

# Projet Séries Temporelles

Evolution mensuelle de de la fabrication d'instruments et de  
fournitures à usage médical et dentaire de janvier 1990 à janvier  
2020

Khairaldin Ahmed, RAZIG Amine



ENSAE Paris  
Palaiseau, France  
Année universitaire 2023-2024

May 22, 2024

## 1 Partie I: Les données

### 1.1 Que représente la série choisie ?

Dans le cadre de ce projet, nous allons étudier l'évolution mensuelle de la fabrication d'instruments et de fournitures à usage médical et dentaire de janvier 1990 à janvier 2020. Cette série représente l'indice de production industrielle en base 100 en 2021, c'est-à-dire qu'un indice supérieur à 100 indique une production supérieure à l'année de référence (2021). De plus, la série est corrigée aux variations saisonnières et des jours ouvrables (CVS-CJO).

La série brute initiale, qu'on note  $X_t$ , est tracée sur la Figure 1:

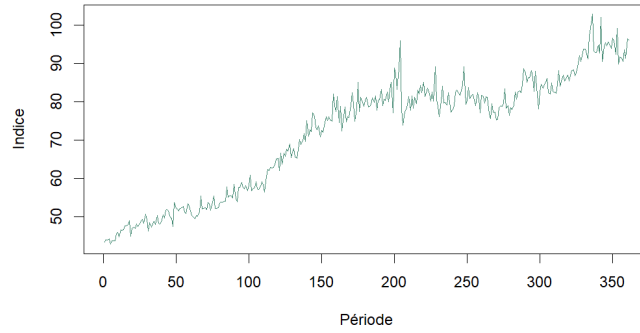


Figure 1: Série Brute Initiale

### 1.2 Transformation de la série

Graphiquement, la série semble avoir une tendance croissante. Cette tendance peut être due à une composante linéaire en  $t$  (déterministe) ou par un processus stochastique non linéaire. Pour corriger ce problème, nous différencierons plus tard notre série. Tout d'abord, vérifions que la série n'est pas stationnaire.

Pour ce faire, nous réalisons un test Augmented Dickey-Fuller dont l'hypothèse nulle est la présence d'une racine unité et donc la non-stationnarité de la série. L'hypothèse alternative est la stationnarité de la série. Ce test consiste en la régression de suivante (avec constante et tendance : cas général):

$$\Delta X_t = c + bt + \beta X_{t-1} + \sum_{\ell=1}^k \phi_{\ell} \Delta X_{t-\ell} + \varepsilon_t \quad (1)$$

où  $k$  représente le nombre de retards de la différence de la série temporelle inclus dans le modèle.

Ce paramètre  $k$  permet de déterminer le nombre de retards qu'il faudrait inclure dans la régression pour que le test soit valide, c'est-à-dire pour que les résidus ne soient pas autocorrélés. Nous vérifions alors pour

chaque  $k \geq 0$  l'autocorrélation des résidus <sup>1</sup>.

Pour supprimer l'autocorrélation des résidus, il a fallu considérer 5 retards au test ADF dont les résultats sont en annexe. On ne rejette pas l'hypothèse nulle à un seuil de 95% : la série n'est donc pas stationnaire.

Nous différencions alors la série :  $Y_t = X_t - X_{t-1}$ . Pour tester la stationnarité de cette série, nous effectuons les mêmes étapes pour notre série initiale. Il a également fallu 5 retards au test ADF pour supprimer l'autocorrélation des résidus. Le test ADF correspondant donne une p-value inférieure à 0.01.

Par ailleurs, nous effectuons un test de Philipps-Perron sur la série différenciée dont l'hypothèse nulle est identique au test ADF. Les hypothèses du test PP sont :

- $H_0$  : La série a une racine unitaire (non-stationnaire).
- $H_A$  : La série est stationnaire.

#### Résultats du test de Phillipps-Perron:

Statistique de test	-418.98
Paramètre Lag	5
Valeur p	0.01
Hypothèse alternative	Stationary

Table 1: Résultats du test de Phillipps-Perron pour la stationnarité

On rejette l'hypothèse nulle au seuil de 1% en faveur de l'hypothèse alternative de stationnarité. Nous considérons par la suite la série  $Y_t$  stationnaire (La série  $X_t$  est intégrée d'ordre 1).

### 1.3 Représentation de la série avant et après transformation

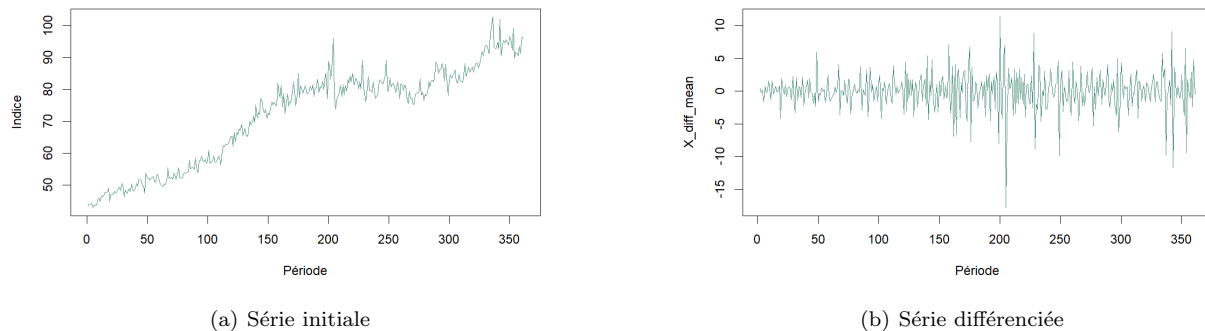


Figure 2: Série avant et après transformation

## 2 Partie II: Modèles ARMA

### 2.1 Modèle ARMA de la série différenciée

Afin de déterminer le modèle ARMA(p,q) approprié à la série différenciée  $Y_t$ , nous nous basons dans un premier temps sur les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF). Elles sont représentées dans la Figure 3:

<sup>1</sup>Comme la série est mensuelle, nous testons l'autocorrélation des résidus jusqu'à l'ordre 24 (2 ans) grâce à des tests de Ljung-Box.

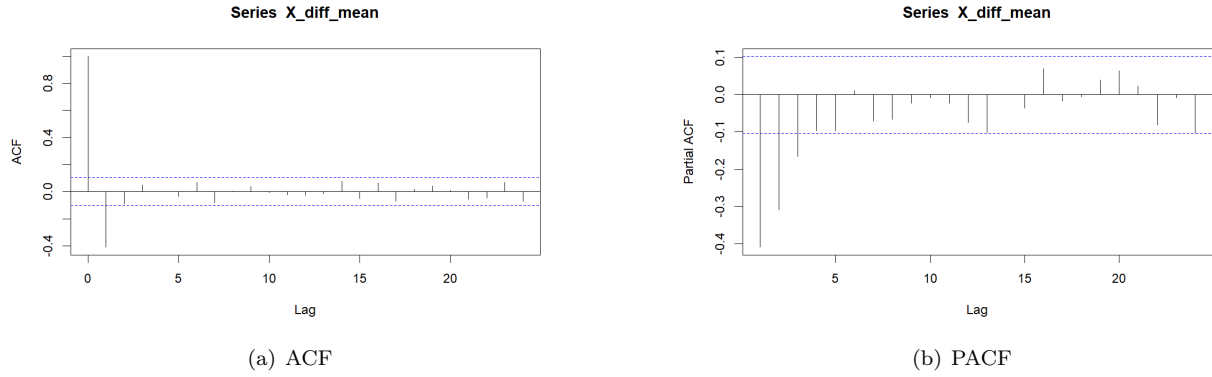


Figure 3: ACF et PACF de la série différenciée

Nous déduisons des deux figures les degrés maximaux :  $p_{max} = 3$  et  $q_{max} = 1$ . Les modèles possibles sont alors tous les ARMA(p,q) tels que  $p \leq 3$  et  $q \leq 1$ . Nous cherchons un modèle :

- bien ajusté, c'est-à-dire que les coefficients sont statistiquement significatifs.
- valide : les résidus ne sont pas autocorrélés.

Dans cette optique, nous effectuons pour chaque modèle possible un test de significativité des coefficients estimés et un test d'autocorrélations des résidus. Pour ce dernier, nous optons pour un test de Ljung-Box vérifiant l'absence jointe d'autocorrélations des résidus jusqu'à un ordre  $k$  donné. Comme dans la partie précédente, nous choisissons ici  $k = 24$ .

A titre d'exemple, les modèles AR(2) et ARMA(2,1) ne valident pas les deux tests (voir annexe). Le premier est bien ajusté mais n'est pas valide: les résidus sont autocorrélés. Tandis que le modèle ARMA(2,1) n'est pas valide, mais est bien ajusté: les coefficients correspondants aux plus hauts degrés sont significatifs.

Parmi tous les modèles possibles, les seuls modèles valides et bien ajustés sont les modèles AR(3) et MA(1) (voir annexe). Afin de déterminer lequel est le meilleur parmi les deux, nous calculons les AIC et BIC correspondants. Nous obtenons les valeurs suivantes des critères AIC et BIC:

	AR(3)	MA(1)
AIC	1708.614	1698.607
BIC	1728.044	1710.265

Table 2: Comparaison des modèles retenus

C'est le modèle MA(1) qui minimise les deux critères. On modélise alors notre série différenciée  $Y_t$  par un ARMA(0,1).

## 2.2 Modèle ARIMA pour la série initiale

Comme nous avons différencié la série initiale à l'ordre 1, nous la modélisons par un ARIMA(0,1,1).

## 3 Partie III: Prédiction

### 3.1 Région de confiance de niveau $\alpha$ sur les valeurs futures

Notre série différenciée (non centrée) suit un ARMA(1,1) qu'on note :

$$Y_t = \mu + \epsilon_t - \theta\epsilon_{t-1} \quad (2)$$

Ainsi, comme on a que  $Y_t = X_t - X_{t-1}$ , notre série initiale s'écrirait:

$$X_t = X_{t-1} + Y_t = \mu + X_{t-1} + \epsilon_t - \theta\epsilon_{t-1} \quad (3)$$

Les prédictions sur les valeurs futures ( $X_{T+1}$ ,  $X_{T+2}$ ) sont données par:

$$\begin{cases} \hat{X}_{T+1|T} = \mu + X_T - \theta\epsilon_T \\ \hat{X}_{T+2|T} = \mu + \hat{X}_{T+1|T} - \theta\epsilon_{T+1} \end{cases}$$

car  $\mathbb{E}[\epsilon_{T+h} | X_T, X_{T-1}, \dots] = 0$

On remplace  $\hat{X}_{T+1|T}$  dans la seconde équation, nous obtenons :

$$\begin{cases} \hat{X}_{T+1|T} = \mu + X_T - \theta\epsilon_T \\ \hat{X}_{T+2|T} = \mu + (\mu + X_T - \theta\epsilon_T) - \theta\epsilon_{T+1} = 2\mu + X_T - \theta\epsilon_T - \theta\epsilon_{T+1} \end{cases}$$

Les erreurs de prédiction sont données alors par:

$$\begin{pmatrix} e_{T+1} \\ e_{T+2} \end{pmatrix} = \mathbf{X} - \hat{\mathbf{X}} = \begin{pmatrix} \epsilon_{T+1} \\ \epsilon_{T+2} + (1 - \theta)\epsilon_{T+1} \end{pmatrix}$$

avec :

$$\mathbf{X} = \begin{pmatrix} X_{T+1} \\ X_{T+2} \end{pmatrix} \quad \text{et} \quad \hat{\mathbf{X}} = \begin{pmatrix} \hat{X}_{T+1|T} \\ \hat{X}_{T+2|T} \end{pmatrix}$$

En supposant que nos résidus suivent une loi gaussienne de variance  $\sigma_\epsilon^2$  alors, le vecteur d'erreurs de prévision est un vecteur gaussien de moyenne nulle et de matrice de variance-covariance :

$$\Sigma = \sigma_\epsilon^2 \begin{pmatrix} 1 & 1 - \theta \\ 1 - \theta & 1 + (1 - \theta)^2 \end{pmatrix}$$

Puisque  $\theta$  est dans le cercle unité (la série  $Y_t$  étant stationnaire), la matrice  $\Sigma$  est inversible si et seulement si  $\sigma_\epsilon > 0$ , ce que nous supposons vrai. Le rang de la matrice est donc égal à 2. D'où,  $\|\Sigma^{-1/2}(\mathbf{X} - \hat{\mathbf{X}})\| \sim \chi^2(2)$  avec  $\|\cdot\|$  la norme euclidienne. La région de confiance de niveau  $\alpha$  est alors donnée par:

$$R_\alpha = \{x \in \mathbb{R}^2 \mid \|\Sigma^{-1/2}(x - \hat{\mathbf{X}})\|^2 \leq \chi_{1-\alpha}^2(2)\}$$

où  $\chi_{1-\alpha}^2(2)$  est le quantile  $1 - \alpha$  de la loi  $\chi^2$  à 2 degrés de liberté.

### 3.2 Hypothèses utilisées

Pour obtenir les résultats précédents, nous avons tout d'abord supposé que le modèle est parfaitement connu et que les coefficients obtenus dans la partie précédente sont les véritables coefficients de notre modèle.

Par ailleurs, comme mentionné au début de cette partie, nous avons supposé la normalité des résidus et le fait que leur variance est strictement positive dans le calcul de la matrice de variance-covariance. La variance qu'on utilise est un estimateur de la variance qu'on suppose égal à la variance théorique des résidus (méthode plug-in).

### 3.3 Représentation graphique de la région de confiance

L'ellipse de confiance représente la région où nous nous attendons à trouver les valeurs futures  $X_{T+1}$  et  $X_{T+2}$  avec une probabilité de 95%. Les prédictions se trouvent au centre de l'ellipse. La taille et la forme de l'ellipse sont déterminées par la variance des résidus et les coefficients de notre modèle.

On remarque que l'ellipse est légèrement allongée et orientée, ce qui suggère qu'il y a une certaine corrélation entre les valeurs prévues à T+1 et T+2. Généralement, si l'ellipse était plus circulaire, cela indiquerait une moindre corrélation.

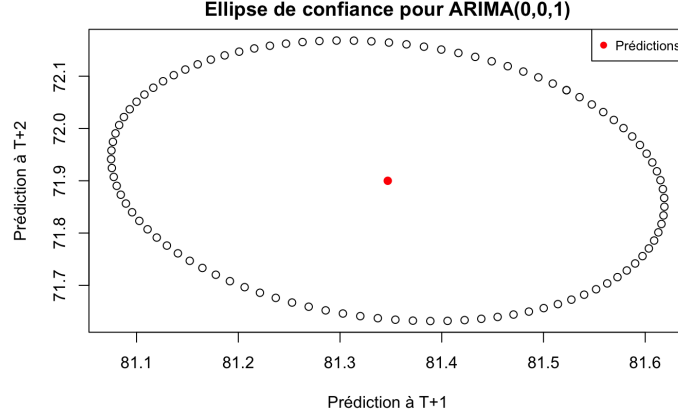


Figure 4: Ellipse de confiance

### 3.4 Question ouverte

Dans le cadre où  $Y_{T+1}$  est disponible plus rapidement que  $X_{T+1}$ , nous pouvons utiliser cette information pour améliorer la prédiction de  $X_{T+1}$  si  $Y$  cause  $X$  instantanément au sens de Granger. La causalité instantanée au sens de Granger signifie que, conditionnellement aux valeurs passées de  $Y_t$  et  $X_t$ , la valeur présente de  $Y_{T+1}$  améliore la prédiction de  $X_{T+1}$ . Plus formellement, on a :

$$\hat{Y}_{t+1|\{X_u, Y_u, u \leq t\}} \cup \{X_{t+1}\} \neq \hat{Y}_{t+1|\{X_u, Y_u, u \leq t\}} \quad (4)$$

Pour tester cette causalité, nous pouvons effectuer un test de causalité instantanée au sens de Granger pour lequel l'hypothèse nulle est : est la non-causalité instantanée de  $Y_t$  et  $X_t$  (dernière slide du cours). S'il y avait uniquement causalité au sens de Granger (non instantanée) :  $\hat{Y}_{t+1|\{X_u, Y_u, u \leq t\}} \neq \hat{Y}_{t+1|\{X_u, Y_u, u \leq t\}}$ , nous pourrions effectuer un test de Wald dont l'hypothèse nulle est la non causalité de  $Y_t$  et  $X_t$ .

## 4 Annexe

	Série Initiale	Série Différenciée
Lag Order	5	5
Dickey-Fuller	-2.0284	-10.0594
P-Valeur	0.5649	0.01

Table 3: Résultats des tests Augmented Dickey-Fuller sur les séries initiale et différenciée

### 4.1 Modèles valides et bien ajustés

Coefficient	P-Value
ma1	0.000
intercept	0.928

Table 4: P-Values pour la nullité des coefficients du modèle ARIMA(0,0,1)

Lag	P-Value	Lag	P-Value
1	NA	13	0.776
2	0.171	14	0.636
3	0.360	15	0.700
4	0.564	16	0.670
5	0.681	17	0.714
6	0.795	18	0.753
7	0.554	19	0.737
8	0.643	20	0.789
9	0.738	21	0.643
10	0.791	22	0.543
11	0.736	23	0.600
12	0.779	24	0.632

Table 5: P-Values pour les différents lags des tests d'absence d'autocorrélation des résidus

Coefficient	P-Value
ar1	0.000
ar2	0.000
ar3	0.001
intercept	0.965

Table 6: P-Values pour la nullité des coefficients du modèle ARIMA(3,0,0)

Lag	P-Value	Lag	P-Value
1	NA	13	0.313
2	NA	14	0.195
3	NA	15	0.254
4	0.015	16	0.233
5	0.046	17	0.275
6	0.090	18	0.312
7	0.070	19	0.320
8	0.122	20	0.382
9	0.184	21	0.269
10	0.240	22	0.174
11	0.246	23	0.212
12	0.238	24	0.247

Table 7: P-Values pour les différents lags des tests d'absence d'autocorrélation des résidus, ARIMA(3,0,0)

### 4.2 Exemples : modèles non retenus

Lag	P-Value	Lag	P-Value
1	NA	13	0.044
2	NA	14	0.021
3	0.000	15	0.032
4	0.001	16	0.037
5	0.002	17	0.040
6	0.005	18	0.048
7	0.005	19	0.047
8	0.009	20	0.065
9	0.016	21	0.047
10	0.028	22	0.028
11	0.027	23	0.037
12	0.029	24	0.049

Table 8: P-Values pour les différents lags des tests d'absence d'autocorrélation, AR(2)

Coefficient	P-Value
ar1	0.000
ar2	0.000
intercept	0.978

Table 9: P-Values pour la nullité des coefficients du modèle ARIMA(2,0,0)

Lag	P-Value	Lag	P-Value
1	NA	13	0.806
2	NA	14	0.649
3	NA	15	0.725
4	0.576	16	0.663
5	0.730	17	0.710
6	0.855	18	0.749
7	0.500	19	0.751
8	0.620	20	0.803
9	0.740	21	0.682
10	0.793	22	0.581
11	0.746	23	0.636
12	0.730	24	0.659

Table 10: P-Values pour les différents lags des tests d'absence d'autocorrélation (ARIMA(2,0,1))

Coefficient	P-Value
ar1	0.641
ar2	0.520
ma1	0.000
intercept	0.921

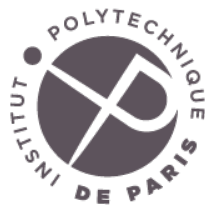
Table 11: P-Values pour la nullité des coefficients du modèle ARIMA(2,0,1)



# Annexe : Code du TP série temporelle

Ahmed Khairaldin et Amine Razig

17/5/2024



**IP PARIS**

# Projet : Fabrication d'instruments et de fournitures à usage médical et dentaire

## Bibliothèques nécessaires:

```
if (!require("httr")) install.packages("httr")
```

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

```
if (!require("ggplot2")) install.packages("ggplot2")
library(httr)
library(ggplot2)
library(zoo)
```

```
##
```

```
## Attachement du package : 'zoo'
```

```
## Les objets suivants sont masqués depuis 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
```

```
##      method      from
```

```
##      as.zoo.data.frame zoo
```

```
library(fUnitRoots)
```

```
library(forecast)
```

```
library(ellipse)
```

```
##
```

```
## Attachement du package : 'ellipse'
```

```
## L'objet suivant est masqué depuis 'package:graphics':
```

```
##
```

```
##      pairs
```

## Import des données :

```
#Importation des données : de janvier 1990 à janvier 2008
```

```
path <- "/Users/amineraizig/Desktop/ENSAE 2A/S2/Projet_Serie_temp/valeurs_mensuelles_dentaire.csv"
```

```
data <- read.csv(path, sep = ";")
```

## Manipulation des données :

```

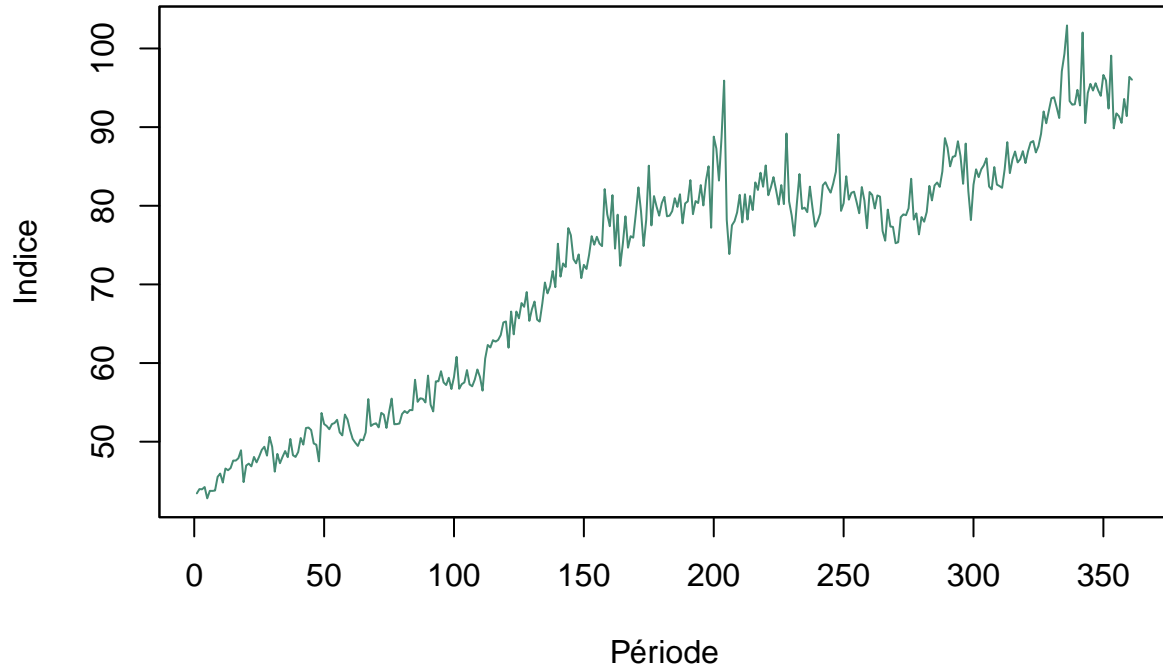
data <- data[-(1:3), ]
colnames(data) <- c("date", "valeur", "code")

# Conversion du format de la date et des valeurs et suppression de la période liée au Covid :
data$valeur <- as.numeric(data$valeur)

# Création de la série au format zoo :
X = data[,2]
X <- zoo(X)

# Vue générale de la série :
plot(X, type='l', col='aquamarine4', xlab="Période", ylab="Indice")

```



## TESTS POUR LA STATIONNARITE (serie initiale) : /! 0 lags ..

```
adf.test(X)
```

```

##
## Augmented Dickey-Fuller Test
##
## data: X
## Dickey-Fuller = -1.922, Lag order = 7, p-value = 0.6098
## alternative hypothesis: stationary

```

L'hypothèse nulle de racine unitaire  $H_0 : \text{Betha} = 0$  est testée par une statistique qui suit une loi de Dickey-Fuller dépendant du nombre d'observation et du cas du test dans lequel on se place. Le test ADF donne

:

```
adf <- adfTest(X, lag=0, type="ct")
```

```
## Warning in adfTest(X, lag = 0, type = "ct"): p-value smaller than printed
## p-value
```

*#Test ADF dans le cas avec constante et tendance. Avant d'interpréter le test, vérifions que les résidu*

```
residuals = adf@test$lm$residuals
```

*# Source : TD4 du cours de série temporelles*

*# tests d'autocorrelation :*

```
Qtests <- function(series, k, fitdf=0) {
  pvals <- apply(matrix(1:k), 1, FUN=function(l) {
    pval <- if (l<=fitdf) NA else Box.test(series, lag=l, type="Ljung-Box", fitdf=fitdf)$
    return(c("lag"=l,"pval"=pval))
  })
  return(t(pvals))
}
```

```
Qtests(adf@test$lm$residuals,24,length(adf@test$lm$coefficients))
```

```
##      lag      pval
## [1,]  1      NA
## [2,]  2      NA
## [3,]  3      NA
## [4,]  4 6.137983e-10
## [5,]  5 3.613019e-09
## [6,]  6 7.940280e-10
## [7,]  7 3.408080e-09
## [8,]  8 4.997377e-09
## [9,]  9 2.701928e-09
## [10,] 10 4.619396e-09
## [11,] 11 9.813409e-09
## [12,] 12 2.080880e-08
## [13,] 13 3.616855e-08
## [14,] 14 4.627600e-09
## [15,] 15 1.080463e-08
## [16,] 16 2.461988e-09
## [17,] 17 5.794519e-09
## [18,] 18 5.221330e-09
## [19,] 19 2.706503e-09
## [20,] 20 3.020007e-09
## [21,] 21 6.496081e-09
## [22,] 22 1.334194e-08
## [23,] 23 3.548236e-09
## [24,] 24 7.174351e-09
```

```

# Ajout d'un retard jusqu'a ce que les résidus soient non corrélés

adfTest_valid <- function(series,kmax,type){ #tests ADF jusqu'a des residus non autocorrelles
k <- 0
noautocorr <- 0
while (noautocorr==0){
cat(paste0("ADF with ",k, " lags: residuals OK? "))
adf <- adfTest(series,lags=k,type=type)
pvals <- Qtests(adf@test$lm$residuals,24,fitdf=length(adf@test$lm$coefficients))[,2]
if (sum(pvals<0.05,na.rm=T) == 0) {
noautocorr <- 1; cat("OK \n")}
else cat("nope \n")
k <- k + 1
}
return(adf)
}

adfTest_valid(X,24,'ct')

```

```
## ADF with 0 lags: residuals OK?
```

```
## Warning in adfTest(series, lags = k, type = type): p-value smaller than printed
## p-value
```

```
## nope
## ADF with 1 lags: residuals OK? nope
## ADF with 2 lags: residuals OK? nope
## ADF with 3 lags: residuals OK? nope
## ADF with 4 lags: residuals OK? nope
## ADF with 5 lags: residuals OK? OK
```

```
##
## Title:
##   Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 5
##   STATISTIC:
##     Dickey-Fuller: -2.0284
##   P VALUE:
##     0.5649
##
## Description:
##   Wed May 22 23:07:28 2024 by user:
```

Serie différenciée :

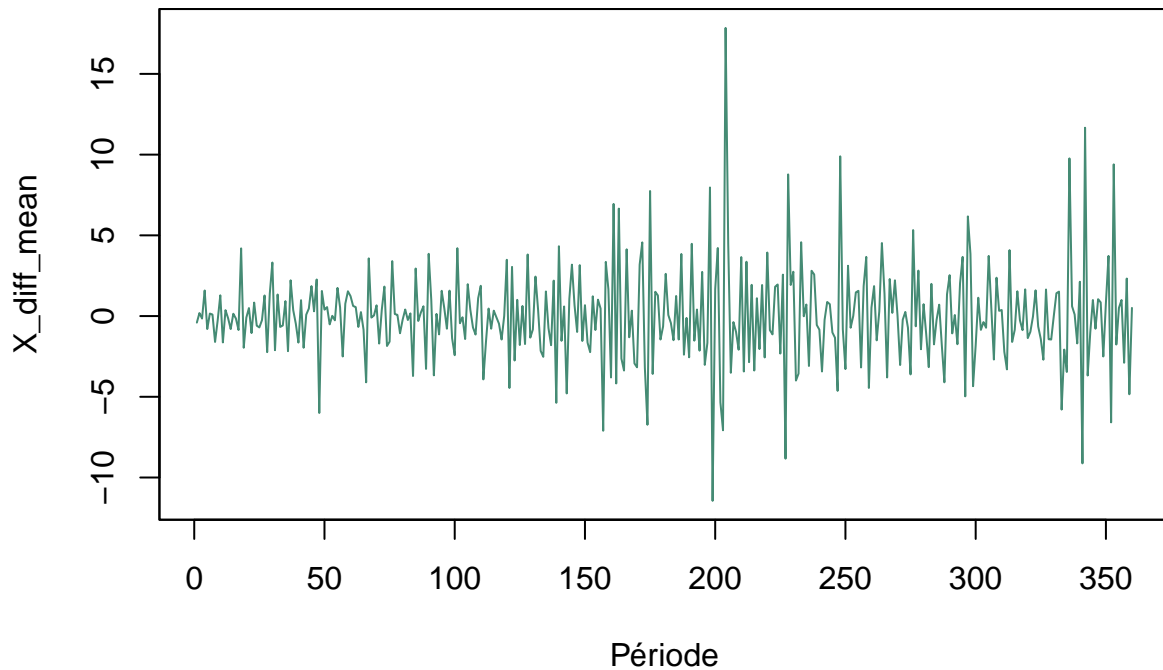
```

## **TENDANCE - Différenciation de la série **
#Différenciation et retranchement de la moyenne la série pour supprimer tendance

```

```
diff_X = X - lag(X,1)
X_diff_mean <- diff_X - mean(diff_X)

plot(X_diff_mean, col='aquamarine4', xlab="Période")
```



## TESTS POUR LA STATIONNARITE (serie differenciée) :

```
adfTest_valid(X_diff_mean,24,'ct')

## ADF with 0 lags: residuals OK?

## Warning in adfTest(series, lags = k, type = type): p-value smaller than printed
## p-value

## nope
## ADF with 1 lags: residuals OK?

## Warning in adfTest(series, lags = k, type = type): p-value smaller than printed
## p-value

## nope
## ADF with 2 lags: residuals OK?
```

```
## Warning in adfTest(series, lags = k, type = type): p-value smaller than printed
## p-value

## nope
## ADF with 3 lags: residuals OK?

## Warning in adfTest(series, lags = k, type = type): p-value smaller than printed
## p-value

## nope
## ADF with 4 lags: residuals OK?

## Warning in adfTest(series, lags = k, type = type): p-value smaller than printed
## p-value

## nope
## ADF with 5 lags: residuals OK?

## Warning in adfTest(series, lags = k, type = type): p-value smaller than printed
## p-value

## OK

##
## Title:
##   Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 5
##   STATISTIC:
##     Dickey-Fuller: -10.0594
##   P VALUE:
##     0.01
##
## Description:
##   Wed May 22 23:07:29 2024 by user:
```

#### Justification du test ADF :

```
# Le test ADF a donc du sens :
adf_diff <- adfTest(X_diff_mean, lag=0, type="ct")
```

```
## Warning in adfTest(X_diff_mean, lag = 0, type = "ct"): p-value smaller than
## printed p-value
```

```
Qtests(adf_diff@test$lm$residuals,24,length(adf_diff@test$lm$coefficients))
```

##		lag	pval
##	[1,]	1	NA
##	[2,]	2	NA
##	[3,]	3	NA
##	[4,]	4	3.751225e-10
##	[5,]	5	2.907159e-09
##	[6,]	6	1.154843e-08
##	[7,]	7	1.606751e-08
##	[8,]	8	5.499177e-08
##	[9,]	9	1.041589e-07
##	[10,]	10	2.984738e-07
##	[11,]	11	5.316090e-07
##	[12,]	12	7.810862e-07
##	[13,]	13	1.846804e-06
##	[14,]	14	1.614754e-06
##	[15,]	15	3.521896e-06
##	[16,]	16	6.293832e-06
##	[17,]	17	7.822019e-06
##	[18,]	18	1.496639e-05
##	[19,]	19	1.316991e-05
##	[20,]	20	2.421731e-05
##	[21,]	21	1.405274e-05
##	[22,]	22	1.671675e-05
##	[23,]	23	2.412165e-05
##	[24,]	24	3.391446e-05

## TESTS POUR LA STATIONNARITE :

```
pp.test(X_diff_mean)
```

```
## Warning in pp.test(X_diff_mean): p-value smaller than printed p-value

##
## Phillips-Perron Unit Root Test
##
## data: X_diff_mean
## Dickey-Fuller Z(alpha) = -418.98, Truncation lag parameter = 5, p-value
## = 0.01
## alternative hypothesis: stationary
```



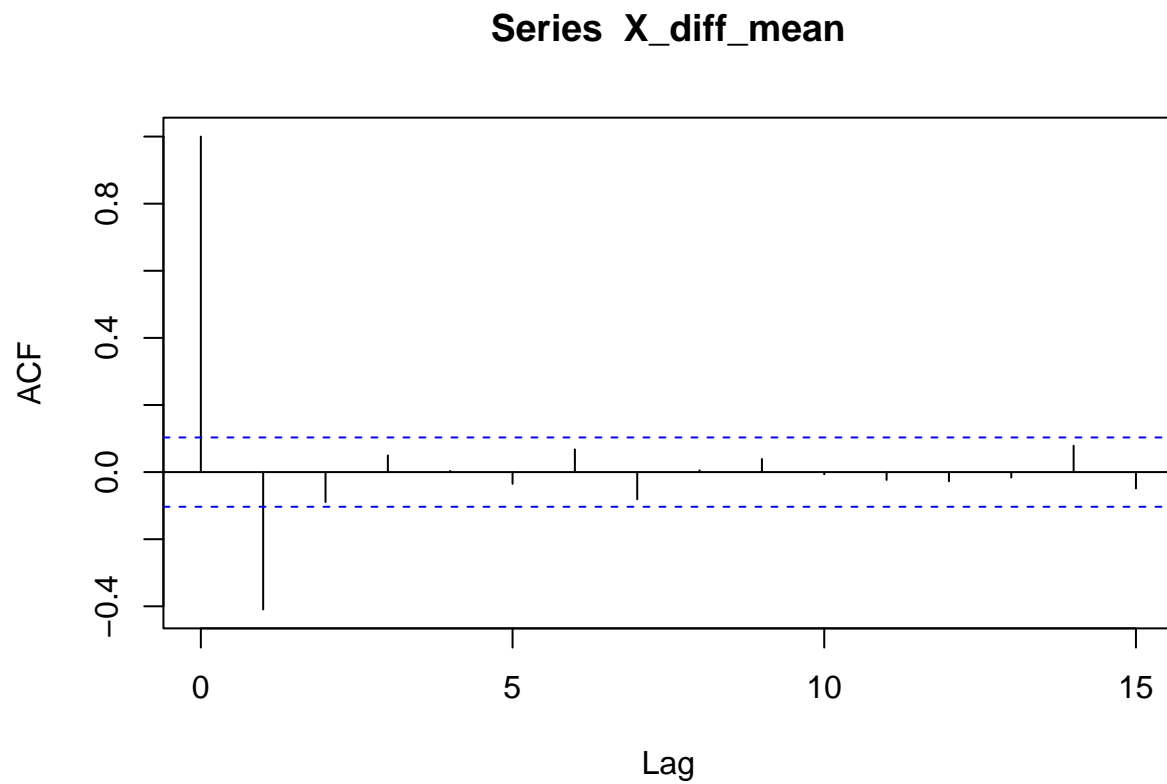
---

## Partie II : Modèles ARMA

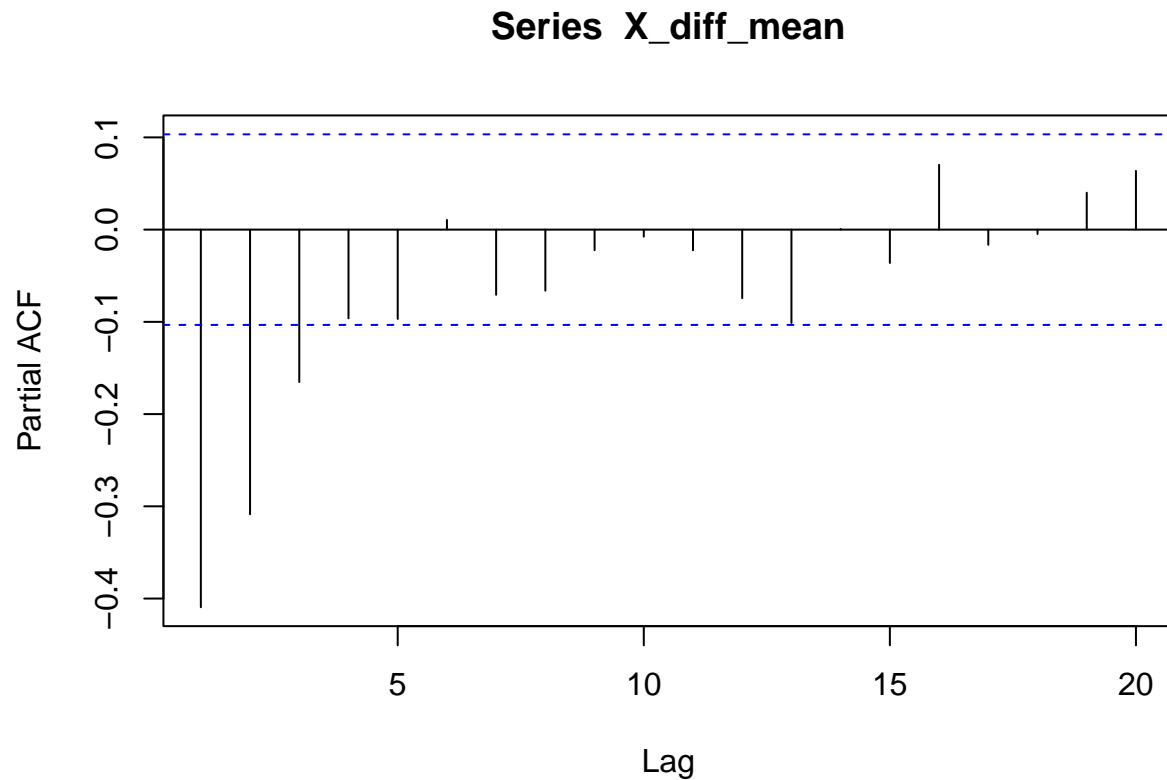
---

Sélection des ordres  $p$  et  $q$  de notre modèle :

```
acf(X_diff_mean, lag.max=15)
```



```
pacf(X_diff_mean, lag.max=20)
```



Pour nous -  $d^* = 0 - q^* = 1 - p^* = 3$

On test donc tous les modèles possible : (0,0,1), (1,0,1), (2,0,1),(3,0,1),(1,0,0), ...

```
#test des significativités individuelles des coefficients :
signif <- function(estim){
  coef <- estim$coef
  se <- sqrt(diag(estim$var.coef))
  t <- coef/se
  pval <- (1-pnorm(abs(t)))*2
  return(rbind(coef,se,pval))
}
arima201 <- arima(X_diff_mean,c(3,0,0))
signif(arima201) #tests de siginificativite de l'ARIMA(2,0,1)
```

```
##          ar1          ar2          ar3  intercept
## coef -0.58619649 -3.976686e-01 -0.165187553 0.002777216
## se    0.05197111  5.678920e-02  0.051967623 0.062910878
## pval  0.00000000  2.513545e-12  0.001479561 0.964788631
```

**Interpretation :** Les coefficients des retards les plus haut AR(2) ne rejette pas leur nullité à 95% ( $p\text{-value} > 0.05$ ), le modèle ARIMA(3,0,2) est donc mal ajusté. Pour MA(2) on rejette la nulité des coefficient.

```
##fonction d'affichage des tests pour la sélection du modèle ARIMA
arimafit <- function(estim){
  adjust <- round(signif(estim),3)
  pvals <- Qtests(estim$residuals,24,length(estim$coef)-1)
  pvals <- matrix(apply(matrix(1:24,nrow=6),2,function(c) round(pvals[c,],3)),nrow=6)
  colnames(pvals) <- rep(c("lag", "pval"),4)
  cat("tests de nullité des coefficients :\n")
  print(adjust)
  cat("\n tests d'absence d'autocorrélation des résidus : \n")
  print(pvals)
}
```

```
# EXEMPLE :
# Modele pas bien ajusté et non valide (absence de corrélation est rejeté)
estim <- arima(X_diff_mean,c(2,0,1)); arimafit(estim)
```

```
## tests de nullité des coefficients :
##      ar1    ar2    ma1 intercept
## coef 0.053 -0.054 -0.676    0.004
## se   0.114  0.084  0.104    0.043
## pval 0.641  0.520  0.000    0.921
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,]   1   NA   7 0.500 13 0.806 19 0.751
## [2,]   2   NA   8 0.620 14 0.649 20 0.803
## [3,]   3   NA   9 0.740 15 0.725 21 0.682
## [4,]   4 0.576 10 0.793 16 0.663 22 0.581
## [5,]   5 0.730 11 0.746 17 0.710 23 0.636
## [6,]   6 0.855 12 0.730 18 0.749 24 0.659
```

```
# POUR TOUTES LES COMBINAISONS :
for (p in 0:3) {
  for (q in 0:1) {
    cat(sprintf("\n\nARIMA(%d,0,%d):\n", p, q))
    # Estimation du modèle ARIMA
    estim <- arima(X_diff_mean,c(p,0,q)); arimafit(estim)
  }
}
```

```
##
##
## ARIMA(0,0,0):
## tests de nullité des coefficients :
##      intercept
## coef    0.000
## se      0.158
## pval    1.000
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,]   1   0   7   0 13   0 19   0
```

```

## [2,] 2 0 8 0 14 0 20 0
## [3,] 3 0 9 0 15 0 21 0
## [4,] 4 0 10 0 16 0 22 0
## [5,] 5 0 11 0 17 0 23 0
## [6,] 6 0 12 0 18 0 24 0
##
##
## ARIMA(0,0,1):
## tests de nullité des coefficients :
##      ma1 intercept
## coef -0.667      0.004
## se    0.043      0.045
## pval  0.000      0.928
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,] 1 NA 7 0.554 13 0.776 19 0.737
## [2,] 2 0.171 8 0.643 14 0.636 20 0.789
## [3,] 3 0.360 9 0.738 15 0.700 21 0.643
## [4,] 4 0.564 10 0.791 16 0.670 22 0.543
## [5,] 5 0.681 11 0.736 17 0.714 23 0.600
## [6,] 6 0.795 12 0.703 18 0.753 24 0.632
##
##
## ARIMA(1,0,0):
## tests de nullité des coefficients :
##      ar1 intercept
## coef -0.408      0.000
## se    0.048      0.102
## pval  0.000      0.999
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,] 1 NA 7 0 13 0 19 0
## [2,] 2 0 8 0 14 0 20 0
## [3,] 3 0 9 0 15 0 21 0
## [4,] 4 0 10 0 16 0 22 0
## [5,] 5 0 11 0 17 0 23 0
## [6,] 6 0 12 0 18 0 24 0
##
##
## ARIMA(1,0,1):
## tests de nullité des coefficients :
##      ar1 ma1 intercept
## coef 0.100 -0.725      0.004
## se    0.081 0.058      0.041
## pval 0.215 0.000      0.917
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,] 1 NA 7 0.563 13 0.839 19 0.801
## [2,] 2 NA 8 0.671 14 0.709 20 0.847
## [3,] 3 0.283 9 0.769 15 0.776 21 0.739
## [4,] 4 0.538 10 0.823 16 0.736 22 0.661

```

```

## [5,]  5 0.715  11 0.789  17 0.771  23 0.709
## [6,]  6 0.802  12 0.774  18 0.808  24 0.728
##
##
## ARIMA(2,0,0):
## tests de nullité des coefficients :
##      ar1    ar2 intercept
## coef -0.535 -0.309      0.002
## se    0.050  0.050      0.074
## pval  0.000  0.000      0.978
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,]  1    NA  7 0.005 13 0.044 19 0.047
## [2,]  2    NA  8 0.009 14 0.021 20 0.065
## [3,]  3 0.000  9 0.016 15 0.032 21 0.047
## [4,]  4 0.001 10 0.028 16 0.037 22 0.028
## [5,]  5 0.002 11 0.027 17 0.040 23 0.037
## [6,]  6 0.005 12 0.029 18 0.048 24 0.049
##
##
## ARIMA(2,0,1):
## tests de nullité des coefficients :
##      ar1    ar2    ma1 intercept
## coef 0.053 -0.054 -0.676      0.004
## se   0.114  0.084  0.104      0.043
## pval 0.641  0.520  0.000      0.921
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,]  1    NA  7 0.500 13 0.806 19 0.751
## [2,]  2    NA  8 0.620 14 0.649 20 0.803
## [3,]  3    NA  9 0.740 15 0.725 21 0.682
## [4,]  4 0.576 10 0.793 16 0.663 22 0.581
## [5,]  5 0.730 11 0.746 17 0.710 23 0.636
## [6,]  6 0.855 12 0.730 18 0.749 24 0.659
##
##
## ARIMA(3,0,0):
## tests de nullité des coefficients :
##      ar1    ar2    ar3 intercept
## coef -0.586 -0.398 -0.165      0.003
## se    0.052  0.057  0.052      0.063
## pval  0.000  0.000  0.001      0.965
##
## tests d'absence d'autocorrélation des résidus :
##      lag pval lag pval lag pval lag pval
## [1,]  1    NA  7 0.070 13 0.313 19 0.320
## [2,]  2    NA  8 0.122 14 0.195 20 0.382
## [3,]  3    NA  9 0.184 15 0.254 21 0.269
## [4,]  4 0.015 10 0.240 16 0.233 22 0.174
## [5,]  5 0.046 11 0.246 17 0.275 23 0.212
## [6,]  6 0.090 12 0.238 18 0.312 24 0.247
##
##

```

```
##
## ARIMA(3,0,1):
## tests de nullité des coefficients :
##      ar1  ar2  ar3  ma1 intercept
## coef 0.169 0.014 0.087 -0.790 0.004
## se   0.119 0.086 0.078 0.104 0.039
## pval 0.155 0.871 0.266 0.000 0.914
##
## tests d'absence d'autocorrélation des résidus :
##      lag  pval lag  pval lag  pval lag  pval
## [1,]  1    NA  7 0.449 13 0.846 19 0.756
## [2,]  2    NA  8 0.616 14 0.656 20 0.811
## [3,]  3    NA  9 0.735 15 0.737 21 0.712
## [4,]  4    NA 10 0.824 16 0.699 22 0.607
## [5,]  5 0.523 11 0.784 17 0.727 23 0.660
## [6,]  6 0.636 12 0.774 18 0.762 24 0.696
```

Nos trois modèles ajustés et valides sont : ARIMA(3,0,0) ; ARIMA(0,0,1)

on peut utiliser un critère d'information, tel que l'AIC ou le BIC (qui dépendent négativement de la log-vraisemblance du modèle pénalisée par le nombre de paramètres à estimer) :

```
ARIMA300= arima(X_diff_mean,c(3,0,0))
ARIMA001= arima(X_diff_mean,c(0,0,1))

models <- c("ARIMA300","ARIMA001"); names(models) <- models
apply(as.matrix(models),1, function(m) c("AIC"=AIC(get(m)), "BIC"=BIC(get(m))))

##      ARIMA300 ARIMA001
## AIC 1708.614 1698.607
## BIC 1728.044 1710.265
```

**Conclusion :** Le ARIMA(0,0,1) minimise à la fois le critère BIC et AIC donc on peut le considérer comme ayant les meilleurs ordres et c'est celui là qu'on retient.

---

## Partie III : Prédiction

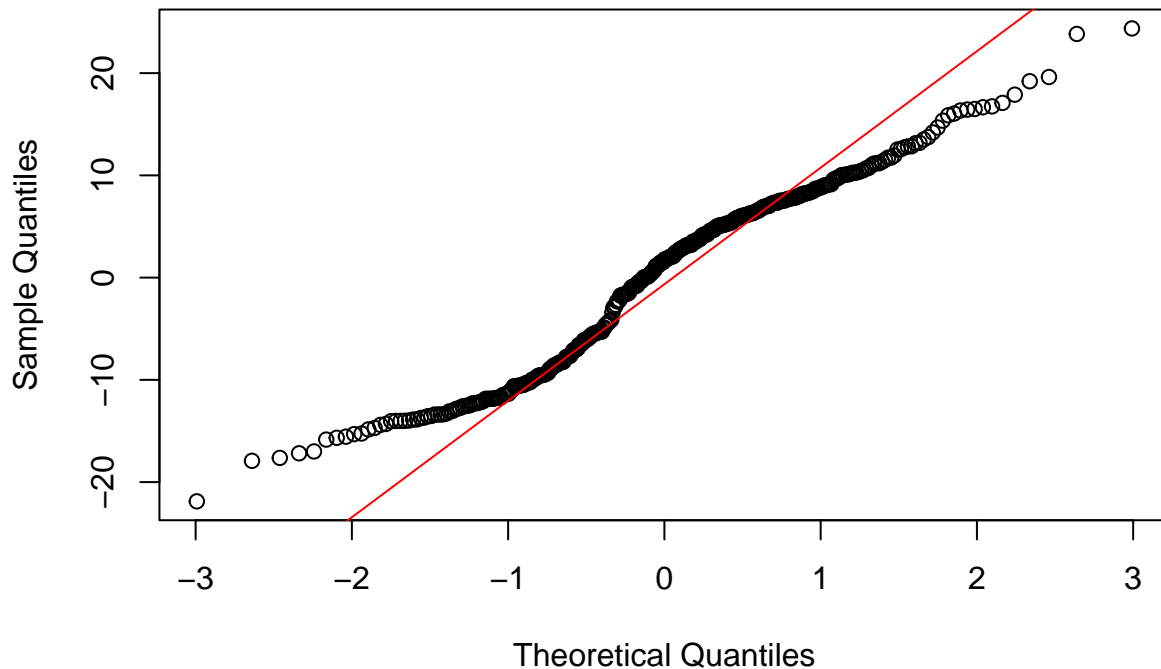
---

```
# On commence par ajuster le modèle sélectionné :
model <- arima(X, c(0, 0, 1))

# On extrait les résidus du modèle
```

```
residuals <- residuals(model)
# QQ plot des résidus (pour savoir si ils suivent une loi normale)
qqnorm(residuals, main = "QQ Plot of ARIMA(0,0,1) Residuals")
qqline(residuals, col = "red")
```

### QQ Plot of ARIMA(0,0,1) Residuals



## 8. Représenter graphiquement de la région pour  $\alpha = 95\%$  :

```
# On determine le coef MA(1) :
model_ma1 <- arima(X, order = c(0, 0, 1))
ma1_coef <- coef(model_ma1)["ma1"]

theta <- ma1_coef

# Variance des résidus (bruit blanc)
residuals_ma1 <- residuals(model_ma1)
variance_residuals <- var(residuals_ma1)

## On obtiens :
## theta = 0.8323871
## Variance du BB = 83.27074

# Calculer les éléments de la matrice Variance Covariance (Sigma)
element11 <- variance_residuals
element12 <- variance_residuals * (1 - theta)
element21 <- variance_residuals * (1 - theta)
element22 <- variance_residuals * (1 + (1 - theta)^2)
```

```

# Créer la matrice Sigma et inversion
Sigma <- matrix(c(element11, element12, element21, element22), nrow=2, ncol=2, byrow=TRUE)
Sigma_inverse <- solve(Sigma)

# Prédiction des valeurs à T+1 et T+2 avec le modèle ARIMA(0,0,1)
valeurs_pred <- predict(model_ma1, n.ahead=2)
prediction_T1 <- valeurs_pred$pred[1]
prediction_T2 <- valeurs_pred$pred[2]

# On trace l'ellipse de confiance
plot(ellipse(Sigma_inverse, centre=c(prediction_T1, prediction_T2), level=0.95),
     xlab="Prédiction à T+1", ylab="Prédiction à T+2",
     main="Ellipse de confiance pour ARIMA(0,0,1)")

points(x=prediction_T1, y=prediction_T2, col="red", pch=19)
legend("topright", legend="Prédictions", pch=19, col="red", cex=0.8)

```

