Projet Méthodes Statistiques pour la Finance

Auteurs :
Alexandre Bronner
Jean-Joseph Quesnot
Pierre Sabouret

19 février 2016

Table des matières

Nous avons choisi de ne pas mettre de code R ou de sortie R (à part les graphes) dans ce rapport pour essayé de rester le plus concis possible. Pour retrouver les éventuelles sorties, notamment les summary pour les différents modèles, il suffit d'éxécuter le code R fourni.

1 Question 1

Une hyposthèse classique de la finance est que les rentabilités d'un titre suivent une loi normale. Dans cette question nous testons cette hypothèse à l'aide des rentabilités journalières, hebdomadaires et mensuelles du Hang Seng. Dans un premier temps nous avons tracé ces différentes rentabilités 1 :

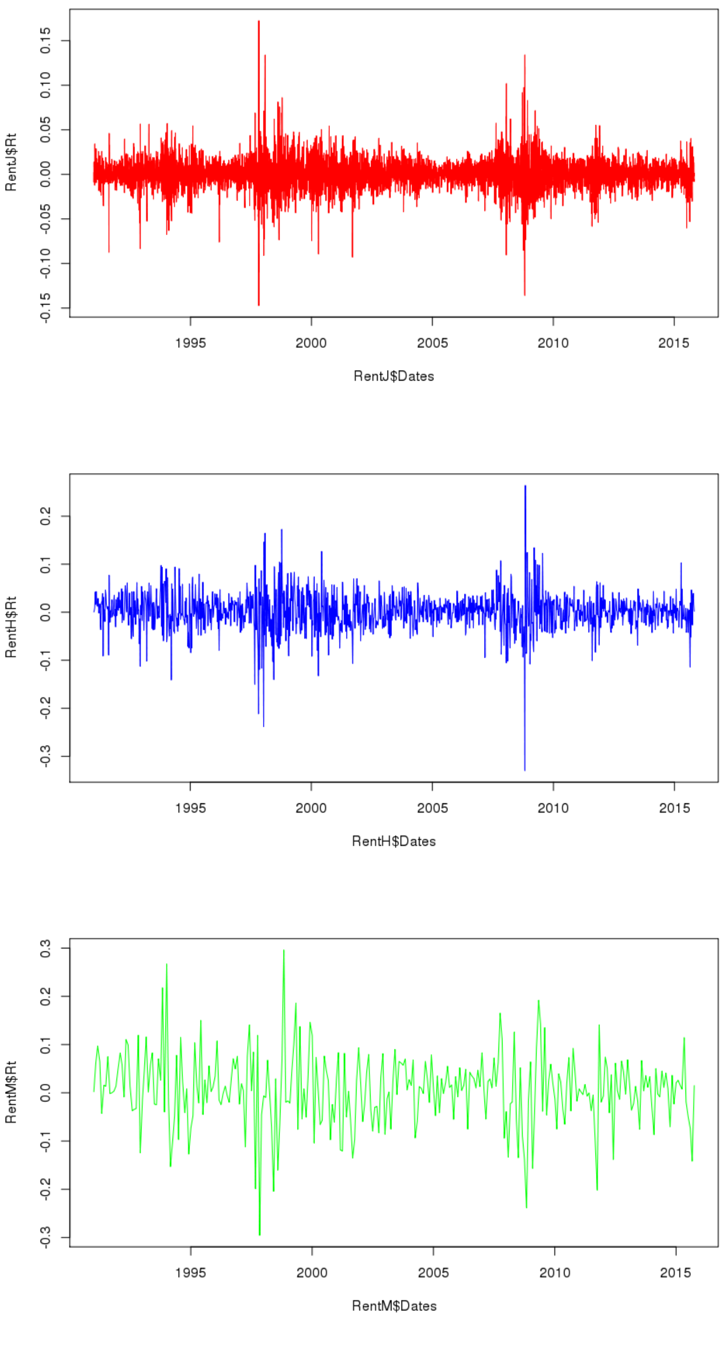


FIGURE 1 – Rentabilités journalières (rouge), Hebdomadaires (bleu) et mensuelles (vert)

Nous savons que les queues de distribution de la loi normale sont minces or ici on peut observerer une fréquence assez importante de "piques", ce qui pousse à étudier l'hypothèse plus en détails. Pour cela on s'intérésse ensuite à la distribution des rentabilités.

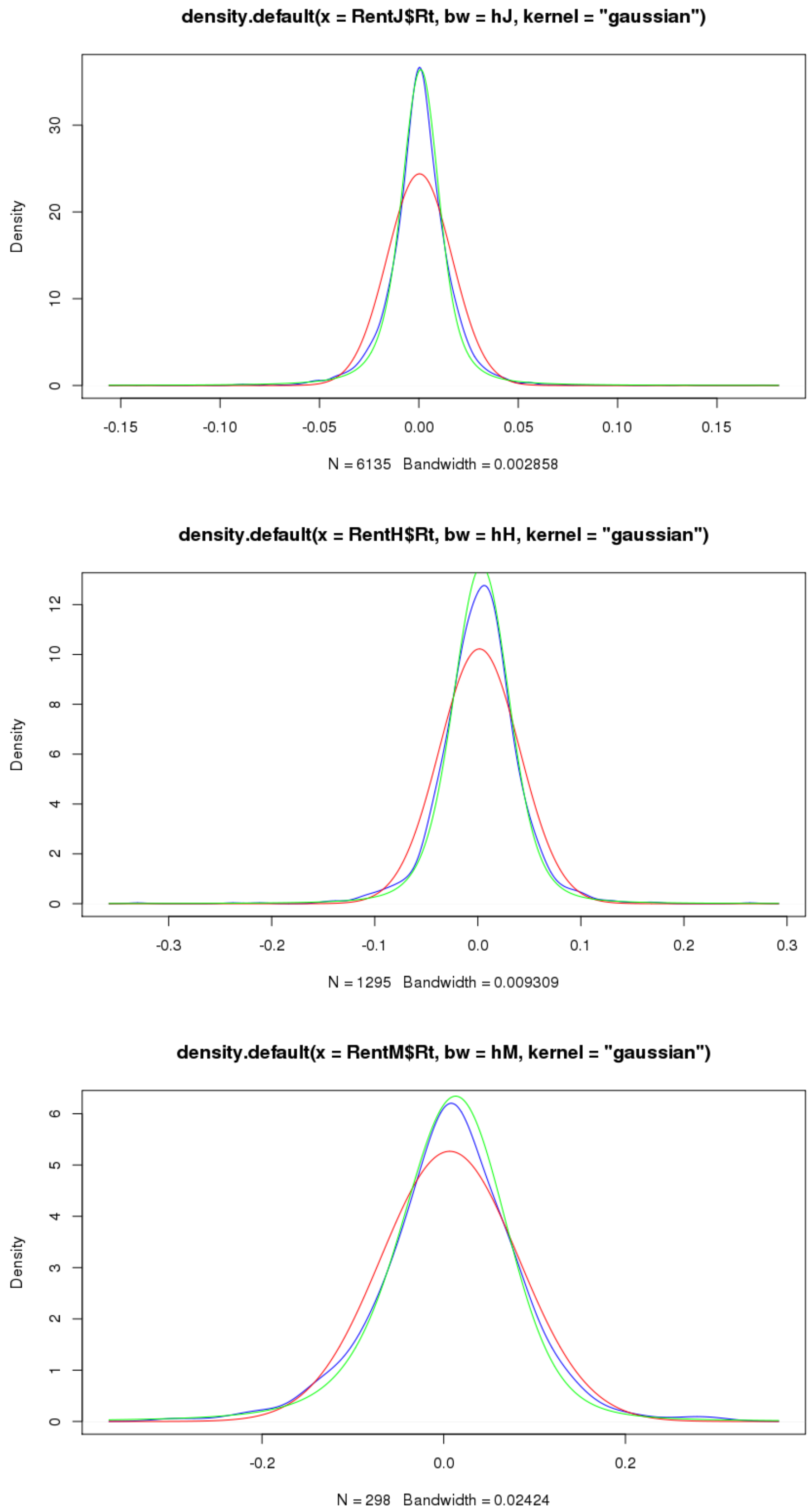


FIGURE 2 – loi normale (rouge), rentabilités observées (bleu) et loi de levy (vert)

Nous remarquons sur ces graphes que les rentabilités observées que l'on considère classiquement suivre une loi normale en fait sont plus proches d'une distribution de Lévy tronquée de paramètres :

Paramètres des distributions de Lévy tronquées		
Journalières	α	1.470
	β	-0.0850
	γ	0.00791
	δ	0.000749
Hebdomadaires	α	1.677
	β	-0.0920
	γ	0.0209
	δ	0.00335
Mensuelles	α	1.642
	β	-0.328
	γ	0.0448
	δ	0.0114

Pour avoir une loi normale comme cas particulier de cette distribution, nous aurions du avoir $alpha = 2.0$; ce qui n'est pas le cas.

Nous remarquons également que plus les rentabilités sont calculées dans un court intervalle de temps, plus les rentabilités sont éloignés d'une loi normale et inversement.

Nous avons également calculés les moments d'ordre 3 et 4 des rentabilités journalières, hebdomadaires et mensuelles.

Moments d'ordre 3 et 4 des rentabilités		
Journalières	γ_3	-0.0018799
	γ_4	12.27967
Hebdomadaires	γ_3	-0.53332
	γ_4	11.17387
Mensuelles	γ_3	-0.143818
	γ_4	4.850249

Ainsi on remarque ques les moments ne correspondent absolument pas à des moments de loi normale en effet on devrait avoir 0 pour le moment d'ordre 3 et 2 pour le moment d'ordre 4. On remarque que le moment d'ordre 3 est négatif, ce qui signifie qu'il y a plus de masse sur le coté gauche de la distribution et donc que l'on a plus souvent des moments de crise que des moments de bénéfices importants.

2 Question 2

Nous avons dans cette partie décidé d'utiliser la fonction eacf de la bibliothèque TSA pour déterminer les coefficients p et q du modèle arma à appliquer.

Nous retenons du résultat de cette fonction que les couples de coefficients maximaux à utiliser sont (2,4) pour les rentabilités journalières, et (2, 1) pour les rentabilités habdomadaires.

Nous avons ensuite testé l'applicaton de modèles arma sur nos données pour des couples coefficients inférieurs aux couples maximum cités plus haut. Nous obtenons les meilleurs résultats pour p, q = 2, 4 concernant les rentabilités journalières. Pour les coefficients de ce modèle, nous avons des intervalles de confiances de l'ordre de 0.001 maximum.

Concernant les rentabilités hedomadaires, nous obtenons les meilleurs résultats pour les coefficients p, q = 3, 1. Nous obtenons alors des intervalles de confiances pour les coefficients de l'ordre de 0.05.

Les P_{value} obtenues sur le set de données hebdomadaires sont satisfaisantes et nous montrent qu'un modèle ARMA colle assez bien à nos données. Les P_{values} obtenues sur le set de données journalières sont très bonnes, et traduisent donc un modèle qui colle très bien aux données. De la même façon, les valeurs AIC obtenues signifient que le modèle posé est un bon compromis entre un faible nombre de paramètres et une bonne log-vraissemblance pour les deux sets de données.

Nous avons de plus calculé le R^2 pour avoir une idée de la performance de notre modèle, afin d'expliciter la part de variance expliquée. Nous obtenons pour les données journalières un résultat médiocre avec une valeur de 0.2%. Le modèle par le calcul de signal sur bruit est donc peu satisfaisant. La valeur qui concerne les données hebdomadaires est quand à elle de l'ordre de 1%, ce qui est mieux dans le calcul de signal sur bruit, mais reste mediocre.

Nous retenons donc que le modèle posé colle parfaitement aux données journalières, mais beaucoup moins pour les données hebdomadaires.

On peut tirer profit de ces de ces dépendances. En effet, si on estime ces coefficients, on peut avoir une idée de la tendance des rentabilités futures et ce jusqu'à des dates supérieures à $p + t_0$ et $q + t_0$. Il s'agit en fait d'utiliser les réalisations connues des variables R_{t0-i} et e_{t_0-i}, pour avoir une idée de la distribution des rentabilités R_{t_0+i}. Cette information peut permettre sur les marchés d'investir dans le bon sens et donc on tire profit de cette information.

Cependant, l'utilisaton d'un modèle aussi simple qui serait efficace serait fortement répendu. Par conséquent on ne peut tirer profit d'une dépendance aussi simple. Dans ce sens, le postulat d'efficience des marchés est respecté.

Nous avons ensuite tracé les histogrammes de la distribution des résidus, afin de verifier si ceux-ci suivent une loi normale. Nous remarquons que cela n'est absolument pas le cas. Que ce soit pour les rentabilités journalières ou hebdomadaires, les dentités des résidus ont un pic plus important en 0, et ont des queues de distributions plus lourdes.

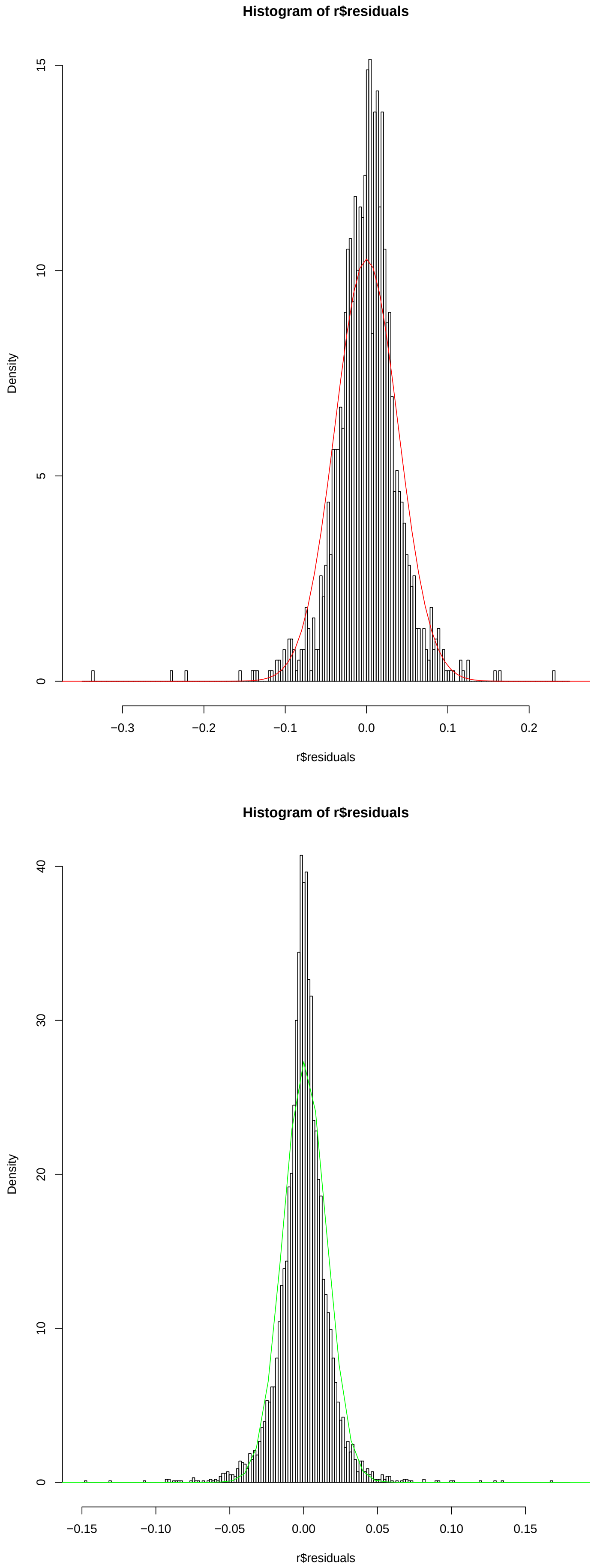
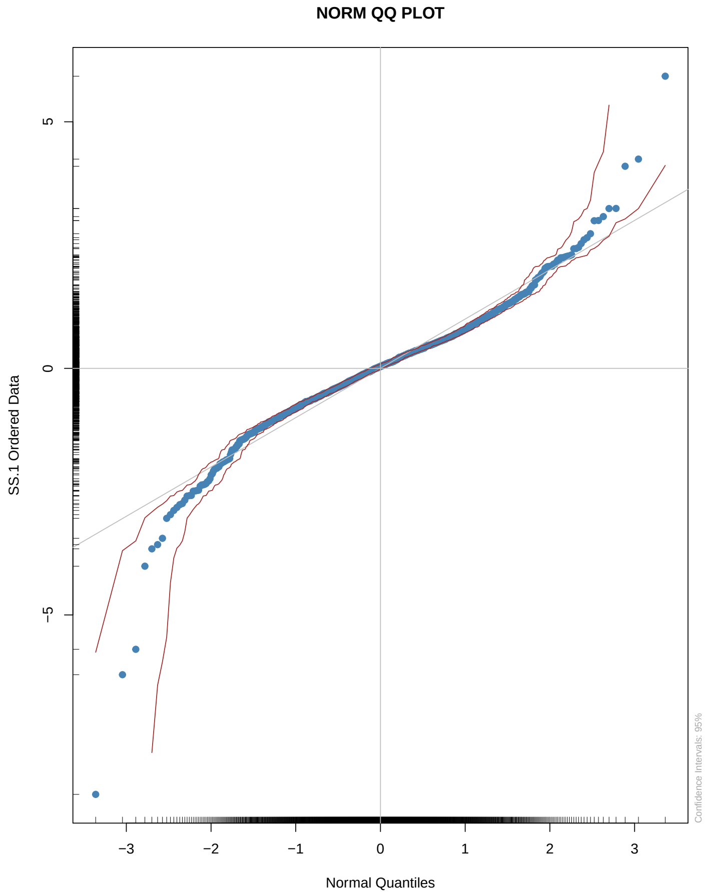


FIGURE 3 – Distribution des résidus



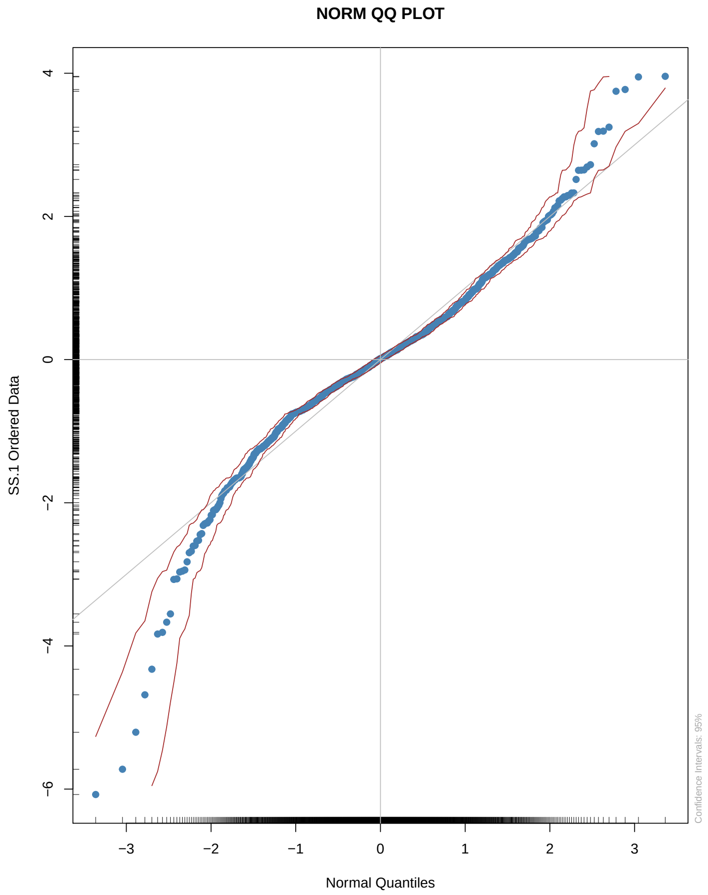


FIGURE 4 – Distribution des résidus

3 Question 3

Pour cette question, nous avons décidé d'afficher dans un premier temps les densités estimées des 3 états. Nous notons donc que ces trois états ont des densités qui se rapprochent de celle de la loi normale. Voici les 3 diagrammes correspondants.

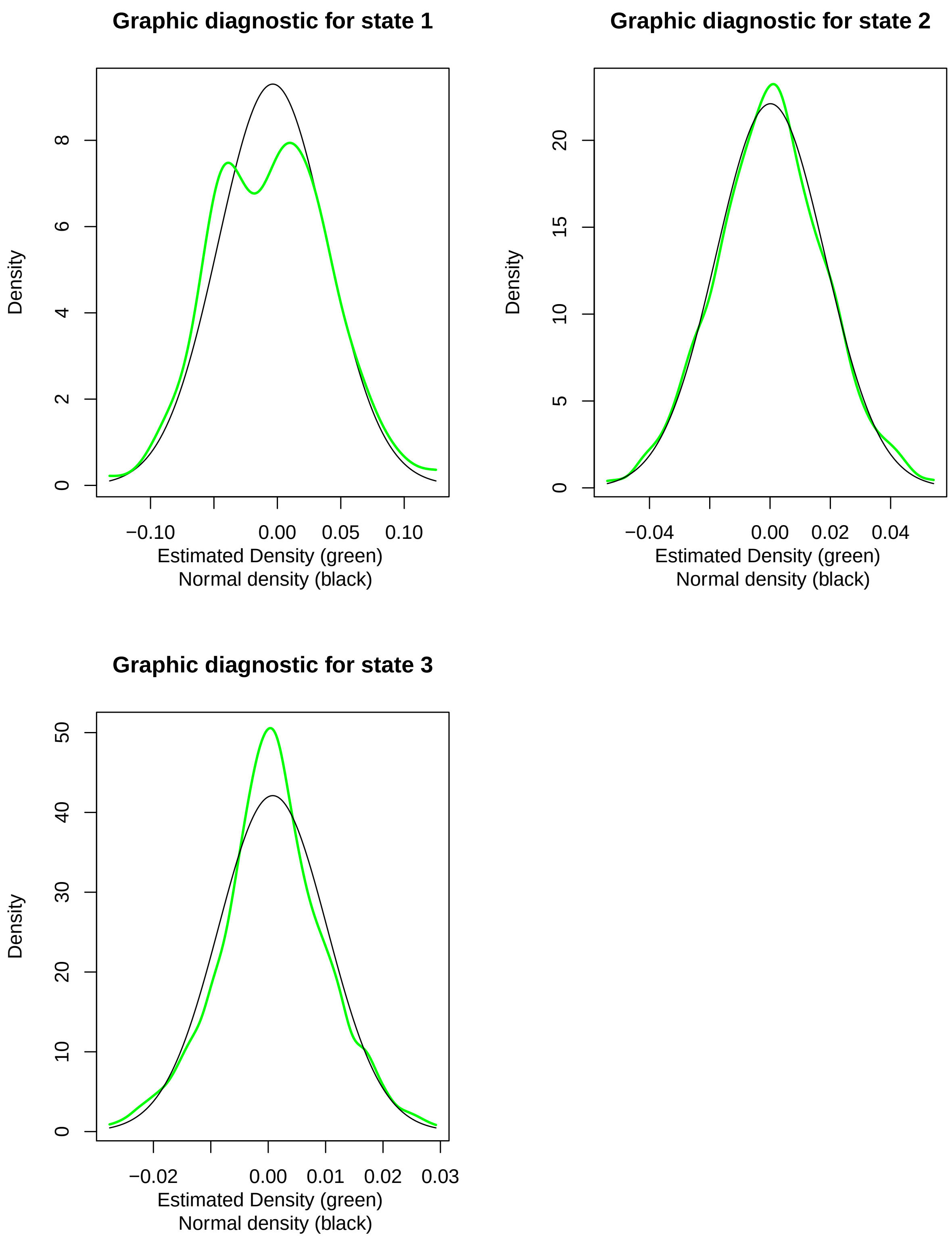


FIGURE 5 – Distribution des Rentabilités sous les 3 états observés

Plus particulièrement, nous avons remarqué que l'état 1, qui correspond à l'etat de marché le plus extrème est plutôt mal représenté par une loi normale. Sous la loi normale, comme prévu, cet état a une variance très importante. Nous remarquons de plus que bien que la loi soit centrée, les données sont plus orientées vers des valeurs négatives. Ceci nous traduit donc que pour cet état, les rentatabilités ont tendance à être négative avec une plus forte probabilité. De plus, les queues de distribution sont moins lourdes que celles d'une loi normale.

D'une tout autre façon, dans l'état 2, les données collent parfaitement à une loi normale. La variance de cette loi est bien moindre que celle de l'état extrème 1. L'état 3 correspond à l'état de marché le plus calme. Dans ce cas, les queues de distributions sont plus lourdes que celle d'une loi normale, qui modélise donc plutôt mal cet état.

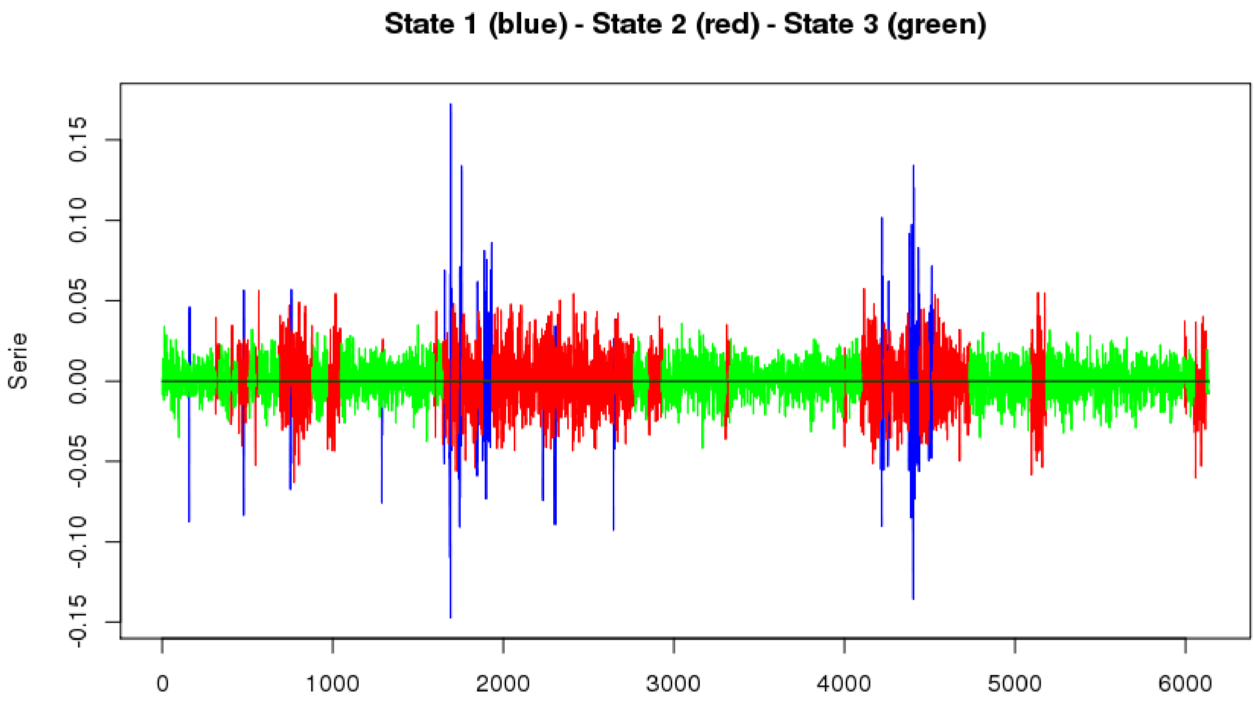


FIGURE 6 – Graphe de la représentatoin des états sur le temps

Nous avons tracé sur ce graphe les rentabilités en fonction du temps. Nous avons de plus coloré en fonction de l'état. Nous avons donc en bleu l'état de maché extrème, en rouge l'état intermédiare et en vert l'état calme de marché. Nous remarquons à tord sur ce diagrame, que les états de marché calme et intermédiaire sont également répartis, à l'inverse de l'état de maché extrème qui est beaucoup plus rare.

La représentation graphique nous induit la en erreur. En effet, bien au contraire, en regardant de plus près les paramètres en sorti de notre test nous observons que l'état le plus calme est en fait le plus présent dans le temps. La présence des autres états est quant à elle négligeable. Nous obtenons en effet les probabilités des différents états suivantes :

$Pi_1 = 4.294e - 174$

$Pi_2 = 1.557e - 51$

$Pi_3 = 1.000e + 00$

Pour ce test, nous obtenons de très bonnes P_{values} pour les différents paramètres ce qui nous traduit une bonne significativité de ceux-ci.

4 Question 4

Nous obtenons les valeurs suivantes pour les moyennes m1 et m2 et les variances v1 et v2 respectivement pour la première et la seconde période.

$m_1 = 3.7 * 10^{-4}$

$m_2 = 2.8 * 10^{-4}$

$v_1 = 3.1 * 10^{-4}$

$v_2 = 2.2 * 10^{-4}$

Nous remarquons donc que sur les deux périodes, les valeurs de ces paramètres varient de l'ordre de 30%. La première période avec des valeurs supérieures était donc plus propice au gain, avec une meilleur rentabilité et une variance plus importante.

Nous appliquons ensuite des tests pour évaluer si ces valeur sont réellements significativement différentes. Concernant les moyennes, nous avons appliqué un test de Student. Ce test nous donne une P_{Value} égale à 0.82, et un intervalle de confiance de 95% qui englobe les deux valeurs avec une marge de 700% de chaque coté. Nous pouvons donc affirmer que ces deux moyennes ne sont pas significativement différentes, contrairement à ce que notre première analyse laissait penser.

Au contraire, nous obtenons une P_{value} très faible pour les tests de Fisher avec $P_{value} = 10^{-15}$. Nous concluons donc que les variances sont significativement différentes, tout comme la première analyse naïve le laissait penser.

Nous tirons de cette information qu'il n'est sans doute pas approprié d'entrainer notre modèle sur les deux périodes concaténées. Les moyennes sur ces deux périodes semblent similaires. Cependant, les variances sont clairement significativement différentes. Il est donc absurde d'entrainer un modèle sur des sets de données qui ont des caractéristiques de bases aussi différentes.

5 Question 5

Dans cette question nous nous intéressons à la correlation entre les valeurs absolues de deux rentabilités journalières consécutives, ainsi qu'entre leurs carrés. Pour cela nous avons tracé les fonctions d'auto-corrélation et les fonctions d'auto-correlation partielles 7 et 8.

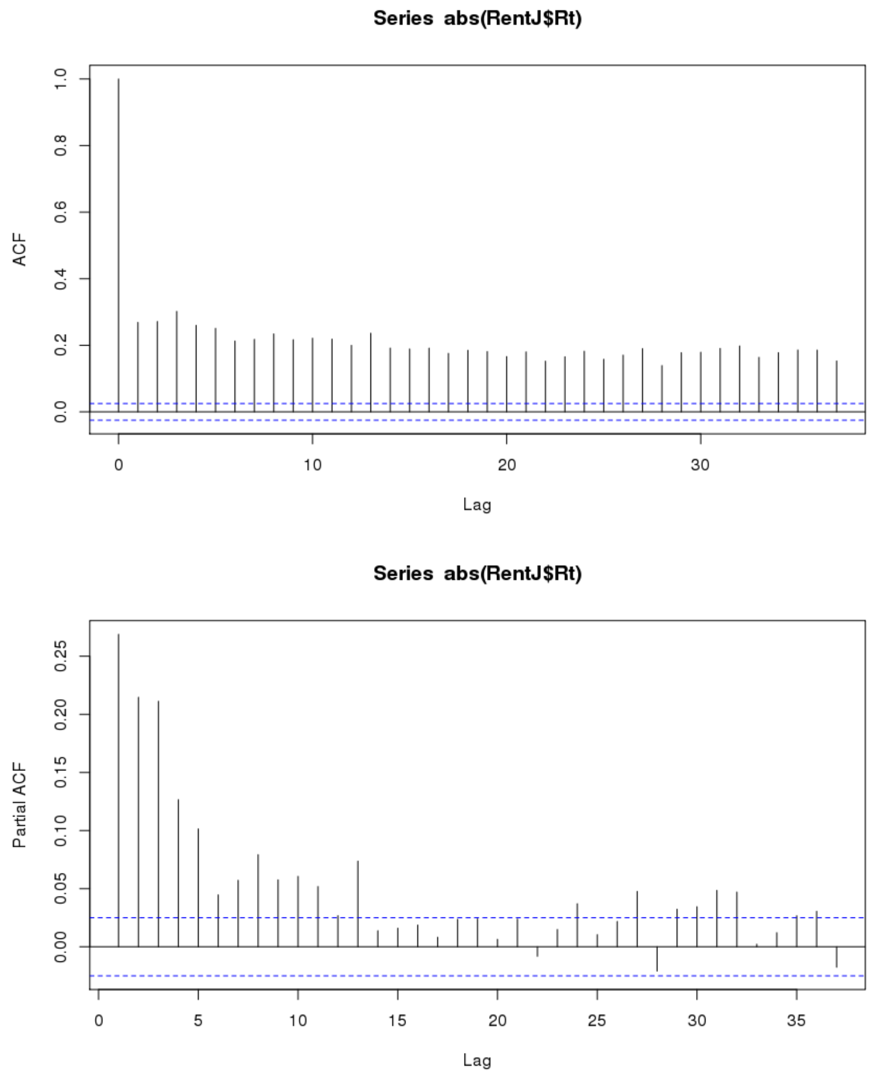


Figure 7 – Autocorrélations (en haut) et autocorrélations partielles (en bas) des rentabilités journalières

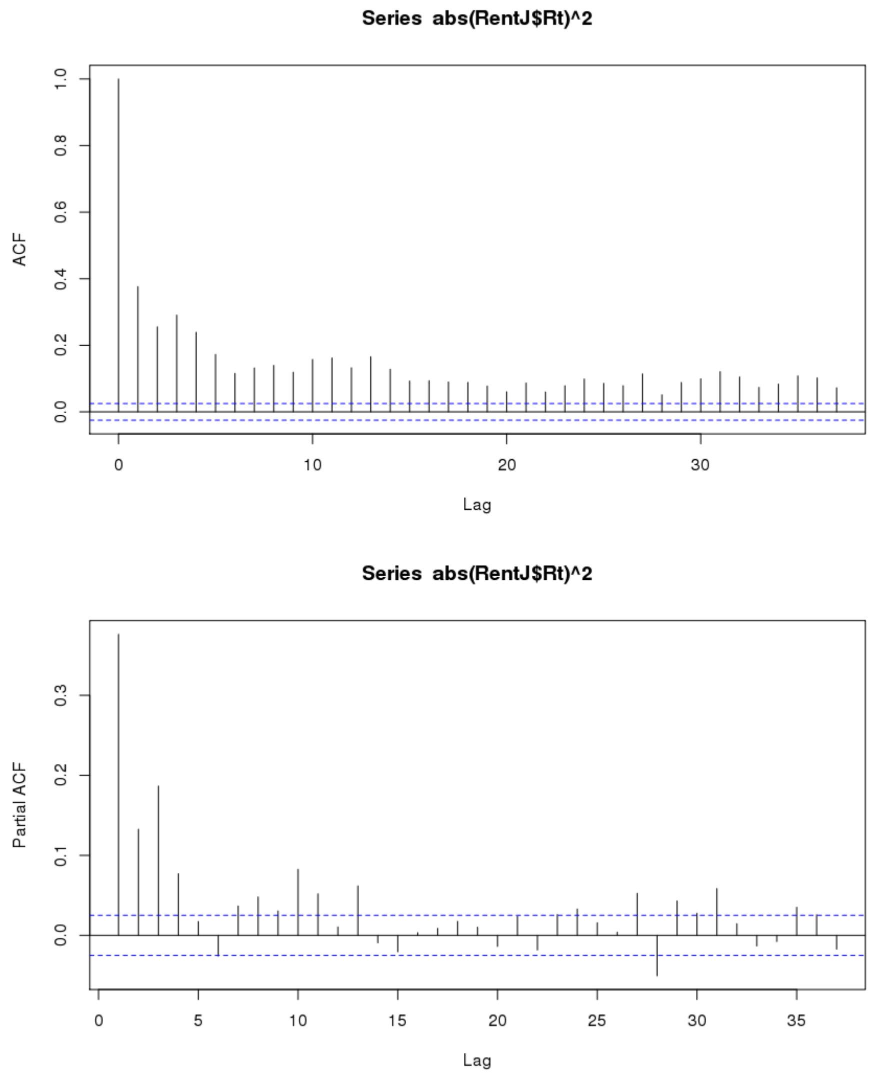


FIGURE 8 – Autocorrélations (en haut) et autocorrélations partielles (en bas) des rentabilités journalières au carré

D'après les graphes de 7 on constate qu'il y a une corrélation positive entre la valeur absolue de la rentabilité d'un jour t donné et la valeur absolue de la rentabilité de la veille. Cette correlation est faible, de l'ordre de 0.25, mais nous indique tout de même qu'un mouvement à un jour t peut être suivi le lendemain d'un mouvement d'une ampleur similaire. Prendre les rentabilités au carré nous donne une information sur la variance des rentabilités. Ainsi les graphes de 8 nous indiquent qu'il y a aussi une corrélation entre les variances des rentabilités journalières.

Un moyen d'exploiter ces résultats et d'affiner notre modèle, en effet les ACF et PACF, en particulier celles de la série $Y(t)^2$ permettent de donner une idée de l'ordre que l'on peut prendre dans un modèle ARCH.

6 Question 6

Nous avons vu dans les questions précédentes que les rentabilités journalières qu'il existe une correlation au niveau de la variance des rentabilités. Pour expliquer cela on peut essayer d'utiliser un modèle GARCH/ARCH en plus du modèle ARMA.

Le problème majeur ici est que la convergence des algorithmes d'optimisation est difficile, dans de nombreux cas il n'y a tout simplement pas convergence et ceci pour tout les algorithmes fournis par $fGarch :: garchFit$.

Pour juger la qualité des modèles nous nous sommes basés sur :

- Valeur du AIC
- Le R^2 ajusté
- Statistique de Ljung-Box
- Auto-corrélation et auto-corrélation partielle des résidus
- Si les coefficients sont significatifs ou non (p-value)

6.1 Rentabilités journalières

En partant du modèle de base ARMA(3,5) aucunes combinaisons avec un GARCH/ARCH ou ARPACH ne convergent et cela pour tout les algorithmes proposés (y compris ceux pour la Hessienne).

Nous avons testé tout les modèles ARMA(i,j)+GARCH(l,m) ou ARMA(i,j)+APARCH(l,m) avec $(i,j) \in [1,7]$ et $(l,m) \in [1,4]$.

La plupart des modèles se valent, leur qualité est très faible, le R^2 est toujours très faible et les résidus sont souvent corrélés et loin d'une loi normale. Nous avons trouvé pour que le "meilleur modèle" est le modèle ARMA(2,4)+GARCH(2,1), nous trouvons un R^2 de $0,5\%$ et un AIC $-5,75$:

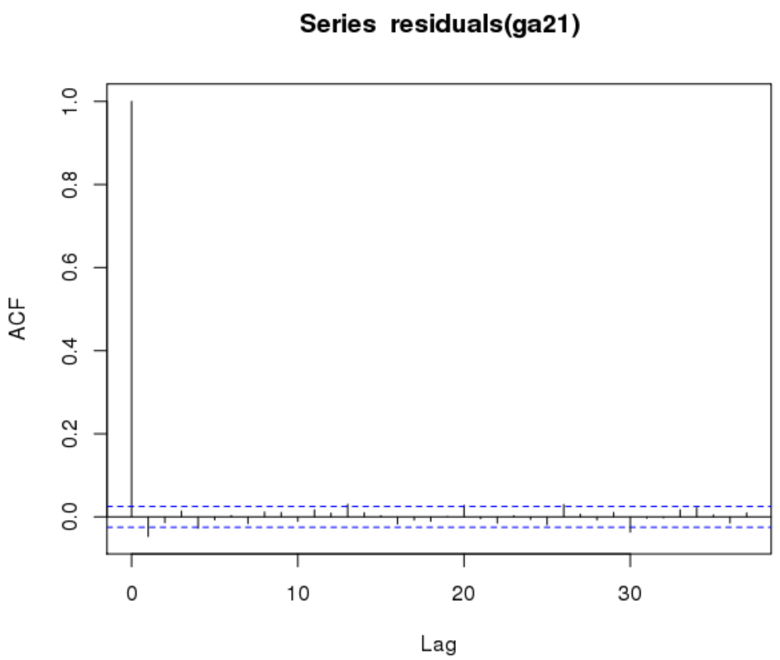


FIGURE 9 – Corrélation des résidus

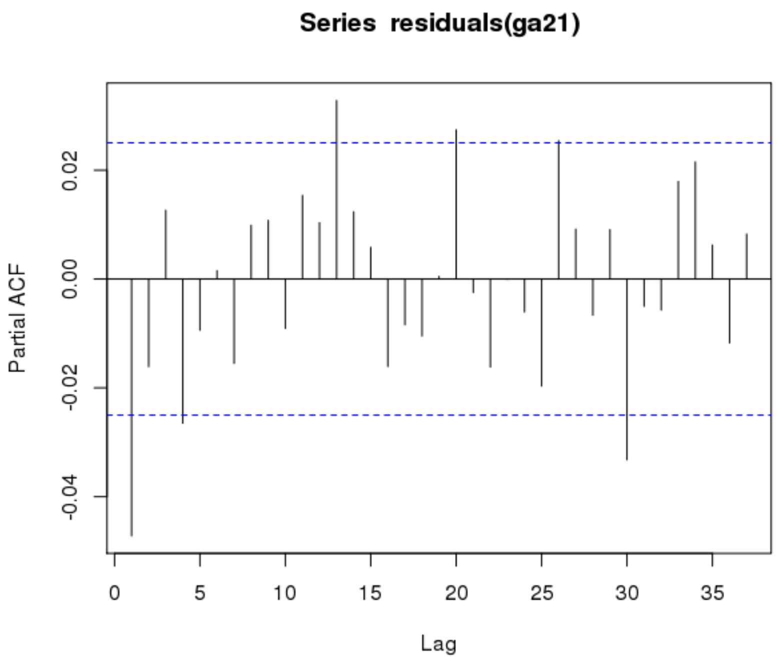


FIGURE 10 – Corrélation partielle des résidus

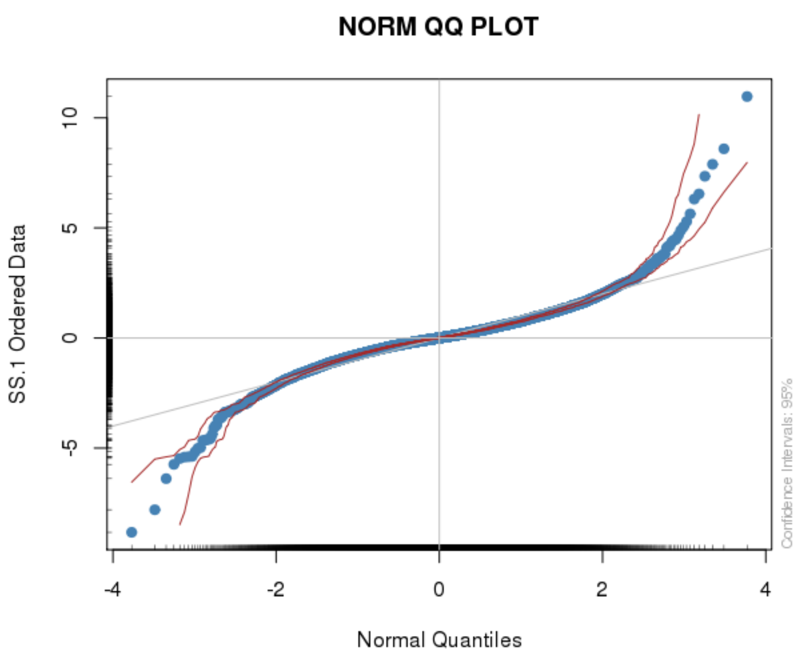


FIGURE 11 – Qqnorm des résidus

Ce modèle est peu performant, il explique très peu de variance, les résidus ne suivent pas une loi normale et ils ne sont pas indépendants on peut le voir avec les graphes 9 10 11. Une fois de plus le modèle est médiocre, ce qui est peu surprenant, si il marchait tout le monde le ferait.

6.2 Rentabilités hebdomadaires

Une fois de plus nous avons eu des problèmes de convergence avec le modèle de base et ceci pour tout les aglorithmes d'optimisation disponible. La plupart des modèles se valent, leur qualité est très faible, le R^2 est toujours très faible et les résidus sont souvent corrélés et loin d'une loi normale. Nous avons trouvé pour que le "meilleur modèle" est le modèle ARMA(1,1)+GARCH(1,1), nous trouvons un R^2 de 1.1% et un AIC $-3,92$. Ce modèle est celui qui maximise le R^2 et minimise l'AIC

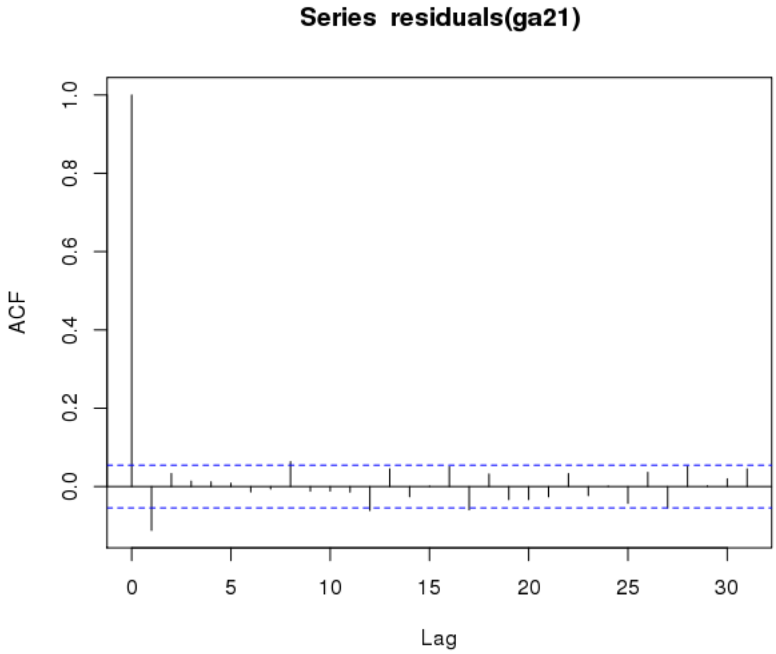


FIGURE 12 – Corrélation des résidus

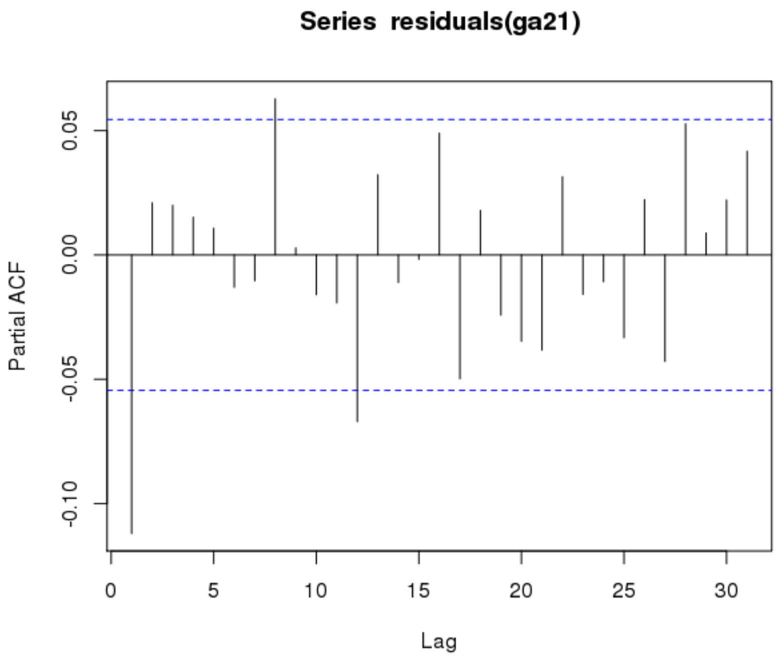


FIGURE 13 – Corrélation partielle des résidus

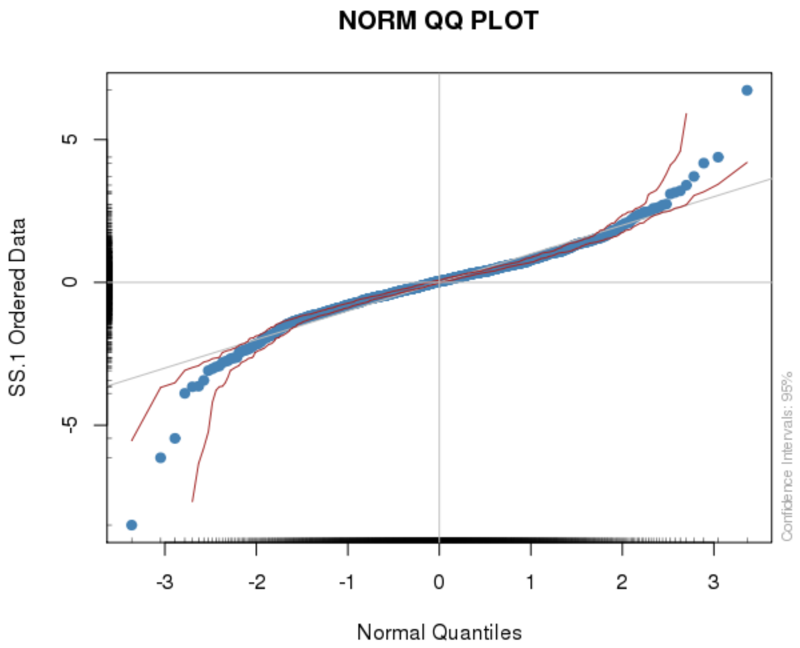


Figure 14 – Qqnorm des résidus

Ce modèle est peu performant, il explique très peu de variance. Les résidus sont assez proches d'une loi normale, la corrélation est aussi beaucoup plus faible que dans le cas journalier, on peut le voir avec les graphes 12 13 14. Une fois de plus le modèle est médiocre, ce qui est peu surprenant cependant il est mieux que dans le cas des rentabilités journalières.

6.3 Conclusion

Les modèles dans le cas des rentabilités journalières et hebdomadaires sont mauvais, il est difficile de trouver un modèle car la convergence des algorithmes restent difficile. Comme on pouvait s'y attendre les modèles sont médiocres, cependant le modèle pour les rentabilités hebdomadaires est meilleur que pour les rentabilités journalières, ce qui était prévisible puisque la distribution des rentabilités est plus proche d'une distribution de loi normale.

7 Question 7

Nous avons étudié l'écart type que donne le modèle ainsi que l'éspérance réalisée par rapport à l'intervalle de confiance que nous donne le modèle.

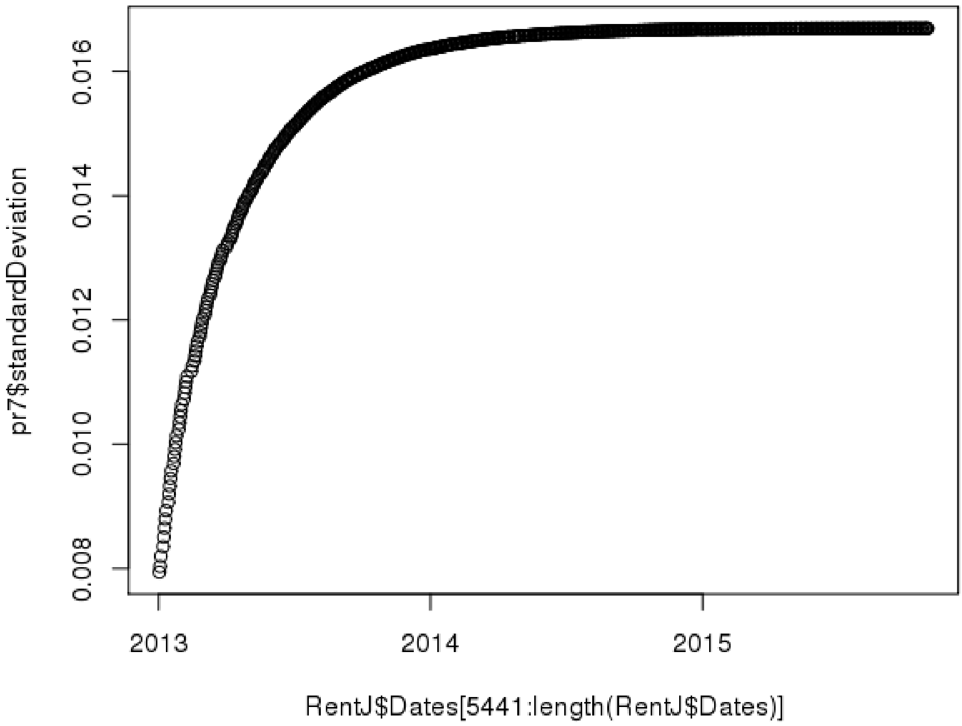


FIGURE 15 – Ecart type du modèle au fil du temps

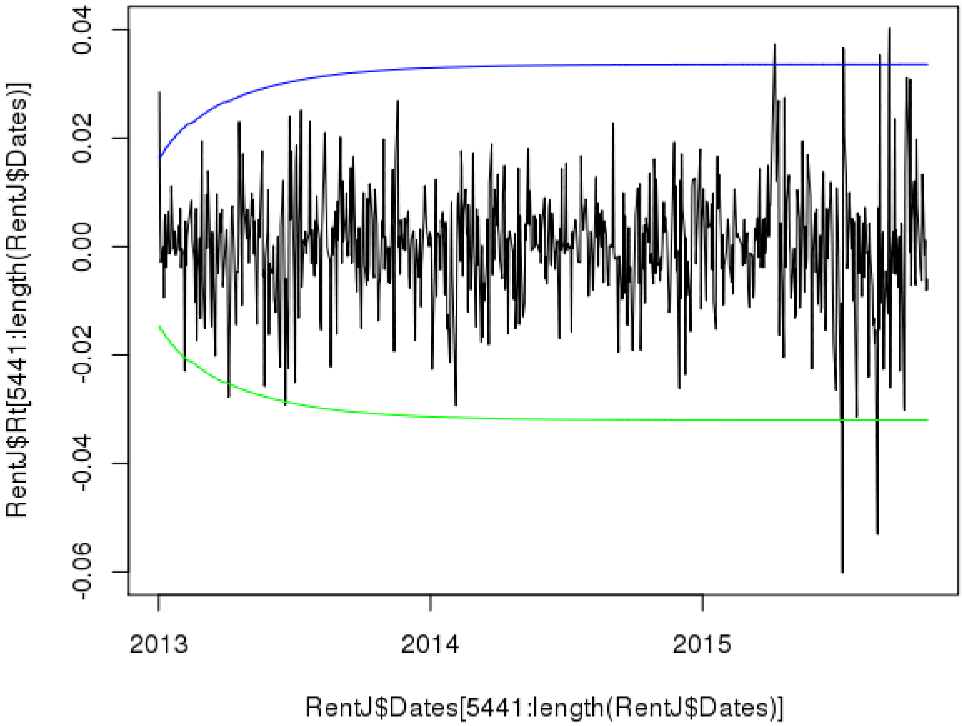


FIGURE 16 – Espèrance réalisées et intervalle de confiance à 95% du modèle

On remarque que les rentabilités réalisées sont en grande partie dans l'intervalle de confiance, on peut penser que le modèle est approprié. Cependant notre modèle n'explique que très peu de chose, le R^2 est extrêmement faible, cet intervalle de confiance est trompeur, notre modèle n'est que du bruit.

8 Question 8

Dans cette partie nous avons estimé les paramètres du modèle sur deux intervalles, le premier de 1991 à 2002 et le second de 2003 à 2015. Cependant les algorithmes d'optimisation ne convergent pas sur l'intervalle [2003,2015] ainsi les coefficients obtenus ne sont pas bons.

Les deux modèles ne peuvent à première vue être identiques, pour vérifier cela nous avons fait un **test de Kolmogorov-Smirnov** sur deux échantillons simulés avec les 2 modèles. La p-value est inférieure à $2 * 10^{-16}$ ainsi on peut rejeter l'hypothèse que les deux modèles sont identiques.

Cela montre que ce modèle n'est pas du tout stable, prédire des rentabilités journalières avec un modèle ARMA-GARCH n'est pas du tout efficace.

Nous avons effectué les mêmes tests pour les rentabilités hebdomadaires, les paramètres semblent plus proches à l'oeil cependant ceci est tromper. Nous obtenons les mêmes résultats pour le test de Kolmogorov. Les modèles sont assez instables. Nous aurions pu effectuer un test de Wald cependant nous n'avons pas accès à la matrice de variance covariance.

9 Conclusion

Nous avons testé différents modèles pour les rentabilités journalières et hebdomadaires de l'indice Hang Seng, cependant il a été difficile de trouver un modèle correct pour les rentabilités journalières, tout les modèles avaient une faible qualité et en utilisant le rapport signal/bruit (avec le R^2) on en déduit que l'espèrance de gain en voulant utiliser ces modèles est faible. Dans ce TP nous avons mis en évidence que les hypothèses standards ne sont pas tenables en pratique, les processus de rentabilité ont une mémoire longue et plus l'intervalle de temps entre les rentabilités est faible plus cela est vrai. Les modèles ne sont pas réellement approprié pour prédire les rentabilités futures notamment à cause de l'instabilité des modèles au fil du temps, comme cela a été montré en quesion 4. Une solution serait d'entrainer nos modèles sur des intervalles plus cours avec des donnés plus nombreuses, par exemple avec des prix séparés par des ticks de 1 min au lieu de 1 jour.