordres de s de l'autocorrélogramme.

1.8.3 Les processus stochastique non linéaire

1.8.3.1 Processus ARCH

Définition 1.8.6. (Modèle ARCH(p)) Un modèle X_t est dit autorégressif conditionnellement hétéroscédastique d'ordre p qu'on note par ARCH(p) qui prend la forme suivante : $X_t = \sigma_t \varepsilon_t$ dont la variance conditionnelle $\sigma_t^2 = \mathbb{E}[X_t^2 \mid X_1, \dots, X_{t-1}]$ satisfaite pour tout $t \in \mathbb{Z}$ et $\alpha_0 > 0$, $\alpha_1, \alpha_2, \dots, \alpha_p$ des constantes positives données la relation suivantes :

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

Où ε_t est un bruit blanc. Sous-hypothèse de stationnarité de X_t , la variance in conditionnelle existe pour $\alpha_0 > 0$, $\alpha_1 + \alpha_2 + \dots + \alpha_p < 1$ et on a $\text{Var}(X_t) = \frac{\alpha_0}{1 - (\alpha_1 + \alpha_2 + \dots + \alpha_p)}$.

1.8.3.2 Processus GARCH

Définition 1.8.7. Un processus GARCH(p, q) est défini par :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Où : $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1, \dots, p, j = 1, \dots, q, \varepsilon_t$ est un bruit blanc.

Proposition 1.8.1. Une condition nécessaire et suffisante de stationnarité du processus GARCH(p, q) s'écrit :

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$$

1.9 Tests sur les résidus

1.9.1 Test d'autocorrélation des résidus

Le test le plus usuel est le test proposé par Box-pièce (1970) il est établi à partir de la statistique :

$$Q = T \sum_{h=1}^k \hat{\rho}^2(h)$$

$\hat{\rho}^2(h)$ étant la corrélation empirique entre les résidus, il permet de vérifier l'hypothèse :

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 \text{ (absence d'autocorrélation)} \\ H_1 : \text{il existe au moins un } \rho_i \text{ significativement différent de } 0 \end{cases}$$

Cette statistique Q en l'absence d'auto-corrélation obéit à un χ^2 à $k-p-q$ degré de liberté, où :

p : l'ordre de la partie AR

q : l'ordre de la partie MA

k : le nombre de retards choisis pour calculer les auto-corrélations, Ce nombre est généralement pris de l'ordre de $T^{1/2}$, d'autres études de simulation suggèrent le choix : nombre de retards $\ln(T)$. Une version asymptotiquement équivalente à été introduite par Ljung-Box est la suivante :

$$Q' = T(T+2) \sum_{h=1}^k \frac{\hat{\rho}^2(h)}{T-h}$$

On rejette H_0 si $Q > \chi_{k-p-q}^2$, et $Q' > \chi_{k-p-q}^2$.

1.9.2 Test de normalité des résidus

On utilise le test de Jarque-Bera(1980), les hypothèses du test sont :

$$\begin{cases} H_0 : \varepsilon_t \text{ suit une loi } N(0, 1) \\ H_1 : \varepsilon_t \text{ ne suit pas la loi } N(0, 1) \end{cases}$$

La statistique du test est :

$$JB = \frac{T}{6}B_1^2 + \frac{T}{24}(B_2 - 3)^2$$

où : $B_1 = \frac{1}{T} \sum_{i=1}^T \left(\frac{X_i - \bar{X}}{S}\right)^3$ et $B_2 = \frac{1}{T} \sum_{i=1}^T \left(\frac{X_i - \bar{X}}{S}\right)^4$

La loi asymptotique de JB est la distribution de Khi deux à deux degrés de liberté tel que :

$\frac{T}{6}B_1\chi_{(1)}^2$ et $\frac{T}{24}(B_2 - 3)^2\chi_{(1)}^2$.

On rejette l'hypothèse de normalité au seuil α si $JB > \chi_{(\alpha,2)}^2$.

1.10 Critères d'information

Revenons sur le choix de l'ordre d'un ARMA(p,q). Nous avons vu précédemment que l'étude des auto-corrélations et auto-corrélations partielles peuvent permettre de présélectionner un certain nombre des modèles, une fois les paramètres de ces modèles sont estimés on peut sélectionner celui qui minimise les critères suivants :

1.10.1 AIC (AKaiKe Information Criterion) [1]

Qui sera généralement préféré si l'objectif de l'étude est de faire de la prévision et qui est définie par :

$$AIC(p, q) = -2 \log L + 2(p + q)$$

1.10.2 BIC (Bayesian Information Criterion)

Sera quand à lui généralement préféré si l'objectif de l'étude est de s'ajuster à la série observée est définie par

$$BIC(p, q) = -2 \log L + (p + q) \log T$$

Où

$\log L$: la log-vraisemblance du modèle ARMA(p, q) estimé.

T : le nombre d'observations.

Les modèles ayant la plus petite valeur du critère devront être choisis.

1.11 Rappel sur le test ARCH

Rappelons que le test *ARCH* a été introduit par Engle[6]. On considère la série X_t générée par le processus suivant :

$$\begin{cases} \Phi(B)X_t = \Theta(B)\varepsilon_t \\ \sigma_\varepsilon^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 \end{cases}$$

Les hypothèses du test :

$$\begin{cases} H_0 : \alpha_1 = \dots = \alpha_p = 0 \\ H_1 : \text{au moins un coefficient } \alpha_i (i = 1, \dots, p) \text{ est différent de zéro} \end{cases}$$

Si l'hypothèse nulle est acceptée, la variance conditionnelle est constante. En revanche, si l'hypothèse nulle est rejetée, les résidus suivent un processus *ARCH*(p).

La mise en oeuvre du test est simple et peut s'effectuer en trois étapes :

- Etape 1 : On estime l'équation de la moyenne, on récupère les résidus estimés $\hat{\varepsilon}_t$ et l'on calcule la série des $\hat{\varepsilon}_t^2$.
- Etape 2 : On régresse $\hat{\varepsilon}_t^2$ sur une constante et sur des p valeurs passées (seuls les retards significatifs sont conservés).
- Etape 3 : On calcule la statistique TR^2 où T est le nombre d'observations et R^2 est le coefficient de détermination associée à la régression de l'étape 2.

Sous l'hypothèse nulle d'homoscédasticité, la statistique TR^2 suit une loi de Khi-deux à p degrés de liberté. La règle de décision est :

- Si $TR^2 < \chi_p^2$ l'hypothèse nulle est acceptée : il n'existe pas d'effet *ARCH*.
- Si $TR^2 \geq \chi_p^2$ on rejette l'hypothèse nulle et en faveur de l'hypothèse alternative d'hétéroscédasticité conditionnelle.

1.12 Modèle AR avec erreurs Hétéroscédastique

On considère dorénavant le résidu d'un modèle linéaire. Prenons l'exemple d'un modèle linéaire autorégressif d'ordre p avec résidus de type $ARCH(q)$:

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

tel que :

$$\varepsilon_t = z_t \sigma_\varepsilon$$

avec :

$$\sigma_\varepsilon^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$$

où $(z_t)_t$ est un bruit blanc.

Les tests de rupture

Introduction

Une rupture est définie comme un changement dans une série temporelle dans le comportement des données à partir d'un certain point. La majeure partie du temps, la moyenne des données avant et après la rupture ne seront pas les mêmes, cela s'appelle une rupture abrupte. Dans d'autre cas, une tendance apparaît après la rupture, d'où le non déclenchement de tendance.

Dans ce travail, nous nous intéressons particulièrement aux ruptures abrupt et le déclenchement de tendance. Plusieurs testes de ruptures existent déjà dans la littérature.

Dans ce chapitre, nous nous focaliserons uniquement sur les tests paramétriques et les tests non paramétriques qui revêtent un intérêt particulier pour notre étude.

2.1 Tests de rupture non paramétrique

Les tests non-paramétriques ne nécessitent pas d'hypothèse sur la distribution des données des séries chronologiques. Ces tests permettent de détecter des tendances ou des changements mais ne sont généralement pas appropriés pour quantifier les pentes des tendances ou l'amplitude des changements.

2.1.1 Test de Corrélacion Sur Le Rang

Le test de corrélation sur le Rang détecter la rupture en tendance est basé sur le calcul du nombre P de paires (X_i, X_j) pour lequel $X_i < X_j$, $(i < j, i = 1, \dots, T - 1)$. Le nombre P est

obtenu en comparant la valeur du premier terme de la série X_1 avec les autres termes suivants jusqu'au dernier X_T et le nombre des termes dont la valeur dépasse X_1 est compté et ainsi de suite pour les termes jusqu'à X_{T-1} .

L'hypothèse de test sont :

$$\begin{cases} H_0 : \text{la série est stationnaire. (n'est ya pas de tendance)}. \\ H_1 : \text{la série n'est pas stationnaire (il ya un tendance)}. \end{cases}$$

On définit la variable W par :

$$W = 4P/(T(T - 1)) - 1$$

Suite une distribution normale de moyenne nulle et de variance égale à :

$$\sigma_t^2 = 2(2T + 5)/9T(T - 1)$$

Pour un risque α de première espèce donnée, l'acceptation de (H_0) est définie par l'appartenance de W à l'intvalle $[-U_{(1-\alpha/2)}\sigma_t, U_{(1-\alpha/2)}\sigma_t]$ ou $U_{(1-\alpha/2)}$ est la valeur de la variable normale réduite de probabilité de non dépassement $1 - \alpha/2$. L'hypothèse alternative de ce teste est celle d'une tendance (celle le cas ou la série chronologique est non stationnaire).

2.1.2 Test de Pettitt

Le test de Pettitt un test permettant de détecter une rupture en moyenne dans une série. Pour décrire ce test, Pettitt considère une séquence de variables aléatoire indépendantes X_1, X_2, \dots, X_T , cette séquence est censée contenir un point de rupture à τ si les X_t pour $t = 1, \dots, \tau$. l'hypothèse nulle de "non rupture" ie $H_0 : \tau = T$ contre l'hypothèse alternative de "rupture" ie $H_1 : 1 \leq \tau < T$

On définit la variable :

$$U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T D_{ij}$$

telle que :

$$D_{ij} = \text{sign}(X_i - X_j)$$

avec :

$$\text{sign}(X) = \begin{cases} 1 & \text{si } X > 0. \\ 0 & \text{si } X = 0. \\ -1 & \text{si } X < 0. \end{cases}$$

la statistique $U_{t,T}$ est considérée pour les valeurs de t comprise entre 1 et T . Pour tester H_0 , Pettitt propose d'utiliser la variable

$$K_T = \max |U_{t,T}|$$

en utilisant la théorie des rangs, Pettitt donne la probabilité de dépassement approximative d'une valeur K par :

$$\text{prob}(K < K_T) \sim 2\exp(-6K^2/(T^3 + T^2))$$

Pour un risque α

On rejete H_0 si cette probabilité est inférieure à α , sinon on accepte H_0 .

si on rejete H_0 , dans ce cas la rupture intervient au temps τ ($t = \tau$) définition K_T ie :

$$\tau = \operatorname{argmax}_{1 < t < T} (|U_{t,T}|)$$

2.1.3 Test de Lombard (1987)

Dans ce test nous supposons que les variables aléatoires X_1, \dots, X_T sont indépendantes. Les statistiques proposées par Lombard (1987) sont basées sur les rangs r_1, \dots, r_T des observations X_1, \dots, X_T et sur les scores associées.

$$\phi(1/T + 1), \dots, \phi(T/T + 1)$$

ou ϕ est une fonction score satisfaisant $0 < \int_0^1 \phi^2(w)dw < \infty$, Lombard suggère d'utiliser la fonction suivante :

$$\phi(u) = 2u - 1 \text{ (rupture en moyenne).}$$

le score associée à l'observation X_i est défini par :

$$S(X_i) = \frac{\phi(r_i/(T + 1)) - \mathbb{E}(\phi)}{\sigma_\phi}$$

$$\phi(r_i/(T + 1)), i = 1, \dots, T.$$

$\mathbb{E}(\phi)$: la moyenne

σ_ϕ : l'écart-type échantillonales de la série.

Si une rupture est intervenue entre les temps τ_1 et τ_2 ($\tau_1 < \tau_2$) alors nous devrions observer une grand valeur pour :

$$L(\tau_1, \tau_2) = \sum_{j=\tau_1+1}^{\tau_2} \sum_{k=1}^j S_{(X_k)}$$

L'hypothese

$$\begin{cases} H_0 : \text{Non rupture.} \\ H_1 : \text{Une rupture.} \end{cases}$$

La statistique du test est défini par :

$$T_N = \sum_{\tau_1=1}^{N-1} \sum_{\tau_2=\tau_1+1}^N L^2(\tau_1, \tau_2)$$

si on observe une grande valeur(égale à 1), on rejette l'hypothèse nulle ,lorsque t tend vers l'infini , la distribution de T_N/T_N^5 converge vers celle de

$$T = \sum_{i=1}^{\infty} Z_i^2 / (\pi i)^4$$

Sous H_0 .

Ou $Z = (Z_1, Z_2, \dots)^t$ est un vecteur de variables aléatoires indépendantes et identiquement distribuées de loi normale standard . les points critiques pour différentes valeur de α et pour toute valeur de T :

α	0.1	0.075	0.05	0.025	0.01
T	0.0287	0.0334	0.0403	0.0525	0.0690
T_T/T_T^5	0.0289	0.0334	0.0402	0.0515	0.0662

TABLE 2.1 – Valeurs de test de Lombard de loi normale pour différentes valeurs de α .

2.1.4 Test de Mann Kendall(Kendall 1938, reprise par Renard 2006)

[11] [12]

Le test de Mann Kendall est appliqué sur toutes les chroniques non stationnaires. l'hypothèse (H_0) testée est l'absence de tendance. la statistique calculée est définie comme

suite :

$$MK = \sum_{i=1}^{T-1} \sum_{j=i+1}^T \text{sign}(X_j - X_i)$$

telle que :

$$\text{sign}(Y) = \begin{cases} 1 & \text{si } X > 0. \\ 0 & \text{si } X = 0. \\ -1 & \text{si } X < 0. \end{cases}$$

avec (X_1, \dots, X_T) est une série. L'utilisation de la statistique normalisée.

$$MK_{(norm)} = \begin{cases} (MK - 1)/\sigma_{MK} & \text{si } MK > 0. \\ 0 & \text{si } MK = 0. \\ (MK + 1)/\sigma_{MK} & \text{si } MK < 0. \end{cases}$$

Mann (1945) et Kendall (1975) ont démontré que $\mathbb{E}(MK) = 0$ et $\sigma_{MK} = \sqrt{(T-1)(2T+5)18}$ s'il ya des ex-aequo dans la série, l'écart-type de MK est corrigé de la façon suivante :

$$\sigma_{MK} = \sqrt{\frac{T(T-1)(2T+5) - \sum_{p=1}^g t_p(p-1)(2p+5)}{18}}$$

Ou t_p est le nombre d'égalités impliquant P valeurs.

ce test permet de détecter des tendances non nécessairement linéaire. pour une taille d'échantillon suffisamment grande ($T > 10$), la statistique $MK_{(norm)}$ est approximativement distribuée suivant une loi normale standard, de sorte que l'on rejette l'hypothèse nulle :

$$Z_{1-\alpha/2} < |MK_{(norm)}|$$

telle que $Z_{1-\alpha/2}$ est la quantile d'ordre $(1 - \alpha/2)$ d'une loi normale standard.

Soit T est le nombre de valeurs dans la chronique est X_1, X_2, \dots, X_T les valeurs ordonnées de la plus petite à la plus grande. Le i ème quartile est défini par interpolation entre X_j et X_{j+1} ou j est la partie entière de $i/4(T-1) + 1$ et g est la partie fractionnaire de ce nombre. on a donc

$$g = i/4(T-1) + 1 - j$$

2.1.5 Test de Mann Kendall Modifié [13]

Ce teste ne peut être appliqué que si la chronique dispose 40 analyse ou plus. Ce test permet de prendre en compte l'autocorrélation des données dans la série chronologique.

La modification du test correspond au fait qu'un échantillon autocorrélé positivement de taille T se comporte comme un échantillon indépendant de taille T^* (et inversement pour un échantillon autocorrélé négativement). Plusieurs méthodes de calcul de MKM sont relevées dans la littérature, elle prend en compte les autocorrélations des résidus, de régression calculées aux différents rangs si celles-ci sont significatives :

$$MKM = 1 + 2/(T(T-1)(T-2)) \sum_{k=1}^{T-1} (T-k)(T-k-2)\rho_k$$

ou ρ_k est l'autocorrélation à l'ordre k , si elle est significative, $\rho_k = 0$ sinon.

Le seuil de significativité choisi pour l'autocorrélation est 0.05.

2.2 Tests de rupture paramétrique

Les tests paramétriques présument que les données des séries chronologiques suivent une distribution particulière (généralement une distribution normale). Cette condition doit être remplie pour pouvoir appliquer ces tests. Ils sont utiles car ils permettent de quantifier les changements dans les données (amplitude du changement de moyenne ou de tendance) et sont généralement plus puissants que les tests non paramétriques.

2.2.1 Le test de Dickey-Fuller (DF)

Dickey et Fuller (1979, 1981) développent trois modèles pour tester le caractère stationnaire d'une série, l'hypothèse nulle est la présence d'une racine unitaire, mais les hypothèses alternatives diffèrent puisqu'elles proposent une représentation particulière de la série stationnaire.

Soit le modèle autorégressif suivant :

$$X_t = \rho X_{t-1} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

Les trois modèles de base sont :

Modèle[1] : modèle sans constante, ni tendance déterministe.

$$(1 - \rho\mathbf{B})X_t = \varepsilon_t \rightarrow X_t = \rho X_{t-1} + \varepsilon_t$$

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

$$(1 - \rho\mathbf{B})(X_t - \mu) = \varepsilon_t, \rightarrow X_t = \rho X_{t-1} + \alpha + \varepsilon_t$$

Modèle[3] : modèles avec constante et composante déterministe.

$$(1 - \rho\mathbf{B})(X_t - \alpha - \beta t) = \varepsilon_t \rightarrow X_t = \rho X_{t-1} + \alpha + \beta t + \varepsilon_t$$

Si $\rho = 1$: cela signifie qu'une des racines de polynômes retard est égale à 1, donc on est en présence d'une racine unitaire, ce qui signifie que X_t est un processus non stationnaire.

On teste l'hypothèse nulle de racine unitaire (X_t non stationnaire) contre l'hypothèse alternative d'absence de racine unitaire (X_t est stationnaire) pour chacun des trois modèles considérés :

Modèle[1] :

$$\begin{cases} H_0 : \rho = 1 \iff X_t = X_{t-1} + \varepsilon_t \\ H_1 : |\rho| < 1 \iff X_t = \rho X_{t-1} + \varepsilon_t. \end{cases}$$

Sous l'hypothèse nulle X_t suit un processus de marche aléatoire sans dérive ; sous l'hypothèse alternative X_t suit un processus $AR(1)$.

Modèle[2] :

$$\begin{cases} H_0 : \rho = 1 \iff X_t = X_{t-1} + \varepsilon_t \\ H_1 : |\rho| < 1 \iff X_t = \rho X_{t-1} + \alpha + \varepsilon_t. \end{cases}$$

Sous l'hypothèse nulle X_t correspond à un processus de marche aléatoire avec dérive et tendance ; sous l'hypothèse alternative X_t suit un processus $AR(1)$ avec dérive.

Modèle[3] :

$$\begin{cases} H_0 : \rho = 1 \iff X_t = X_{t-1} + \alpha + \varepsilon_t \\ H_1 : |\rho| < 1 \iff X_t = \rho X_{t-1} + \alpha + \beta t + \varepsilon_t. \end{cases}$$

Sous l'hypothèse nulle, X_t suit une marche aléatoire avec dérive, sous l'hypothèse alternative .

Règle de décision

On estime par les *MCO* le paramètre ρ noté $\hat{\rho}_1$ pour les modèles [1] ,[2], [3] l'estimation des coefficients et des écarts types du modèle par les *MCO* fournit $t_{\hat{\rho}_1}$, qui est analogue à la statistique de Student (rapport du coefficient sur son écart type).

Si $t_{\text{tabulé}} \leq t_{\hat{\rho}_1}$ on accepte l'hypothèse H_0 , il existe une racine unitaire, le processus est non stationnaire.

Si $t_{\text{tabulé}} > t_{\hat{\rho}_1}$ on accepte l'hypothèse H_0 , le processus est stationnaire.

En pratique on estime les modèle sous la forme suivante :

Modèle[1] :

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

modèle sans constante, ni tendance déterministe.

Modèle[2] :

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

modèle avec constante, sans tendance.

Modèle[3] :

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

modèle avec constante et tendance.

Avec pour chaque modèle $\phi = (\rho - 1)$, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

On teste l'hypothèse nulle $\phi = 0$ (la non stationnarité) contre l'hypothèse alternative $\phi < 0$ (la stationnarité) en se référant aux valeurs tabulées par Fuller (1976) et Dickey et Fuller (1979, 1981).

Les valeurs tabulées sont présentées dans les tableaux suivantes :

Modèle[1] :

T	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
100	-2.60	-1.95	-1.61
250	-2.58	-1.95	-1.62
500	-2.58	-1.95	-1.62
∞	-2.58	-1.95	-1.62

TABLE 2.2 – Les valeurs tabulées de modèle [1]

Modèle[2] :

T	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
100	-3.51	-2.89	-2.58
250	-3.46	-2.88	-2.57
500	-3.44	-2.88	-2.57
∞	-3.43	-2.87	-2.57

TABLE 2.3 – Les valeurs tabulées de modèle [2]

Modèle[3] :

T	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
100	-4.04	-3.45	-3.15
250	-3.99	-3.43	-3.13
500	-3.98	-3.42	-3.13
∞	-3.96	-3.41	-3.12

TABLE 2.4 – Les valeurs tabulées de modèle [3]

2.2.2 Test de Jarušková (1997)

Le test de Jarušková (1997) est un test paramétrique de type «maximum». Tout comme le test de Lombard (1987), supposons que l'on observe des variables aléatoires X_1, \dots, X_T . L'hypothèse nulle stipule que $H_0 : \mu_1 = \mu_2$. L'hypothèse alternative stipule qu'il existe un point dans le temps $\tau \in \{1, \dots, T-1\}$ qui fait en sorte que le modèle est :

$$X_i \sim \begin{cases} N(\mu_1, \sigma^2) & \text{si } i = 1, \dots, \tau \\ N(\mu_2, \sigma^2) & \text{si } i = \tau + 1, \dots, T \end{cases}$$

u $\mu_1 \neq \mu_2$. En supposant que σ^2 inconnu, le test de Jarušková noté $J(T)$ est le suivant :

$$J(T) = \max_{1 < \tau < T} |J_\tau| = \max_{1 < \tau < T} \sqrt{\frac{(T-\tau)\tau}{T}} |\bar{X}_\tau - \bar{X}_\tau^*| \frac{1}{S_\tau}$$

$$\bar{X}_\tau = \sum_{j=1}^{\tau} X_j / \tau$$

$$\bar{X}_\tau^* = \sum_{j=\tau+1}^T X_j / (T - \tau)$$

$$S_\tau = \sqrt{\frac{\sum_{i=1}^{\tau} (X_i - \bar{X}_\tau)^2 + \sum_{j=\tau+1}^T (X_j - \bar{X}_\tau^*)^2}{(T - 2)}}$$

L'hypothèse nulle est rejetée lorsque la statistique $J(T)$ est "plus grande qu'une" certaine valeur critique. Si l'hypothèse nulle est vraie, alors \bar{X}_τ et \bar{X}_τ^* devraient être d'environ la même valeur. Jarušková (1997) obtient celles-ci par simulations. Une statistique tronquée peut aussi être utilisée :

$$J_1(T) = \max_{t_0 T < \tau < (1-t_0)T} |J_\tau|$$

où $t_0 \in [0, 0.5]$ (Jarušková (1997) utilise $t_0 = 0.05$). Jarušková (1997) obtient aussi les valeurs critiques par simulations et sont aussi présentées :

Valeurs critiques de $J(T)$ et $J_1(T)$

pour $\alpha = 0.05$:

T	$J(T)$	$J_1(T)$
50	3.15	3.08
100	3.16	3.06
200	3.19	3.07
300	3.21	3.08
500	3.24	3.09

TABLE 2.5 – Valeurs critiques de test de Jarušková pour $\alpha = 0.05$.

pour $\alpha = 0.01$:

T	$J(T)$	$J_1(T)$
50	3.76	3.69
100	3.71	3.62
200	3.72	3.61
300	3.73	3.62
500	3.73	3.62

TABLE 2.6 – Valeurs critiques de test de Jarušková pour $\alpha = 0.01$.

2.2.3 Tests de Reeves et al(2007) [16].

Les tests de Reeves et al. (2007) utilisent les sommes au carré des résidus et comparent leur statistique de test à une loi de Fisher dont les degrés de liberté changent dépendamment du test utilisé. Tout comme les autres tests, supposons que l'on a des variables aléatoires X_1, \dots, X_T et que celles-ci, sous l'hypothèse nulle, soient indépendantes et identiquement distribuées tandis que, sous l'hypothèse alternative, il existe un point τ tel que la moyenne des variable X_1, \dots, X_τ soit différente des variables $X_{\tau+1}, \dots, X_T$. Reeves et al. (2007) proposent les tests (méthodes) LR (Modified Lund and Reeves TPR method) et XLW (Modified Wang's TPR method) où la méthode TPR (two-phase regression) avec un point de rupture au point τ est simplement :

$$X_i = \begin{cases} \mu_1 + \beta_1 x_i + \varepsilon_i & \text{si } i = 1, \dots, \tau. \\ \mu_2 + \beta_2 x_i + \varepsilon_i & \text{si } i = \tau + 1, \dots, T. \end{cases}$$

ainsi que $x_1 \leq \dots \leq x_T$, $\varepsilon_i \sim N(0, \sigma^2)$ pour $i = 1, \dots, T$ et $\beta_1, \beta_2, \mu_1, \mu_2, \tau$ sont inconnus. Il y a cependant une contrainte de continuité de la régression au point τ se traduisant par $\mu_2 = \mu_1 + (\beta_1 - \beta_2)x_\tau$. Les deux tests suivants proposent quelques modifications au test TPR, tel que l'égalité de

β_1 et β_2 .

2.2.4 Lund et Reeves (2007)LR

Lund et Reeves (2007) ont modifié le modèle TPR en cessant d'imposer la contrainte de continuité. Le modèle LR suppose que les x_i sont des entiers. Le modèle LR est donc :

$$X_i = \begin{cases} \mu_1 + \beta_1 x_i + \varepsilon_i & \text{si } i = 1, \dots, \tau. \\ \mu_2 + \beta_2 x_i + \varepsilon_i & \text{si } i = \tau + 1, \dots, T. \end{cases}$$

où on peut alors détecter une rupture dans la moyenne ($\mu_1 \neq \mu_2$) et dans la tendance ($\beta_1 \neq \beta_2$). Les erreurs suivent une loi normale de moyenne nulle et de variance inconnue. Ce qui revient donc à confronter les hypothèses suivantes :

$$\begin{cases} H_0 : \mu_1 = \mu_2, \text{ et } \beta_1 = \beta_2 \\ H_1 : \mu_1 \neq \mu_2, \text{ et/ou } , \beta_1 \neq \beta_2 \end{cases}$$

Si le point de rupture τ est connu et fixé, le test devient simplement :

$$F_\tau = \frac{(SSE_0 - SSE_A)/2}{SSE_A/(T-4)} \sim F(2, T-4)$$

où SSE_0 et SSE_A sont respectivement les sommes des erreurs au carré sous l'hypothèse nulle et alternative. Sous H_0 , la statistique de test devrait suivre une loi de Fisher à (2, T-4) degrés de liberté. On rejette l'hypothèse nulle pour de grandes valeurs de F_τ .

Si le point de rupture τ est inconnu, la statistique devient

$$F_{max} = \max_{1 < \tau < T} |F_\tau|$$

et on rejette H_0 pour une grande valeur de F_{max} . Comme cette statistique ne suit pas de loi connue, Lund et Reeves (2007) obtiennent les valeurs critiques par simulations et sont présentées dans le tableau 2.5

2.2.5 Test de XLW [16]

Wang (2007) a apporté une petite modification au test LR car il trouvait que celui-ci expliquait mal les phénomènes climatiques. Son modèle devient donc :

$$X_i = \begin{cases} \mu_1 + \beta_1 t + \varepsilon_i & \text{si } i = 1, \dots, \tau. \\ \mu_2 + \beta_2 t + \varepsilon_i & \text{si } i = \tau + 1, \dots, T. \end{cases}$$

où les termes sont définis précédemment. Les hypothèses deviennent donc :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Si le point de rupture τ est connu et fixé, le test devient simplement :

$$F_\tau = \frac{(SSE_0 - SSE_A)}{SSE_A/(T-3)} \sim F(1, T-3)$$

où les termes sont aussi définis précédemment.

Si τ est inconnu la statistique F_{max} est une fois de plus utilisée. Wang (2007) obtient aussi les valeurs critiques par simulations et sont présentées dans le tableau .

Valeur critique pour $\alpha = 0.05$ des tests LR et XLW

T	$LR(F_{max})$	$XLW(F_{max})$
25	11.67	7.37
50	11.07	6.92
75	11.06	6.88
100	11.09	6.91
200	11.21	7.01
500	11.54	7.24

TABLE 2.7 – Valeurs critiques des tests des LR et WLW pour $\alpha = 0.05$.

Applications sur R

Définition 2.2.1. *R est un logiciel et un langage de programmation gratuits et à source ouverte développés en 1995 à l'Université d'Auckland en tant qu'environnement informatique statistique et graphique (Ihaka et Gentleman, 1996). Depuis lors, R est devenu l'un des environnements logiciels dominants pour l'analyse de données et est utilisé par diverses disciplines scientifiques, notamment les sciences du sol, l'écologie et la géoinformatique (vue Envirometrices CRAN ; vue Spatial CRAN). R est particulièrement populaire pour ses fonctionnalités graphiques , mais il est également prisé pour ses fonctionnalités SIG, qui facilitent relativement la génération de modèles à base de raster. Plus récemment, R a également acquis plusieurs packages spécialement conçus pour l'analyse des données da sol.*

- *un environnement logiciel : statistiques, graphiques, la programmation, calculatrice,...*
- *un langage pour explorer, résumer et modéliser des données.*

L'utilisation de R

- c'est un logiciel multiplateforme, qui fonctionne aussi bien sur des systèmes Linux, Mac OS X ou Windows.
- c'est un logiciel libre, développé par ses utilisateurs et modifiable par tout un chacun .
- c'est un logiciel très puissant, dont les fonctionnalités de base peuvent être étendues à l'aide de plusieurs milliers d'extensions .
- les possibilités de manipulation de données sous R sont en général largement supérieures à celles des autres logiciels usuels d'analyse statistique .
- c'est un logiciel avec d'excellentes capacités graphiques et de nombreuses possibilités d'export .
- R est de plus utilisé dans tous les secteurs scientifiques, y compris dans le domaine des analyses d'enquêtes et, plus généralement, des sciences sociales.

- le logiciel, la documentation de référence et les principales ressources sont en anglais. Il est toutefois parfaitement possible d'utiliser R sans spécialement maîtriser cette langue.

2.3 Package "change point"

Implémente diverses méthodes de point de changement classiques et spécialisées pour la recherche d'un ou de plusieurs points de changement dans les données. De nombreuses méthodes populaires et fréquentistes sont incluses.

Les fonctions `cpt.mean()`, `cpt.var()`, `cpt.meanvar()` devraient être votre premier point d'appel.

2.3.1 Test de changement de moyenne

```
set.seed(10).
x=c(rnorm(200,0,1),rnorm(200,2,3)).
mmean=cpt.mean(x).
plot(mmean,cpt.col='blue').
print(mmean).
```

Dans ce cas, on a changé la moyenne telle que le premier échantillon est de loi normale $N(0,1)$ et l'autre est de loi normale $N(2,3)$.

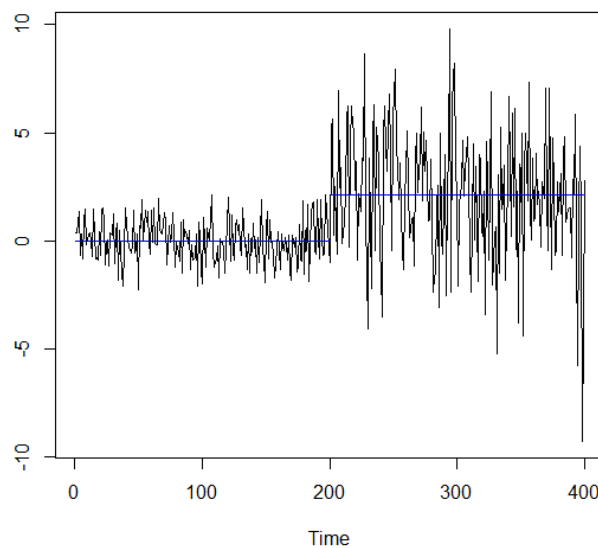


FIGURE 2.1 – L'exécution de test de changement de moyenne sur R

On remarque au milieu de la série il y a une rupture.

2.3.2 Test de changement de variance

```
w=c(rnorm(200,0,1),rnorm(200,5,6)).  
vvar=cpt.var(w).  
plot(vvar).  
print(vvar).
```

Dans ce cas, en changé la variance telle que le premier échantillon est de loi normale (0,1) et l'autre est de loi normale (5,6).

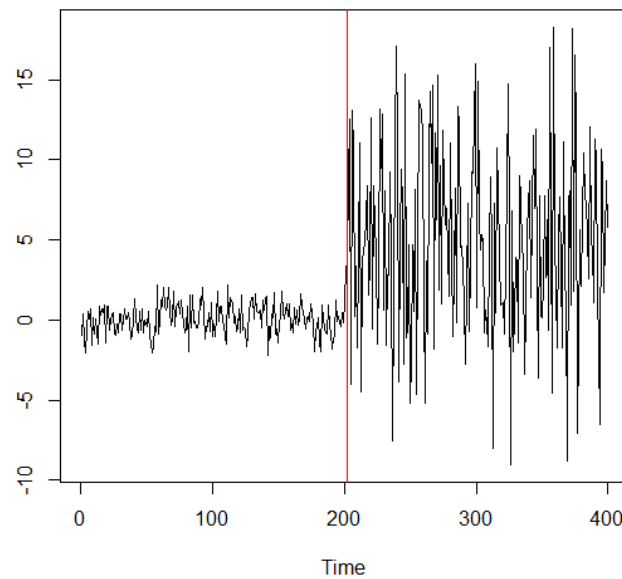


FIGURE 2.2 – L'exécution de test de changement de variance sur R

On remarque au milieu de la série il y a une rupture.

2.3.3 Test de changement de moyenne et variance

```
y=c(rnorm(200,0,1),rnorm(200,3,5)).  
meannvar=cpt.meanvar(y).  
plot(meannvar,cpt.width=3).  
print(meannvar).
```

Dans ce cas, en changé la moyenne et la variance telle que le premier échantillon est de loi normale (0,1) et l'autre est de loi normale (3,5).

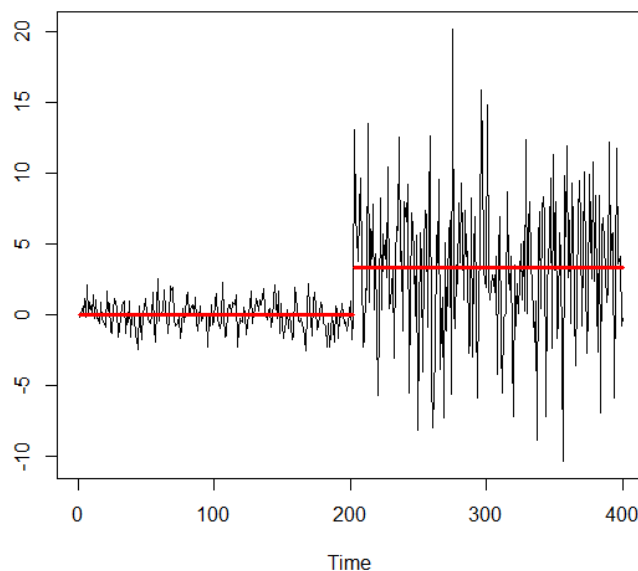


FIGURE 2.3 – Le graphe du changement de moyenne et la variance sur R

On remarque au milieu de la série il y a une rupture.

2.4 Application

Nous intéressons à appliquer la méthode de Box-Jenkins sur des données réelles notées "Pst" qui représentent le prix SPOT du pétrole brut, Nous disposons de la série mensuelle des prix spot du pétrole brut avec un total de données de 190 observations exprimées en Dollars US par baril extrait des bulletins statistiques annuels de l'OPEP couvrant la période : janvier 2000 à Octobre 2015. Dans cette section nous présenterons les données des prix SPOT du pétrole brut et leurs propriétés statistiques ensuite on essaye d'aborder l'étape de modélisation.

Les données des prix spot du pétrole, les graphes et les calculs sont effectués par le logiciel R.

2.4.1 Analyse préliminaire de la série P

2.4.1.1 Représentation graphique de la série P

Cette première étape qui précède la modélisation est très importante. Car elle repose sur l'analyse visuelle du graphe retenu des données historiques de l'évolution des prix spot du pétrole brut, et aussi sur l'étude de la stationnarité qui est la partie la plus importante dans la

phase de la modélisation de la série des prix spot du pétrole étudié. Le graphe suivant représente l'évolution des prix spot du pétrole (P) pendant la période de janvier 2000 à octobre 2015.

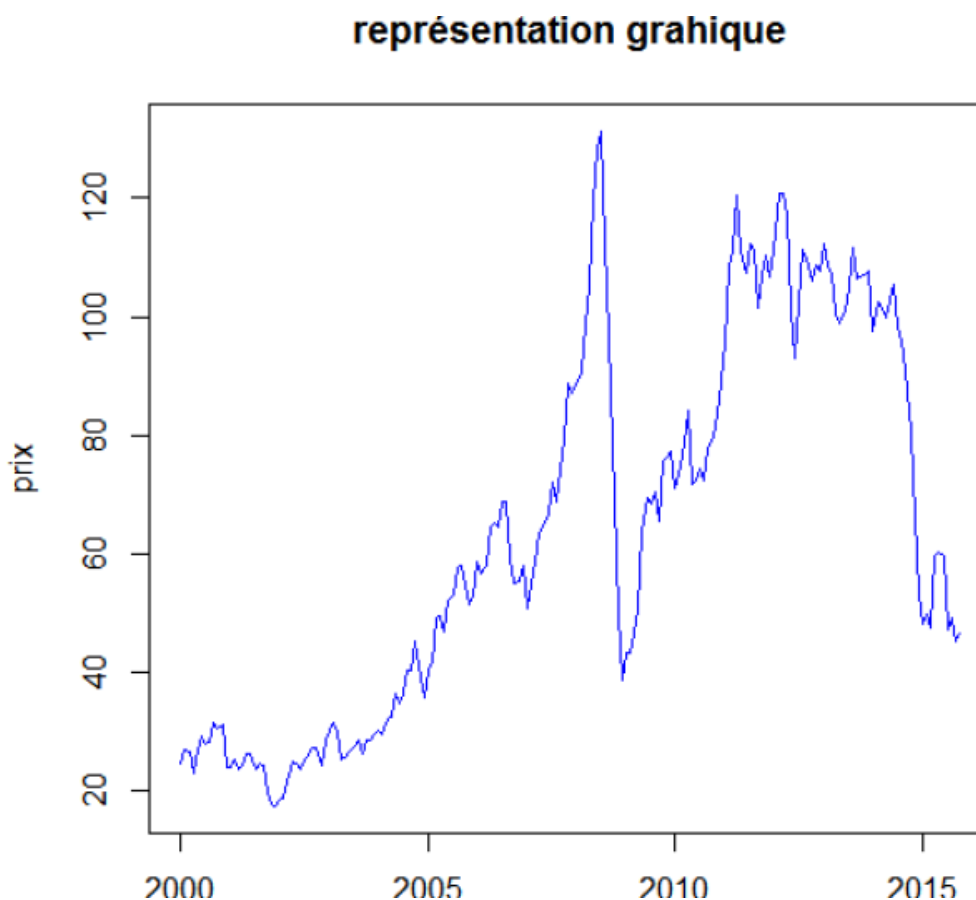


FIGURE 2.4 – L'évolution de la série (P) entre (2000-2015)

La lecture visuelle du graphe ci-dessus montre une évolution des prix spot du pétrole brut dans le temps. On observe que la série des prix spot du pétrole brut semble non stationnaire. Selon le graphe, on remarque qu'il y'a un mouvement léger à moyen terme de croissance ou décroissance révélant la présence d'une composante Tendance ou trend qui exprime donc l'évolution générale de la série.

Bien entendu, on constate également des fluctuations plus ou moins importantes que l'on appelle irrégulières ou mouvements résiduels « Stochastique ». Ces fluctuations irrégulières sont dues à des facteurs exceptionnels imprévisibles comme la crise financière (la crise de subprime 2008).

En remarque qu'il n'y'a pas de périodicité dans les valeurs observées révélant qu'il n'y a pas

une présence de saisonnalité.

Tout cela va être vérifié par le test ADF.

2.4.1.2 Etude du corrélogramme de la série P

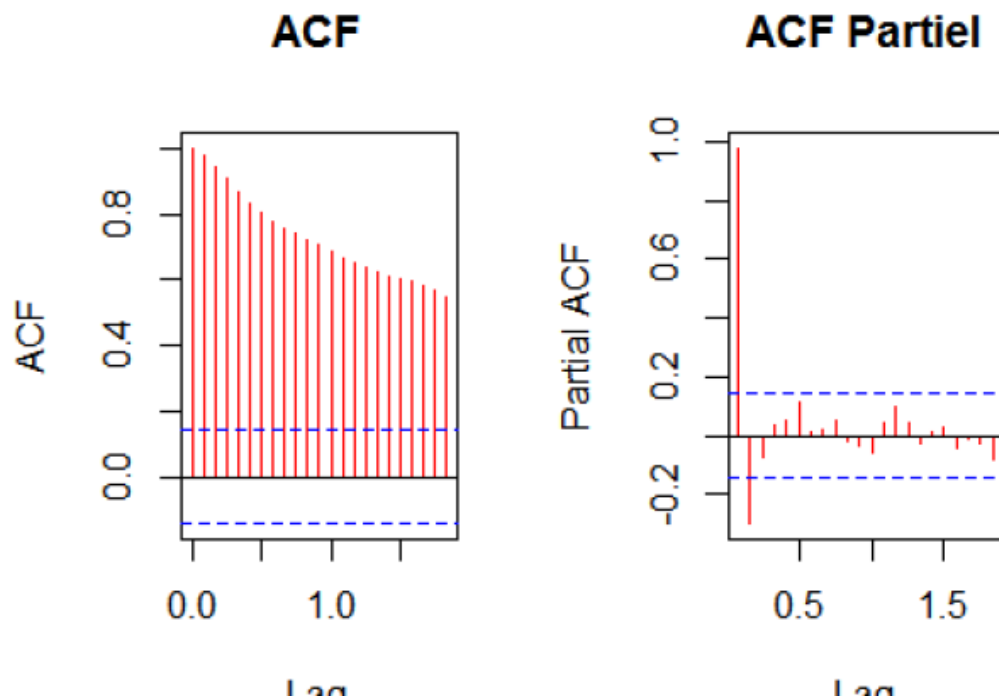


FIGURE 2.5 – Corrélogramme et corrélogramme partiel de la série P

Le Corrélogramme de la série brute des prix spot du pétrole montre des fonctions d'auto-corrélations dont les valeurs sont proches l'une de l'autre, et décroissent très lentement vers le zéro, qui sont les signes d'une non-stationnarité en tendance.

2.4.1.3 Traitement de la stationnarité

Test de la racine unitaire (Dickey-Fuller) La série a une racine unitaire car la $p - value = 0.7806 > 0.05$, (la probabilité critique affectée à la Constante est supérieure à 0.05).

D'après les résultats obtenus par le test de (Dickey-Fuller), on conclut que la série (P) est non stationnaire.

2.4.2 Le teste de Mann-Kendall

mk.test(P) Sample estimates :

<i>MK</i>	<i>varMK</i>	<i>tau</i>
$1.110400e^4$	$7.680720e^5$	$6.184866e^{-1}$

TABLE 2.8 – Les valeur critique de $MK_{(norm)}$

2.4.3 Le test de Pettitt

pettitt.test(P)

K_T	<i>P – value</i>
8241	$2.2e^{-16}$

TABLE 2.9 – Les valeur critique K_T

alternativehypothesis :two,sided.

Sample estimates :

probable change point at time $K = 75$.

$$K_{(T)} = 2exp\left[-\frac{6K^2}{(T^2 + T^3)}\right] = 1.9902343591$$

On a $\alpha(5\%) < K_{(T)}$ donc on accepte H_0 , alors on a la l'absence de rupture de la moyenne dans la série.

2.4.4 Différenciation de la série P

On note la série différenciée de la série P par (DP).

$$DP = P_t - P_{t-1}$$

2.4.4.1 La représentation graphique de la série DP

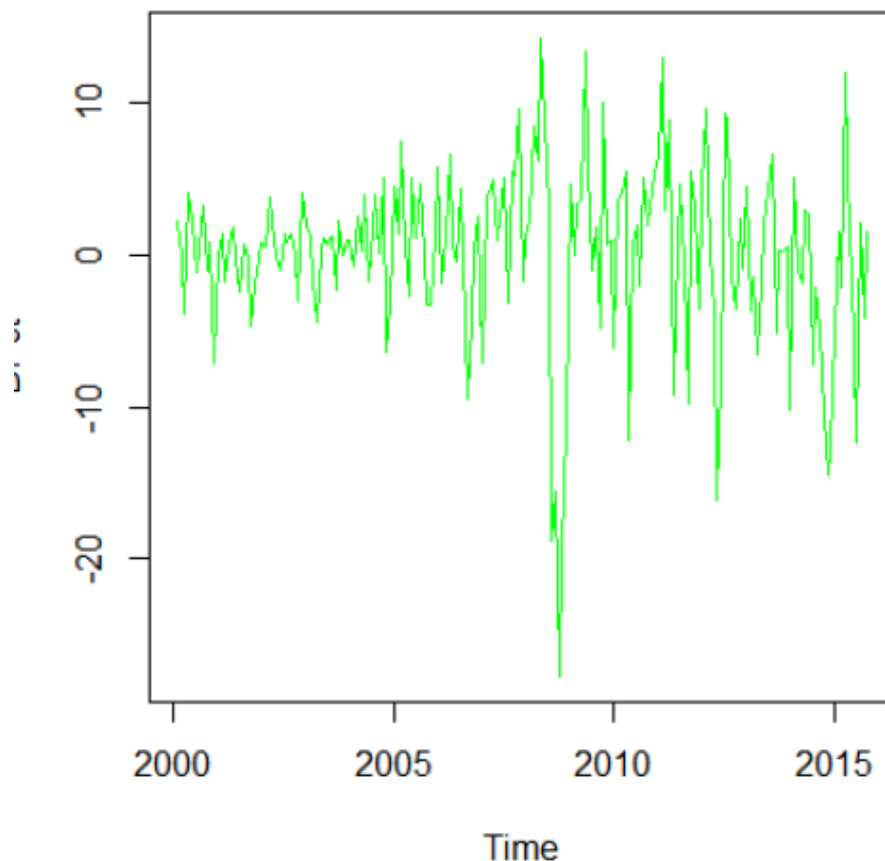


FIGURE 2.6 – La représentation graphique de la série DP

Le graphe de la série DP montre que la série est stationnaire. On utilise aussi le test de racine unitaire (test ADF) pour confirmer la stationnarité, ce test donne une p -value = $0.01 < 0.05$ (la probabilité critique affectée à la constante est inférieure à 0.05).

Le corrélogramme et le corrélogramme partiel de la série DP

2.4.5 La méthodologie de Box-Jenkins

2.4.5.1 Identification des modèles possibles :

Cette étape est effectuée par le biais de l'étude des fonctions d'autocorrélation et d'auto-corrélation partielle de la série DP. Retournons sur le corrélogramme simple et partiel de la série DP, on remarque que les deux ACF et PACF contiennent des pics significatifs, cela nous a permis de proposer plusieurs modèles.

S'appuyant sur le corrélogramme et corrélogramme partiel des fonctions d'auto-corrélations et

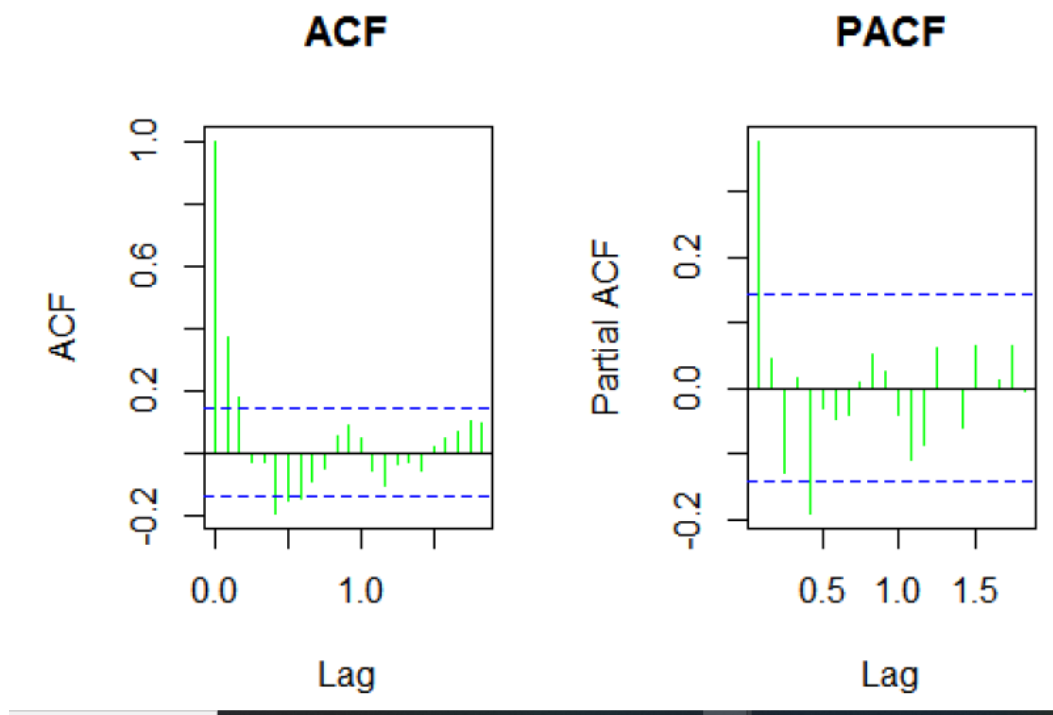


FIGURE 2.7 – Corrélogramme simple et partiel de la série DP

d'auto-corrélations partielles

- ▶ Le corrélogramme simple de la fonction d'auto-corrélation nous fournit la valeur de q .
- ▶ Le corrélogramme partiel de la fonction d'auto-corrélation partielle nous donne la valeur de p .

Ainsi on déduit les modèles suivants :

- ▶ Premier modèle : $MA(1)$
- ▶ Deuxième modèle : $AR(1)$
- ▶ Troisième modèle : $ARMA(1,1)$

2.4.5.2 Estimation et validation du modèle

Après avoir identifié l'ordre des processus MA et AR, il convient d'estimer les paramètres du modèle, puis de vérifier à la base de certains nombres de tests statistiques que l'estimation du modèle est valable pour faire les prévisions.

Le tableau présente l'estimation des modèles $ARMA(1, 1)$, $AR(1)$, $MA(1)$ avec et sans constant.

Modèles		Avec constant				Sans constant	
		coefficient		constant		coefficient	
		Estimation	P(>)	Estimation	P(>)	Estimaton	P(>)
ARMA(1,1)	AR(1)	0.43400	0.00185	0.05696	0.87725	0.43447	0.00182
	MA(1)	-0.06801	0.064373			-0.06836	0.64189
AR(1)		0.37516	2.62e-08	0.06416	0.871	0.3753	2.57e-08
MA(1)		0.29251	2.51e-07	0.11267	0.828	0.29255	2.49e-07

TABLE 2.10 – Table d'estimation des modèles ARMA(1,1),AR(1),MA(1)

Remarque 2.4.1. *En comparant toujours les p-values avec la valeur 0.05 (5%). Les modèles où les p-values sont supérieures à 0.05 seront écartés, si un des paramètres (ou des constantes) ne soient pas significativement différents de 0 le modèle sera alors rejeté.*

D'après le tableau, on remarque que tout les modèles avec constants sont rejetés car les constants ne sont pas significativement différents de 0 (p-value supérieures à 0.05).

Pour les modèles sans constants on rejette le modèle ARMA(1,1) car la p-value du coefficient supérieur à 0.05.

Les modèles AR(1) et MA(1) sans constants reste candidat dans la mesure où les coefficients sont significativement différents de zéro (les p-values sont inférieures à 0.05).

2.4.5.3 Validation des résidus

Pour tester l'autocorrélation des résidus, on fait le test de **Box-Pierce** et on trace les corrélogrammes et corrélogrammes partiels.

Modèles	Q_h	P-value	Khi-deux
AR(1)	0.052036	0.8196	21.03
MA(1)	0.85072	0.3563	21.03

TABLE 2.11 – Résultat de test de Box-pierce

Les corrélogrammes simples et partiels des résidus des modèles MA(1) et AR(1) sont :

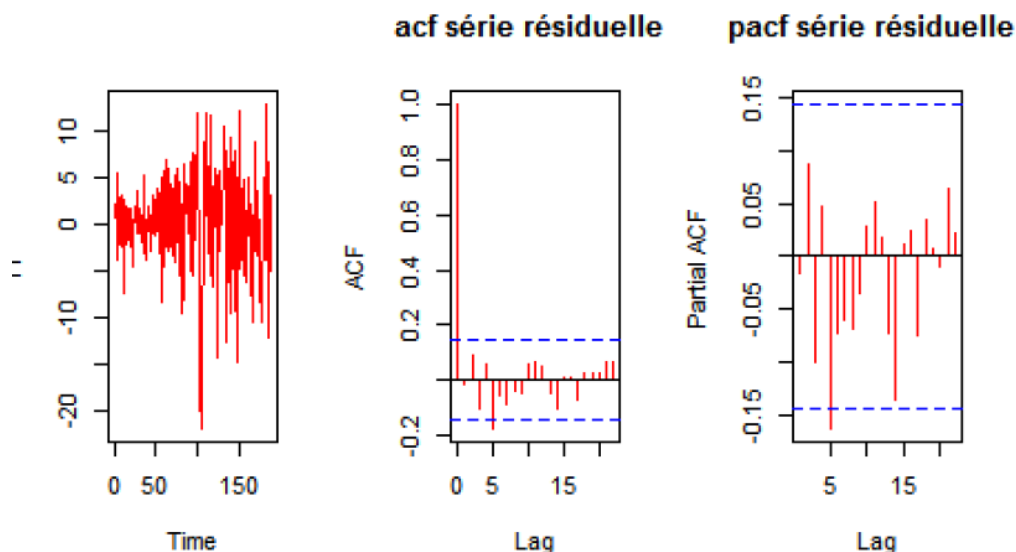


FIGURE 2.8 – Le graphe, l'ACF et PACF des résidus du modèle AR(1)

$$Q_h = 0.052036, \text{Khi-deux}(h) = \text{Khi-deux}(12) = 21.03 \text{ (au seuil 5\%)}$$

Et $p\text{-value} = 0.8196 > 0.05$ (accepter H_0), alors les pics ne sont pas significatifs. Donc on accepte l'hypothèse nulle, absence d'auto-corrélation du résidu, ce qui implique qu'il peut être assimilé à un bruit blanc. Alors le modèle est valide.

$$Q_h = 0.85072, \text{Khi-deux}(h) = \text{Khi-deux}(12) = 21.03 \text{ (au seuil 5\%)}$$

Et $p\text{-value} = 0.3563 > 0.05$ (accepter H_0), alors les pics ne sont pas significatifs. Donc on accepte l'hypothèse nulle, absence d'auto-corrélation du résidu, ce qui implique qu'il peut être assimilé à un bruit blanc. Alors le modèle est valide.

2.4.5.4 choix de modèle

Pour comparer les modèles entre eux, on cherche à minimiser les critères AIC et BIC des deux modèles. La table suivante présente ces critères :

D'après ces critères, il ressort que le modèle AR(1) dispose d'une qualité supérieure car les critères d'information de ce modèle sont minimales.

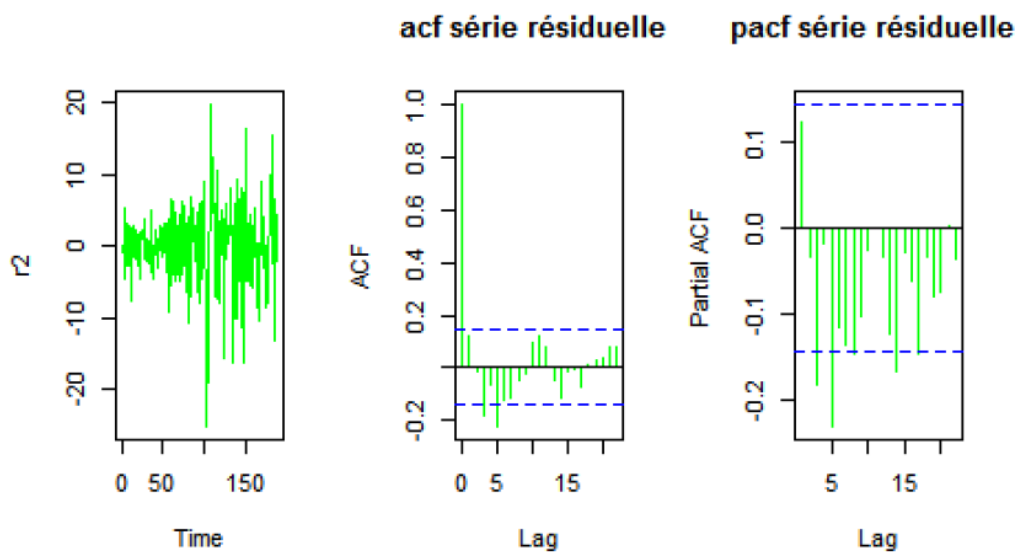


FIGURE 2.9 – Le graphe, l'ACF et PACF des résidus du modèle MA(1)

<i>Modèles</i>	<i>AIC</i>	<i>BIC</i>	σ^2
<i>AR(1)</i>	1179.31	1182.869	29.7
<i>MA(1)</i>	1185.76	1213.747	30.73

TABLE 2.12 – la table des critères du choix du meilleur modèle

2.4.5.5 Test de normalité pour les résidus du modèle AR(1)

Nous testons les hypothèses suivantes : $H_0 : s_1 = 0$ (symétrie) et $s_2 = 0$ (aplatissement normal)

► **Test de Skewness**

On a : $s_1 = -0.7049296$, donc on rejette l'hypothèse de la symétrie des résidus.

► **Test de Kurtosis**

On a : $s_2 = 1.85064$, donc on rejette l'hypothèse de l'aplatissement normal.

► **Test de Jarque Bera**

```
Jarque Bera Test
data : res
X-squared = 44.404, df = 2, p-value = 2.279e-10
```

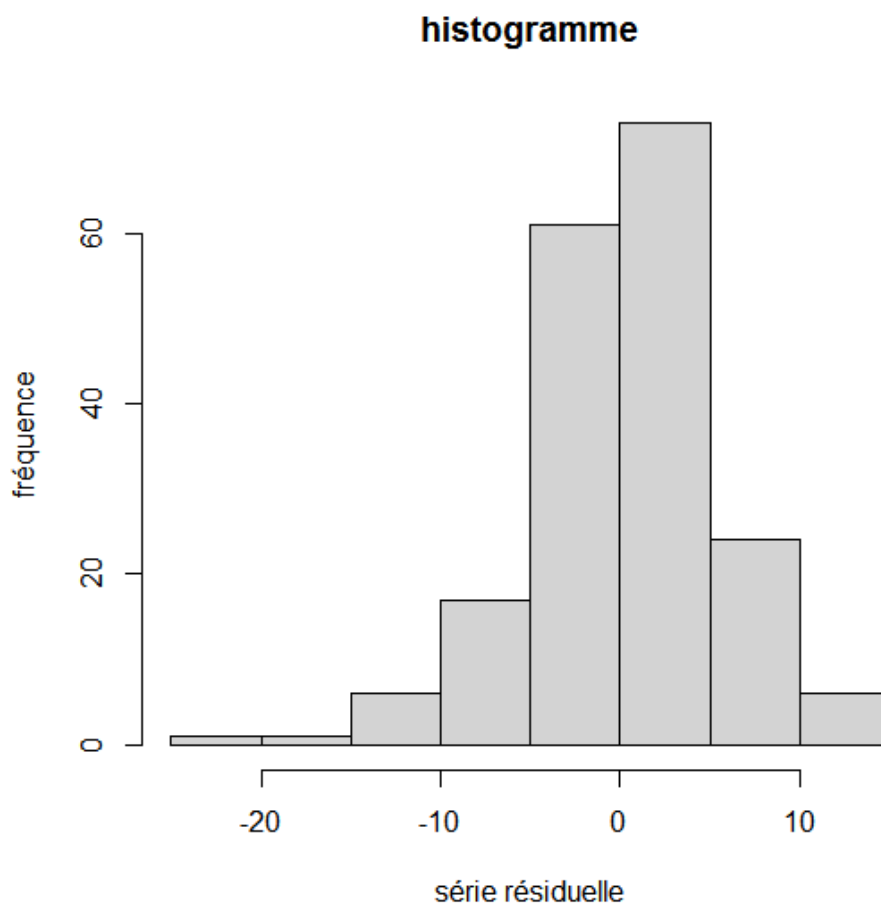


FIGURE 2.10 – l’histogramme des résidus pour le modèle AR(1)

On a $p\text{-value} < 5\%$, donc on rejette l’hypothèse de normalité des résidus. Donc la distribution des erreurs est un bruit blanc non gaussien.

2.4.5.6 Application de la modélisation Autorégressif Conditionnellement Hétéroscédastique

Plusieurs termes de l’autocorrélation partielle sont significativement différents de zéro, on retient un nombre de retards p pour réaliser un test ARCH. On corrobore ce test par celui du multiplicateur de Lagrange.

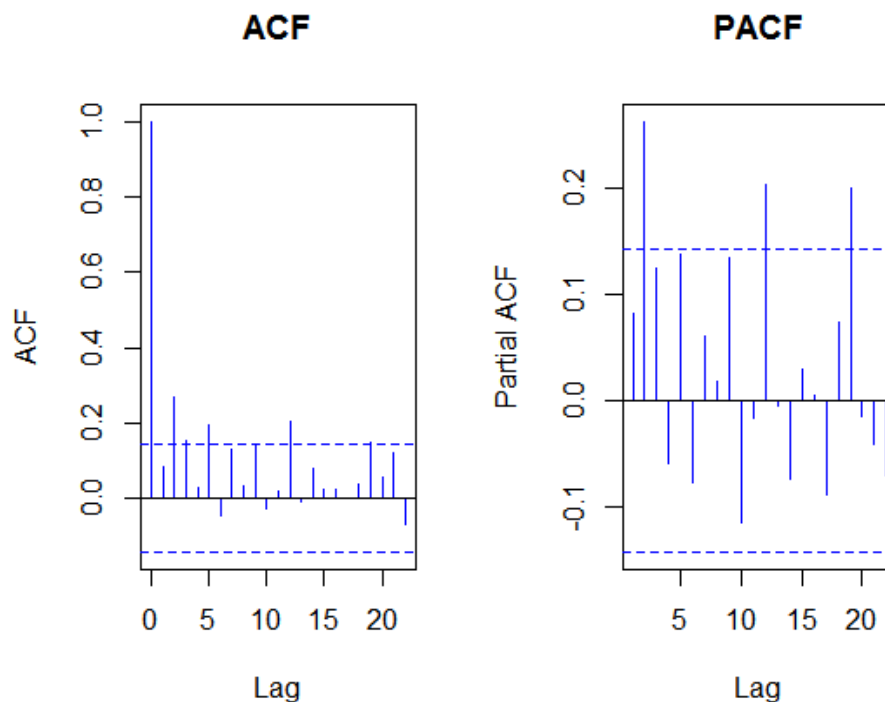


FIGURE 2.11 – Le corrélogramme des résidus carrés

► Test ARCH

Nous appliquons le test ARCH mentionné au chapitre précédent. Les résultats sont présentés ci après :

$$\text{On a : } \begin{cases} H_0 : \alpha_1 = \alpha_2 = \dots \alpha_p \text{ (homoscédasticité)} \\ H_1 : \text{au moins un coefficient est différent de 0 (hétéroscédasticité)} \end{cases}$$

Lagrange-Multiplier test		
order	LM	p.value
4	82.98	0.00e+00
8	34.56	1.35e-05
12	19.42	5.40e-02
16	11.09	7.46e-01

TABLE 2.13 – Résultats du test ARCH

Alors, on accepte l'hypothèse alternative d'hétéroscédasticité conditionnelle en faveur de l'hypothèse d'homoscédasticité.

2.4.6 Estimation des modèles GARCH sur les résidus du modèles ARMA

Error Analysis		
	Estimate	Std. Error
alpha0	0.22920	0.29099
alpha1	0.14503	0.04526
beta1	0.86629	0.03774

TABLE 2.14 – Estimation du modèle GARCH(1,1)

Error Analysis		
	Estimate	Std. Error
alpha0	26.79553	3.46302
alpha1	0.09016	0.09047

TABLE 2.15 – Estimation du modèle ARCH(1)

2.4.7 Validation du modèle

► les critères du choix de modèle

	ARCH(1)	GARCH(1,1)
AIC	6.230946	6.042943
BIC	6.265250	6.094399
SIC	6.230725	6.042449
HQIC	6.244844	6.063789

TABLE 2.16 – La table des critères

D'après ces critères, le modèle GARCH(1,1) est plus adéquat.

Alors, l'application de la méthodologie de Box-Jenkins nous a conduit à retenir un processus AR(1) avec erreur GARCH(1,1). Telle que :

$$DP_t = 0.3753DP_{t-1} + \varepsilon_t$$

Avec

$$\varepsilon_t = z_t \sigma_\varepsilon$$

$$\sigma_\varepsilon^2 = 0.22920 + 0.14503\varepsilon_{t-1}^2 + 0.86629\sigma_{t-1}^2$$

où, $(z_t)_t$ est une suite de v.a $N(0,1)$.

Conclusion :

En guise de conclusion générale, nous allons tenter d'établir une synthèse globale sur le travail qui a été réalisé dans ce mémoire. L'objectif principal de ce mémoire est d'étudier des tests de rupture qui ont illustrés par des résultats que nous avons aboutis à travers des données réelles qui sont des séries statistiques qui mesurent le prix spot du pétrole brut de puis Janvier 2000 a Octobre 2015. Les graphes et les calculs sont effectués par le logiciel R.

Premièrement, une recherche bibliographique a été établie sur les notions de base sur les séries temporelles. Ainsi en deuxième partie nous présentons les différents tests de ruptures notamment les tests de rupture paramétrique et non paramétrique par la suite nous avons démontré l'existence d'une rupture dans une série chronologique, cette rupture est expliquée par une rupture de la tendance au point $t = 75$.

Bibliographie

- [1] **Akaike, H** : Fitting autoregressive models for prediction. *Anal. of the institute of statistical mathematics*, 21(1), 243-247. (1969)

- [2] **Bollerslev, T** : Generalized autoregressive conditional heteroskedasticity. *J.Econometrics*. 31, 307-327, (1986).

- [3] **Box G. E. P, Jenkins G. M,(1976)** : *Time series Analysis : Forecasting and control*, Holden Day, San Francisco.

- [4] **Brockwell, P, J, et Davis, R.A, (1996)** : *Introduction to Time series and Forecasting*. Springer, Newyork.

- [5] **Corinne, P** : *Séries chronologiques : Quelques ment du cours*, Paris. (2005).

- [6] **Engle,R.F, (1982)** : Autoregressive conditional heteroskedastique models with estimates of the variance of the United Kingdom inflation.*Econometrica*. 50,987-1007.

- [7] **Eshel, G** : the Yule walker equations for the AR coefficients. *Internet resource*, 2, 68-73. (2003).

-
- [8] **Jarušková ,(1997)** Some Problems With Application Of Change-point Detection Methods To Environmental Data, *Environmetrics*.Vol. 8, pp. 469-483
- [9] **Florence, N** : généralités sur les séries chronologiques. (2006).
- [10] **Hamilton J. D.** Time series analysis. Princeton University Press, Princeton (1994).
- [11] **Kendall, M. (1975)**. Rank Correlation Methods. Charles Griffin, London, 4ème, édition.
- [12] **Kendall, M.G, (1955)**, Rank Correlation Méthode, New York : Hafner publishing Co.
- [13] **Khaled, H, Hamed, A,(1998)** A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*.
- [14] **Lombard, F. (1987)** Rank tests for changepoint problems. *Biometrika*. Vol. 74, No.3, pp. 615-624.
- [15] **Pettitt, A. (1979)**. A non-parametric approach to the change point problem. *Applied Statistics*, 28(2) :126–135.
- [16] **Reeves, Chen, Wang,Lund and Lu(2007)**, A Review and Comparison of Changepoint Detection Techniques for Climate Data, *Journal Of Applied Meteorology And Climatology*, Vol. 46, pp. 900-915.
- [17] **Reeves, J. K, Pierie, K, Beral, V, Green, J, Spencer, E, Bull, D, (2007)** . Cancer incidence and mortality in relation to body mass index in the Million Women Studay : cohort study. *Bmj*, 335(7630),1134.

Annex

Le programme sous R :

Application

```
x=c(24.58,26.84,26.71,22.93,26.94,29.12,27.94,28.3,31.48,30.42,31.22,24.13,24.06,25.41,23.7,24.38
,26.25,26.1,23.73,24.46,24.29,19.64,17.65,17.53,18.33,18.89,22.64,24.88,24.76,23.8,25.13,25.99,27.38,27.32,
,26.74,27.43,28.63,26.32,28.54,28.45,29.44,30.33,29.56,32.05,32.35,36.27,34.61,36.29,40.27,40.36,45.37,38.9
6.62,57.87,64.44,65.11,64.6,68.89,68.81,59.34,54.97,55.42,57.95,50.79,54.56,58.59,63.55,64.48,66.89,71.89,6
,68.59,70.37,65.55,75.56,76.20,77.16,71.01,74.6,78.7,84.13,71.88,72.49,74.43,72.39,77.48,79.42,83.65,88.99,
,112.18,111.4,101.57,107.12,110.38,106.84,111.21,120.79,120.89,117.2,101.06,92.99,102.22,111.17,109.68,10
,100.65,98.78,100.78,104.97,111.55,106.47,106.75,107.07,107.67,97.49,102.59,101.58,99.74,102.71,
105.37,98.17,95.95,91.16,80.54,66.15,53.27,48.24,49.76,47.6,59.63,60.3,59.47,47.12,49.2,45.09,46.59)
library(aTSA)
library(tseries)
library(e1071)
library(fGarch)
P=ts(x,start=c(2000,1),end=c(2015,10),frequency=12)
# transformer la à une série temporelle
par(mfcol=c(1,2))
plot(P, main="représentation graphique", col="blue", xlab="time", ylab="prix")
# tracer le graphe de la série P
acf(P, main="ACF",col="red")
# tracer les autocorrélations de la série P
pacf(P, main="ACF Partiel",col="red")
# les tests de Mann Kendall est Pettitt
```

```
library (trend)
mk.test (P)
pettitt.test(P)
# tracer les autocorrélations partielle de la série P
adf.test(P)
DP=diff(P)
plot(DP, main="Graphe de DP", col="blue")
# tracer le graphe de la série DP
acf(DP, main=" ACF",col="green")
# tracer les autocorrélations de la série DP
pacf(DP, main="PACF",col="green")
# tracer les autocorrélations partielle de la série DP
adf.test(DP)
library(forecast)
library(FitARMA)
# pour lancer le package "FitARMA"
M1=arma(DP,order=c(1,0,0),include.intercept=FALSE)
summary(M1)
L1=arma(DP,order=c(1,0,0),include.intercept=TRUE)
summary(L1)
mod=estimate(DP, p=1, d=0, q=0, intercept=FALSE)
mod1=estimate(DP, p=1, d=0, q=0, intercept=FALSE)
mod2=estimate(DP, p=0, d=1, q=1, intercept=FALSE)
r1=mod1$residuals
r2=mod2$residuals
Box.test(r1)
Box.test(r2)
par(mfcol=c(1,3))
plot(r1,col="red")
plot(r2,col="green")
acf(r1,col="red",main=" acf série résiduelle ")
pacf(r1,main="pacf série résiduelle ",col="red")
acf(r2,col="green",main=" acf série résiduelle ")
```

```
pacf(r2,main="pacf série résiduelle ",col="green")
res=mod$residuals
# donner la série des résids d'un modèle ARMA(1,0)
plot(res,xlab="temps",main="série résiduelle",col="blue")
# tracer le graphe de la série résiduelle
acf(res,main="acf série résiduelle ",col="blue")
# tracer le corrélogramme de la série résiduelle
pacf(res,main="pacf série résiduelle ",col="blue")
# tracer le corrélogramme partiel de la série résiduelle
Box.test(res)
# le test de bruit blanc
jarque.bera.test(res)
#le test de normalité des résidus
hist(res,xlab="série résiduelle",ylab="fréquence",main="histogramme")
#l'histogramme des résidus
skewness(res)
kurtosis(res)
par(mfcol=c(1,2))
EPS=res2
# calculer le carré des résidus
acf(EPS, main="ACF",col="blue")
# tracer les autocorrélations de la série EPS
pacf(EPS, main="PACF",col="blue")
# tracer les autocorrélations partielle de la série EPS
arch.test(mod)
# appliquer le test ARCH
garch11=garchFit(formula = garch(1, 1)
data=res, include.mean=FALSE)
# L'estimation du modèle GARCH(1,1)
summary(garch11)
# afficher les caractéristiques statistiques de garch11
plot(garch11)
arch1=garchFit(formula = garch(1, 0), data=res, include.mean=FALSE)
```

```
# L'estimation du modèle ARCH(1)
summary(arch1)
# afficher les caractéristiques statistiques de arch1
plot(arch1)
```

L'évolution historique des prix spot du pétrole brut en dollar/baril.

Années\mois	P	Anées\mois	P	Anées\mois	P
2000" 1"	24.58	2002" 6"	23.8	2004" 11"	38.96
2000" 2"	26.84	2002" 7"	25.13	2004" 12"	35.7
2000" 3"	26.71	2002" 8"	25.99	2005" 1"	40.24
2000" 4"	22.93	2002" 9"	27.38	2005" 2"	41.68
2000" 5"	26.94	2002" 10"	27.32	2005" 3"	49.07
2000" 6"	29.12	2002" 11"	24.29	2005" 4"	49.63
2000" 7"	24.94	2002" 12"	28.39	2005" 5"	46.96
2000" 8"	28.3	2003" 1"	30.34	2005" 6"	52.04
2000" 9"	31.48	2003" 2"	31.54	2005" 7"	53.13
2000" 10"	30.42	2003" 3"	29.78	2005" 8"	57.82
2000" 11"	31.22	2003" 4"	25.34	2005" 9"	57.88
2000" 12"	24.13	2003" 5"	25.6	2005" 10"	54.63
2001" 1"	24.06	2003" 6"	26.74	2005" 11"	51.29
2001" 2"	25.41	2003" 7"	27.43	2005" 12"	52.65
2001" 3"	23.7	2003" 8"	28.63	2006" 1"	52.48
2001" 4"	24.38	2003" 9"	26.32	2006" 2"	56.62
2001" 5"	26.25	2003" 10"	28.54	2006" 3"	57.87
2001" 6"	26.1	2003" 11"	28.45	2006" 4"	64.44
2001" 7"	23.73	2003" 12"	29.44	2006" 5"	65.11
2001" 8"	24.46	2004" 1"	30.33	2006" 6"	64.6
2001" 9"	24.29	2004" 2"	29.56	2006" 7"	68.89
2001" 10"	19.64	2004" 3"	32.05	2006" 8"	68.81
2001" 11"	17.65	2004" 4"	32.35	2006" 9"	59.34
2001" 12"	17.53	2004" 5"	36.27	2006" 10"	54.97
2002" 1"	18.33	2004" 6"	34.61	2006" 11"	55.42
2002" 2"	18.89	2004" 7"	36.29	2006" 12"	57.95
2002" 3"	22.64	2004" 8"	40.27	2007" 1"	50.79
2002" 4"	24.88	2004" 9"	40.36	2007" 2"	54.56
2002" 5"	24.76	2004" 10"	45.37	2007" 3"	58.59

2007" 4"	63.55	2009" 6"	69.56	2011" 8"	111.4	2013" 10"	106.75
2007" 5"	64.48	2009" 7"	68.59	2011" 9"	101.57	2013" 11"	107.07
2007" 6"	66.89	2009" 8"	70.37	2011" 10"	107.12	2013" 12"	107.67
2007" 7"	71.89	2009" 9"	65.55	2011" 11"	110.38	2014" 1"	97.49
2007" 8"	68.70	2009" 10"	75.56	2011" 12"	106.84	2014" 2"	102.59
2007" 9"	74.18	2009" 11"	76.20	2012" 1"	111.21	2014" 3"	101.58
2007" 10"	79.31	2009" 12"	77.16	2012" 2"	120.79	2014" 4"	99.74
2007" 11"	88.84	2010" 1"	71.01	2012" 3"	120.89	2014" 5"	102.71
2007" 12"	87.05	2010" 2"	74.6	2012" 4"	117.2	2014" 6"	105.37
2008" 1"	88.35	2010" 3"	78.7	2012" 5"	101.06	2014" 7"	98.17
2008" 2"	90.64	2010" 4"	84.13	2012" 6"	92.99	2014" 8"	95.95
2008" 3"	99.03	2010" 5"	71.88	2012" 7"	102.22	2014" 9"	91.16
2008" 4"	105.16	2010" 6"	72.49	2012" 8"	111.17	2014" 10"	80.54
2008" 5"	119.39	2010" 7"	74.43	2012" 9"	109.68	2014" 11"	66.16
2008" 6"	128.33	2010" 8"	72.39	2012" 10"	106.16	2014" 12"	53.27
2008" 7"	131.22	2010" 9"	77.48	2012" 11"	108.59	2015" 1"	48.24
2008" 8"	112.41	2010" 10"	79.42	2012" 12"	107.76	2015" 2"	49.76
2008" 9"	96.85	2010" 11"	83.65	2013" 1"	112.3	2015" 3"	47.6
2008" 10"	69.16	2010" 12"	88.99	2013" 2"	108.62	2015" 4"	59.63
2008" 11"	49.76	2011" 1"	95.53	2013" 3"	107.23	2015" 5"	60.3
2008" 12"	38.6	2011" 2"	108.5	2013" 4"	100.65	2015" 6"	59.47
2009" 1"	43.3	2011" 3"	111.42	2013" 5"	98.78	2015" 7"	47.12
2009" 2"	43.22	2011" 4"	120.35	2013" 6"	100.78	2015" 8"	49.2
2009" 3"	46.65	2011" 5"	111.2	2013" 7"	104.97	2015" 9"	45.09
2009" 4"	50.36	2011" 6"	107.5	2013" 8"	111.55	2015" 10"	46.59
2009" 5"	63.71	2011" 7"	112.18	2013" 9"	106.47		