
CENTRALESUPÉLEC

TP DE SÉRIES CHRONOLOGIQUES

Modèle ARMA(p,q) et Méthode des moindres carrés ordinaires

Auteurs :
Lorenzo Gasparollo

24 février 2019

Table des matières

1	Question 1	3
2	Question 2	3
3	Question 3	4
4	Question 4	5
4.1	Test de Fisher	6
5	Question 5	6
5.0.1	Plot densité et test de Shapiro-Wilk	6
5.1	QQ plot	7
5.2	Plot autocorrelation	8
5.3	Hétéroscédasticité	9
6	Question 6	10
7	Question 7	11
8	Question 8	12
9	Question 9	14
10	Question 10	15
11	Question 11 et 11 bis	15
12	Question 12	17
13	Question 13	18

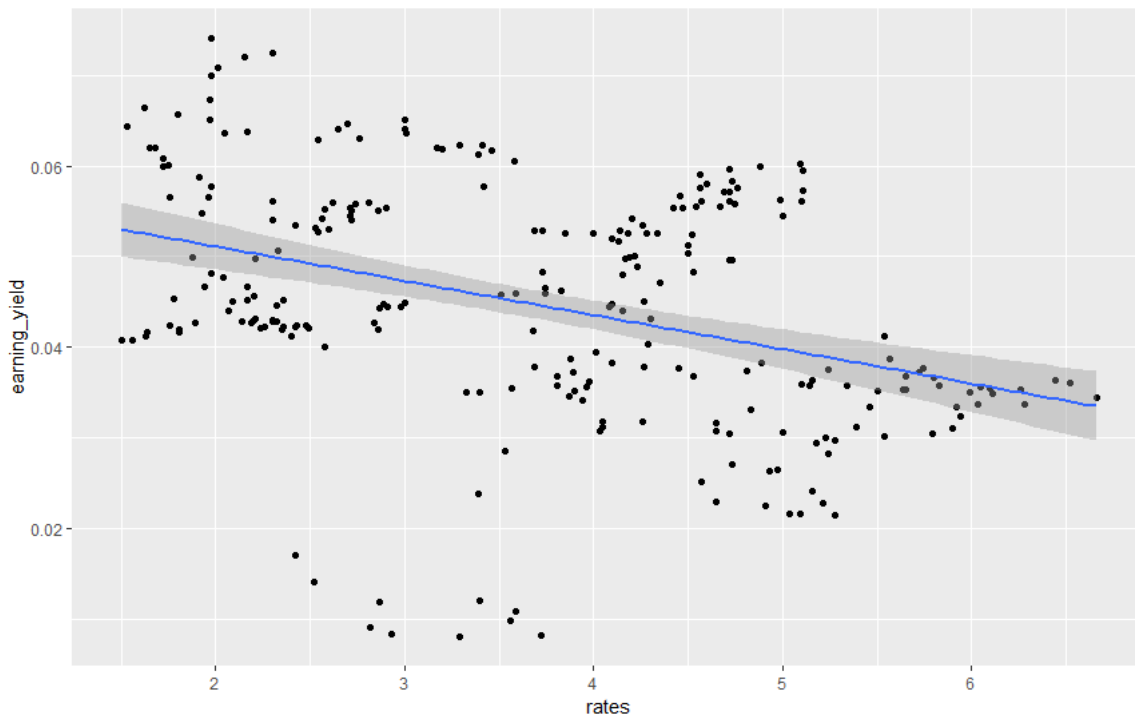
Introduction

Cadre : Le Fed model est un modèle empirique que l'on peut classer dans la catégorie de modèle d'arbitrage. Le Fed model repose sur l'idée qu'il existe une relation d'arbitrage entre la rémunération d'une obligation d'état, jugée sans risque, et le taux de rendement d'une action. Le prolongement de cette logique induit l'existence d'une relation d'équilibre entre les deux variables. Le modèle économétrique est donné par la relation ci-dessous :

$$\frac{E_t}{P_t} = \alpha + \beta.r + \epsilon_t \quad (1)$$

1 Question 1

Voici à ce qu'il ressemble le nuage de points liant le earning yield au taux sans risque.



on voit clairement que la droite n'a pas de pouvoir explicatif sur le jeu de données.

2 Question 2

Considerons un nuage de points $\{y_i, x_i\}_{i=1, \dots, n}$, $y_i \in \mathbb{R}$ et $x_i \in \mathbb{R}^{n \times p-1}$. il s'agit de les ajuster selon une relation linéaire, i.e.,

$$\begin{cases} y = f_{\theta}(x^*) + \epsilon \\ x^* = [1, x]^T \\ f_{\theta}(x^*) = \langle \theta, x^* \rangle \\ \theta \in \mathbb{R}^p \\ \epsilon \perp x \mid \mathbb{E}[\epsilon] = 0, V[\epsilon] = \sigma_{\epsilon}^2 \end{cases}$$

Définissons préliminairement Y comme le vecteur colonne des observations de la variable explicité et $X \in \mathbb{R}^{n \times p}$ la matrice des variables explicative dont la i -ième composante est égale à x_i^* .

La méthode des moindres carrés ordinaires vise à trouver l'estimateur à travers un processus de minimisation de la fonction de perte définie par :

$$\begin{aligned} L(\theta) &= \|Y - X\theta\|_2^2 \\ &= \sum_{i=1}^n (y_i - f_\theta(x_i))^2 \end{aligned} \quad (2)$$

La solution de MCO est donnée par :

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

qui se réduit pour $p = 2$ à :

$$\begin{cases} \theta_1 = \frac{\bar{y} \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \theta_2 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{cases} \quad (3)$$

L'estimateur des moindres carrés présentent les propriétés suivantes :

1. Meilleur estimateur linéaire sans biais.
2. Dans le cas où les observations sont iid et sous l'hypothèse de normalité de perturbation, l'estimateur MCO coïncide avec l'estimateur du maximum de vraisemblance et donc il hérite des propriétés de cet estimateur.

3 Question 3

Dans le notre cadre on a : $x_i = r_i$, $y_i = \left(\frac{E}{P}\right)_i$, $\theta_1 = \alpha$, $\theta_2 = \beta$.

Par conséquent :

$$\begin{aligned} Cov(y, x) &\triangleq \frac{1}{n} \sum_i^n (y_i - \bar{y})(x_i - \bar{x}) \\ &= \frac{1}{n} \left(\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + n \sum_{i=1}^n \bar{x} \bar{y} \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \right) \end{aligned} \quad (4)$$

$\Rightarrow \hat{\beta} = \frac{Cov(\frac{E}{P}, r)}{\hat{\sigma}_r^2}$ où $\hat{\sigma}_r^2$ désigne l'estimateur de la variance de r .

De plus,

$$\begin{aligned} \hat{\alpha} &= \frac{\bar{y} (\sum_{i=1}^n x_i^2 - \bar{x}^2) + \bar{y} \bar{x}^2 - \bar{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \bar{y} - \bar{x} \frac{(\bar{y} \bar{x} - \sum_{i=1}^n x_i y_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \bar{y} - \bar{x} \hat{\beta} \end{aligned}$$

$$\Rightarrow \hat{\alpha} = \frac{\bar{E}}{\bar{P}} - \bar{r}\hat{\beta}$$

Calcul de la variance asymptotique :

$$\begin{aligned}
\lim_{n \rightarrow \infty} V(\hat{\beta}) &= \lim_{n \rightarrow \infty} V \left[\frac{Cov(y, x)}{\hat{\sigma}_x^2} \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n^2} \frac{1}{\hat{\sigma}_x^4} V \left[\sum_{i=1}^n (y_i - \bar{y}) (x_i - \bar{x}) \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n^2} \frac{1}{\hat{\sigma}_x^4} V \left[\sum_{i=1}^n y_i x_i - n \bar{x} \bar{y} \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n^2} \frac{1}{\hat{\sigma}_x^4} \mathbb{E} \left[V \left[\sum_{i=1}^n y_i x_i - n \bar{x} \bar{y} \mid x = x^* \right] \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n^2} \frac{1}{\hat{\sigma}_x^4} \mathbb{E} \left[V \left[\sum_{i=1}^n y_i (x_i - \bar{x}) \mid x = x^* \right] \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n^2} \frac{\sigma_\epsilon^2}{\hat{\sigma}_x^4} \mathbb{E} \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\sigma_\epsilon^2}{\hat{\sigma}_x^2} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\sigma_\epsilon^2}{\sigma_x^2} \\
&= 0
\end{aligned} \tag{5}$$

4 Question 4

Les valeurs calculées sont :

$$\begin{cases} \hat{\alpha} = 0.05869356 \\ \hat{\beta} = -0.003781593 \end{cases}$$

Coefficient de détermination

Le coefficient de détermination représente la proportion de la variance contenue dans la variable explicative qu'est prévisible à part de la variable explicative.

Si le R est nul, cela signifie que l'équation de la droite de régression détermine 0 % de la distribution des points. Si le R vaut 1, cela signifie que l'équation de la droite de régression est capable de déterminer la totalité distribution des points. Dans notre cas : $R = 0.1383392$. Cela signifie que le modèle mathématique utilisé n'explique absolument pas la distribution des points.

Maintenant on va fixer un niveau de significativité $\alpha = 0.05$ et on va étudier deux tests.

Test student

Le test de Student permet de déterminer si une variable joue un rôle explicatif dans un modèle. Pour faire un test de Student, il faut vérifier préliminairement que les erreurs suivent une loi normale, c.à.d., $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon)$ $\forall i$.

Pour $j \in \{1 \dots p\}$ on a le test est le suivant :

$$\begin{cases} H_0 = \beta_j = 0 \\ H_1 = \beta_j \neq 0 \end{cases}$$

On a la règle de décision suivante : Si $\mathbb{P}[|t| > t^*] < \alpha$, alors on rejette H_0 et on accepte H_1 : le coefficient est significativement différent de zéro et la variable joue un rôle explicatif dans le modèle. Par ailleurs, dans notre modèle on trouve pour l'intercept $\mathbb{P}[|t| > t^*] = 2\text{e-}16$, $t^* = 24.935$ et pour β $\mathbb{P}[|t| > t^*] = 1.38\text{e-}09$, $t^* = -6.297$. Pour les deux cas on rejette H_0 .

4.1 Test de Fisher

Le test Fisher est le suivant :

$$\begin{cases} H_0 = \beta_j = 0 \forall j \in \{1 \dots p\} \\ H_1 = \exists j \text{ t.q. } \beta_j \neq 0 \end{cases}$$

Ce test est peu utilisé car lorsqu'il indique qu'il y a au moins un coefficient non nul, il ne précise pas lesquels. Il est moins précis que le test de Student. Dans notre cas $F = 1.375\text{e-}09$. On rejette donc H_0 .

5 Question 5

5.0.1 Plot densité et test de Shapiro-Wilk

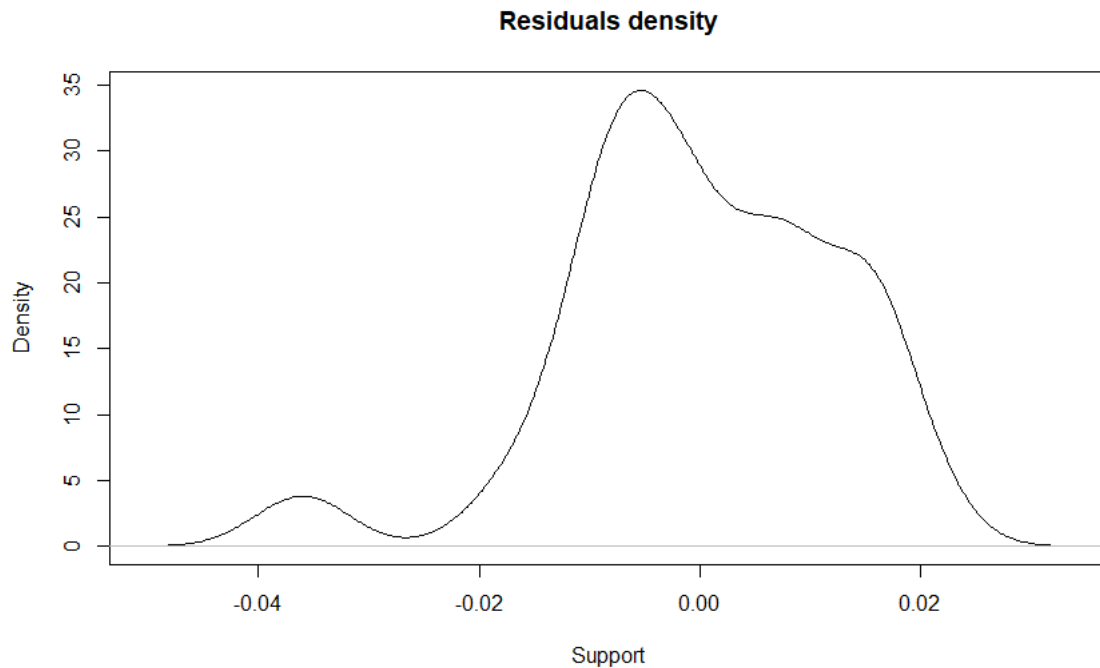


FIGURE 1

On remarque que la densité :

- n'est pas centrée en 0.
- n'est pas symétrique.
- ne présente pas la forme d'une "cloche".

En outre, afin de donner un test objective sur la normalité de la distribution des résidus, on procède à un test de normalité appelé Shapiro test. Le test de shapiro donne une p-value égale à $1.233\text{e-}07$.

On conclut qu'il ne s'agit pas d'une variable normale.

5.1 QQ plot

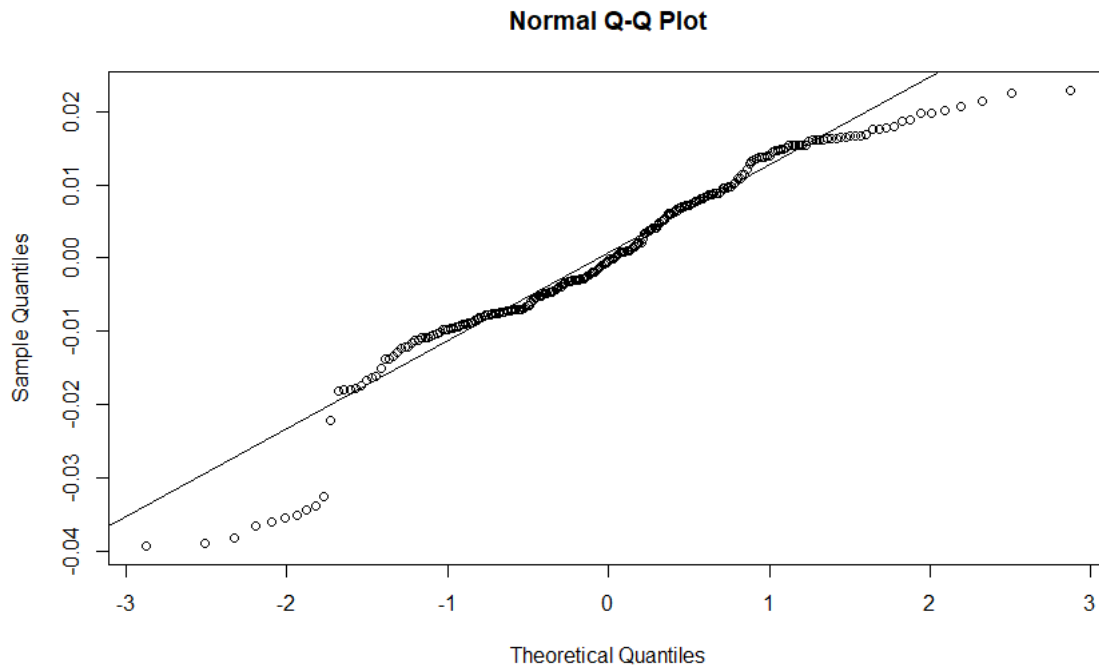


FIGURE 2

La Q-Q plot est une méthode qui trace les quantiles estimées face aux quantiles théoriques : une distribution gaussienne devrait être une droite. On remarque que la distribution des écarts dans l'intervalle $[-1, +1]$ ressemble à celle d'une v.a. normal. Néanmoins, en dehors de cet interval, c.à.d. en correspondance de la queue de distribution, on remarque un comportement très différent et antisymétrique : notamment la queue gauche réste considérablement plus éloignée que celle à droite.

5.2 Plot autocorrelation

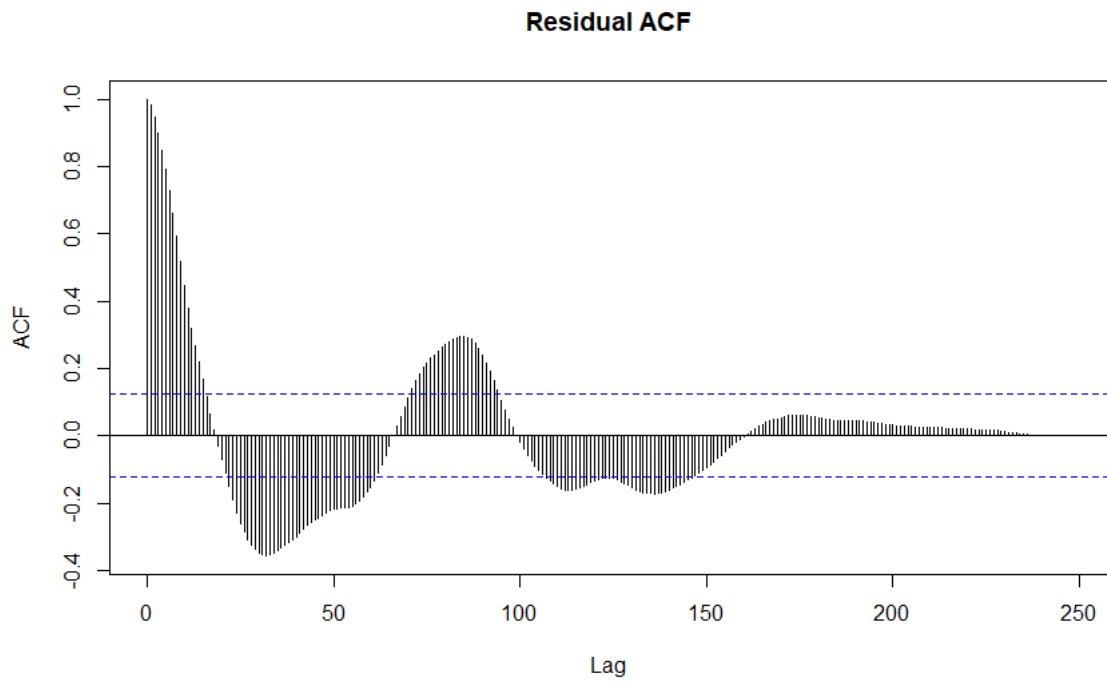


FIGURE 3

D'après le TCL, on sait que :

$$\mathbb{P} \left[\sqrt{n} \frac{\gamma(h)}{\sigma_\epsilon} \in I_{1-\alpha} \right] = 1 - \alpha$$

où un interval de confiance de niveau $1 - \alpha$ d'une loi gaussienne standard. La fonction `acf` en R, plot les intervals de confiance à 95%. On voit qu'il n'y a pas d'indépendance entre les résidus.

5.3 Hétéroscédasticité

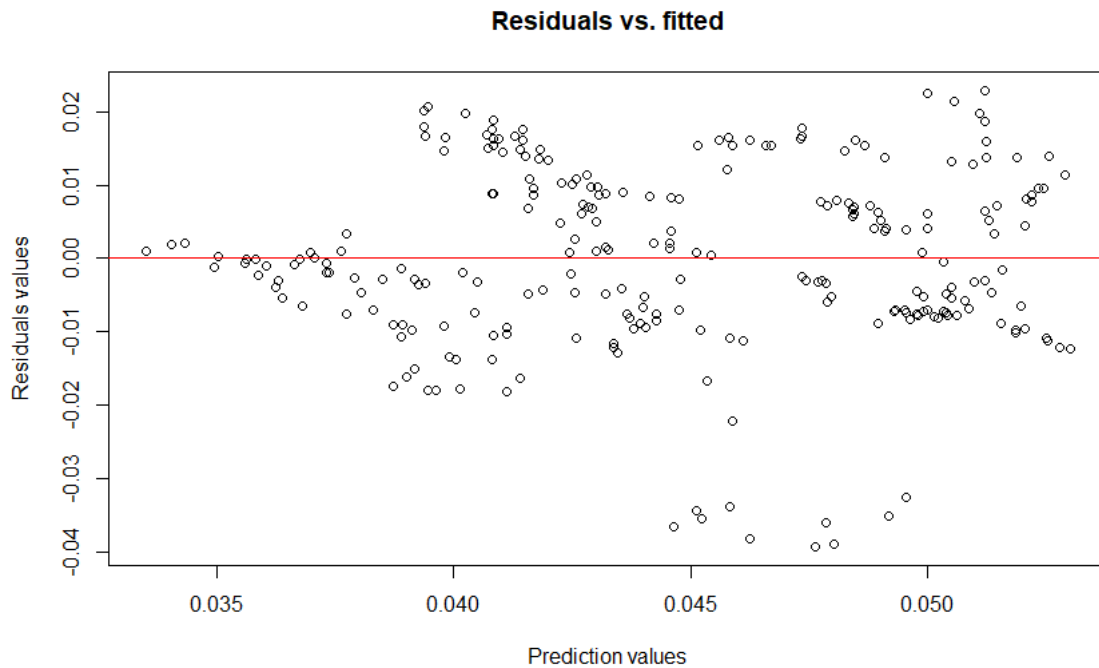


FIGURE 4

On parle d'hétéroscédasticité lorsque les variances des résidus des variables examinées sont différentes. La figure montre clairement la dépendance de la valeur des résidus avec la valeur de la prédiction : plus la valeur des prédictions augmente plus les écart avec la droite $y = 0$ augmentent, ce qu'il entraîne une augmentation de la variance.

Conclusion de l'analyse partielle

Pour conclure on a montré que :

- les résidus ne suivent pas une loi normale.
- les résidus ne sont pas indépendants les uns des autres.
- on a pas d'homoscédasticité : les variances des résidus augmentent avec la valeur des prédictions.

6 Question 6

La méthode des moindres carrés ordinaires appliquée à cette nouvelle spécification améliore le pouvoir explicatif du modèle. Voyons pourquoi.

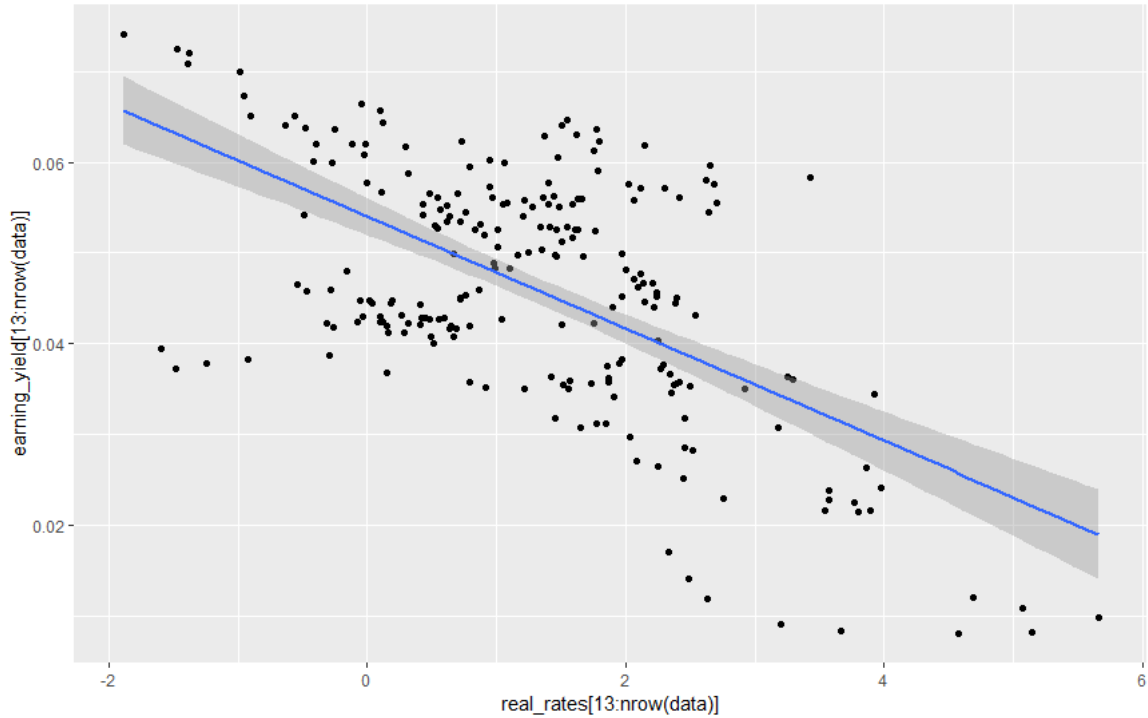


FIGURE 5

La p-valeur pour test t-student donne $< 2e - 16$ et $< 2e - 16$ respectivement pour l'intercept α et β . On rejette donc l'hypothèse nulle ($\beta = 0$) pour le deux paramètre α, β .

Le coefficient de spécificité vaut $R = 0.39$. Par ailleurs, les mêmes considérations sur la normalité, la corrélation et l'hétéroscédasticité, faites dans la question précédente sont valides.

L'introduction de la variable CPI est intéressante afin de donner un modèle qui soit mieux interprétable et permet d'améliorer les performances de l'estimateur à MCO (+26 % pour le coefficient de spécificité). Néanmoins les hypothèses sous-jacents d'optimalité ne sont pas vérifiées.

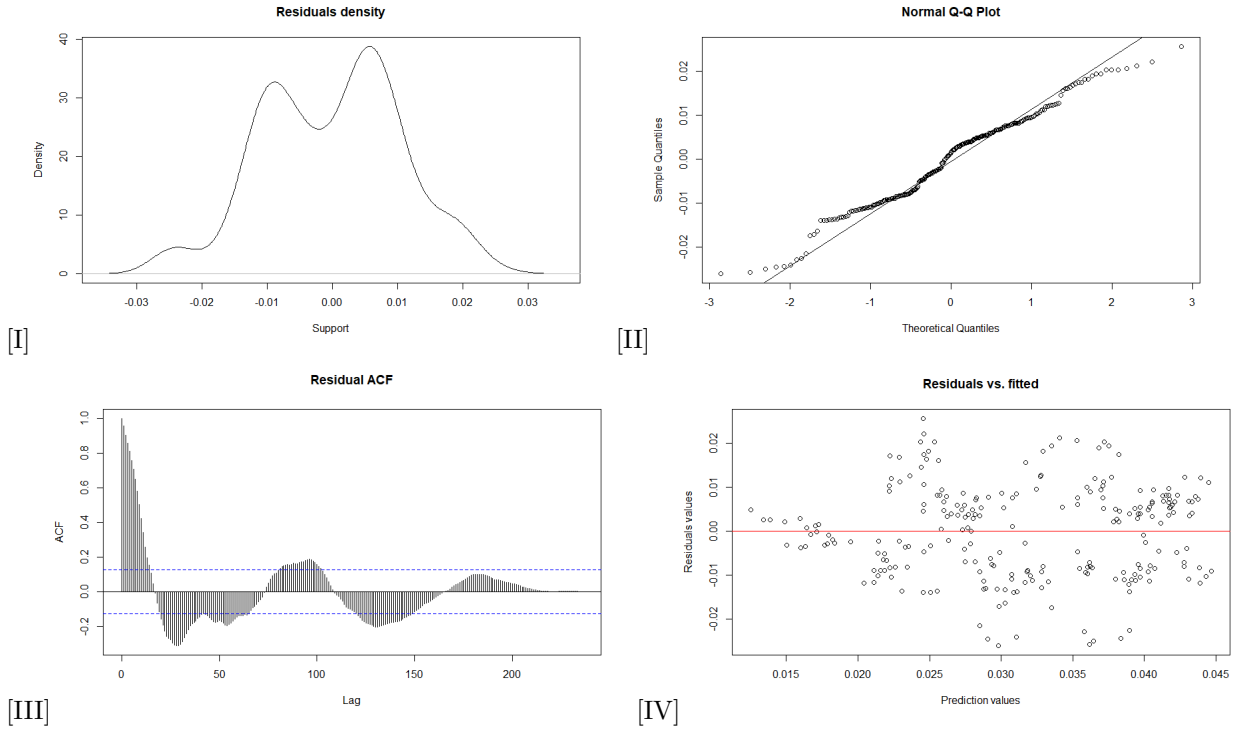


FIGURE 6 – [I] - Densité des résidus, [II] - Normalité, QQ plot, [III] - ACF des résidus, [IV] - Distribution des résidus en fonction de la valeur des prédictions.

7 Question 7

Les processus ARMA(p,q) [auto-regressive (AR) moving average (MA)] généralise les modèles autorégressifs et moyennes mobiles. Il s'agit d'un modèle très utilisés en pratique pour modéliser des séries réelles comme il nécessite de moins de paramètres que les modèles AR ou MA simples.

Définition. La SC (X_t) est appelé série ARMA(p,q) si (X_t) satisfait l'équation aux différences :

$$X_t = \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

où les coefficients ϕ_k et θ_k sont complexes et où ϵ_i représente une suite de VA complexes décorrelées et t.q. $\mathbb{E}[\epsilon] = 0$, $V[\epsilon] = \sigma_\epsilon^2$, notée $BB(0, \sigma_\epsilon^2)$. L'équation précédente peut être re-exprimé en forme plus compacte comme :

$$\begin{cases} \phi(B) X_t = \theta(B) \epsilon_t \\ \phi(z) = 1 - \phi_1 z^1 - \phi_2 z^2 - \dots - \phi_p z^p \\ \theta(z) = 1 + \theta_1 z^1 + \theta_2 z^2 + \dots + \theta_q z^q \end{cases}$$

avec B l'opérateur retard.

Sous les hypothèses suivantes [HI] : ϕ et θ n'ont pas de zéros en commun et $\phi(z) \neq 0$ pour $z \in D_1 = \{z \in \mathbb{C} \mid |z| \leq 1\}$ on a que :

1. X_t est causale et la série admet la représentation suivante :

$$X_t = \frac{\theta(B)}{\phi(B)} \epsilon_t = \sum_{j=0}^{+\infty} c_j \epsilon_{t-j}$$

où c_j sont les coefficients du développement en série de Laurent du $\frac{\theta(z)}{\phi(z)}$ pour $z \in D_1$.
 2. De plus X_t admet densité spectrale f continue dont l'expression est :

$$f(v) = \frac{\sigma_\epsilon^2 |\theta(e^{-i\nu})|^2}{2\pi |\phi(e^{-i\nu})|^2}$$

Sous l'[H1] et sous $\theta(z) \neq 0$ pour $z \in D_1$ on a :
 3. X_t est inversible, ϵ_t admet la représentation :

$$\epsilon_t = \frac{\phi(B)}{\theta(B)} X_t = \sum_{j=0}^{+\infty} a_j X_{t-j}$$

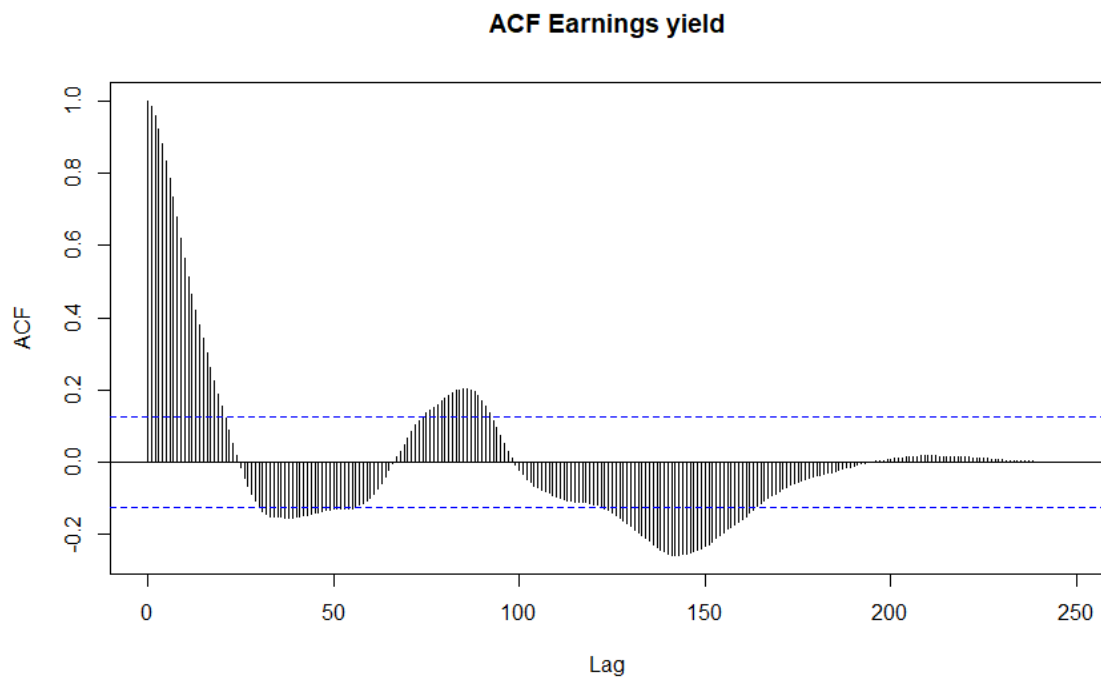
où a_j sont les coefficients du développement en série de Laurent du $\frac{\phi(z)}{\theta(z)}$ pour $z \in D_1$. En outre $f, f^{-1} \in \mathcal{L}^\infty$.
 4. L'innovation de X est ϵ .
 5. La fonction d'autocovariance vérifie :

$$\gamma(h) + \sum_{j=1}^p \phi_j \gamma(h-j) = 0 \quad \forall h \geq q+1 \quad (6)$$

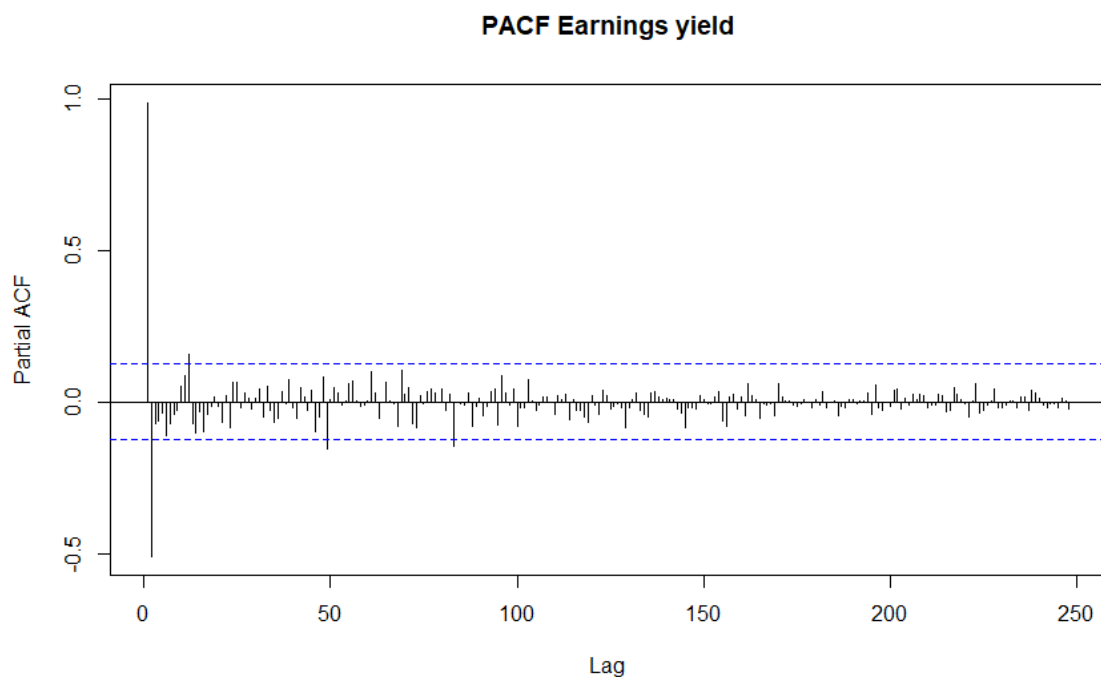
8 Question 8

Une approche parcimonieuse consiste à étudier préliminairement les deux modèles simples dont on connaît des caractérisations à travers la fonction d'autocorrelation et la fonction d'autocorrelation partielle. En particulier on peut étudier les :

1. modèles MA à travers la fonction d'autocorrelation.
2. modèles AR à travers la fonction d'autocorrelation partielle.



[I]



[II]

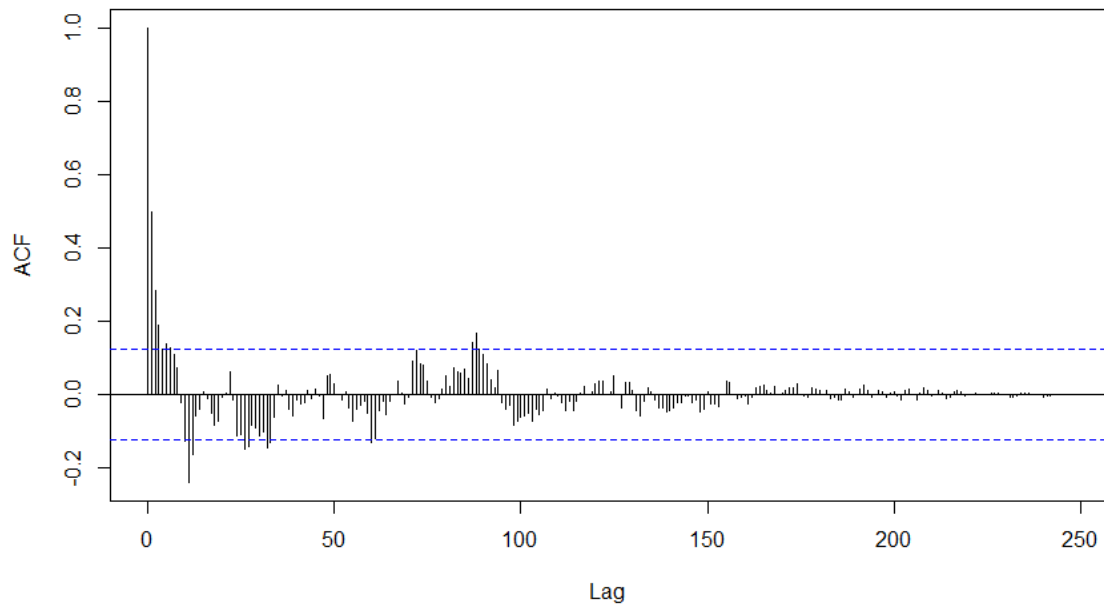
On constate que la valeur absolue de la fonction d'autorrelation tend à se stabiliser vers 0 pour $h \simeq 200$. Sous l'hypothèse de stationnarité, d'après le Th. 3.7 [Poly du cours] le modèle $MA(p)$ pour représenter la SC devrait avoir un ordre $p \simeq 200$, qui n'est pas une quantité raisonnable. On peut remarquer que pour $p > 2$ on a $\phi_{p,p} \simeq 0$. Pour une SC causale, d'après le Th. 4.2. [Poly du cours] on estime $X_t = AR(2)$. Les résultats obtenus nous suggèrent donc qu'on a pas de stationnarité.

9 Question 9

Le modèle identifié par la fonction `auto.arima` est $ARIMA(1,1,0)$. L'analyse conduite dans la question 8 n'est pas donc opportune : on avait implicitement fait l'hypothèse que le processus était stationnaire, alors qu'il présente une tendance constante.

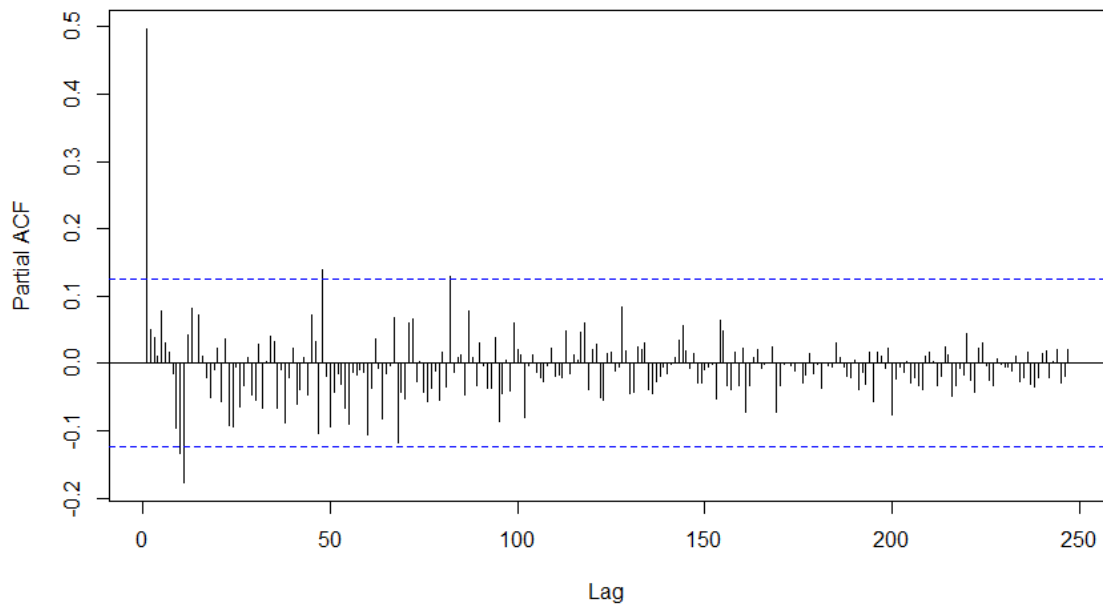
Pour enlever l'effet de la tendance et conduire une analyse similaire à celle de la Question 8, on applique l'opération de différenciation. Les résultats obtenus sont les suivants :

Differentiation - ACF Earnings yield



[I]

Differentiation - PACF Earnings yield



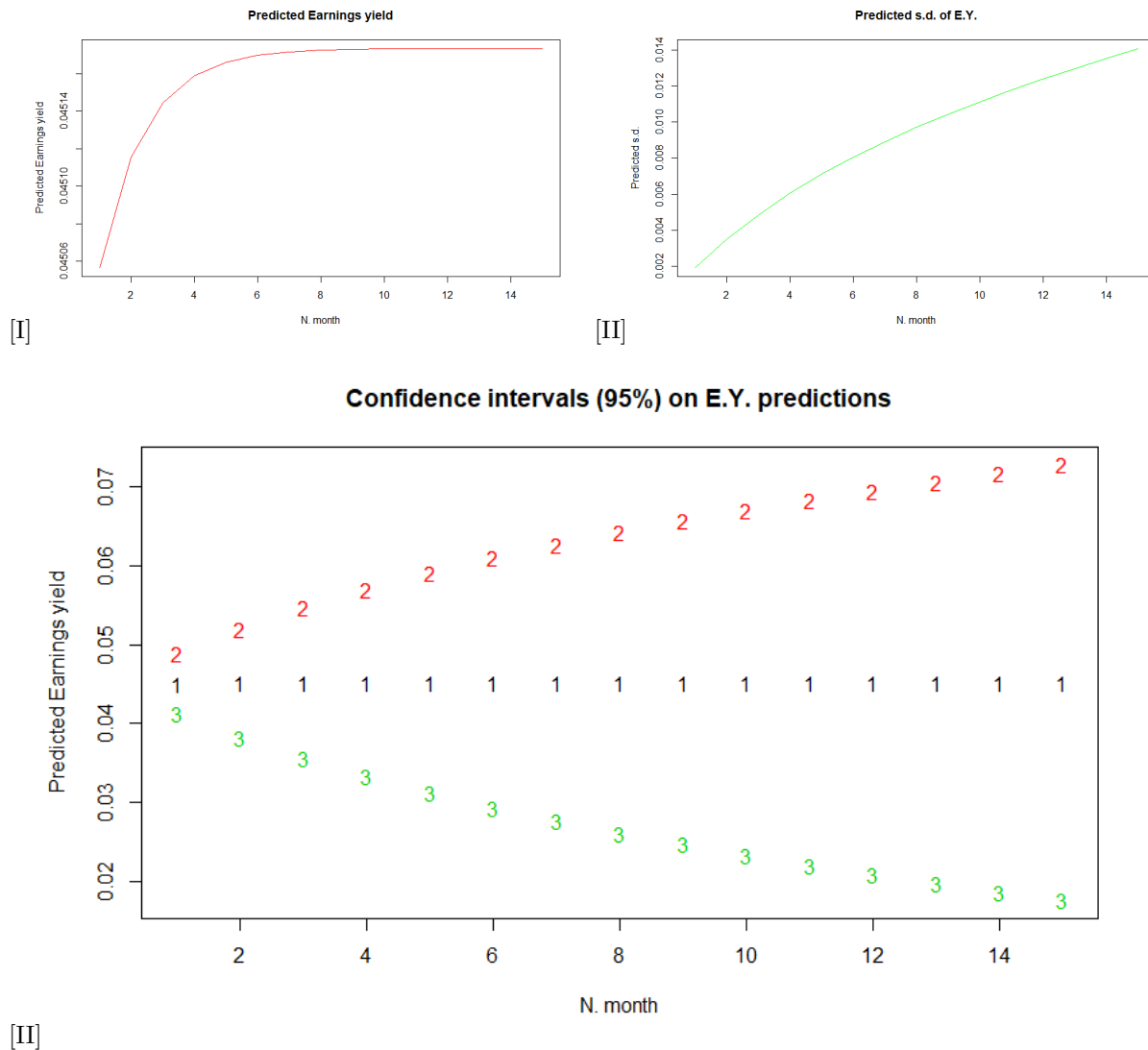
[II]

En effet, en posant $q = 0$, série se réduit à une série $AR(p)$. Or, on remarque que pour $p > 1$ on a $\phi_{p,p} \simeq 0$. D'après le Th. 4.2. [Poly du cours] on estime $X_t = AR(1)$.

Les estimations sont donc en accord avec celles de la fonction `auto.arima`.

10 Question 10

Voici les valeur des prévision pour une fenêtre de dimension $W = 15$.



L'écart type croît d'environ un ordre de grandeur dans cette fenêtre. Les bandes de confiance sont d'autant plus étendue que le temps augmente.

11 Question 11 et 11 bis

Les résultats dépendent de le choix du nombre des échantillon qu'on dispose pour fitter le problème. Clairement, le plus de données qu'on dispose le mieux le modèle ARIMA pourra reproduire le comportement de la série.

N training samples	MAE	RMSE
25	0.0148	0.0175
50	0.0261	0.0281
100	0.0111	0.0156
150	0.0153	0.0177
200	0.0101	0.0106
240	0.0019	0.0023
247	0.0032	0.0035

A l'aide d'un modèle arima dont les coefficients ont été estimés à partir de N training samples, on a calculé les deux statistiques sur les échantillons restants. On obtient toujours des meilleurs résultats avec MAE.

Pour ce qui concerne le test de Diebold et Mariano on a trouvé les résultats suivants :

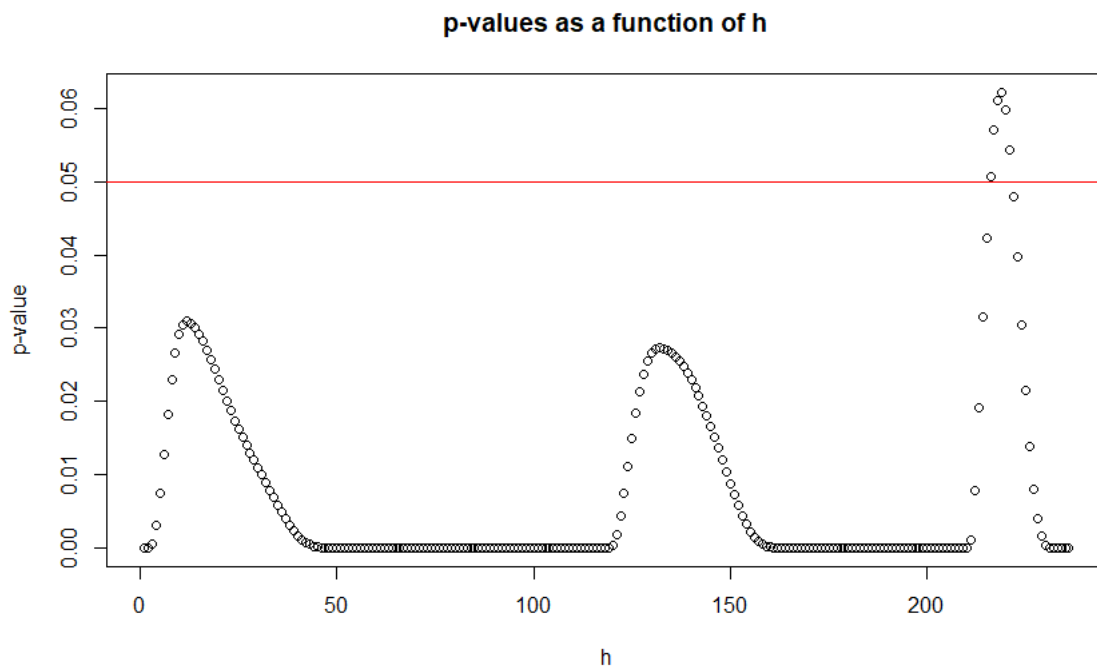


FIGURE 7

Par conséquent on rejette presque tous les horizon de prévision h l'hypothèse H_0 au niveau de significativité de $\alpha = 0.05$: les prévisions ne sont pas équivalentes.

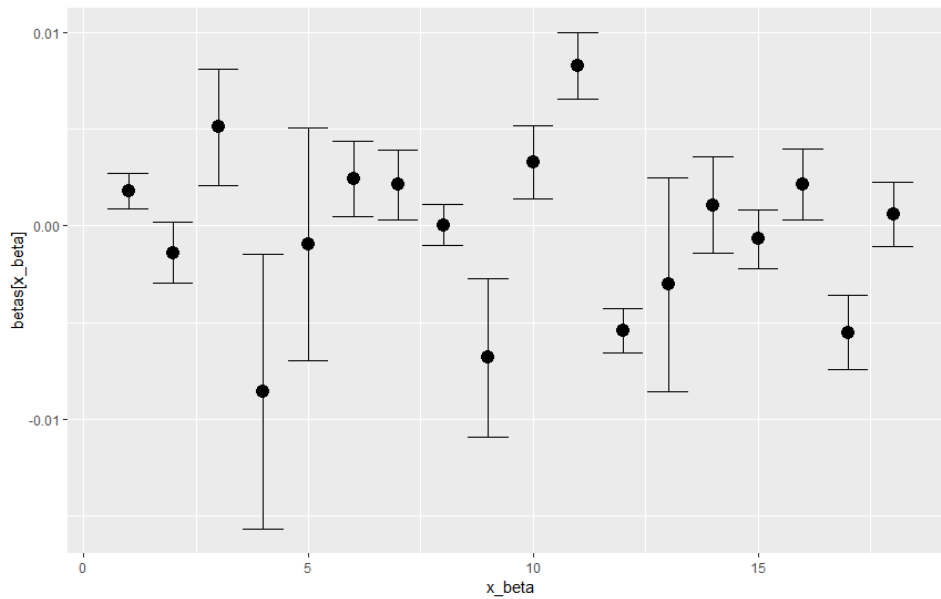
12 Question 12

On a considéré une période T de 12 mois et on a calculé β pour chaque cas et les intervalles de confiance qui lui sont associés.

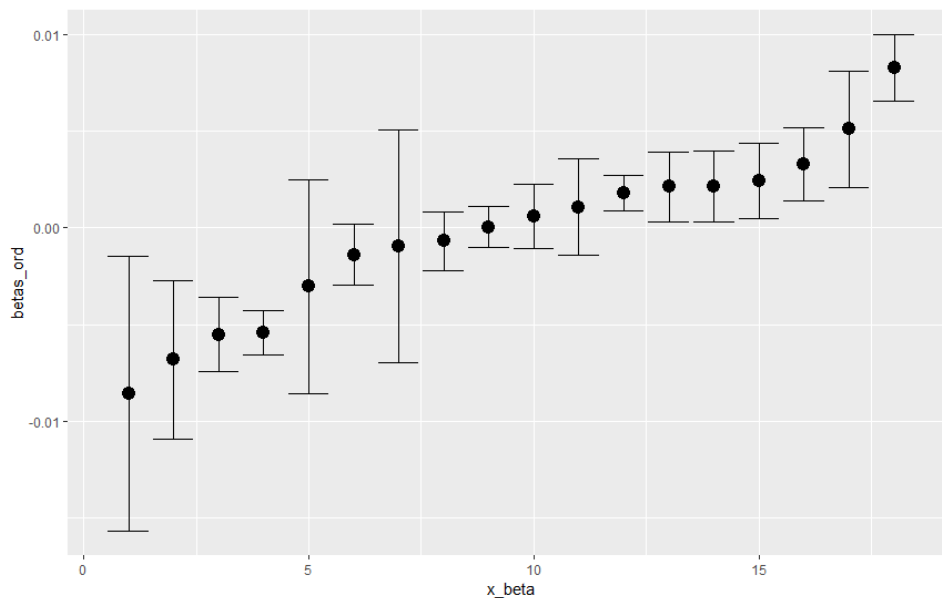
On remarque un très grande variabilité entre les différents périodes soit de la valeur du parametre, aussi bien que de la largeur de l'intervall de confiance relatif.

En outre, on remarque que cette dernière n'est pas forcément liée à la valeur de β contrairement à ceux qu'on avait rémarque dans la Question 5.

Il y donc une forte variabilité qui dépende de la période considérée : l'hypothèse de stationnariété n'est pas valide.



[I]



[II]

FIGURE 8 – [I] - betas (en fonction du temps) et leur intervalles de confiance, [II] - betas (ordonné selon leur valeur) et leurs intervalles de confiance

13 Question 13

CUSUM (Cumulative sum) est une technique utilisée en séries temporelles et utilisé pour le détecter des changements au sein du modèle. Le CUSUM est fondé sur la somme cumulée des résidus récursifs. On note ϵ_t les résidus normalisés. Pour $t \in [0, 1]$ la statistique S_t du CUSUM est définie par :

$$S_t = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i$$

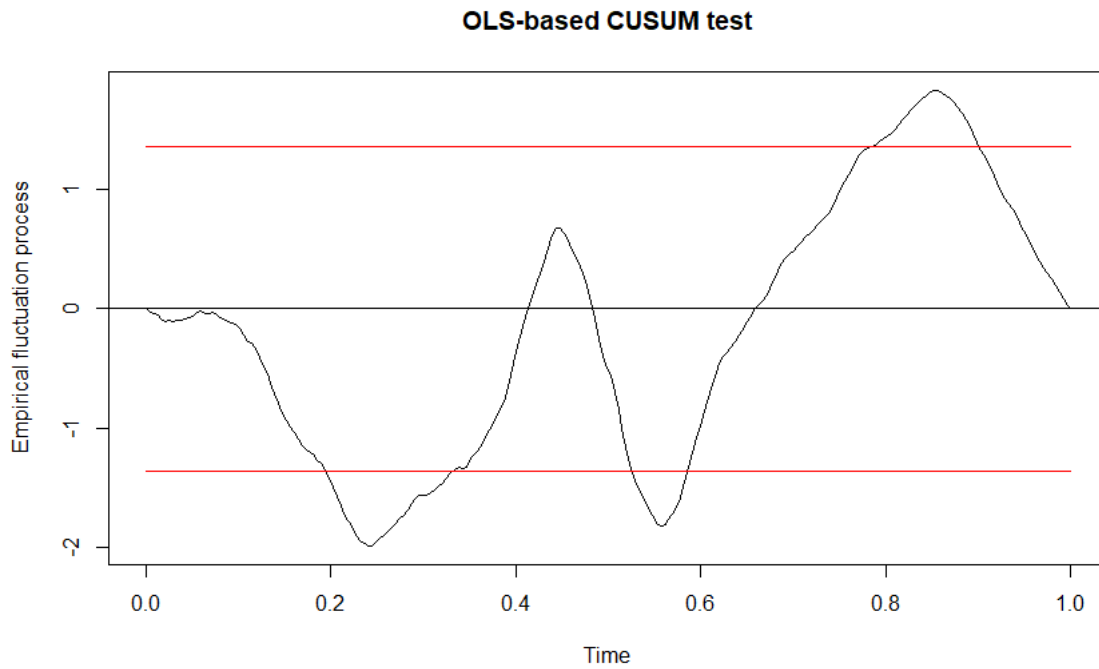


FIGURE 9

Sous l'hypothèse nulle [les coefficients sont stables au cours du temps], la limite du processus est le pont brownien standard. Sous l'hypothèse alternative, la statistique doit dévier à un instant t^* si il y a un changement. On fixe un niveau de significativité $\alpha = 0.05$ et on trace les bandes de confiance correspondantes. Le processus dépasse trois fois les bandes à intervalles réguliers ($t \simeq 0.3, t \simeq 0.6, t \simeq 0.9$) : il y a un changement structurel du modèle.