Rapport - Séries Temporelles

Philippe Real

06 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

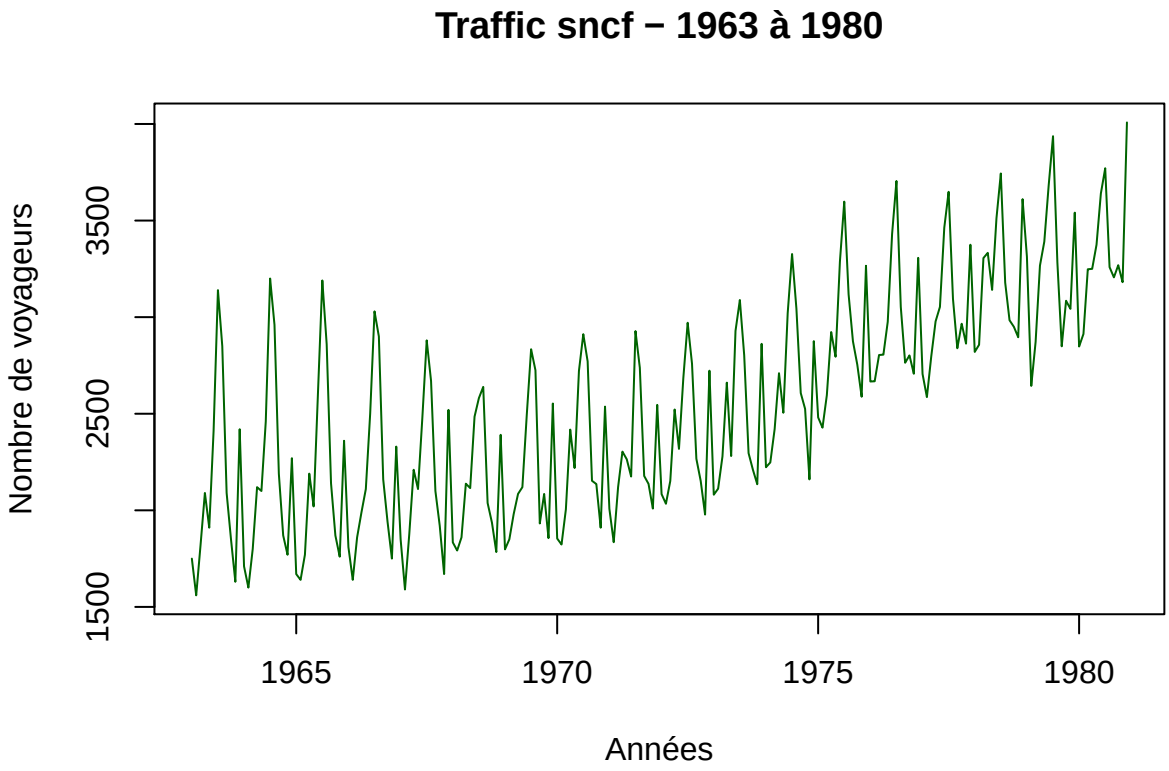
1.1.1 Lecture des données

```
##       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
```

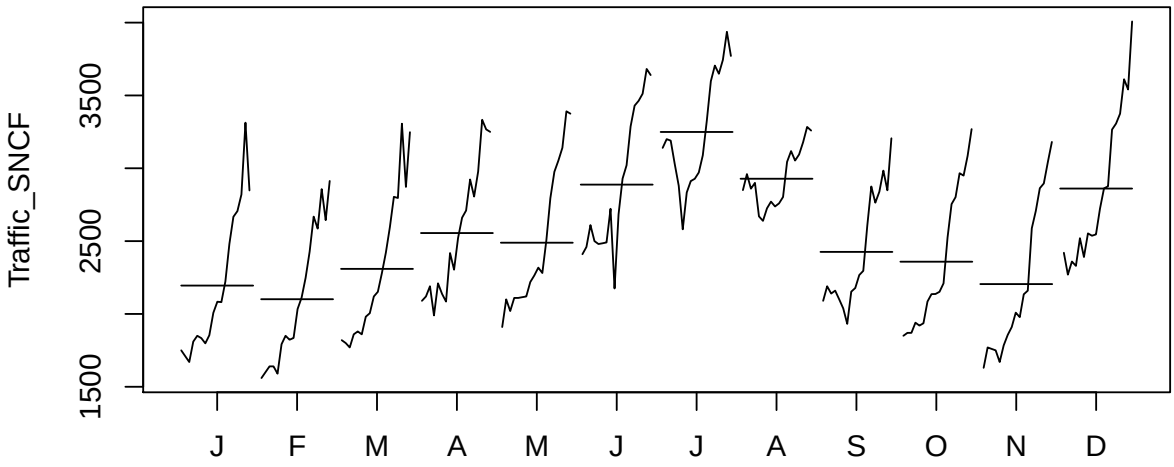
1.1.2 Chronogramme de la séries temporelles - sncf

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).

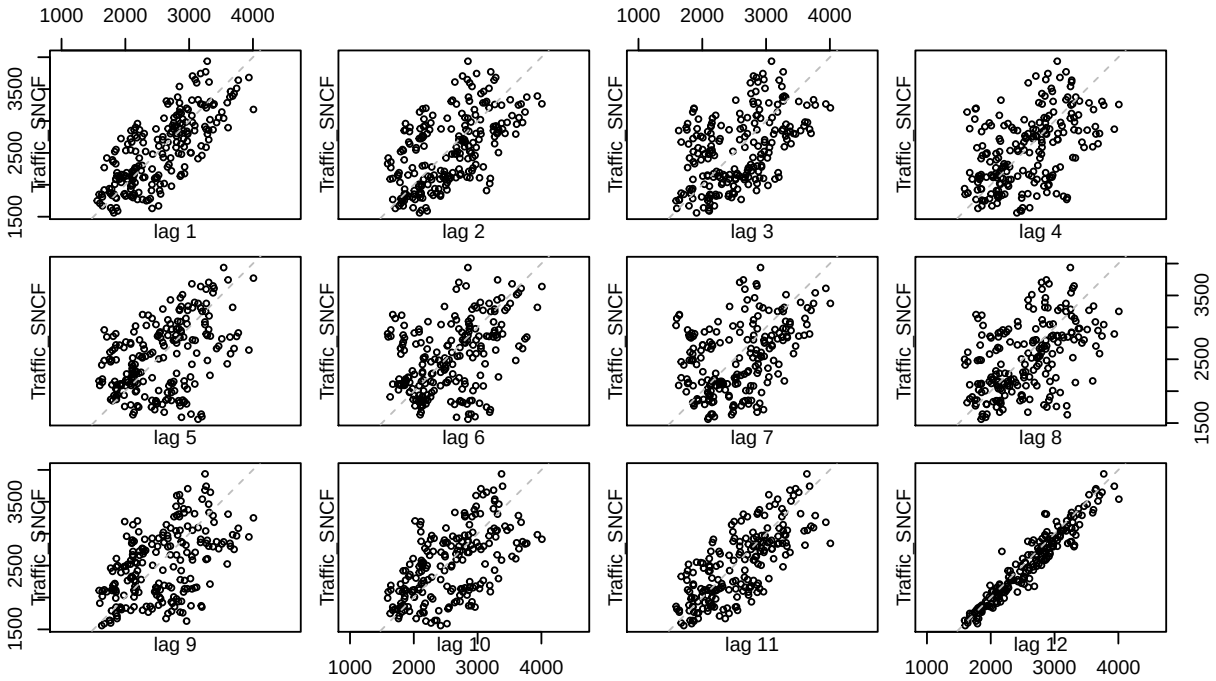


1.1.3 Représentations graphiques : month-plot et lag-plot

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.

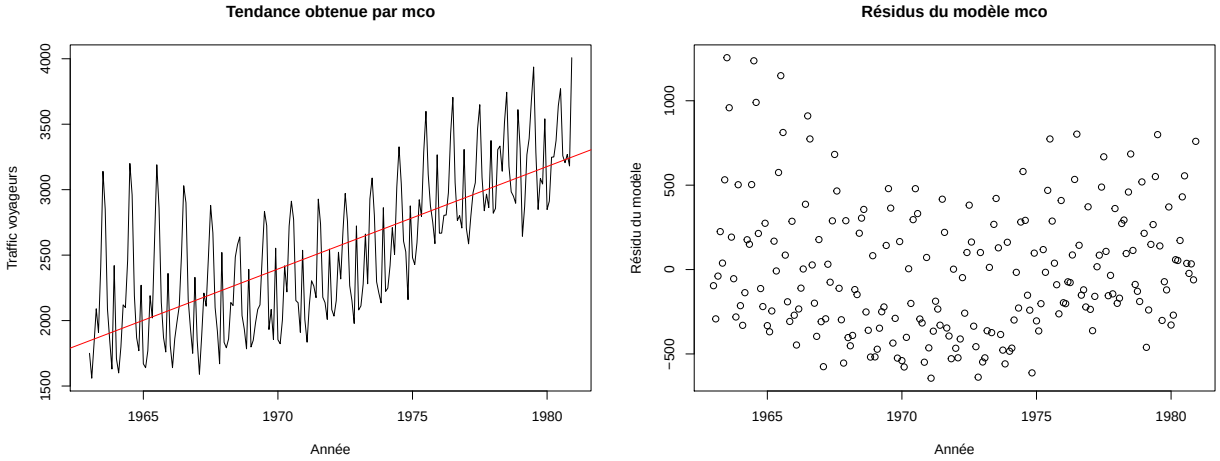
1.1.4 Tendance et saisonnalité

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

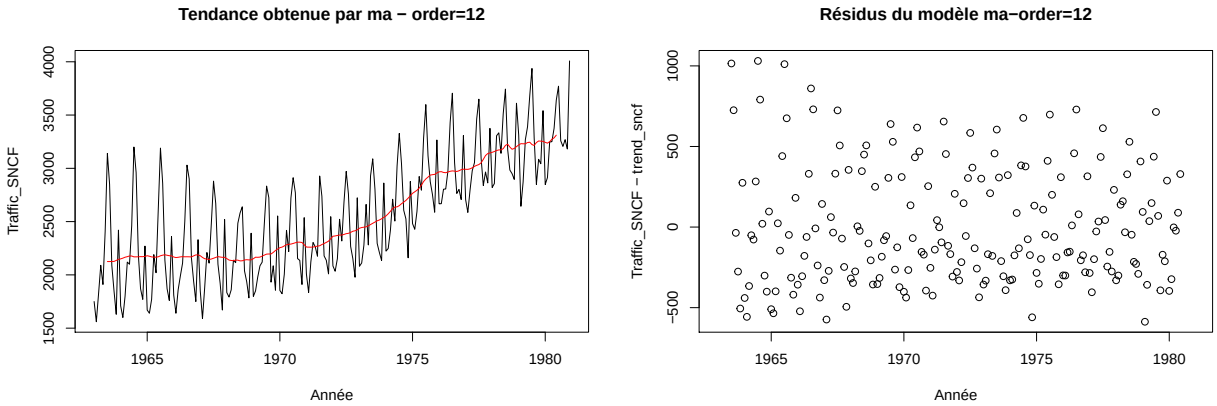
On suppose que la série est de la forme $X_t = m_t + z_t$ avec $m_t = \beta_0^* + \beta_1^* t$ et z_t l'ereur ou résidus. On cherche à estimer par l'estimateur des moindres carrés, les paramètres β_0^* et β_1^* à partir de la série des observations.

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



- Estimation de la tendance par moyennes mobiles

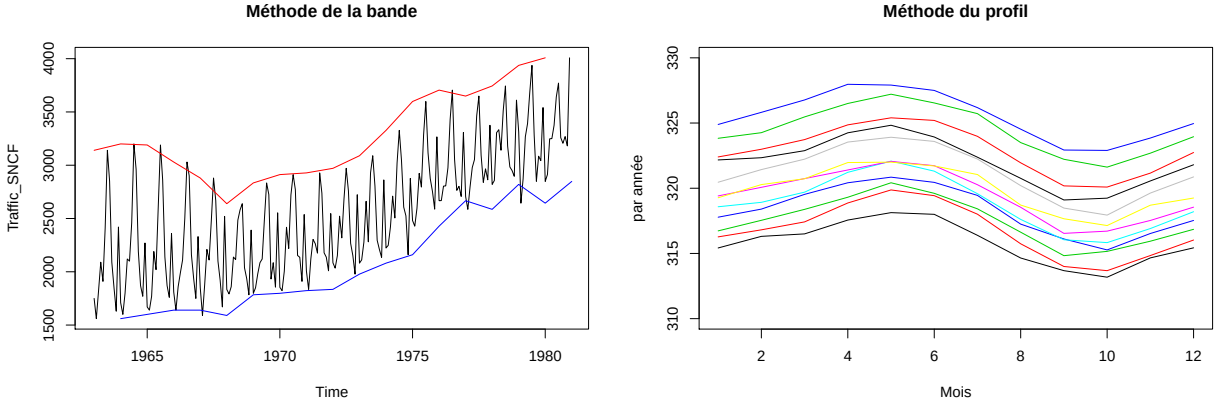
On cherche ici à ajuster un modèle à la courbe observée et parmis les nombreuses méthodes statistiques disponibles (ondelettes, noyaux, splines...) on va utiliser la méthode des moyennes mobiles qui est bien adaptée aux série temporelles. Pour celà on utilise la fonction "ma" du package "forcast" de 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.

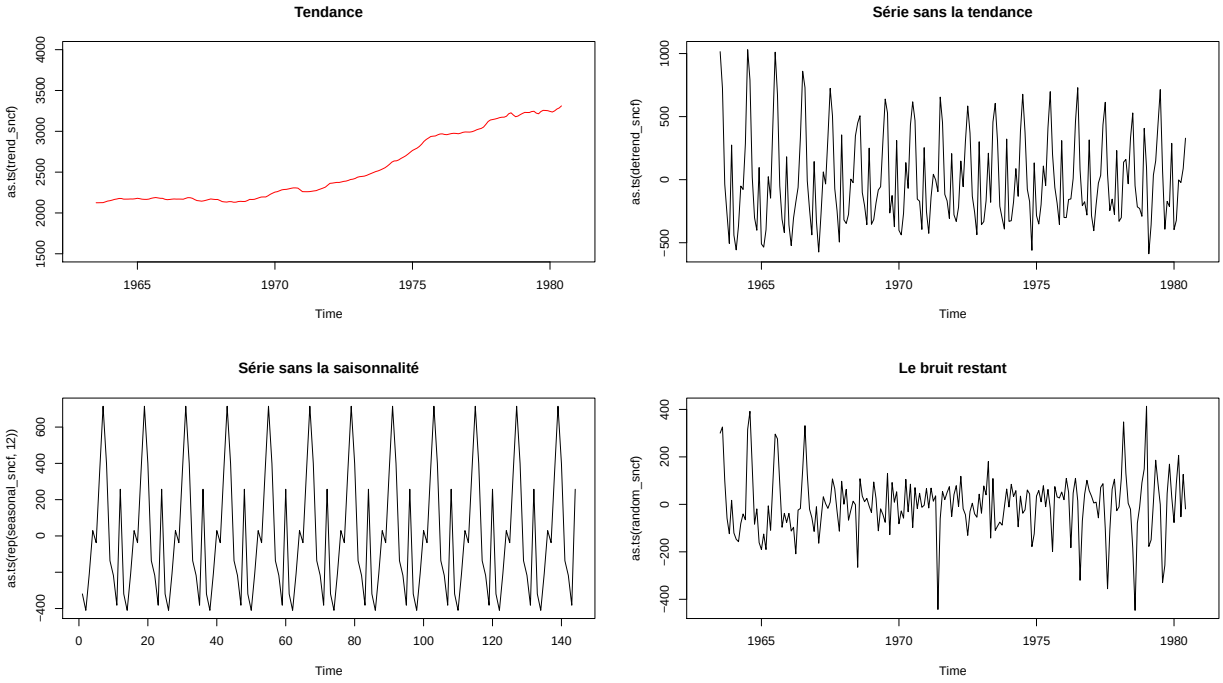
- Série décomposée - Tendance, Saisonnalité, Résidus

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.



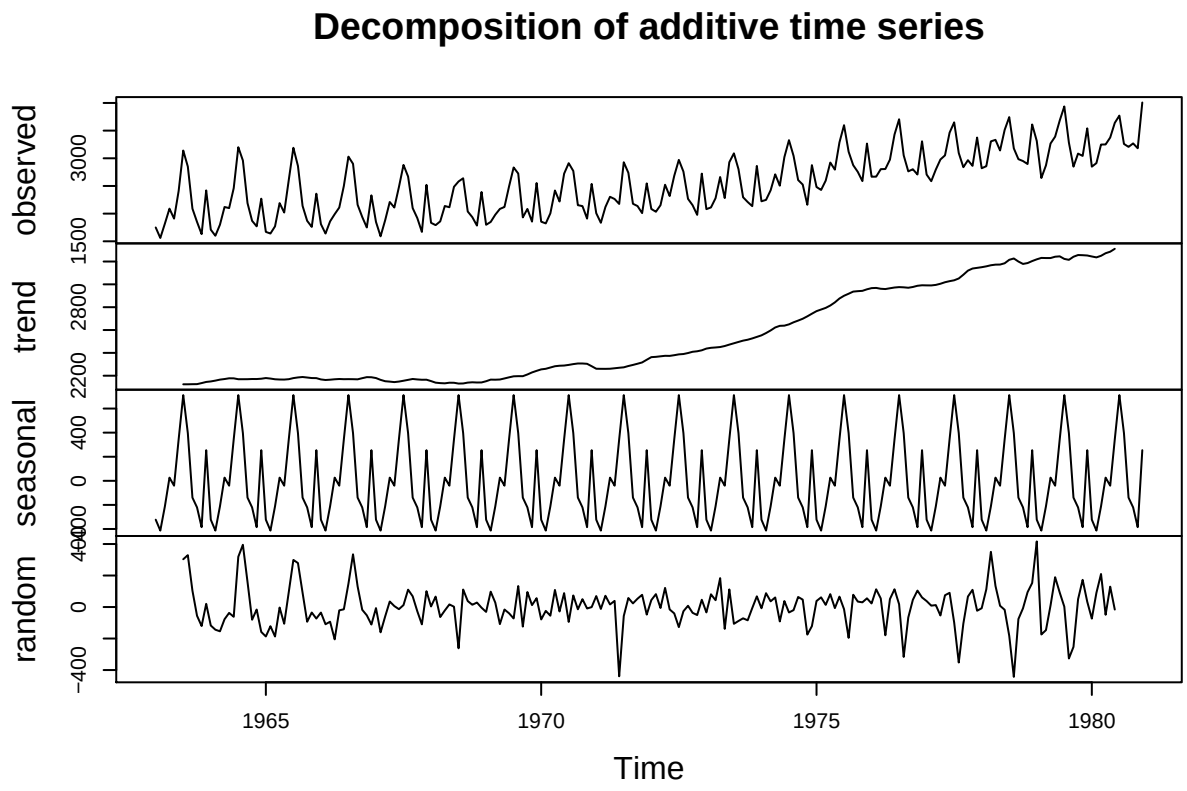
Dans notre cas la méthode de la bande indiquerait un modèle additif à partir de l'année 1968. La méthode du profil plaide aussi plutôt pour un modèle additif.

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

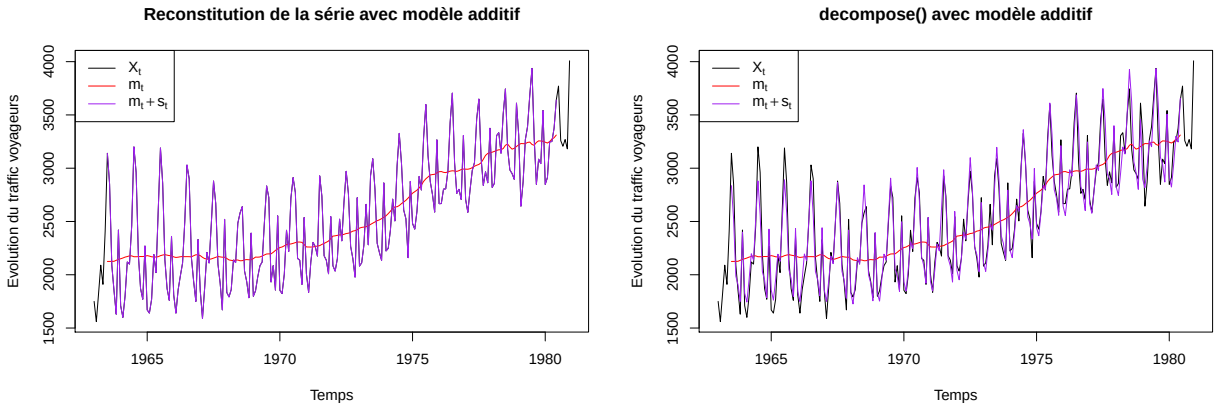
On va décomposer la série temporelles en utilisant la fonction décompose de R de façon à avoir une idée générale de la tendance (trend) saisonalité et bruit. On remarque que les graphiques obtenus sont très similaire avec ceux obtenus précédement.



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 :

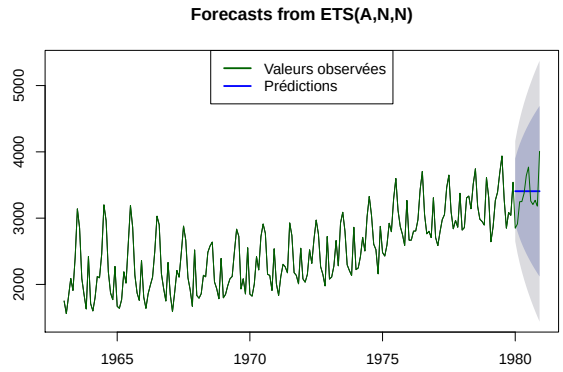
Type de lissage Exponentiel	AIC	AICc	BIC
Lissage Simple (A,N,N)	3517.8041845	3517.9241845	3527.7585445
Lissage Double (A,A,N)	3522.0129751	3522.3160054	3538.6035751
Holt-Winters (A,A,A)	3093.312925	3096.6032476	3149.7209649

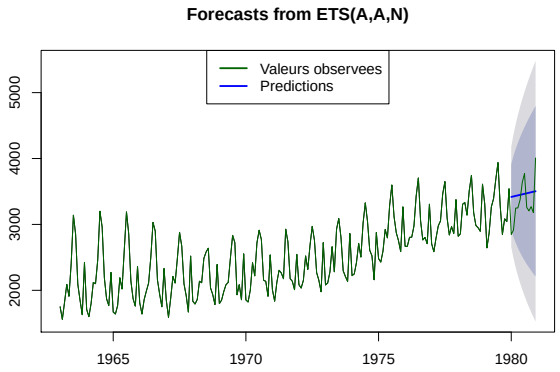
Nous remarquons que le modele est trop basique dans notre cas pour predire a l'horizon 12 et que l'intervalle de confiance a 80%, bien que tres large, ne contient pas toutes les vraies valeurs de la serie. On peut utilser predict(fitLES,12) prediction a horizon 1 an.

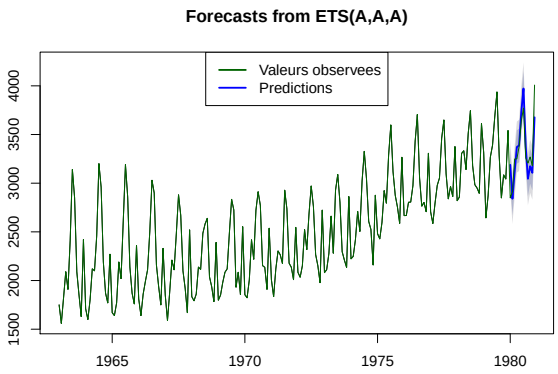
On remarque que le lissage double n'apporte pas vraiment d'amélioration par raport au simple. Les critères AIC et BIC sont plus élevés dans le cas du lissage double par rapport au lissage simple. Le lissage exponentiel triple ou de Holt-Winters apporte une amélioration.

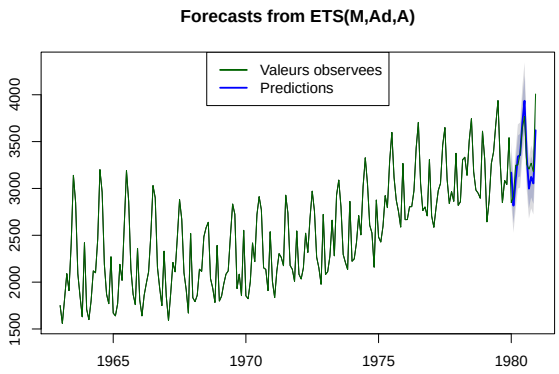
On peut considérer d'autres méthodes de lissage en jouant sur le caractères Additif/Multiplicatif des composantes. On a vu précédemment, que la tendance était plutôt additive, on va donc fixer la tendance=A et faire varier les autres paramètres.

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







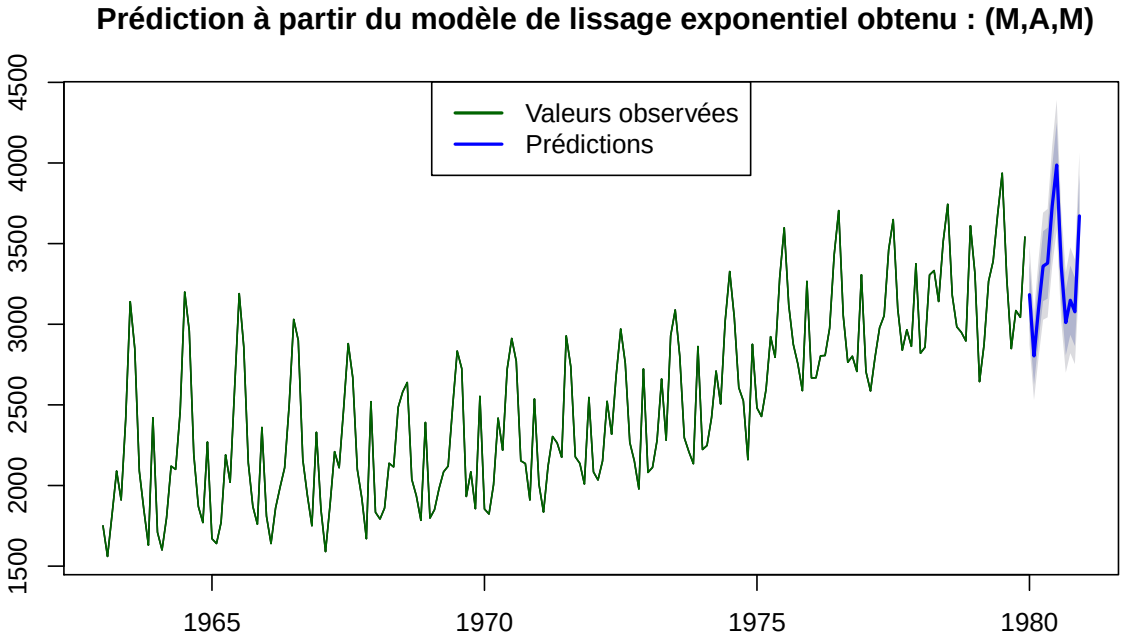


- Procédure automatique - modèles ajustés par ets.

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



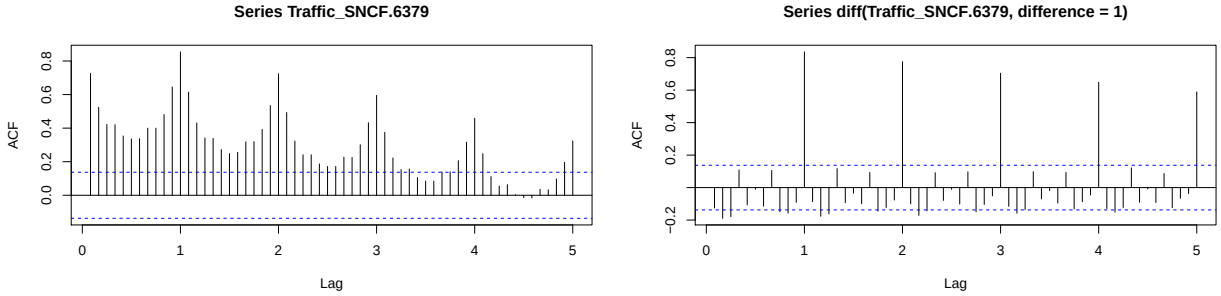
1.3 Modélisation

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)

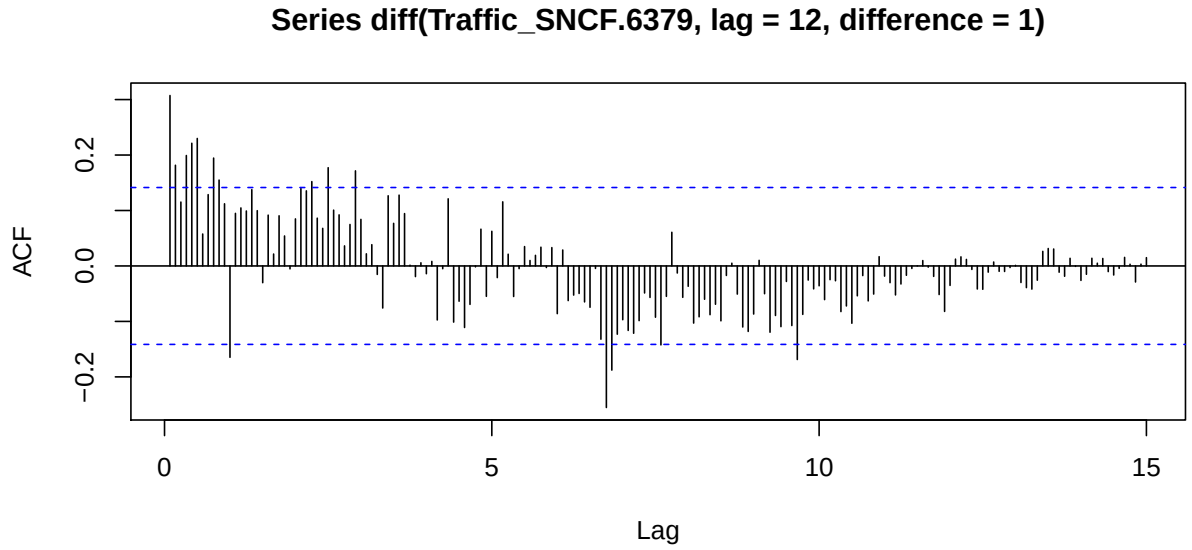
1.3.1 Identification du modèle

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 $(1-B)X_t$



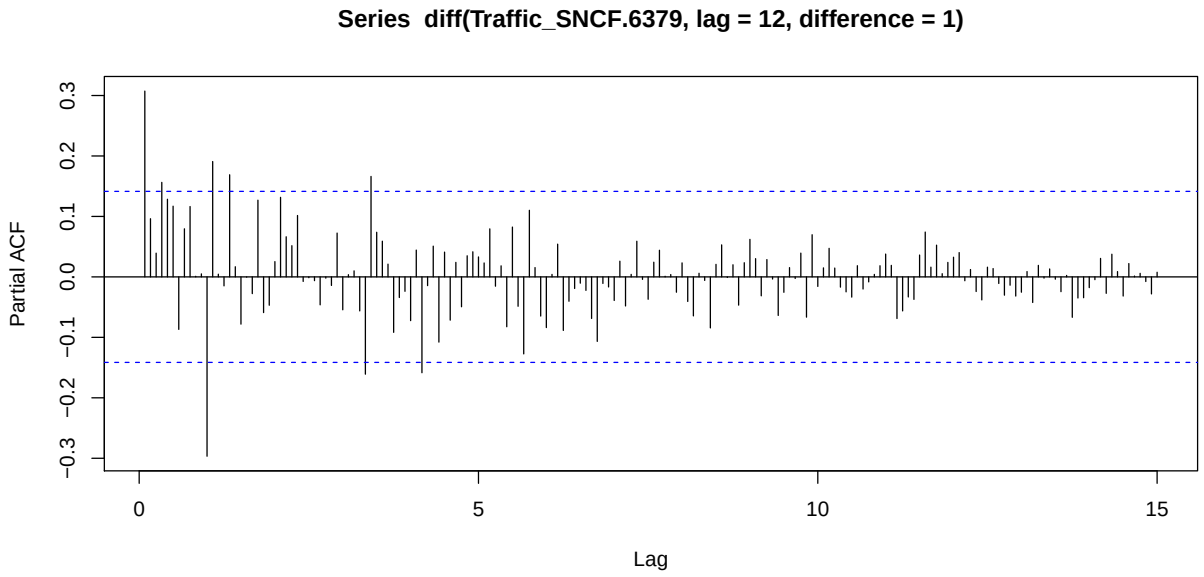
La fonction d'autocorrélation estimée est positive. On remarque une périodicité de 1 (12 mois) (graphique de gauche).On peut essayer de différentier la série au moins une fois (graphique de droite). On remarque des autocorrélations importantes pour les valeurs de h de 1 période (année), tout les 12 mois. 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.

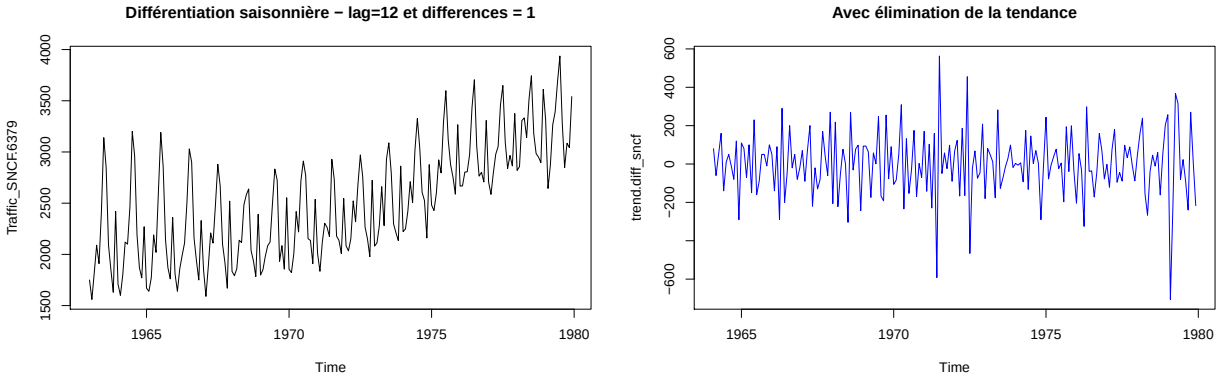


L'autocorrélation partielle suggère un terme d'ordre q=1 (12ème mois) soit un terme moyenne mobile du type : $(1-\theta_1 B)(1-\theta_2 B^{12})\epsilon_t)$

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.

Et le test du Portmanteau ou test de blancheur sur R n'est quant à lui pas concluant et donne une p_value=$5.7890803 \times 10^{-11}$

On va maintenant estimer plusieurs modèles SARIMA(p,1,q)(r,1,s) en faisant varier les paramètres p et q de 0 à 2 et la saisonnalité (r,1,s) de 0 à 1.

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	2456.2986855	2456.3625153	2462.8032323
010_010	2503.2666857	2503.2878498	2506.5189592
010_110	2465.9531798	2466.0170095	2472.4577266
010_111	2458.0372766	2458.1656189	2467.7940969
011_011	2841.6159974	2841.7366004	2851.5556153
011_010	2425.0373649	2425.1011947	2431.5419117
011_110	2394.2238938	2394.3522361	2403.9807141
011_111	2391.0242495	2391.2393033	2404.0333432
110_011	2429.4544314	2429.5827737	2439.2112517
110_010	2470.9624385	2471.0262683	2477.4669854
110_110	2435.4894728	2435.6178151	2445.2462931
110_111	2430.383853	2430.5989068	2443.3929467
111_011	2385.9962789	2386.2113327	2399.0053726
111_010	2422.0205318	2422.1488741	2431.7773521
111_110	2391.3997308	2391.6147846	2404.4088246
111_111	2387.4722775	2387.7966018	2403.7336446
112_011	2387.9550467	2388.2793711	2404.2164139
112_010	2424.0047984	2424.2198521	2437.0138921
112_110	2393.3930052	2393.7173295	2409.6543723
112_111	2389.4450772	2389.9015989	2408.9587177
211_011	2387.913895	2388.2382193	2404.1752622
211_010	2423.9980506	2424.2131044	2437.0071443
211_110	2393.3857032	2393.7100276	2409.6470704
211_111	2389.4174055	2389.8739273	2408.9310461

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

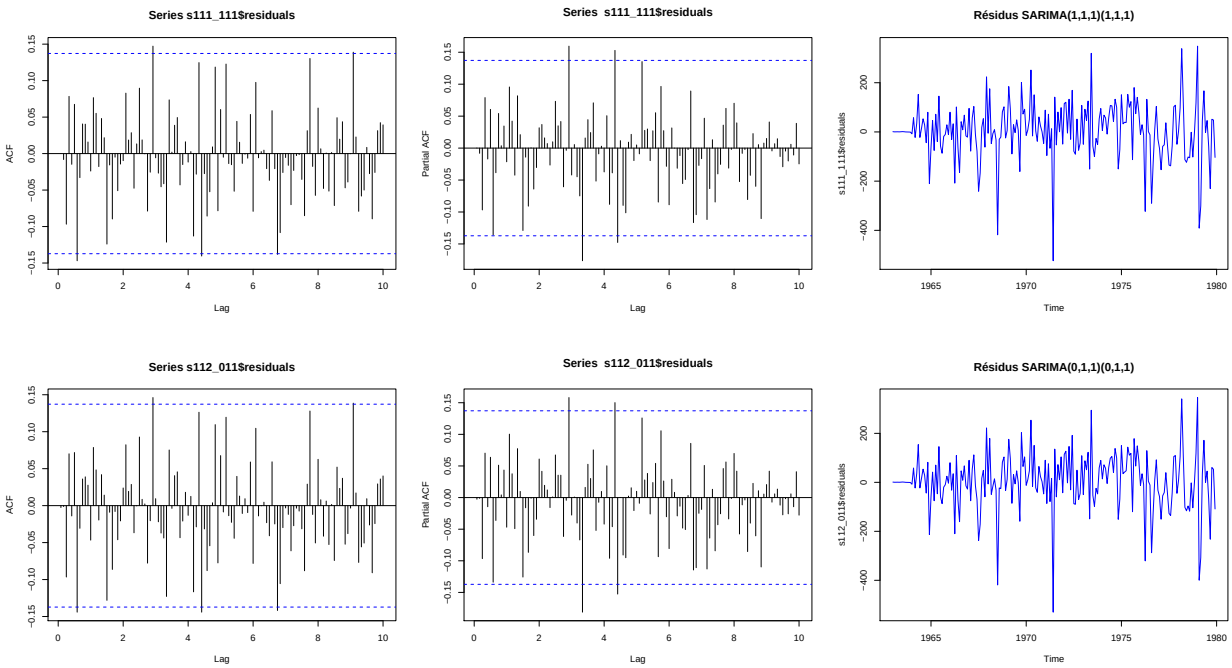
Avant de paser à la prédiction, on va maintenant valider ou invalider les modèles obtenus.

- 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 3 modèles SARIMA(2,1,1)(0,1,1), SARIMA(1,1,2)(0,1,1) et SARIMA(1,1,1)(1,1,1) on accepte la blancheur des résidus comme le montre le tableau ci-dessous.

SARIMA	p-value
211_011	0.6724191
211_011	0.6669889
111_111	0.6602705
011_011	0

- ACF et PACF des résidus



- Statistiques

SARIMA(1,1,2)(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
```

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

```
##             ar1       ma1       ma2     sma1
## t.stat 0.509628 -2.918658 -0.205239 -7.45299
## p.val  0.610312  0.003515  0.837385  0.00000
```

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

```
##             ar1       ma1      sar1      sma1
## t.stat 2.393756 -23.20904 -0.732900 -2.865292
## p.val  0.016677   0.00000  0.463619  0.004166
```

- Corrélations

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

```
##             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
```

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

```
##              ar1         ma1          ma2         sma1
## ar1   1.00000000 -0.96718272  0.956553485 -0.024956127
## ma1  -0.96718272  1.00000000 -0.990059515 -0.022626529
## ma2   0.95655349 -0.99005951  1.000000000  0.009845505
## sma1 -0.02495613 -0.02262653  0.009845505  1.000000000
```

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

```
##              ar1          ma1        sar1         sma1
## ar1   1.00000000 -0.494295626  0.05457001 -0.099011985
## ma1  -0.49429563  1.000000000 -0.04839928 -0.009262029
## sar1  0.05457001 -0.048399278  1.00000000 -0.862729803
## sma1 -0.09901198 -0.009262029 -0.86272980  1.000000000
```

1.4 Modélisation automatique avec R

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

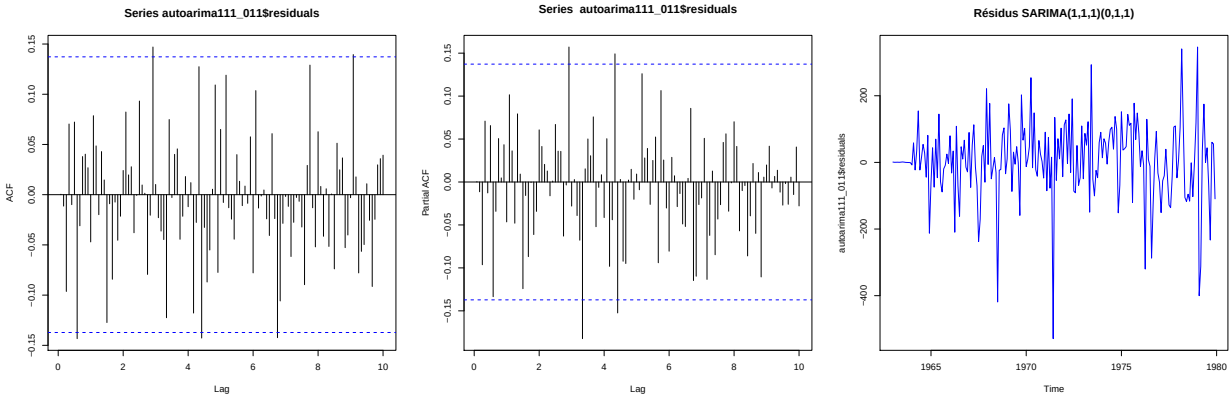
```
## Series: Traffic_SNCF.6379
## ARIMA(1,1,1)(0,1,1)[12]
##
## Coefficients:
##          ar1      ma1     sma1
##       0.2021  -0.8907  -0.4985
## s.e.  0.0828   0.0375   0.0669
##
## sigma^2 estimated as 14801:  log likelihood=-1189
## AIC=2386   AICc=2386.21   BIC=2399.01
```

- Validation du modèle obtenu :

```
##             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
```

Le test de blancheur des résidus ou test Box-Pierce est accepté avec la p-value : 0.6762469

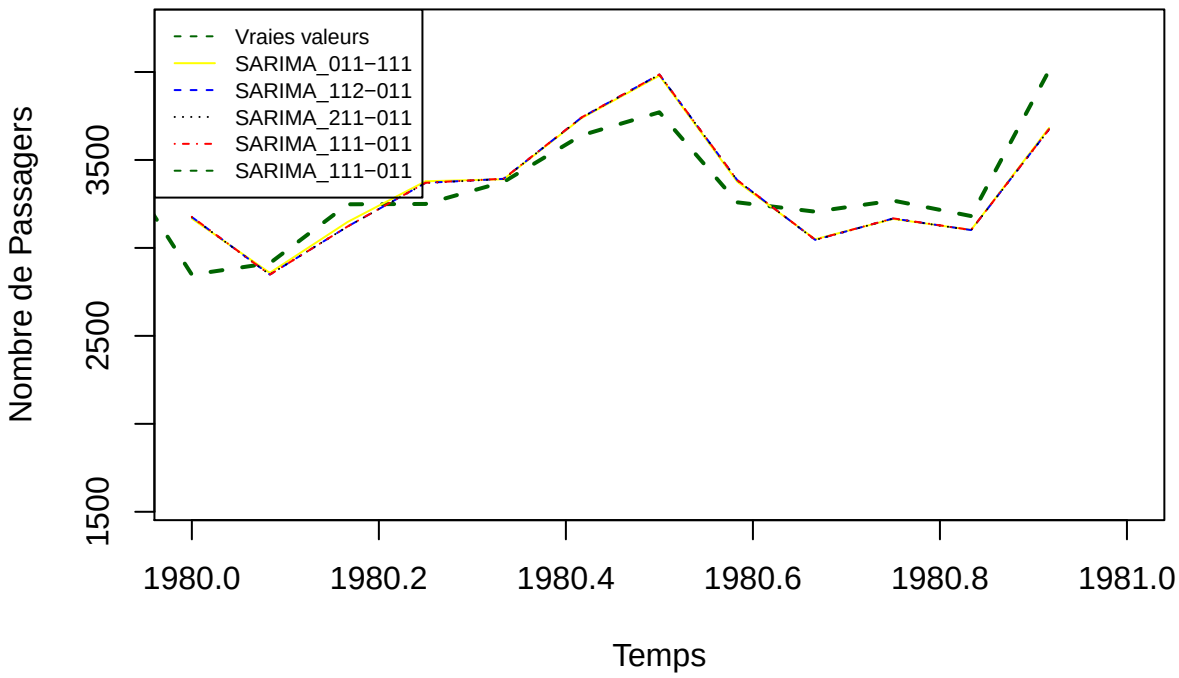


1.5 Prévisions et comparaison des modèles obtenus

- Prédictions SARIMA(1,1,1)(0,1,1)

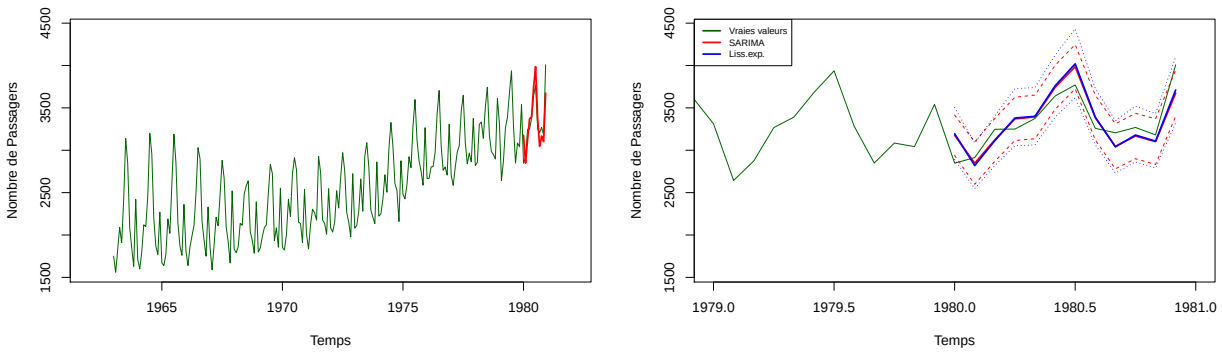
```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1980       3176.918 3021.002 3332.833 2938.466 3415.369
## Feb 1980       2847.791 2684.491 3011.092 2598.045 3097.538
## Mar 1980       3122.217 2956.723 3287.710 2869.116 3375.317
## Apr 1980       3370.292 3203.281 3537.303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

- Comparaison entre les différents modèles SARIMA



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(1,1,1)(0,1,1) 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.

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

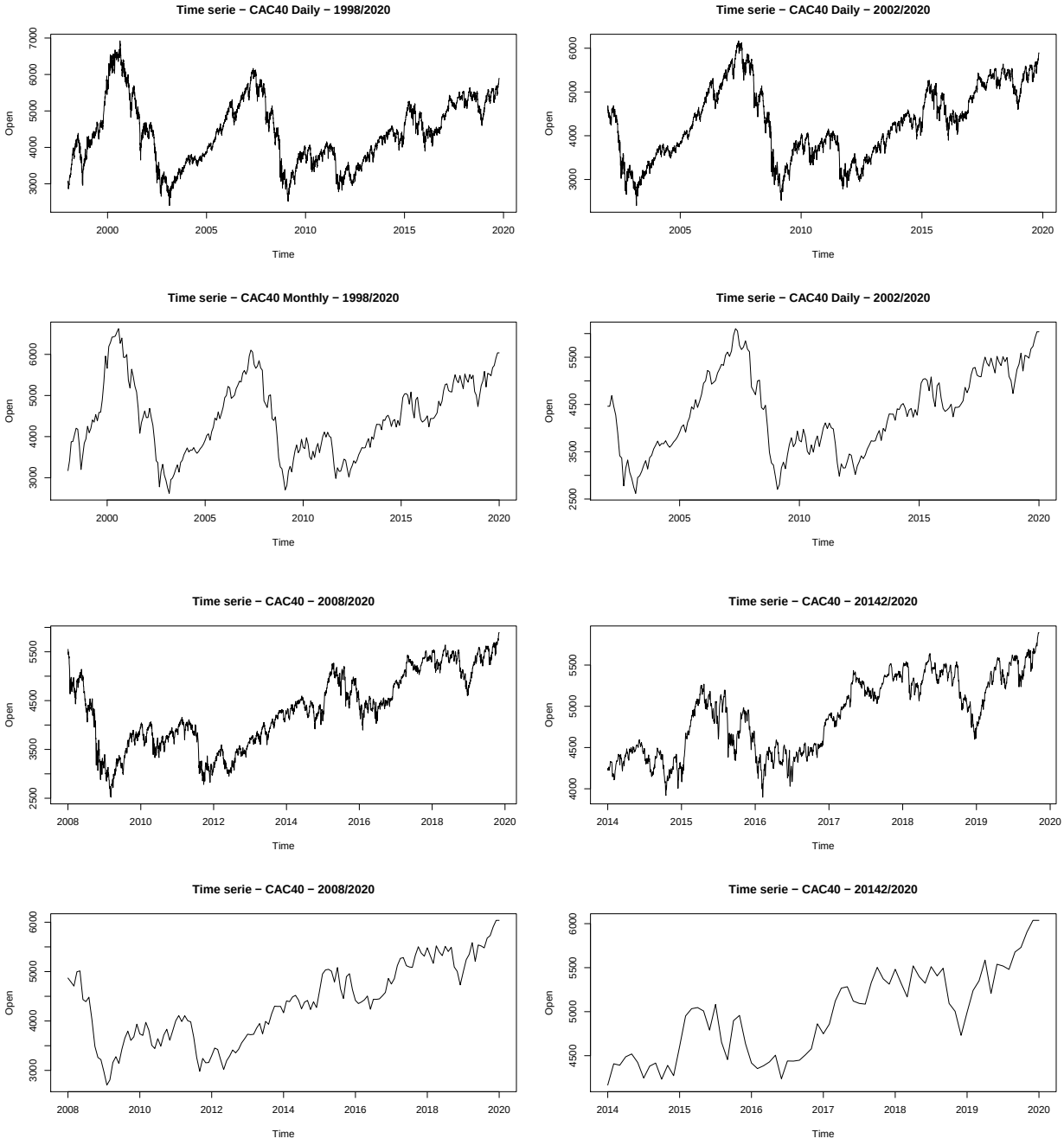
2.3 Analyse des séries temporelles obtenues

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)

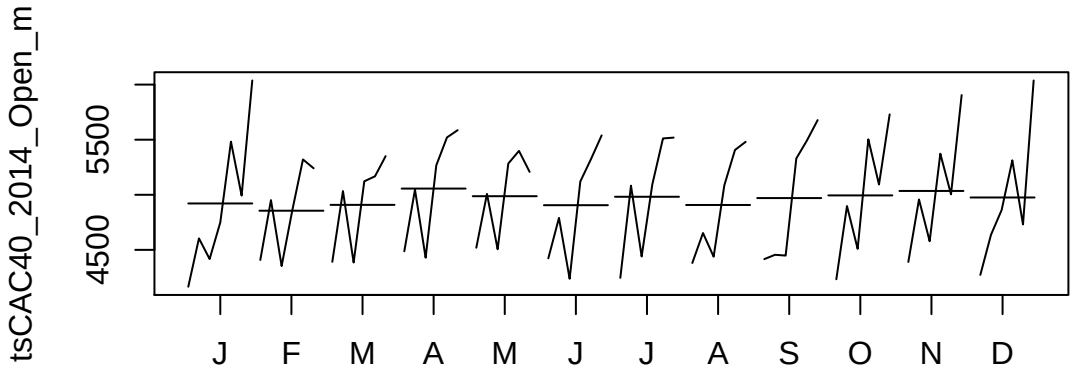
On a 4 séries temporelles possibles en fonction du choix de la quantité observée (High, Low, Open, Close, Volume). On va s'intéresser à la valeur à la fermeture pour la cotation de l'indice CAC40 (Close).

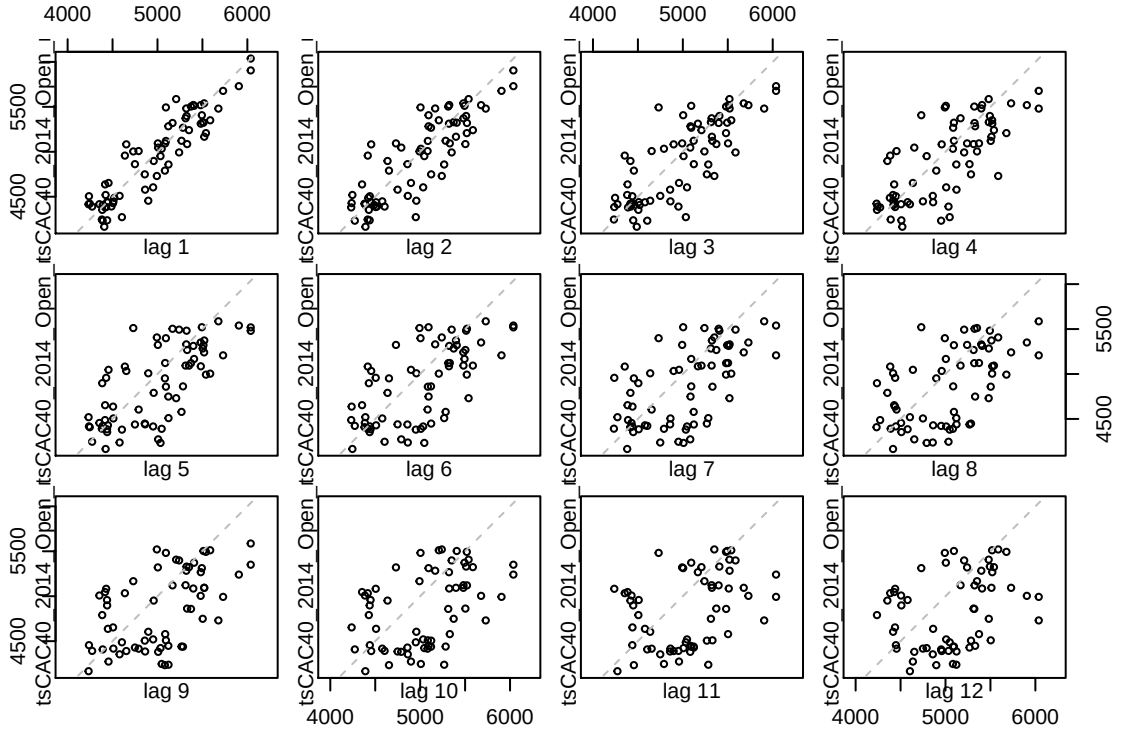


Dans le cas des données mensuelles On retrouve biensûr la forme globale de la série mais moins bruitée.

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.

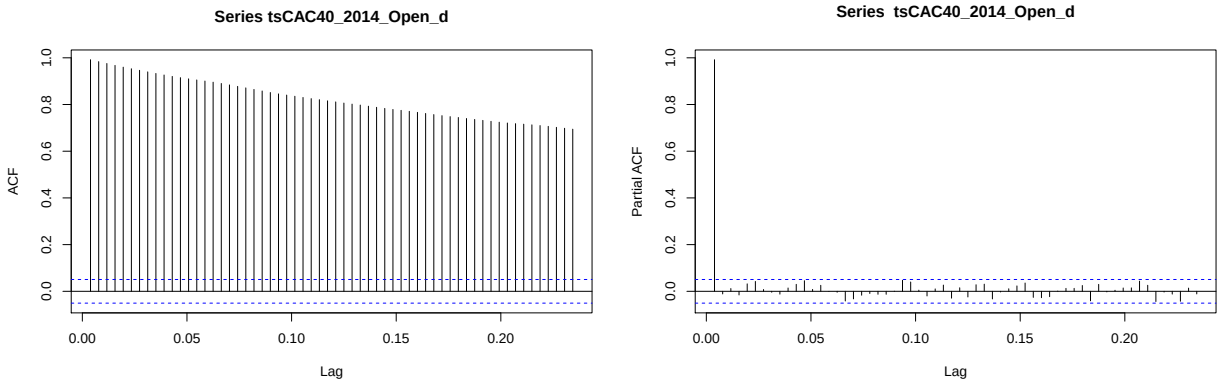




2.3.3 Etude de la stationarité

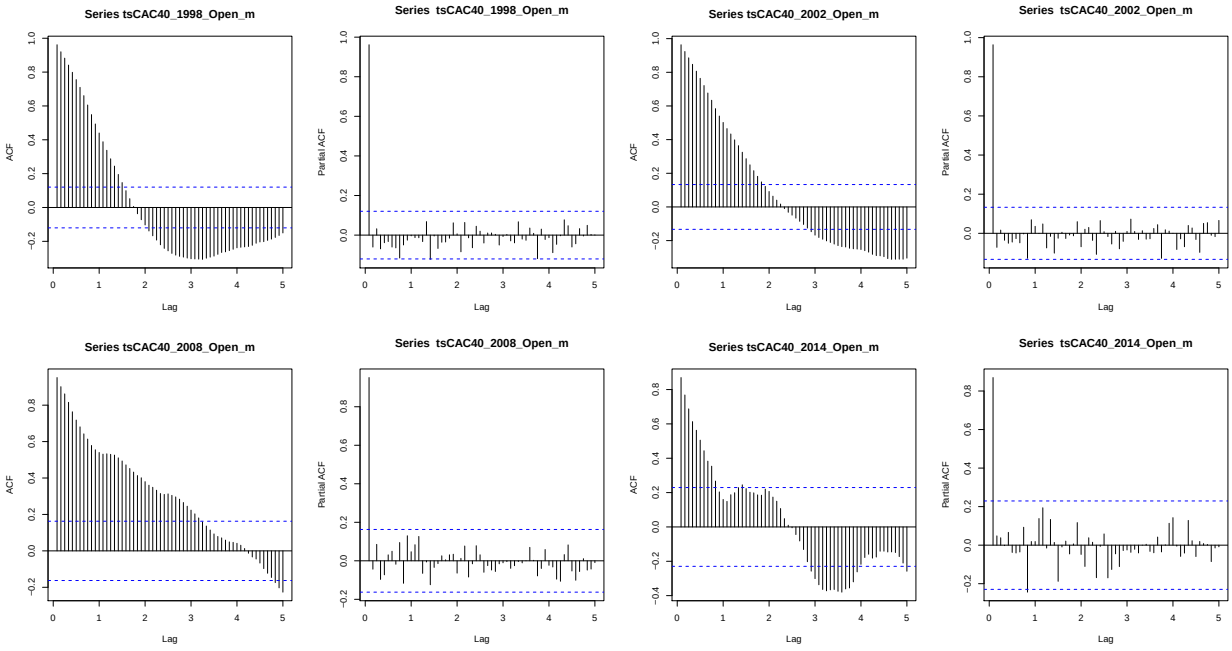
La stationnarité est la stationnarité du processus au sens faible. Un tel processus doit avoir les propriétés suivantes : La moyenne et la variance ne varient pas au cours du temps et le processus n a pas de tendance. Pour vérifier ces hypothèses, on s'appuiera sur une analyse des graphique d'autocorrélation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

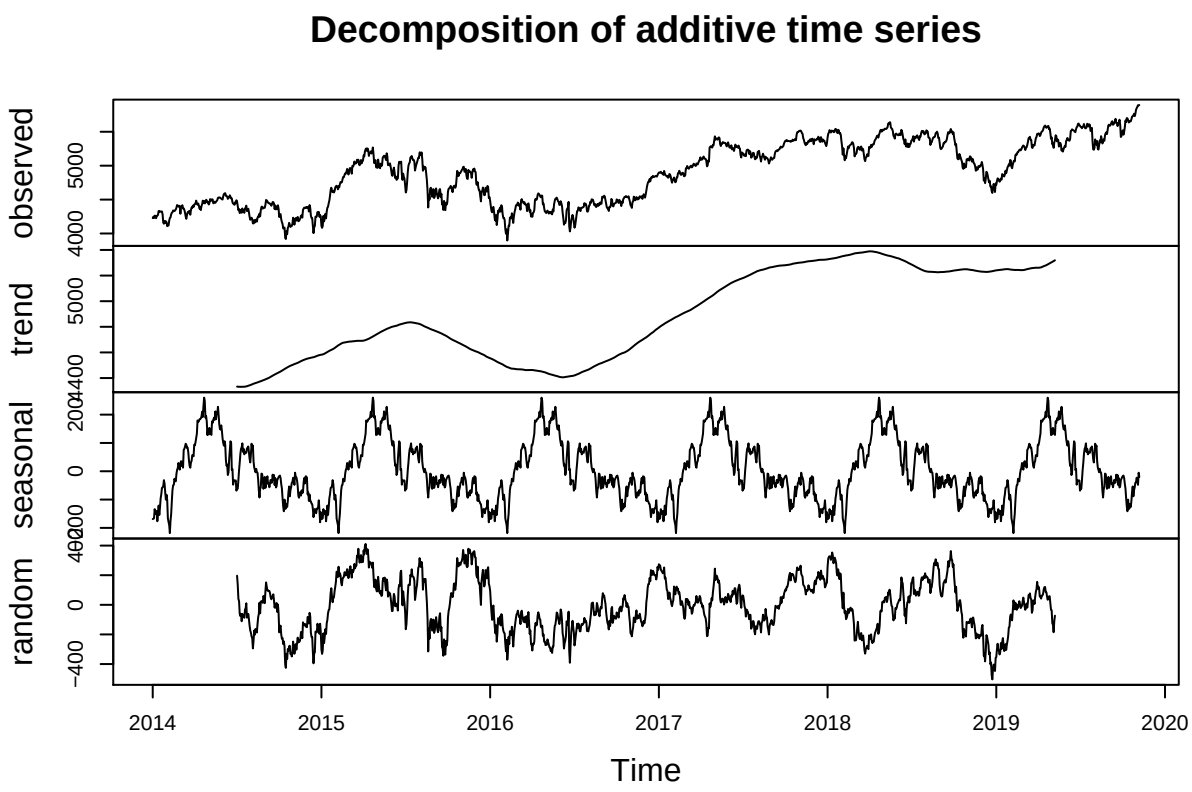
série	Dickey-Fuller	Lag order	p-value
1998	-2.3092612	6	0.4456761
2002	-2.1857017	5	0.4980833

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

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 t2Z 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

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

Avec l'opérateur retard cette équation se réécrit :

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

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

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

Avec l'opérateur retard cette équation se réécrit :

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

Un processus MA(q) est toujours stationnaire quelles que soient les valeurs des α_i , il est de plus de moyenne μ. 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 :

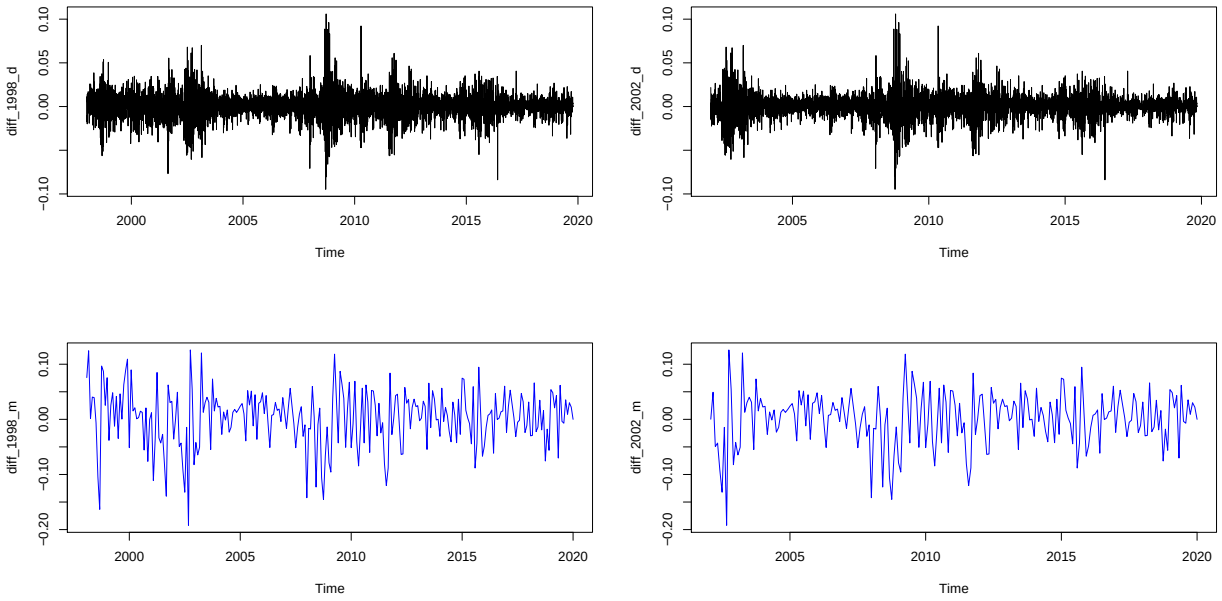
$$\forall t = 1, \ldots, n \quad X_t = \beta + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + ... + \alpha_q \varepsilon_{t-q} \text{ (où les } \varepsilon_t \text{ sont des bruits blanc centré)}$$

On cherche maintenant à se ramener à un processus sationnaire en utilsant 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 differentiation des séries

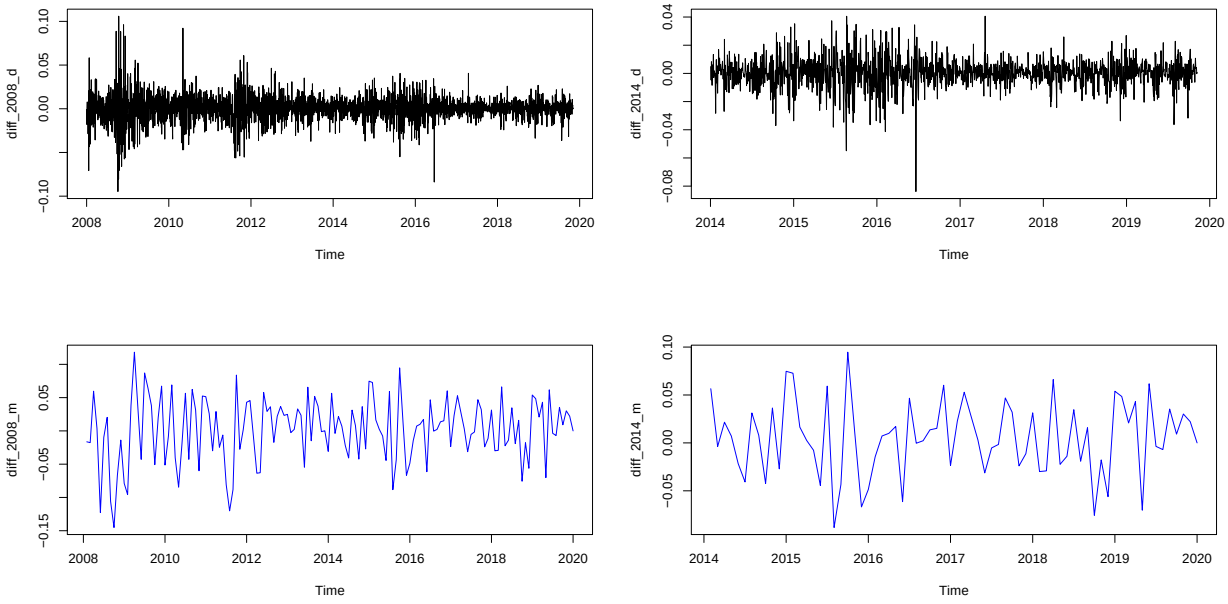
Pour tenter de rendre la série stationnaire, on applique la méthode de différentiation. On utilise la function R diff. En paramètre on passe un facteur de 1 pour la différence et de 0 pour la saisonnalité. On retrouve l'idée de la transformation Log-return R_t des prix X_t où $R_t = Log(X_t/X_{t-1})$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

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



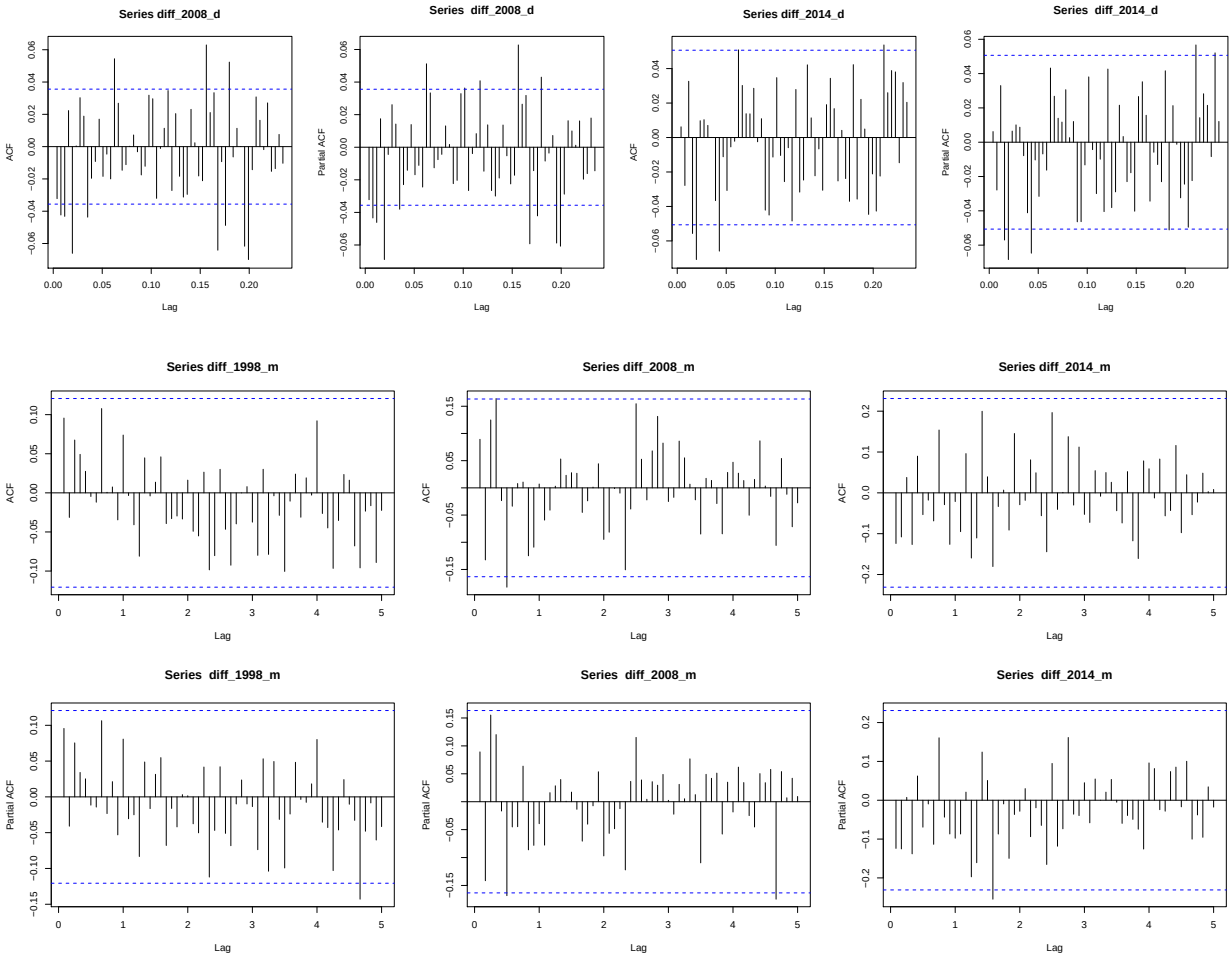
L'hypothèse de stationnarité du processus ainsi transformé, étant vérifié on peut envisager une modélisation ARIMA. On passe maintenat à la détermination des paramètres p et q du mopdèle ARMA(p,q).

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

2.4.3 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 possdent 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.

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

On remarque que selon ce critère plusieurs modéles 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

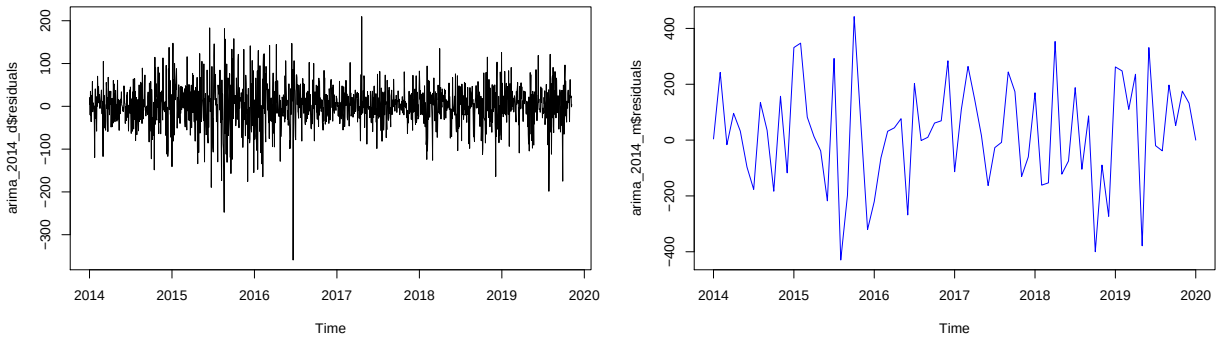
série	X-squared	df	p-value
1998	12.6097304	20	0.8934941
2008	14.8918527	20	0.782562
2014	17.2944661	20	0.6337782

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 alternatve. On verra qu'ils sont mieux adaptées à la modélisation des séries financières.

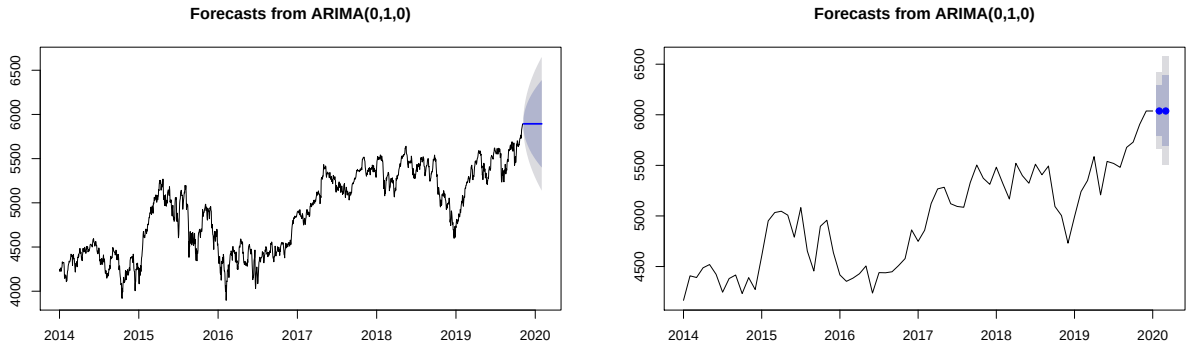
2.5.3 Normalité des résidus

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 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.

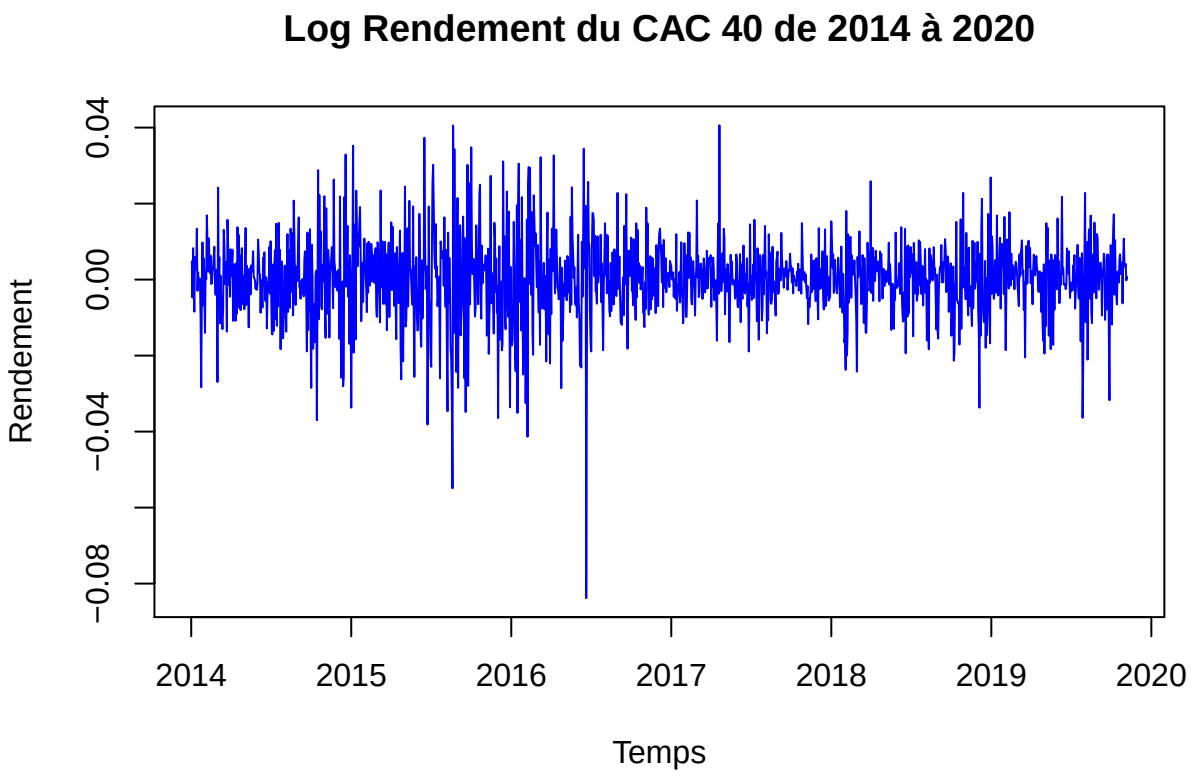
Comme ont la fait dans le caxs de la modélisation ARIMA on transforme la série originelle par différentiation. Pour obtenir une série stationnaire. Ici on va considérer comme c'est souvent le cas en finance le log-return. C'est à dire la quantité déduite du prix X_t de la manière suivante : Log-return 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

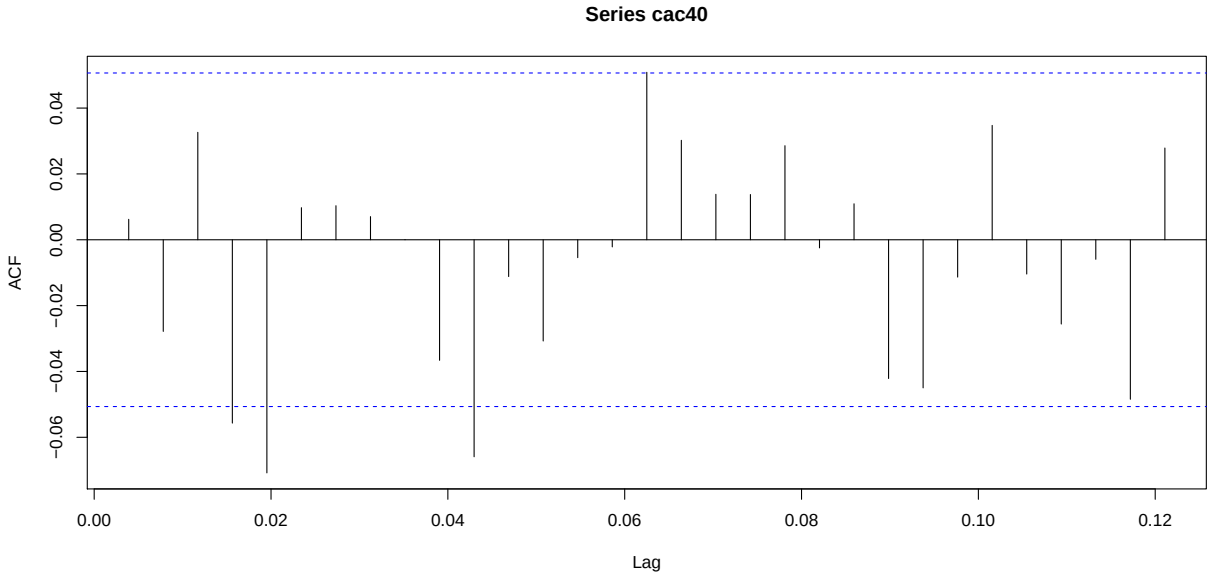
D'après D3 page 96 : Au regard de l'autocorrélogramme partiel du log return, une modélisation à l'aide d'un modèle GARCH(1, 1) semblerait possible. En effet, l'autocorrélogramme et l'autocorrélogramme partiel sont significativement nuls à partir des premiers retards (i.e. p = q = 1) Dans notre cas vu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

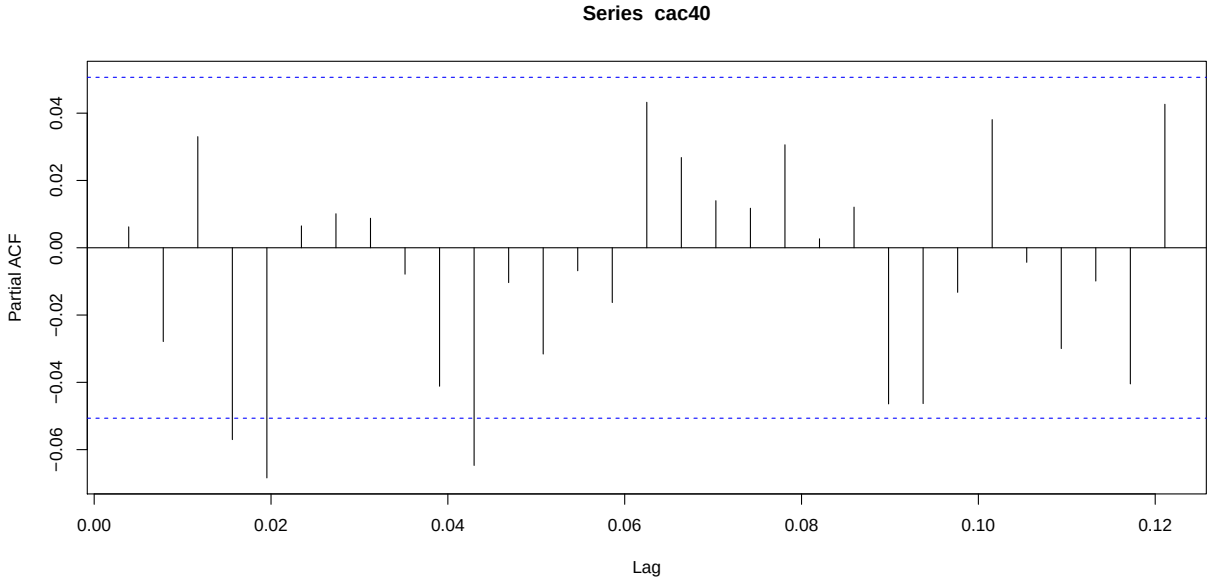
- Obtention de la série des log return

```
cac40 <- 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
```

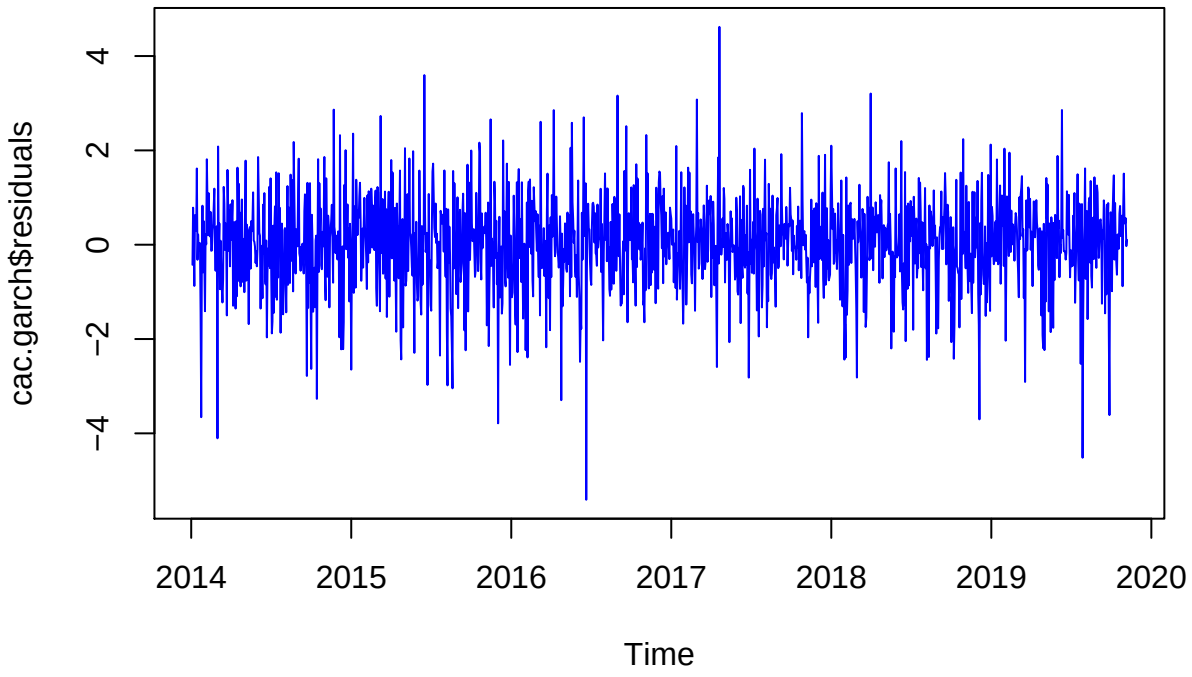
```
##      3   17 -6.095e+03  3.06e-03  4.45e-03  4.5e-01  2.0e+00  8.0e-02  9.30e+00
##      4   20 -6.130e+03  5.64e-03  4.20e-03  7.4e-01  1.9e+00  3.2e-01  5.37e-01
##      5   22 -6.140e+03  1.66e-03  1.50e-03  7.8e-02  2.0e+00  6.4e-02  1.06e+02
##      6   24 -6.162e+03  3.64e-03  3.43e-03  1.3e-01  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 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/