

Caractéristiques des rendements de l'action Colgate-Palmolive

ROBIN Alexandre

A rendre le : 09/11/2022

Contents

Introduction	2
Description des données	2
Test sur la moyenne des rendements de <i>rte</i>	3
<u>Propriété 1 : Asymétrie Perte / Gain</u>	4
<u>Propriété 2 : Epaisseur des queues de distribution</u>	5
<u>Propriété 3 : Autocorrélation des rendements</u>	5
Création d'un modèle ARMA(p,q)	6
1. Détermination de p et q de l'ARMA(p,q) à l'aide de l'eacf	6
2. Estimation du modèle trouvé via l'eacf avec la fonction Arima() du package forecast . .	7
3. Espérance et autocorrélation des aléas du modèle MA(4)	7
Tentative d'AR(25) ou MA(25)	8
<u>Propriété 4 : Clusters de volatilité</u>	8
<u>Propriété 5 : Queues épaisses conditionnelles</u>	8
<u>Propriété 6 : Effet de levier</u>	10
<u>Propriété 7 : Saisonnalité</u>	11
7.1 Effet Week-end	11
7.2 Effet Janvier	12
<u>Propriété 8 : Stationnarité</u>	13
Tests de racine unitaire	13
1. Test de Dickey Fuller (DF)	13
2. Test de Dickey Fuller Augmenté (ADF)	15
3. Test de Zivot Andrews (ZA)	18
4. Test de Lee Strazicich (LS)	20

Introduction

Ce rapport s'inscrit dans l'objectif d'estimer une mesure de risque : la **Value-at-Risk** du titre Colgate Palmolive.

Le groupe Colgate Palmolive est un groupe qui trouve son origine dans la première entreprise de savon de William Colgate, créée il y a plus de 200 à New York, Etats-Unis. Aujourd'hui le groupe rassemble plusieurs marques de soins bucaux, d'hygiène personnelle, de produits d'entretien ménagers et de nourriture pour animaux de compagnie. Avec certaines des marques du groupe, Colgate Palmolive détient de nombreuses parts de marché. On compte parmi ces marques : Colgate, Palmolive, Sanex, Ajax, Elmex, et bien d'autres.

Site web : groupe Colgate Palmolive

Il nous faut d'abord explorer les caractéristiques de notre série temporelle des prix et rendements de ce titre sur la période Janvier 2009 - Novembre 2022 (soit presque 13 ans).

Nous utiliserons dans ce projet un niveau de confiance à 95% pour les tests statistiques. L'exploration de ces caractéristiques correspond aux 8 principales propriétés des rendements logarithmiques distinguées par **Charpentier** (2002) :

1. Asymétrie perte/gain
2. Queues de distribution épaisses
3. Autocorrélation des rendements
4. Clusters de volatilité
5. Queues épaisses conditionnelles
6. Effet de levier
7. Saisonnalité
8. Stationnarité

On procèdera à du **backtesting** pour vérifier que notre estimation est correcte. Il nous faut alors pour cela diviser notre échantillon complet de 3481 observations en 2 échantillons:

- Echantillon d'apprentissage, noté *rte* couvrant la période 2009 - 2017 (inclus);
- Echantillon de test, noté *rtt* couvrant le reste des données.

On appliquera toutes les propriétés aux deux échantillons. Les résultats obtenus pour *rte* seront dans le corps de ce rapport. Ceux pour *rtt* se trouveront à la fin du rapport dans un tableau récapitulatif.

Description des données

On commence par se poser les 3 questions suivantes en nous appuyant sur les 3 chronogrammes suivants :

1. Observe-t-on une tendance dans la série ? Si oui, est-elle croissante ou décroissante ? Sinon, autour de quelle valeur la série varie-t-elle ?
2. Est-ce que les fluctuations autour de la tendance (ou de la constante) varient dans le temps ?
3. Observe-t-on des clusters de volatilité dans les données? C'est-à-dire des regroupements de valeurs fortement positives et/ou fortement négatives pour des dates qui se succèdent.

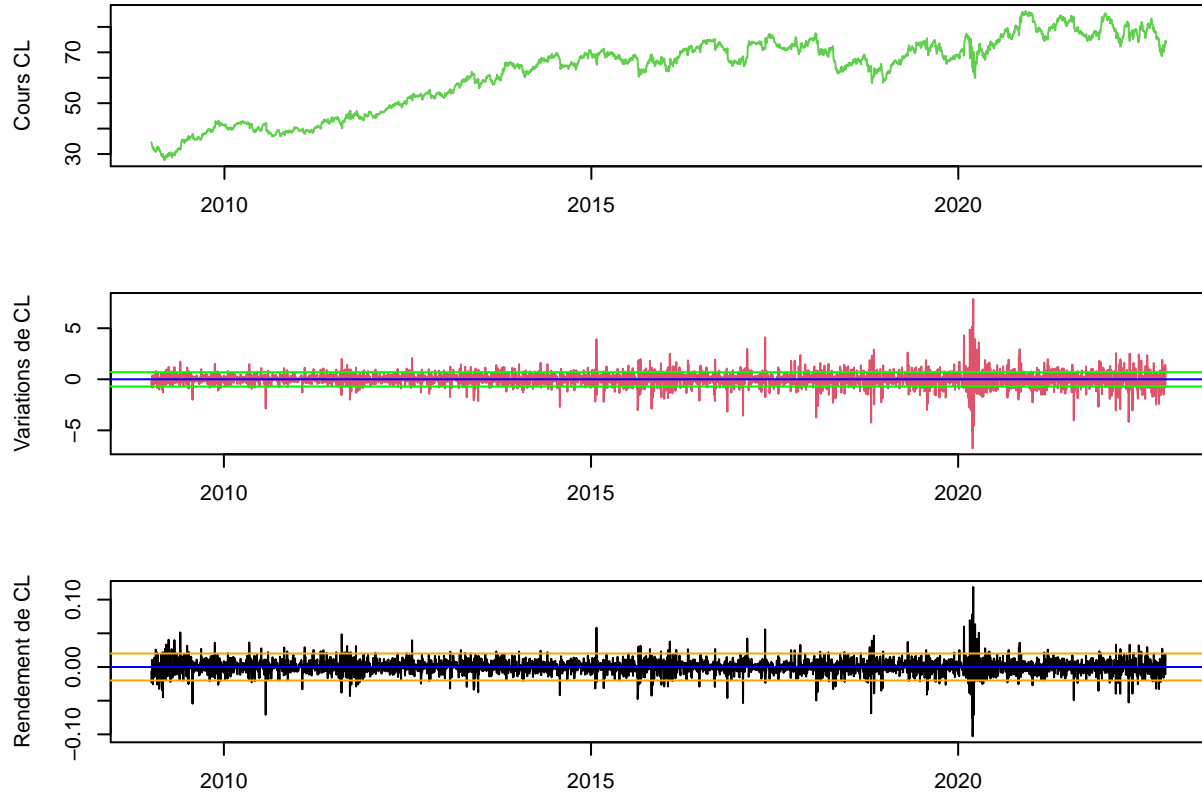


Figure 1: Action Colgate-Palmolive : ses valeurs, ses variations et son rendement logarithmique

1. Tendance (chronogramme 1, vert):

- On observe une tendance croissante sur la courbe verte des prix de l'action Colgate Palmolive, les prix ont augmenté avec les années.

2. Variations autour de la tendance (chronogramme 2, rouge) :

- Les écarts sont de plus en plus grand, la variance augmente donc au fil du temps.

3. Clusters (chronogramme 3, noir) :

- Nous observons un cluster de volatilité (hausse brève et soudaine dans la volatilité des rendements) début 2020, correspondant à la période du confinement mondial.

Notons également que l'on observe à priori pas de saisonnalité dans cette série. Nous pouvons aussi dire que l'action Colgate-Palmolive n'a pas trop souffert de la crise de la Covid 19. Excepté début 2020 où le prix a chuté, le cours est assez rapidement revenu à son niveau précédent et a continué d'augmenter légèrement pour arriver dans un intervalle de valeurs qui se répète maintenant depuis mi-2020.

Test sur la moyenne des rendements de rte

La moyenne empirique des rendements de 2009 à 2017 (inclus) vaut 4.352312×10^{-4} . Le test de Student nous indique une pvalue < 0.05 , on rejette donc l'hypothèse nulle du test. La moyenne des rendements rte est **significativement différente de 0** et vaut environ 0.004896.

Propriété 1 : Asymétrie Perte / Gain

Ici nous testons la nullité de la skewness de notre série *rte* pour savoir comment notre série est distribuée. Si la skewness est statistiquement nulle, la distribution est symétrique. On teste :

$$H_0 : E \left[\left(\frac{X - E(X)}{\sigma_X} \right)^3 \right] = 0 \quad (1)$$

versus

$$H_a : E \left[\left(\frac{X - E(X)}{\sigma_X} \right)^3 \right] \neq 0 \quad (2)$$

Distribution de *rte*

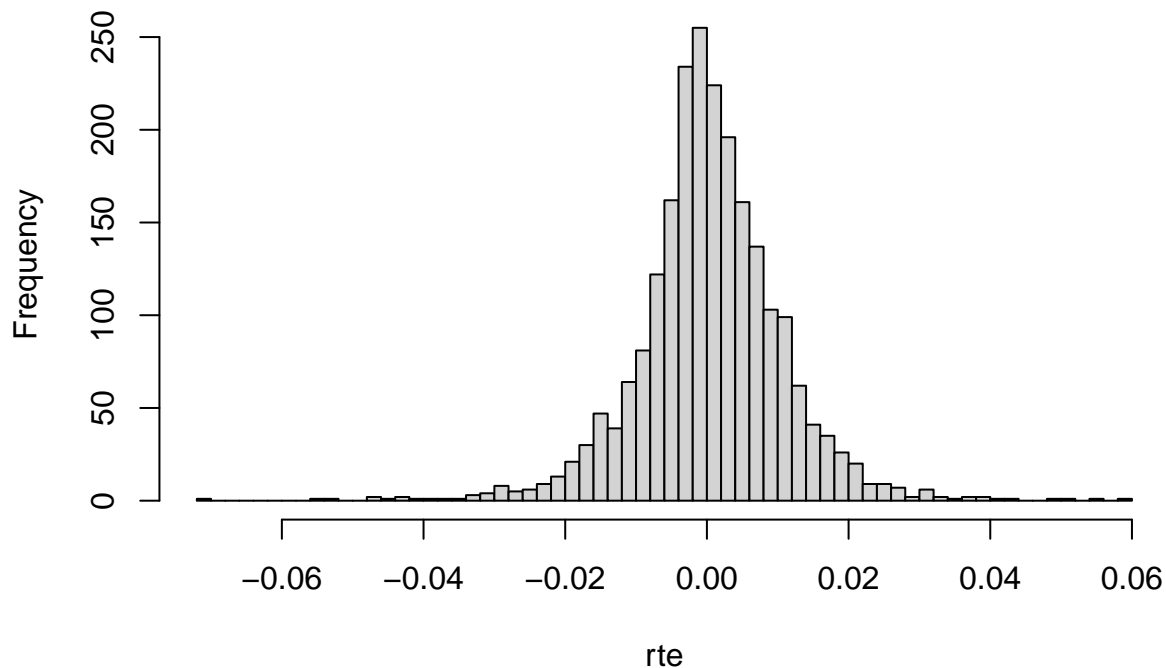


Figure 2: Histogramme de *rte*

```
##  
## D'Agostino skewness test  
##  
## data:  rte  
## skew = -0.1798, z = -3.4774, p-value = 0.0005063  
## alternative hypothesis: data have a skewness
```

Nous procédons au **Test D'Agostino** qui nous donne une p-value < 0.05 (0.0005). Nous pouvons rejeter H_0 , le coefficient de skewness est donc **significativement différent** de 0, la skewness n'est pas statistiquement nulle. Et puisqu'elle est significative et négative (-0.1798), la probabilité de gains est supérieure à la probabilité de pertes quand on détient un titre Colgate Palmolive. Précisons que les gains (fréquents) seront faibles et que les pertes (rares) seront fortes.

Propriété 2 : Epaisseur des queues de distribution

On s'intéresse maintenant à l'épaisseur des queues de la distribution de nos rendements logarithmiques. Pour cela on va tester si le kurtosis vaut 3. On teste alors :

$$H_0 : E \left[\left(\frac{X - E(X)}{\sigma_X} \right)^4 \right] = 3 \quad (3)$$

versus

$$H_a : E \left[\left(\frac{X - E(X)}{\sigma_X} \right)^4 \right] \neq 3 \quad (4)$$

Si on rejette H_0 et que la kurtosis > 3 alors la distribution est leptokurtique, si elle est < 3 alors la distribution est platikurtique.

```
##
##  Anscombe-Glynn kurtosis test
##
## data:  rte
## kurt = 7.1728, z = 14.8747, p-value < 2.2e-16
## alternative hypothesis: kurtosis is not equal to 3
```

La pvalue du test d'Anscombe modifié par Agostino et Zar est < 0.05 . On rejette H_0 , la distribution de la série est leptokurtique (kurtosis > 3). Les queues de distribution des rendements logarithmiques sont plus épaisses que celles d'une loi normale.

Propriété 3 : Autocorrélation des rendements

Ici on veut voir s'il y a de l'autocorrélation dans les rendements logarithmiques *rte*.

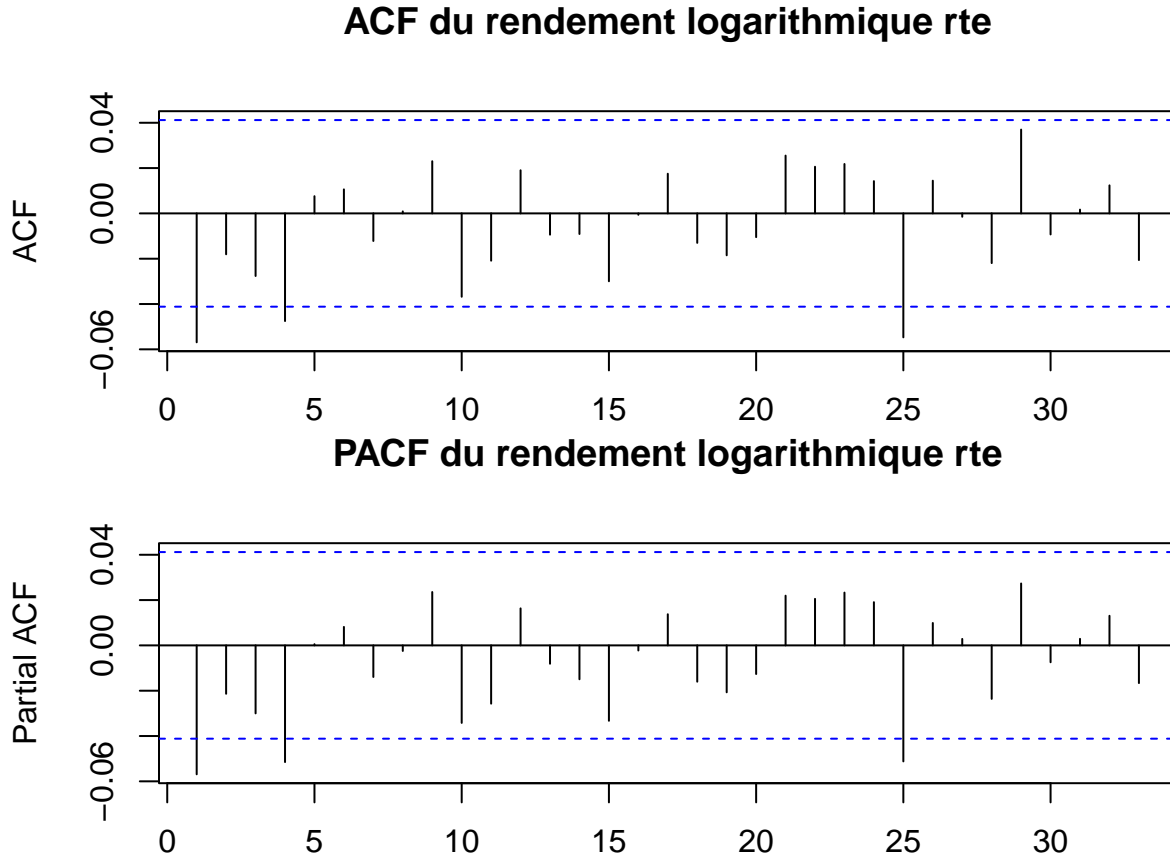


Figure 3: Corrélogrammes *rte*

De l'autocorrélation apparaît aux *lags* 1, 4, et 25. Nous pouvons affirmer qu'il y a de l'autocorrélation à l'aide de la **statistique de Ljung-Box**. Cette statistique teste $H_0 : \rho(k) = 0$ pour $k = 1$ jusqu'à K versus $H_a : \rho(k) \neq 0$ pour au moins une valeur de k comprise entre 1 et K .

La pvalue obtenue vaut $0.007 < 0.05$, on rejette donc l'hypothèse nulle, on peut conclure la présence d'autocorrélation (cela ne nous permet pas de savoir à quel ordre).

Nos rendements logarithmiques *rte* sont donc autocorrélés et nous allons employer un modèle $\text{ARMA}(p,q)$ pour modéliser cette caractéristique. Sa mise en œuvre contient 3 étapes.

Création d'un modèle $\text{ARMA}(p,q)$

1. Détermination de p et q de l' $\text{ARMA}(p,q)$ à l'aide de l'eacf

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o x o o o o o o o o o
## 1 x o o x o o o o o o o o o
## 2 x o o x o o o o o o o o o
## 3 x x x o o o o o o o o o
## 4 o x x o o o o o o o o o
## 5 o o x o x o o o o o o o
## 6 x x x x o x o o o o o o
## 7 x x x o o x x o o o o o
```

On trouve $p = 0$ et $q = 4$, on a donc un $MA(4)$.

On observe également de l'autocorrélation au *lag* 25, nous essaierons aussi de modéliser un $MA(25)$ et un $AR(25)$.

2. Estimation du modèle trouvé via l'eacf avec la fonction `Arima()` du package `forecast`

Rappelons que la pvalue du Test de Student est < 0.05 et cela nous indique donc que la moyenne de *rte* est significativement différente de 0. Nous n'utiliserons pas `include.mean=FALSE`.

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value Pr(>|z|)
## ma1      -0.05999325  0.02101258 -2.8551 0.004302 **
## ma2      -0.02011222  0.02108145 -0.9540 0.340071
## ma3      -0.03169677  0.02096417 -1.5119 0.130547
## ma4      -0.04764241  0.02101165 -2.2674 0.023364 *
## intercept  0.00043626  0.00018454  2.3640 0.018077 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ici des coefficients ne sont pas significatifs (**ma2** et **ma3**), nous allons donc les enlever un à un du modèle afin que tous les coefficients soient significativement différents de zéro (leur pvalue associée est < 0.05).

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value Pr(>|z|)
## ma1      -0.06055809  0.02150463 -2.8160 0.004862 **
## ma4      -0.04848372  0.02087831 -2.3222 0.020222 *
## intercept  0.00043566  0.00019559  2.2274 0.025923 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Après avoir retiré individuellement **ma2** puis **ma3** des coefficients reste égaux à 0 selon le test. On retire donc **ma2** et **ma3**. On va alors poursuivre avec ce modèle dont tous les coefficients sont significativement différents de 0. Notons que l'on remarque que la valeur de la moyenne : 0.004357 est très proche de la moyenne calculée de *rte* : 0.0004352.

Ce modèle $MA(4)$ parvient à modéliser l'autocorrélation détectée dans *rte* à condition que les coefficients soient significatifs (nous venons de le vérifier) et que les aléas du modèle $MA(4)$ aient une espérance nulle et ne soient pas autocorrélés.

3. Espérance et autocorrélation des aléas du modèle $MA(4)$

(a) **Test d'espérance nulle des aléas du modèle $MA(4)$** On teste $H_0 : E(\epsilon) = 0$ versus $H_a : E(\epsilon) \neq 0$, avec ϵ les résidus de la régression $MA(4)$. On utilise le test de Student et on obtient une pvalue > 0.05 , ce qui ne nous confère pas le droit de rejeter l'hypothèse nulle. L'espérance des aléas du modèle $MA(4)$ est bien **nulle** selon ce test.

(b) Test d'absence d'autocorrélation dans les aléas du MA(4)

```
## [1] 0.9626429 0.6169290 0.3391022 0.4990372 0.6155198 0.7196813 0.7633448
## [8] 0.8431708 0.8210263 0.5988828 0.5618198 0.5870426 0.6568893 0.6947358
## [15] 0.5729263 0.6445500 0.6608286 0.6987999 0.6962790 0.7393150
```

Aucune pvalue n'est inférieure à 5%, les aléas du MA(4) ne sont pas autocorrélés.

Notre modèle de régression MA(4) conservé est modélise bien l'autocorrélation détectée dans *rte*.

Tentative d'AR(25) ou MA(25)

(a) AR(25) Il y a présence d'autocorrélation dans les aléas du AR(25), nous ne pouvons pas conserver ce modèle de régression.

MA(25) Il y a également présence d'autocorrélation dans les aléas du MA(25), nous ne pouvons pas conserver ce modèle de régression.

Propriété 4 : Clusters de volatilité

Dans cette section on souhaite savoir s'il y a des effets ARCH (c'est-à-dire des effets de clusters). Il y a des effets ARCH dans une situation d'homoscédasticité conditionnelle. Situation qui postule que la variance de l'aléa au temps t ne dépend pas de l'importance des aléas au carré des périodes précédentes. On réalise le test ARCH d'Engle (1982). Avec ϵ_t le t ième aléa d'un modèle ARMA(p, q) représentant l'équation de la moyenne conditionnelle de *rte* et ϵ les résidus associés à son estimation. Alors le modèle ARCH(m) représentant l'équation de la volatilité de *rte* s'écrit

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_m \epsilon_{t-m}^2 \quad (5)$$

avec σ_t^2 la variance de l'aléa au temps t .

On teste $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$ donc homoscédasticité conditionnelle versus H_a : au moins 1 α_i est différent de 0 avec $i \neq 0$ donc hétéroscédasticité conditionnelle. On utilise la fonction **ArchTest()** pour la variable *rte* à différents *lags* en commençant par le premier.

```
##
## ARCH LM-test; Null hypothesis: no ARCH effects
##
## data: as.numeric(rte)
## Chi-squared = 34.727, df = 1, p-value = 3.794e-09
```

On peut affirmer qu'il y a des effets ARCH dès le premier ordre. En effet, la pvalue est < 0.05 , on peut alors rejeter l'hypothèse nulle : pas d'effet ARCH, avec confiance.

Propriété 5 : Queues épaisses conditionnelles

Les propriétés précédentes nous ont permis de :

- Estimer la moyenne de *rte* avec un MA(4)

- Vérifier que les aléas du MA(4) ont une moyenne nulle et ne sont pas autocorrélés
- Détecter des clusters de volatilité sur *rte*

Nous allons ici estimer un modèle GARCH(1,1) sur les résidus du MA(4) :

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (6)$$

```
##
## Call:
## garch(x = residuv, order = c(1, 1))
##
## Model:
## GARCH(1,1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.50133 -0.52239 -0.01701  0.54104  6.60107
##
## Coefficient(s):
##      Estimate Std. Error  t value Pr(>|t|)
## a0   0.13913    0.01820   7.645 2.09e-14 ***
## a1   0.12025    0.01275   9.431 < 2e-16 ***
## b1   0.74617    0.02546  29.306 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##  Jarque Bera Test
##
## data:  Residuals
## X-squared = 2174.2, df = 2, p-value < 2.2e-16
##
##
##  Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.018092, df = 1, p-value = 0.893
```

D'après les pvalues, les 3 coefficients **a0**, **a1**, **b1** correspondant respectivement à α_0 , α_1 et β_1 sont significativement différents de zéro (pvalues < 0.05).

Remarques :

- Le test de Jarque Bera nous indique que l'on peut rejeter l'hypothèse nulle : les aléas du GARCH(1,1) sont distribués selon une loi normale. Ils ne sont alors pas distribués selon une loi normale.
- Le test de Ljung-Box nous indique quant à lui que l'on ne peut pas rejeter l'hypothèse nulle du test : absence d'autocorrélation dans les aléas du GARCH(1,1).

Nous nous posons dorénavant la question de savoir si le modèle GARCH(1,1) a réussi à prendre en compte toute l'hétéroscédasticité conditionnelle présente dans nos données ?

```
##          Pvalue
## 1  0.8930563
## 2  0.8052250
## 3  0.7524863
## 4  0.8527662
## 5  0.8417981
## 6  0.8786615
## 7  0.8128716
## 8  0.8837663
## 9  0.9281277
## 10 0.9573455
## 11 0.9650947
## 12 0.9784746
## 13 0.9870802
## 14 0.9924458
## 15 0.9109316
## 16 0.9324284
## 17 0.9377089
## 18 0.9517930
## 19 0.9556730
## 20 0.8122524
```

Toutes les pvalues jusqu'à un lag de 20 sont très proche de 1 et nous permettent de ne pas rejeter H_0 avec confiance. Avec notre MA(4) couplé à un GARCH(1,1) nous avons modélisé l'autocorrélation ainsi que l'hétéroscédasticité conditionnelle présentes dans le rendement logarithmique.

Nous souhaitons maintenant savoir si les queues de distribution des aléas de notre ARMA-GARCH sont plus épaisses que celles d'une loi normale. Pour cela, nous réitérons le test d'Anscombe :

```
##
##  Anscombe-Glynn kurtosis test
##
## data:  volat$res
## kurt = 7.7721, z = 15.7042, p-value < 2.2e-16
## alternative hypothesis: kurtosis is not equal to 3
```

La pvalue du test est inférieure à 0.05, nous rejettons H_0 . Alors on a encore des queues de distribution épaisses conditionnelles plus épaisses que celles d'une loi normale.

Propriété 6 : Effet de levier

Dans cette propriété nous voulons voir si la plus grande baisse dans la valeur a induit davantage de volatilité dans le prix du titre que la plus grande hausse dans le prix de l'action Colgate Palmolive. Si cela s'avère être le cas, on peut dire qu'il y a un effet de levier. Nous allons nous appuyer sur la figure suivante :

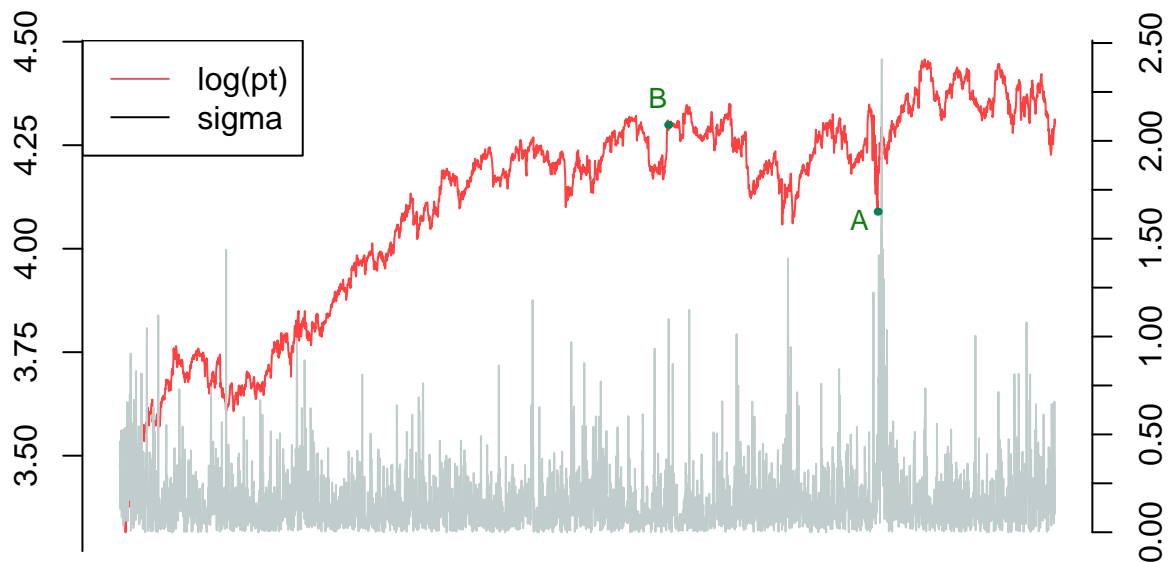


Figure 4: Logarithme de l'action Colgate Palmolive journalier et écart-type récursif journalier des rendements de Colgate Palmolive

On observe donc la plus forte baisse pour $\log(pt)$ au point A, la volatilité associée est la grande barre grise qui double toutes les autres barres de σ . Il est complexe d'identifier clairement la plus grande hausse pour $\log(pt)$ sur la période, mais on peut dire qu'elle se situe au point B. Si ce n'est pas le bon point, cela n'a pas trop d'incidence sur l'interprétation puisque la volatilité associée à la plus forte baisse (point A) est plus grande que pour toutes les hausses qu'a pu vivre $\log(pt)$ sur la période. Il y a donc un effet de levier.

Propriété 7 : Saisonnalité

7.1 Effet Week-end

Nous dirons qu'il y a un effet week-end si la variance des rendements :

- Augmente à partir de mercredi (French et Roll (1986), Baillie et Bollerslev (1989));
- Est plus forte le lundi (French et Roll (1986)).

##	lundi	mardi	mercredi	jeudi
## moyenne en %	0.02190221	0.087450096	0.02693912	0.02486348
## écart-type annuel en %	15.63565155	14.795155948	16.26429837	18.64852009
## skewness	0.13329908	-0.004480288	0.14891565	-0.59296709

```
## kurtosis          3.75826017  0.727319818  3.34097270  6.54410851
##                  vendredi
## moyenne en %      0.05462587
## écart-type annuel en % 17.11395001
## skewness          -0.23635306
## kurtosis          2.91170422
```

Ce résultat correspond à des statistiques descriptives. Nous pouvons conclure qu'il n'y a a priori pas d'effet week-end puisque la variance n'est pas la plus forte lundi et n'augmente pas de mercredi à vendredi.

7.2 Effet Janvier

En général les mois les plus positifs en terme de performances sont les mois d'Avril, suivi de Janvier et Décembre. Nous souhaitons observer si c'est le cas pour notre série. Jettons un oeil au `monthplot` des rendements de Colgate Palmolive.

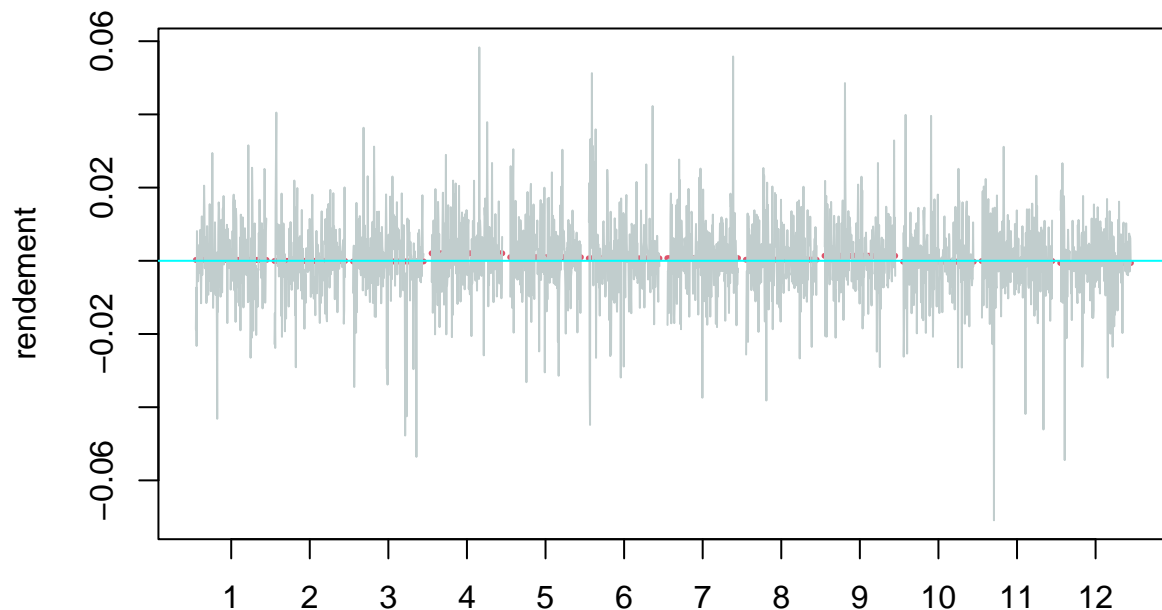


Figure 5 : Rendements logarithmiques de l'action Colgate Palmolive par mois

Nous observons en effet que le mois d'Avril se dégage du reste des mois de l'année pour des performances supérieures et positives. Nous sommes à nouveau dans l'a priori puisque ce sont encore des statistiques descriptives. Notons que les mois avec les plus grands écarts à la moyenne sont les mois de Mars, Juin et Novembre.

Propriété 8 : Stationnarité

Dans cette dernière propriété nous allons étudier la stationnarité de nos processus stochastiques *rte* associés aux rendements. Pour cela nous devons effectuer des tests de racine unitaire afin de définir la nature du PGD (l'équation qui a créé nos données, que l'on ne connaîtra jamais, on essaie de s'en approcher le plus possible) de nos rendements. Nos rendements peuvent alors être issus d'un processus stationnaire, DS, ou TS. Les prix d'une action sont en général issus d'un PGD DS (si un choc dans l'économie modifie les prix, il faudra mettre en place des politiques monétaires ou publiques pour revenir au niveau d'avant le choc) ou d'un PGD TS (après un choc le prix reviendra au niveau d'avant-choc sans effort, le choc est transitoire). Mais les rendements sont en général issus d'un PGD stationnaire. Et c'est ce que nous allons vérifier en procédant aux tests de racines unitaires à l'aide de 4 méthodes :

- Test de Dickey Fuller (DF)
- Test de Dickey Fuller Augmenté (ADF)
- Test de Zivot Andrews (ZA)
- Test de Lee Strazicich (LS)

Tests de racine unitaire

1. Test de Dickey Fuller (DF)

Avec la méthode DF, on teste H_0 : *présence de racine unitaire* versus H_a : *absence de racine unitaire*. On commence par la spécification **trend**, on estime :

$$\Delta X_t = \beta_0 + \beta_1 trend + (\rho - 1)X_{t-1} + \epsilon_t \quad (7)$$

On utilise la pvalue pour tester : $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$.

```
##
## #####
## # 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
## -0.071999 -0.005128 -0.000228  0.005334  0.057441
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.741e-04  4.375e-04   1.541   0.123
## z.lag.1      -1.057e+00  2.098e-02 -50.379 <2e-16 ***
## tt           -1.814e-07  3.344e-07  -0.542   0.588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.0104 on 2261 degrees of freedom
## Multiple R-squared:  0.5289, Adjusted R-squared:  0.5285
## F-statistic: 1269 on 2 and 2261 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -50.3793 846.0267 1269.04
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.96 -3.41 -3.12
## phi2  6.09  4.68  4.03
## phi3  8.27  6.25  5.34
```

La pvalue = 0.588 > 0.05, on ne peut pas rejeter H_0 . Le coefficient β_1 n'est pas significativement différent de zéro.

Note : nous savons que le processus ne peut être TS, seule la spécification **trend** peut mener à cette conclusion.

On utilise désormais la spécification **drift**, on estime :

$$\Delta X_t = \beta_0 + (\rho - 1)X_{t-1} + \epsilon_t \quad (8)$$

On utilise la pvalue pour tester : $H_0 : \beta_0 = 0$ versus $H_a : \beta_0 \neq 0$.

```
##
## #####
## # 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
## -0.071864 -0.005098 -0.000198  0.005355  0.057371
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0004686  0.0002187   2.142  0.0323 *
## z.lag.1      -1.0569030  0.0209768 -50.384 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0104 on 2262 degrees of freedom
## Multiple R-squared:  0.5288, Adjusted R-squared:  0.5286
## F-statistic: 2539 on 1 and 2262 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -50.3843 1269.289
##
## Critical values for test statistics:
```

```
##      1pct  5pct 10pct
## tau2 -3.43 -2.86 -2.57
## phi1  6.43  4.59  3.78
```

La pvalue du coefficient $\beta_0 = 0.03 < 0.05$, cela nous permet de rejeter H_0 et de dire que le coefficient β_0 est significativement différent de zéro.

Nous testons maintenant $H_0 : \rho - 1 = 0$ versus $H_a : \rho - 1 \neq 0$. On regarde la statistique t calculée (**t value**), **t value** = - 50.38 < - 2.86 la valeur critique du test statistique. Nous pouvons donc conclure que selon la méthode DF, la série *rte* est stationnaire. Cependant, la conclusion obtenue avec DF n'est valable que si les ϵ_t ne sont pas autocorrélés. Testons cela avec l'autocorrélogramme partiel (PACF) des résidus des régressions DF.

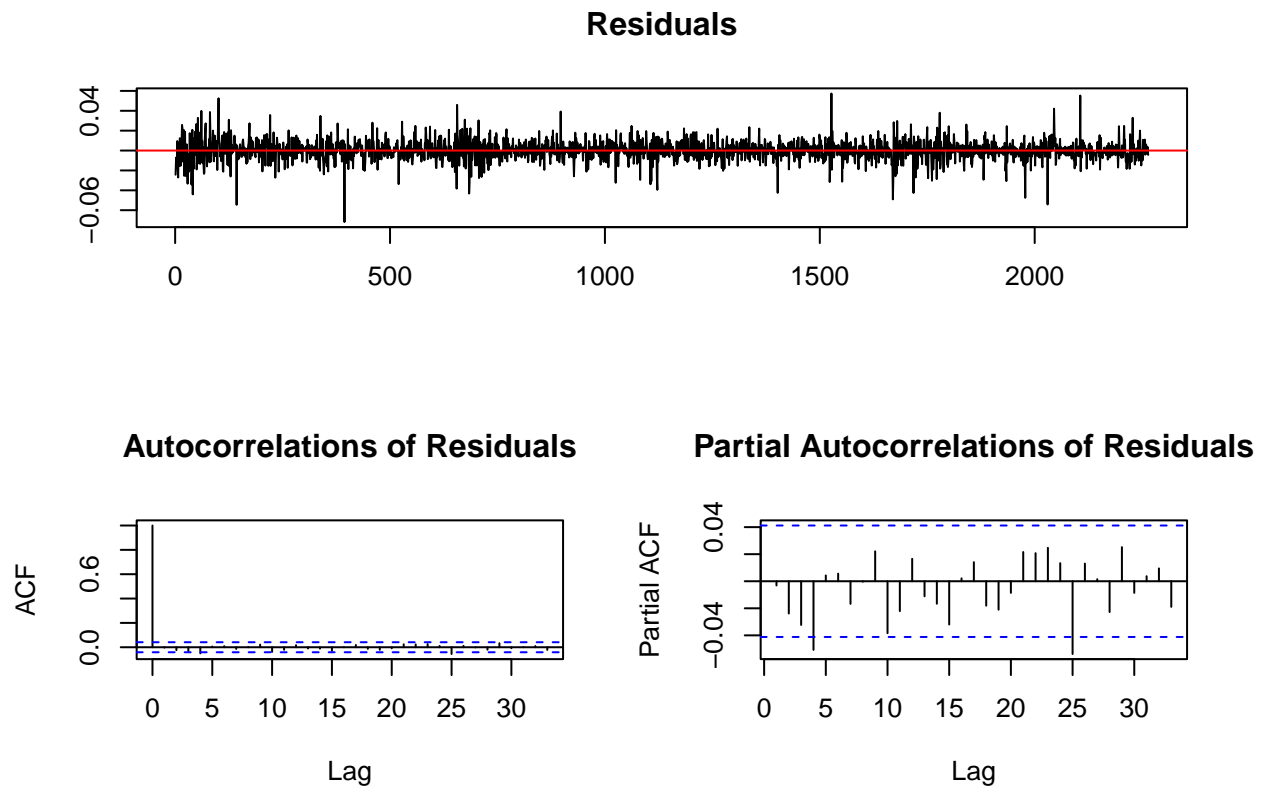


Figure 6: Corrélogramme des résidus de la régression DF 'drift'

Ici on observe de l'autocorrélation à l'ordre 4 et à l'ordre 25. On doit prendre en compte cette autocorrélation en réalisant le test ADF. Cela signifie que les rendements à l'instant t dépendent de la valeur des rendements X_{t-4} et X_{t-25} .

2. Test de Dickey Fuller Augmenté (ADF)

Le test d'hypothèses est le même qu'en méthode DF : on teste H_0 : *présence de racine unitaire* versus H_a : *absence de racine unitaire*. Nous reprenons la spécification employée en DF, c'est-à-dire la spécification **drift**. ADF c'est DF avec des variables explicatives en plus qui sont la variable dépendante retardée jusqu'à

l'ordre P. On estime donc :

$$\Delta X_t = \beta_0 + (\rho - 1)X_{t-1} + \sum_{p=1}^p \gamma_p \Delta X_{t-p} + \epsilon_t \quad (9)$$

P est la valeur maximale de retards à introduire, on se servira de Pmax, donné par la formule de Schwert (1989). P le nombre de retards à introduire, se situe entre 0 et Pmax. Cette valeur de P peut être obtenue en faisant le MAIC.

Le Pmax calculé par méthode de Schwert est 26.

```
## Augmented DF test
##                               ADF test
## t-test statistic:            -50.2327
## p-value:                     0.0001
## Max lag of the diff. dependent variable: 0.0000
##
## Call:
## dynlm(formula = formula(model), start = obs.1, end = obs.T)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.071936 -0.005085 -0.000216  0.005291  0.057300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0005104  0.0002183   2.338  0.0195 *
## L(y, 1)      -1.0603870  0.0211095 -50.233  0.0001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01031 on 2236 degrees of freedom
## Multiple R-squared:  0.5302, Adjusted R-squared:  0.53
## F-statistic:      NA on NA and NA DF, p-value: NA
```

Max lag of the diff. dependent variable: 0, on emploi donc le BIC.

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.071775 -0.005119 -0.000177  0.005371  0.057113
##
## Coefficients:
```



```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0005193  0.0002185   2.376  0.0176 *
## z.lag.1      -1.0792357  0.0307849 -35.057 <2e-16 ***
## z.diff.lag   0.0177702  0.0211243   0.841  0.4003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01032 on 2235 degrees of freedom
## Multiple R-squared:  0.5303, Adjusted R-squared:  0.5299
## F-statistic: 1262 on 2 and 2235 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -35.0573 614.5076
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.43 -2.86 -2.57
## phi1  6.43  4.59  3.78
```

Cette méthode nous donne un $P = 1$. On introduit 1 retard γ_1 mais ce dernier n'est pas significativement différent de zéro. En effet, sa $|t \text{ value}| = 0.841 < 1.6$. Cela revient à dire que la méthode du BIC nous indique finalement $P = 0$. Nous allons donc partir de $P = P_{\max} = 26$ et ôter un à un les γ_p jusqu'à obtenir le dernier γ_p significatif.

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.070350 -0.004990 -0.000036  0.005301  0.055558
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0006034  0.0002258   2.673  0.00758 **
## z.lag.1      -1.2823644  0.1282271 -10.001 < 2e-16 ***
## z.diff.lag1   0.2199697  0.1257085   1.750  0.08028 .
## z.diff.lag2   0.1995320  0.1229765   1.623  0.10483
## z.diff.lag3   0.1575870  0.1200481   1.313  0.18942
## z.diff.lag4   0.1076180  0.1169639   0.920  0.35762
## z.diff.lag5   0.1071517  0.1136209   0.943  0.34575
## z.diff.lag6   0.1048540  0.1102299   0.951  0.34159
## z.diff.lag7   0.0862636  0.1068362   0.807  0.41950
## z.diff.lag8   0.0858649  0.1033841   0.831  0.40632
## z.diff.lag9   0.1070393  0.0996655   1.074  0.28295
## z.diff.lag10  0.0759665  0.0958016   0.793  0.42789
## z.diff.lag11  0.0459363  0.0919926   0.499  0.61758
```

```

## z.diff.lag12  0.0571655  0.0881691  0.648  0.51682
## z.diff.lag13  0.0541589  0.0842446  0.643  0.52037
## z.diff.lag14  0.0352799  0.0800056  0.441  0.65928
## z.diff.lag15  0.0002241  0.0757900  0.003  0.99764
## z.diff.lag16  0.0059604  0.0715726  0.083  0.93364
## z.diff.lag17  0.0186477  0.0669442  0.279  0.78061
## z.diff.lag18 -0.0011557  0.0620797 -0.019  0.98515
## z.diff.lag19 -0.0177901  0.0570091 -0.312  0.75503
## z.diff.lag20 -0.0272103  0.0513248 -0.530  0.59606
## z.diff.lag21 -0.0075335  0.0449570 -0.168  0.86694
## z.diff.lag22  0.0138420  0.0382624  0.362  0.71756
## z.diff.lag23  0.0373834  0.0306691  1.219  0.22300
## z.diff.lag24  0.0519206  0.0210156  2.471  0.01356 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01031 on 2214 degrees of freedom
## Multiple R-squared:  0.5361, Adjusted R-squared:  0.5308
## F-statistic: 102.3 on 25 and 2214 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -10.0007 50.0081
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.43 -2.86 -2.57
## phi1  6.43  4.59  3.78

```

Notre coefficient γ_{24} est significativement différent de zéro : $|t \text{ value}| = 2.471 > 1.6$. Nous avons également β_0 significativement différent de zéro ($pvalue < 0.05$) et $(\rho - 1) \neq 0$ ($-10 < -2.86$). Nous avons donc:

$$\Delta X_t = \beta_0 + (\rho - 1)X_{t-1} + \sum_{p=1}^{24} \gamma_p \Delta X_{t-p} + \epsilon_t \quad (10)$$

Et la série est stationnaire selon la méthode ADF.

Remarque : Nous gardons ces conclusions en tête mais ne pouvons pas nous y fier à 100%. En effet, les méthodes DF et ADF ne prennent pas en compte les changements structurels dans des séries. Or, les prix d'une action sont très sensibles aux crises, aux changements politiques dans certains pays / groupes de pays.

Afin d'intégrer ces aspects nous pouvons utiliser la méthode Zivot Andrews qui introduit des variables muettes pour modéliser ces changements structurels. Nous considérons 2 modèles possibles, le modèle A ("**crash**" = "**intercept**" selon Perron), dans lequel, le niveau de la série est touché par la crise, et le modèle C ("**both**" selon Perron), dans lequel le niveau et la croissance de la série sont touchés par le changement structurel. On note T_B la date du changement structurel, DU_t (respectivement DT_t) la variable à ajouter pour modéliser un changement dans le niveau (respectivement la pente) de la partie déterministe de la série. Ces variables valent 0 avant la crise, 1 après.

3. Test de Zivot Andrews (ZA)

Avec la méthode ZA on effectue le test suivant : H_0 : la série est *DS sans changement structurel* versus H_a : la série est *TS avec un unique changement structurel* (non H_0). On estime :

$$\Delta X_t = \beta_0 + \beta_1 trend + \rho X_{t-1} + \delta_1 DU_t(T_B) + \delta_2 DT_t(T_B) + \sum_{p=1}^p \gamma_p \Delta X_{t-p} + \epsilon_t \quad (11)$$

On commence par identifier si on a un modèle A ou un modèle C. On débute avec un nombre de retards égal à $P_{\max} = 26$, et on regarde en premier la significativité de δ_1 et δ_2 . On teste : $H_0 : \delta_2 = 0$ versus $H_a : \delta_2 \neq 0$. A ce stade, puisque le modèle dans lequel on ne retrouve que DT n'est pas pertinent pour une série économique, on ne fera pas ce même test pour δ_1 . Soit on a DU et DT soit on a seulement DU.

Nous pouvons garder le modèle C (**both**) puisque les coefficients de DU et DT sont significativement différents de zéro (pvalues < 0.05) et nous arrivons à un nombre de retards = 25. En descendant de un en un à partir de γ_{26} , γ_{25} a sa $|t \text{ value}| = 0.204 < 1.6$ et γ_{24} a sa $|t \text{ value}| = 1.774 > 1.6$. Nous avons le modèle suivant :

$$\Delta X_t = \beta_0 + \beta_1 trend + \rho X_{t-1} + \delta_1 DU_t(T_B) + \delta_2 DT_t(T_B) + \sum_{p=1}^{25} \gamma_p \Delta X_{t-p} + \epsilon_t \quad (12)$$

Pour répondre au test d'hypothèses de la méthode ZA, on regarde maintenant la statistique calculée associée à $H_0 : \rho - 1 = 0$, et on la compare à la valeur critique à 5% donnée par ZA. Ici la statistique calculée $= -10.77 < -5.08$ la valeur critique à 5% donc on rejette H_0 . La série n'est pas DS sans changement structurel. Le processus est peut-être TS avec un unique changement structurel. Mais on ne peut pas conclure que c'est TS. En ZA, accepter H_a c'est aussi rejeter H_0 : *la série est DS avec unique changement structurel*, et nous avons observé au début que les rendements n'ont pas de tendance et semblaient être stationnaires. Notons que les coefficients β_0 et β_1 sont significativement différents de zéro.

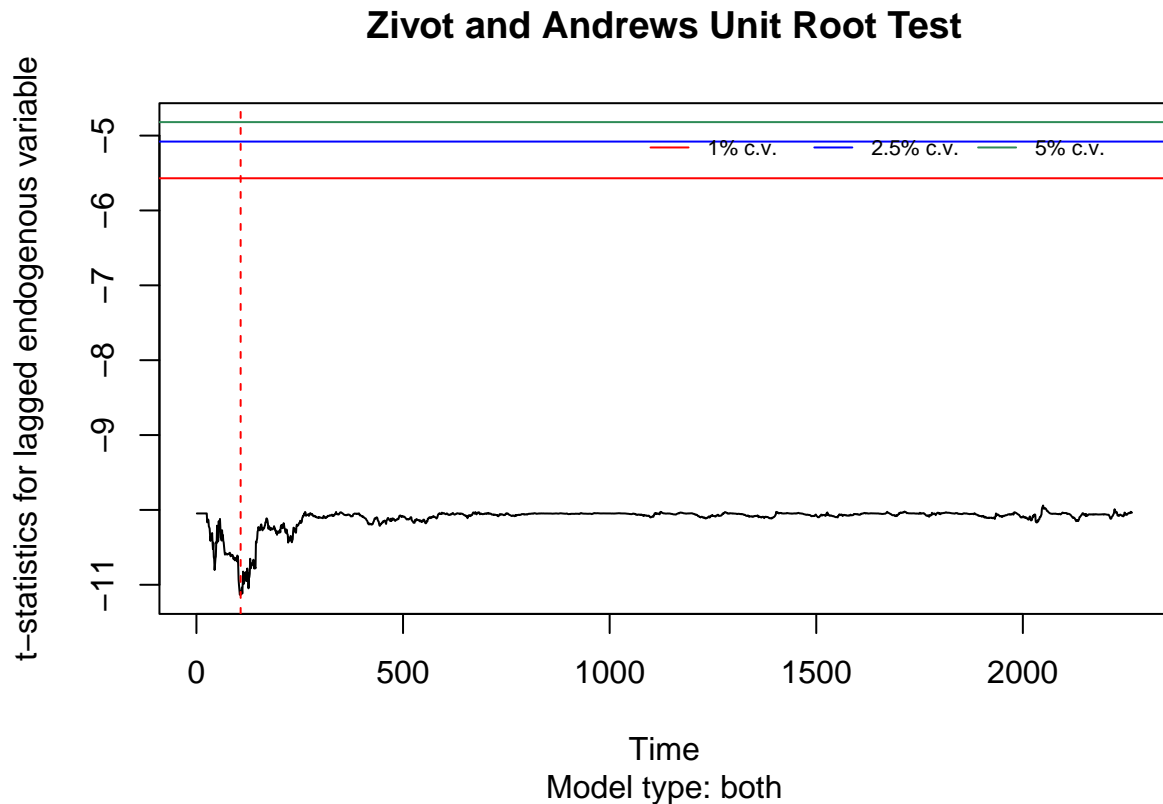


Figure 7: Détermination de la date de rupture en ZA

La méthode ZA donne une date de crise potentielle en position 107. Cela correspond au 5 juin 2009

Une limite de ce test de Zivot Andrews est qu'il ne prend en compte qu'un seul changement structurel dans les données. Pour palier cette limite nous allons utiliser la dernière des 4 méthodes : la méthode Lee Strazicich.

4. Test de Lee Strazicich (LS)

En LS le PGD comprend des changements structurels sous H_0 et sous H_a . On effectue le test suivant : H_0 : la série est DS avec changement structurel versus H_a : la série est TS avec changement structurel. Le test LS permet d'introduire des changements structurels jusqu'à 2 dates possibles.

En LS, on reprend la classification des changements structurels de Perron, **crash** et **both**, que l'on a utilisé dans le test ZA, que Lee et Strazicich appellent respectivement **crash** et **break**. Ici on reprendra alors **break**. On a le modèle suivant :

$$X_t = \delta' Z_t + e_t \quad (13)$$

$$e_t = \beta e_{t-1} + \epsilon_t \quad (14)$$

Avec ϵ Normale($0, \sigma^2$) et Z la matrice des variables exogènes.

Soit T_{B1} la date du premier changement structurel et T_{B2} la date du second. On utilisait **both** en ZA, on va alors utiliser **break** pour LS. On a donc : $Z_T = [\tau, tendance, DU_{1t}, DU_{2t}, DT_{1t}, DT_{2t}]'$, avec $DT_{jt} = t - T_{Bj}$ si $t \geq T_{Bj} + 1$ pour $j = 1, 2$ et 0 sinon.

Ainsi, pour le modèle **break**, avec v_{jt} stationnaire, on a :

$$X_t = \mu_0 + d_1 B_{1t} + d_2 B_{2t} + d_3 D_{1t} + d_4 D_{2t} + X_{t-1} + v_{1t} \text{ sous } H_0 \quad (15)$$

$$X_t = \mu_1 + \gamma trend_t + d_1 D_{1t} + d_2 D_{2t} + d_3 DT_{1t} + d_4 DT_{2t} + v_{2t} \text{ sous } H_a \quad (16)$$

Ici T le nombre d'observations est grand (2265), nous n'utiliserons pas la méthode LS bootstrap (Chou 2007).

Puisque l'on avait δ_1 et δ_2 significatifs dans ZA, on choisit **break**. On utilisera les λ qui nous permettront de localiser le ou les points de rupture.

On fait un premier test de racine unitaire selon la méthode LS avec 1 date de rupture.

```
## [1] -14.15167
## [1] "First possible structural break at position: 1236"
## [1] "The location of the first break - lambda_1: 0.5 , with the number of total observations: 2265"
## Critical values - Crash model:
##      1%      5%     10%
## [1,] -4.239 -3.566 -3.211
## [1] "Number of lags determined by general-to-specific lag selection: 5"
## Runtime:
## Time difference of 0.113835 mins
```

On obtient un $\lambda_1 = 0.5$ et pour savoir si notre série est DS avec changement structurel, on doit observer que la statistique calculée est supérieure à la valeur critique dans la grille de λ , pour $\lambda_1 = 0.5$. On a $-14.15 < -3.56$, alors on rejette H_0 . Et, étant donné que l'on a observé que nos rendements varient autour de zéro, et n'ont donc pas de tendance, on peut s'attendre à ce que le PGD dont notre série de rendements est issue soit stationnaire.

Nous allons maintenant procéder au test de racine unitaire selon la méthode LS avec 2 points de rupture. Notons bien que **la conclusion que nous retiendrons est celle que nous venons d'obtenir avec la méthode LS 1 rupture** puisque LS à 2 points de ruptures nous donne une conclusion moins fiable. Le résultat que nous obtiendrons de la méthode à 2 ruptures pourra ceci-dit conforter le résultat précédemment obtenu s'il est le même.

```
## [1] -14.9323
## [1] "First possible structural break at position: 907"
## [1] "The location of the first break - lambda_1: 0.4 , with the number of total observations: 2265"
```

```

## [1] "Second possible structural break at position: 1386"
## [1] "The location of the second break - lambda_2: 0.6 , with the number of total observations: 2265"
## Critical values:
##          Break 2 - 0.4 - 1% Break 2 - 0.4 - 5% Break 2 - 0.4 - 10%
## Break 1 - 0.2          -6.16          -5.59          -5.27
## Break 1 - 0.4           NA           NA           NA
## Break 1 - 0.6           NA           NA           NA
##          Break 2 - 0.6 - 1% Break 2 - 0.6 - 5% Break 2 - 0.6 - 10%
## Break 1 - 0.2          -6.41          -5.74          -5.32
## Break 1 - 0.4          -6.45          -5.67          -5.31
## Break 1 - 0.6           NA           NA           NA
##          Break 2 - 0.8 - 1% Break 2 - 0.8 - 5% Break 2 - 0.8 - 10%
## Break 1 - 0.2          -6.33          -5.71          -5.33
## Break 1 - 0.4          -6.42          -5.65          -5.32
## Break 1 - 0.6          -6.32          -5.73          -5.32
## [1] "Number of lags determined by general-to-specific lag selection: 5"
## Runtime:
## Time difference of 173.2638 mins

## [1] "2014-07-07"

```

Tableau récapitulatif pour les propriétés étudiées		rte (2009-2017)	rtt (2018-2022)
Description des données	Tendance	oui, croissante	oui, croissante
	Variations autour de la tendance	la variance augmente au fil du temps	la variance est la même au fil du temps
	Clusters	non	oui 1 en 2020
Test de Student sur la moyenne	Moyenne < 0 ou > 0	moyenne différente de zéro	moyenne égale à zéro
Propriété 1 : Asymétrie perte / gain	Coefficient de skewness	skewness différent de zéro, négatif	skewness égal à zéro
	Distribution	concentrée sur la droite, étalée sur la gauche	symétrique
	Gain / Perte	probabilité de gains > probabilité de pertes	probabilité de gains = probabilité de pertes
Propriété 2 : Epaisseur des queues de distribution	Coefficient de kurtosis	différent de 3, supérieur à 3	différent de 3, supérieur à 3
	Distribution	leptokurtique	leptokurtique
	Epaisseur des queues	plus épaisses que pour la loi normale	plus épaisses que pour la loi normale
Propriété 3 : Autocorrélation des rendements	PACF	autocorrélation aux lags 4 et 25	autocorrélation aux lags 1,2,3,4,7,9 et 10
	Ljung-Box	il y a de l'autocorrélation	il y a de l'autocorrélation
	ARMA(p,q)	MA(4)	MA(10)
Propriété 4 : Clusters de volatilité		oui	oui
Propriété 5 : Queues épaisses conditionnelles		tous les coefficients du GARCH(1,1) sont significatifs	tous les coefficients du GARCH(1,1) sont significatifs
		MA(4)+GARCH(1,1) modélisent l'autocorrélation	MA(10)+GARCH(1,1) modélisent l'autocorrélation
	Queues de distribution des aléas du ARMA-GARCH	leptokurtiques, plus épaisses que pour une loi normale	leptokurtiques, plus épaisses que pour une loi normale
Propriété 6 : Effet de levier		oui	oui
Propriété 7 : Saisonnalité	Effet week-end	non	oui, variance forte le lundi
	Effet janvier	oui seulement pour Avril	non
Propriété 8 : Stationnarité	Dickey Fuller	stationnaire	stationnaire
	Dickey Fuller Augmenté	stationnaire	stationnaire
	Zivot Andrews	(Ha) PAS DS avec unique changement structurel	(Ha) PAS DS avec unique changement structurel
	Lee Strazicich (1 et 2 ruptures)	(Ha) PAS DS avec changement structurel	(Ha) PAS DS avec changement structurel