Rapport - Séries Temporelles

Philippe Real

10 janvier, 2020

Abstract

This is my abstract.

Contents

1 Partie I - Exemple de modélisation appliqué au traffic voyageur

1.1 Lecture des données et premières analyses de la série temporelle

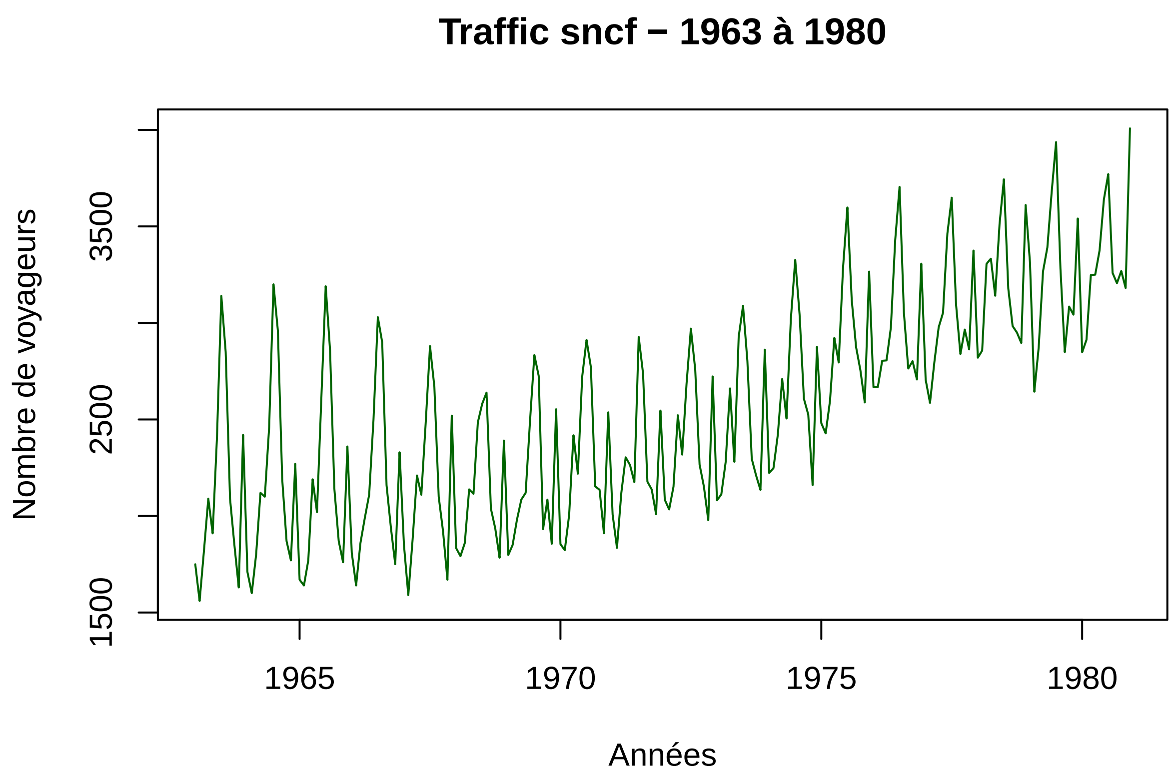
1.1.1 Lecture des données

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1963 1750 1560 1820 2090 1910 2410 3140 2850 2090 1850 1630 2420
## 1964 1710 1600 1800 2120 2100 2460 3200 2960 2190 1870 1770 2270
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1560    2098    2531    2547    2934    4008
```

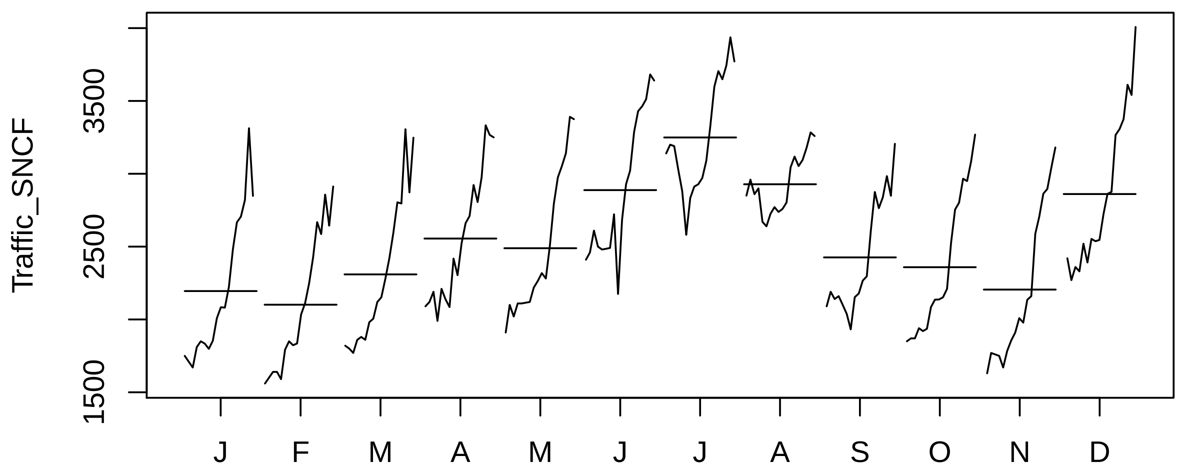
1.1.2 Chronogramme de la séries temporelles - sncf

On a 4 séries temporelles possibles en fonction du choix de la quantité observée (High, Low, Open, Close, Volume). On va s'intéresser à la valeur à la fermeture pour la cotation de l'indice CAC40 (Close).

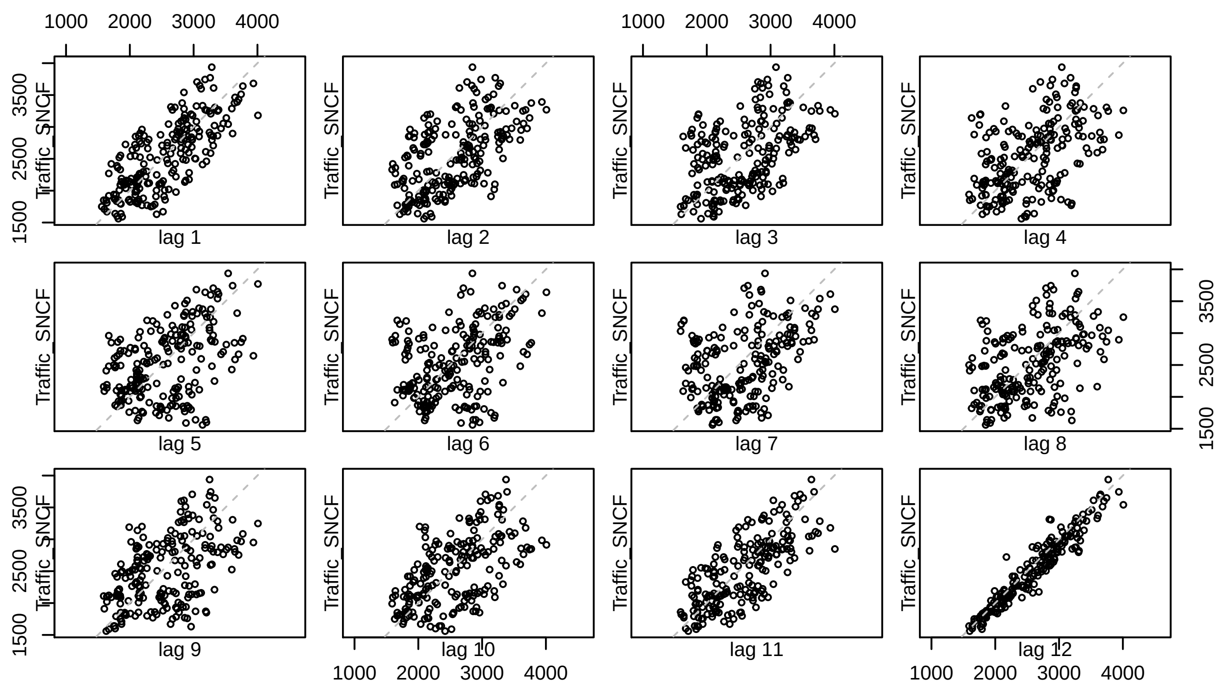


1.1.3 Représentations graphiques : month-plot et lag-plot

Si le diagramme retardée suggére une corrélation entre les deux séries, on dit que la série présente une autocorrélation d'ordre k.Ce diagramme permet de comprendre la dépendance de la série par rapport à son passée. Il donne une vision locale de la série, si y a une corrélation entre la série a un instant et la série 1, 2... instants avant.



Les tracés du chronogramme et du diagramme par mois montrent un motif saisonnier global avec une tendance à l'augmentation du nombre du traffic en juillet août ainsi que décembre.



Le lag plot indique une saisonnalité de 1 an (période T=12 mois) marquée.

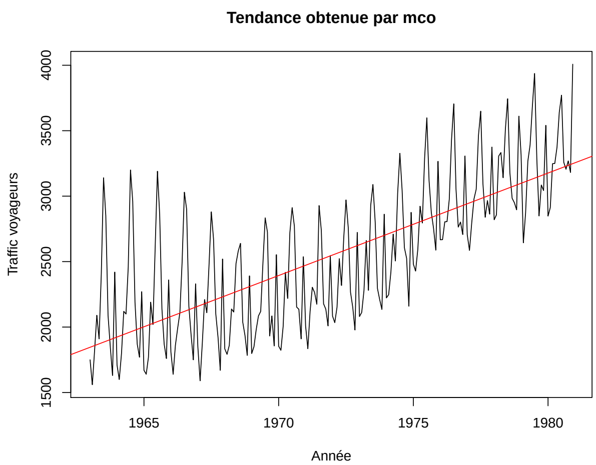
1.1.4 Tendance et saisonnalité

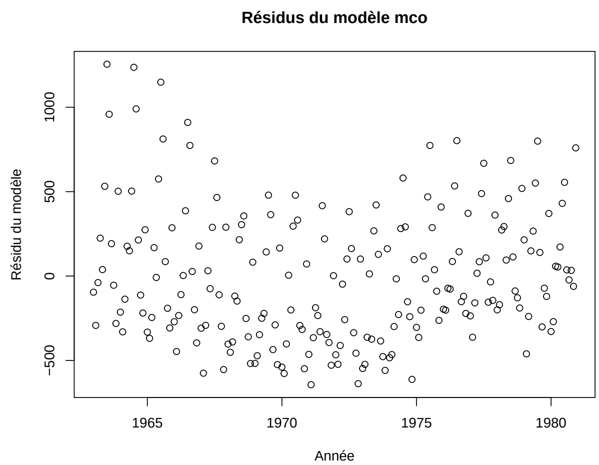
On cherche ici à analyser la série et à déterminer une tendance (allure moyenne) aisni qu'un comportement périodique ou saisonnalité aisni que des variations exeptionnelles, qu'il faut alors expliquer.

- Estimation de la tendance par moindre carré ordinaire

On suppose que la série est de la forme $X_t = m_t + z_t$ avec $m_t = \beta_0^* + \beta_1^* t$ et z_t l'ereur ou résidus. On cherche à estimer par l'estimateur des moindres carrés, les paramètres β_0^* et β_1^* à partir de la série des observations.

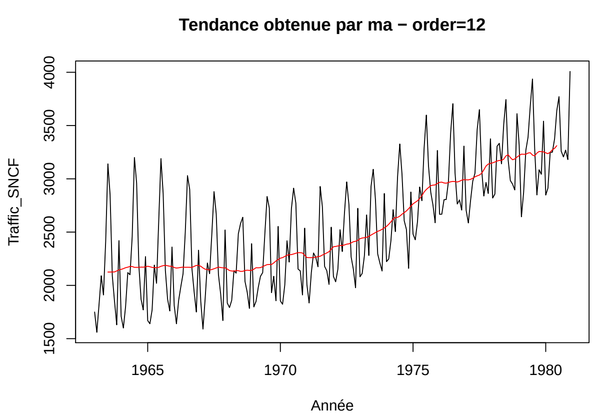
La tendance obtenu est une droite, la droite de régression par mco.

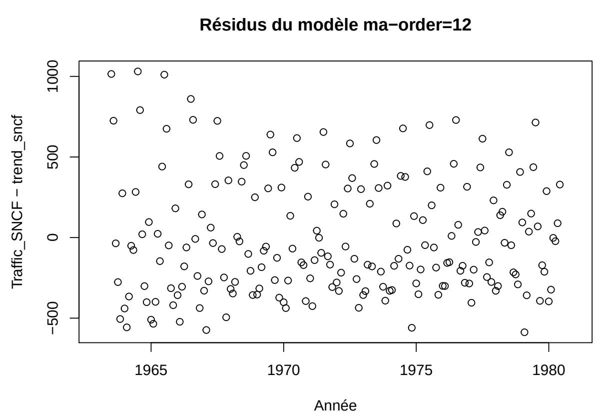




- Estimation de la tendance par moyennes mobiles

On cherche ici à ajuster un modèle à la courbe observée et parmis les nombreuses méthodes statistiques disponibles (ondelettes, noyaux, splines. . .) on va utiliser la méthode des moyennes mobiles qui est bien adaptée aux série temporelles. Pour celà on utilise la fonction "ma" du package "forcast" de R.
On a remarqué une saisonnalité de 12 mois (1 an) on efffectue ici une moyenne mobile d'ordre 12 pour obtenir la tendance (ma avec le paramètre order=12).

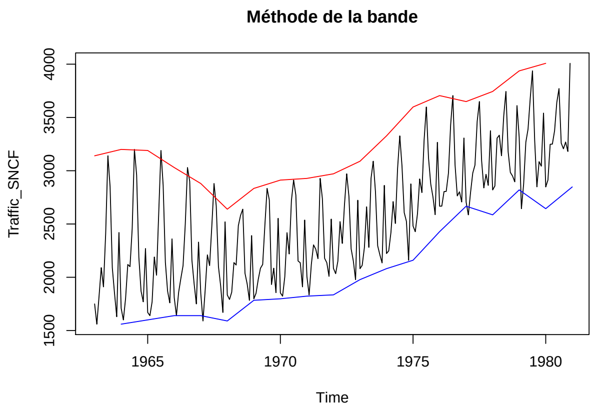


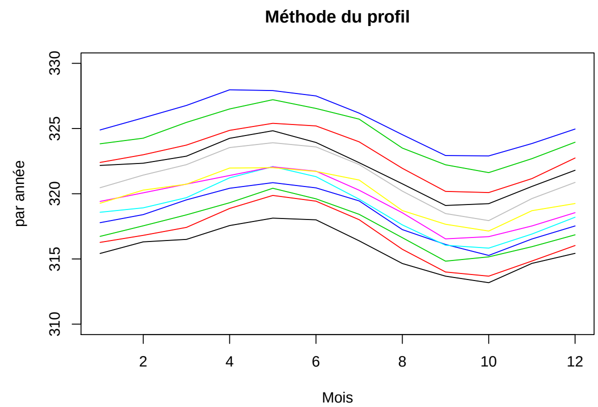


L'ajustement à l'évolution globale de la courbe est meilleur mais les résidus ont peu évolués et toujours aussi peu centrés en 0.

- Série décomposée - Tendance, Saisonnalité, Résidus

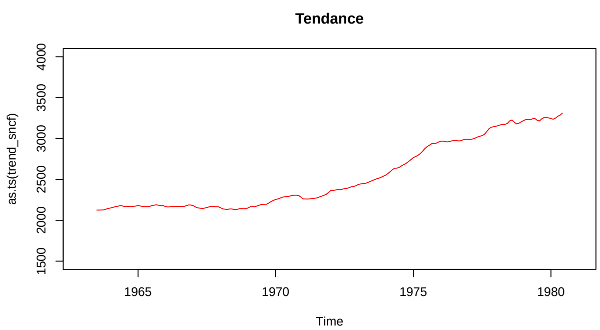
Avant d'établir les différentes coposantes de la série, on va déterminer la catégorie du modèle : additif ou multiplicatif. Pour savoir quel modèle est le plus adapté entre additif et multiplicatif on peut utiliser la méthode de la bande ou du profil. Dans la méthode de la bande on regarde si les 2 droites sont à peu près parrallèles et dans la méthode du profil si les c'est le cas pour les différentes courbes on conclut alors à un modèle est additif. Et multiplicatif dans le cas contraire.

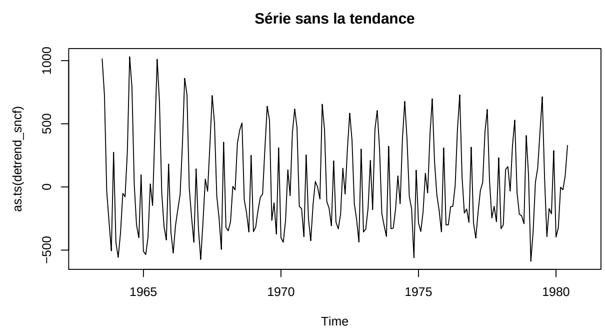


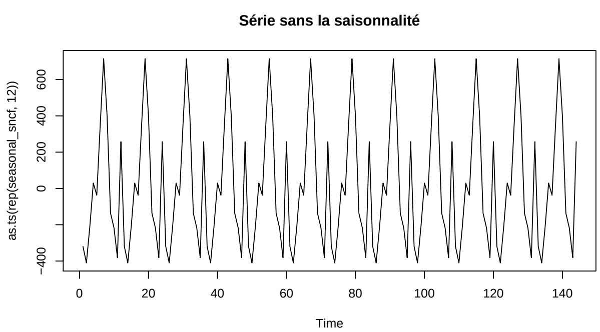


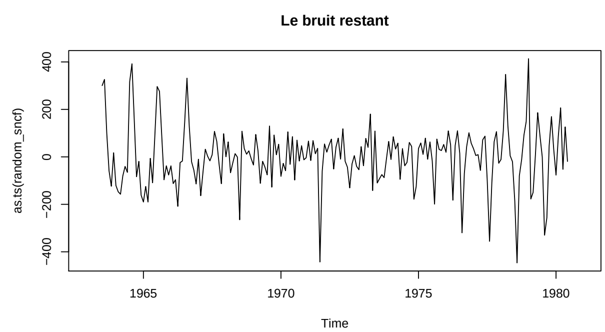
Dans notre cas la méthode de la bande indiquerait un modèle additif à partir de l'année 1969/1970. La méthode du profil plaide aussi plutôt pour un modèle additif.

On va donc utiliser un modèle additif, c'est à dire que l'on va décomposer la série sous la forme $X_t = m_t + s_t + z_t$ avec m_t : La tendance (orientation à long terme), s_t : La saisonnalité (phénomène, composante périodique ou saisonnière) et z_t : L'erreur ou résidu, dont la variation doit être faible par rapport aux 2 autres.



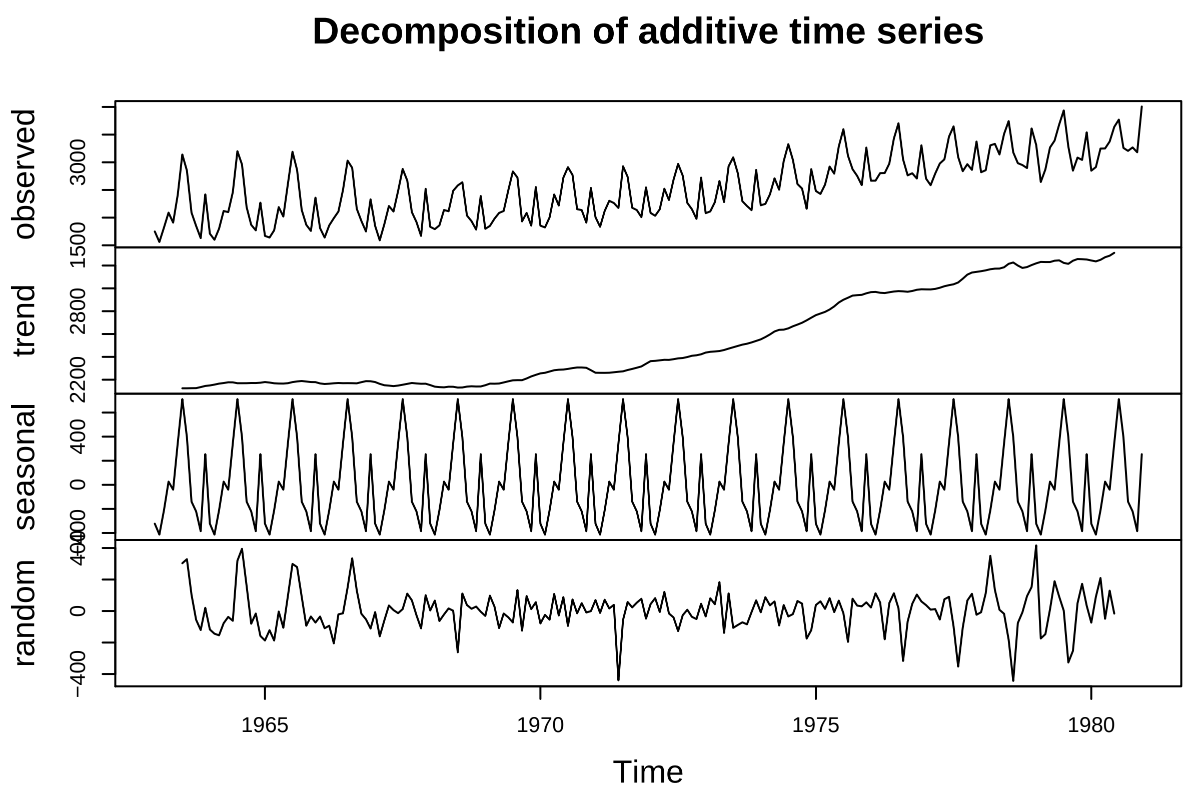






- Decomposition des séries temporelles avec la fonction *decompose* de R

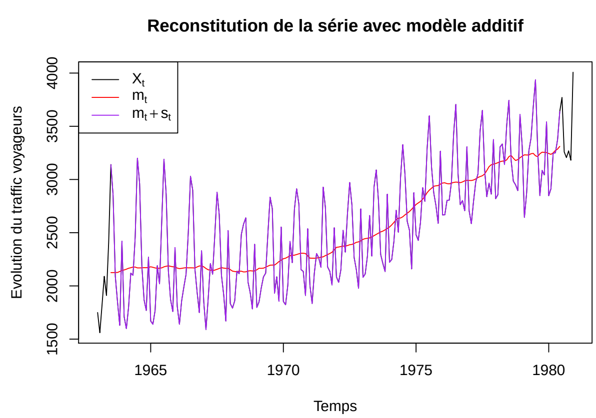
On va décomposer la série temporelles en utilisant la fonction décompose de R de façon à avoir une idée générale de la tendance (trend) saisonalité et bruit. On remarque que les graphiques obtenus sont très similaire avec ceux obtenus précédement.

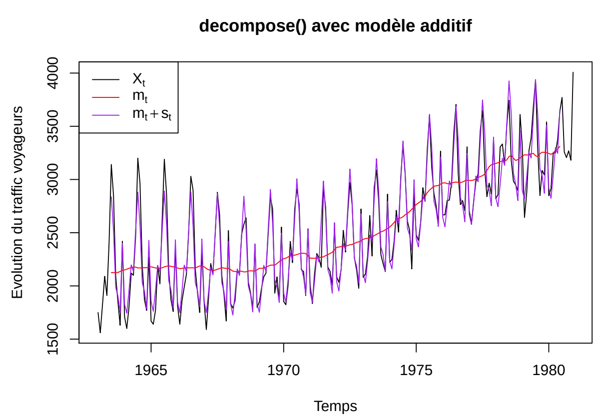


La tendance est nette, on a aussi une saisonnalité qui semble marquée. Par contre le bruit présente une structure. La modélisation doit être améliorée. La fonction decompose en modèle multiplicatif n'apporte pas d'amélioration au niveau de la distribution des résidus, qui semble toujours dépendre du temps.

- Reconstitution de la série

A partir des différentes composantes calculées précédemment, on peut recontruire la série.





Pour évaluer la performance de prédiction, nous allons estimer les paramètres du modèle sur la série allant de janvier 1962 jusqu'à décembre 1979 et garder les observations de l'année 1980 pour les comparer avec les prévisions.

1.2 Prévision par lissage exponentiel

On obtient les différents lissages à partir de la fonction ets de R.

Le lissage exponentiel simple (ANN) est obtenu à partir du paramètre model=ANN où : La première lettre A de model="ANN" signifie que l'erreur est additive. La deuxième lettre concerne la tendance, N indique qu'il n'y en a pas. La troisième lettre concerne la saisonnalitée, N indique qu'il n'y en a pas.

Dans le lissage exponentiel double model=AAN on considère une tendance additive. Et pour le lissage exponentiel triple ou de Holt-Winters on considére qu'il y a en plus une saisonnalité additive model=AAA

On peut comparer ces méthodes de lissage, en terme de critères AIC, AICc et BIC à partir du tableau suivant :

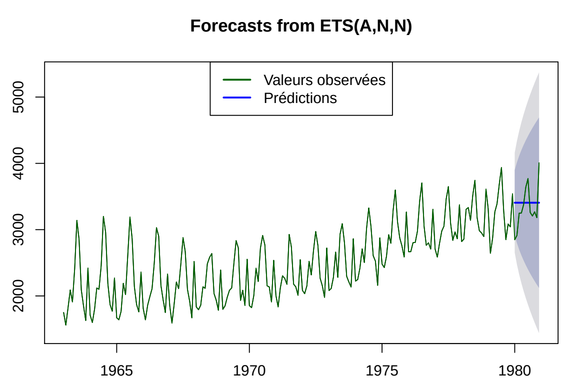
Type de lissage Exponentiel	AIC	AICc	BIC
Lissage Simple (A,N,N)	3517.8041845	3517.9241845	3527.7585445
Lissage Double (A,A,N)	3522.0129751	3522.3160054	3538.6035751
Holt-Winters (A,A,A)	3093.312925	3096.6032476	3149.7209649

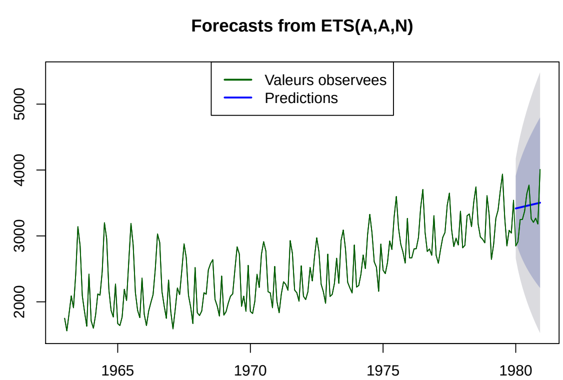
Nous remarquons que le modele est trop basique dans notre cas pour predire a l'horizon 12 et que l'intervalle de confiance a 80%, bien que tres large, ne contient pas toutes les vraies valeurs de la serie. On peut utilser predict(fitLES,12) prediction a horizon 1 an.

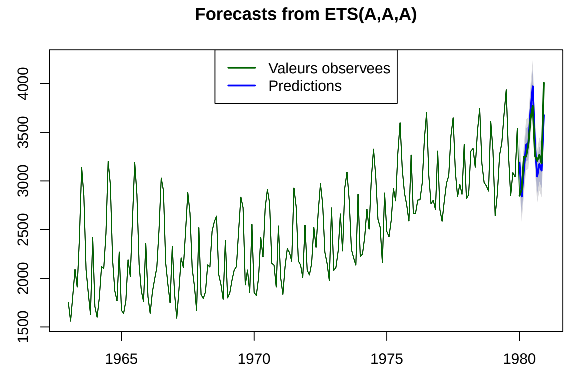
On remarque que le lissage double n'apporte pas vraiment d'amélioration par raport au simple. Les critères AIC et BIC sont plus élevés dans le cas du lissage double par rapport au lissage simple. Le lissage exponentiel triple ou de Holt-Winters apporte une amélioration.

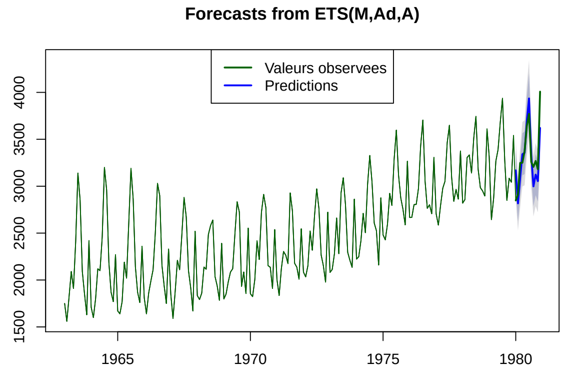
On peut considérer d'autres méthodes de lissage en jouant sur le caractères Additif/Multiplicatif des composantes. On a vu précédemment, que la tendance était plutôt additive, on va donc fixer la tendance=A et faire varier les autres paramètres.

Modèle	AIC	AICc	BIC
(M,A,A)	3076.9495014	3080.6467987	3136.6756613
(M,A,M)	3059.8613864	3063.151709	3116.2694263







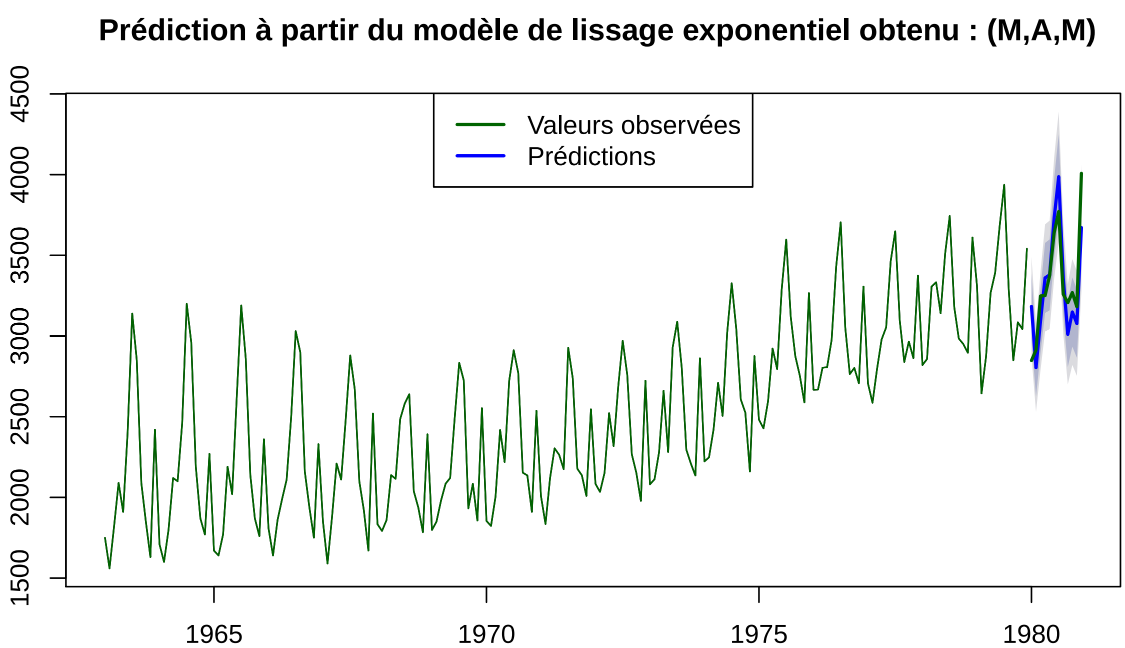


- Procédure automatique - modèles ajustés par ets.

La fonction ets permet aussi un ajustement automatique du modéle, lorsqu'aucun modèle n'est spécifié.

Le modèle sélectionné est le modèle avec tendance additive et avec erreur et saisonnalité multiplicatives (M,A,M). C'est aussi ce modèle qui minimise le critère AIC, AICc et BIC. En effet, en spécifiant le critère à minimiser: AIC, AICc et BIC avec le paramètre ic="aic"/"aicc" ou bic" de la fonction ets, on obtient toujours le même modèle.

Modèle	AIC	AICc	BIC
(M,A,M)	3059.8613864	3063.151709	3116.2694263



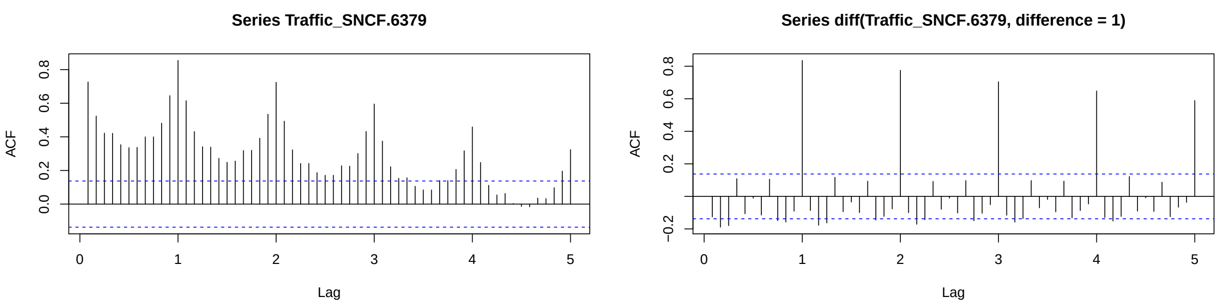
1.3 Modélisation

On cherche ici à modéliser la série par un processus stationnaire ARMA(p,q) ou bien SARMA(p,q). Si besoin on cherchera à stationnariser la série en utilisant l'opérateur de différentiation. On obtiendra alors une modélisation à partir de processus ARIMA (p,d,q) ou SARIMA(p,d,q)

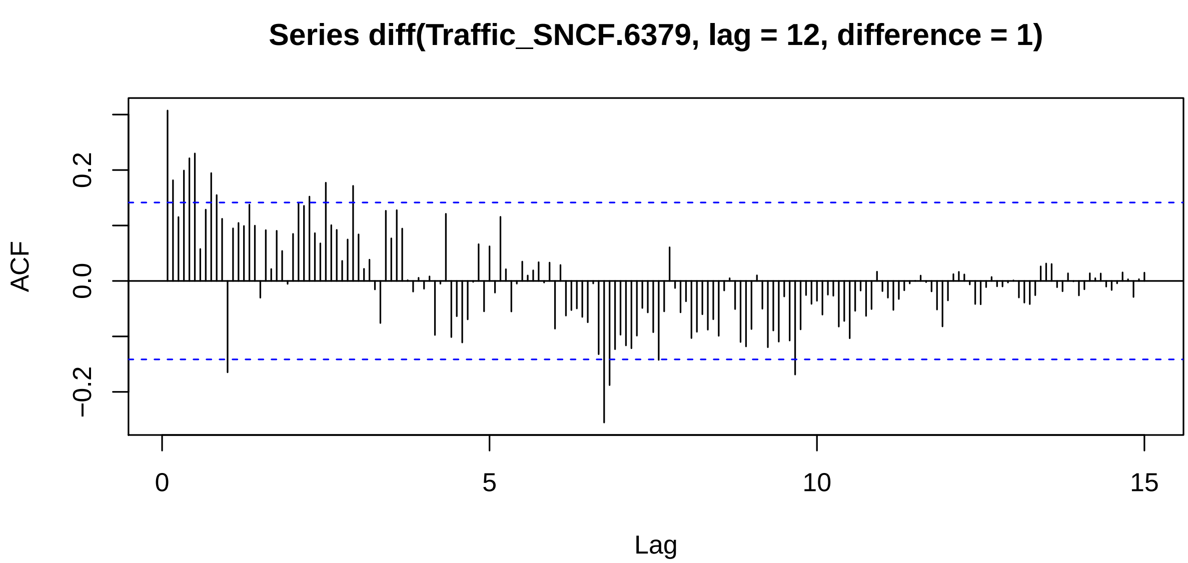
1.3.1 Identification du modèle

La première étape est l'étude de la stationnarité du processus régissant la série. Pour identifier le modèle on commence par une étude de la stationarité en traçant le corrélogramme de la série, la valeur de $\rho_X(h)$ en fonction de h. On va voir que l'on observe une périodicité annuelle, lorsque h = 12 dans le cas ici de données mensuelles. Pour mettre en évidence ce phénomène, on trace le corrélogramme de la série et de la série différentiée.

- Corrélogramme de X_t et de $(1 - B)X_t$

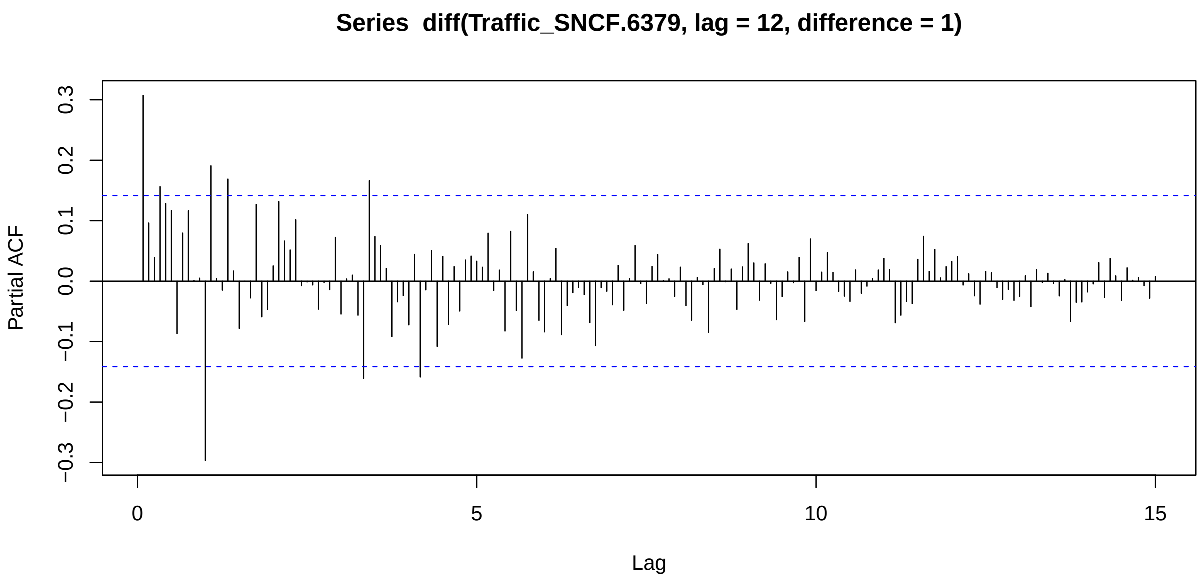


La fonction d'autocorrélation estimée est positive. On remarque une périodicité de 1 (12 mois) (graphique de gauche).On peut essayer de différentier la série au moins une fois (graphique de droite). On remarque des autocorrélations importantes pour les valeurs de h de 1 période (année), tout les 12 mois. C'est aussi ce que l'on avait remarqué précedemment avec le lag plot. On va donc appliquer l'opérateur $(1 - B^{12})$ à la série précédente, transformée par diffrentiation : $(1 - B)X_t$. Et lon trace le corrélogramme associé.



Le corrélogramme de la série obtenue par différentiation: $(1 - B)(1 - B^{12})X_t$ ne présente plus de fortes amplitudes pour les petites valeurs de h. Ni pour h multiple de 12 comme c'était le cas pour la série brute. On peut considérer que la série ainsi transformé est issue d'un processus stationnaire. Il y a encore cependant encore d'assez fortes valeurs pour $\hat{\rho}(1)$ ce qui indique d'ajouter un terme dans la partie MA du modèle.

On peut regarder l'autocorrélation partielle pour avoir une idée du terme degrès q du terme moyenne mobile MA(q) du modèle.

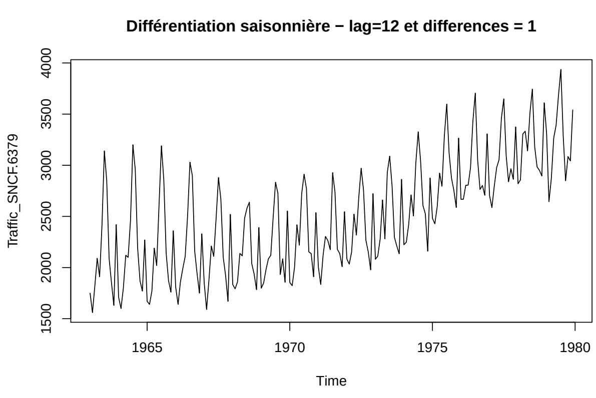


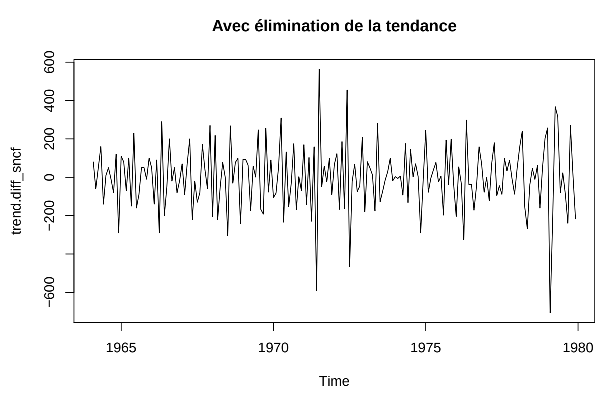
L'autocorrélation partielle suggère un terme d'ordre q=1 (12ème mois) soit un terme moyenne mobile du type : $(1-\theta_1 B)(1-\theta_2 B^{12})\epsilon_t)$

On obtient ainsi un modèle du type SARIMA(0,1,1)(0,1,1)

$(1-B)(1-B^{12})X_t = (1-\theta_1 B)(1-\theta_2 B^{12})\epsilon_t$ où $E\epsilon_t = 0$ et $V\epsilon_t = \sigma^2$

- Elimination de la tendance





- Validation du modèle obtenu par différentiation saisonnière

On confirme cette hypothèse à l'aide d'un test de $Dickey - Fuller$ on obtient une p_value=0.0198766 On a donc réussit à amélioré la staionnarité de la série avec une différentiantion saisonnière. Par contre le modèle différentié et sans tendance obtenu à une p_value=0.01 et n'est donc pas stationnaire.

Et le test du Portmanteau ou test de blancheur sur R n'est pas concluant et donne une p_value=$5.7890803 \times 10^{-11}$ pour le modèle différentié (différence=1, lag=12) et une p_value=$5.7890803 \times 10^{-11}$ pour le modèle auquel ont a enlevé en plus la tendance.

On va maintenant estimer plusieurs modèles SARIMA(p,1,q)(r,1,s) en faisant varier les paramètres p et q et la saisonnalité (r,1,s). On fera ensuite un choix de modèles en se basant sur les critères AIC, AICc, et BIC. Pour celà on estime ici le modèle de manière automatique en utilisant la fonction *auto.arima* de R. On regarde la trace et on sélectionne les meilleurs modèles.

```
autoarima63<-auto.arima(Traffic_SNCF.6379,trace=TRUE,allowdrift=FALSE)
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,1,2)(1,1,1)[12]                    : 2260.634
##  ARIMA(0,1,0)(0,1,0)[12]                    : 2369.937
##  ARIMA(1,1,0)(1,1,0)[12]                    : 2306.258
##  ARIMA(0,1,1)(0,1,1)[12]                    : 2254.534
##  ARIMA(0,1,1)(0,1,0)[12]                    : 2291.471
##  ARIMA(0,1,1)(1,1,1)[12]                    : 2261.867
##  ARIMA(0,1,1)(0,1,2)[12]                    : 2255.756
##  ARIMA(0,1,1)(1,1,0)[12]                    : 2265.126
##  ARIMA(0,1,1)(1,1,2)[12]                    : 2263.955
##  ARIMA(0,1,0)(0,1,1)[12]                    : 2321.209
##  ARIMA(1,1,1)(0,1,1)[12]                    : 2250.587
##  ARIMA(1,1,1)(0,1,0)[12]                    : 2288.057
##  ARIMA(1,1,1)(1,1,1)[12]                    : 2259.635
##  ARIMA(1,1,1)(0,1,2)[12]                    : 2252.119
##  ARIMA(1,1,1)(1,1,0)[12]                    : 2262.136
##  ARIMA(1,1,1)(1,1,2)[12]                    : 2259.435
##  ARIMA(1,1,0)(0,1,1)[12]                    : 2294.803
##  ARIMA(2,1,1)(0,1,1)[12]                    : 2254.112
##  ARIMA(1,1,2)(0,1,1)[12]                    : 2252.357
##  ARIMA(0,1,2)(0,1,1)[12]                    : 2250.829
##  ARIMA(2,1,0)(0,1,1)[12]                    : 2287.066
##  ARIMA(2,1,2)(0,1,1)[12]                    : 2254.18
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(1,1,1)(0,1,1)[12]                    : 2386.211
##
##  Best model: ARIMA(1,1,1)(0,1,1)[12]
```

SARIMA	AIC	AICc	BIC	sigma2	Log-Likelihood
010_010	2503.2666857	2503.2878498	2506.5189592	2.8503211×10^4	-1250.6333429
111_011	2385.9962789	2386.2113327	2399.0053726	1.4801476×10^4	-1188.9981395
012_011	2386.2057925	2386.4208463	2399.2148862	1.4821928×10^4	-1189.1028962
111_012	2387.4692901	2387.7936144	2403.7306572	1.4835149×10^4	-1188.734645
112_011	2387.9550467	2388.2793711	2404.2164139	1.4877499×10^4	-1188.9775234
212_011	2388.5219013	2388.9784231	2408.0355419	1.4839547×10^4	-1188.2609507
211_011	2387.913895	2388.2382193	2404.1752622	1.487462×10^4	-1188.9569475

En terme de minimisation des critères AIC, AICc et BIC les 6 meilleurs modèles sont les modèles : SARIMA(1,1,1)(0,1,1), SARIMA(0,1,2)(0,1,1), SARIMA(1,1,1)(0,1,2), SARIMA(1,1,2)(0,1,1), SARIMA(2,1,2)(0,1,1), SARIMA(2,1,1)(0,1,1) et le modèle initial SARIMA(0,1,1)(0,1,1) que l'on conserve pour l'analyser.

1.3.2 Validation des modèles SARIMA obtenus

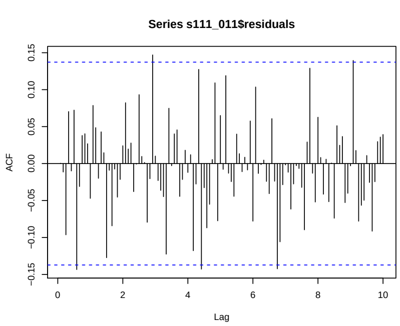
Avant de passer à la prédiction, on va maintenant valider ou invalider les modèles obtenus.

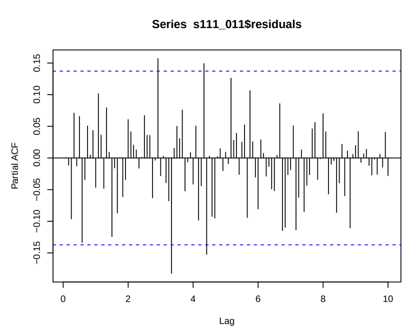
- Test de Box-Pierce

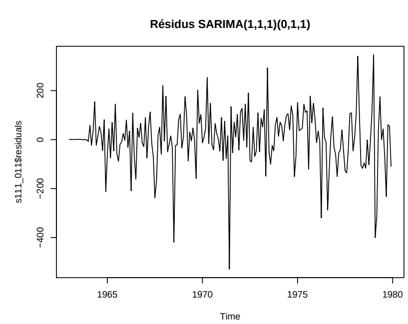
Le test de blancheur des résidus rejette nettement le modèles SARIMA(0,1,1)(0,1,1) avec une p-value $<$ 2.2e-16. Pour les autres modèles on accepte la blancheur des résidus comme le montre le tableau ci-dessous.

SARIMA	p-value
111_011	0.6762469
012_011	0.6436838
111_012	0.667892
112_011	0.6724191
212_011	0.7596604
211_011	0.6669889

- ACF et PACF des résidus







- Statistiques - qualité d'estimation des coefficients pour le modèle SARIMA(1,1,1)(0,1,1) et corrélations

```
##              ar1       ma1      sma1
## t.stat 2.441972 -23.74509 -7.446865
## p.val  0.014607   0.00000  0.000000
```

```
##             ar1        ma1       sma1
## ar1   1.0000000 -0.4852660 -0.1182038
## ma1  -0.4852660  1.0000000 -0.0928834
## sma1 -0.1182038 -0.0928834  1.0000000
```

Les résultats pour les autres modèles sont en annexe-partie I en fin de document.

1.4 Estmation d'un modèle à partir des données tronquées

On va ici restreindre les données pour la modélisation et ne considérer que les données qu'à partir de 1970.

On estime à nouveau le modèle de manière automatique en utilisant la fonction *auto.arima* de R.

```
autoarima70<-auto.arima(Traffic_SNCF.7079,trace=TRUE,allowdrift=FALSE)
```

```
##
##  ARIMA(2,0,2)(1,1,1)[12]                    : Inf
##  ARIMA(0,0,0)(0,1,0)[12]                    : 1440.493
##  ARIMA(1,0,0)(1,1,0)[12]                    : 1403.111
##  ARIMA(0,0,1)(0,1,1)[12]                    : 1421.316
##  ARIMA(1,0,0)(0,1,0)[12]                    : 1408.003
##  ARIMA(1,0,0)(2,1,0)[12]                    : 1404.244
##  ARIMA(1,0,0)(1,1,1)[12]                    : 1404.725
##  ARIMA(1,0,0)(0,1,1)[12]                    : 1403.201
##  ARIMA(1,0,0)(2,1,1)[12]                    : 1406.318
##  ARIMA(0,0,0)(1,1,0)[12]                    : 1439.439
##  ARIMA(2,0,0)(1,1,0)[12]                    : 1398.447
##  ARIMA(2,0,0)(0,1,0)[12]                    : 1407.986
##  ARIMA(2,0,0)(2,1,0)[12]                    : 1398.189
##  ARIMA(2,0,0)(2,1,1)[12]                    : 1400.43
##  ARIMA(2,0,0)(1,1,1)[12]                    : 1398.873
##  ARIMA(3,0,0)(2,1,0)[12]                    : 1395.87
##  ARIMA(3,0,0)(1,1,0)[12]                    : 1397.707
##  ARIMA(3,0,0)(2,1,1)[12]                    : 1398.158
##  ARIMA(3,0,0)(1,1,1)[12]                    : 1396.991
##  ARIMA(4,0,0)(2,1,0)[12]                    : 1383.709
##  ARIMA(4,0,0)(1,1,0)[12]                    : 1385.988
##  ARIMA(4,0,0)(2,1,1)[12]                    : 1385.999
##  ARIMA(4,0,0)(1,1,1)[12]                    : 1384.18
##  ARIMA(5,0,0)(2,1,0)[12]                    : 1385.196
##  ARIMA(4,0,1)(2,1,0)[12]                    : 1382.473
##  ARIMA(4,0,1)(1,1,0)[12]                    : 1383.819
##  ARIMA(4,0,1)(2,1,1)[12]                    : Inf
##  ARIMA(4,0,1)(1,1,1)[12]                    : Inf
##  ARIMA(3,0,1)(2,1,0)[12]                    : Inf
##  ARIMA(5,0,1)(2,1,0)[12]                    : 1388.351
##  ARIMA(4,0,2)(2,1,0)[12]                    : 1382.513
##  ARIMA(3,0,2)(2,1,0)[12]                    : Inf
##  ARIMA(5,0,2)(2,1,0)[12]                    : Inf
##
##  Best model: ARIMA(4,0,1)(2,1,0)[12]
```

Ici les meilleurs modèles sont très différents, et leur degrès pour la partie autoregressive est plus élevé. Les modèles ARIMA(4,0,1)(2,1,0) et ARIMA(4,0,2)(2,1,0) sont les meilleurs modèles.

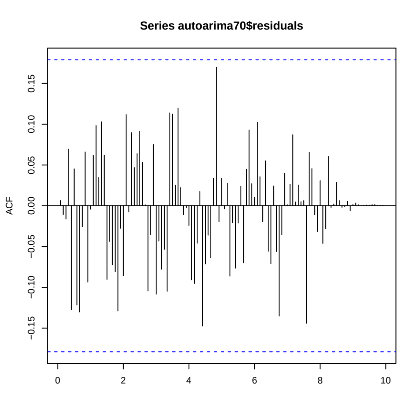
```
## Series: Traffic_SNCF.7079
## ARIMA(4,0,1)(2,1,0)[12]
##
## Coefficients:
##          ar1      ar2      ar3     ar4      ma1     sar1     sar2
##       1.0248  -0.1697  -0.0521  0.1796  -0.7023  -0.5600  -0.2406
## s.e.  0.2275   0.1605   0.1381  0.1448   0.2286   0.1123   0.1228
##
## sigma^2 estimated as 18438:  log likelihood=-682.51
## AIC=1381.02   AICc=1382.47   BIC=1402.48
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 12.37287 124.5736 89.43233 0.4101417 3.280327 0.6142643
##                     ACF1
## Training set 0.006374917
```

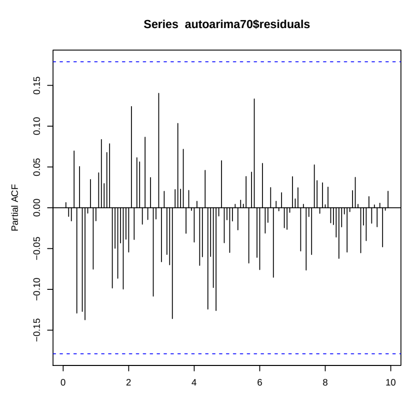
- Validation du modèle obtenu :

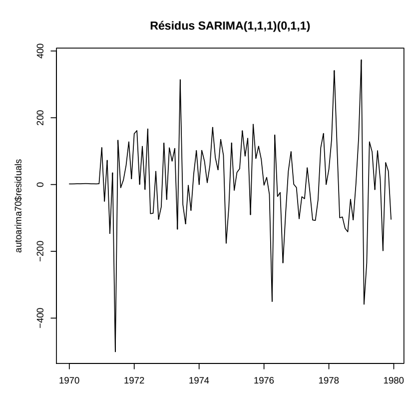
```
##             ar1       ar2       ar3      ar4       ma1      sar1      sar2
## t.stat 4.505142 -1.057059 -0.376882 1.240316 -3.071810 -4.985137 -1.959892
## p.val  0.000007  0.290484  0.706261 0.214859  0.002128  0.000001  0.050008
```

```
##                ar1         ar2         ar3         ar4         ma1
## ar1   1.0000000000 -0.73056684  0.08076506 -0.72393287 -0.90874996
## ar2  -0.7305668440  1.00000000 -0.52080014  0.47379303  0.51713037
## ar3   0.0807650550 -0.52080014  1.00000000 -0.49951003 -0.01891398
## ar4  -0.7239328698  0.47379303 -0.49951003  1.00000000  0.74990856
## ma1  -0.9087499638  0.51713037 -0.01891398  0.74990856  1.00000000
## sar1 -0.0004537662 -0.07668600  0.04376366  0.02562435  0.00827915
## sar2 -0.0718970323  0.07653393 -0.14255582  0.14183224  0.05380081
##               sar1        sar2
## ar1  -0.0004537662 -0.07189703
## ar2  -0.0766860004  0.07653393
## ar3   0.0437636601 -0.14255582
## ar4   0.0256243518  0.14183224
## ma1   0.0082791503  0.05380081
## sar1  1.0000000000  0.23710506
## sar2  0.2371050648  1.00000000
```

Le test de blancheur des résidus ou test Box-Pierce est accepté avec la p-value : 0.8498373



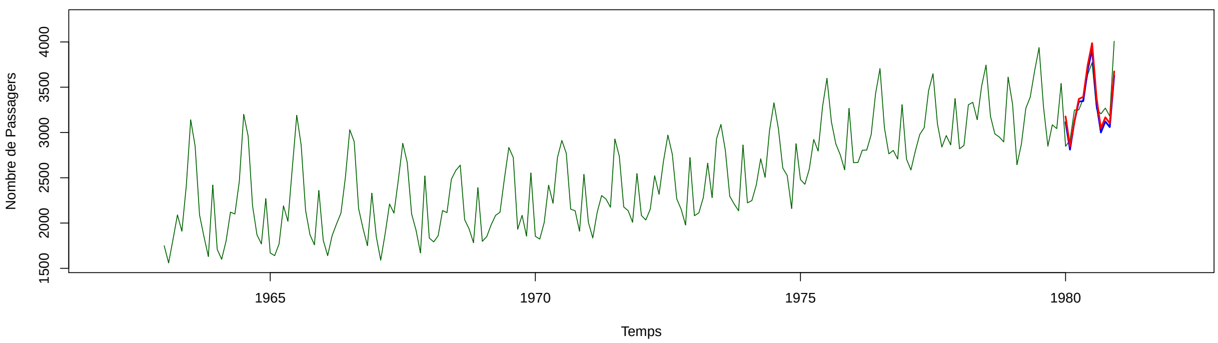


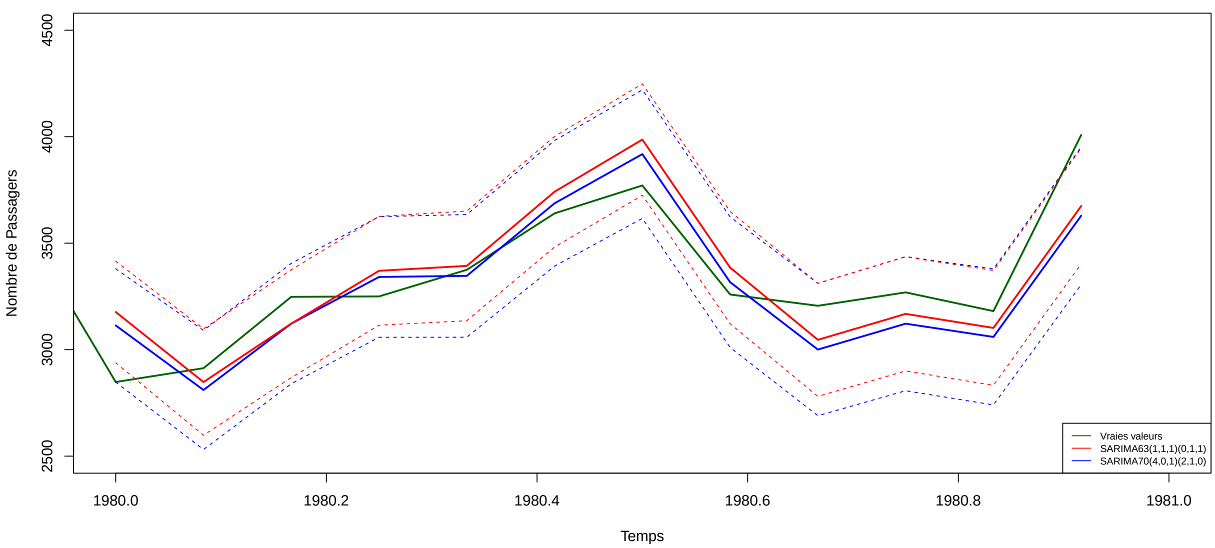


1.5 Prévisions et comparaison des modèles obtenus

Pour les prédictions ont ne va s'intéresser qu'aux 2 modèles obtenus SARIMA(1,1,1)(0,1,1) et SARIMA(4,0,1)(2,1,0) Les prédictions obtenus avec les autres modèles se révèelent être très proche du modéle SARIMA(1,1,1)(0,1,1).

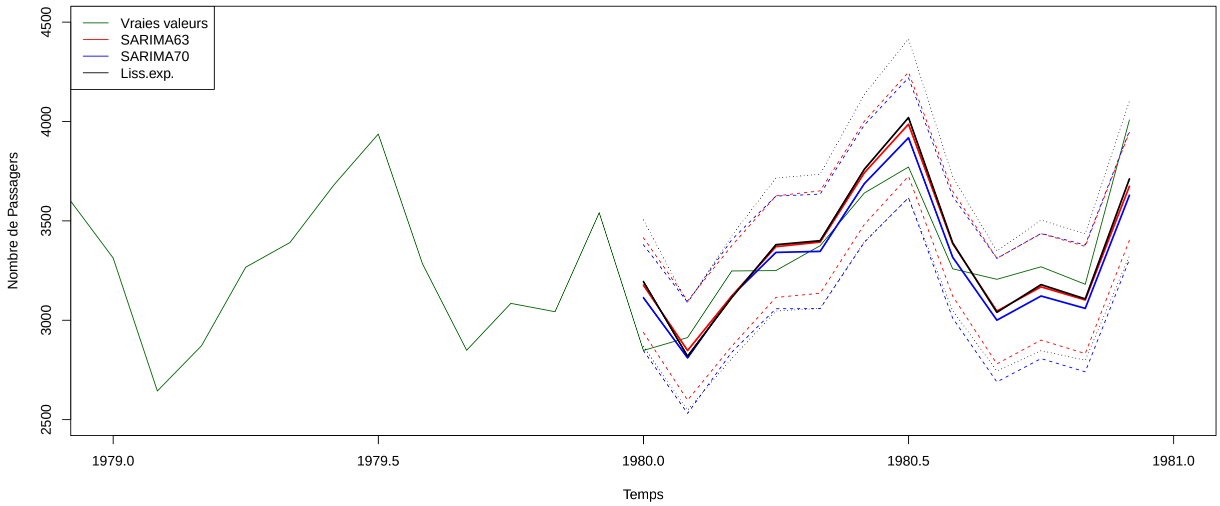
- Comparaison entre les différents modèles SARIMA(1,1,1)(0,1,1) et SARIMA(4,0,1)(2,1,0)





Les modèles sont quasiment confondus, excepté le SARIMA(0,1,1)(1,1,1) qui est très légèrement décalé. Ils suivent plutôt bien la courbe des données des vraies valeurs.

- Comparaison entre SARIMA et lissage exponentiel



Remarque : la stabilisation de la variance en utilisant la fonction log n'apporte pas l'atténuation souhaité, et est sans effet ici.

2 Partie II - Tentative de modélisation d'un indice boursier de type action à l'aide de processus ARIMA

2.1 Introduction

On cherche dans cette partie à modéliser par des processus de type ARIMA ou assimilé (SARIMA) l'évolution du prix d'indice boursier de type acion. En fait un poretefeuille d'actions, ou indice type CAC40, DAX, Eurostoxx, SP500… Pour cette première étude on va se baser sur l'indice CAC40. A parir de la modélisation (présupposée possible) obtenu on va chercher à prévoir l'évolution de l'indice en question, à horizon 3 mois, 6 mois voire 1 an.

Notre but est double :

- La définition de stress test de type action : A partir de la modélisation obtenue, et de l'intervalle de confiance sous jacent, on va chercher à déterminer une valeur de choc absolue à la hausse et à la baisse. Cette méthodologie de définition d'un choc absolue associé à un niveau de confiance devrait nous aider à définir un scénario économique plausible (avec un certain seuil de confiance) à horizon 3mois, 6mois et 1an. Ainsi cette modélisation devrait pouvoir nos guider dans la détermination de stress test de type action. Pour ête complet quant à la définition de stress test de type financier, il faudrait parvenir à définir une méthodologie équivalente pour les produits de types taux ou courbes de taux d'intérêt. Ce dernier cas est plus complexe dans la mesure où on cherche à modéliser une surface et les séries temporelles ne sont peut ête pas appropriées. Plus précisemment on cherche à modéliser un faisceau de courbes aléatoires qui dépendent les unes des autres. Le mécanisme de dépendance étant en parti connu, ou plutôt des modèles éxistent.
- Elaboration d'un portefeuille simplifié : Une fois les principaux indices modélisé, on va chercher à décomposer nos portefeuilles sur ces indices et ainsi constituer un portefeuille simplifié. Ce portefeuille simplifié serait la base d'un indice benchmark du portefeuille étudié.

Dans un premier temps on va étudier la série temporelle associée à l'évoultion du prix de l'indice étudié le CAC40 : représenttion graphique, saisonnalité, tendance, stationarité… Pour entrer dans le cadre d'un modèle ARMA(p,q), on va dans un premier temps, étudier la stationarité de notre série. Et la rendre stationnaire le cas échéant. A partir de là on cherchera à déterminer les paramètres p et q du processus auto régressif AR(p) et moyenne mobile MA(q) sous jacent à parir des graphiques ACF et PACF. Enfin on ajustera les coefficient pour obtenir notre modèle. On terminera l'étude en validant le modèle: blancheur des résidus, ndépendance, normalité. On pourra alors après validation l'utiliser pour nos prédictions.

2.2 Lecture des données et premières analyses

Les données ont été récupérées sur le site Yahoo Finance. Ticker "^FCHI" pour les données de l'indice CAC40. On considère un jeu de données quotidienne et un autre mensuel. Avec dans les 2 cas un historique de Janvier 1998 à Janvier 2020. A partir de cet historique de 22 ans on va construire différentes séries de profondeur d'historique différente. Après avoir annalysé ces séries on essaiera de construire un modèle de type ARIMA pour chacune d'elles.

2.2.1 Traitement des données

Dans le cas des données journalières, il y a des données manquantes. On va les supprimer. La variable Date est aussi connvertit en structure date.

```
##          Date           Open           High           Low
##  1997-11-12:   1   Min.   :2453   Min.   :2518   Min.   :2401
```

```
##  1997-11-13:   1   1st Qu.:3695   1st Qu.:3724   1st Qu.:3667
##  1997-11-14:   1   Median :4341   Median :4374   Median :4309
##  1997-11-17:   1   Mean   :4392   Mean   :4424   Mean   :4358
##  1997-11-18:   1   3rd Qu.:5106   3rd Qu.:5135   3rd Qu.:5074
##  1997-11-19:   1   Max.   :6929   Max.   :6945   Max.   :6839
##  (Other)   :5660   NA's   :54     NA's   :54     NA's   :54
##      Close         Adj.Close         Volume
##  Min.   :2403   Min.   :2403   Min.   :        0
##  1st Qu.:3697   1st Qu.:3697   1st Qu.:        0
##  Median :4341   Median :4341   Median : 90349500
##  Mean   :4392   Mean   :4392   Mean   : 83842758
##  3rd Qu.:5106   3rd Qu.:5106   3rd Qu.:129391650
##  Max.   :6922   Max.   :6922   Max.   :531247600
##  NA's   :54     NA's   :54     NA's   :54
```

```
##         Date   Open   High    Low  Close Adj.Close Volume
## 1 1997-11-12 2688.8 2701.0 2649.5 2694.5    2694.5      0
## 2 1997-11-13 2691.6 2712.2 2681.8 2700.7    2700.7      0
## 3 1997-11-14 2735.9 2751.4 2691.9 2698.9    2698.9      0
## 4 1997-11-17 2772.1 2779.6 2760.1 2773.0    2773.0      0
## 5 1997-11-18 2787.2 2793.6 2762.6 2782.6    2782.6      0
## 6 1997-11-19 2753.0 2792.3 2753.0 2790.6    2790.6      0
```

Dans le cas des données mensuelles ont n'a pas de problème de données manquantes.

2.2.2 conversion des données en objet *time series*

Ici on convertit les données en objet R ts (time series) Dans le cas des données journalières on utilise pour le parmaètre de fréquence (nb jours dans l'année) la valeur 256 Ce qui correspond au nombre de jours par an (jours ouvrès sans les jours de fermeture) que l'on obtient une fois les NA supprimés.(On remarque que cette valeur de fréquence influe la vitesse de traitement lors de l'appel de la fonction auto.arima)

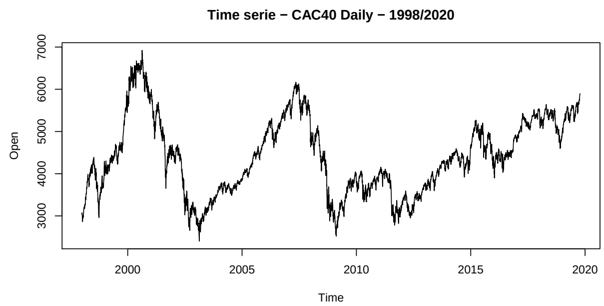
2.3 Analyse des séries temporelles obtenues

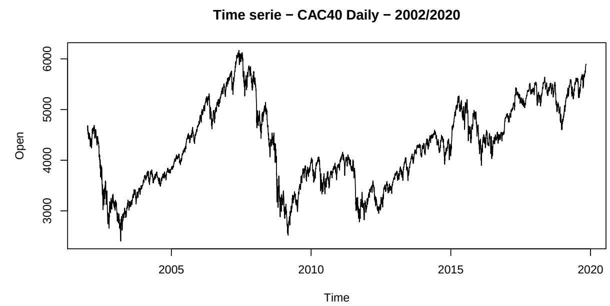
Comme déja énnoncé on va étudier plusieurs profondeur d'historique.

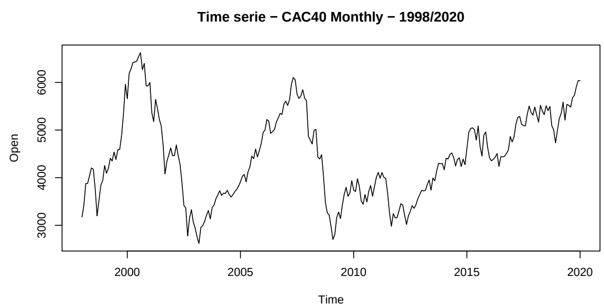
- Toute la série de janvier 1998 à janvier 2020 soit 22 années de profondeur d'historique.
- A partir de Janvier 2002 jusqu'à janvier 2020 soit 18 années de profondeur d'historique.
- A partir de Janvier 2008 jusqu'à janvier 2020 soit 12 années de profondeur d'historique.
- A partir de Janvier 2014 jusqu'à janvier 2020 soit 5 années de profondeur d'historique. Et on considére 2 jeux de données, avec une fréquence quotidienne et mensuelle.

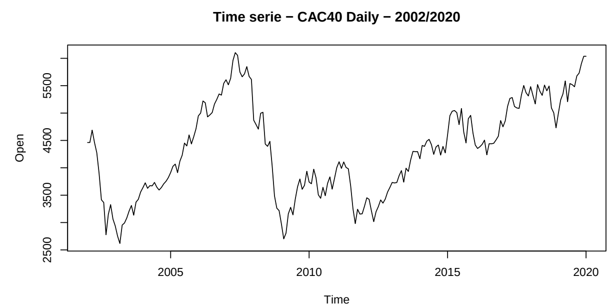
2.3.1 Graphique des séries temporelles - valeur observée Prix à la fermeture (Close)

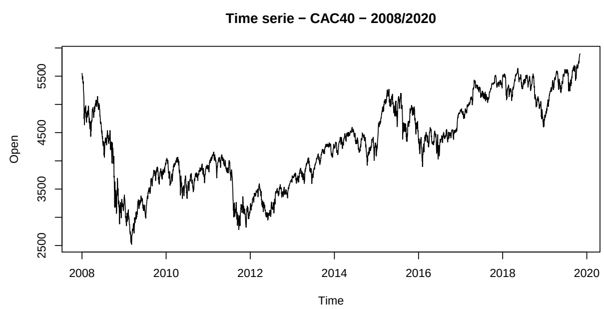
On a 4 séries temporelles possibles en fonction du choix de la quantité observée (High, Low, Open, Close, Volume). On va s'intéresser à la valeur à la fermeture pour la cotation de l'indice CAC40 (Close).

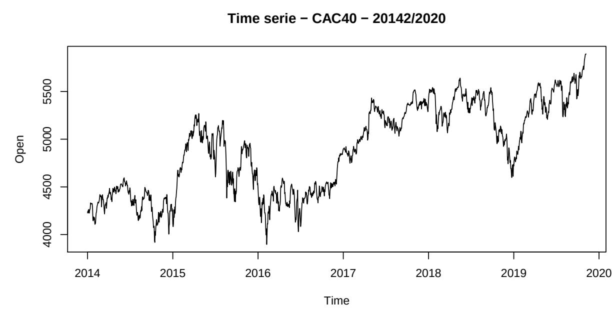


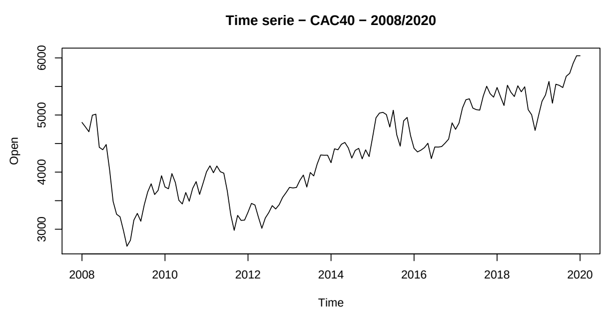


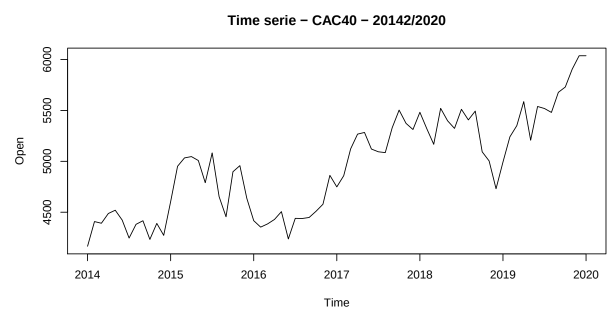








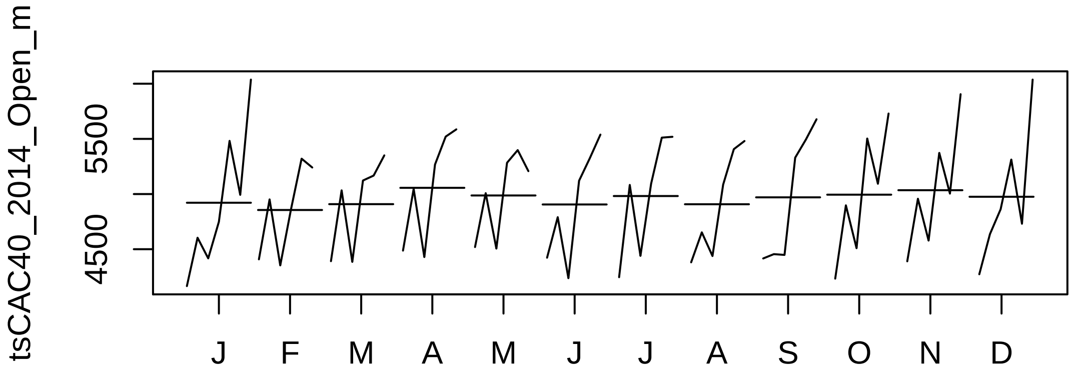


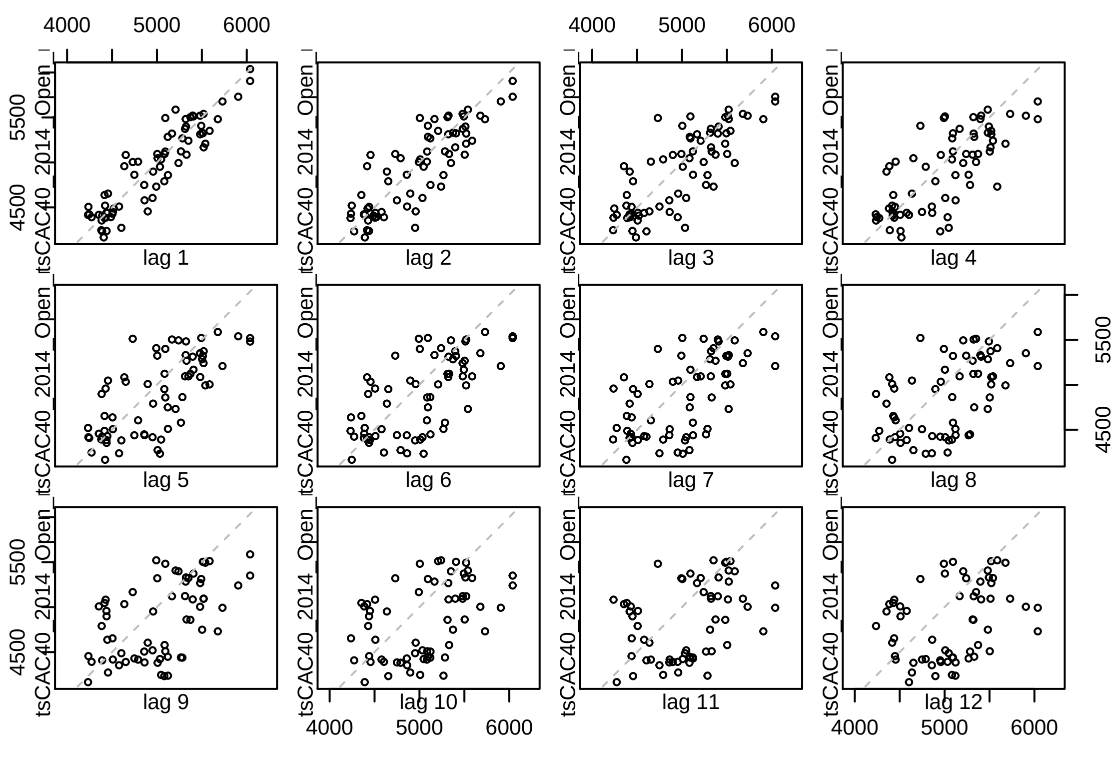


Dans le cas des données mensuelles On retrouve biensûr la forme globale de la série mais moins bruitée.

2.3.2 Représentations graphiques : month-plot et lag-plot

Si le diagramme retardée suggére une corrélation entre les deux séries, on dit que la série présente une autocorrélation d'ordre k.Ce diagramme permet de comprendre la dépendance de la série par rapport à son passée. Il donne une vision locale de la série, si y a une corrélation entre la série a un instant et la série 1, 2... instants avant.

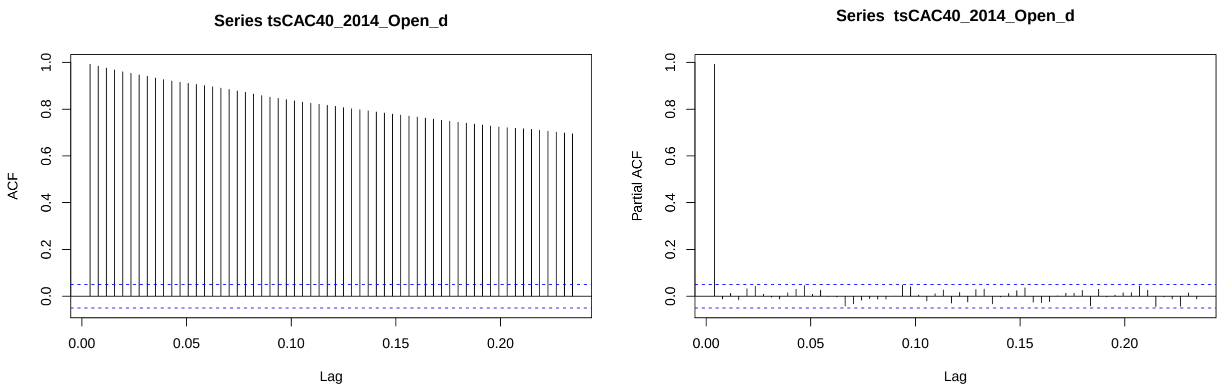




2.3.3 Etude de la stationarité

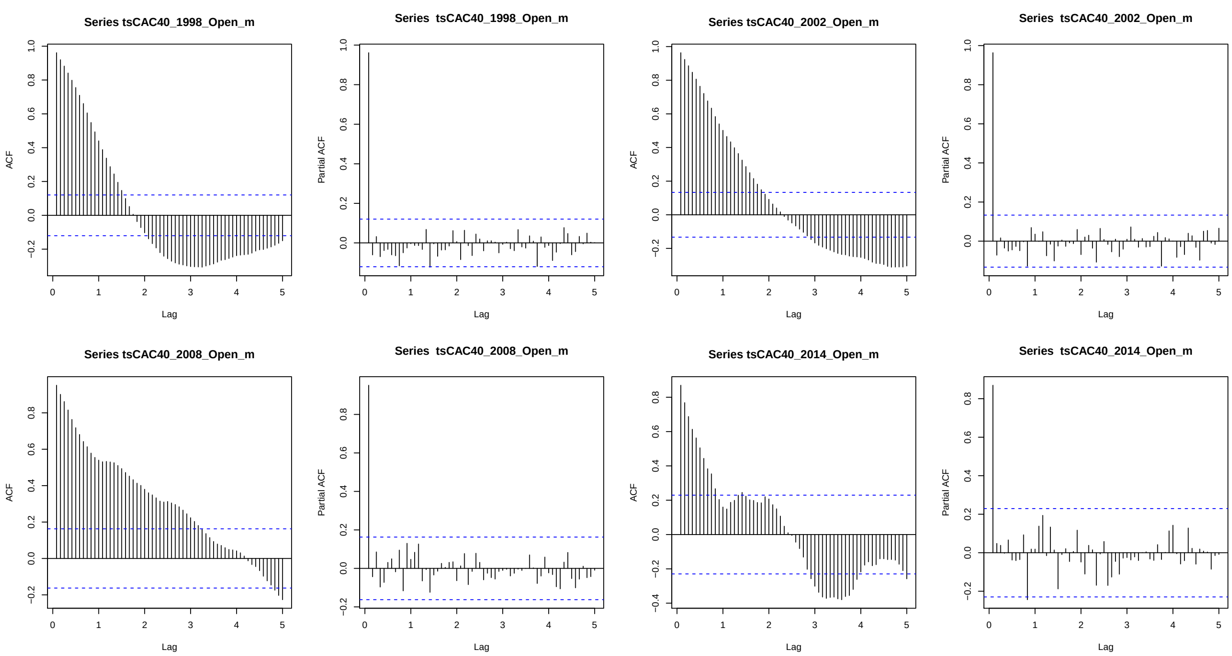
La stationnarité est la stationnarité du processus au sens faible. Un tel processus doit avoir les propriétés suivantes : La moyenne et la variance ne varient pas au cours du temps et le processus n a pas de tendance. Pour vérifier ces hypothèses, on s'appuiera sur une analyse des graphique d'autocorrélation ACF et d'autocorrélation partielle PACF ainsi que sur le test de $Dickey - Fuller$.

- Fonction d'autocorrélation ACF et PACF



On constate que les variables sont liées entre elles, i.e. les données ne semblent pas être stationnaires.

On confirme cette hypothèse à l'aide d'un test de $Dickey - Fuller$ on p_value=0.2639764.



Là aussi on constate que les variables sont liées entre elles, i.e. les données ne semblent pas être stationnaires. On confirme cette hypothèse à l'aide d'un test de $Dickey - Fuller$.

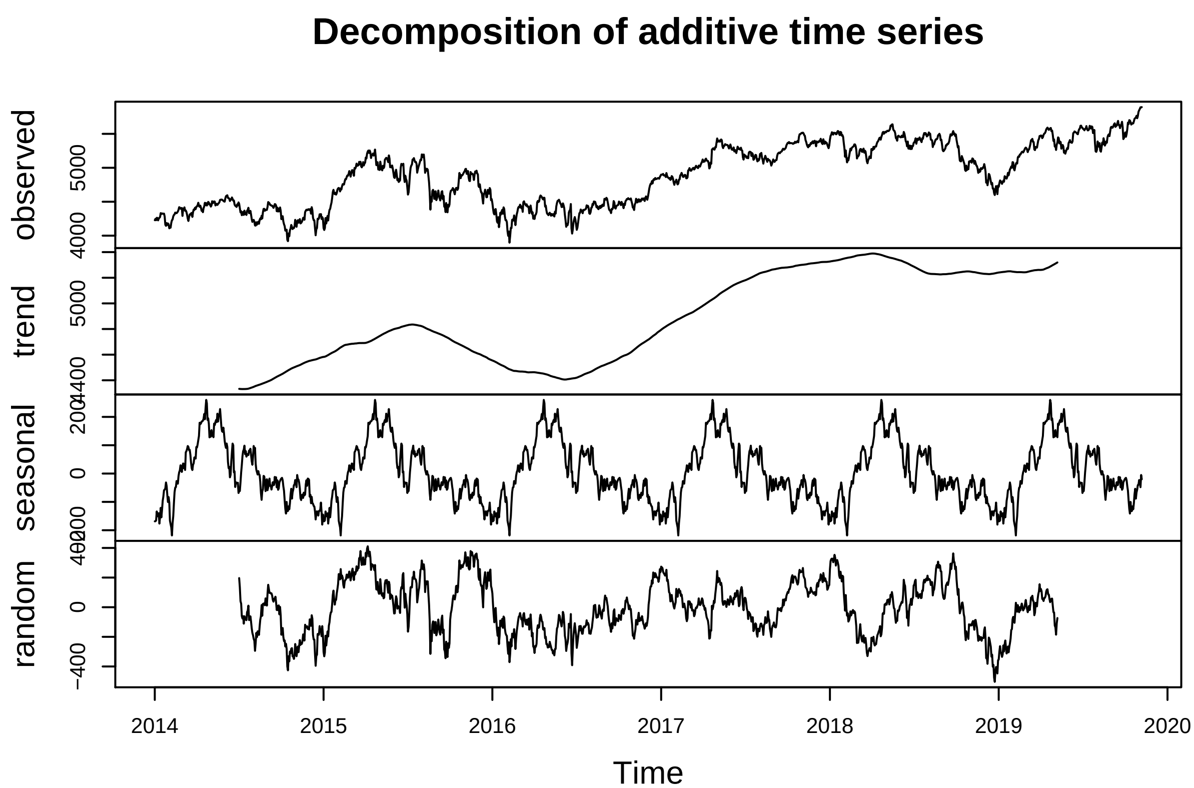
série	Dickey-Fuller	Lag order	p-value
1998	-2.3092612	6	0.4456761
2002	-2.1857017	5	0.4980833

série	Dickey-Fuller	Lag order	p-value
2008	-4.2978772	5	0.01
2014	-2.1515395	4	0.5136323

La p-valeur de ce test est importante et confirme donc que les données ne sont pas stationnaires. Il y a une cependant exception pour la série 2008.

2.3.4 Decomposition des séries temporelles :

Ici on va décomposer la série temporelles en utilisant la fonction décompose de R de façon à avoir une idée générale de la tendance (trend) saisonalité et bruit.



On retrouve les formes générales mais mon bruitées.La saisonnalité ne semble pas très nette.

On va essayer de rendre stationnaire nos séries. C'est un prérequis pour pouvoir effectuer une modilisation de type ARMA En utilisant la différentiation on va essayer de se ramener à un processus ARMA.Ainsi on va essayer de modéliser l'évolution du prix de l'indice CAC4O par un processus ARIMA. On commence donc par différentier les séries. Le facteur utilsé est de 1.

2.4 Détermination des modèles ARIMA

Les procssus ARIMA sont des processus non stationnaire.
Un processus X_t t Z est un processus ARIMA(p,d,q) si $\Delta^d X$ est un processus ARMA(p,q).

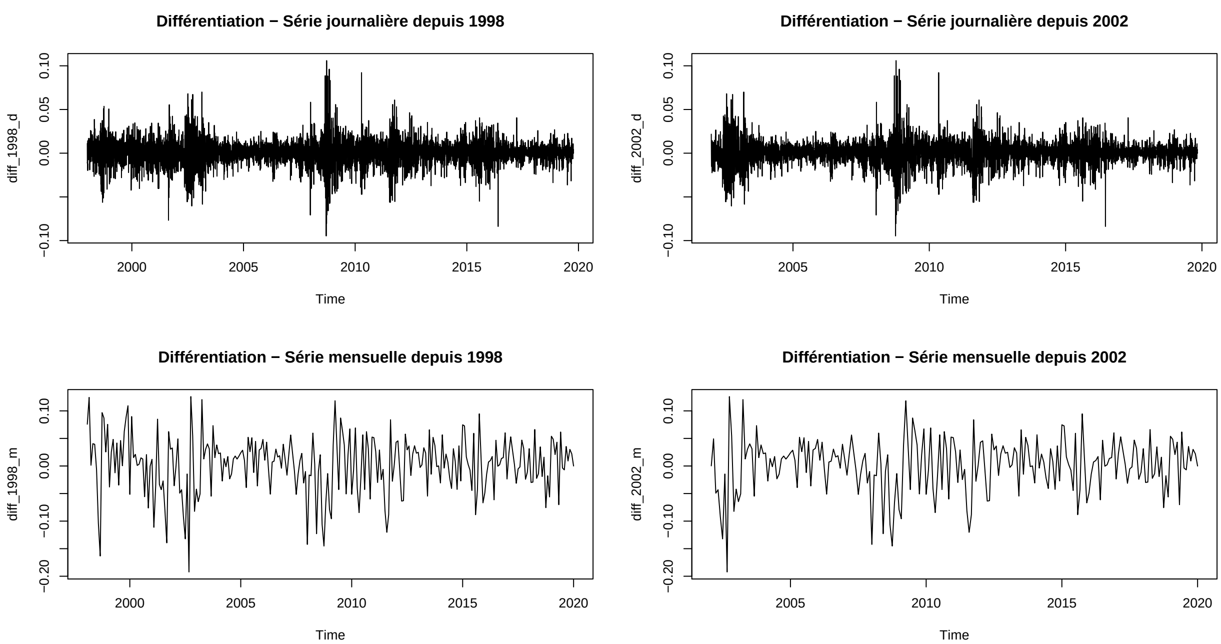
Les processus ARMA(p,q) font parti d'une famille très large de processus stationnaires. Ces processus sont composés des processus auto-régressifs AR(p) et de moyennes mobiles (“moving average”) MA(q). Un processus ARMA est la combinaison des processus autorégressifs et moyennes mobiles.

On cherche dans un premeir temps à se ramener à un processus sationnaire en utilisant la différentiation. Si l'on est bien dans le cadre d'un modèle ARIMA, après l'opération de différentiation on se ramènera à l'étude d'un processus ARMA(p,q).

2.4.1 Stationarisation des processus par differentiation des séries

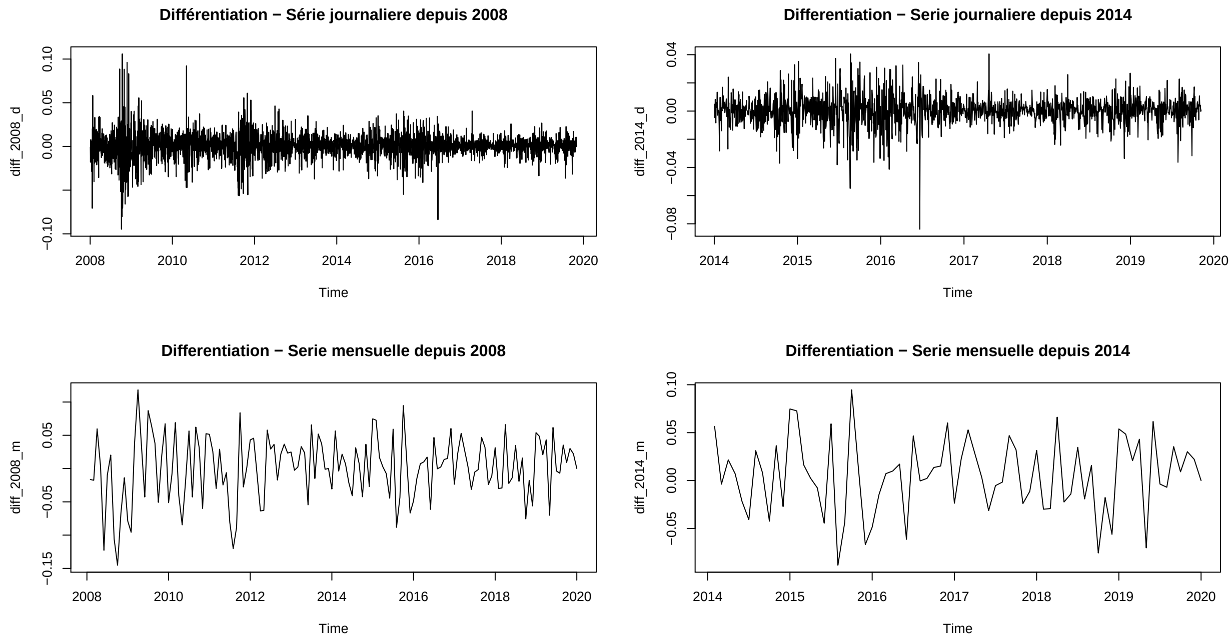
Pour tenter de rendre la série stationnaire, on applique la méthode de différentiation. On utilise la function R diff. En paramètre on passe un facteur de 1 pour la différence et de 0 pour la saisonnalité. On retrouve l'idée de la transformation Log-return R_t des prix X_t où $R_t = Log(X_t/X_{t-1})$ qui est classique en finance.

- Séries obtenues par differentiation pour les 2 jeux de données quotidien et mensuel à partir de 1998 et 2002

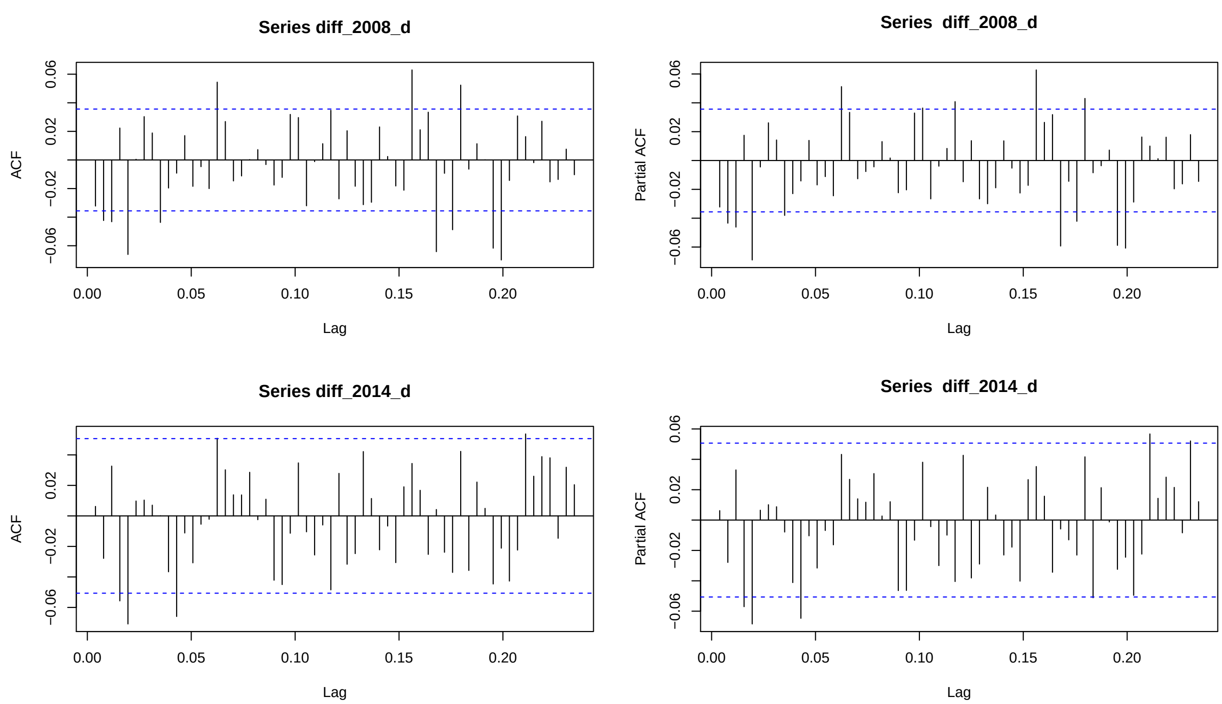


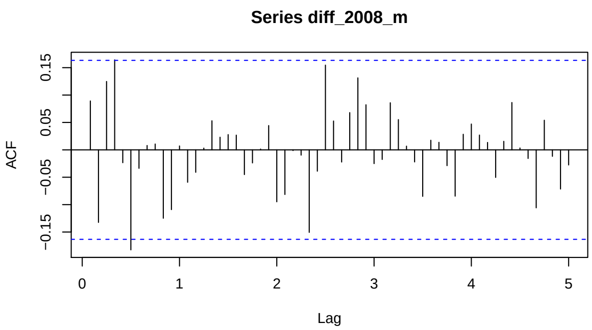
Au sens de la stationnarité faible, les séries semblent bien stationnaire. On retrouve bien une moyenne nulle et cosntante dans le temps. Ainsi qu'aussi une variance constante bien que cet aspect soit moins évident. On affaibliera cetre dernière hypothèse loes de l'étude des modèles GARCH au dernier paragraphe.

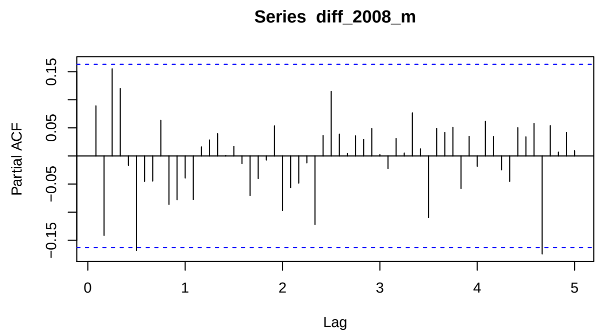
- Séries obtenues par differentiation pour les 2 jeux de données quotidien et mensuel à partir de 2008 et 2014

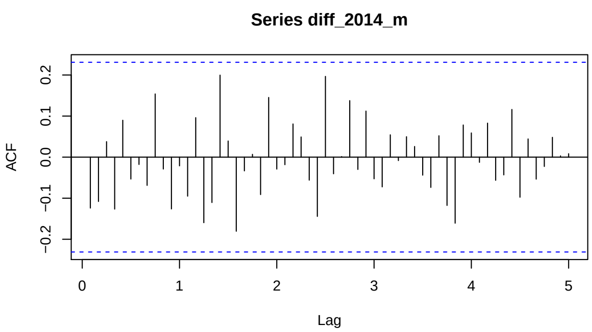


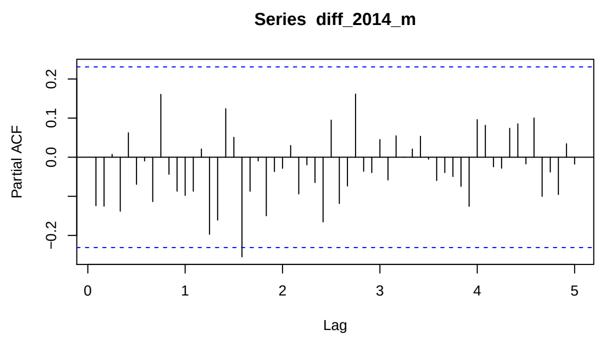
Les chorélogrames des séries différentiés n'ont pas de fortes valeurs.











L'hypothèse de stationnarité du processus ainsi transformé étant acceptable, on peut envisager une modélisation ARIMA. Compte tenu de l'allure des autocorrélogrammes de Yt, nous pouvons penser modéliser la série Xt par un processus ARMA (p, q).

On confirme cette impression avec différents tests de stationnarité.

- Test ADF - Augmented Dickey-Fuller Test (Robuste à l'autocorrélation)

série	Dickey-Fulle r	Lag order	p-value
1998 j	-17.6300806	17	0.01
2002 j	-16.1984318	16	0.01
2008 j	-16.1037353	14	0.01
2014 j	-12.6670123	11	0.01
1998 m	-5.5541872	6	0.01
2002 m	-5.607552	5	0.01
2008 m	-5.1084101	5	0.01
2014 m	-3.8500507	4	0.0214233

- Test PP - Phillips-Perron Unit Root Test (Robuste à l'hétéroscédasticité)

série	Dickey-Fuller Z(alpha)	Truncation Lag order	p-value
1998 j	-5150.43149	10	0.01
2002 j	-4243.9501495	10	0.01
2008 j	-2877.2657544	9	0.01
2014 j	-1411.017738	7	0.01
1998 m	-240.4025467	5	0.01
2002 m	-201.4496515	4	0.01

série	Dickey-Fuller Z(alpha)	Truncation Lag order	p-value
2008 m	-130.1214991	4	0.01
2014 m	-74.412947	3	0.01

- Test de KPSS

série	KPSS Level	Truncation Lag order	p-value
1998 j	0.0768448	10	0.1
2002 j	0.1132662	10	0.1
2008 j	0.2473177	9	0.1
2014 j	0.0301087	7	0.1
1998 m	0.0747491	5	0.1
2002 m	0.1190819	4	0.1
2008 m	0.2058661	4	0.1
2014 m	0.0524481	3	0.1

Ces tests confirment la stationnarité de la série différentiée (difference=1). A titre comparatif, la série Zt obtenue en différenciant 2 fois donne des résultats ne semblant pas significativement différents. Aussi, différencier 1 fois suffit pour obtenir un modèle stationnaire.

L'hypothèse de stationnarité du processus ainsi transformé, étant vérifié on peut envisager une modélisation ARIMA(p,1,q). On passe maintenant à la détermination des paramètres p et q du mopdèle ARMA(p,q).

L'estimation de p et de q se fait simplement en lisant le graphe des fonctions d'auto-corrélation et d'auto-corrélation partielle. Le graphe de la fonction d'auto-corrélation nous fournit la valeur de q. Le graphe de la fonction d'autocorrélation partielle nous donne la valeur de p. On peut alors essayer d'améliorer le modèle en prenant des valeurs de p et q plus petites que les valeurs obtenues précédemment, en utilisant notamment les critères d'AIC ou de BIC.

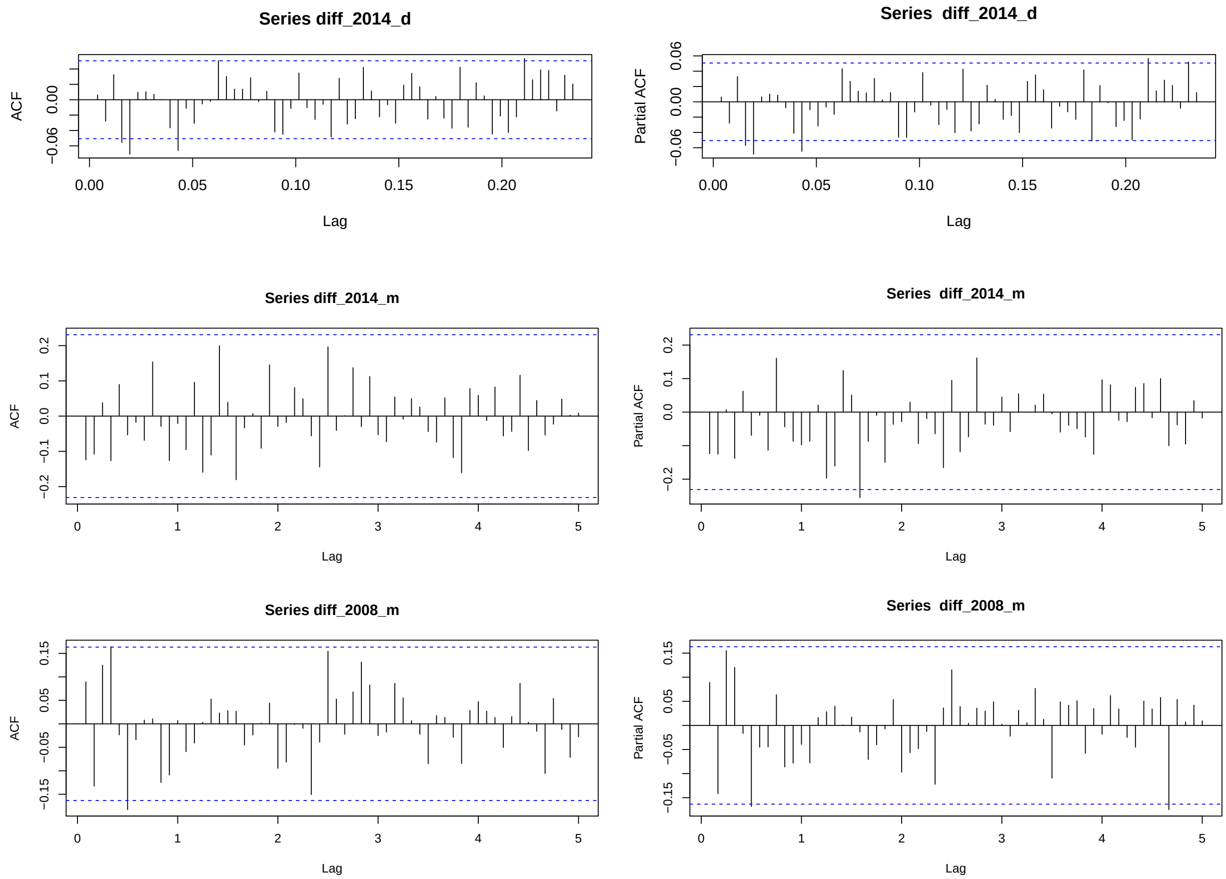
Les autres paramètres α_i et α_i se font pas minimisation / regression

2.4.2 Détermination des paramètres p et q, études des corrélogrammes et autocorrélations partiels

L'estimation de p et de q se fait simplement en lisant le graphe des fonctions d'auto-corrélation et d'auto-corrélation partielle. Le graphe de la fonction d'auto-corrélation nous fournit la valeur de q (ACF). Le graphe de la fonction d'autocorrélation partielle nous donne la valeur de p (PACF).

- Etudes des corrélogrammes et autocorrélations partiels (acf et pacf)

Ces graphiques, peuvent permettre d'identifier les composantes AR(p) et MA(q) d'un ARMA(p,q). Dans le cas d'un AR(p) : Dans le cas e'un MA(q) : L'ACF est nulle à partir de l'ordre q + 1. Si une ACF empirique semble nulle à partir d'un certain ordre q + 1, on peut penser qu'il s'agit de l'ACF d'une série MA(q). Cependant pour certains processus, ni la fonction d'autocorrélation, ni la fonction d'autocorrélation partielle ne possèdent de point de rupture. Dans de tels cas, il faut construire un modele mixte.



La PACF d'un processus qui a une composante moyenne mobile a une décroissance exponentielle. Ainsi la PACF d'un ARMA(p,q), $q > 0$ présente une décroissance exponentielle. Ici la PACF ne décroît pas exponentiellement, et rien de très net ne ressort des différents graphiques. De manière similaire, la fonction d'autocorrélation (ACF) d'un AR(p) montre une décroissance exponentielle avec ou sans oscillations vers 0 et La fonction d'autocorrélation partielle (PACF) d'un AR(p) est nulle à partir de l'ordre p+1. Ici aucunes des valeures n'est importante. Le modèle qi ressort de l'étude de ces corrélogrammes serait un modèle ARIMA(0,1,0) soit une marche aléatoire.

D'autres méthodes et fonction existent sous R, comme la méthode des coins, l'utilisation de la fonction eacf - d'autocorrélation étendue (Tsay, & Ciao) Cf Référence D3, Modèles de prévision Séries temporelles (A. Charpentier). Fonction R : armaselect et armasubsets.

On va plutôt utiliser la métode automatique : auto.arima basé sur les critères AIC,AICc et BIC.

Une première étude à montré que les différents modéles résultants ne sont pas validés pour les donneés quotidiennes, quelque soit la profondeur d'historique le tests de blancheur des résidus échoue. Mais il y a cependant une amélioration (p-value = 0.09348) pour la profondeur d'historique la plus faible (données à partir de 2014). La profondeur de l'historique et la forme de la courbe associé semble jouer un rôle assez important.

En ce qui concerne les données mensuelles tous les modèles obtenus: ARIMA(0,1,0) pour 2008 et 2014, SARIMA(0,1,0)(1,0,0) pour 2002 et SARIMA(1,1,0)(2,0,2) pour 1998 valident le test de blancheur des résidus avec des p-values importantes (> 0.63). Par contre les tests de normalités des résidus ne sont validés que par le modèle obtenu avec la plus faible profondeur d'historique (données à partir de 2014).

Dorénavant on ne considère plus que les données mensuelles à partir de 2014. Dans le dernier § on essaiera d'autres types de modèles sur les données quotidienne (toujours à partir de 2014) : les modèles à volatilités stochastique. Ces modèles semblent en effet bien mieux adaptéS aux données financières, en particulier pour les données journalières. Dans le cas des données quotidiennes on a une variance non constante. On va utiliser les modèles GARCH comme alternatve. On verra qu'ils sont mieux adaptées à la modélisation des séries financières.

2.4.3 Méthode automatique de calibration d'un modèles ARIMA sur les données 2014 mensuelles

La fonction auto.arima renvoie les meilleurs modèles ARIMA en considérant les critères AIC, AICc ou BIC value. Ici on va sélectionner

```
##
##  ARIMA(2,1,2)(1,0,1)[12]                    : Inf
##  ARIMA(0,1,0)                               : 874.9688
##  ARIMA(1,1,0)(1,0,0)[12]                    : 878.1471
##  ARIMA(0,1,1)(0,0,1)[12]                    : 877.8189
##  ARIMA(0,1,0)(1,0,0)[12]                    : 877.0904
##  ARIMA(0,1,0)(0,0,1)[12]                    : 877.0901
##  ARIMA(0,1,0)(1,0,1)[12]                    : 879.2919
##  ARIMA(1,1,0)                               : 875.9791
##  ARIMA(0,1,1)                               : 875.673
##  ARIMA(1,1,1)                               : 876.9957
##
##  Best model: ARIMA(0,1,0)
```

```
## Series: cac40.1418
## ARIMA(0,1,0)
##
## sigma^2 estimated as 39793:  log likelihood=-436.45
## AIC=874.91   AICc=874.97   BIC=877.08
```

On peut utiliser une approche empirique et regarder le critère AIC AICc et BIC des différents modèle obtenues on prendra celui qui minimise ces critères.

- Voici le résultat obtenu pour la série 2014 quotidienne :

ARIMA	AIC	AICc	BIC	sigma2	log-likelihood
010	874.9052585	874.9687506	877.0796458	3.9792802×10^4	-436.4526293
110	875.7855632	875.9791115	880.1343377	3.9713265×10^4	-435.8927816
011	875.4794941	875.6730425	879.8282687	3.9520002×10^4	-435.7397471
111	876.6022466	876.9956892	883.1254084	3.9556106×10^4	-435.3011233
012	876.7147223	877.1081649	883.2378841	3.9658478×10^4	-435.3573611
112	877.3045697	877.9712364	886.0021188	3.9357096×10^4	-434.6522849
210	876.9189347	877.3123773	883.4420965	3.9792401×10^4	-435.4594673
211	878.5100484	879.176715	887.2075975	4.0131698×10^4	-435.2550242

SARIMA	AIC	AICc	BIC	sigma2	log-likelihood
010_101	876.8968377	877.0903861	881.2456123	4.0408082×10^4	-436.4484189
010_001	876.8965669	877.0901153	881.2453415	4.0407834×10^4	-436.4482835

SARIMA	AIC	AICc	BIC	sigma2	log-likelihood
010_101	878.8984363	879.2918789	885.4215981	4.105066×10^4	-436.4492182
011_001	877.425435	877.8188776	883.9485968	4.0103861×10^4	-435.7127175
110_100	877.7536494	878.1470921	884.2768113	4.0317808×10^4	-435.8768247
212_101	882.6977577	884.66267	897.9184686	3.9770954×10^4	-434.3488789

On remarque que selon ce critère plusieurs modéles sont très proches. On va sélectionner les modèles suivants ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1)

2.5 Validation des modèles obtenus

2.5.1 Statistiques

Qualité d'évaluation des coefficients du modèles.

ARIMA(0,1,0)

Marche aléatoire, pas de coéfficients auto-regressif ou Moyenne mobile

ARIMA(0,1,1)

```
##               ma1
## t.stat -1.218144
## p.val   0.223169
```

ARIMA(1,1,0)

```
##               ar1
## t.stat -1.062969
## p.val   0.287796
```

ARIMA(1,1,1)

```
##             ar1       ma1
## t.stat 1.657491 -2.468702
## p.val  0.097420  0.013560
```

2.5.2 Corrélations

ARIMA(1,1,1)

```
##            ar1        ma1
## ar1  1.0000000 -0.9745172
## ma1 -0.9745172  1.0000000
```

2.5.3 Blancheur des résidus

Test de Box-Pierce

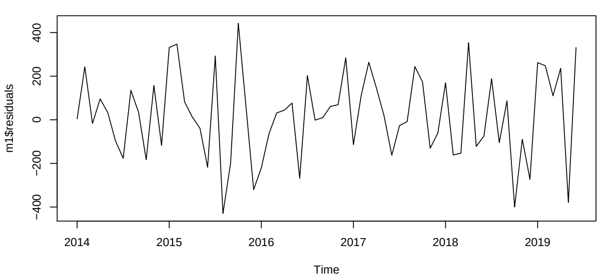
série	X-squared	df	p-value
010	14.0908242	20	0.8258612
011	12.3646978	20	0.9029516
110	12.8963073	20	0.8817886
111	12.2646538	20	0.9066656

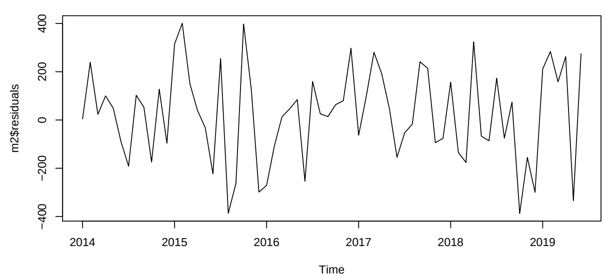
Test de Ljung-Box

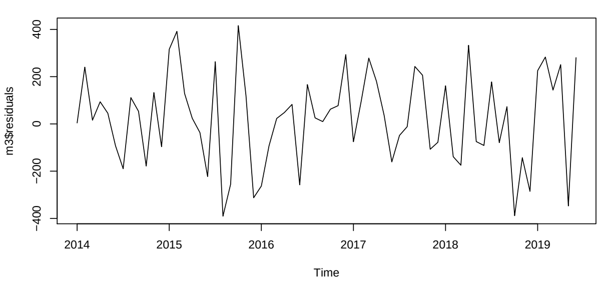
série	X-squared	df	p-value
010	17.9379723	20	0.5914944
011	16.1756863	20	0.7056669
110	16.8036489	20	0.6656843
111	16.0374931	20	0.7142955

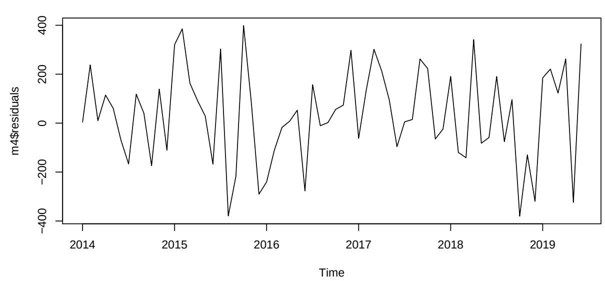
Dans le cas des modèles sur données mensuelles le test de blancheur des résidus accepte le modèle, et ce dans tous les cas. Dans le cas des données quotidienne au contraire le modèle n'est pas validé la p-value la plus importante est obtenu pour les données 2014*

2.5.4 Graphiques de résidus obtenus à partir des différents modèles









2.5.5 Normalité des résidus

On regarde les tests classiques de normalités dans le cadre des données mensuelles.

Test de Shapiro-Wilk normality test

série	p-value
010	0.834067

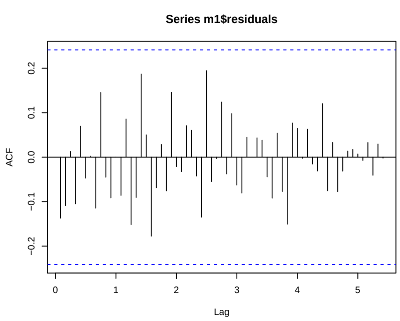
série	p-value
011	0.4754316
110	0.5873255
111	0.5176835

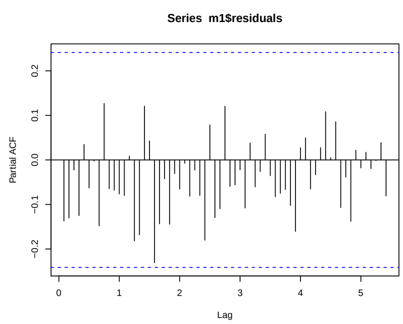
Test de Lilliefors (Kolmogorov-Smirnov) normality test

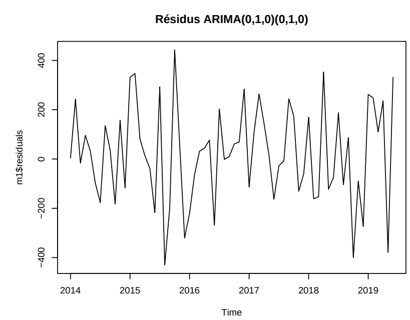
série	p-value
010	0.845679
011	0.918646
110	0.9011202
111	0.9957237

Les modéles suivent bien une loi de Gauss.

2.5.6 ACF et PACF des résidus

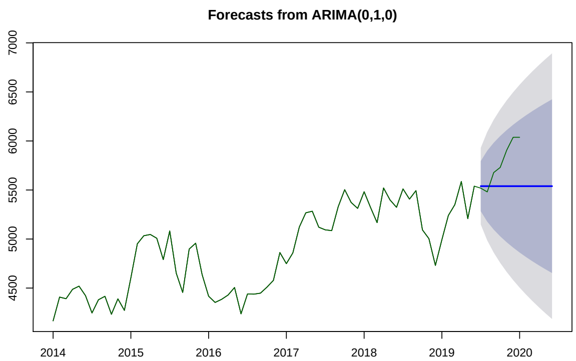


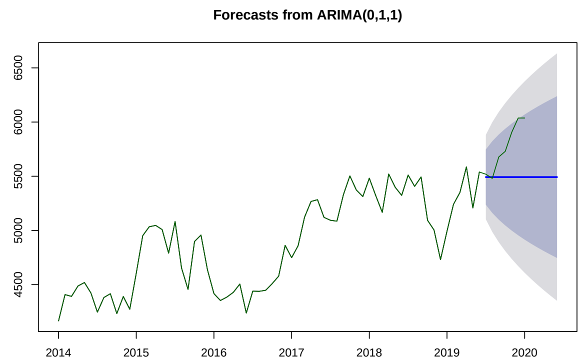


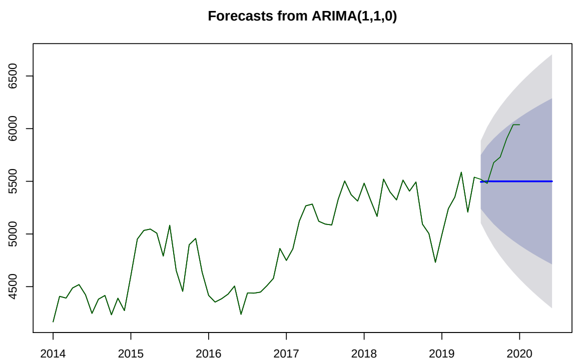


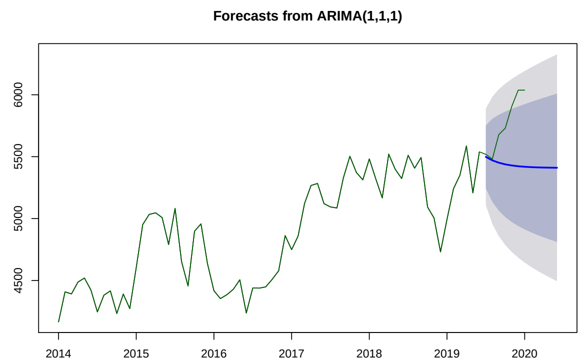
2.5.7 Prévisions à partir des modèles obtenus

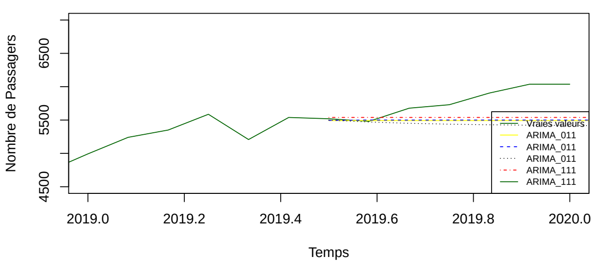
Ces prévisions seront utilisées pour nous permettre de déterminer les valeurs possibles prise par l'indice CAC40 à horizon 1 mois, 2 mois, 6 mois On souhaite utiliser l'intervalle de confiance obtenu pour nos aider à déterminer un scénario économique possible sur les actionS. Vue que l'on est dans le cadre de stress tests on cherche à déterminser un choc absolue plausible et non pas obtenir la valueur du CAC à horizon.

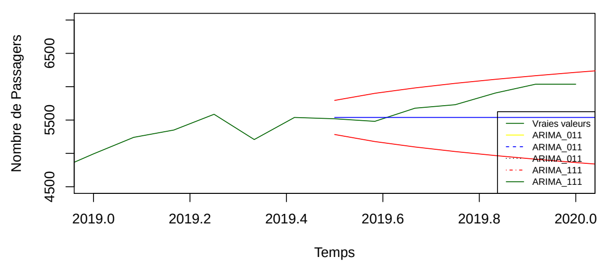












Les erreurs de prédictions obtenues semble suivre une normale centré et de variance assez constante. Le modèle ARIMA semble être adpaté pour la prédiction.

The forecast errors seem to be normally distributed with mean zero and constant variance, the ARIMA model does seem to provide an adequate predictive model Here we looked at how to best fit ARIMA model to univariate time series. Next thing that I'll work on is Multivariate Time Series Forecasting using neural net.

Cependant on remarque que la

2.6 Alternative au modèle de type ARIMA, les modèles GARCH

Ces modèles prenne en compte l'hétérocédasticité. Ils sont mieux adapotés aux séries financières.

Les séries financières comme ont a pu le voir, ne sont pas stationaire. Et on observe une tendance locale B1 p227.

Comme ont la fait dans le cas de la modélisation ARIMA on transforme la série originelle par différentiation. Pour obtenir une série stationnaire. Ici on va considérer comme c'est souvent le cas en finance le log-return. C'est à dire la quantité déduite du prix X_t de la manière suivante : Log-return R_t des prix X_t où $R_t = Log(X_t/X_{t-1})$. Cette appproche est bien adaptée au cadre de la théorie de Black-Scholes. On pourra se reporter au Document D3 page

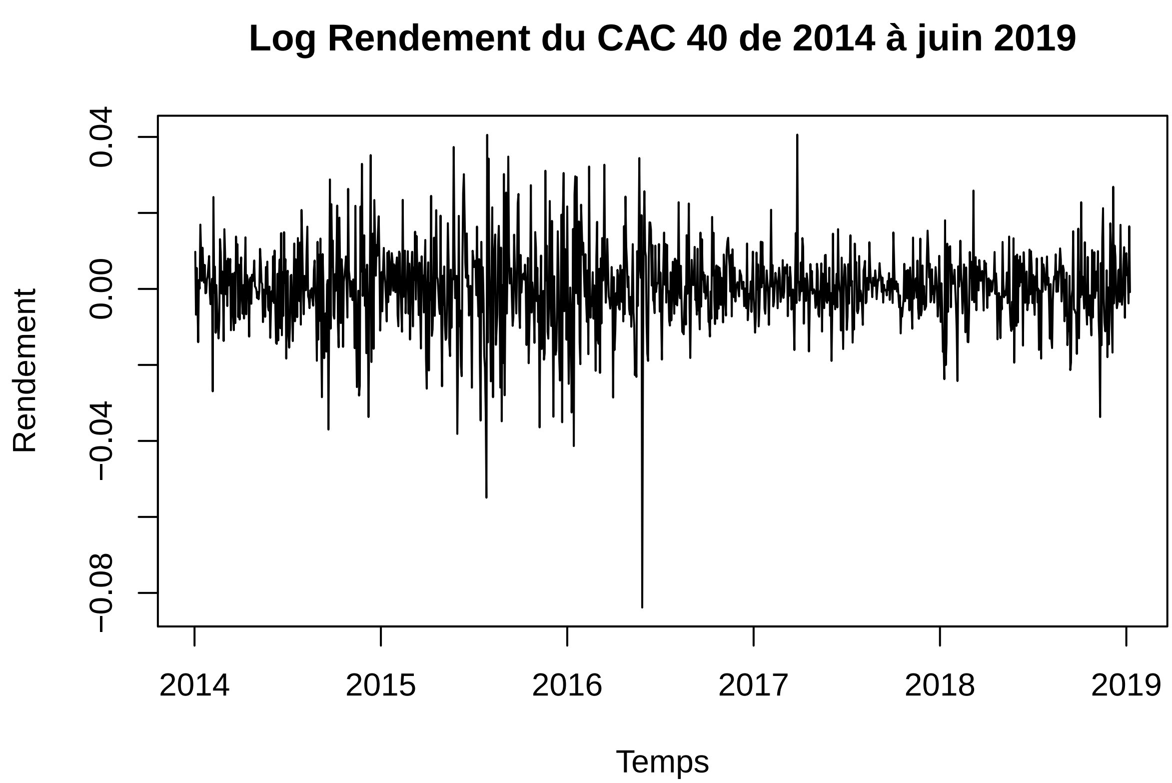
D'après D3 page 96 : Au regard de l'autocorrélogramme partiel du log return, une modélisation à l'aide d'un modèle GARCH(1, 1) semblerait possible. En effet, l'autocorrélogramme et l'autocorrélogramme partiel sont significativement nuls à partir des premiers retards (i.e. p = q = 1) Dans notre cas vu que l'indice et la période diffère.

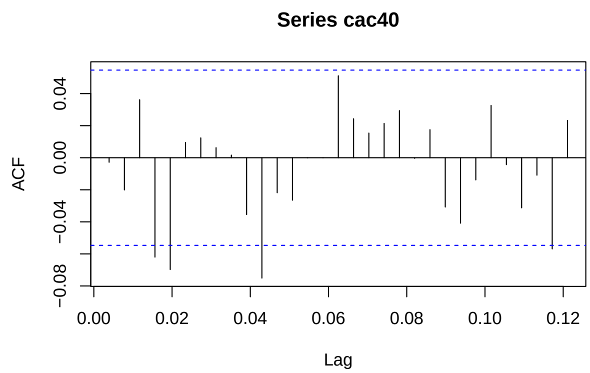
- Obtention de la série des log return

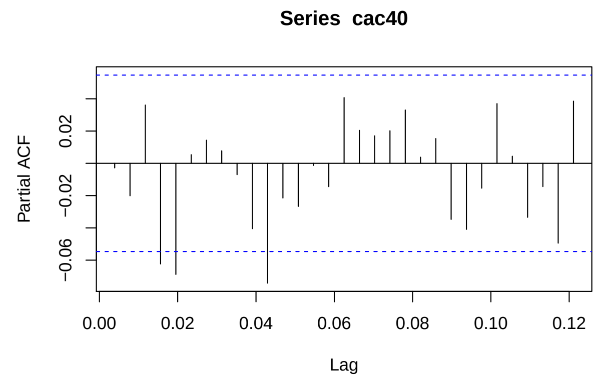
```
cac40.1418d <- window(tsCAC40_1998_Open_d,start=c(2014,1),end=c(2019,6))
```

```
cac40 <- diff(log(cac40.1418d))
```

- Graphique de la série obtenue








```
##
##  ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##
##      I     INITIAL X(I)        D(I)
##
##      1     1.071458e-04     1.000e+00
##      2     5.000000e-02     1.000e+00
##      3     5.000000e-02     1.000e+00
##
##     IT   NF      F         RELDF    PRELDF    RELDX   STPPAR   D*STEP   NPRELDF
##      0    1 -5.179e+03
##      1    7 -5.180e+03  4.31e-05  5.31e-04  1.0e-04  2.8e+10  1.0e-05  7.31e+06
##      2    8 -5.180e+03  1.09e-04  1.25e-04  4.9e-05  2.0e+00  5.0e-06  8.49e+00
##      3    9 -5.180e+03  1.39e-06  1.31e-06  5.0e-05  2.0e+00  5.0e-06  8.54e+00
##      4   17 -5.197e+03  3.26e-03  4.69e-03  4.5e-01  2.0e+00  8.2e-02  8.52e+00
##      5   20 -5.228e+03  6.00e-03  4.52e-03  7.4e-01  1.9e+00  3.3e-01  5.11e-01
##      6   22 -5.238e+03  1.85e-03  1.61e-03  7.9e-02  2.0e+00  6.6e-02  8.17e+01
##      7   24 -5.259e+03  3.99e-03  3.76e-03  1.3e-01  2.0e+00  1.3e-01  3.30e+03
##      8   26 -5.263e+03  8.06e-04  8.13e-04  2.2e-02  2.0e+00  2.6e-02  2.76e+05
##      9   32 -5.264e+03  3.12e-05  7.72e-05  7.6e-07  6.7e+02  9.1e-07  5.42e+01
##     10   33 -5.264e+03  1.07e-06  1.08e-06  7.4e-07  2.0e+00  9.1e-07  1.07e+01
##     11   43 -5.277e+03  2.51e-03  4.40e-03  1.1e-01  2.0e+00  1.5e-01  1.07e+01
##     12   44 -5.277e+03  6.07e-05  2.21e-03  1.7e-02  0.0e+00  3.0e-02  2.21e-03
##     13   45 -5.283e+03  1.16e-03  9.30e-04  1.0e-02  1.4e+00  1.5e-02  9.94e-04
##     14   46 -5.285e+03  3.47e-04  5.16e-04  8.7e-03  1.7e+00  1.5e-02  2.36e-03
##     15   48 -5.289e+03  6.89e-04  5.72e-04  3.2e-02  3.3e-01  6.0e-02  6.08e-04
##     16   49 -5.292e+03  5.63e-04  7.16e-04  2.7e-02  9.6e-01  6.0e-02  1.13e-03
##     17   61 -5.292e+03  1.73e-05  3.09e-05  1.0e-07  3.5e+00  1.7e-07  5.71e-05
##     18   62 -5.292e+03  1.35e-07  1.42e-07  7.1e-08  2.0e+00  1.7e-07  1.41e-05
##     19   82 -5.292e+03 -3.09e-15  2.28e-14  1.4e-14  3.0e+04  2.4e-14  1.41e-05
##
##  ***** FALSE CONVERGENCE *****
##
##  FUNCTION    -5.291806e+03   RELDX        1.387e-14
##  FUNC. EVALS      82         GRAD. EVALS      19
##  PRELDF       2.277e-14      NPRELDF      1.415e-05
##
##      I      FINAL X(I)        D(I)          G(I)
##
```

```
##      1    2.624121e-06     1.000e+00     5.035e+03
##      2    1.217364e-01     1.000e+00     1.175e+01
##      3    8.626932e-01     1.000e+00    -3.410e+00
```

```r
summary(cac.garch)
```

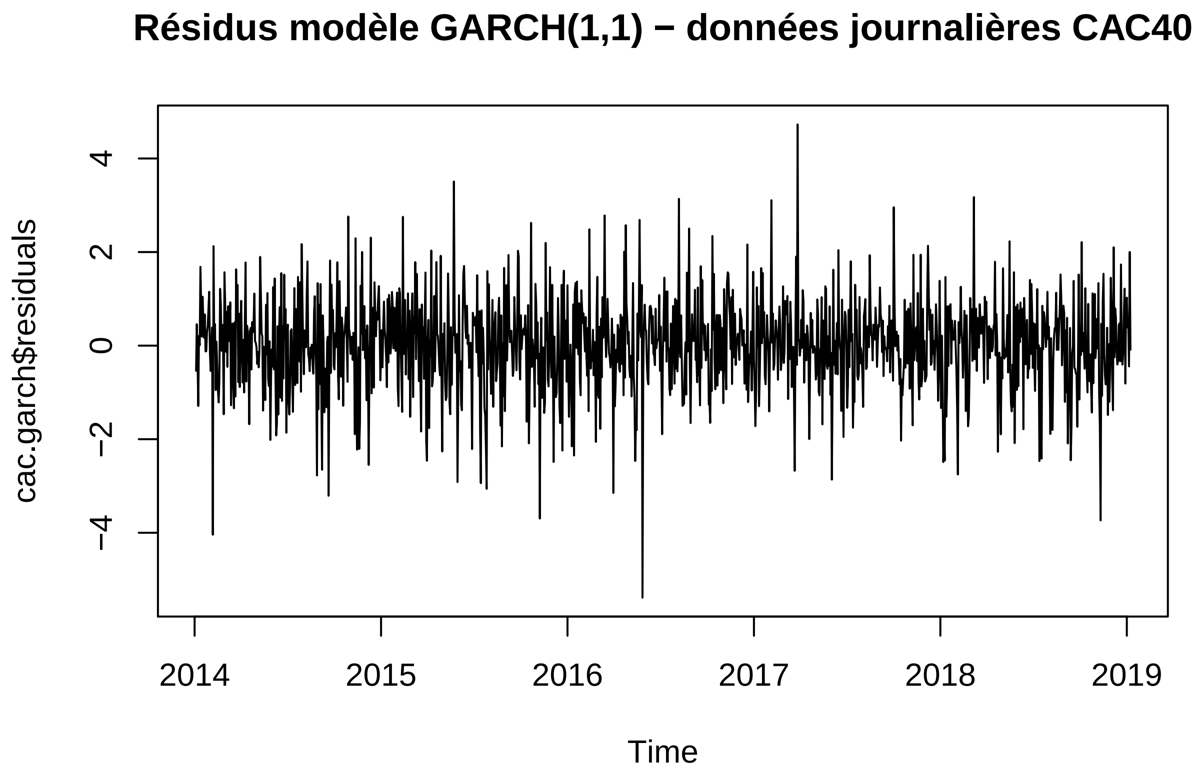
```
##
## Call:
## garch(x = cac40)
##
## Model:
## GARCH(1,1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.38618 -0.55726  0.03614  0.61375  4.72777
##
## Coefficient(s):
##     Estimate  Std. Error  t value Pr(>|t|)
## a0 2.624e-06   6.118e-07    4.289 1.79e-05 ***
## a1 1.217e-01   1.229e-02    9.905  < 2e-16 ***
## b1 8.627e-01   1.388e-02   62.157  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##  Jarque Bera Test
##
## data:  Residuals
## X-squared = 174.58, df = 2, p-value < 2.2e-16
##
##
##  Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 0.037894, df = 1, p-value = 0.8457
```

2.6.1 Validation des modèles obtenus

- Blancheur des résidus

```
##
##  Box-Pierce test
##
## data:  cac.garch$residuals
## X-squared = 20.738, df = 20, p-value = 0.4127
```

- Graphique des résidus



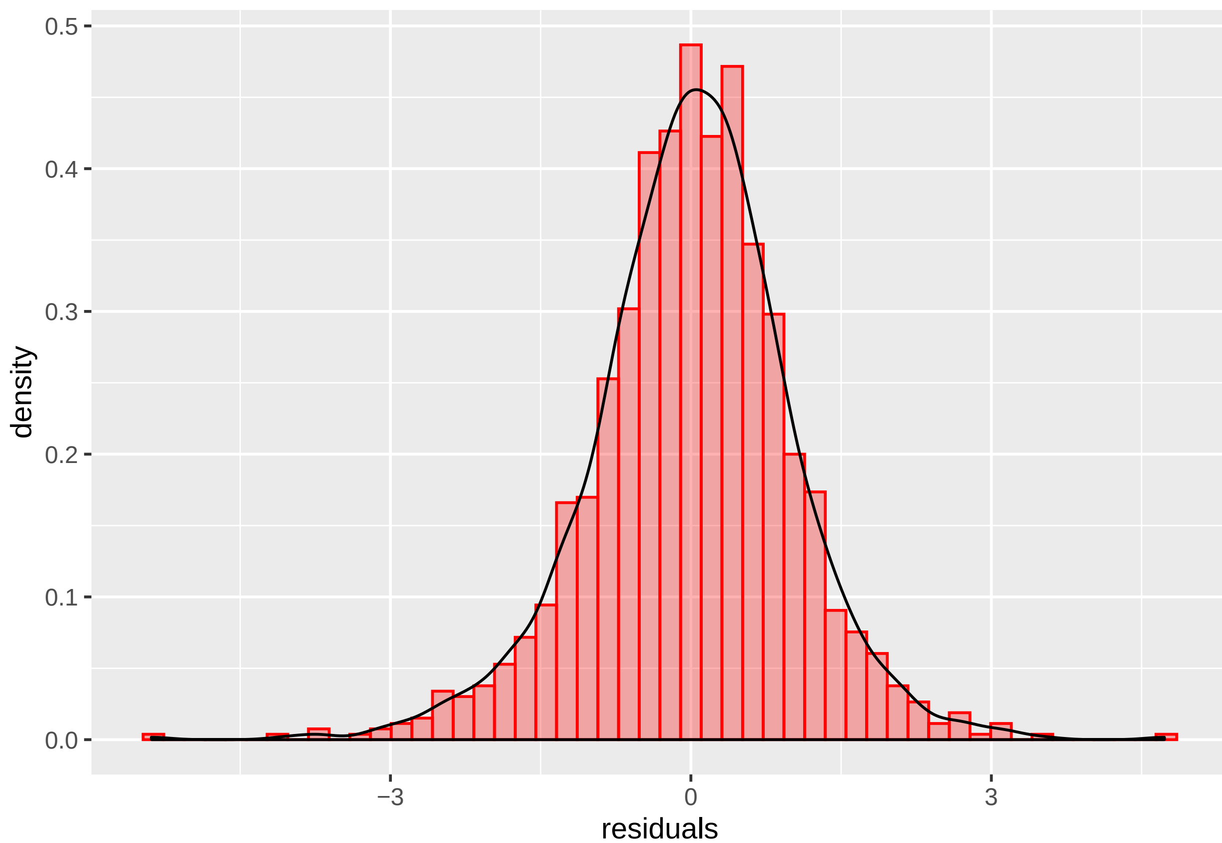
- Normalité des résidus

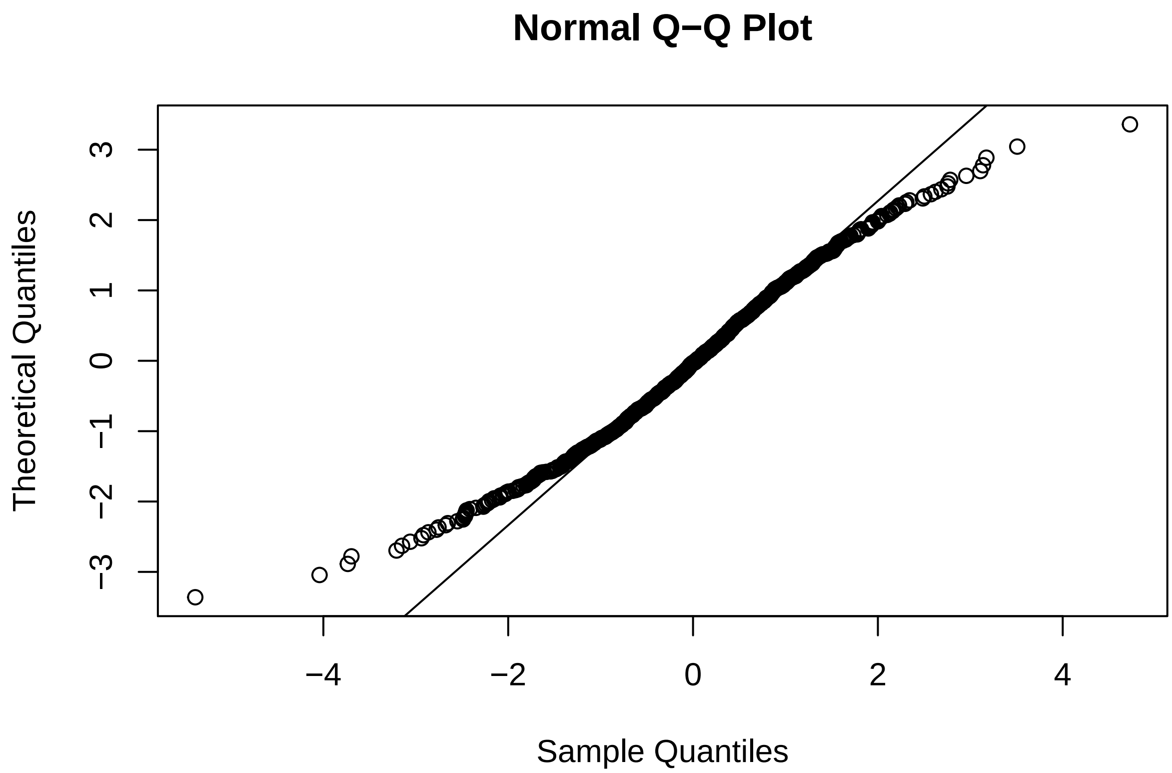
On regarde les tests classiques de normalités dans le cadre des données mensuelles.

```
##
##  Shapiro-Wilk normality test
##
## data:  cac.garch$residuals
## W = 0.98445, p-value = 1.684e-10
```

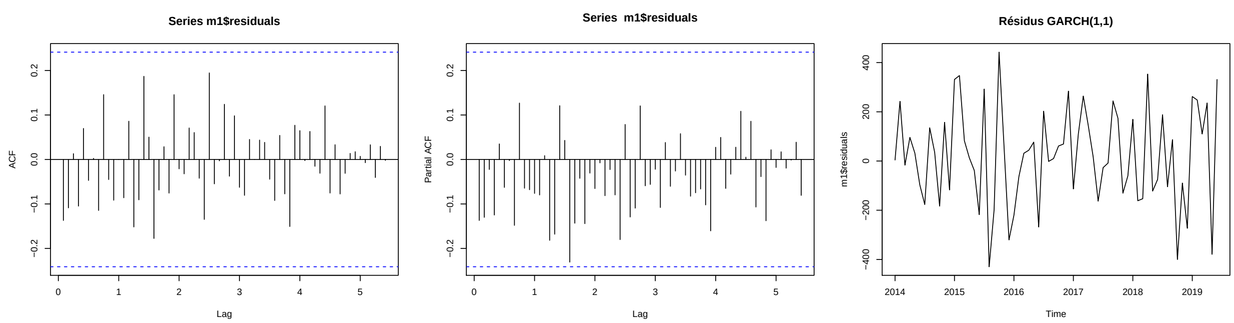
```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  cac.garch$residuals
## D = 0.039301, p-value = 8.099e-05
```

```
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
```


0.5
0.4
0.3
0.2
0.1
0.0
density
−3
0
3
residuals



- ACF et PACF des résidus



2.6.2 Prévision obtenus à partir du modèle GARCH(1,1)

```
#plot(cac.garch, ask = non-interactive())

#predict(cac.garch,30)
```

3 Annexe - partie I

3.1 Statistiques - qualité d'estimation des coefficients.

SARIMA(1,1,1)(0,1,1) :

```
##             ar1       ma1      sma1
## t.stat 2.441972 -23.74509 -7.446865
## p.val  0.014607   0.00000  0.000000
```

SARIMA(0,1,2)(0,1,1) :

```
##              ma1       ma2      sma1
## t.stat -10.12399 -2.447916 -7.466775
## p.val    0.00000  0.014369  0.000000
```

SARIMA(1,1,1)(0,1,2) :

```
##             ar1       ma1      sma1     sma2
## t.stat 2.380009 -23.14191 -6.638139 0.729410
## p.val  0.017312   0.00000  0.000000 0.465751
```

SARIMA(1,1,2)(0,1,1) :

```
##             ar1       ma1       ma2      sar1      sma1
## t.stat 0.521659 -2.890177 -0.167427 -0.723238 -2.870492
## p.val  0.601908  0.003850  0.867034  0.469534  0.004098
```

SARIMA(2,1,2)(0,1,1) :

```
##              ar1      ar2       ma1       ma2      sma1
## t.stat -3.435100 2.674158 -0.528898 -5.095158 -7.523646
## p.val   0.000592 0.007492  0.596876  0.000000  0.000000
```

SARIMA(2,1,2)(0,1,1) :

```
##             ar1       ar2       ma1      sma1
## t.stat 2.411968 -0.287502 -20.62754 -7.458516
## p.val  0.015867  0.773728   0.00000  0.000000
```

3.2 Corrélations - entre mes processus AR et MA

SARIMA(1,1,1)(0,1,1) :

```
##             ar1        ma1       sma1
## ar1   1.0000000 -0.4852660 -0.1182038
## ma1  -0.4852660  1.0000000 -0.0928834
## sma1 -0.1182038 -0.0928834  1.0000000
```

SARIMA(0,1,2)(0,1,1) :

```
##              ma1         ma2        sma1
## ma1   1.0000000 -0.80388631 -0.17532237
## ma2  -0.8038863  1.00000000  0.09559027
## sma1 -0.1753224  0.09559027  1.00000000
```

SARIMA(1,1,1)(0,1,2) :

```
##               ar1         ma1        sma1        sma2
## ar1   1.00000000 -0.49600612 -0.06105877 -0.06824537
## ma1  -0.49600612  1.00000000 -0.11872741  0.07391491
## sma1 -0.06105877 -0.11872741  1.00000000 -0.48889947
## sma2 -0.06824537  0.07391491 -0.48889947  1.00000000
```

SARIMA(1,1,2)(0,1,1) :

```
##               ar1         ma1         ma2        sar1        sma1
## ar1   1.00000000 -0.96848841  0.95785668 -0.03323734  0.01523959
## ma1  -0.96848841  1.00000000 -0.99025973  0.04521456 -0.04737931
## ma2   0.95785668 -0.99025973  1.00000000 -0.05279417  0.04692459
## sar1 -0.03323734  0.04521456 -0.05279417  1.00000000 -0.86340116
## sma1  0.01523959 -0.04737931  0.04692459 -0.86340116  1.00000000
```

SARIMA(2,1,2)(0,1,1) :

```
##               ar1         ar2         ma1         ma2        sma1
## ar1   1.00000000  0.25137615 -0.91158356  0.82084098  0.04735981
## ar2   0.25137615  1.00000000  0.04539109 -0.23962854 -0.16985606
## ma1  -0.91158356  0.04539109  1.00000000 -0.91073926 -0.13365536
## ma2   0.82084098 -0.23962854 -0.91073926  1.00000000  0.09669396
## sma1  0.04735981 -0.16985606 -0.13365536  0.09669396  1.00000000
```

SARIMA(2,1,2)(0,1,1) :

```
##              ar1         ar2         ma1        sma1
## ar1   1.0000000  0.04573990 -0.46680107 -0.12264148
## ar2   0.0457399  1.00000000 -0.41921113  0.02816471
## ma1  -0.4668011 -0.41921113  1.00000000 -0.08582669
## sma1 -0.1226415  0.02816471 -0.08582669  1.00000000
```

3.3 Vraies valeurs année 1980 et prédictions

- Vraies valeurs Année 1980

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1980 2848 2913 3248 3250 3375 3640 3771 3259 3206 3269 3181 4008
```

- Prédictions SARIMA(1,1,1)(0,1,1)

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1980       3176.918 3021.002 3332.833 2938.466 3415.369
## Feb 1980       2847.791 2684.491 3011.092 2598.045 3097.538
## Mar 1980       3122.217 2956.723 3287.710 2869.116 3375.317
## Apr 1980       3370.292 3203.281 3537.303 3114.871 3625.713
## May 1980       3393.392 3224.993 3561.791 3135.848 3650.936
## Jun 1980       3741.669 3571.916 3911.422 3482.054 4001.283
## Jul 1980       3986.431 3815.340 4157.523 3724.769 4248.093
## Aug 1980       3385.328 3212.908 3557.747 3121.635 3649.020
## Sep 1980       3046.176 2872.439 3219.912 2780.469 3311.882
## Oct 1980       3167.998 2992.955 3343.042 2900.293 3435.704
## Nov 1980       3102.021 2925.680 3278.361 2832.330 3371.711
## Dec 1980       3674.376 3496.747 3852.005 3402.716 3946.036
```

- Prédictions (4,0,1)(2,1,0)

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1980       3113.693 2939.675 3287.710 2847.556 3379.829
## Feb 1980       2810.667 2627.827 2993.506 2531.037 3090.296
## Mar 1980       3122.394 2937.427 3307.361 2839.512 3405.276
## Apr 1980       3341.407 3156.166 3526.649 3058.105 3624.710
## May 1980       3346.517 3158.196 3534.837 3058.505 3634.528
## Jun 1980       3687.187 3494.312 3880.061 3392.210 3982.163
## Jul 1980       3918.179 3720.903 4115.455 3616.472 4219.887
## Aug 1980       3316.835 3116.401 3517.269 3010.298 3623.372
## Sep 1980       3000.411 2797.245 3203.577 2689.695 3311.127
## Oct 1980       3122.026 2916.142 3327.909 2807.154 3436.898
## Nov 1980       3059.735 2851.042 3268.428 2740.566 3378.903
## Dec 1980       3628.918 3417.482 3840.354 3305.554 3952.282
```

4 Références

Books :

- (B1) Statistics of finantial Markets (J. Franke, W.K. Härdle, C. M. Hafner)
- (B2) Series-Temporelles-avec-R-methodes et cas (Y. Aragon)

Documents:

- (D1) Modèles GARCH et à volatilité stochastique (Christian. Francq)
- (D2) Time Series Analysis with ARIMA – ARCH/GARCH model in R (L-Stern Group - Ly Pham)
- (D3) Séries Temporelles et test d'adéquation d'un modèle GARCH(1,1) (Y. Djabrane)
- (D4) Rapport ISFA - Les Momentums et leur application dans le cadre des marchés boursiers (M. Adil Rahimi)

Blog/Internet:

- (I1) https://tradingninja.com/2017/03/sp-500-exponential-garch-volatility-model-using-r/
- (I2) https://tradingninja.com/2016/01/financial-time-series-modelling-using-arima-plus-garch-models/