Série temporelles: TD 1

Quentin Moret
quentin.moret@student.ecp.fr

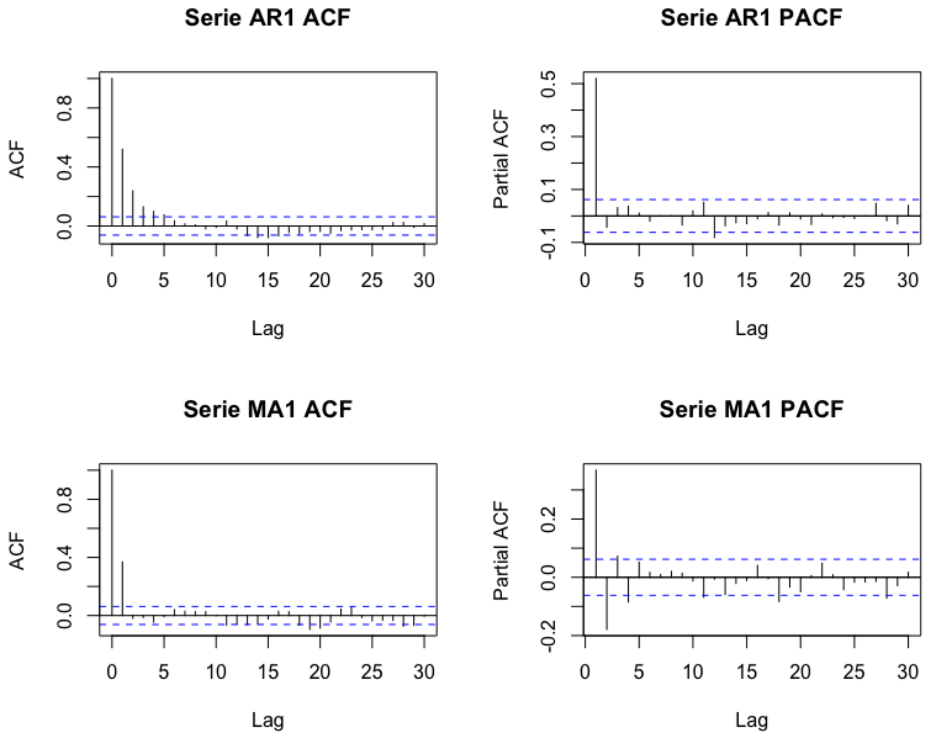
Louis Martin
louis.martin@student.ecp.fr

March 15, 2017

Exercice 1 : Généralités autour des processus stationnaires

Partie A: Simulations de processus AR(p) et MA(q)

Question 1 : Simuler un processus MA(1) puis AR(1) stationnaire.



Question 2 : **Tracer sur un même graphique les fonctions d'autocorrélations et les fonctions d'autocorrélations partielles des processus simulés. Confronter vos résultats à vos connaissances théoriques (vues en cours).**

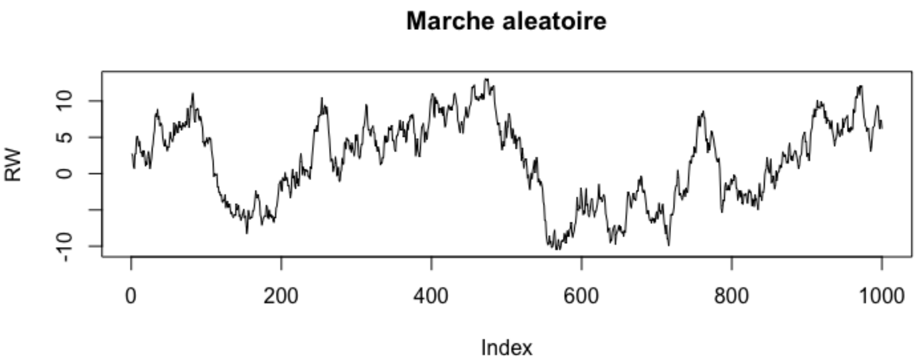
- **AR(1):** $X_t = \phi_1 X_{t-1} + \epsilon_t$
  La courbe d'autocorrélation décroit progressivement. En effet on peut exprimer une variable en fonction des précédentes par application récursive de la formule.
  Exemple: $X_t = \phi_1 X_{t-1} + \epsilon_t = \phi_1^2 X_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t$ (avec la corrélation qui décroit exponentiellement si $\phi_1 < 1$).

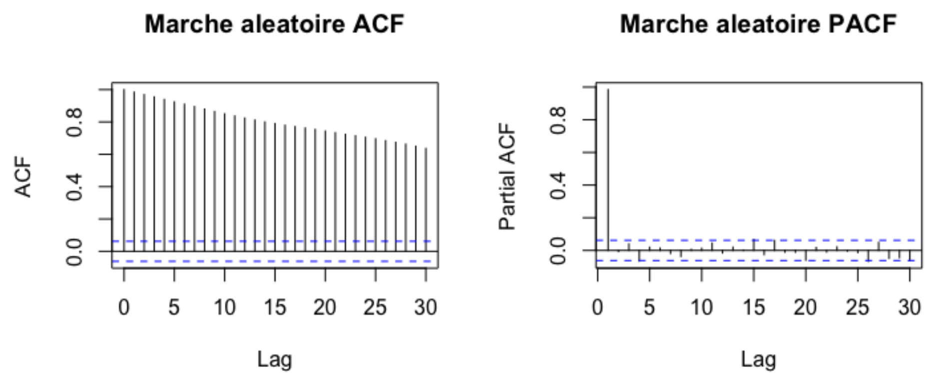
  Ce n'est donc pas l'ACF qui va nous indiquer l'ordre AR de la série. En revanche l'auto corrélation partielle retire à la corrélation la partie du aux dépendances à des variables extérieures. On peut donc très bien voir l'ordre 1 sur cette courbe : le premier pic est très significatif (très forte corrélation à l'ordre 1) et les autres très faibles (pas de corrélations d'ordre supérieur).

- **MA(1):** $X_t = \epsilon_t + \theta_1 \epsilon_{t-1}$
  Pour une série MA(1), c'est le contraire. L'ACF joue le rôle de la PACF pour l'AR et respectivement pour la PACF.

  On retrouve clairement l'ordre 1 sur l'ACF et une décroissance lente sur la PACF qui bien est la signature d'une série MA.

Question 3 : **Simuler une marche aléatoire (minimum 1000 points). Tracer la séries simulées**





Question 4 : Tracer le corrélogramme de la marche aléatoire puis le comparer avec ceux obtenus précédemment (MA(1) puis AR(1)). Le corrélogramme peut-il être utilisé comme un outil permettant de se prononcer stationnarité d'une série

- **Marche aléatoire:** $X_t = X_{t-1} + \epsilon_t$
  On peut réécrire la marche aléatoire comme une somme de bruits blancs:

  $$X_t = X_{t-1} + \epsilon_t = X_{t-2} + \epsilon_{t-1} + \epsilon_t = \sum_{i=0}^{t} \epsilon_i$$

  C'est d'ailleurs de cette manière que nous avons simulé la marche aléatoire de la figure. On retrouve le même comportement qu'une série AR(1) car elle en a l'équation. La corrélation décroît très lentement à cause du coefficient $\phi_1 = 1$, i.e. de la racine unitaire. Cette décroissance lente peut donc nous indiquer que la série n'est pas stationnaire.

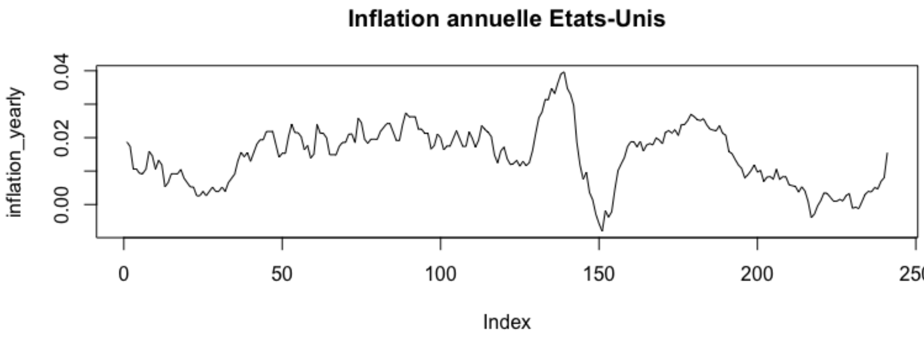
Partie B: Modèle ARMA(p,q) - Applications à l'inflation

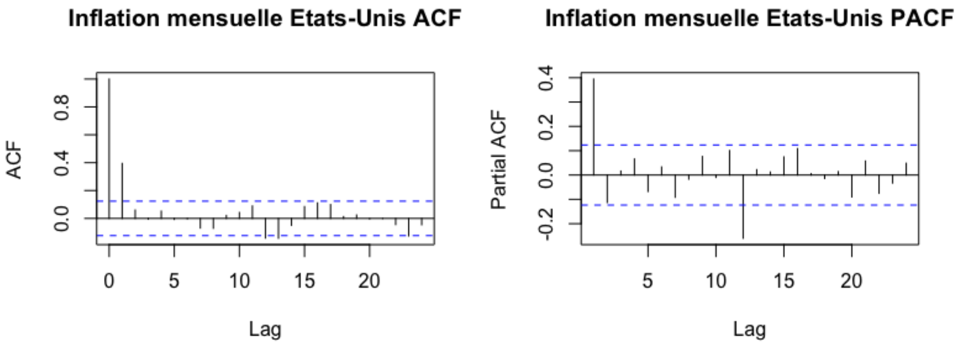
Question 1 : Charger les indices de prix à la consommation pour la France, l'Espagne, les Etats-Unis, la Suède, les Pays-Bas, l'Irelande, l'Italie, la Corée, la Pologne, le Chili, le Danemark et la Suisse sur la période 1996-2017

Nous avons choisi l'inflation des États-Unis.

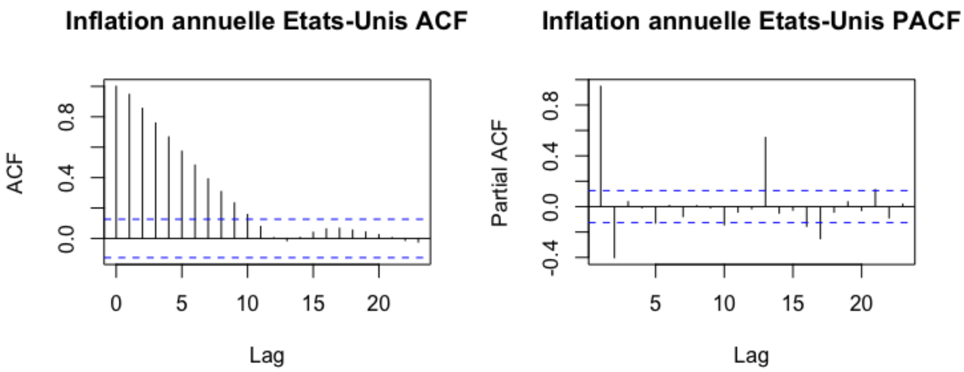
Question 2 : Calculer la variation mensuelle de l'indice de prix à la consommation d'un pays de votre choix. Tracer le corrélogramme et sa fonction d'autocorrélation partielle de la série d'inflation. Commenter vos résultats. Que constatez vous concernant la périodicité de certaines séries.

On remarque clairement sur la PACF un pic significatif à l'ordre 1 pour l'inflation mensuelle et un autre à l'ordre 12. On remarque un comportement similaire sur l'ACF. Cela laisse entendre qu'il y a une dépendance sur la variable précédente (AR(1)) ainsi qu'une périodicité annuelle.





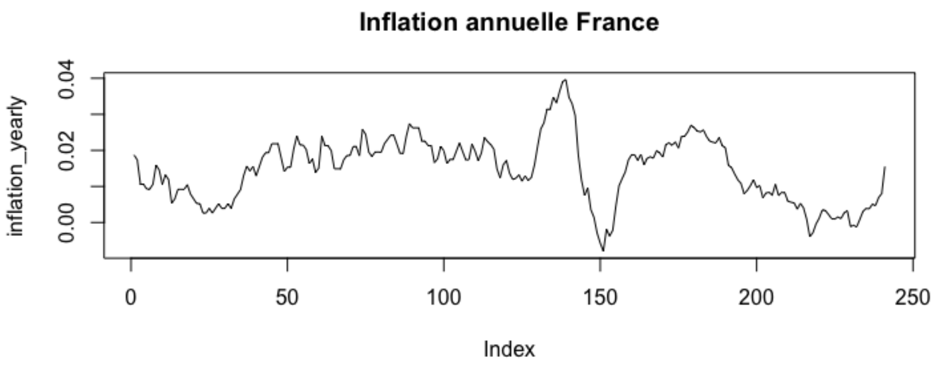
Question 3 : Reprenez la question précédente en calculant le glissement annuelle de l'inflation soit la variation sur 12 mois. Tracer votre glissement annuel ainsi que ses fonctions d'autocorrélation. Quels constats faites-vous ?

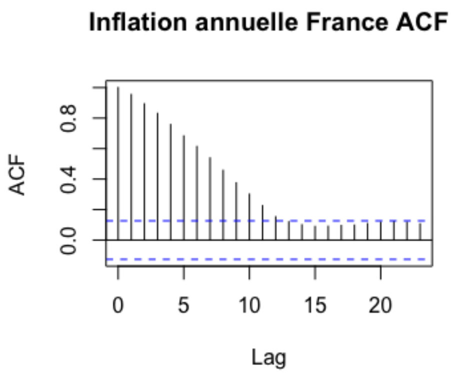


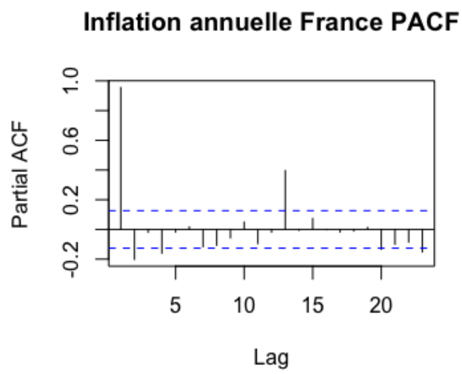
En prenant le glissement annuel on observe une ACF qui décroît progressivement jusqu'à ne plus être significatif à l'ordre 12. La PACF est très significative à l'ordre 1 et 2, et quasiment plus après.

On se propose désormais de modéliser le taux d'inflation en France à l'aide d'un processus ARMA(p,q) et de réaliser une prévision d'inflation.

Question 4 : **Déterminer l'ordre du processus (autrement dit les valeurs de p et de q) à l'aide du corrélogramme et du corrélogramme partiel.**







La France présente un comportement similaire aux États-Unis. En reprenant les constats précédents, un modéle adapté aux observations serait un model ARMA(1, 12) ou ARMA(2, 12).

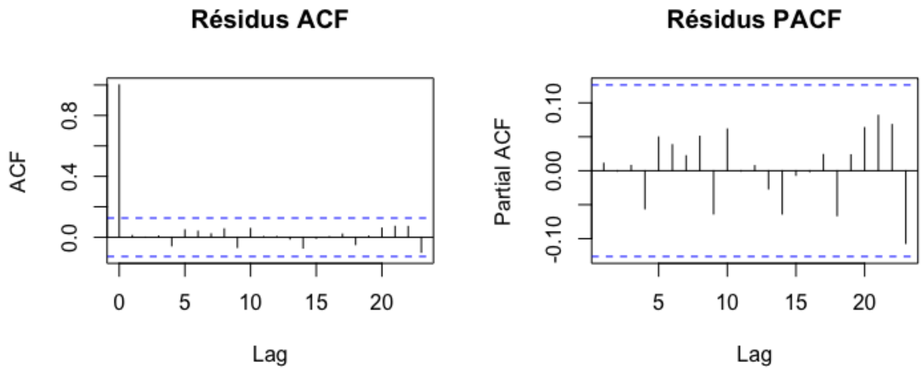
Question 5 : **Confronter vos résultats avec ceux données par les critères d'information AIC. A partir des résultats obtenus, proposer une spécification pour la modélisation de l'inflation française.**

On obtient un score AIC le plus bas avec une modélisation ARMA(0, 14) ou ARMA(3, 11). On retrouve bien le q proche de 12 et un p proche de 1. L'inflation française semble dépendre de l'inflation des quelques mois précédents ainsi que des valeurs d'ϵ_t sur la dernière année.

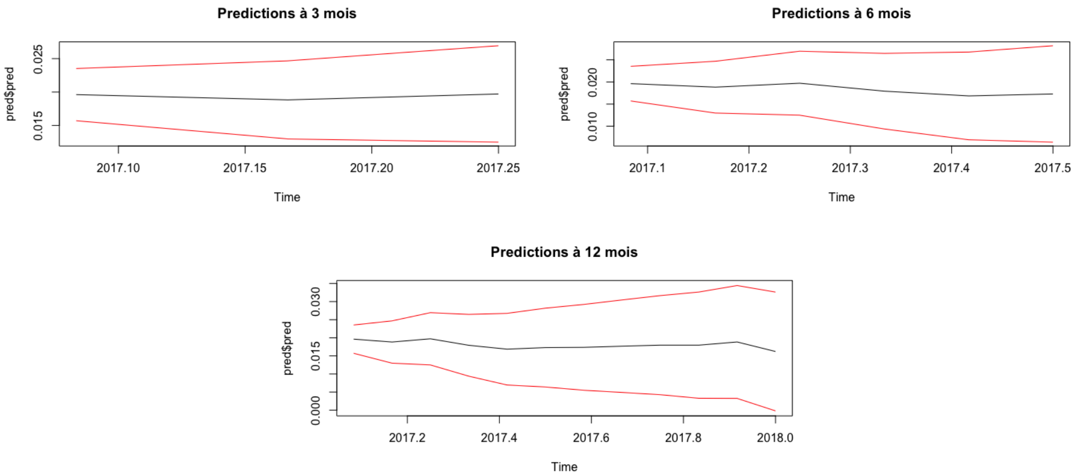
Question 6 : **A l'aide de la fonction "arima" directement implémentée dans R, évaluer par maximum de vraisemblance les coefficients du modèle ARMA(p, q) sur les données d'inflation sur la période 1996 - 2017.**

Voir code.

Question 7 : **Effectuer une simulation de la série l'inflation sur la période d'estimation. Calculer les erreurs de prévision. Tracer le corrélogramme des résidus puis commenter. Peut-on envisager un exercice de prévision dans ce cadre ? On souhaite faire une prévision de l'inflation à trois mois, à six mois et à 12 mois? Effectuer ces prévisions sur les différents horizons indiqués (fonction predict). Tracer graphiquement vos prévisions augmentées de leurs intervalles de confiance à 5%?**



Les résidus ne semblent pas avoir de pics significatifs, ils sont donc probablement IID, on a donc bien capturé toute l'information de la série avec notre modélisation. La prévision sera donc possible et pertinente.



On constate naturellement sur les prévisions que les intervalles de confiance à 95% s'élargissent avec l'horizon de prédiction. Les prédictions à 12 mois semblent lisses car les intervalles de confiances sont grands en valeur relative.

Question 8 : Quels sont les tests adéquats qui permettent de valider le choix d'un modèle. Calculer et tracer les 4 résidus puis procéder à un test d'autocorrélation des erreurs (tests de Durbin Watson et test de Ljung-Box) . Commenter vos résultats. Que concluez-vous? Les modèles ARMA vous semblent-t-ils adaptés à nos besoins ?

Pour valider le choix d'un modèle on peut étudier les erreurs de prédictions / résidus. Si les résidus ne contiennent plus d'information, i.e. sont décorrélés, on conclue que notre modéle est pertinent. On peut utiliser les tests de Durbin-Watson et Ljung-Box. On peut aussi utiliser d'autres méthodes statistiques telles que l'étude du coefficient de détermination.

Le test de Durbin-Watson nous retourne une variable entre 0 et 4: plus la valeur est proche de 2 moins les résidus sont décorrélés (en dessous de 2 les variables sont positivement corrélées et en dessous négativement). On obtient une valeur de 1.96, les résidus sont donc très peu corrélés (légèrement corrélés positivement) ce qui valide notre modèle.

Concernant le test de Ljung-Box, les tests pour différents lags donnent tous des p-values supérieures à 0.8 ce qui nous confirme que les résidus sont bien indépendants.

Les modèles ARMA semblent donc bien modéliser ici la série d'inflation de la France.

Exercice 2 : Le modele du Capital Asset Pricing model (CAPM) - estimation par des methodes standards et flexibles

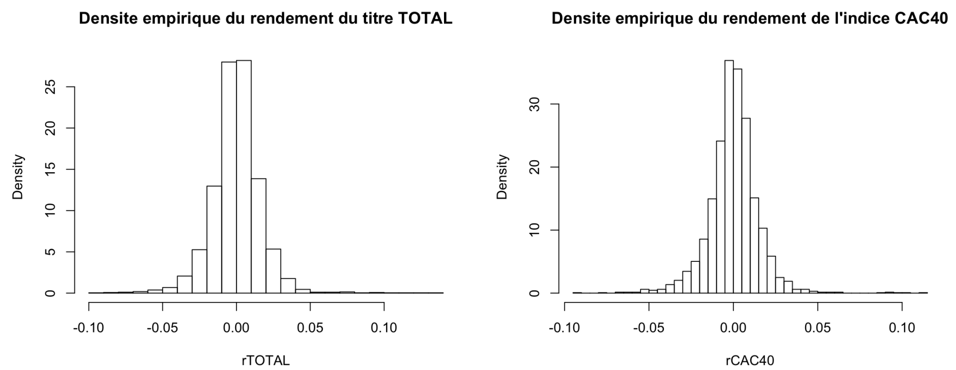
Partie A: CAPM et approche standard

Question 1 :

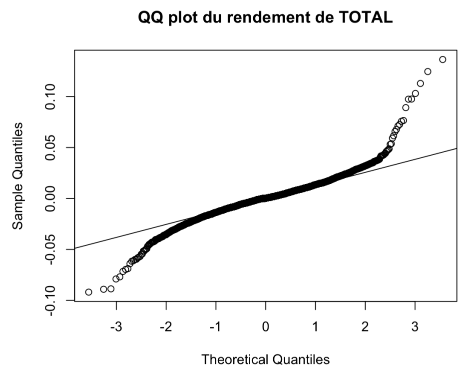
Nous avons choisi le titre 'TOTAL', noté t par la suite. Le rendement du titre est obtenu par la relation

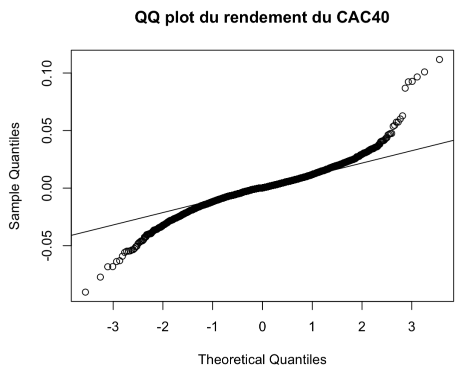
$$r[i] = \frac{t[i+1] - t[i]}{t[i]}$$

Les densités empiriques paraissent celles de gaussiennes.



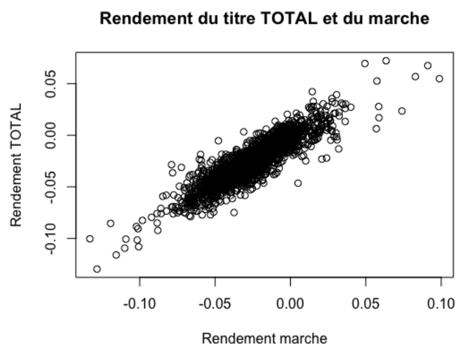
Les queues de distribution ne sont toutefois pas conformes à des gaussiennes, on ne peut conclure sur la normalité de ces distributions.





Question 2 :

Le taux sans risque est ici le "10y".



On observe une nette corrélation entre le rendement du titre total et le rendement du taux sans risque. L'utilisation de la méthode des MCO parait ainsi cohérente.

Question 3 :

En fittant une régression linéaire à l'aide de la fonction lm de R, nous obtenons les résultats suivants :

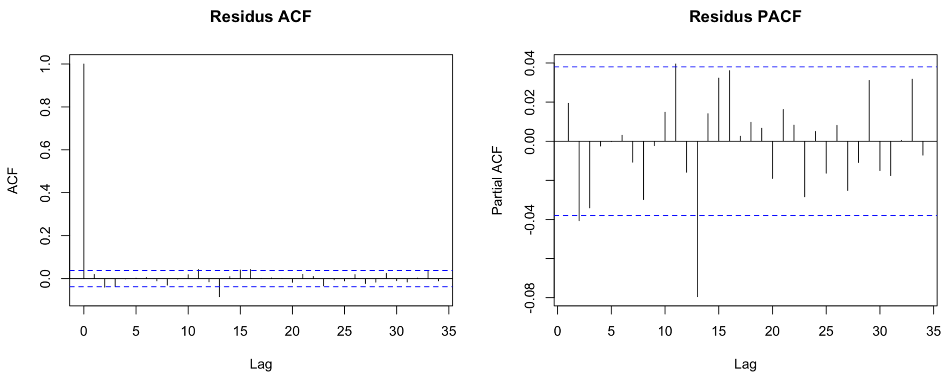
```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.050464 -0.004986 -0.000185  0.004842  0.052223 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -0.0009134  0.0002893  -3.157  0.00161 ** 
exces_marche  0.9649373  0.0090442 106.691  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009266 on 2659 degrees of freedom
Multiple R-squared:  0.8106,	Adjusted R-squared:  0.8106 
F-statistic: 1.138e+04 on 1 and 2659 DF,  p-value: < 2.2e-16
```

On obtient un intercept quasiment négligeable, et un coefficient de régression proche de 1. Ce qui est sans surprise après le précédent graphique, et montre que le rendement de l'action de TOTAL suit le rendement du marché, en étant un peu plus faible. La statistique de Student indique que le coefficient de la régression non nul de manière statistiquement significative. La régression est de plus justifiée dans son ensemble puisque le Test de Fisher indique aussi rejeter la nullité de l'ensemble des coefficients.

Cette méthode permet d'obtenir une estimation d'un actif financier de manière assez grossière, et ne s'adapte pas s'il y a des évolutions de ratio entre l'actif à prédire et l'actif explicatif au cours du temps.

Question 4 :

On remarque un léger pic sur le corrélogramme des résidus du modèle de régression linéaire en $h = 13$. Il semble donc rester une tendance proche d'une saisonnalité annuelle dans les résidus, inexpliquée par le modèle linéaire. Le test de Ljung-Box nous donne des valeurs inférieures à 0.8, ce qui confirme que le modèle n'est pas adapté.



Question 5 :

En utilisant le critère AIC, nous testons différentes combinaisons de p et q. L'AIC le plus faible est obtenu pour $(p, q) = (5, 10)$