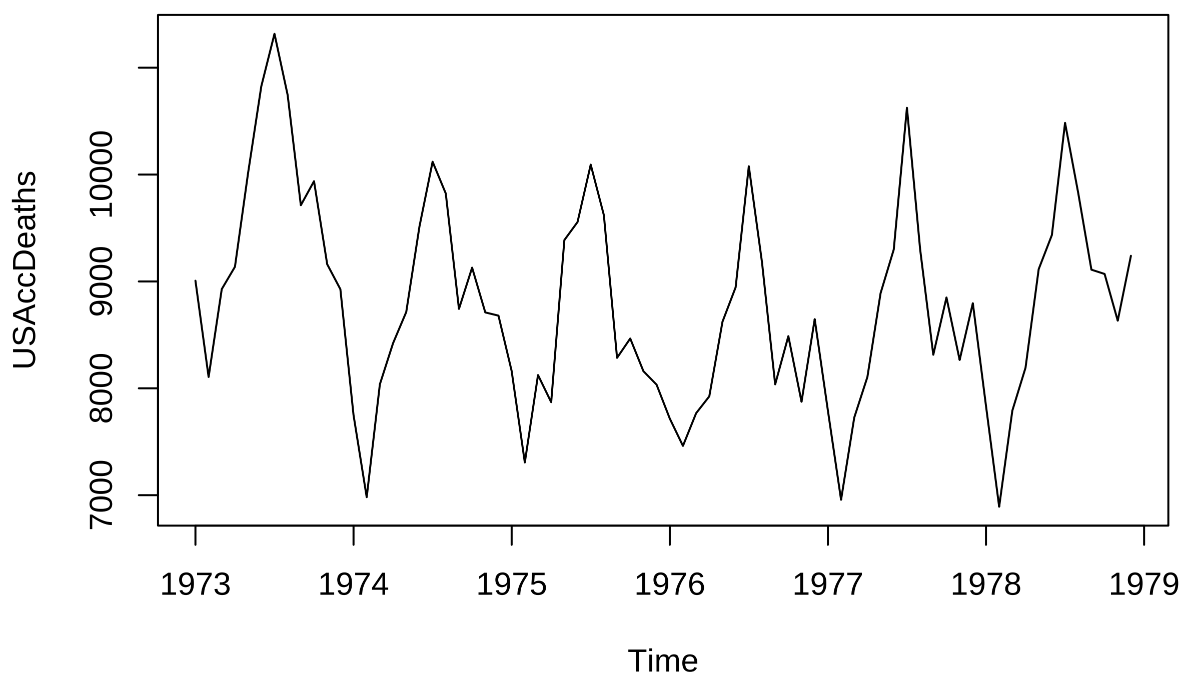
Projet Séries Temporelles

CONSTANZA Corentin HOUFAF KHOUFAF RAFIQ

Introduction

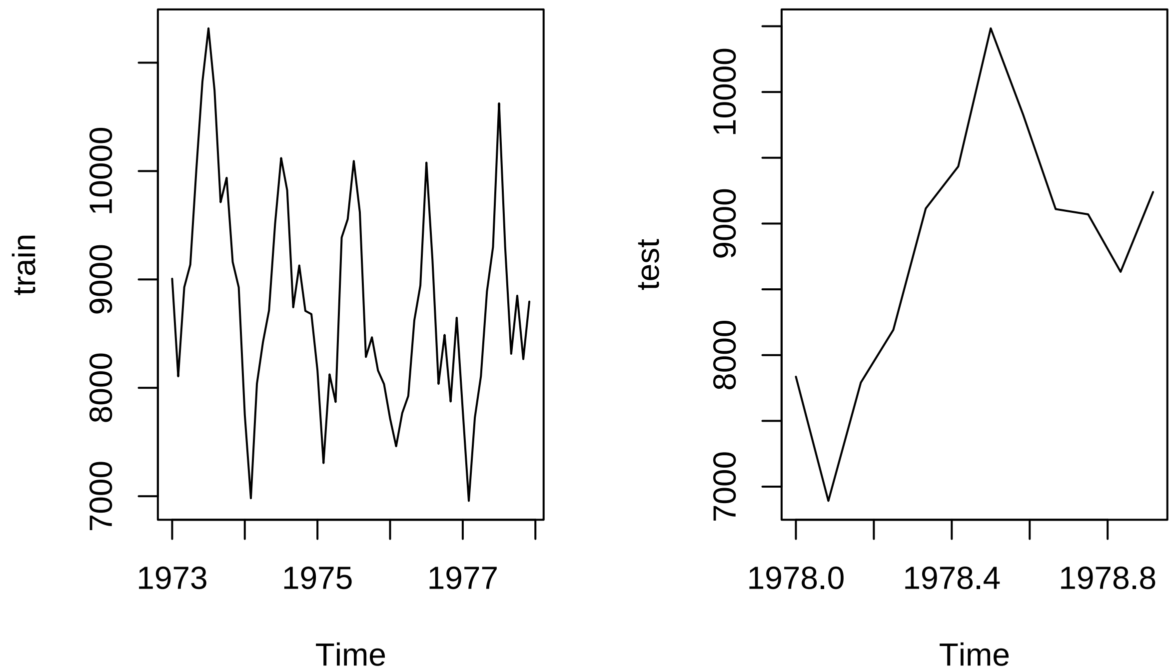
Cette étude vise à etudier une série temporelle afin de comprendre ses variations ainsi que le phénomène qu'elle représente. On cherchera ensuite à prédire des valeurs futures par le biais de plusieurs modélisations : - SARIMA - Lissage exponentiel - Régression Linéaire

Ci-dessous, le chronogramme de la série temporelle que nous allons modéliser.



Ces données correspondent à la au nombre d'accident de voiture entre 1973 et 1979. Nous pouvons observer que cette série ne **semble** pas avoir de tendance une variance qui augmente. On va donc privilégier un modèle multiplicatif. Nous n'avons donc pas besoin de passer la serie au Log. On observe aussi une saisonalitée annuelle (12 mois).

Pour commencer, nous allons diviser notre série en deux parties : *train* et *test*. La partie train nous servira pour entrainer nos modèles, et la partie test nous servira pour verifier la précision de nos prédictions avant d'appliquer notre modèle sur tout le jeu de données.



#Partie 1 : SARIMA

Introduction à la méthode de Box-Jenkins.

La méthodologie la plus connue pour modéliser une série temporelle est la méthode de Box-Jenkins.

Elle suit les étapes suivantes :

- Stationnarisation de la série
- Identification des modèles potentiels
- Estimation des modèles potentiels
- Vérification des modèles potentiels
- Choix du modèle
- Prévision du modèle
- Analyse à Posteriori

Cross - Validation :

Pour pouvoir entrainer notre modèle et valider les prévisions nous allons couper notre série en deux parties.

USAccDeaths.M : Partie de la série dédiée à l'entrainement, contiendra les années et mois de janvier 1973 à décembre 1977.

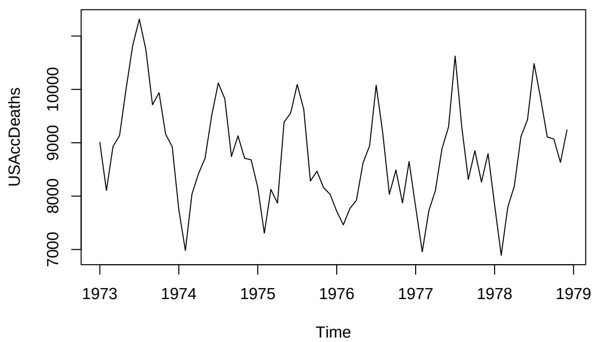
USAccDeaths.T : Partie de la série dédiée au test, contiendra la dernière année.

```
## [1] "USAccDeaths.T : "
```

```
##         Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
## 1978  7836  6892  7791  8192  9115  9434 10484  9827  9110  9070  8633  9240
```

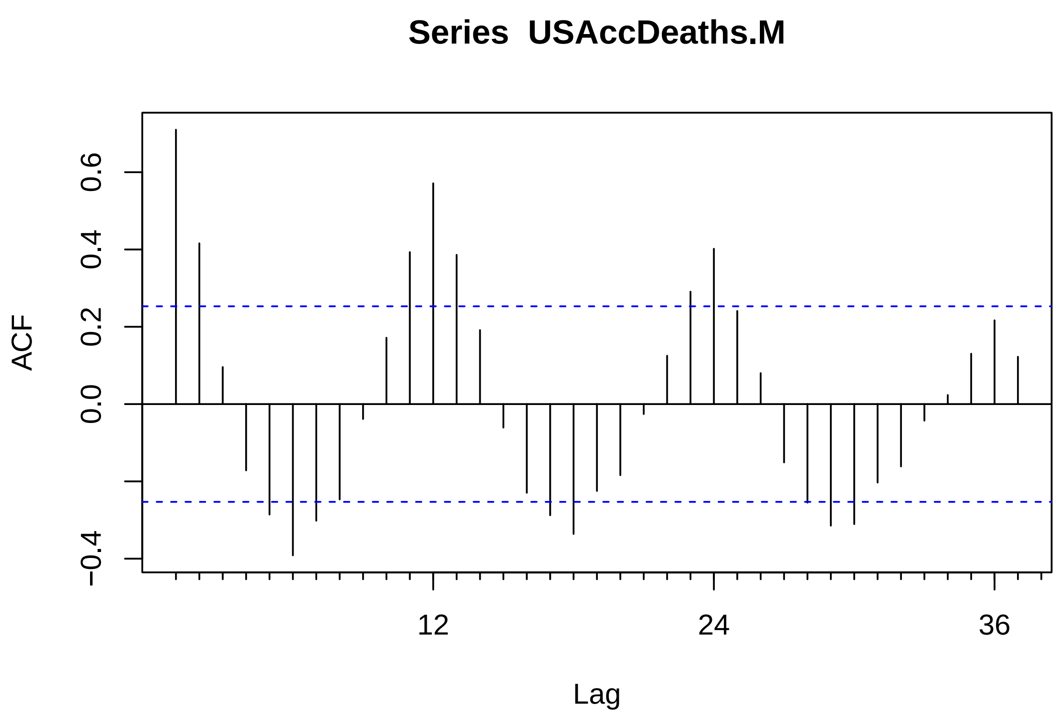
Stationnarisation de la série :

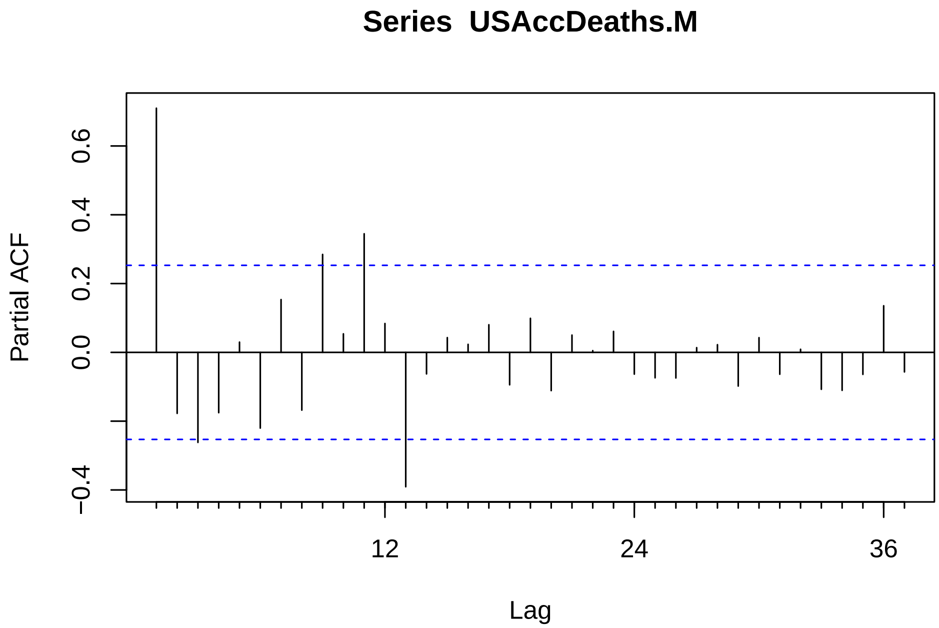
Tout d'abord, affichons la série et voyons ce qu'on peut en conjecturer :



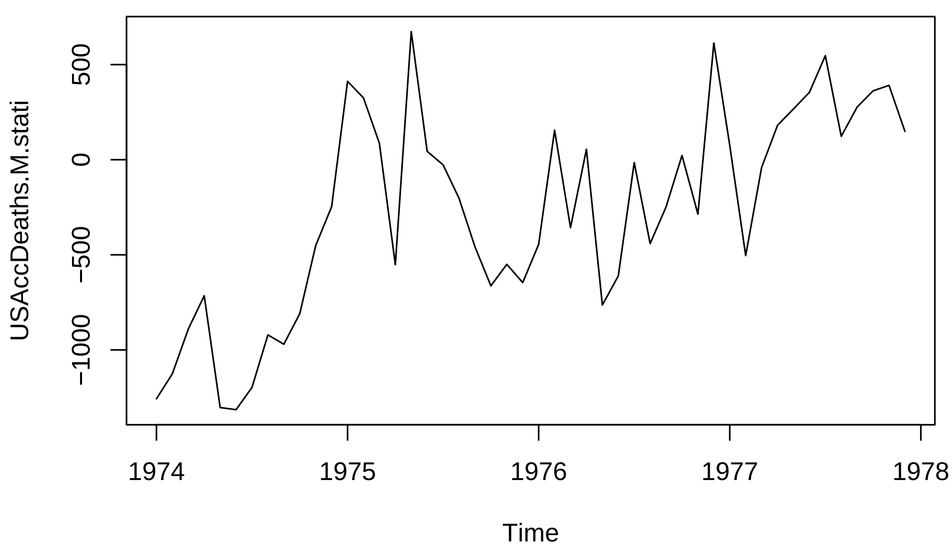
On peut conjecturer qu'il y'a une seasonalité douze mais il ne semble pas y avoir de tendance. La série garde aussi une variance relativement similaire tout au long, donc elle n'est pas hétéroscedastique il n'est pas donc pas nécessaire de passer la série au log ou à la racine.

Etudions à présent l'**ACF** et le **PACF** pour essayer de voir s'il y'a des possibilités de différenciation/stationnarisation.





On remarque, sur l'ACF, des pics qui se répétent tous les temps en douze. Ce qui indique une **séasonnalité en 12**. Nous allons donc différencier la série en 12, en appliquant : $\nabla_s^D = (I - B^s)^D$ avec $s = 12$ à la série.

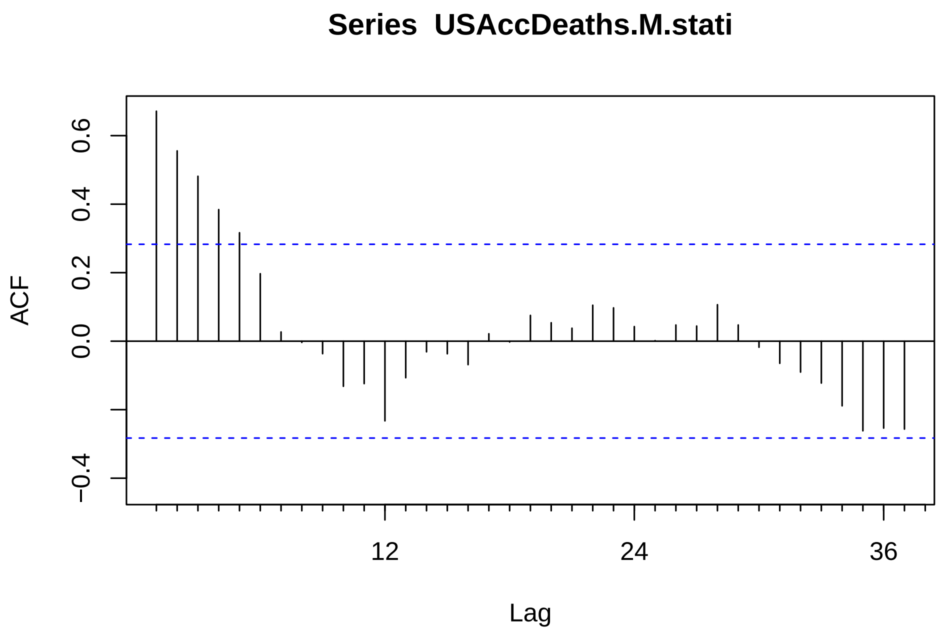


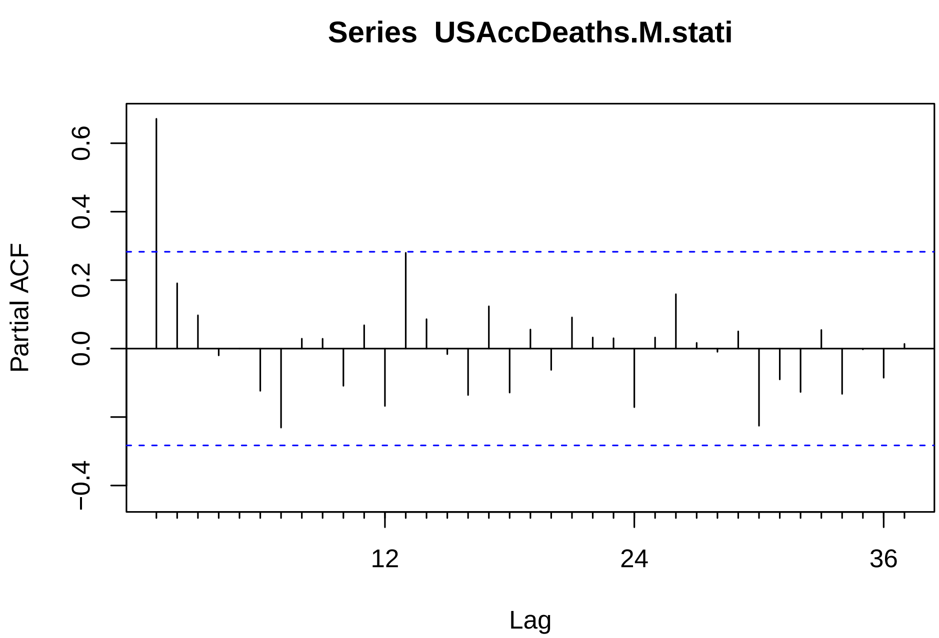
Modélisation de la série :

Pour la modélisation de la série, nous analyserons les ACF et PACF comme suite :

$Modèle$	ACF	$PACF$
$AR(p)$	$Décroissance\ exponentielle$	$0\ \forall h > p$
$MA(q)$	$0\ \forall h > q$	$Décroissance\ exponentielle$

Etudions à nouveau l'ACF et le PACF :



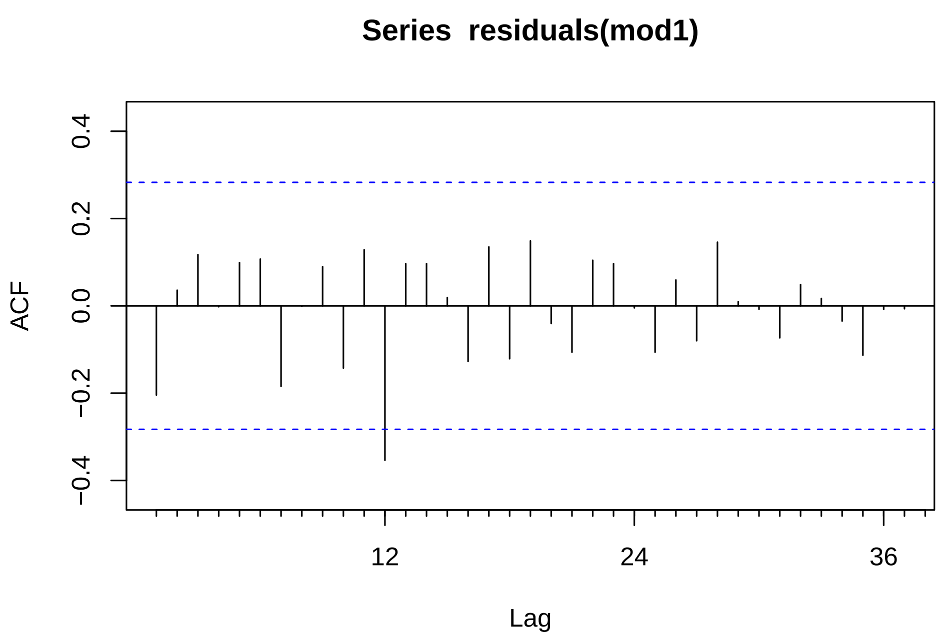


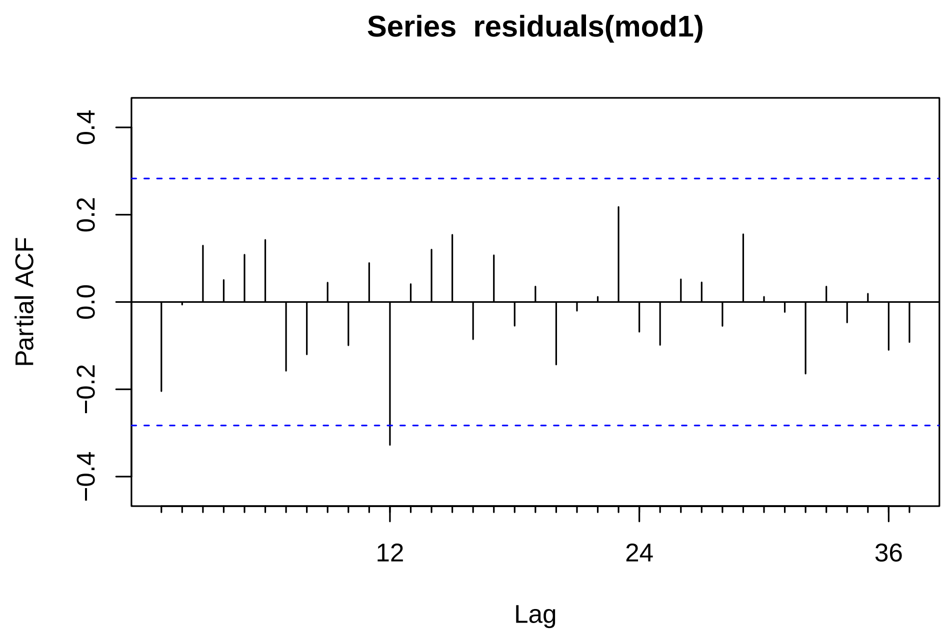
On a un ACF en décroissance exponentielle et PACF nul pour $\forall\ h > 1$. C'est les propriétés d'un AR(1).

On va donc modéliser notre série par un $AR(1)$.

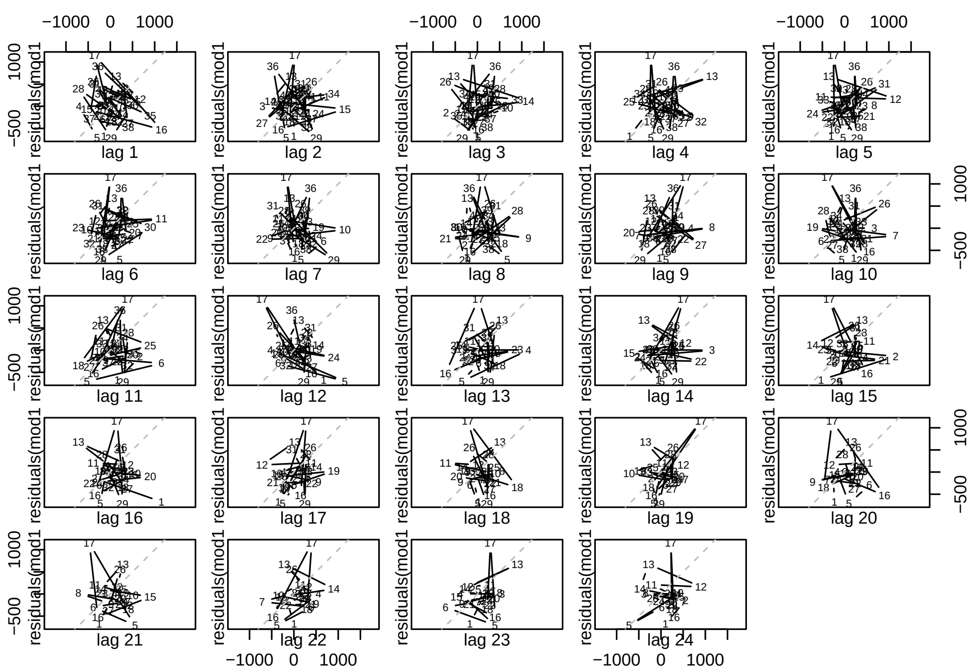
```
## Series: USAccDeaths.M.stati
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1       mean
##       0.7155  -295.9645
## s.e.  0.1037   185.7716
##
## sigma^2 = 153809:  log likelihood = -354.09
## AIC=714.18   AICc=714.72   BIC=719.79
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
## Training set 20.41559 383.9269 295.0419 -35.52115 218.1819 0.4385067 -0.2043327
```

Etude des résidus : A présent nous allons étudier l'ACF et le $PACF$ des résidus, afin de vérifier si il y'a possibilité d'affiner le modèle.





Analysons le lagplot des résidus :

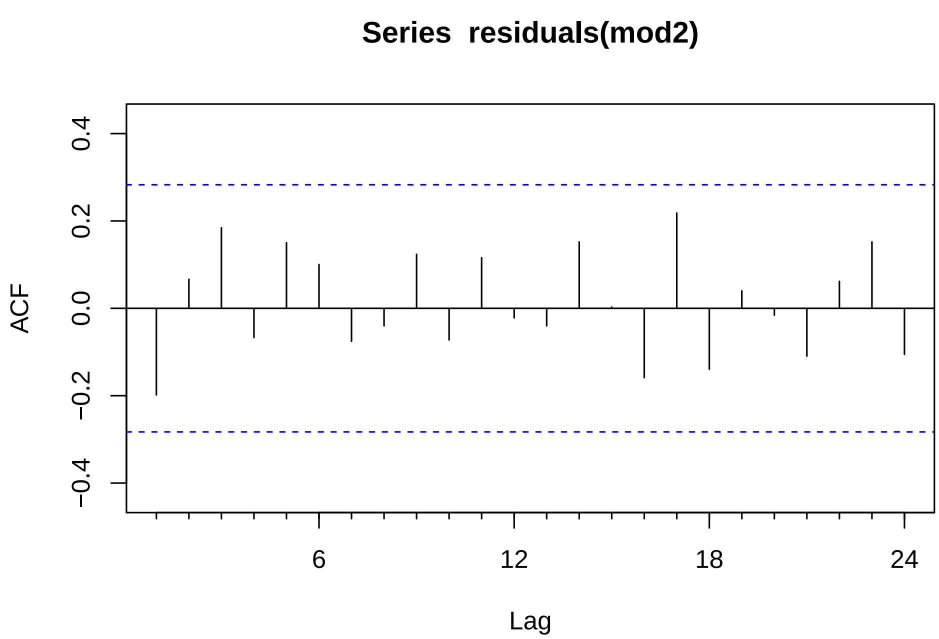


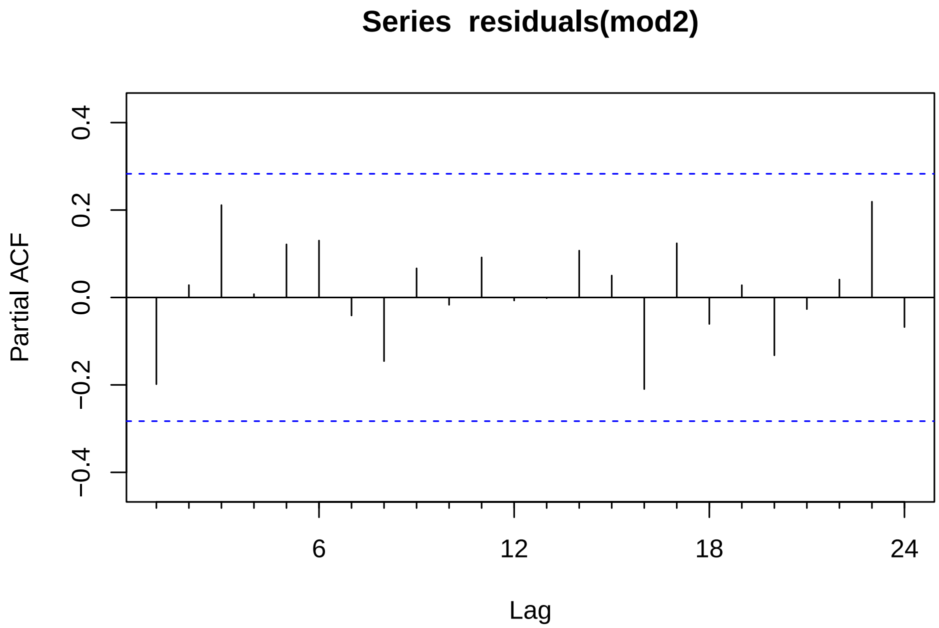
Deuxième modèle :

On remarque un ACF nul pour $\forall\ h > 12$. On peut ajouter un $MA(12)$ à nos résidus, d'où ce deuxième modèle :

```
## Series: USAccDeaths.M.stati
## ARIMA(1,0,0)(0,0,1)[12] with non-zero mean
##
## Coefficients:
##          ar1     sma1       mean
##       0.7190  -0.5431  -246.7623
## s.e.  0.1066   0.2169   105.4082
##
## sigma^2 = 121599:  log likelihood = -350
## AIC=708   AICc=708.93   BIC=715.48
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -22.27174 337.6378 265.0958 -38.38768 179.0221 0.3939992
##                    ACF1
## Training set -0.1984447
```

Etudions les résidus :





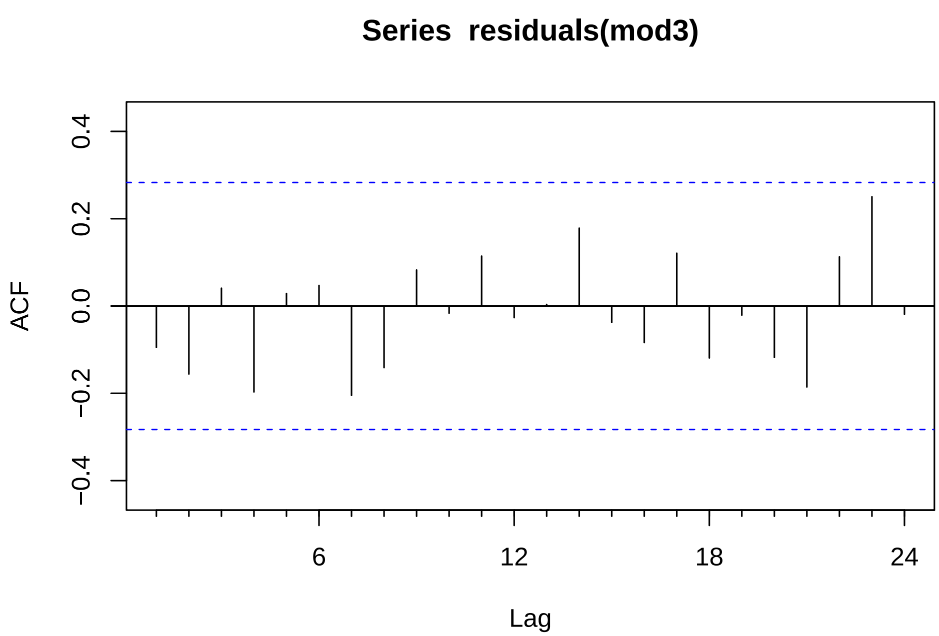
L'ACF et le PACF sont tous les deux nuls, il n'y a plus d'affinage possible sur ce modèle. On obtient des **résidus bruits blancs**.

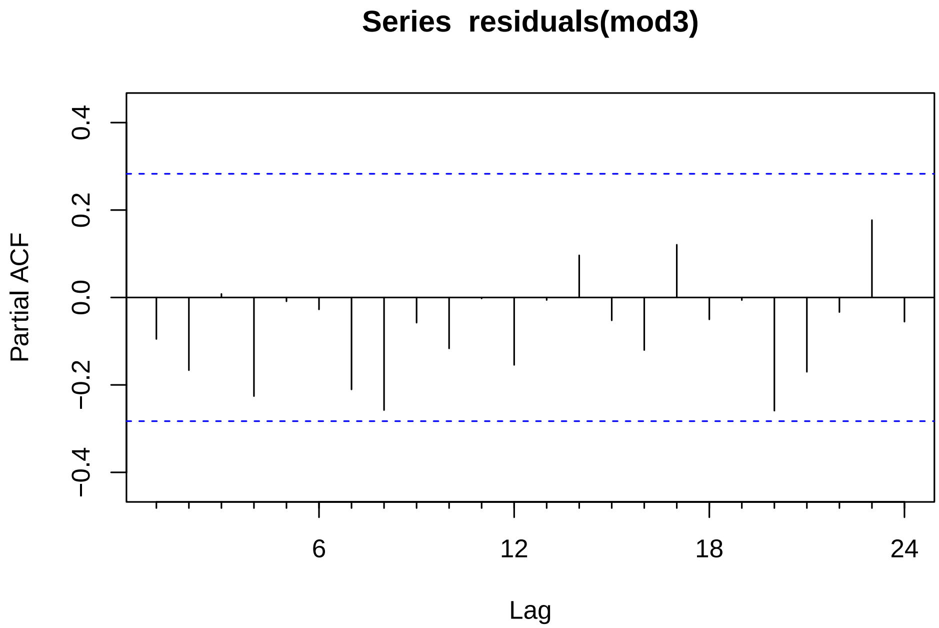
Troisième modèle :

On remarque qu'une fois différenciée en seasonalité, la série présente une tendance linéaire. Nous allons donc appliquer l'opérateur $\nabla^d = (I - B)^d$, en plus des modifications précédemments effectuées :

```
## Series: USAccDeaths.M.stati
## ARIMA(1,1,0)(0,0,1)[12]
##
## Coefficients:
##           ar1     sma1
##       -0.3516  -0.4868
## s.e.   0.1352   0.1884
##
## sigma^2 = 121794:  log likelihood = -342.54
## AIC=691.08   AICc=691.64   BIC=696.63
##
## Training set error measures:
##                    ME     RMSE     MAE       MPE     MAPE      MASE        ACF1
## Training set 58.55308 337.9079 265.695 -31.94911 199.7225 0.3948898 -0.09492094
```

Etude des résidus :



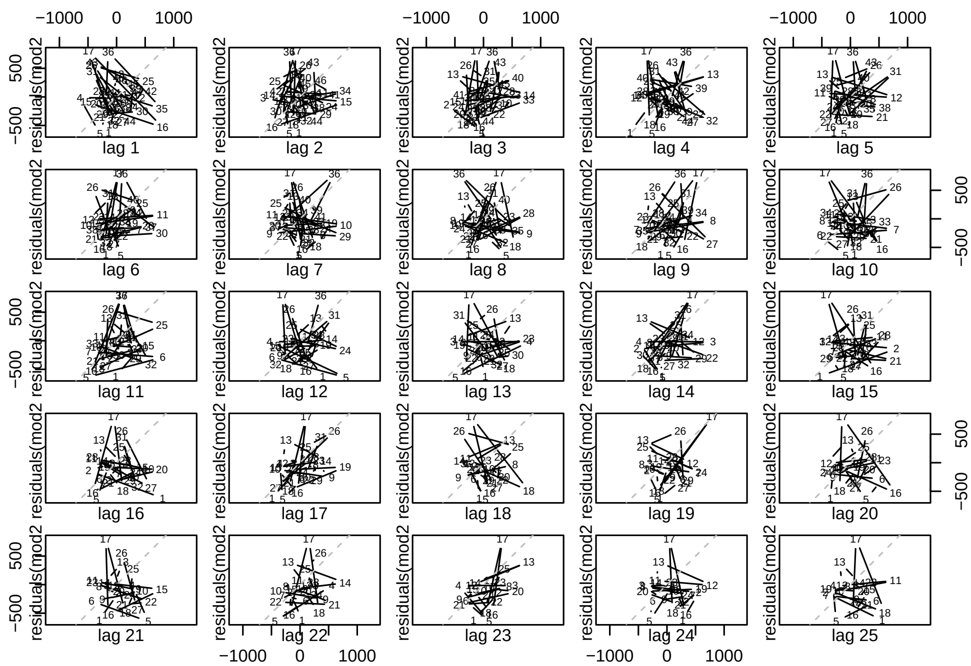


Les résidus sont **des bruits blancs** pour les mêmes raisons que précédemment.

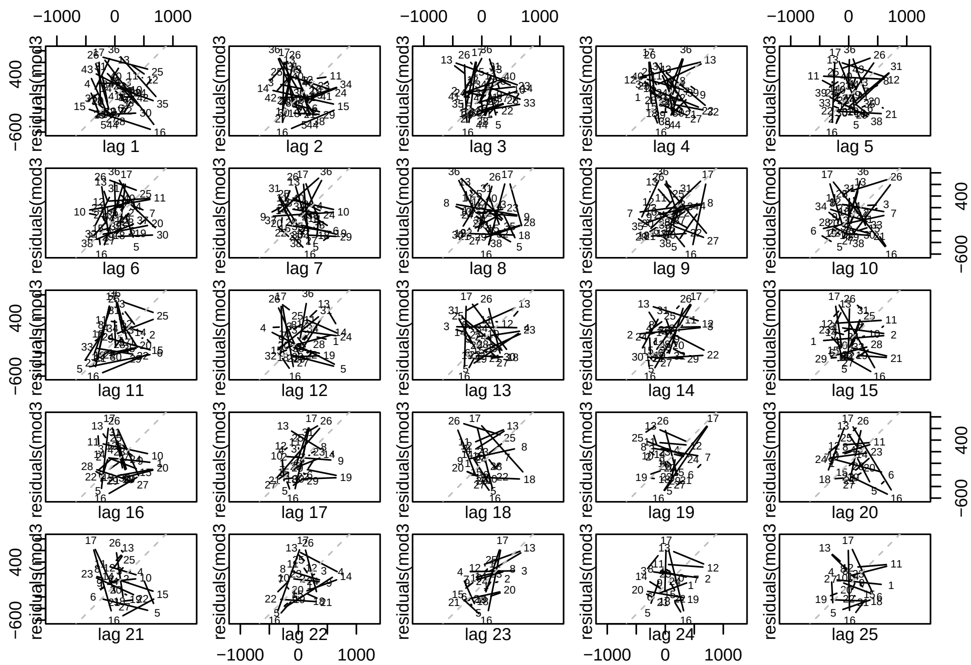
Choix du modèle :

Nous avons à présent deux modèles valides (dont les résidus sont des bruits blancs). Afin de les départager nous allons faire plusieurs vérifications. Premièrement, que les résidus sont bien des bruits blancs via le lag-plot. Deuxièmement, que les coefficients de la partie AR ne sont pas proche de 1. Et finalement, étudier le critère d'Akaike (AIC).

Vérification des résidus :



Modèle 2 :



Modèle 3 :

Conclusion : Les lag-plots montre qu'il n'a aucune corrélation entre les résidus des deux modèles, ce n'est donc pas possible de les départager avec cette première vérification.

Coefficients de la partie AR :

modèle 2 :

```
##           ar1          sma1    intercept
##     0.7190289   -0.5431059 -246.7623487
```

modèle 3 :

```
##        ar1       sma1
## -0.3516217 -0.4868450
```

Conclusion : Les deux modèles ont des coefficients AR assez loin de 1. Donc la partie stationnaire modélisée est bien stationnaire.

Critère d'Akaike :

Le critère d'Akaike, $AIC(p,q) = log(\sigma^2) + 2\frac{p+q}{T}$, en minimisant ce critère cela permettra d'avoir à la fois le modèle le **mieux ajustée** et le plus **parcimonieux**.

```
## [1] "AIC du modèle 2 : 707.999834"
```

```
## [1] "AIC du modèle 3 : 691.077229"
```

C'est le troisième modèle qui présente l'AIC le plus faible.

Conclusion de la comparaison des modèles :

Nos deux modèles sont valides; leurs résidus sont bien des bruits blancs et les coefficients de la partie AR sont loin de 1. Ce qui va permettre de les départager c'est le critère d'Akaike, en effet, c'est le modèle 3 qui minimise ce critère et c'est pourquoi on va passer à la partie prévision avec ce modèle.

Prévisions du modèle :

Le modèle final est le suivant :

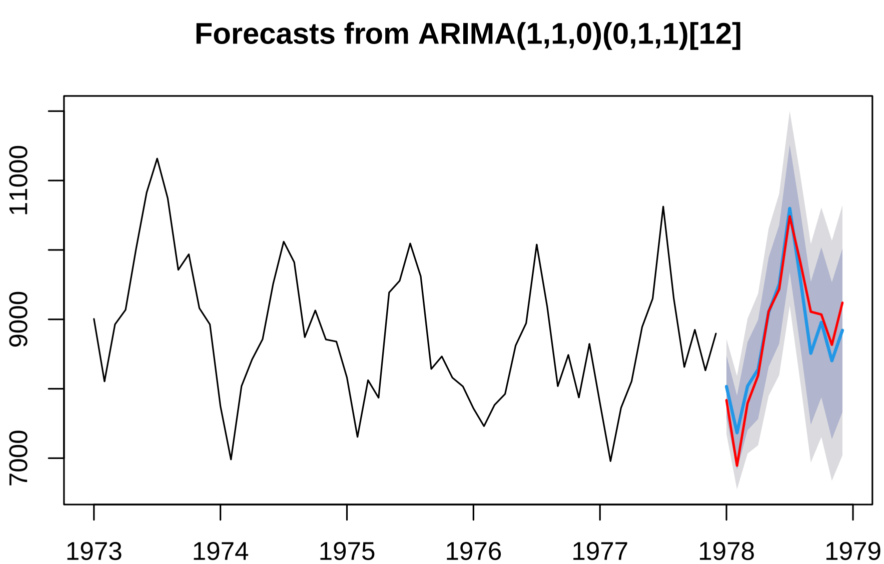
$\phi(\beta)\nabla^d X_t = \theta(\beta^s)\nabla_s^D \epsilon_t$

Avec :

- $\phi(B) = I - B$, partie $AR(1)$.
- $\nabla^d = (I - B)^d$, avec $d = 1$. Partie stationnarisation.
- $\theta(B^S) = (I - B^s)$, avec $s = 12$. $MA(12)$.
- $\nabla_s^D = (I - B^s)^D$, avec $s = 12$, $D = 1$. Partie différenciation.

Nous allons effectuer des prévisions sur 12 mois, et comparer les données prédites aux données réels.

```
## Series: USAccDeaths.M
## ARIMA(1,1,0)(0,1,1)[12]
##
## Coefficients:
##           ar1     sma1
##       -0.3516  -0.4869
## s.e.   0.1352   0.1884
##
## sigma^2 = 121822:  log likelihood = -342.54
## AIC=691.08   AICc=691.63   BIC=696.63
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE     MASE        ACF1
## Training set 46.44133 302.2689 213.6019 0.5238382 2.501191 0.443445 -0.08853031
```



Analyse à Posteriori :

Pour analyser la qualité de la prédiction, on va calculer le $RMSE$ et le $MAPE$.

- $RMSE = \sqrt(\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{X}_i)^2)$
- $MAPE = \sqrt(\frac{1}{n}\sum_{i=1}^{n}(|\frac{X_i - \hat{X}_i}{X_i}|))$

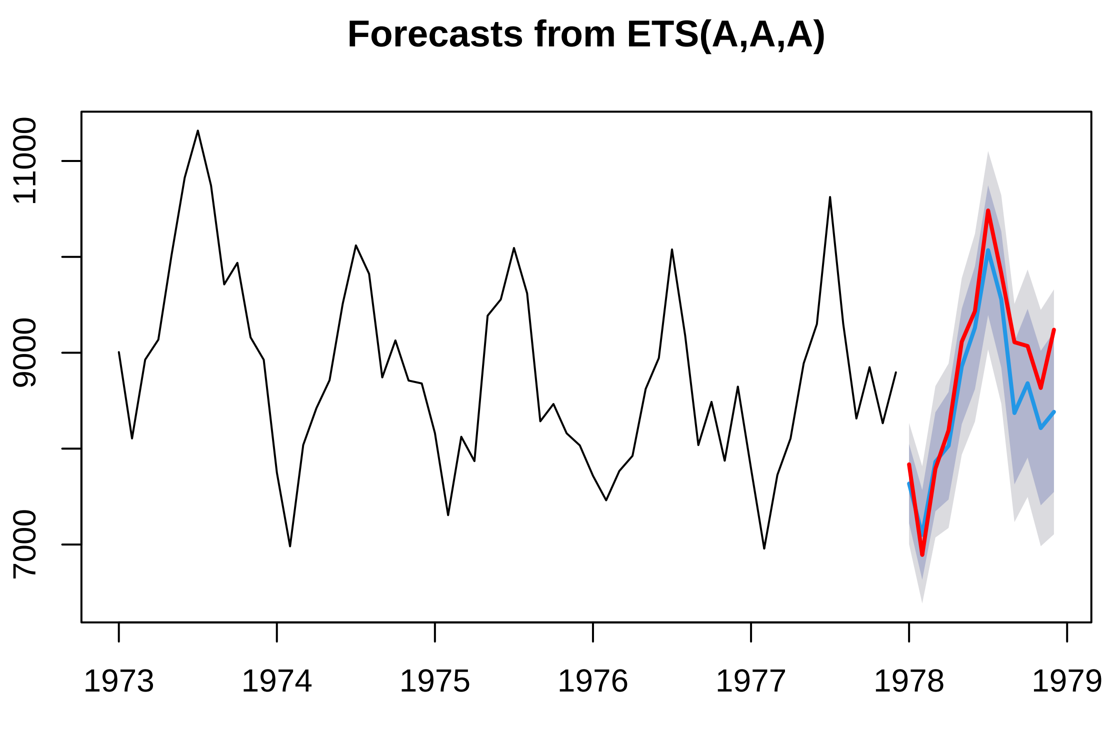
```
## [1] "RMSE : 287.149069"
```

```
## [1] "MAPE : 0.027322"
```

Partie 2 : Prédiction par Lissage Exponentiel

Comme la série étudiée possède une saisonalité, on sait que les lissages exponentiels simples et doubles ne sont pas adaptés. On va donc directement faire un lissage exponentiel de Holt-Winters. De plus, comme la variance de nos données ne semble pas augmenter avec le temps, on préférera la version additive à la version multiplicative.

Modèle de Holt-Winters additif sans amortissement




```
## ETS(A,A,A)
##
```

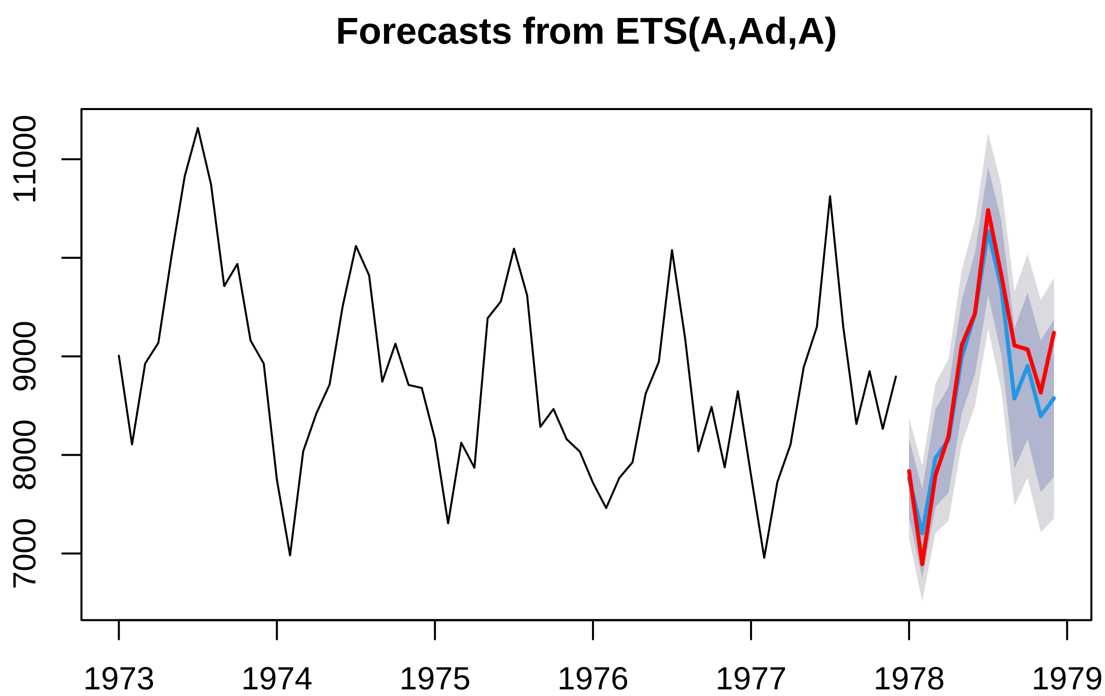
```
## Call:
##  ets(y = train, model = "AAA", damped = FALSE)
##
##   Smoothing parameters:
##     alpha = 0.5282
##     beta  = 1e-04
##     gamma = 1e-04
##
##   Initial states:
##     l = 9936.7902
##     b = -20.6133
##     s = -5.1696 -194.7218 252.405 -78.9006 1084.819 1578.012
##            748.6061 323.7468 -523.0428 -711.4698 -1494.45 -979.8336
##
##   sigma:  322.6749
##
##      AIC     AICc      BIC
## 954.2488 968.8203 989.8527
```

On regarde les critères AIC et BIC pour avoir une idée de la pertinence de notre modèle. On peut donc voir que l'on a un AIC de 954.2488 et un BIC de 989.8527 .

```
## [1] 414.4098
```

```
## [1] 0.03862091
```

Modèle de Holt-Winters additif avec amortissement




```
## ETS(A,Ad,A)
##
## Call:
##  ets(y = train, model = "AAA", damped = TRUE)
##
##   Smoothing parameters:
##     alpha = 0.5215
##     beta  = 1e-04
##     gamma = 1e-04
##     phi   = 0.9629
##
##   Initial states:
##     l = 9921.2144
##     b = -53.8002
##     s = -63.8679 -249.2478 256.4361 -77.4269 1045.642 1615.327
##            770.3008 327.1222 -516.2155 -709.3152 -1476.457 -922.2988
##
##   sigma:  311.428
##
##      AIC     AICc      BIC
## 950.6122 967.2951 988.3104
```

On regarde les critères AIC et BIC pour avoir une idée de la pertinence de notre modèle. On peut donc voir que l'on a un AIC de 950.6122 et un BIC de 988.3104 .

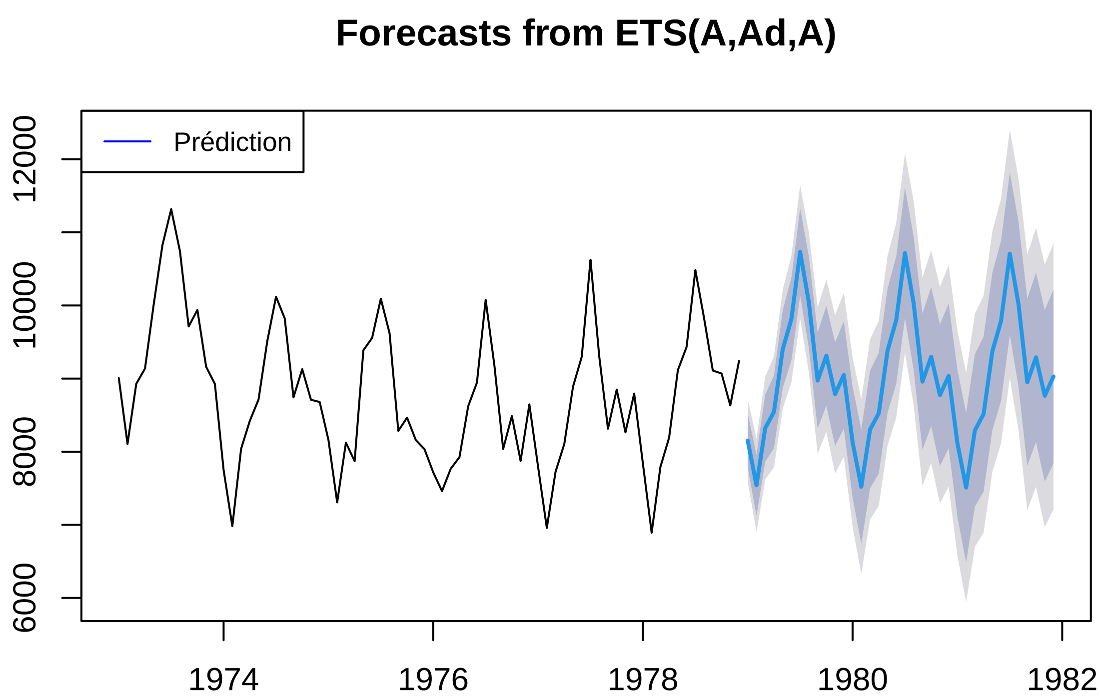
```
## [1] 292.3417
```

```
## [1] 0.02542135
```

On peut donc voir d'après les critères ci-dessus, que le modèle avec amortissement est meilleur que le modèle sans ammortissement.

Prédiction

Maintenant que l'on a sélectionné le modèle qui semblait être plus performant (additif avec ammortissement), on peut donc l'utiliser pour faire une prédiction sur les 3 prochaines années.

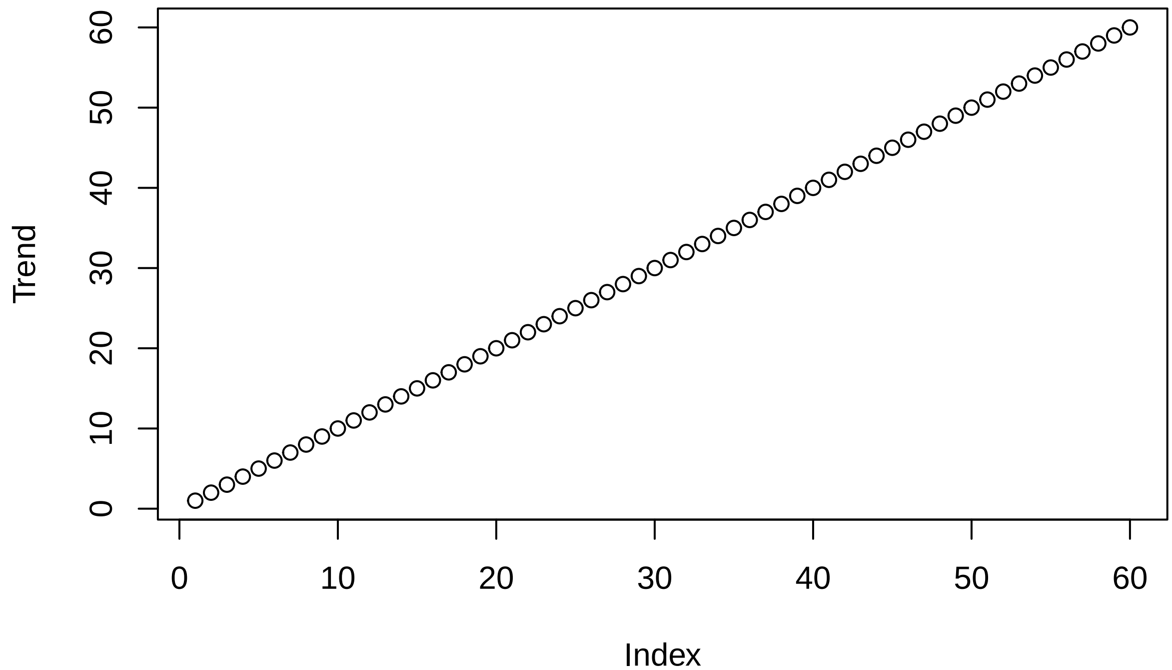


Partie 3 : Prédiction par régression

Dans cette dernière partie, nous souhaitons expliciter la tendance en fonction de t. Comme dit precedamment, on ne semble pas vraiment avoir de tendance. Cependant comme vu lors de la partie SARIMA, ceci est peut-être faux.

Nous allons commencer par supposer que la tendance est lineaire puis nous supposerons qu'elle est quadratic et nous compareront les résultats.

Tendance linéaire



La série étudiée possède une tendance linéaire et une saisonalité de 12. On aura donc :

$$T_t = a + bt \quad et \quad S_t = \alpha icos(\frac{2\pi it}{12}) + \beta isin(\frac{2\pi it}{12})$$

On va donc avoir une matrice de la forme :

$$\begin{pmatrix} 1 & 1 & cos(\frac{2\pi}{12}) & cos(\frac{2\pi*2}{12}) & . & sin(\frac{2\pi*1}{12}) & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & T & cos(\frac{2\pi T}{12}) & cos(\frac{2\pi*T*2}{12}) & . & sin(\frac{2\pi*T*1}{12}) & . \end{pmatrix}$$

```
##  [1]  1.224606e-16 -2.449213e-16  3.673819e-16 -4.898425e-16  6.123032e-16
##  [6] -7.347638e-16  8.572244e-16 -9.796851e-16  1.102146e-15 -1.224606e-15
## [11]  4.899781e-15 -1.469528e-15 -1.960725e-15 -1.714449e-15  5.389623e-15
## [16] -1.959370e-15 -1.470883e-15 -2.204291e-15  5.879466e-15 -2.449213e-15
## [21] -9.810403e-16 -9.799561e-15  6.369308e-15 -2.939055e-15 -4.911978e-16
## [26]  3.921451e-15  6.859151e-15 -3.428898e-15 -1.355253e-18 -1.077925e-14
## [31]  7.348993e-15 -3.918740e-15  4.884873e-16  2.941766e-15  7.838836e-15
## [36] -4.408583e-15  9.783298e-16 -1.175893e-14  8.328678e-15 -4.898425e-15
## [41]  1.567903e-14  1.962081e-15  8.818521e-15 -1.959912e-14  1.958015e-15
## [46] -1.273862e-14 -4.902491e-15 -5.878110e-15  1.665871e-14  9.823956e-16
## [51]  9.798206e-15  7.842902e-15  2.937700e-15 -1.371830e-14 -3.922806e-15
## [56] -6.857796e-15  1.763840e-14  2.710505e-18  1.077789e-14 -2.155849e-14
```

On peut voir que la 12ème colonne est d'ordre 10ˆ-14 à 10ˆ-18. On va donc la retirer.

On fait la matrice de régression :

```
##   V1            V2   V3            V4   V5            V6 V7           V8
## 1  1  8.660254e-01  0.5  6.123032e-17 -0.5 -8.660254e-01 -1 5.000000e-01
## 2  2  5.000000e-01 -0.5 -1.000000e+00 -0.5  5.000000e-01  1 8.660254e-01
## 3  3  6.123032e-17 -1.0 -1.836910e-16  1.0  3.061516e-16 -1 1.000000e+00
## 4  4 -5.000000e-01 -0.5  1.000000e+00 -0.5 -5.000000e-01  1 8.660254e-01
## 5  5 -8.660254e-01  0.5  3.061516e-16 -0.5  8.660254e-01 -1 5.000000e-01
## 6  6 -1.000000e+00  1.0 -1.000000e+00  1.0 -1.000000e+00  1 1.224606e-16
##              V9           V10           V11           V12
## 1  8.660254e-01  1.000000e+00  8.660254e-01  5.000000e-01
## 2  8.660254e-01  1.224606e-16 -8.660254e-01 -8.660254e-01
## 3  1.224606e-16 -1.000000e+00 -2.449213e-16  1.000000e+00
## 4 -8.660254e-01 -2.449213e-16  8.660254e-01 -8.660254e-01
## 5 -8.660254e-01  1.000000e+00 -8.660254e-01  5.000000e-01
## 6 -2.449213e-16  3.673819e-16 -4.898425e-16  6.123032e-16
```

Ce qui nous donne le model suivant :

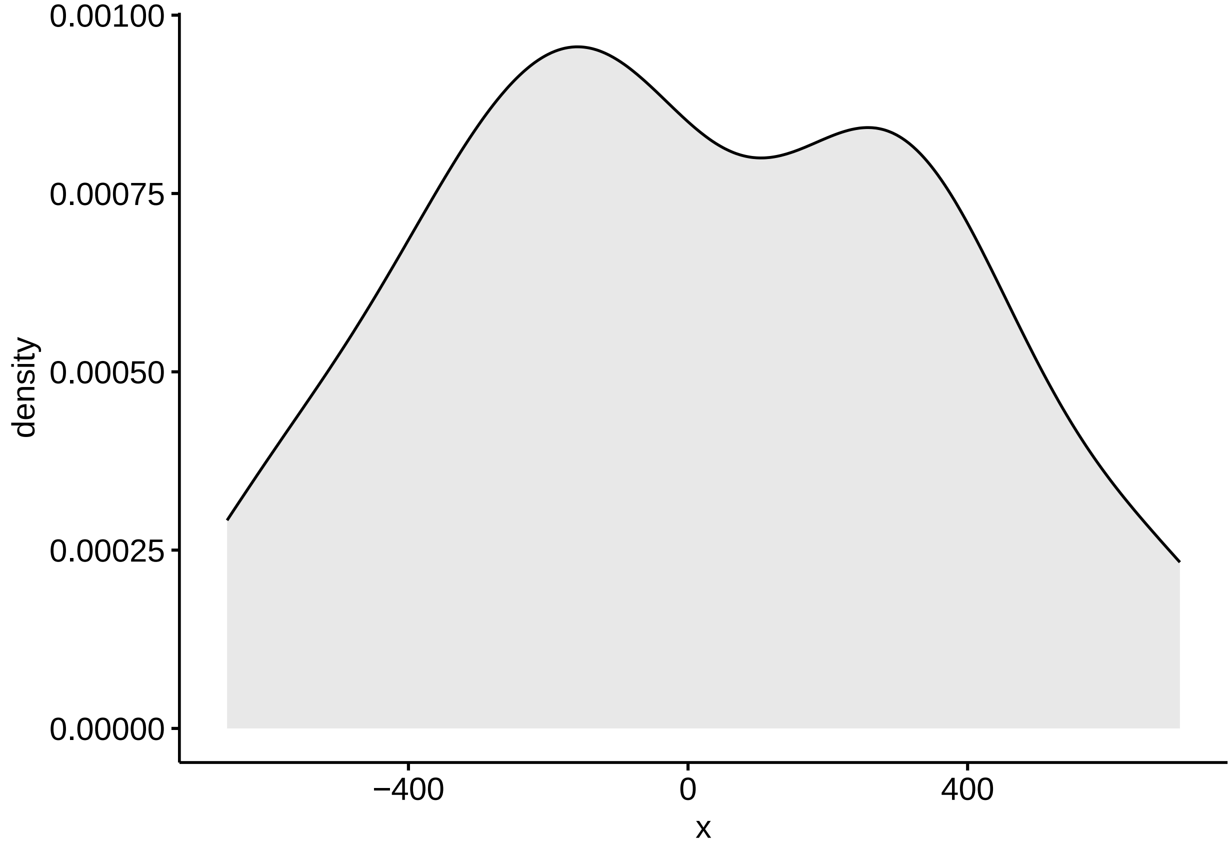
```
##
## Call:
## lm(formula = train ~ ., data = Regresseur)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -659.5 -282.3  -63.9  285.3  703.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9414.768    106.190  88.660  < 2e-16 ***
## V1           -20.610      3.042  -6.774 1.80e-08 ***
## V2          -729.314     73.081  -9.980 3.42e-13 ***
## V3           416.377     73.081   5.697 7.70e-07 ***
## V4           133.677     73.081   1.829  0.07372 .
## V5           -20.690     73.081  -0.283  0.77834
## V6           152.069     73.081   2.081  0.04292 *
## V7            -8.111     51.654  -0.157  0.87589
## V8          -740.368     73.895 -10.019 3.01e-13 ***
## V9            80.003     73.207   1.093  0.28004
## V10         -215.110     73.081  -2.943  0.00503 **
## V11          152.992     73.039   2.095  0.04162 *
## V12          212.527     73.022   2.910  0.00550 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 399.9 on 47 degrees of freedom
## Multiple R-squared:  0.8618, Adjusted R-squared:  0.8265
## F-statistic: 24.42 on 12 and 47 DF,  p-value: 3.382e-16
```

```
## Le chargement a nécessité le package : ggplot2
```

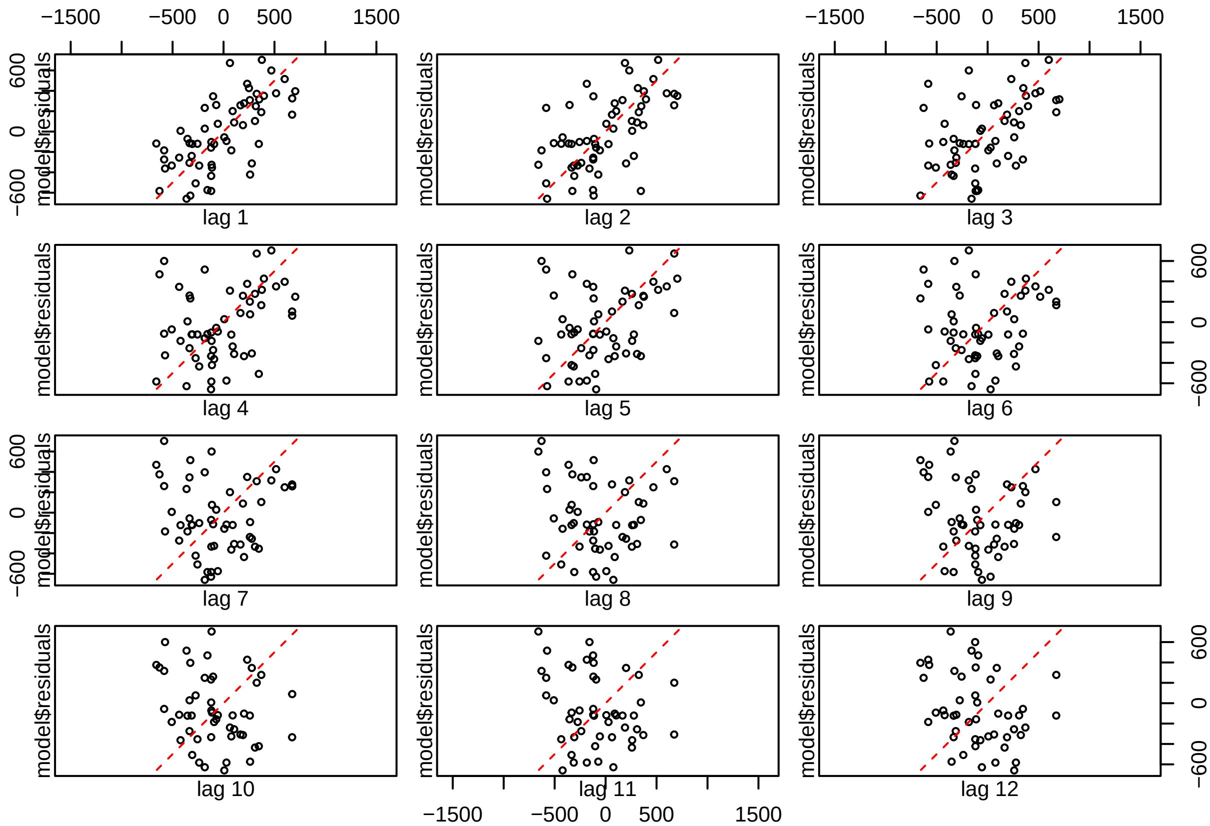
```
##
## Attachement du package : 'ggpubr'
```

```
## L'objet suivant est masqué depuis 'package:forecast':
##
##     gghistogram
```

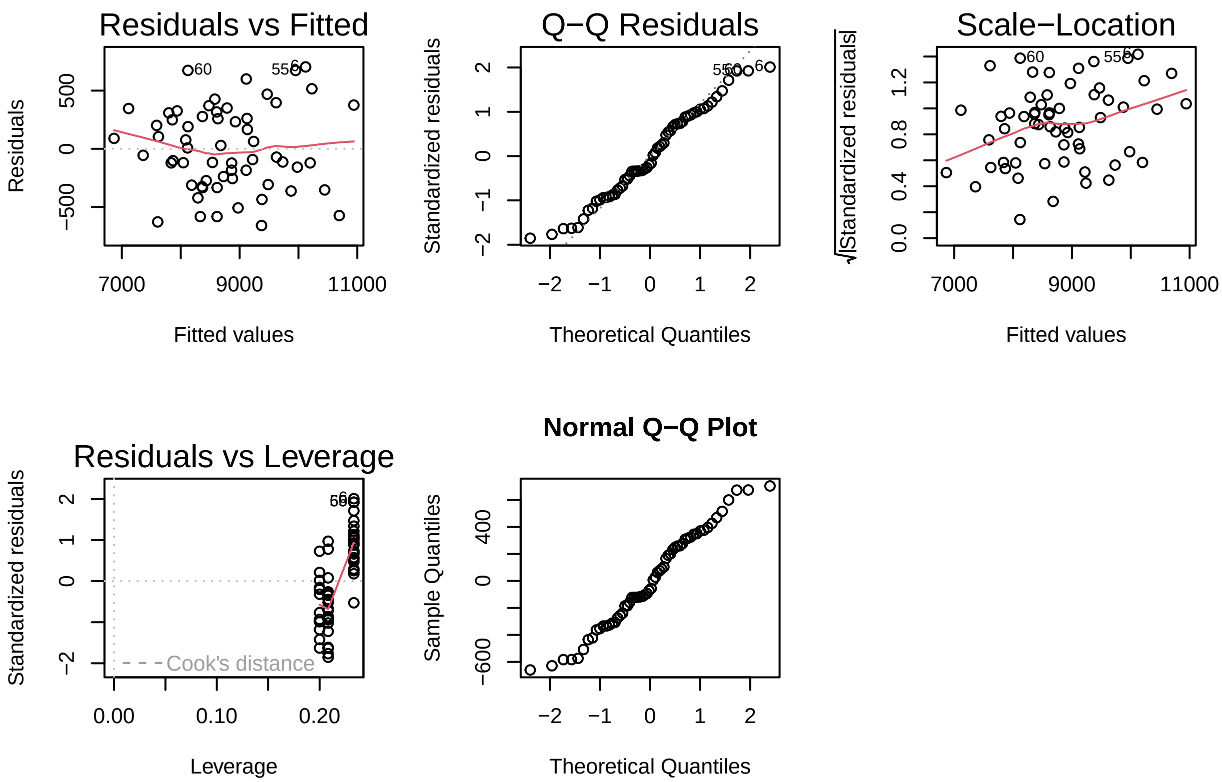
Voici la densité des résidu :



Cette densité nous montre que nos résidus ne semble pas distribués de manière gausienne et il semble cependant centrés en 0. Vérifions que nos résidus sont des bruits blancs(peut probable au vu du graph précedent).



−1500
−500
0
500
1500
600
0
−600
model$residuals
lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12

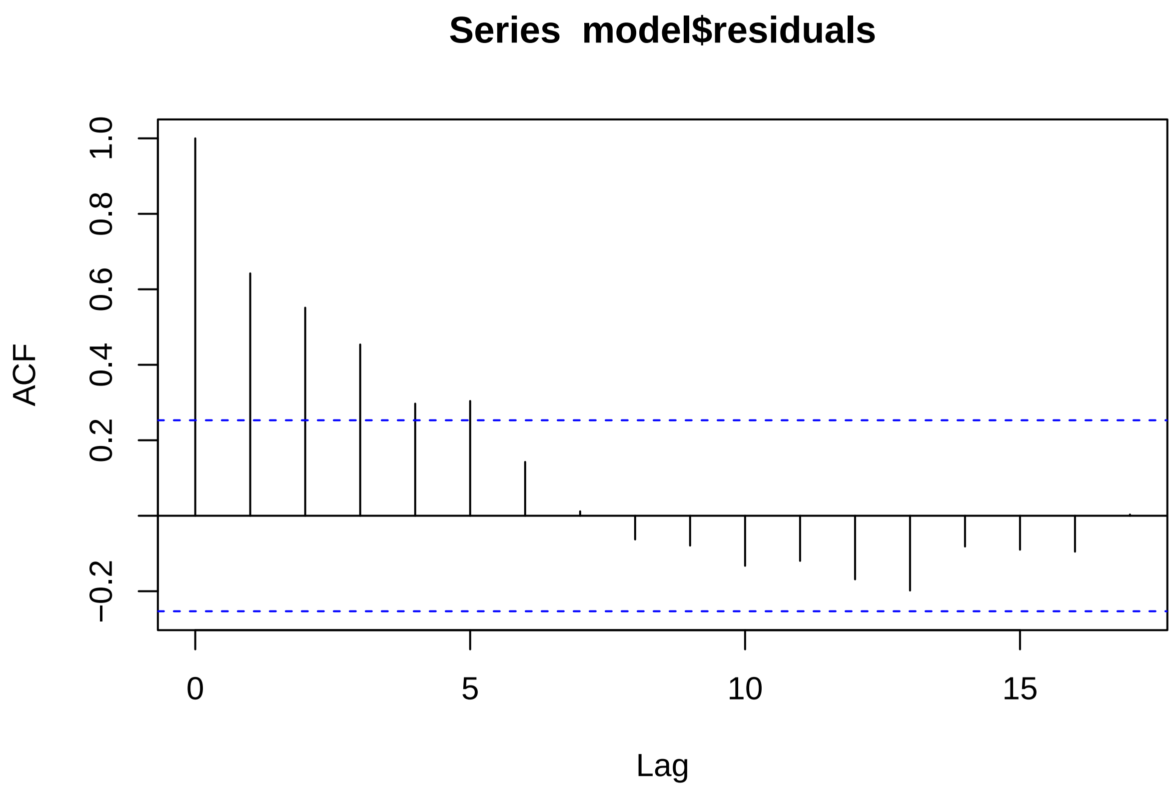


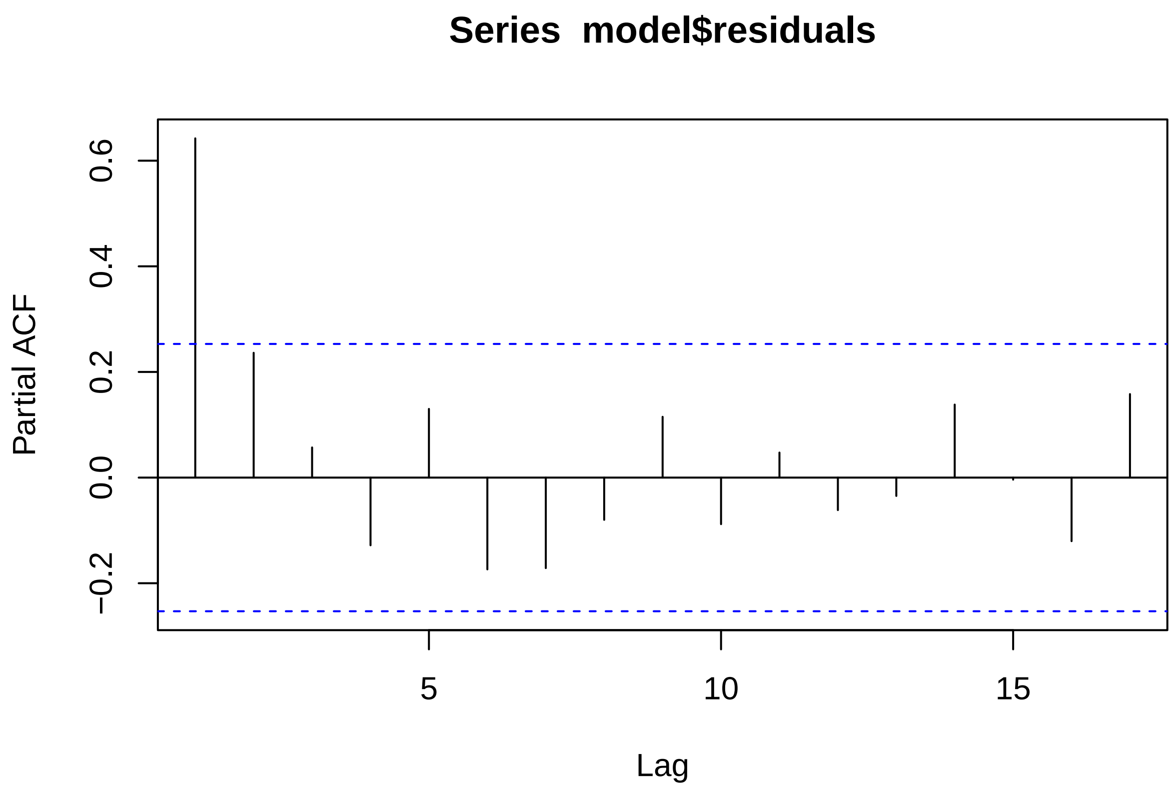
On voit que les résidus suivent globalement une droite sur au moins 3 des lagplots. De même le Q-Q plot ne semble pas correspondre à ce que l'on attends d'un bruit blanc (une belle droite en x=y).

Ils sont donc corrélés et ce ne sont pas des bruit blanc.

On va essayer de décorréler les résidus afin d'avoir un bruit blanc.

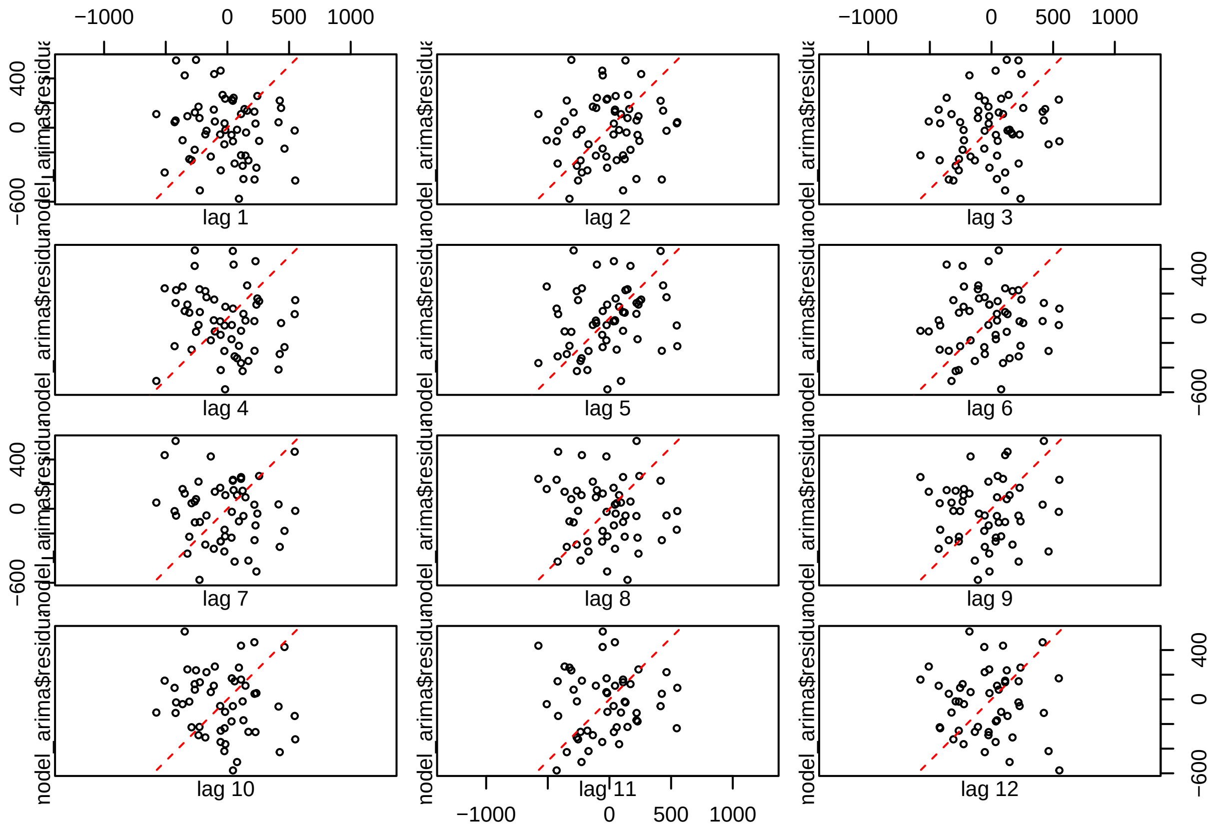
Traçons l'ACF et le PACF.


Series model$residuals
ACF
1.0
0.8
0.6
0.4
0.2
−0.2
0
5
10
15
Lag

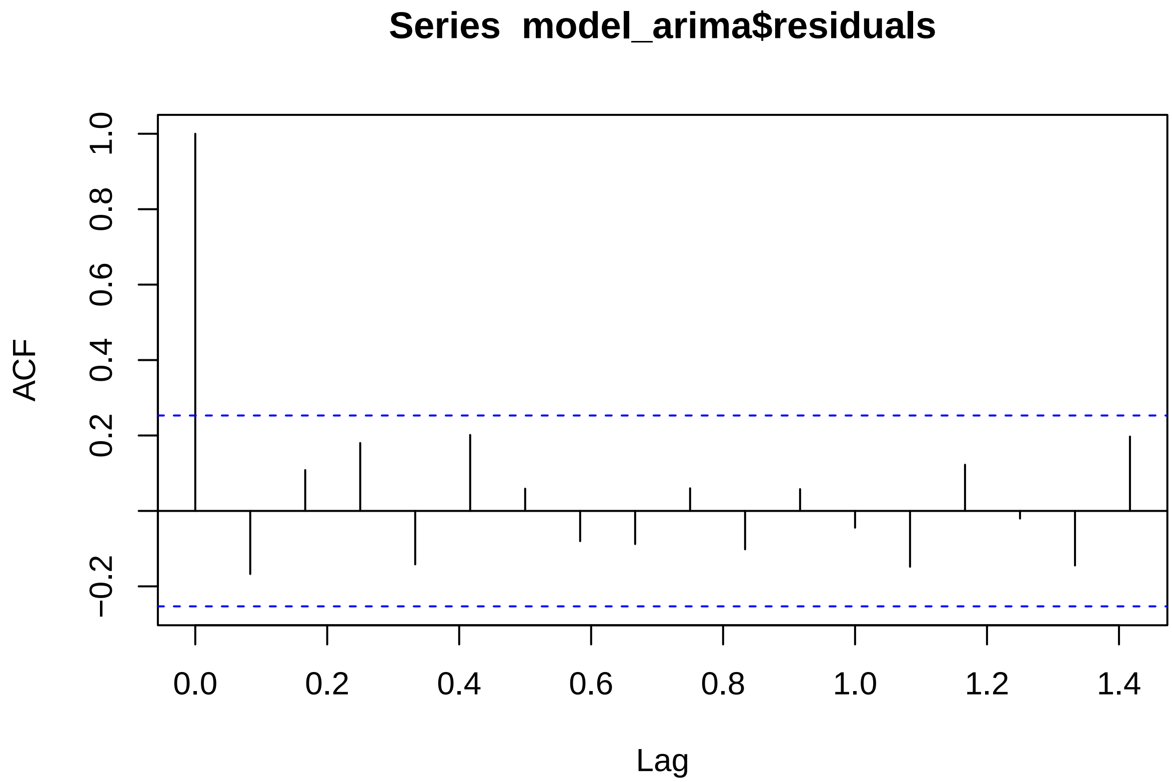


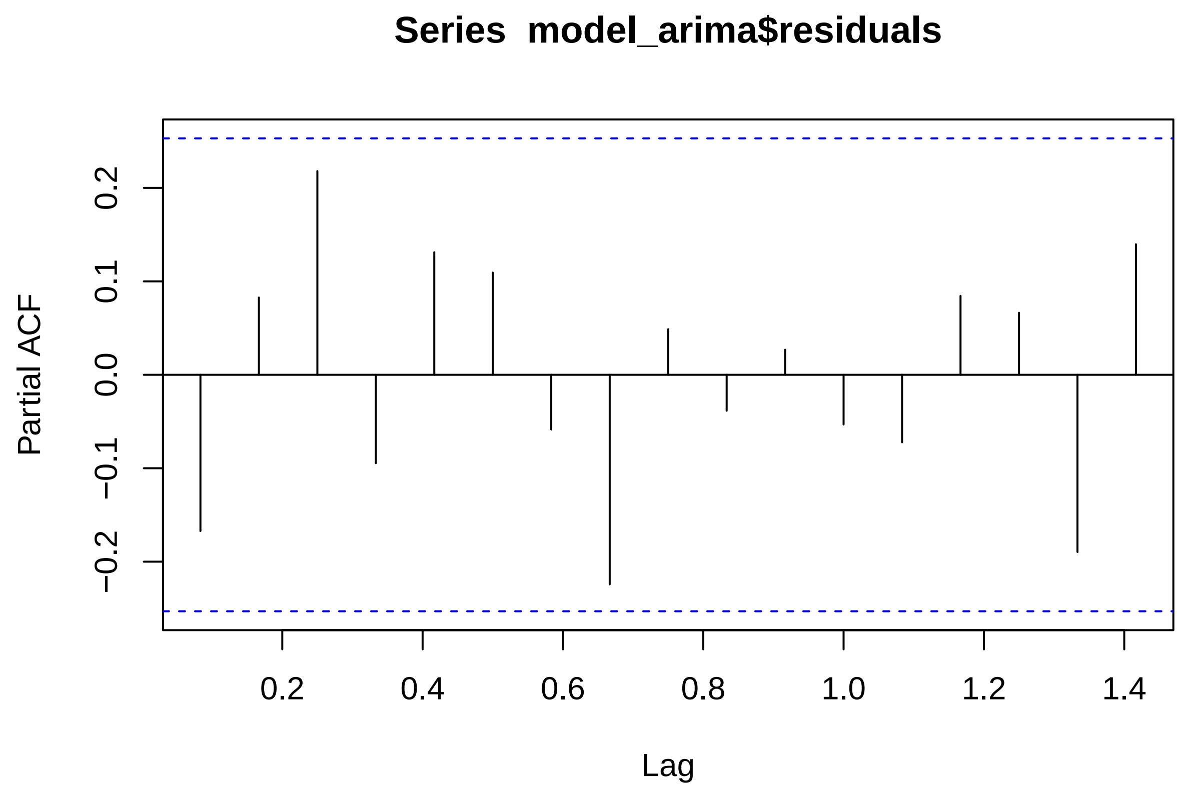
En suivant les mêmes rêgles que lors de la partie I on peut voir qu'il semble judicieux d'utiliser un AR(1):

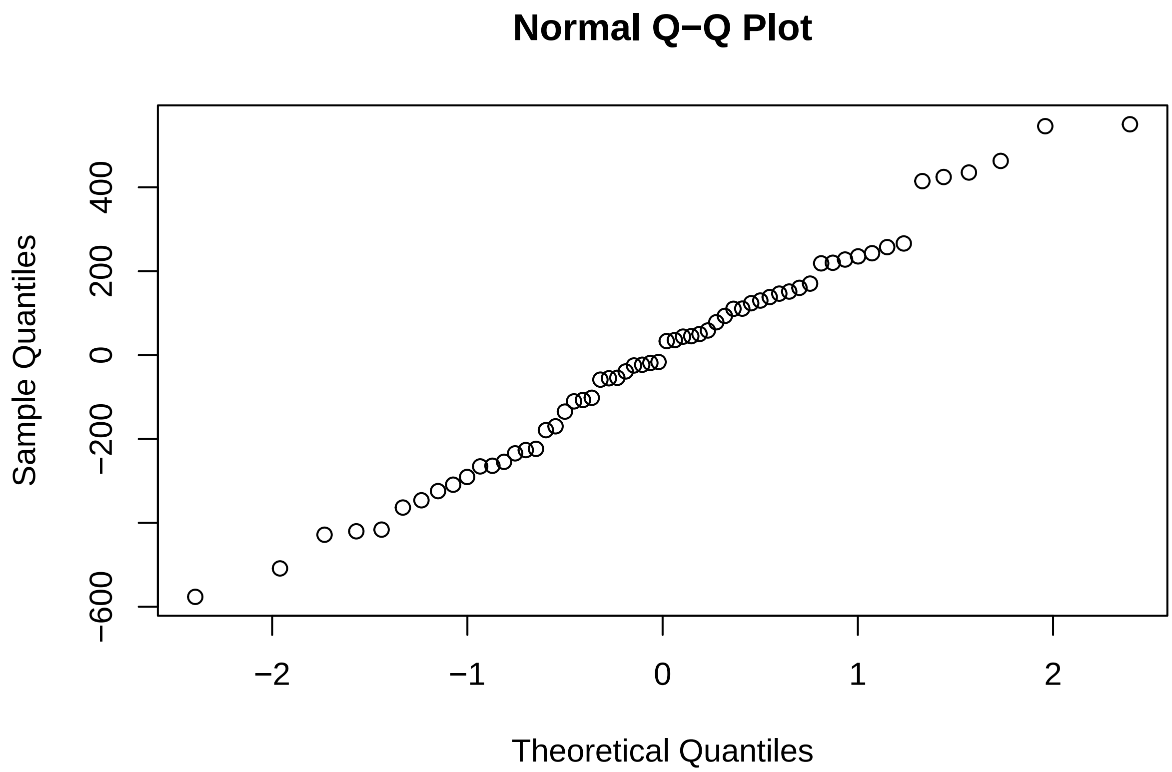
```
## Series: train
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##          ar1  intercept        V1         V2        V3        V4        V5
##       0.6903  9418.7040  -19.5581  -709.8543  418.5333  131.9044  -24.1281
## s.e.  0.0968   204.0229    5.7396    88.5117   53.8949   39.6131   32.8592
##             V6        V7        V8        V9        V10       V11       V12
##       148.0782  -10.1989  -722.510  90.8244  -208.7409  156.6989  214.2874
## s.e.   29.6828   20.1519    91.532  54.4536    39.5967   32.5689   29.2492
##
## sigma^2 = 89273:  log likelihood = -419.47
## AIC=868.94   AICc=879.85   BIC=900.36
##
## Training set error measures:
##                     ME     RMSE      MAE        MPE     MAPE      MASE
## Training set -5.993031 261.6156 212.1185 -0.1468995 2.414701 0.4403654
##                   ACF1
## Training set -0.167253
```



−1000
0
500
1000
400
0
−600
lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12

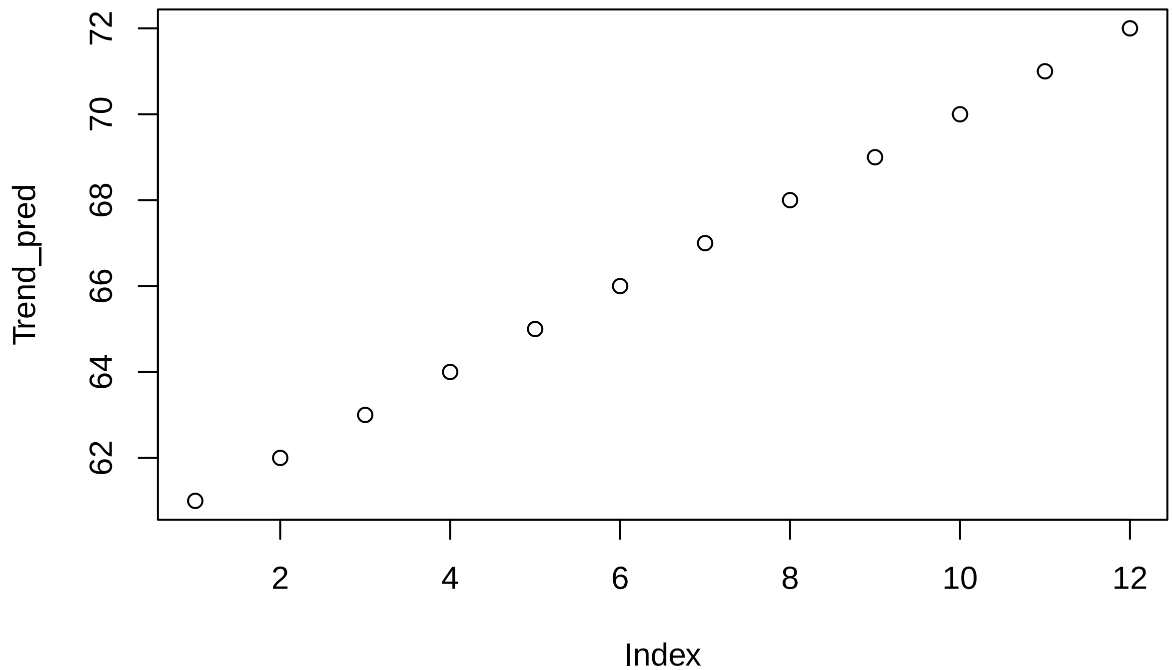

Series model_arima$residuals
ACF
−0.2
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Lag


Series model_arima$residuals
Partial ACF
0.2
0.1
0.0
−0.1
−0.2
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Lag



Test pour savoir si les résidus sont bien des bruits blancs.

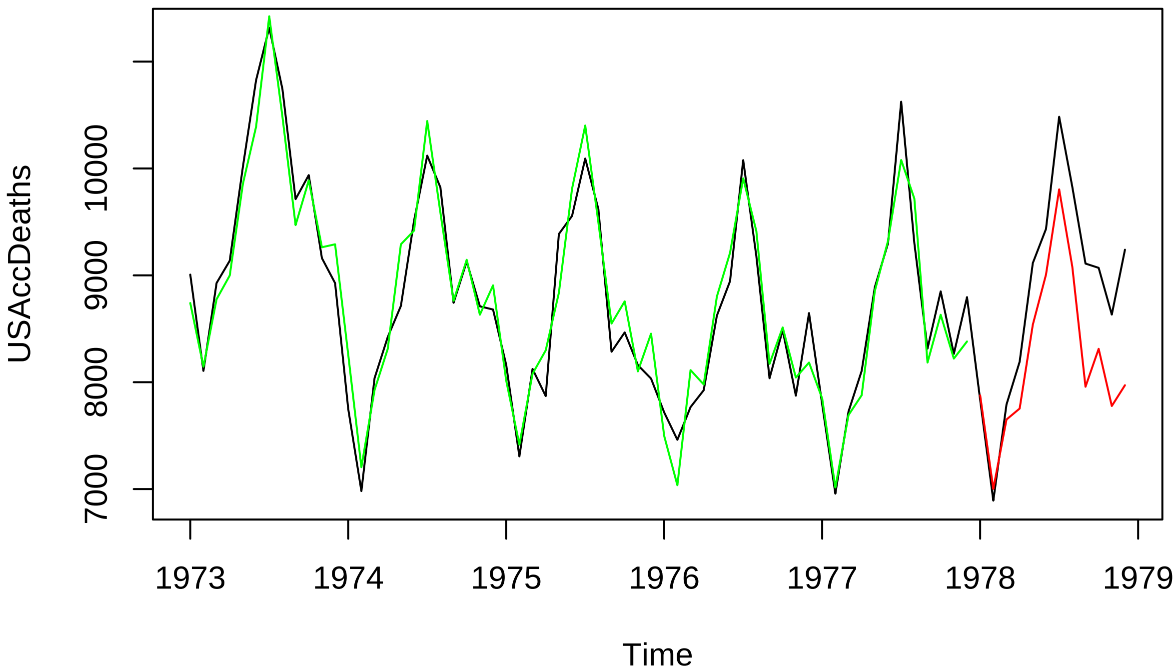
On va ensuite essayer de faire une prédiction avec le modèle où les résidus sont corrélés et ensuite avec les résidus non corrélés.




```
##                [,1] [,2]          [,3] [,4]          [,5] [,6]         [,7]
## [1,]  8.660254e-01  0.5  1.958693e-15 -0.5 -8.660254e-01   -1 5.000000e-01
## [2,]  5.000000e-01 -0.5 -1.000000e+00 -0.5  5.000000e-01    1 8.660254e-01
## [3,] -4.905202e-16 -1.0  1.471561e-15  1.0 -2.452601e-15   -1 1.000000e+00
## [4,] -5.000000e-01 -0.5  1.000000e+00 -0.5 -5.000000e-01    1 8.660254e-01
## [5,] -8.660254e-01  0.5  9.309041e-15 -0.5  8.660254e-01   -1 5.000000e-01
## [6,] -1.000000e+00  1.0 -1.000000e+00  1.0 -1.000000e+00    1 4.899781e-15
##                [,8]          [,9]         [,10]         [,11]
## [1,]  8.660254e-01  1.000000e+00  8.660254e-01  5.000000e-01
## [2,]  8.660254e-01  7.348993e-15 -8.660254e-01 -8.660254e-01
## [3,] -9.810403e-16 -1.000000e+00  1.962081e-15  1.000000e+00
## [4,] -8.660254e-01 -3.918740e-15  8.660254e-01 -8.660254e-01
## [5,] -8.660254e-01  1.000000e+00 -8.660254e-01  5.000000e-01
## [6,] -9.799561e-15  4.884873e-16 -1.959912e-14 -3.922806e-15
```

```
##   V1            V2   V3            V4   V5            V6 V7           V8
## 1 61  8.660254e-01  0.5  1.958693e-15 -0.5 -8.660254e-01 -1 5.000000e-01
## 2 62  5.000000e-01 -0.5 -1.000000e+00 -0.5  5.000000e-01  1 8.660254e-01
## 3 63 -4.905202e-16 -1.0  1.471561e-15  1.0 -2.452601e-15 -1 1.000000e+00
## 4 64 -5.000000e-01 -0.5  1.000000e+00 -0.5 -5.000000e-01  1 8.660254e-01
## 5 65 -8.660254e-01  0.5  9.309041e-15 -0.5  8.660254e-01 -1 5.000000e-01
## 6 66 -1.000000e+00  1.0 -1.000000e+00  1.0 -1.000000e+00  1 4.899781e-15
##             V9           V10           V11           V12
## 1  8.660254e-01  1.000000e+00  8.660254e-01  5.000000e-01
## 2  8.660254e-01  7.348993e-15 -8.660254e-01 -8.660254e-01
```

```
## 3 -9.810403e-16 -1.000000e+00  1.962081e-15  1.000000e+00
## 4 -8.660254e-01 -3.918740e-15  8.660254e-01 -8.660254e-01
## 5 -8.660254e-01  1.000000e+00 -8.660254e-01  5.000000e-01
## 6 -9.799561e-15  4.884873e-16 -1.959912e-14 -3.922806e-15
```



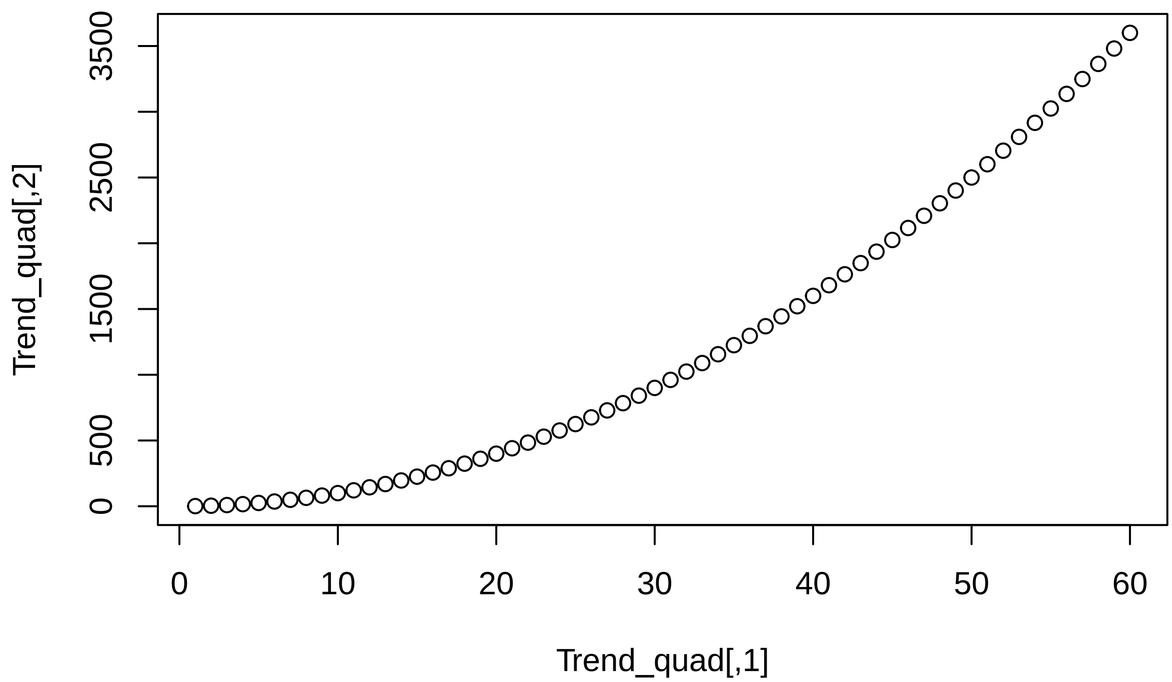
On va calculer le RMSE et le MAPE :

```
## [1] 707.4032
```

```
## [1] 0.06572168
```

Tendance Quadratique

Nous allons maintenant supposer une tendances quadratique :



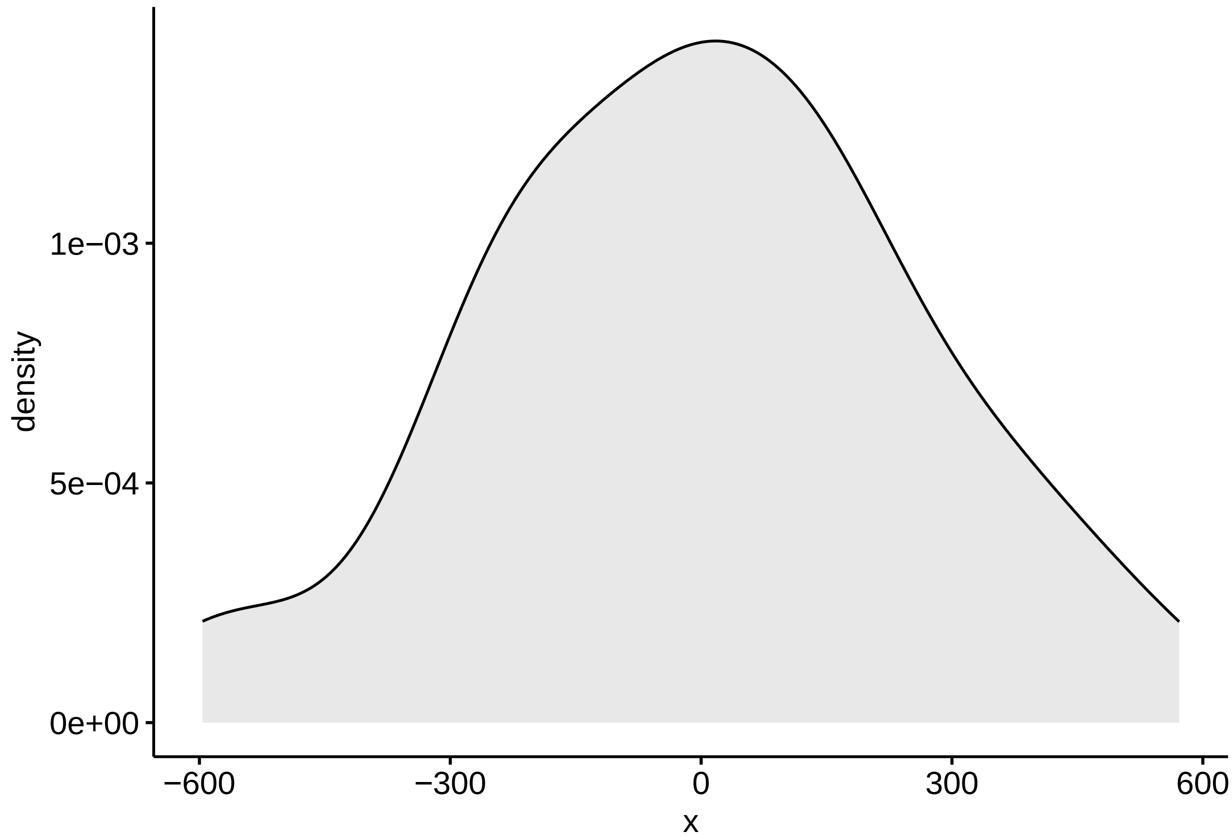
Notre matrice sera cette fois de la forme :

$$\begin{pmatrix} 1 & 1^2 & cos(\frac{2\pi}{12}) & cos(\frac{2\pi*2}{12}) & . & sin(\frac{2\pi*1}{12}) & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & T^2 & cos(\frac{2\pi T}{12}) & cos(\frac{2\pi*T*2}{12}) & . & sin(\frac{2\pi*T*1}{12}) & . \end{pmatrix}$$

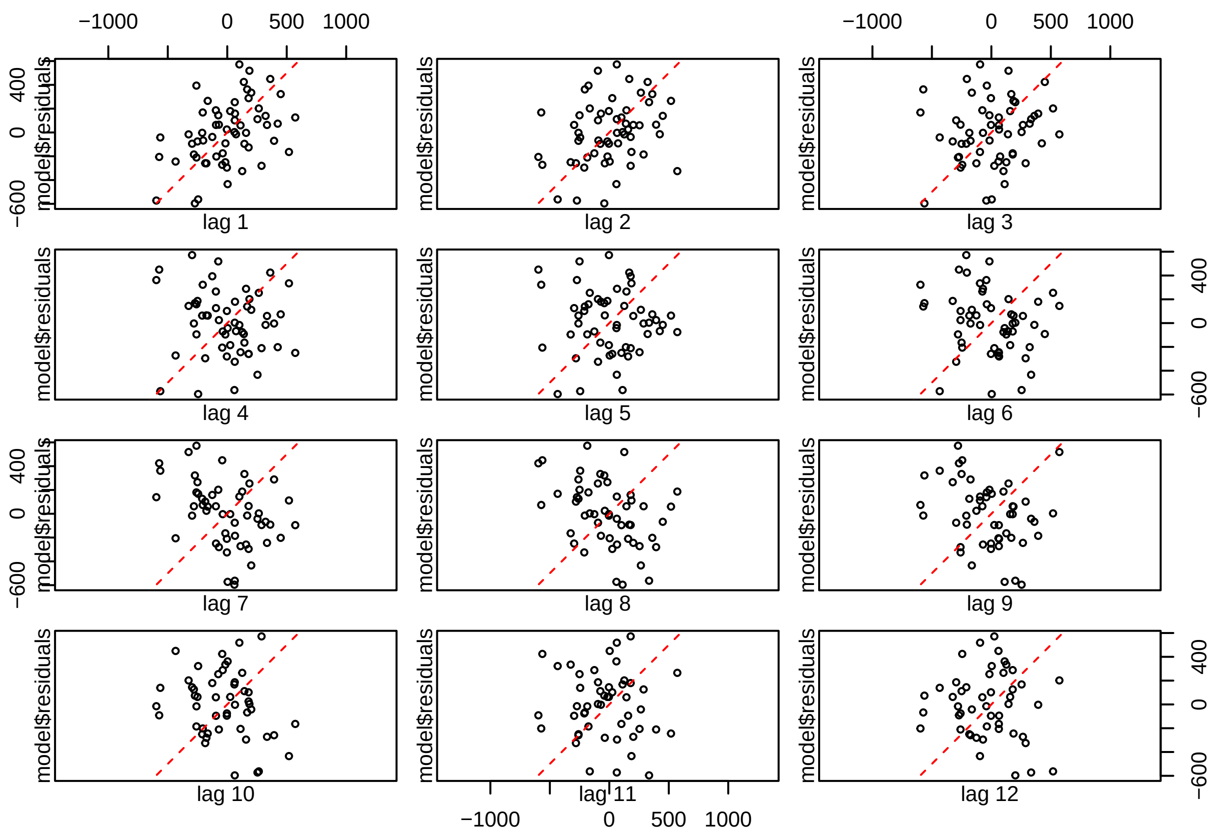
```
##   V1 V2            V3   V4            V5   V6            V7 V8           V9
## 1  1  1  8.660254e-01  0.5  6.123032e-17 -0.5 -8.660254e-01 -1 5.000000e-01
## 2  2  4  5.000000e-01 -0.5 -1.000000e+00 -0.5  5.000000e-01  1 8.660254e-01
## 3  3  9  6.123032e-17 -1.0 -1.836910e-16  1.0  3.061516e-16 -1 1.000000e+00
## 4  4 16 -5.000000e-01 -0.5  1.000000e+00 -0.5 -5.000000e-01  1 8.660254e-01
## 5  5 25 -8.660254e-01  0.5  3.061516e-16 -0.5  8.660254e-01 -1 5.000000e-01
## 6  6 36 -1.000000e+00  1.0 -1.000000e+00  1.0 -1.000000e+00  1 1.224606e-16
##             V10           V11           V12           V13
## 1  8.660254e-01  1.000000e+00  8.660254e-01  5.000000e-01
## 2  8.660254e-01  1.224606e-16 -8.660254e-01 -8.660254e-01
## 3  1.224606e-16 -1.000000e+00 -2.449213e-16  1.000000e+00
## 4 -8.660254e-01 -2.449213e-16  8.660254e-01 -8.660254e-01
## 5 -8.660254e-01  1.000000e+00 -8.660254e-01  5.000000e-01
## 6 -2.449213e-16  3.673819e-16 -4.898425e-16  6.123032e-16
```

```
##
## Call:
```

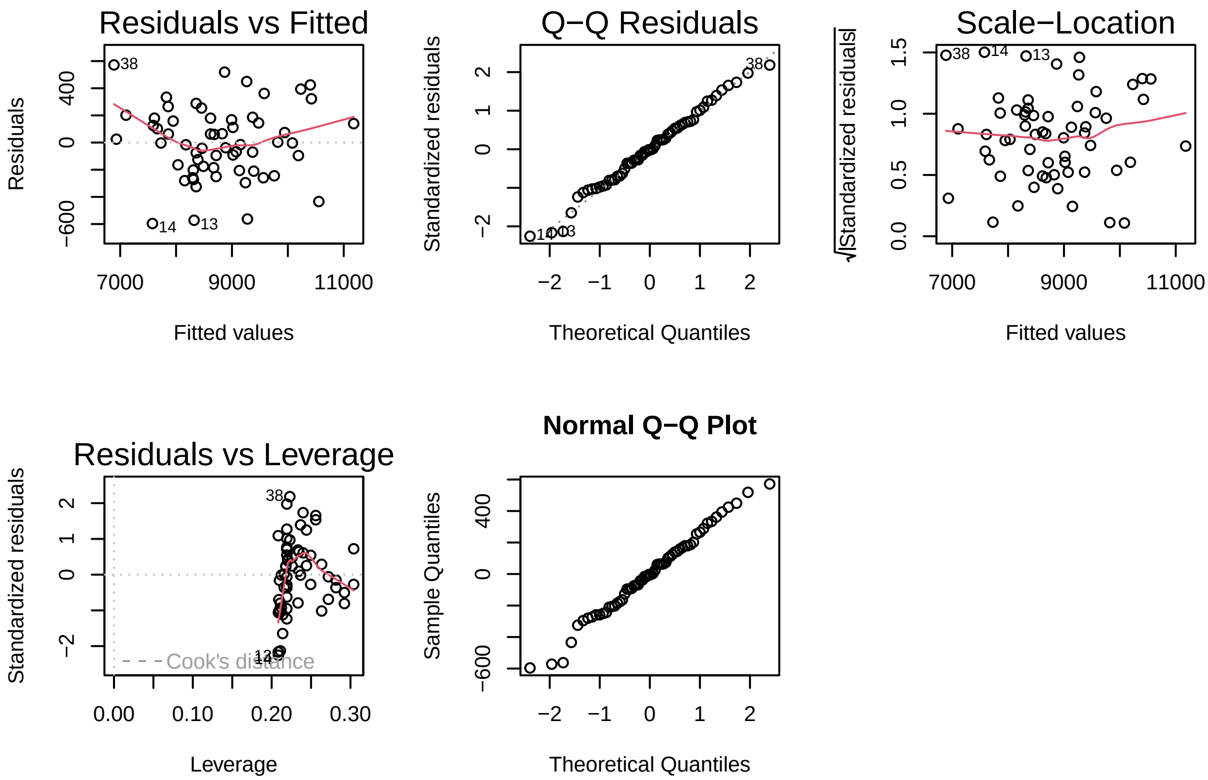
```
## lm(formula = train ~ ., data = Regresseur)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -596.29 -189.41   -3.24  170.78  571.80
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9979.2569   119.8942  83.234  < 2e-16 ***
## V1           -75.2384     9.0230  -8.338 9.42e-11 ***
## V2             0.8955     0.1432   6.254 1.20e-07 ***
## V3          -741.7877    54.3445 -13.650  < 2e-16 ***
## V4           413.6905    54.3095   7.617 1.09e-09 ***
## V5           132.7815    54.3080   2.445 0.018375 *
## V6           -20.9881    54.3079  -0.386 0.700934
## V7           152.0042    54.3078   2.799 0.007467 **
## V8            -8.1115    38.3848  -0.211 0.833571
## V9          -743.7105    54.9155 -13.543  < 2e-16 ***
## V10           78.4516    54.4024   1.442 0.156060
## V11         -216.0060    54.3080  -3.977 0.000245 ***
## V12          152.4748    54.2765   2.809 0.007267 **
## V13          212.2867    54.2642   3.912 0.000300 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 297.2 on 46 degrees of freedom
## Multiple R-squared:  0.9253, Adjusted R-squared:  0.9042
## F-statistic: 43.83 on 13 and 46 DF,  p-value: < 2.2e-16
```



Cette fois, on obtient une densité qui nous laisse penser que nos résidus sont distribués de manière gausienne et centrés en 0. Il faut maintenant que l'on vérifie que nos résidus sont des bruits blancs.

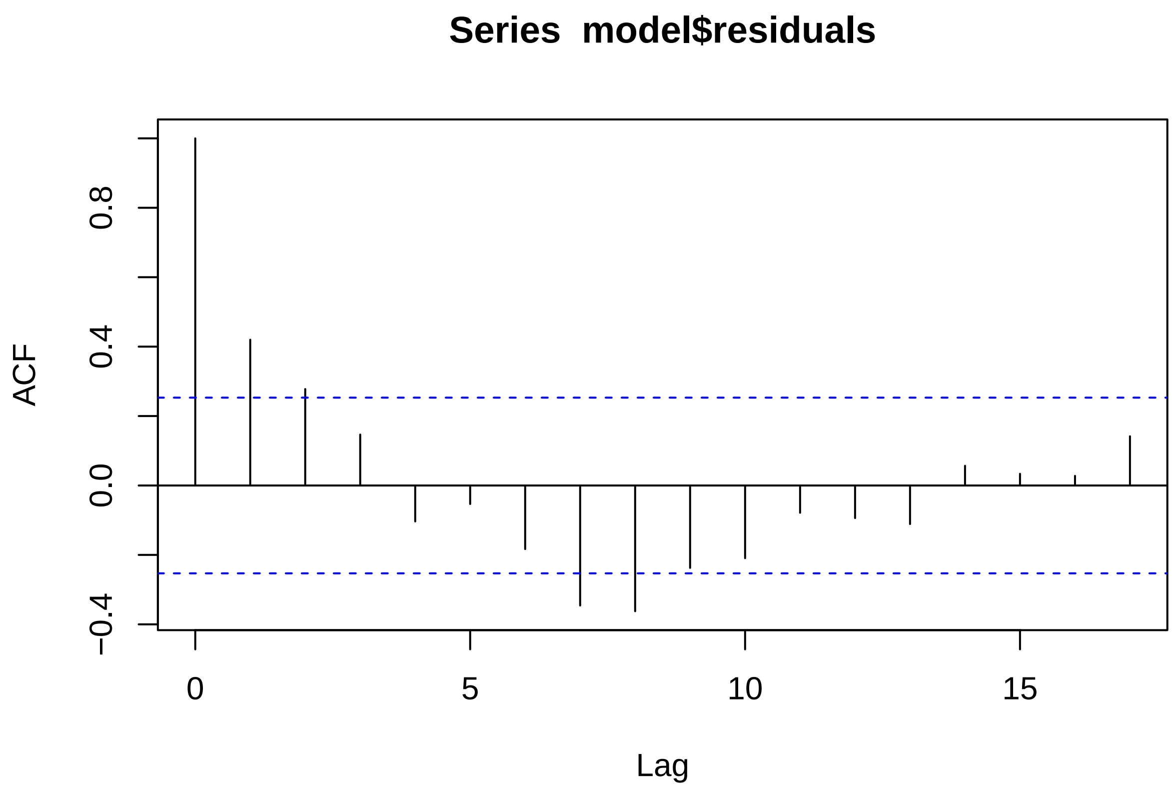


−1000
0
500
1000
400
0
−600
model$residuals
lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12

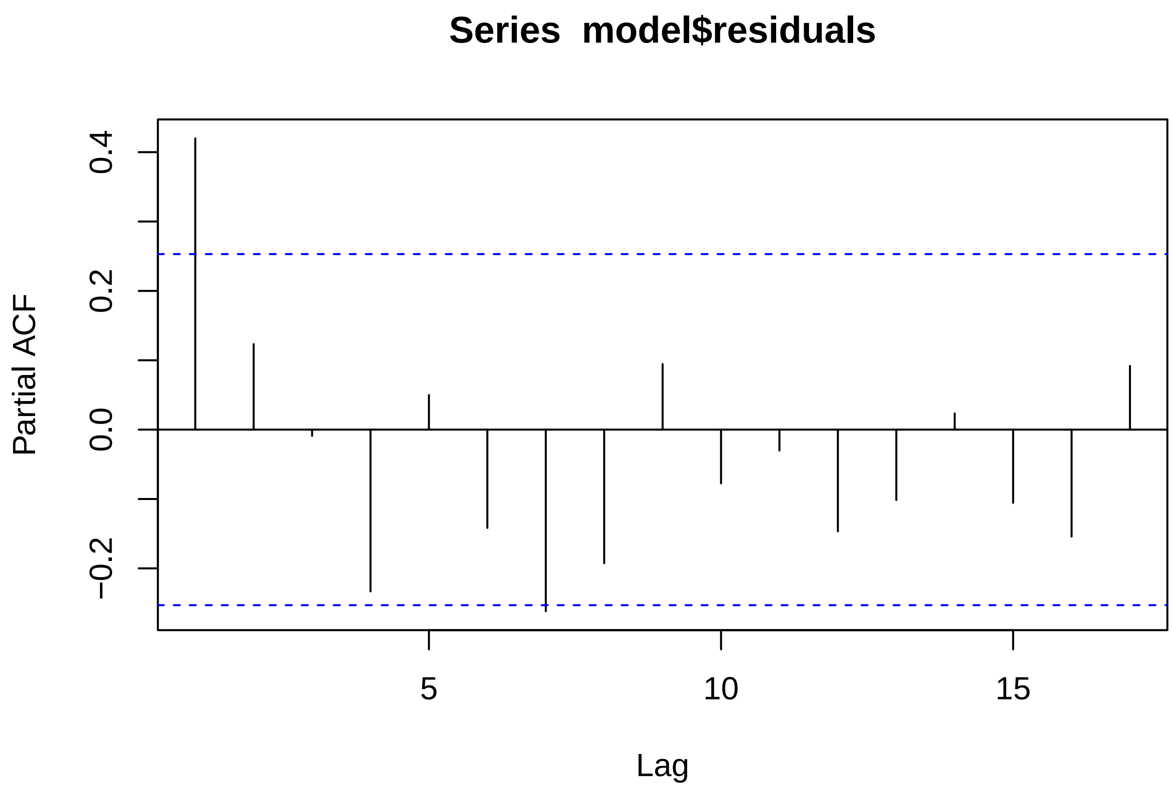


Encore une fois on observe que nous n'avons pas des bruit blanc, il semble y avoir une correlation dans au moins au lagplot, et le QQplot ne suit pas totalement une droite.

Traçon l'ACF et le PACF afin de voir comment décorréler nos résidus.

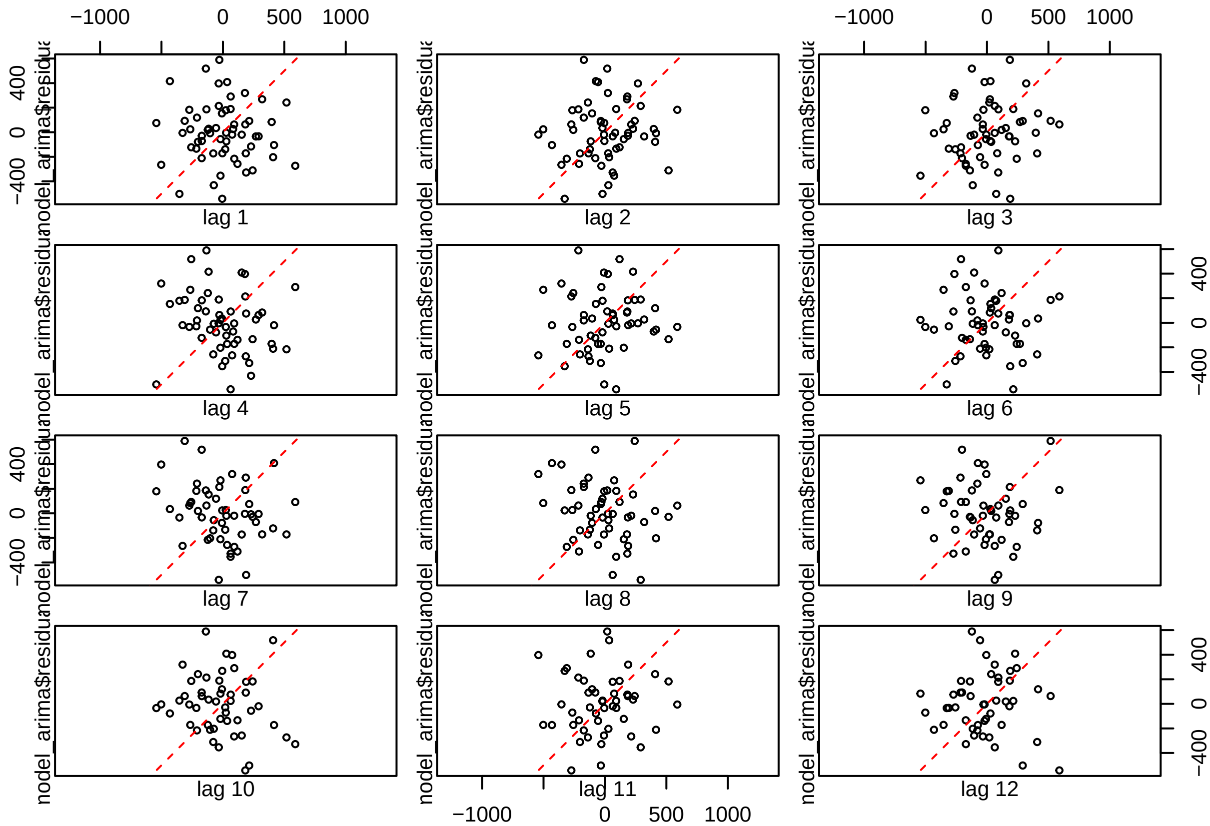


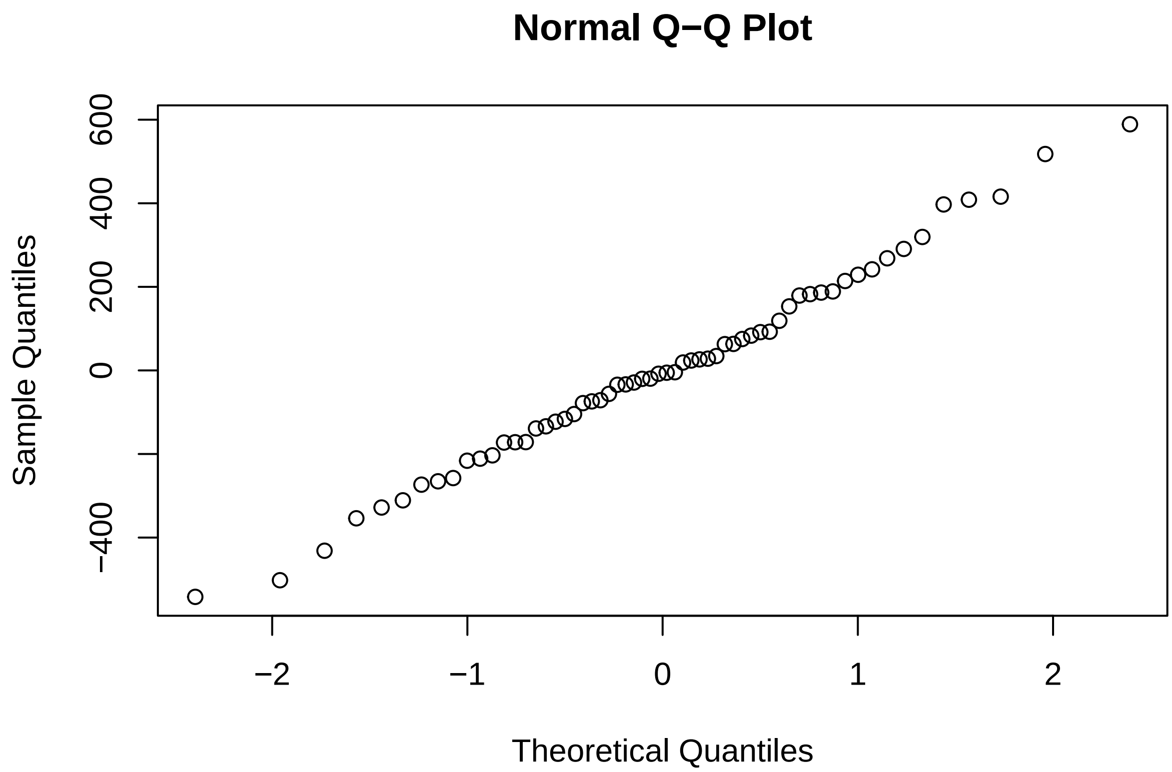
Series model$residuals
ACF
0.8
0.4
0.0
−0.4
0
5
10
15
Lag

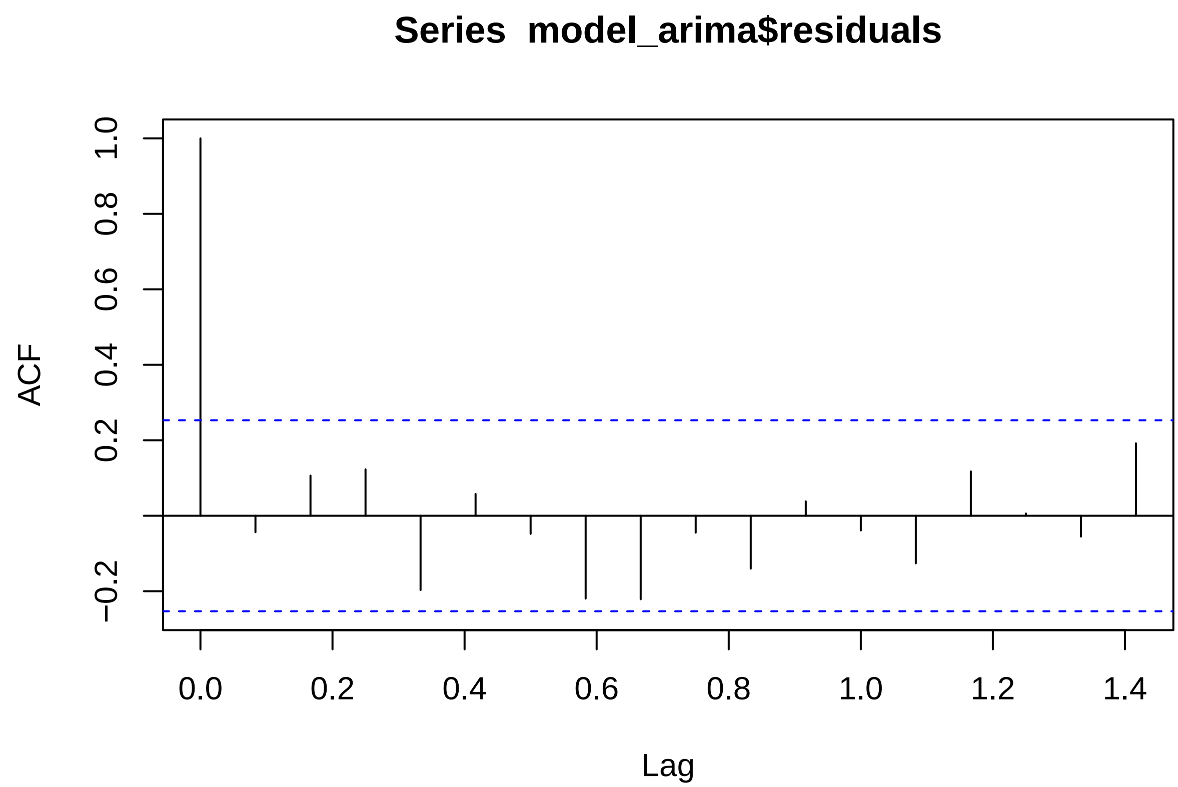


Encore une fois il semble que l'on doit utiliser un AR(1) :

```
## Series: train
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##          ar1  intercept        V1      V2         V3        V4        V5
##       0.4186  9975.3890  -75.4896  0.9058  -741.4618  412.3476  130.5787
## s.e.  0.1164   154.9489   11.6849  0.1853    63.5625   49.4032   39.8903
##             V6        V7       V8         V9      V10        V11       V12
##       -23.5840  149.2310  -9.5578  -739.1768  82.3025  -213.2374  154.2026
## s.e.   34.3994   31.5849  21.5931    65.0387  49.8234    39.8893   34.1822
##            V13
##       213.1464
## s.e.   31.2559
##
## sigma^2 = 74165:  log likelihood = -413.02
## AIC=858.05   AICc=870.7   BIC=891.56
##
## Training set error measures:
##                     ME     RMSE      MAE         MPE     MAPE      MASE
## Training set 0.6711556 235.8461 182.7107 -0.06099885 2.095723 0.3793137
##                     ACF1
## Training set -0.04368597
```


−1000
0
500
1000
400
0
−400
lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12


Normal Q−Q Plot
Sample Quantiles
−400
0
200
400
600
−2
−1
0
1
2
Theoretical Quantiles

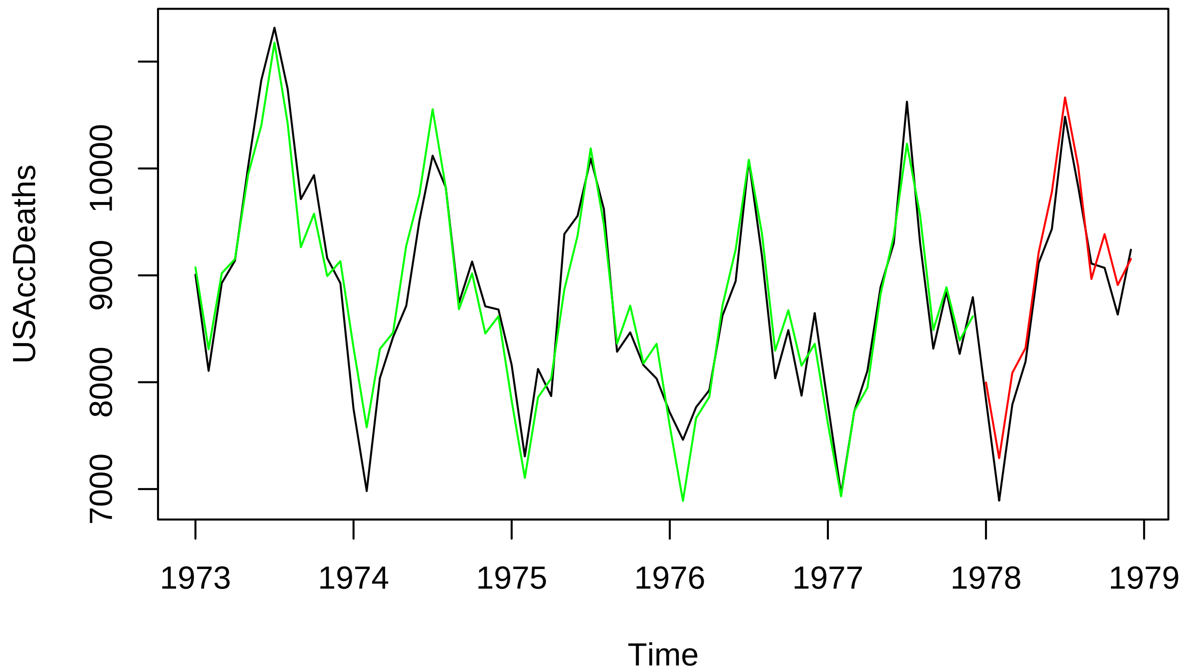


On pourrait aussi faire de nouveau un test comme précédement avec la bilbiothèque qui ne fonctionne pas : Test pour savoir si les résidus sont bien des bruits blancs.

On va maintenant faire une prédiction avec le modèle où les résidus sont non corrélés.

```
##    V1   V2            V3   V4            V5   V6            V7 V8            V9
## 1  61 3721  8.660254e-01  0.5  6.123032e-17 -0.5 -8.660254e-01 -1  5.000000e-01
## 2  62 3844  5.000000e-01 -0.5 -1.000000e+00 -0.5  5.000000e-01  1  8.660254e-01
## 3  63 3969  6.123032e-17 -1.0 -1.836910e-16  1.0  3.061516e-16 -1  1.000000e+00
## 4  64 4096 -5.000000e-01 -0.5  1.000000e+00 -0.5 -5.000000e-01  1  8.660254e-01
## 5  65 4225 -8.660254e-01  0.5  3.061516e-16 -0.5  8.660254e-01 -1  5.000000e-01
## 6  66 4356 -1.000000e+00  1.0 -1.000000e+00  1.0 -1.000000e+00  1  1.224606e-16
## 7  67 4489 -8.660254e-01  0.5 -4.286122e-16 -0.5  8.660254e-01 -1 -5.000000e-01
## 8  68 4624 -5.000000e-01 -0.5  1.000000e+00 -0.5 -5.000000e-01  1 -8.660254e-01
## 9  69 4761 -1.836910e-16 -1.0  5.510729e-16  1.0 -2.694812e-15 -1 -1.000000e+00
## 10 70 4900  5.000000e-01 -0.5 -1.000000e+00 -0.5  5.000000e-01  1 -8.660254e-01
## 11 71 5041  8.660254e-01  0.5 -2.449890e-15 -0.5 -8.660254e-01 -1 -5.000000e-01
## 12 72 5184  1.000000e+00  1.0  1.000000e+00  1.0  1.000000e+00  1 -2.449213e-16
##              V10           V11           V12           V13
## 1   8.660254e-01  1.000000e+00  8.660254e-01  5.000000e-01
## 2   8.660254e-01  1.224606e-16 -8.660254e-01 -8.660254e-01
## 3   1.224606e-16 -1.000000e+00 -2.449213e-16  1.000000e+00
## 4  -8.660254e-01 -2.449213e-16  8.660254e-01 -8.660254e-01
## 5  -8.660254e-01  1.000000e+00 -8.660254e-01  5.000000e-01
## 6  -2.449213e-16  3.673819e-16 -4.898425e-16  6.123032e-16
## 7   8.660254e-01 -1.000000e+00  8.660254e-01 -5.000000e-01
## 8   8.660254e-01 -4.898425e-16 -8.660254e-01  8.660254e-01
```

```
## 9   3.673819e-16  1.000000e+00 -7.347638e-16 -1.000000e+00
## 10 -8.660254e-01  6.123032e-16  8.660254e-01  8.660254e-01
## 11 -8.660254e-01 -1.000000e+00 -8.660254e-01 -5.000000e-01
## 12 -4.898425e-16 -7.347638e-16 -9.796851e-16 -1.224606e-15
```



On calcule le RMSE et le MAPE :

```
## [1] 240.0805
```

```
## [1] 0.02569034
```

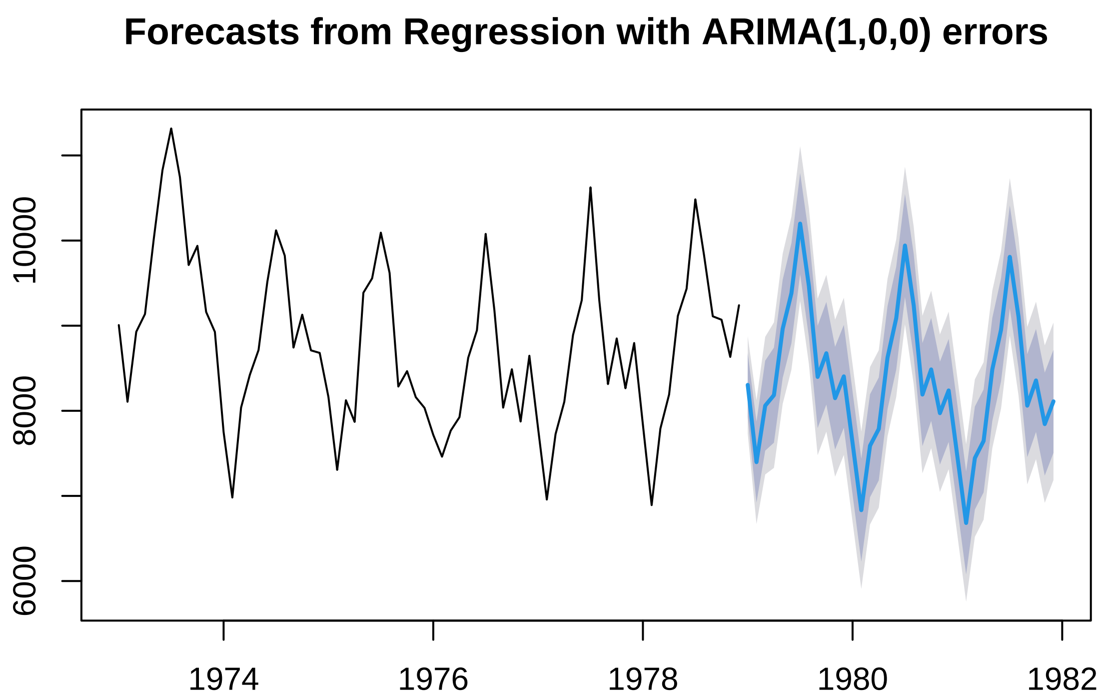
Prédiction sur 3 ans

Il constate donc que le modèle avec la tendance quadratique semble être meilleur que le modèle avec la tendance linéaire.

Nous allons maintenant effectuer des prévisions sur 3 ans avec le modèle quadratique :

```
##   V1            V2   V3            V4   V5            V6 V7            V8
## 1 73  8.660254e-01  0.5  9.798884e-15 -0.5 -8.660254e-01 -1  5.000000e-01
## 2 74  5.000000e-01 -0.5 -1.000000e+00 -0.5  5.000000e-01  1  8.660254e-01
## 3 75 -2.455989e-16 -1.0 -6.368631e-15  1.0  2.008829e-14 -1  1.000000e+00
## 4 76 -5.000000e-01 -0.5  1.000000e+00 -0.5 -5.000000e-01  1  8.660254e-01
## 5 77 -8.660254e-01  0.5  2.938378e-15 -0.5  8.660254e-01 -1  5.000000e-01
## 6 78 -1.000000e+00  1.0 -1.000000e+00  1.0 -1.000000e+00  1 -1.960725e-15
```

```
##                V9           V10           V11           V12
## 1  8.660254e-01  1.000000e+00  8.660254e-01  5.000000e-01
## 2  8.660254e-01  9.783298e-16 -8.660254e-01 -8.660254e-01
## 3 -4.911978e-16 -1.000000e+00  9.823956e-16  1.000000e+00
## 4 -8.660254e-01 -1.175893e-14  8.660254e-01 -8.660254e-01
## 5 -8.660254e-01  1.000000e+00 -8.660254e-01  5.000000e-01
## 6  3.921451e-15  8.328678e-15  7.842902e-15  1.861808e-14
```



Conclusion

Afin de déterminer quel modèle est le plus performant, nous allons comparer les critères d'erreurs de chacun de ces modèles.

$SARIMA(1,1,0)(0,1,1)_12$: - RMSE = 287.149069 - MAPE = 0.027322

Lissage exponentiel (HW additif avec ammortissement) : - RMSE = 292.3417 - MAPE = 0.02542135

Régression linéaire : - RMSE = 240.0805 - MAPE = 0.02569034

Le modèle par régression est clairement le plus performant, notamment grâce au modèle ayant une tendance quadratique. Les résultats du SARIMA et du lissage exponentiel sont très proche, il est donc dur de determiner quel est le meilleur modèle entre les deux.