

# Rapport - Séries Temporelles

*Philippe Real*

*12/01/2020*

## Contents

<b>1</b>	<b>Partie I - Exemple de modélisation appliqué au trafic voyageur</b>	<b>3</b>
1.1	Lecture des données et premières analyses de la série temporelle . . . . .	3
1.1.1	Lecture des données . . . . .	3
1.1.2	Chronogramme de la séries temporelles - sncf . . . . .	3
1.1.3	Représentations graphiques : month-plot et lag-plot . . . . .	3
1.1.4	Tendance et saisonnalité . . . . .	5
1.2	Prévision par lissage exponentiel . . . . .	7
1.2.1	Lissage exponentiel simple . . . . .	7
1.2.2	Lissage exponentiel double . . . . .	8
1.2.3	Méthode de Holt-Winters - Lissage exponentiel triple . . . . .	9
1.2.4	Autres méthodes . . . . .	10
1.2.5	Procédure automatique - modèles ajustés par ets. . . . .	10
1.2.6	Prédiction à partir du modèle de lissage exponentiel obtenu : (M,A,M) . . . . .	11
1.3	Modélisation . . . . .	11
1.3.1	Identification du modèle . . . . .	11
1.3.2	Validation des modèles SARIMA obtenus . . . . .	14
1.4	Modélisation automatique avec R . . . . .	16
1.5	Prévisions et comparaison des modèles obtenus . . . . .	17
<b>2</b>	<b>Partie II - Tentative de modélisation d'un indice boursier de type action à l'aide de processus ARIMA</b>	<b>19</b>
2.1	Introduction . . . . .	19
2.2	Lecture des données et premières analyses . . . . .	19
2.2.1	Traitement des données . . . . .	19
2.2.2	conversion des données en objet <i>time series</i> . . . . .	20
2.3	Analyse des séries temporelles obtenues . . . . .	20
2.3.1	Graphique des séries temporelles - valeur observée Prix à la fermeture (Close) . . . . .	20
2.3.2	Représentations graphiques : month-plot et lag-plot . . . . .	22
2.3.3	Etude de la stationarité . . . . .	23
2.3.4	Decomposition des séries temporelles : . . . . .	24
2.4	Détermination des modèles ARIMA . . . . .	25

2.4.1	Définitions . . . . .	25
2.4.2	Stationarisation des processus par differentiation des séries . . . . .	25
2.4.3	Détermination des paramètres p et q, études des corrélogrammes et autocorrélations partiels . . . . .	27
2.4.4	Méthode automatique de calibration d'un modèles ARIMA . . . . .	28
2.5	Validation des modèles obtenus . . . . .	29
2.5.1	Blancheur des résidus . . . . .	29
2.5.2	Graphiques de résidus obtenus à partir des différents modèles . . . . .	29
2.5.3	Normalité des résidus . . . . .	29
2.5.4	Prévisions à partir des modèles obtenus . . . . .	30
2.6	Alternative au modèle de type ARIMA, les modèles GARCH . . . . .	31
2.6.1	Graphique des résidus . . . . .	34
2.6.2	Prévision obtenus à partir du modèle GARCH(1,1) . . . . .	34
2.7	Références . . . . .	34

# 1 Partie I - Exemple de modélisation appliqué au trafic voyageur

## 1.1 Lecture des données et premières analyses de la série temporelle

### 1.1.1 Lecture des données

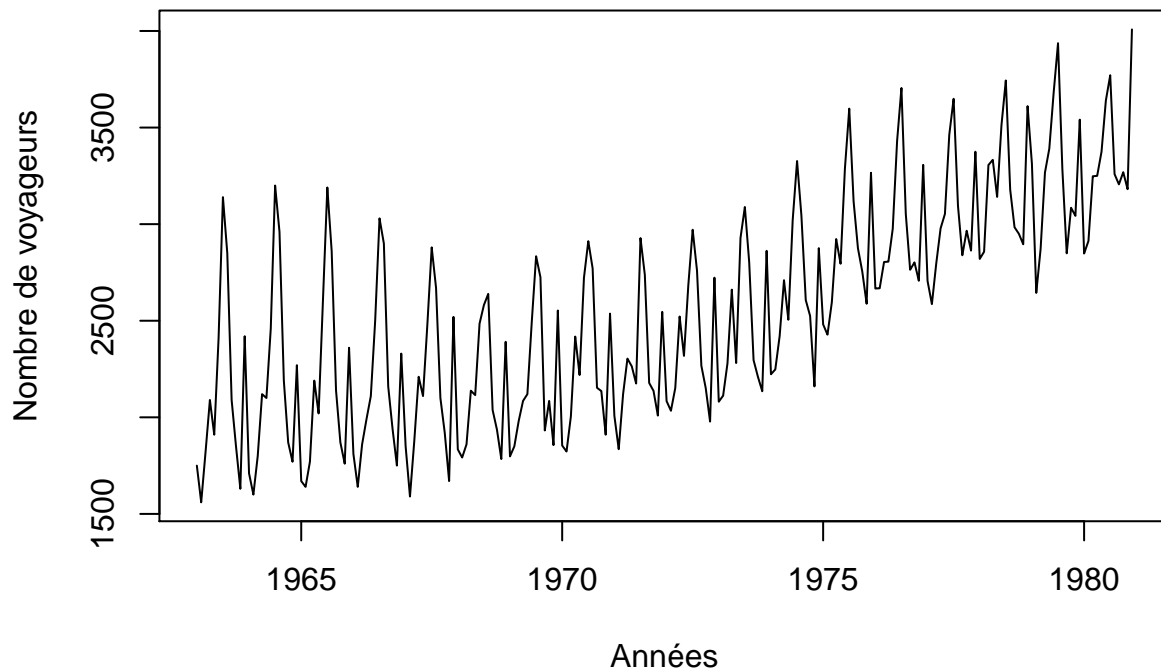
```
##      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1963 1750 1560 1820 2090 1910 2410 3140 2850 2090 1850 1630 2420
## 1964 1710 1600 1800 2120 2100 2460 3200 2960 2190 1870 1770 2270
```

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

### 1.1.2 Chronogramme de la séries temporelles - sncf

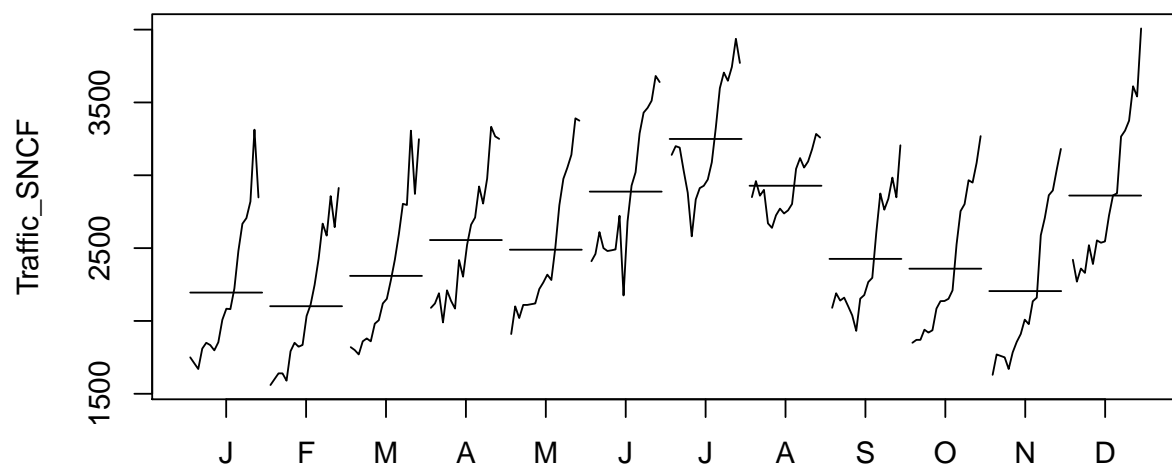
On a 4 séries temporelles possibles en fonction du choix de la quantité observée (High, Low, Open, Close, Volume). On va s'intéresser à la valeur à la fermeture pour la cotation de l'indice CAC40 (Close).

### Traffic sncf – 1963 à 1980

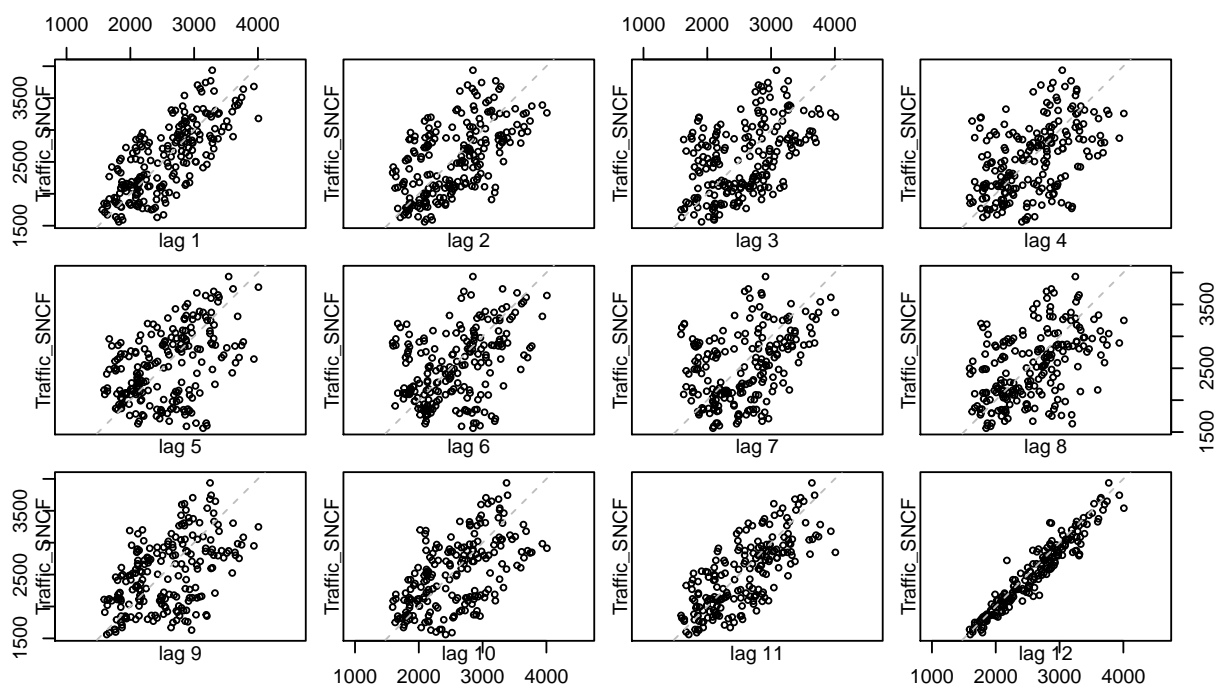


### 1.1.3 Représentations graphiques : month-plot et lag-plot

Si le diagramme retardée suggère 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 à 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 trafic en juillet août ainsi que décembre.

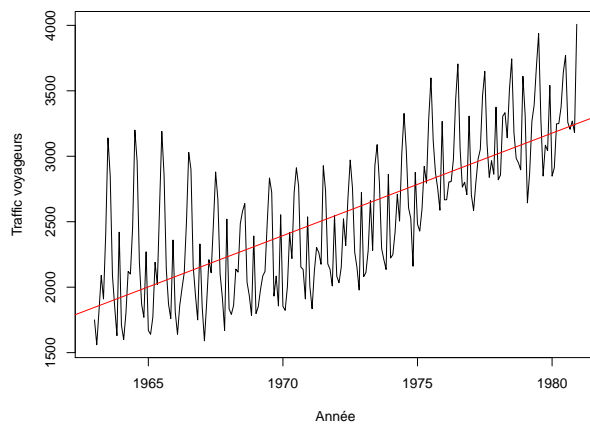


Le lag plot indique une saisonnalité de 1 an marquée.

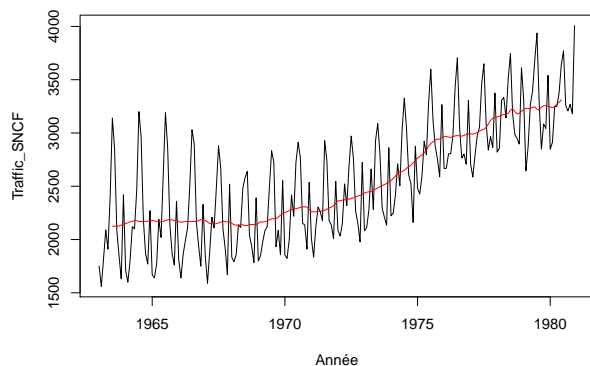
### 1.1.4 Tendance et saisonnalité

- Estimation de la tendance par moindres carrés ordinaires La tendance obtenue est une droite, la droite de régression par mco.

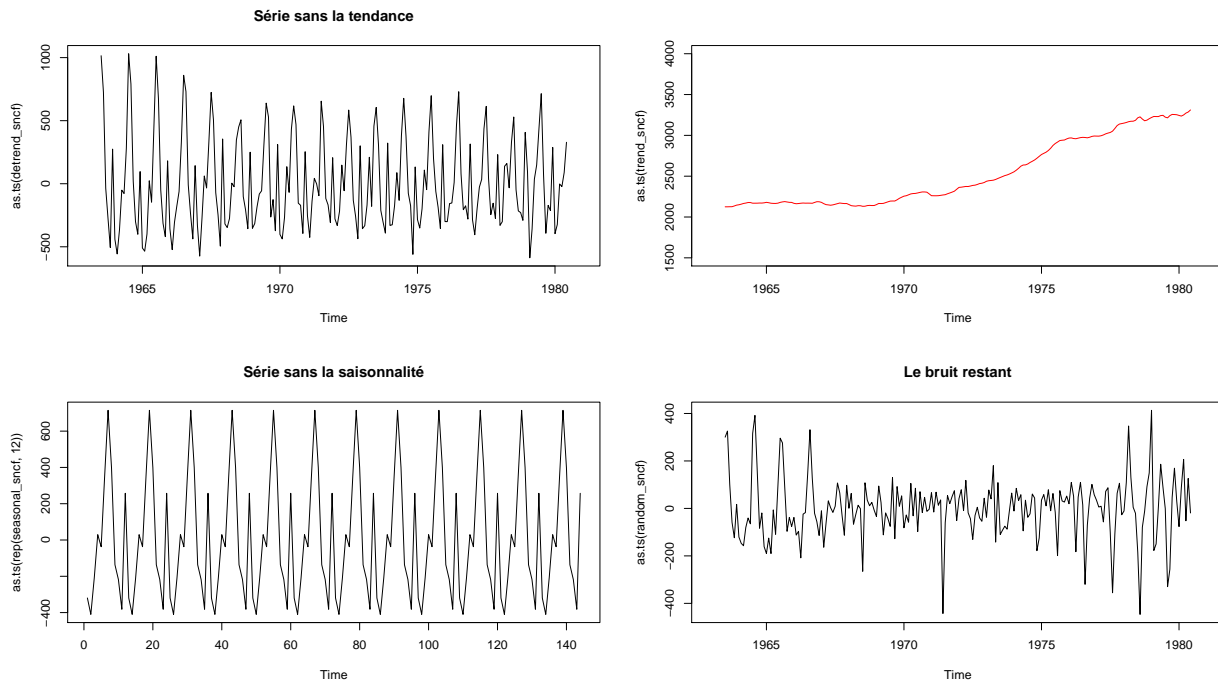
```
##  
## Call:  
## lm(formula = Traffic_SNCF ~ time(Traffic_SNCF))  
##  
## Coefficients:  
##      (Intercept)  time(Traffic_SNCF)  
##      -151842.79         78.29
```



- Estimation de la tendance par moyennes mobiles On a remarqué une saisonnalité de 12 mois (1 an) on effectue ici une moyenne mobile d'ordre 12 pour obtenir la tendance.

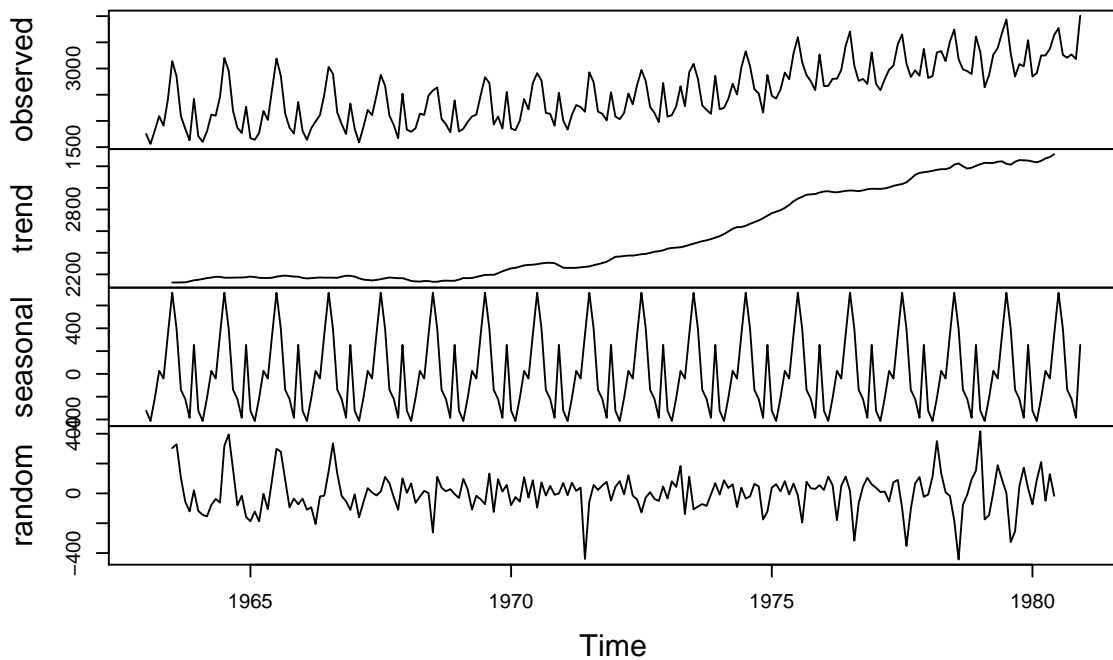


- Série décomposé : sans la tendance, sans la saisonnalité



- 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) saisonnalité et bruit.

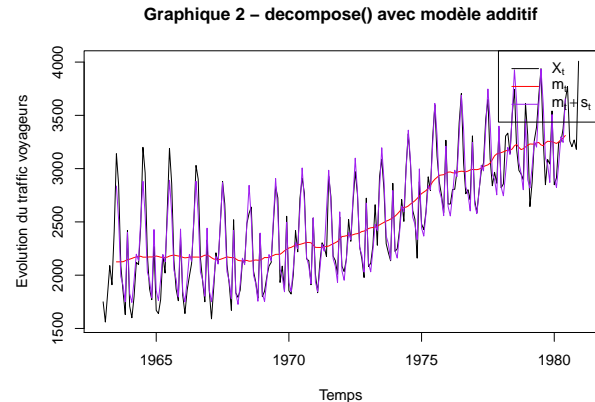
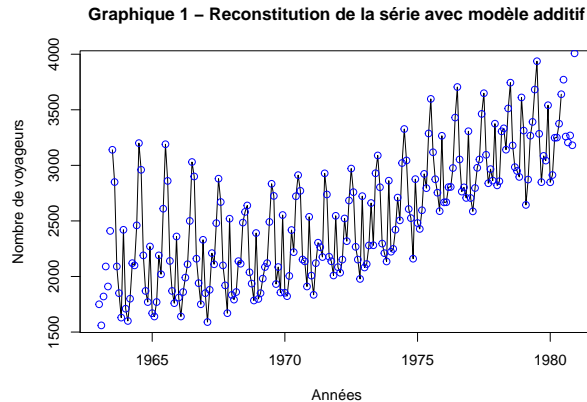
## Decomposition of additive time series



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



Pour évaluer la performance de prédiction, nous allons estimer les paramètres du modèle sur la série allant de janvier 1973 jusqu'à décembre 1977 et garder les observations de l'année 1978 pour les comparer avec les prévisions.

- Reconnaître le bon modèle - Méthode de la bande p77

## 1.2 Prédiction par lissage exponentiel

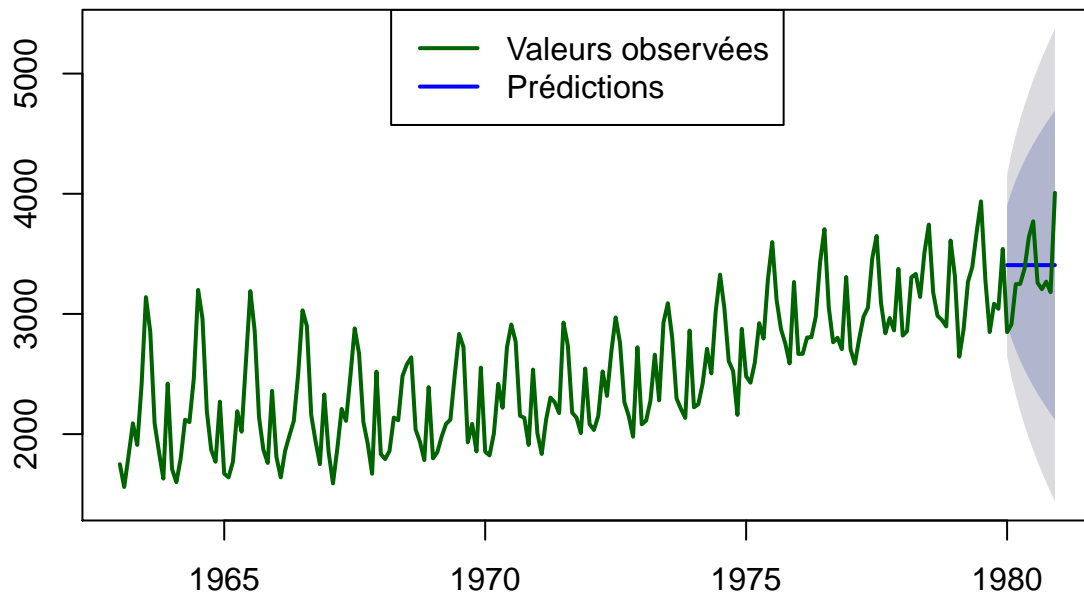
### 1.2.1 Lissage exponentiel simple

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 saisonnalité, N indique qu'il n'y en a pas.

```
## ETS(A,N,N)
##
## Call:
## ets(y = Traffic_SNCF.6379, model = "ANN")
##
## Smoothing parameters:
##   alpha = 0.726
##
## Initial states:
##   l = 1721.5989
##
## sigma: 384.9946
##
##      AIC      AICc      BIC
## 3517.804 3517.924 3527.759
##
```

```
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 11.3736 383.1027 322.424 -1.103926 13.24554 2.727351
##           ACF1
## Training set 0.07613418
```

## Forecasts from ETS(A,N,N)



Nous remarquons que le modèle est trop basique dans notre cas pour prédire à l'horizon 12 et que l'intervalle de confiance à 80%, bien que très large, ne contient pas toutes les vraies valeurs de la série. On peut utiliser `predict(fitLES,12)` pour la prédiction à horizon 1 an.

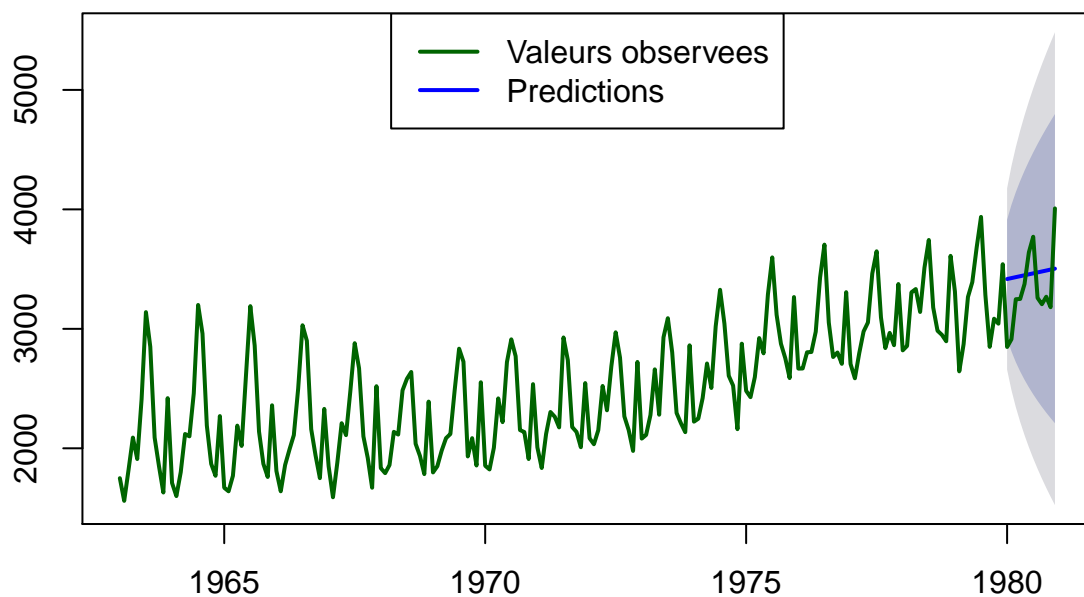
### 1.2.2 Lissage exponentiel double

```
## ETS(A,A,N)
##
## Call:
## ets(y = Traffic_SNCF.6379, model = "AAN")
##
## Smoothing parameters:
##   alpha = 0.7258
##   beta  = 1e-04
##
## Initial states:
##   l = 1933.6411
##   b = 7.9164
##
## sigma: 387.1129
```



```
##
##      AIC      AICc      BIC
## 3522.013 3522.316 3538.604
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.8968736 383.2988 323.7674 -1.641413 13.35702 2.738714
##              ACF1
## Training set 0.07743178
```

### Forecasts from ETS(A,A,N)



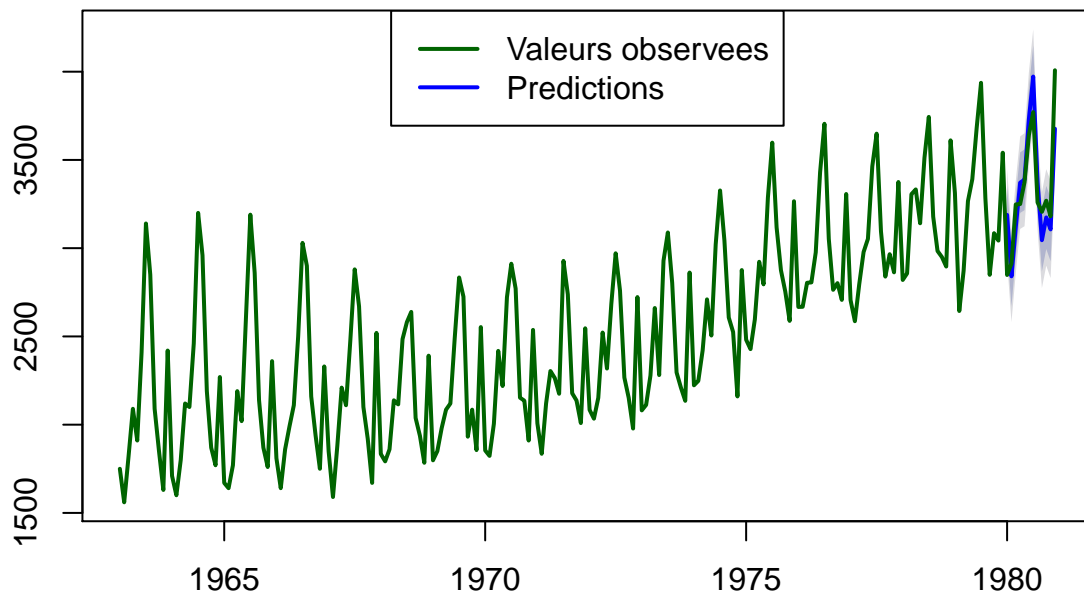
Pas vraiment d'amélioration. AIC et BIC sont plus grand.

#### 1.2.3 Méthode de Holt-Winters - Lissage exponentiel triple

```
## ETS(A,A,A)
##
## Call:
## ets(y = Traffic_SNCF.6379, model = "AAA")
##
## Smoothing parameters:
##   alpha = 0.0928
##   beta  = 0.0046
##   gamma = 0.479
##
## Initial states:
##   l = 2180.8865
```

```
##      b = 1.7331
##      s = 335.8738 -424.1515 -281.0965 -92.9194 504.7844 715.2654
##          315.7252 30.5647 20.1066 -256.943 -426.3811 -440.8287
##
##      sigma: 131.6432
##
##      AIC      AICc      BIC
## 3093.313 3096.603 3149.721
##
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 6.372573 126.3753 91.91021 0.1744532 3.681543 0.7774589
##           ACF1
## Training set 0.2434013
```

### Forecasts from ETS(A,A,A)



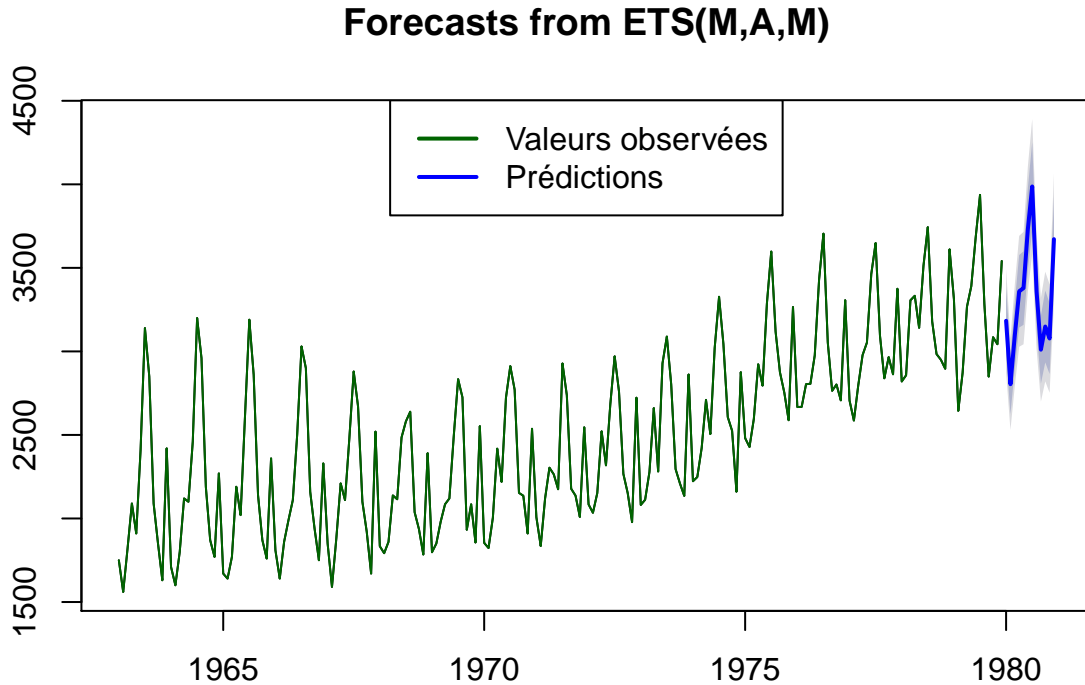
#### 1.2.4 Autres méthodes

#### 1.2.5 Procédure automatique - modèles ajustés par ets.

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.

AIC	AICc	BIC
3059.861	3063.152	3116.269

### 1.2.6 Prédiction à partir du modèle de lissage exponentiel obtenu : (M,A,M)



## 1.3 Modélisation

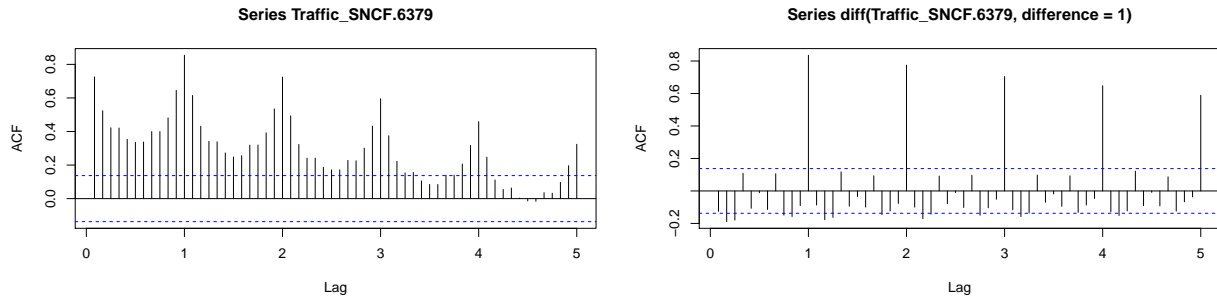
On cherche ici à modéliser la série par un processus stationnaire ARMA(p,q) ou bien SARMA(p,q). Si besoin on cherchera à stationnariser la série en utilisant l'opérateur de différentiation. On obtiendra alors une modélisation à partir de processus ARIMA (p,d,q) ou SARIMA(p,d,q)

### 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 stationnarité en traçant le corrélogramme de la série.

la valeur de  $\rho_X(h)$  varie, mais est positive et (presque) maximale en correspondance avec certaines valeurs de  $h$ . Si par exemple on observe une périodicité annuelle, lorsque  $h = 12$  dans le cas des données mensuelles, alors qu'elle est plus faible ou négative pour les autres valeurs de  $h$ . Cela signifie que les valeurs d'un instant donné ou de la période de l'année sont fortement corrélées à celles des mêmes instants ou périodes des années précédentes, de sorte que le phénomène varie au cours de chaque année et de manière similaire d'une année à l'autre.

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

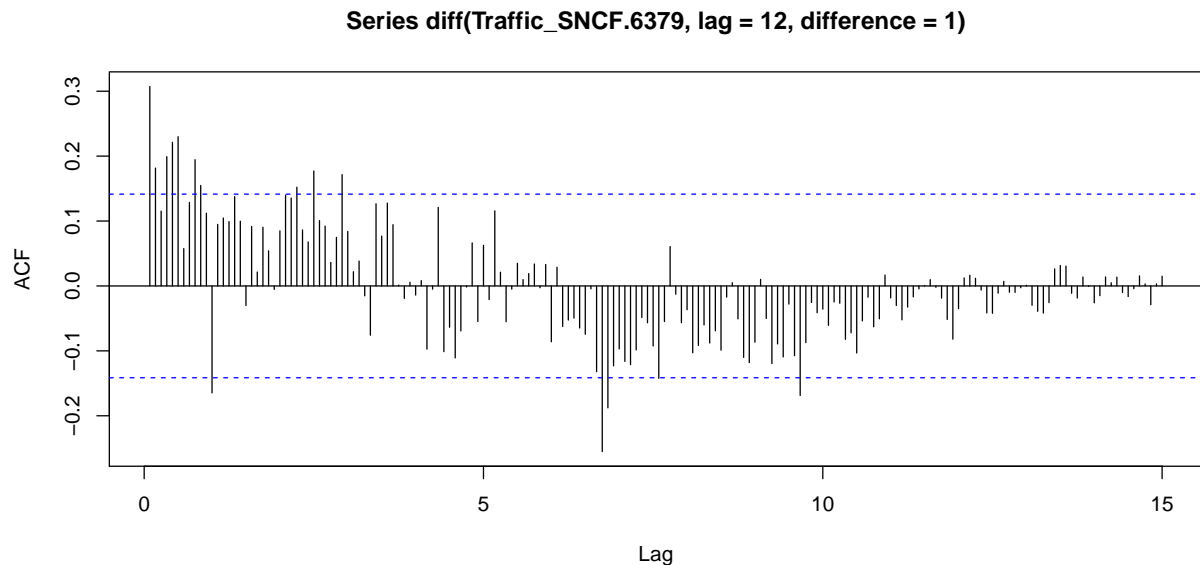


La fonction d'autocorrélation estimée est positive. On remarque une périodicité de 1 (12 mois) (graphique de gauche).

On peut essayer de différencier 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 précédemment avec le lag plot.

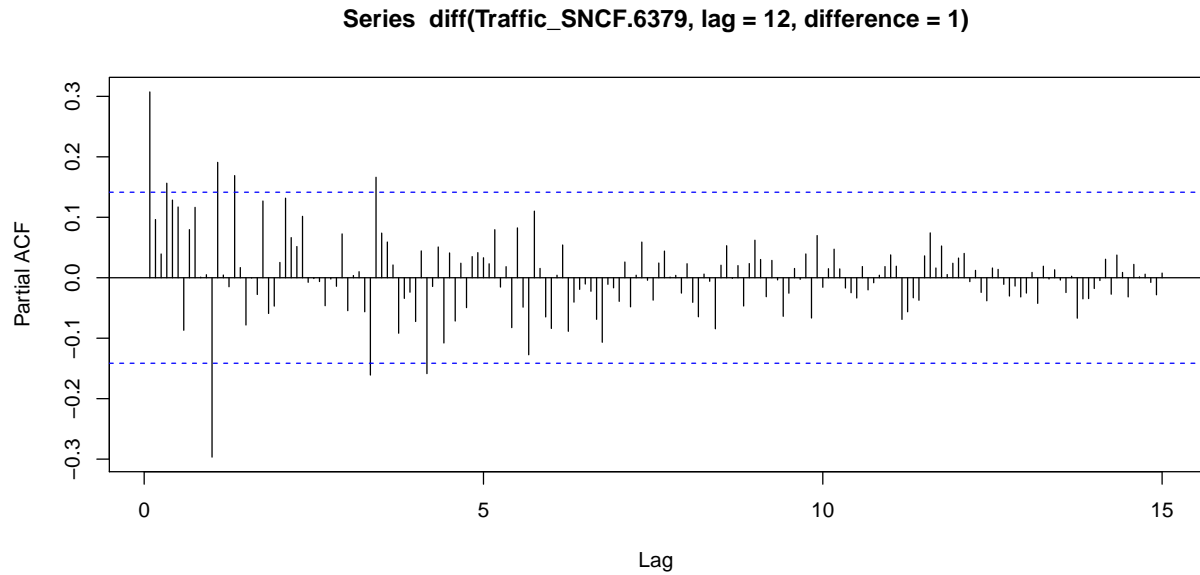
On va donc appliquer l'opérateur  $(1 - B^{12})$  à la série précédente, transformée par différenciation :  $(1 - B)X_t$

- Corrélogramme de  $(1 - B^{12})$



Le corrélogramme de la série obtenue par différenciation:  $(1 - B)(1 - B^{12})X_t$  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 transformée 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 regarde l'autocorrélation partielle pour avoir une idée du terme de degré  $q$  du terme moyen mobile  $MA(q)$  du modèle.

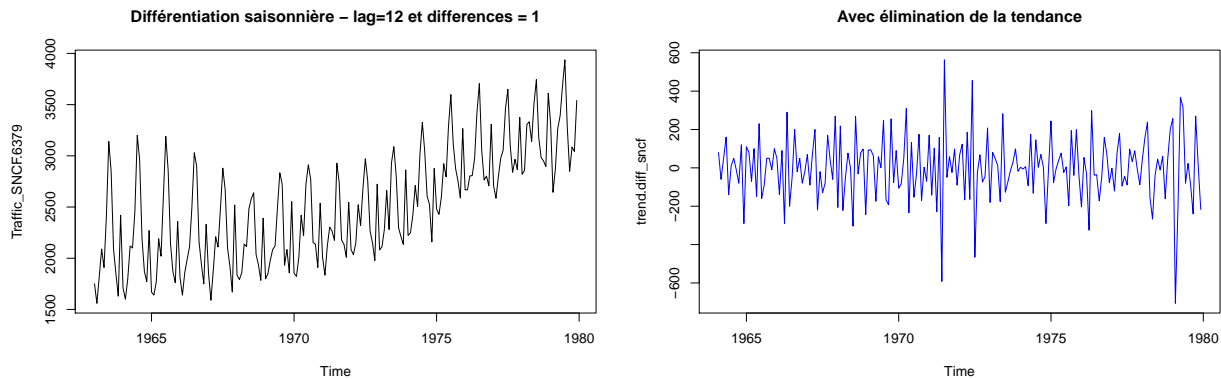


L'autocorrélation partielle suggère un terme d'ordre  $q=1$  (12ème mois) soit un terme moyenne mobile du type :  $(1 - \theta_1 B)(1 - \theta_2 B^{12})\epsilon_t$

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 et élimination de la tendance

On confirme cette hypothèse à l'aide d'un test de *Dickey – Fuller* on obtient une  $p\_value=0.01$

Et le test du Portmanteau ou test de blancheur sur R est lui aussi concluant et donne une  $p\_value=1.1473045 \times 10^{-12}$

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.

SARIMA	AIC	AICc	BIC
010_011	2561.9659697	2562.0345412	2568.3295368
010_010	2702.1440097	2702.166737	2705.3257933
010_110	2609.1956787	2609.2642501	2615.5592458
010_111	2533.4001704	2533.5381015	2542.9455211
011_011	2453.2494647	2453.378497	2462.9905369
011_010	2520.3334234	2520.4019948	2526.6969905
011_110	2433.7094859	2433.8474169	2443.2548365
011_111	2360.2735629	2360.5047768	2373.0006971
110_011	2474.0090396	2474.1469707	2483.5543903
110_010	2602.8286469	2602.8972183	2609.192214
110_110	2518.1824904	2518.3204215	2527.7278411
110_111	2444.5382951	2444.7695089	2457.2654293
111_011	2357.7012056	2357.9324194	2370.4283398
111_010	2482.0405757	2482.1785068	2491.5859264
111_110	2403.9671832	2404.198397	2416.6943174
111_111	2329.9176024	2330.2664396	2345.8265201
112_011	2317.0145286	2317.3633658	2332.9234463
112_010	2441.0490063	2441.2802202	2453.7761405
112_110	2365.0778833	2365.4267205	2380.9868011
112_111	2291.5359338	2292.0271619	2310.6266351
211_011	2348.221645	2348.5704822	2364.1305628
211_010	2473.3487972	2473.5800111	2486.0759314
211_110	2397.675979	2398.0248162	2413.5848968
211_111	2321.559769	2322.0509971	2340.6504703

En terme de minimisation des critères AIC, AICc et BIC les 3 meilleurs modèles sont les modèles : SARIMA(1,1,2)(0,1,1), SARIMA(2,1,1)(1,1,1), SARIMA(1,1,1)(1,1,1) et le modèle initial SARIMA(0,1,1)(0,1,1)

### 1.3.2 Validation des modèles SARIMA obtenus

```
##          ar1          ma1      ma2          sma1
## t.stat 1.731819 -31.64589 15.51 -15.97807
## p.val  0.083306  0.00000  0.00  0.00000

##          ar1          ar2          ma1          sma1
## t.stat -7.073639 -3.449901 -45.81864 -15.41762
## p.val  0.000000  0.000561  0.00000  0.00000

##          ar1          ma1          sar1          sma1
## t.stat -5.975373 -44.66276 -6.002373 -11.33519
## p.val  0.000000  0.00000  0.000000  0.00000

##          ma1          sma1
## t.stat -66.08548 -8.426866
## p.val  0.00000  0.000000

##          ar1          ma1          ma2          sma1
## ar1  1.000000e+00 -4.583318e-01  5.005176e-01  5.915924e-07
## ma1 -4.583318e-01  1.000000e+00 -9.951144e-01 -2.238240e-06
```

```

## ma2    5.005176e-01 -9.951144e-01  1.000000e+00  2.215732e-06
## sma1    5.915924e-07 -2.238240e-06  2.215732e-06  1.000000e+00

##          ar1          ar2          ma1          sma1
## ar1    1.000000e+00  4.163572e-01 -5.163373e-06 -5.264744e-06
## ar2    4.163572e-01  1.000000e+00 -4.485734e-06  3.488872e-06
## ma1   -5.163373e-06 -4.485734e-06  1.000000e+00  1.684997e-07
## sma1  -5.264744e-06  3.488872e-06  1.684997e-07  1.000000e+00

##          ar1          ma1          sar1          sma1
## ar1    1.000000e+00 -3.610557e-07  4.683976e-02  1.343425e-06
## ma1   -3.610557e-07  1.000000e+00 -3.046597e-07  3.228988e-07
## sar1    4.683976e-02 -3.046597e-07  1.000000e+00 -3.061937e-05
## sma1    1.343425e-06  3.228988e-07 -3.061937e-05  1.000000e+00

##          ma1          sma1
## ma1    1.000000e+00 -5.602926e-06
## sma1   -5.602926e-06  1.000000e+00

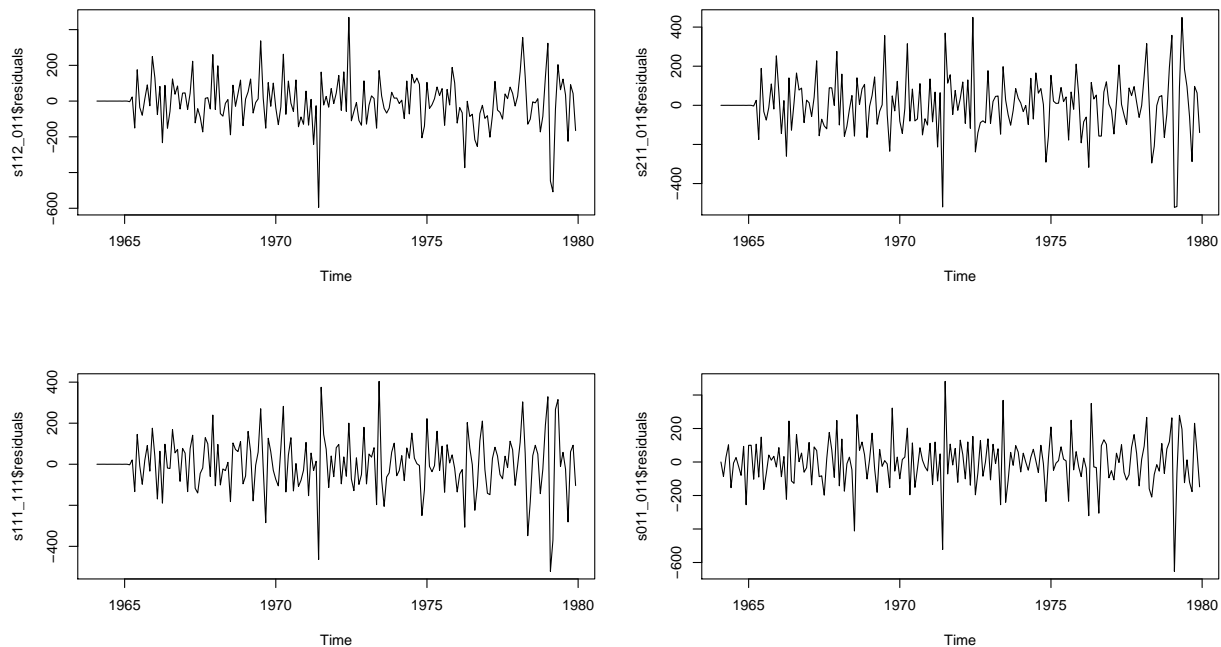
##
## Box-Pierce test
##
## data:  s112_011$residuals
## X-squared = 41.503, df = 20, p-value = 0.003209

##
## Box-Pierce test
##
## data:  s211_011$residuals
## X-squared = 72.877, df = 20, p-value = 6.133e-08

##
## Box-Pierce test
##
## data:  s111_111$residuals
## X-squared = 54.268, df = 20, p-value = 5.278e-05

##
## Box-Pierce test
##
## data:  s011_011$residuals
## X-squared = 48.773, df = 20, p-value = 0.000331

```



## 1.4 Modélisation automatique avec R

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

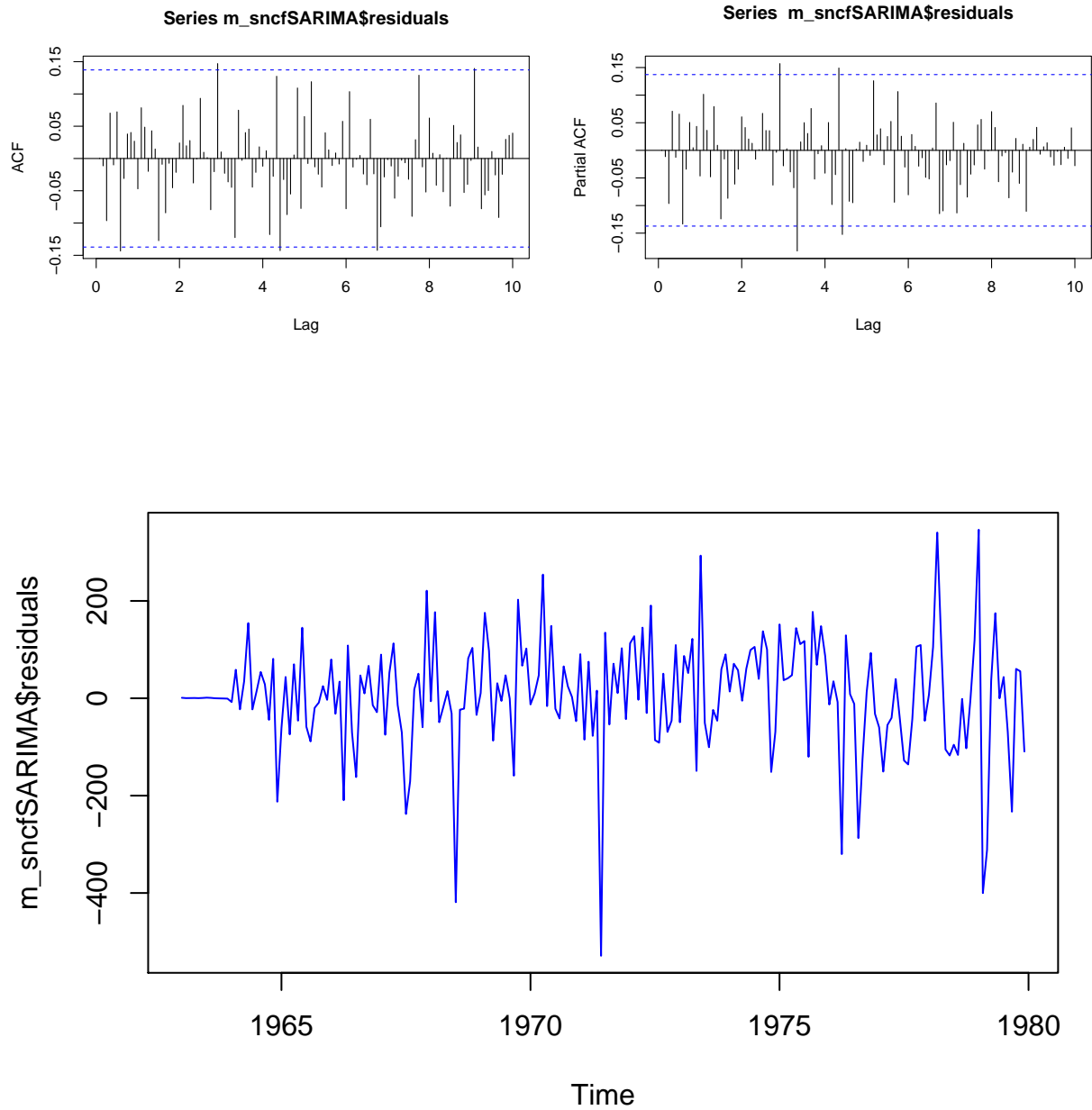
```
## Series: Traffic_SNCF.6379
## ARIMA(1,1,1)(0,1,1)[12]
##
## Coefficients:
##      ar1      ma1      sma1
##    0.2021 -0.8907 -0.4985
## s.e. 0.0828 0.0375 0.0669
##
## sigma^2 estimated as 14801: log likelihood=-1189
## AIC=2386  AICc=2386.21  BIC=2399.01

##      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

##
## Box-Pierce test
##
## data: m_sncfSARIMA$residuals
## X-squared = 16.639, df = 20, p-value = 0.6762
```



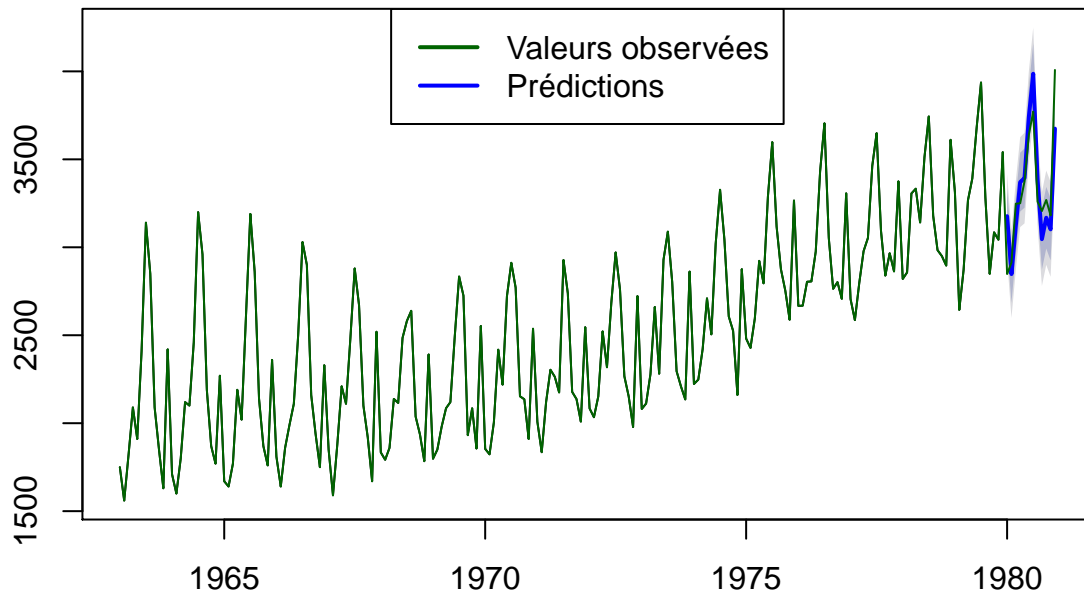


## 1.5 Prévisions et comparaison des modèles obtenus

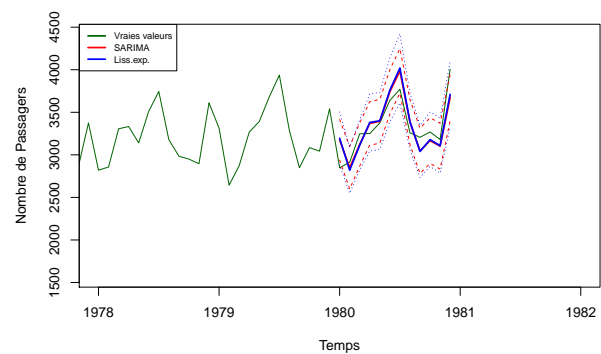
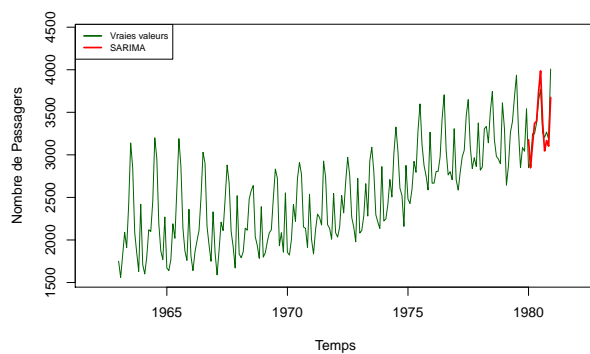
##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Jan 1980	3176.918	3021.002	3332.833	2938.466	3415.369
## Feb 1980	2847.791	2684.491	3011.092	2598.045	3097.538
## Mar 1980	3122.217	2956.723	3287.710	2869.116	3375.317
## Apr 1980	3370.292	3203.281	3537.303	3114.871	3625.713
## May 1980	3393.392	3224.993	3561.791	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

### Forecasts from ARIMA(1,1,1)(0,1,1)[12]



\* Comparaison entre SARIMA et lissage exponentiel



- Remarque : la stabilisation de la variance en utilisant la fonction log n'apporte pas l'atténuation souhaitée, 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 action. En fait un portefeuille d'actions, ou indice type CAC40, DAX, Eurostoxx, SP500... Pour cette première étude on va se baser sur l'indice CAC40. A partir 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 nous guider dans la détermination de stress test de type action. Pour être 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 être pas appropriées. Plus précisément 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 existent.
- Elaboration d'un portefeuille simplifié : Une fois les principaux indices modélisés, 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'évolution du prix de l'indice étudié le CAC40 : représentation 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 partir des graphiques ACF et PACF. Enfin on ajustera les coefficients pour obtenir notre modèle. On terminera l'étude en validant le modèle: blancheur des résidus, indépendance, 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 analysé ces séries on essaiera de construire un modèle de type ARIMA pour chacune d'elles.

```
dataCAC40_raw_d <-read.table("Daily_Data_CAC40_1997-2019.csv", sep=";", dec=".",header=T, na.strings = "  
dataCAC40_raw_m <-read.table("Monthly_Data_CAC40_1997-2019.csv", sep=";", dec=".",header=T, na.strings
```

#### 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 convertit 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 on 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 paramètre de fréquence (nb jours dans l'année) la valeur 256 Ce qui correspond au nombre de jours par an (jours ouvrés 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*)

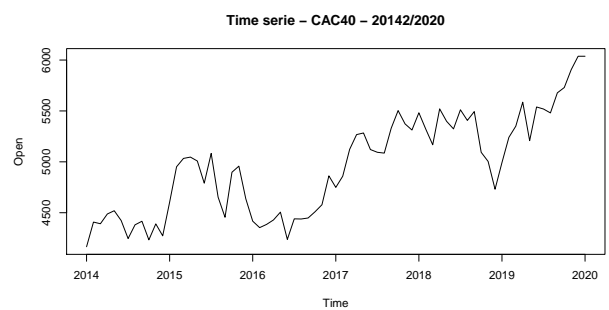
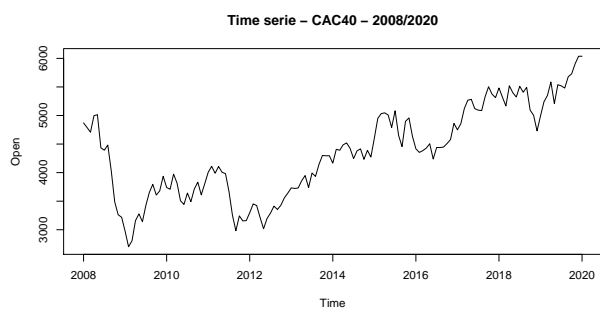
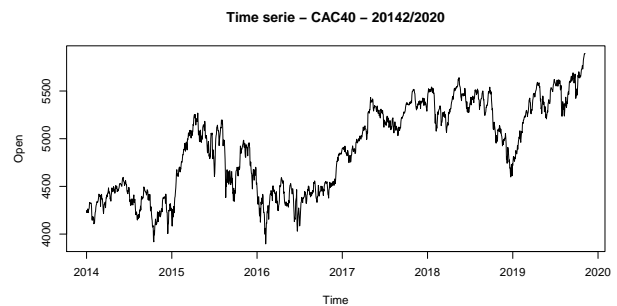
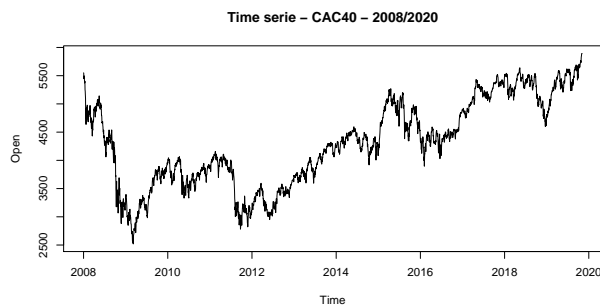
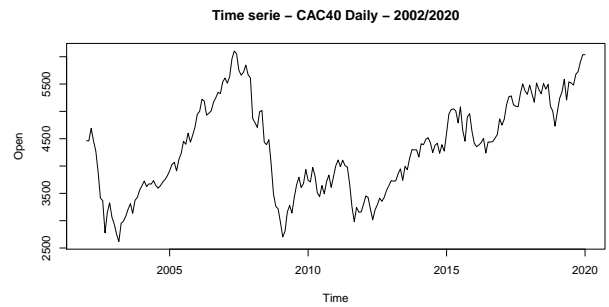
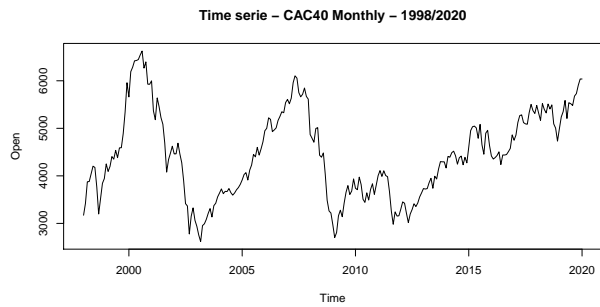
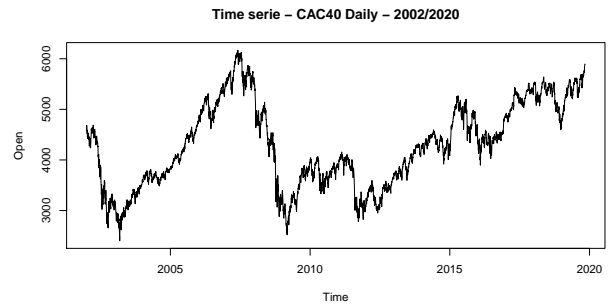
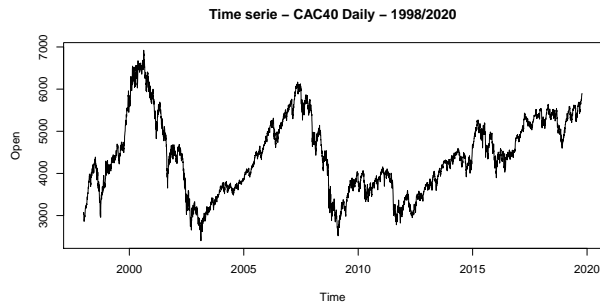
## 2.3 Analyse des séries temporelles obtenues

Comme déjà énoncé on va étudier plusieurs profondeurs 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 considère 2 jeux de données, avec une fréquence quotidienne et mensuelle.

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

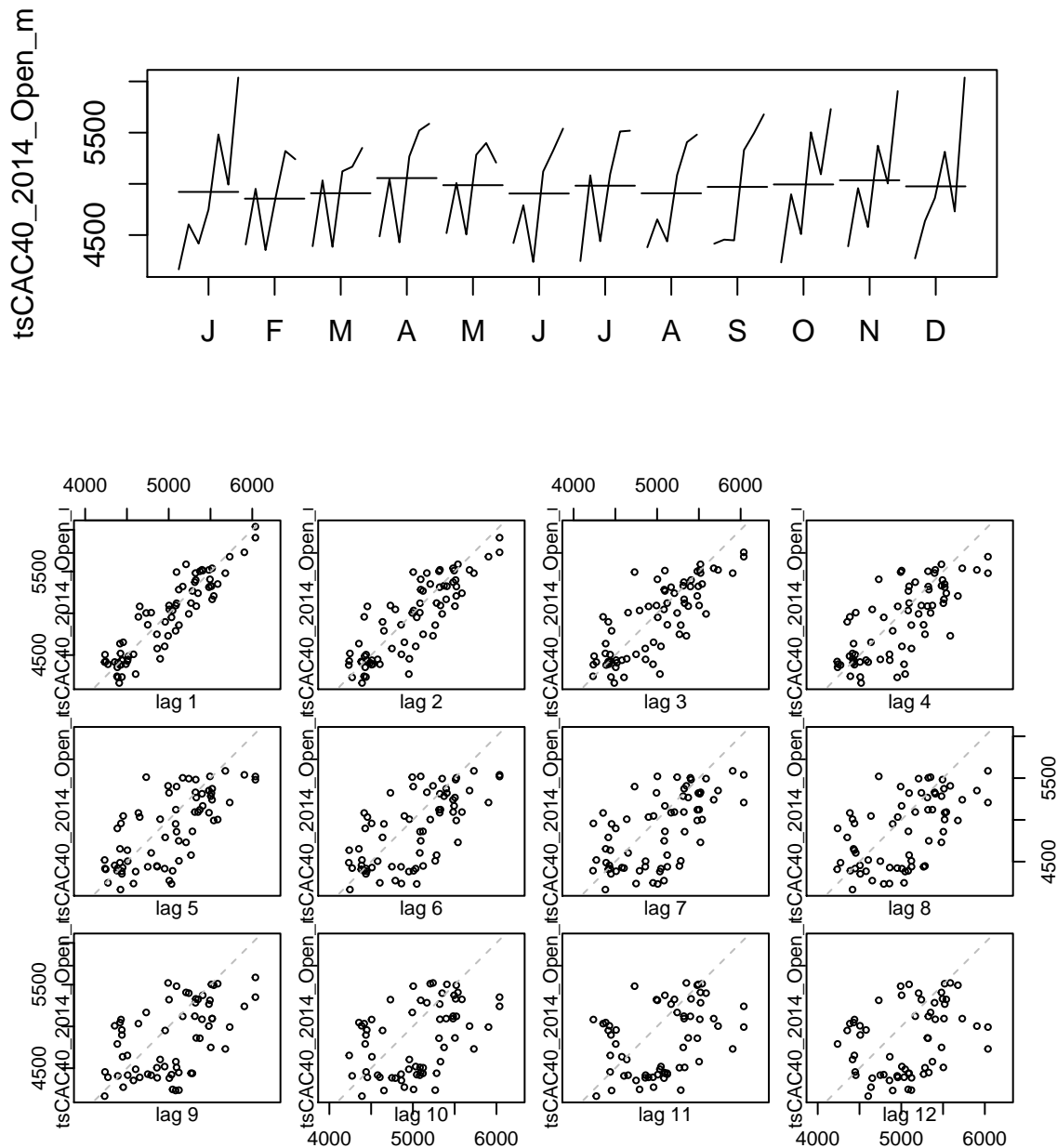
On a 4 séries temporelles possibles en fonction du choix de la quantité observée (High, Low, Open, Close, Volume). On va s'intéresser à la valeur à la fermeture pour la cotation de l'indice CAC40 (Close).



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

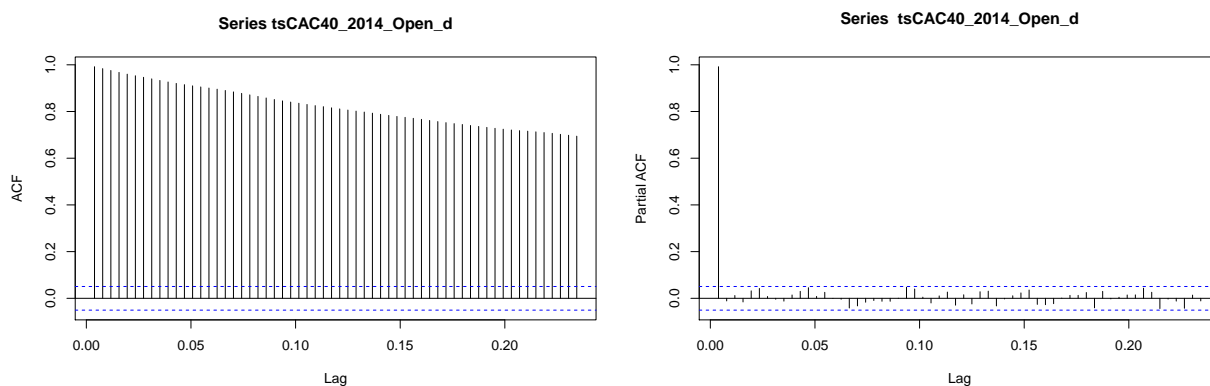
Si le diagramme retardée suggère une corrélation entre les deux séries, on dit que la série présente une autocorrélation d'ordre  $k$ . Ce diagramme permet de comprendre la dépendance de la série par rapport à son passé. Il donne une vision locale de la série, si y a une corrélation entre la série à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



### 2.3.3 Etude de la stationarité

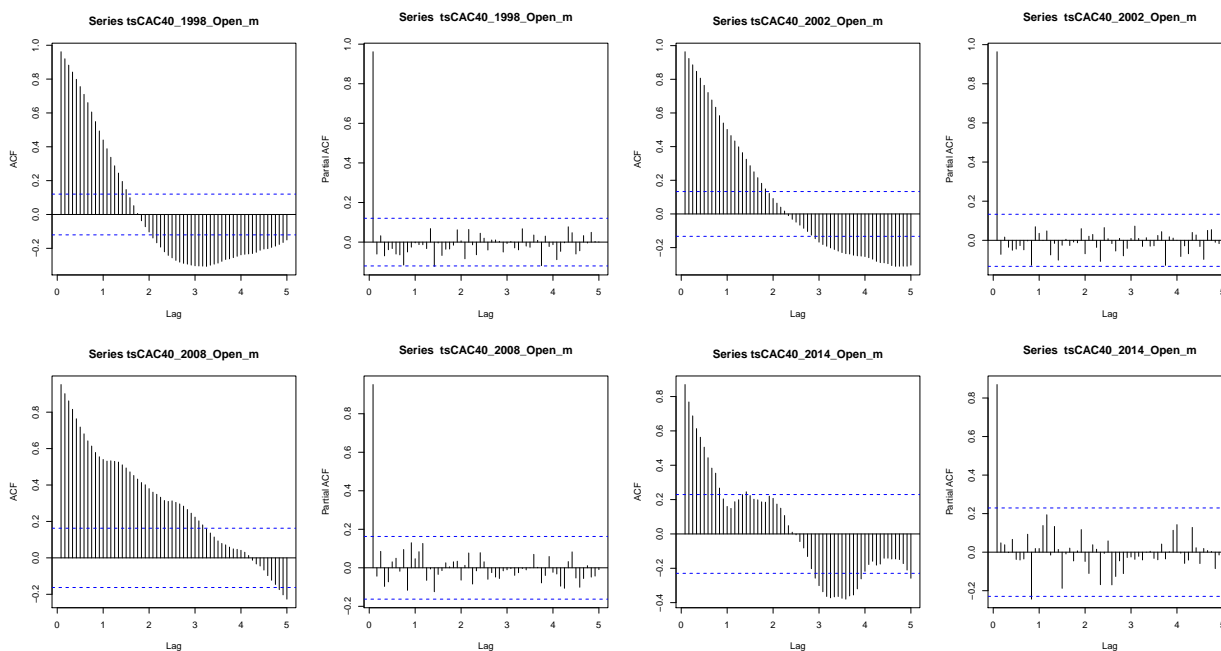
La stationnarité est la stationnarité du processus au sens faible. Un tel processus doit avoir les propriétés suivantes : La moyenne et la variance ne varient pas au cours du temps et le processus n'a pas de tendance. Pour vérifier ces hypothèses, on s'appuiera sur une analyse des graphiques d'autocorrélation 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

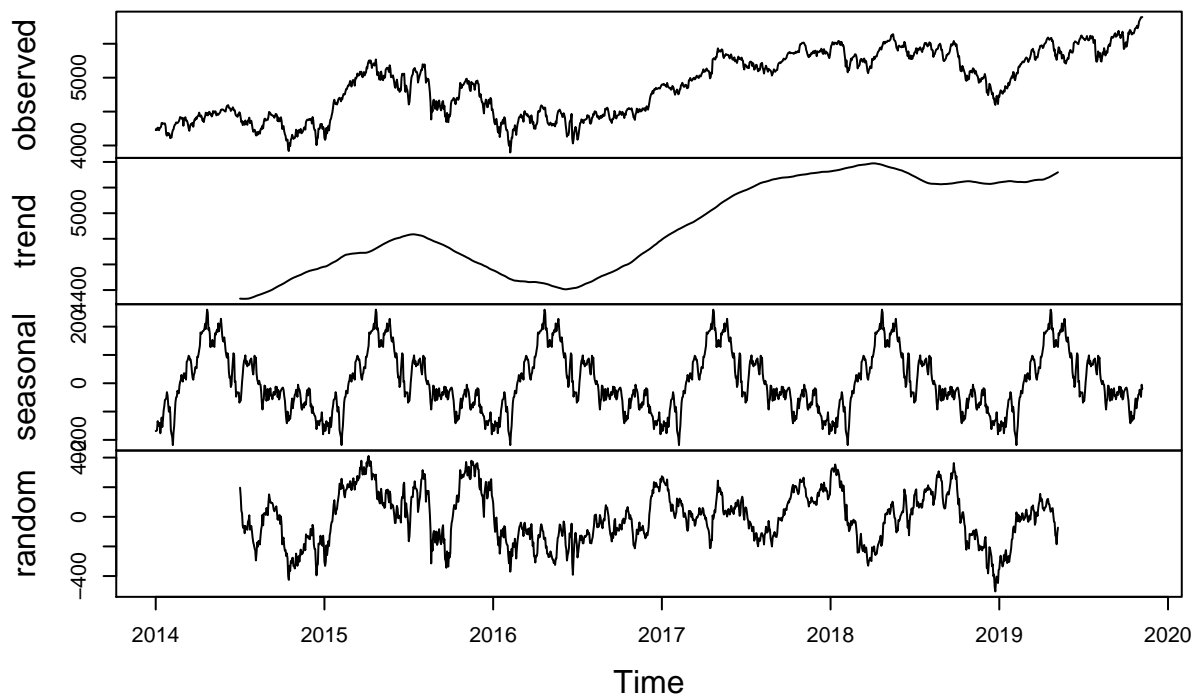
série	Dickey-Fuller	Lag order	p-value
2008	-4.2978772	5	0.01
2014	-2.1515395	4	0.5136323

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

#### 2.3.4 Décomposition 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) saisonnalité et bruit.

### Decomposition of additive time series



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 modélisation 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 CAC40 par un processus ARIMA. On commence donc par différencier les séries. Le facteur utilisé est de 1.



## 2.4 Détermination des modèles ARIMA

Après avoir rappelé la définition d'un processus ARIMA(p,d,q) et ARMA(p,q) on cherchera à déterminer les paramètres p et q.

### 2.4.1 Définitions

Un processus  $X_t$   $t \in \mathbb{Z}$  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 des processus stationnaires. Ces processus sont composés des processus auto-régressifs AR(p) et de moyennes mobiles ("moving average") MA(q).

- Processus AR(p) - Processus auto régressif d'ordre p

$$\forall t = 1, \dots, n \quad X_t = \beta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t \text{ (où } \varepsilon_t \text{ est un bruit blanc)}$$

Avec l'opérateur retard cette équation se réécrit :

$$\forall t = 1, \dots, n \quad X_t = \beta + \varepsilon_t + (\alpha_1 B^1 + \alpha_2 B^2 + \dots + \alpha_q B^q) X_t$$

La fonction d'autocorrélation (ACF) d'un AR(p) montre une décroissance exponentielle avec ou sans oscillations vers 0. La fonction d'autocorrélation partielle (PACF) d'un AR(p) est nulle à partir de l'ordre p+1.

- Processus MA(q) - Processus auto régressif d'ordre q

Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations d'un bruit blanc.

$$\forall t = 1, \dots, n \quad X_t = \beta + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \dots + \alpha_q \varepsilon_{t-q} \text{ (où les } \varepsilon_t \text{ sont des bruits blanc centré)}$$

Avec l'opérateur retard cette équation se réécrit :

$$\forall t = 1, \dots, n \quad X_t = \mu + (\alpha_1 B^1 + \alpha_2 B^2 + \dots + \alpha_q B^q) \varepsilon_t$$

Un processus MA(q) est toujours stationnaire quelles que soient les valeurs des  $\alpha_i$ , il est de plus de moyenne  $\mu$ . L'ACF d'un processus MA(q) 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).

Un processus ARMA est la combinaison des processus autorégressifs et moyennes mobiles. Ainsi, avec les notations précédentes,  $(X_t)$  est un processus ARMA(p,q), si :

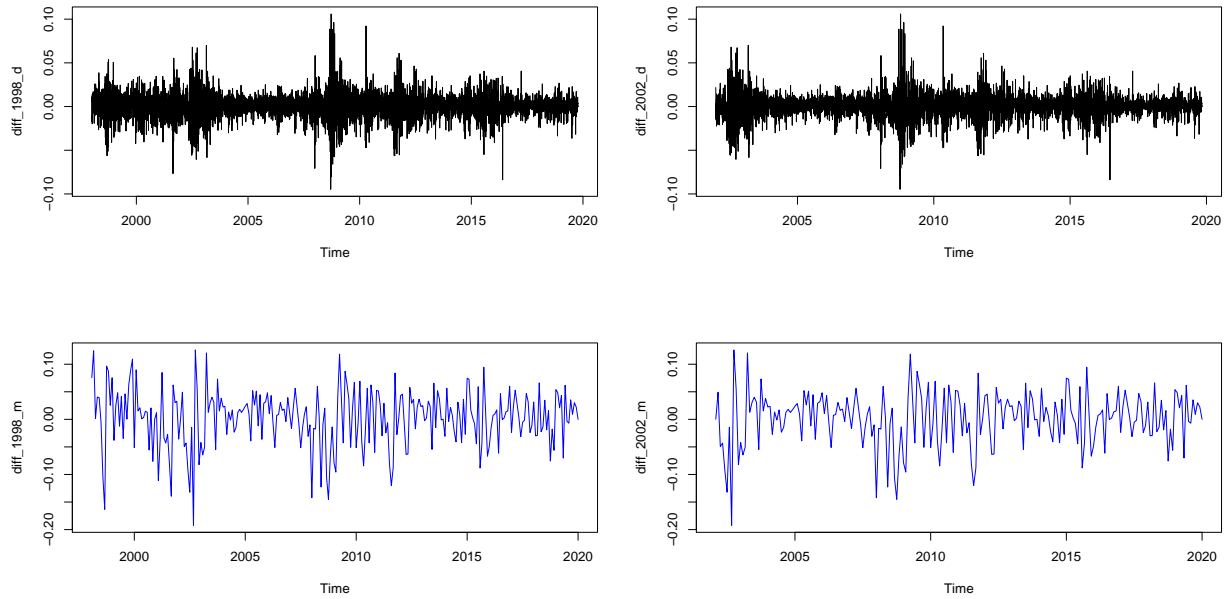
$$\forall t = 1, \dots, n \quad X_t = \beta + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \dots + \alpha_q \varepsilon_{t-q} \text{ (où les } \varepsilon_t \text{ sont des bruits blanc centré)}$$

On cherche maintenant à se ramener à un processus stationnaire 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).

### 2.4.2 Stationarisation des processus par différentiation des séries

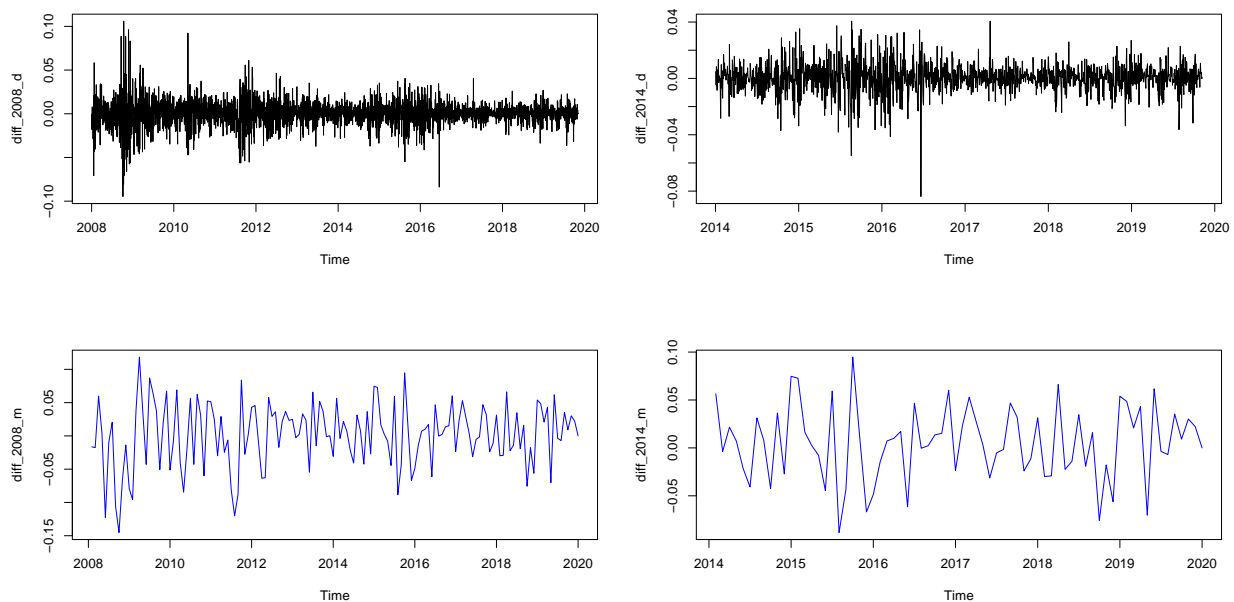
Pour tenter de rendre la série stationnaire, on applique la méthode de différentiation. On utilise la fonction R diff. En paramètre on passe un facteur de 1 pour la différence et de 0 pour la saisonnalité. On retrouve l'idée de la transformation Log-return  $R_t$  des prix  $X_t$  où  $R_t = \text{Log}(X_t/X_{t-1})$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

- Séries obtenues par différentiation 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 constante dans le temps. Ainsi qu'aussi une variance constante bien que cet aspect soit moins évident. On affaiblira cette dernière hypothèse lors de l'étude des modèles GARCH au dernier paragraphe.

#### 2.4.2.1 Séries obtenues par différentiation pour les 2 jeux de données quotidien et mensuel à partir de 2008 et 2014



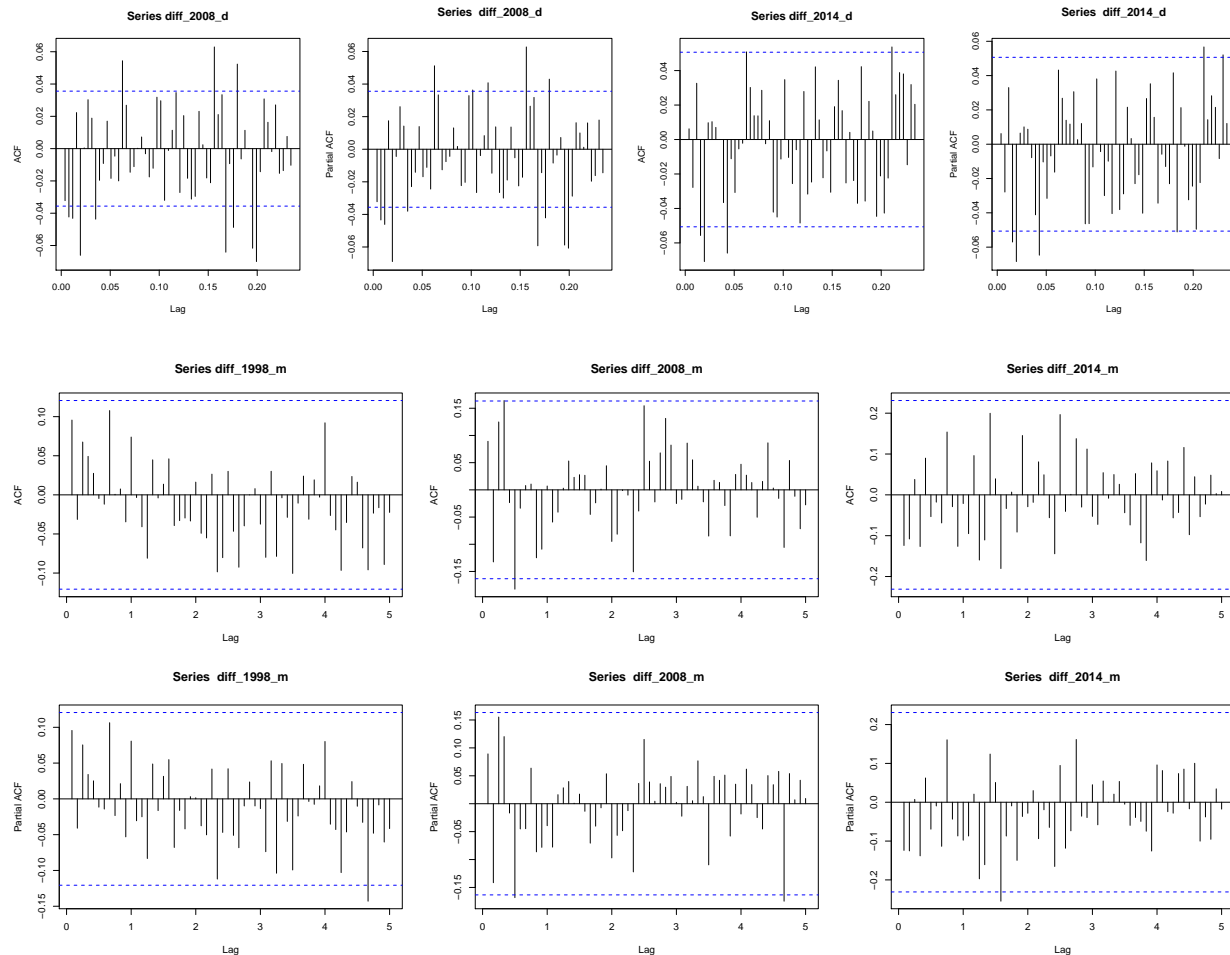
L'hypothèse de stationnarité du processus ainsi transformé, étant vérifié on peut envisager une modélisation ARIMA. On passe maintenant à la détermination des paramètres  $p$  et  $q$  du modèle ARMA( $p,q$ ).

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  $\alpha_i$  et  $\alpha_i$  se font par minimisation / regression

### 2.4.3 Détermination des paramètres $p$ et $q$ , études des corrélogrammes et autocorrélations partiels

The next step is to select appropriate ARIMA model, which means finding the most appropriate values of  $p$  and  $q$  for an ARIMA( $p,d,q$ ) model. You usually need to examine the correlogram and partial correlogram of the stationary time series for this. To plot a correlogram and partial correlogram, we can use the `acf()` and `pacf()` functions in R, respectively.



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 modèle 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.

#### 2.4.4 Méthode automatique de calibration d'un modèles ARIMA

R provides a function `auto.arima`, which returns best ARIMA model according to either AIC, AICc or BIC value. The function conducts a search over possible model within the order constraints provided.

Pour les 2 jeux de données à partir de 2008 on obtient un SARIMA(0,1,0)(1,0,0) dans le cas des données quotidiennes et un SARIMA(1,1,0)(2,0,2) pour les données mensuelles.

```
## Series: tsCAC40_2008_Open_m
## ARIMA(0,1,0)
##
## sigma^2 estimated as 39166: log likelihood=-965.77
## AIC=1933.53 AICc=1933.56 BIC=1936.5
```

Pour les 2 jeux de données à partir de 2014 on obtient un ARIMA(0,1,0) soit une marche aléatoire.

```
## Series: tsCAC40_2014_Open_d
## ARIMA(0,1,0)
##
## sigma^2 estimated as 2486: log likelihood=-7976.35
## AIC=15954.7 AICc=15954.7 BIC=15960.01
```

```
## Series: tsCAC40_2014_Open_m
## ARIMA(0,1,0)
##
## sigma^2 estimated as 37198: log likelihood=-481.03
## AIC=964.06 AICc=964.11 BIC=966.33
```

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
010	$1.6959674 \times 10^4$	$1.6959674 \times 10^4$	$1.6959674 \times 10^4$
110	$1.6571636 \times 10^4$	$1.6571636 \times 10^4$	$1.6571636 \times 10^4$
011	$1.5953606 \times 10^4$	$1.5953606 \times 10^4$	$1.5953606 \times 10^4$
111	$1.5955273 \times 10^4$	$1.5955273 \times 10^4$	$1.5955273 \times 10^4$
012	$1.5955254 \times 10^4$	$1.5955254 \times 10^4$	$1.5955254 \times 10^4$
112	$1.5955838 \times 10^4$	$1.5955838 \times 10^4$	$1.5955838 \times 10^4$
210	$1.6364777 \times 10^4$	$1.6364777 \times 10^4$	$1.6364777 \times 10^4$
211	$1.5956244 \times 10^4$	$1.5956244 \times 10^4$	$1.5956244 \times 10^4$
212	$1.5959274 \times 10^4$	$1.5959274 \times 10^4$	$1.5959274 \times 10^4$

On remarque que selon ce critère plusieurs modèles sont très proches : ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(0,1,2), ARIMA(1,1,1), ARIMA(1,1,2), ARIMA(2,1,1), ARIMA(2,1,2)

## 2.5 Validation des modèles obtenus

### 2.5.1 Blancheur des résidus

*Test de Box-Pierce*

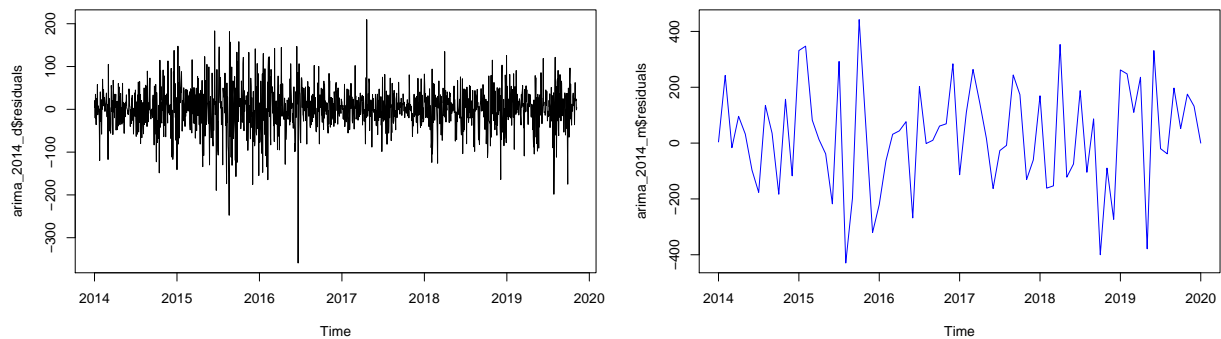
série	X-squared	df	p-value
1998	12.6097304	20	0.8934941
2008	14.8918527	20	0.782562
2014	17.2944661	20	0.6337782

*Test de Ljung-Box*

série	X-squared	df	p-value
1998	13.2677654	20	0.865602
2008	15.8861252	20	0.7236633
2014	21.4096029	20	0.3733861

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: 0.0934832. La profondeur de l'historique et la forme de la courbe associé semble aussi jouer un rôle.

### 2.5.2 Graphiques de résidus obtenus à partir des différents modèles



Dans le cas des données quotidiennes on a une variance non constante. On va utiliser les modèles GARCH comme alternative. On verra qu'ils sont mieux adaptées à la modélisation des séries financières.

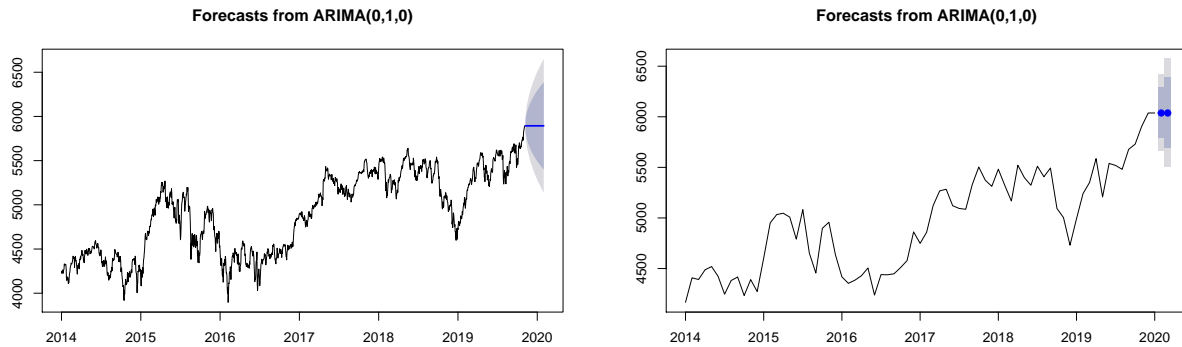
### 2.5.3 Normalité des résidus

*Test de Shapiro-Wilk normality test*

série	p-value
1998	0.0011782
2008	0.0996264
2014	0.8733718

#### 2.5.4 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 nous aider à déterminer un scénario économique possible sur les actions. Vu que l'on est dans le cadre de stress tests on cherche à déterminer un choc absolu plausible et non pas obtenir la valeur du CAC à horizon.



Les erreurs de prédictions obtenues semblent suivre une normale centrée et de variance assez constante. Le modèle ARIMA semble être adapté 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 prennent en compte l'hétéroscédasticité. Ils sont mieux adaptés aux séries financières.

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

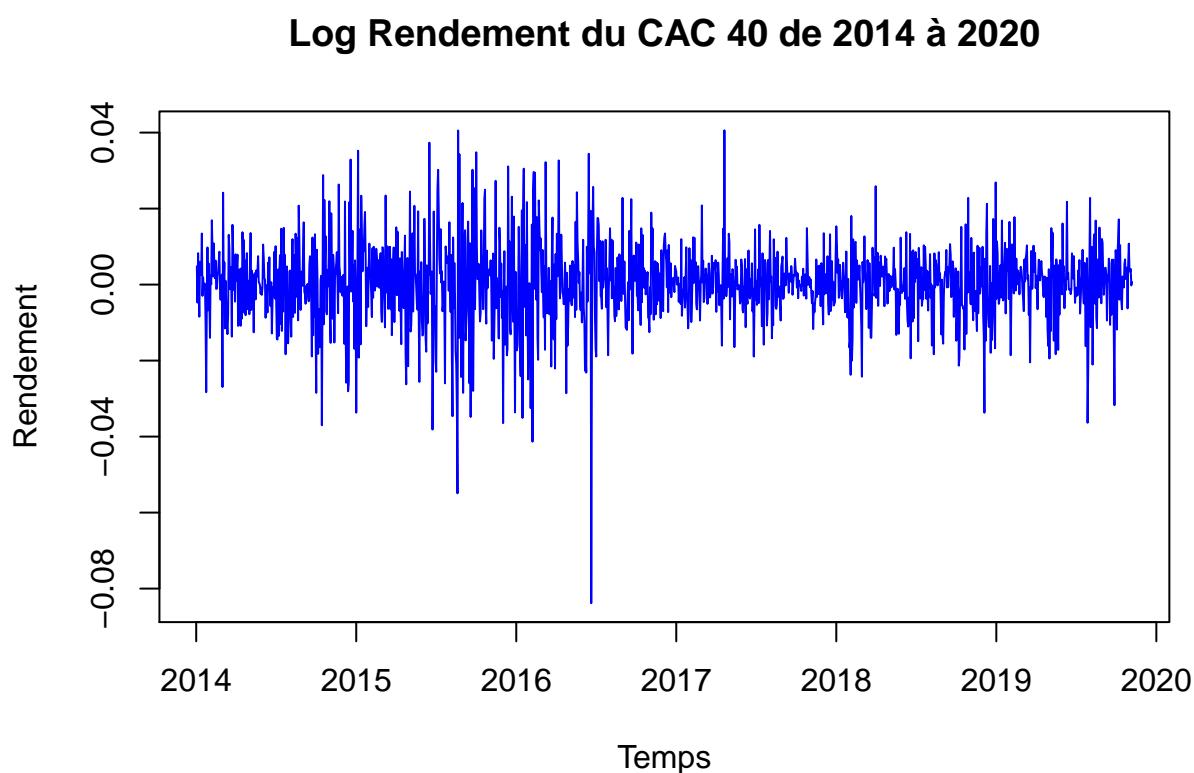
Comme on l'a fait dans le cas de la modélisation ARIMA on transforme la série originale par différenciation. Pour obtenir une série stationnaire. Ici on va considérer comme c'est souvent le cas en finance le log-return. C'est à dire la quantité déduite du prix  $X_t$  de la manière suivante : Log-return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

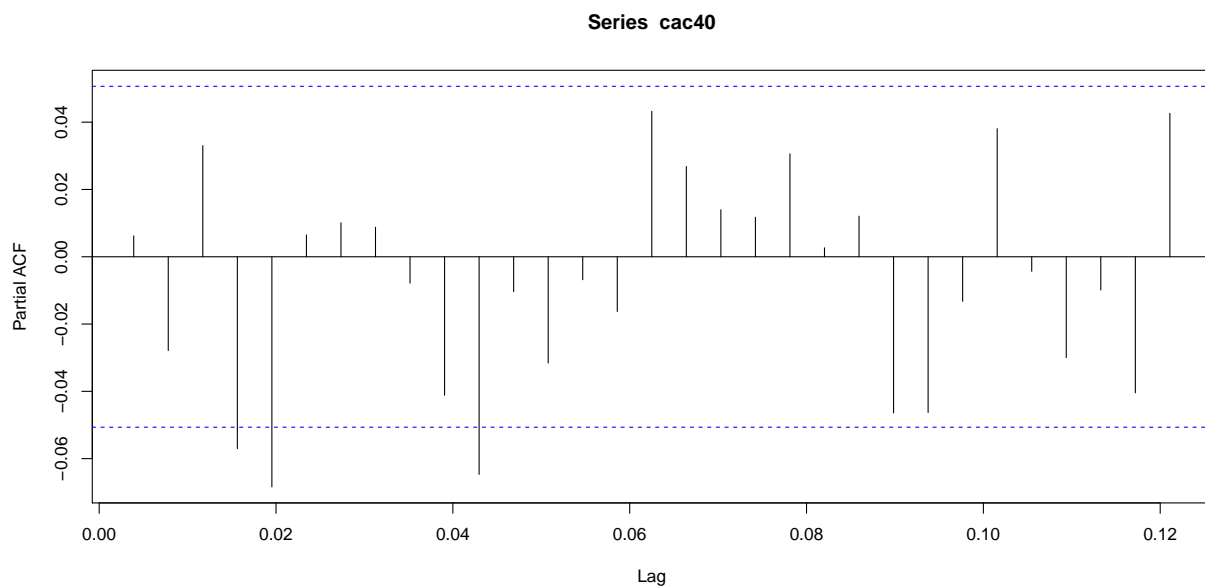
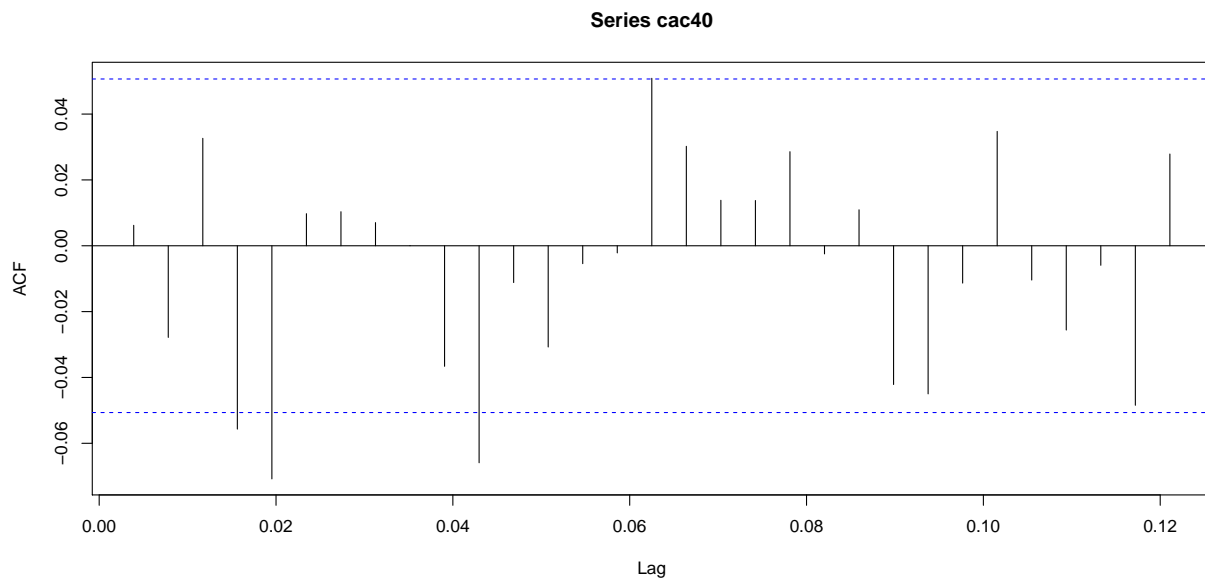
D'après D3 page 96 : Au regard de l'autocorrélogramme partiel du log return, une modélisation à l'aide d'un modèle GARCH(1, 1) semblerait possible. En effet, l'autocorrélogramme et l'autocorrélogramme partiel sont significativement nuls à partir des premiers retards (i.e.  $p = q = 1$ ) Dans notre cas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

- Obtention de la série des log return

```
cac40 <- diff(log(tsCAC40_2014_Open_d))
```

- Graphique de la série obtenue





```
##
## ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##
##      I      INITIAL X(I)      D(I)
##
##      1      1.013071e-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 -6.076e+03
##      1   8 -6.077e+03  1.41e-04  2.65e-04  4.9e-05  6.7e+10  4.9e-06  8.85e+06
##      2   9 -6.077e+03  4.38e-07  4.64e-07  4.9e-05  2.0e+00  4.9e-06  9.34e+00
```



```

##      3      17 -6.095e+03  3.06e-03  4.45e-03  4.5e-01  2.0e+00  8.0e-02  9.30e+00
##      4      20 -6.130e+03  5.64e-03  4.20e-03  7.4e-01  1.9e+00  3.2e-01  5.37e-01
##      5      22 -6.140e+03  1.66e-03  1.50e-03  7.8e-02  2.0e+00  6.4e-02  1.06e+02
##      6      24 -6.162e+03  3.64e-03  3.43e-03  1.3e-01  2.0e+00  1.3e-01  4.63e+03
##      7      26 -6.167e+03  7.37e-04  7.41e-04  2.2e-02  2.0e+00  2.6e-02  3.07e+05
##      8      36 -6.168e+03  1.40e-04  2.66e-04  1.4e-06  5.1e+00  1.6e-06  6.22e+03
##      9      37 -6.168e+03  3.34e-07  3.53e-07  1.4e-06  2.0e+00  1.6e-06  4.28e+03
##     10      47 -6.180e+03  1.98e-03  2.14e-03  5.8e-02  2.0e+00  7.4e-02  4.29e+03
##     11      55 -6.180e+03  1.30e-05  2.93e-05  3.3e-07  9.7e+00  4.4e-07  1.37e-02
##     12      56 -6.180e+03  3.48e-07  3.91e-07  3.3e-07  2.0e+00  4.4e-07  1.14e-02
##     13      67 -6.187e+03  1.12e-03  1.37e-03  4.4e-02  1.2e+00  6.1e-02  1.14e-02
##     14      68 -6.190e+03  4.52e-04  6.35e-04  3.9e-02  6.5e-01  6.1e-02  8.52e-04
##     15      69 -6.193e+03  4.40e-04  4.31e-04  2.3e-02  0.0e+00  4.6e-02  4.31e-04
##     16      70 -6.195e+03  4.59e-04  5.21e-04  2.5e-02  7.9e-01  4.6e-02  7.05e-04
##     17      72 -6.196e+03  3.61e-05  5.13e-05  6.3e-03  1.1e+00  1.3e-02  7.86e-05
##     18      73 -6.196e+03  5.21e-07  2.32e-06  8.6e-04  0.0e+00  1.8e-03  2.32e-06
##     19      74 -6.196e+03  7.61e-07  7.14e-07  5.2e-04  0.0e+00  8.9e-04  7.14e-07
##     20      90 -6.196e+03 -1.10e-14  1.62e-14  1.8e-14  2.1e+04  3.0e-14  2.46e-10
##
## ***** FALSE CONVERGENCE *****
##
## FUNCTION      -6.195664e+03      RELDX      1.786e-14
## FUNC. EVALS      90      GRAD. EVALS      20
## PRELDF      1.617e-14      NPRELDF      2.462e-10
##
##      I      FINAL X(I)      D(I)      G(I)
##
##      1      3.544011e-06      1.000e+00      3.325e+03
##      2      1.307012e-01      1.000e+00      9.388e-02
##      3      8.438273e-01      1.000e+00      2.096e-01

```

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

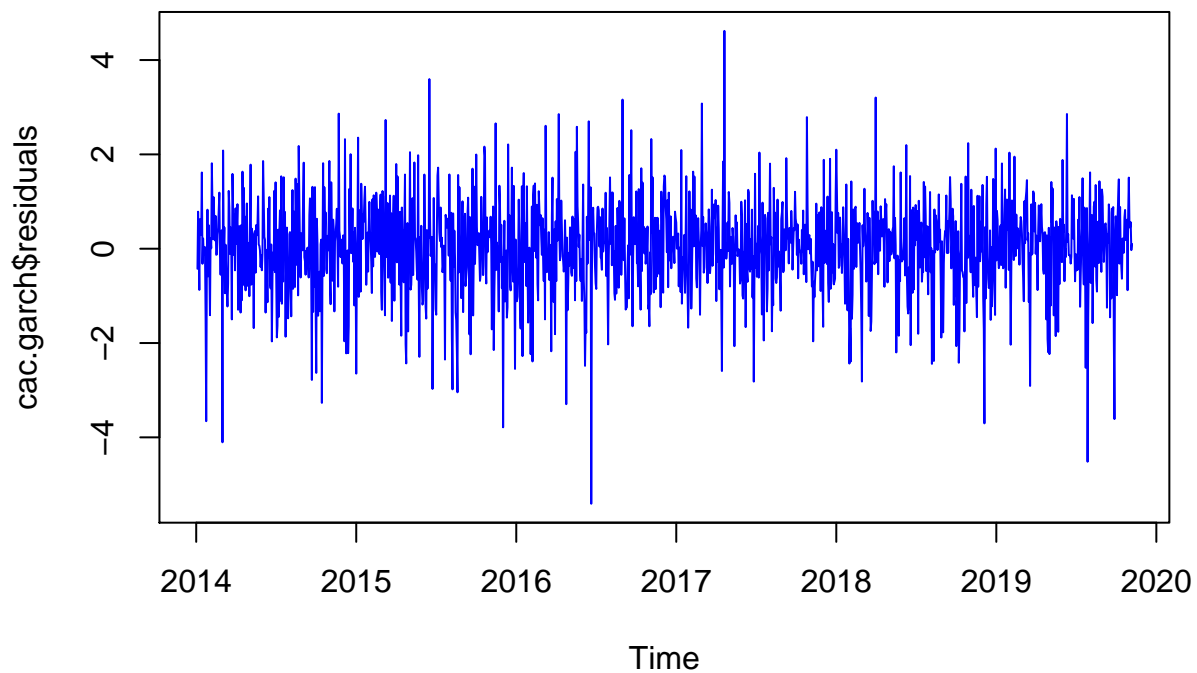
```

##
## Call:
## garch(x = cac40)
##
## Model:
## GARCH(1,1)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -5.40797 -0.52192  0.06096  0.62007  4.61860
##
## Coefficient(s):
##      Estimate Std. Error t value Pr(>|t|)
## a0 3.544e-06  6.782e-07   5.226 1.73e-07 ***
## a1 1.307e-01  1.269e-02  10.302 < 2e-16 ***
## b1 8.438e-01  1.468e-02  57.469 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
## Jarque Bera Test

```

```
##
## data: Residuals
## X-squared = 273.67, df = 2, p-value < 2.2e-16
##
##
## Box-Ljung test
##
## data: Squared.Residuals
## X-squared = 0.043021, df = 1, p-value = 0.8357
```

### 2.6.1 Graphique des résidus



Les résidus obtenus sont plus conforme à un bruit blanc.

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

## 2.7 Références

Books :

- (B1) Statistics of financial 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/>