Projet de Séries Temporelles

Modélisation ARIMA d'une série temporelle

Auteurs :

Pierre Rouillard

Thomas Kientz

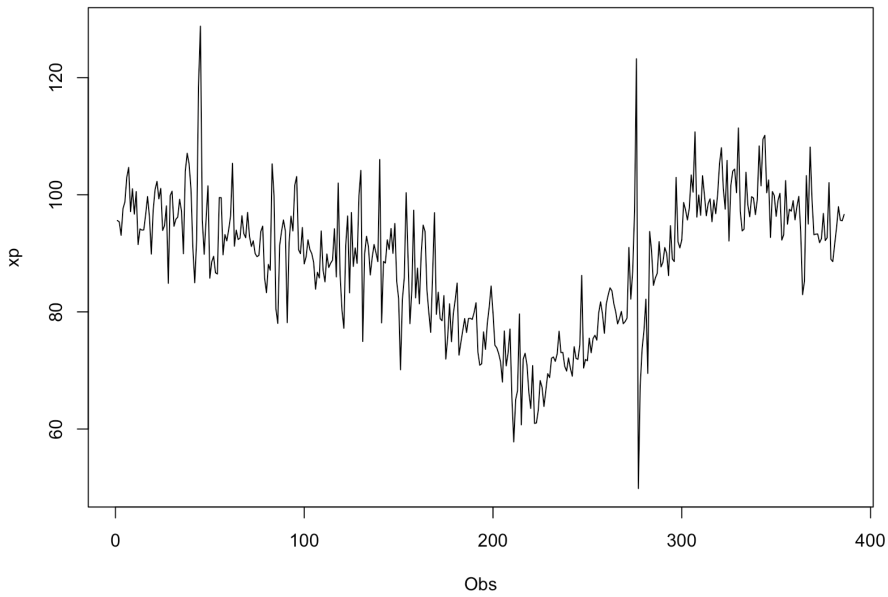
3 mai 2022

1 Les données

Q.1

La série choisie représente l'indice CVS-CJO de la fabrication de bière entre janvier 1990 et février 2022 en France métropolitaine [données mensuelles, 386 observations] et a pour année de référence 2015 (c'est-à-dire moyenne de 100 pour 2015). Ce type d'indice permet de suivre l'évolution industrielle de la France et fournit une information pour le suivi conjoncturel de l'économie française.
La série initiale est tracée sur la figure1 ci-après.

Figure 1 – Série brute initiale



Q.2 [Annexe.A]

Graphiquement on suspecte la présence d'une tendance, afin de s'en assurer on considère donc la série des différences premières $xpd_t = xp_t - xp_{t-1}$ $\forall t \in \{1, ..., T\}$. La série obtenue xpd_t ne présente elle plus de tendance déterministe [fig.7]. La tendance semble donc bien avoir été corrigée. L'objectif est maintenant de voir si la série xpd_t est stationnaire : on étudie donc ses autocorrélogrammes [fig.8&.9]. Les autocorrélations (totales ou partielles) dépassant les intervalles de confiance à 95% étant au maximum d'environ -0.35, ce qui est très faible et loin d'être égal à ± 1, les fonctions d'autocorrélation totale ou partielle ne montrent donc aucune corrélation avec une autre date : la série xpd_t semble donc être stationnaire, et on retient comme série de travail la série définie par :

$$xpd_t = xpd_t - xpd_{t-1} = (1 - L)xpd_t \ , \forall t \in T \tag{1}$$

Test de stationnarité

Afin de s'assurer de la bonne stationnarité xpd_t on effectue les tests classiques de racine unité : test de Dickey-Fuller augmenté[fig.10] et test de Perron-Phillips[fig.11]. On rappelle que pour les tests ADF et PP l'hypothèse nulle correspond à la présence de racine unité et donc de non stationnarité. La première partie de la question semble nous indiquer que la série xpd_t ne présente pas de tendance, on va donc effectuer un test de Dickey-Fuller augmenté sans tendance.

L'hypothèse de racine unité est rejetée au seuil de 95% dans les deux cas puisque les tests donnent des p-values strictement inférieures à 0,05. Ainsi, cela concorde avec la précédente conclusion et nous pouvons supposer pour la suite de l'étude que la série xpd_t est stationnaire.

Q.3

On représente la série initiale choisie xp_t possédant une tendance (linéaire) ainsi que la série transformée xpd_t stationnaire.

FIGURE 2 – Série initiale

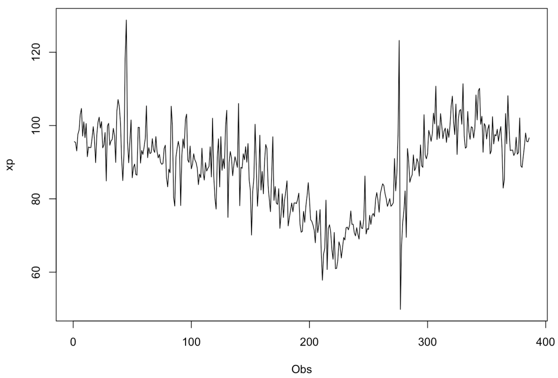
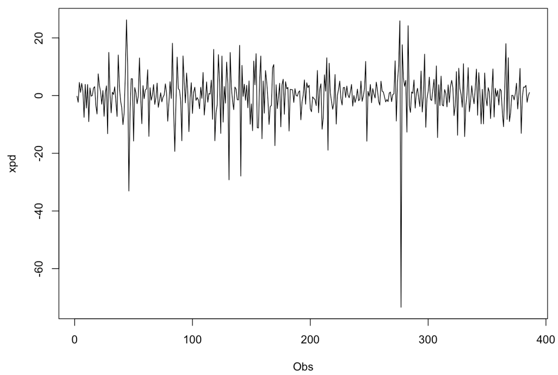


FIGURE 3 – Série transformée



2 Modèles ARMA

Q.4 [Annexe.B]

Soit X_t la série xpd_t centrée, on va choisir un modèle $ARMA(p,q)$ pour cette dernière.

La méthodologie appliquée est celle de Box-Jenkins :

Etape.0 : transformation de la série initiale afin de supprimer d'éventuelles non-stationnarités ;

Etape.1 : identification des ordres p et q ;

Etape.2 : ajustement du modèle i.e. estimation et significativité des coefficients ;

Etape.3 : validité du modèle i.e. étude des résidus ;

Etape.4 : sélection du modèle par critère d'informations

Identification des ordres

Les autocorrélations sont significatives, c'est-à-dire plus grandes que les bornes de confiance $\pm 1.96/\sqrt{n}$ d'un test de nullité de l'autocorrélation au seuil de 95%, jusqu'à $q = 2$ pour les totales et jusqu'à $p = 9$ pour les partielles [fig.12 & fig.13].

Ainsi, si X_t suit un $ARMA(p,q)$ il suit au plus un $ARMA(p = 9, q = 2)$. Les modèles possibles sont tous les sous-modèles $ARMA(p \leq 9, q \leq 2)$.

Ajustement & validité du modèle

On commence par vérifier si le modèle est valide i.e. que les résidus ne sont pas autocorrélés. On effectue un test de *Portemanteau*[B] afin de vérifier la blancheur des résidus estimés en testant la nullité jointe des autocorrélations jusqu'à un ordre k. On considère la statistique de Ljung-Box de l'hypothèse nulle de nullité jointe des autocorrélations des résidus (plus précisément sous *(H0)* les résidus sont indépendants et identiquement distribués donc leurs autocorrélations sont biens nulles) en corrigeant le degré de liberté de la loi suivie asymptotiquement par cette statistique du nombre de régresseurs $p+q = 11$. On ne peut donc effectuer le test pour un $lag \leq 11$. La p-value de ce dernier est de 0.59 [fig.14], ainsi l'hypothèse de nullité jointe de l'autocorrélation des résidus n'est pas rejetée au seuil de 95% : il y a absence d'autocorrélations des résidus jusqu'à 12 lags. Pour s'assurer de cette nullité jointe on pousse l'étude jusqu'à 20 lags (plus de un an et demi) : la blancheur des résidus semble alors vérifiée [fig.15]. En revanche, en testant la significativité des coefficients estimés le modèle apparaît comme mal ajusté : au moins l'un des coefficients des ordres les plus élevés n'est pas signififcatif au seuil de 95% [fig.16]. On doit donc s'intéresser aux sous-modèles.

On crée une fonction qui permet : de faire un test de Ljung-Box pour étudier l'absence d'autocorrélation des résidus (encore jusqu'à 20 lags) et d'étudier la significativité des coefficients pour les ordres les plus élevés : *arimafit* [fig.17]. Cette étude de significativité est faite pour chaque coefficient avec un test de Student dont l'hypothèse nulle est la nullité du-dit coefficient, on compare la t-statistique de test avec le quantile $q_{1-0.05/2}$ de la loi normale centrée réduite pour une significativité au seuil de 95%. Avec cette fonction nous pouvons donc :

1. Étudier la significativité des coefficients au seuil de 95%
2. Vérifier la blancheur des résidus

Après avoir fait l'ensemble ces tests sur les sous-modèles, deux d'entre eux sont valides et bien ajustés (cf code commenté pour les résultats). Afin de choisir un modèle pour la série X_t nous pourrions utiliser un critère d'information (AIC ou BIC) mais malheureusement ils ne s'accordent pas sur un des deux modèles proposés[fig.18].

On décide alors de garder celui qui donne la meilleure prévision sur l'échantillon. On définit le R^2 ajusté : $R^2 = 1 - \left(\frac{1}{n-p-q-1}\sum_{i=0}^{n}(X_i - \widehat{X}_i)^2\right) / \left(\frac{1}{n-1}\sum_{i=0}^{n}(X_i - \overline{X})^2\right)$ où $\overline{X} = 0$ série centrée.

Les résultats sont présentés sur la figure 19. L'$ARIMA(2,0,2)$ a le R^2 ajusté le plus élevé, il donne donc la meilleure prévision sur l'échantillon. On le garde comme meilleur modèle au final. Les résultats de la spécification pour le modèle retenu se retrouvent sur la figure suivante.

FIGURE 4 – Résultat de la spécification

```
Modèle estimé

                   ar1           ar2        ma1          ma2
coef_est 1.022184e+00 -2.385687e-01 -1.7346691 7.768956e-01
se_est   1.519028e-01  5.959678e-02  0.1527955 1.265533e-01
tstat    6.729198e+00  4.003047e+00 11.3528795 6.138883e+00
pvalue   1.706013e-11  6.253197e-05  0.0000000 8.310368e-10
```

Q.5

En conclusion de cette partie le modèle estimé pour X_t est donc un $ARIMA(2,0,2)$ prenant la forme suivante, où les $(\phi_i)_i$ sont les coefficients estimés de la partie autorégressive et les $(\psi_j)_j$ ceux de la partie moyenne mobile :

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} \tag{2}$$

$$\underline{X_t = 1.02\, X_{t-1} - 0.24\, X_{t-2} + \varepsilon_t - 1.73\, \varepsilon_{t-1} + 0.78\, \varepsilon_{t-2}} \tag{3}$$

3 Prévision

Q.6

On note T la longueur de la série et on suppose que les résidus sont gaussiens, i.e. que $\varepsilon_t \sim \mathcal{N}(0,\, \sigma_\varepsilon^2)$, avec $\sigma_\varepsilon^2 > 0$. On a également $X_t \sim ARIMA(2,2)$ selon l'équation (2). On cherche à établir une région de confiance de niveau α sur les valeurs futures $(X_{T+1}, X_{T+2}) \in \mathbb{R}^2$.

On a EL$(\varepsilon_{T+h}|X_T, X_{T-1}, \ldots)$= 0 $\forall h > 0$ (modèle causal et inversible), ce qui nous permet de déterminer les **prévisions optimales en T** :

$$\begin{cases} \widehat{X_{T+1|T}} = \phi_1 X_T + \phi_2 X_{T-1} + \psi_1 \varepsilon_T + \psi_2 \varepsilon_{T-1} \\ \widehat{X_{T+2|T}} = \phi_1 \widehat{X_{T+1|T}} + \phi_2 X_T + \psi_2 \varepsilon_T \end{cases}$$

On pose alors :

$$X = \begin{pmatrix} X_{T+1} \\ X_{T+2} \end{pmatrix} \quad \text{et} \quad \widehat{X} = \begin{pmatrix} \widehat{X_{T+1|T}} \\ \widehat{X_{T+2|T}} \end{pmatrix} \quad \text{soit} \quad X - \widehat{X} = \begin{pmatrix} X_{T+1} - \widehat{X_{T+1|T}} \\ X_{T+2} - \widehat{X_{T+2|T}} \end{pmatrix}$$

Le vecteur $X - \widehat{X}$ représente alors le vecteur des erreurs de prévision.

$$X - \widehat{X} = \begin{pmatrix} 1 & 0 \\ \phi_1 + \psi_1 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{T+1} \\ \varepsilon_{T+2} \end{pmatrix} = \boldsymbol{A\varepsilon} \quad avec \quad \varepsilon_t \sim \mathcal{N}(0,\, \sigma_\varepsilon^2)$$

$\Longrightarrow$ **Ainsi**, $X - \widehat{X} \sim \mathcal{N}(0,\, \sigma_\varepsilon^2 \boldsymbol{AA}')$, où :

$$\sigma_\varepsilon^2 \boldsymbol{AA}' = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \phi_1 + \psi_1 \\ \phi_1 + \psi_1 & 1 + (\phi_1 + \psi_1)^2 \end{pmatrix} = \boldsymbol{\Sigma} \tag{4}$$

Puisque $Det(\boldsymbol{\Sigma}) = \sigma_\varepsilon^2$, la matrice $\boldsymbol{\Sigma}$ (matrice de variance-covariance du vecteur des erreurs de prévision) est inversible si et seulement si $\sigma_\varepsilon^2 > 0$ qui est supposé vrai. Ainsi, on a directement :

$$\Rightarrow (X - \widehat{X})' \boldsymbol{\Sigma}^{-1} (X - \widehat{X}) \sim \chi^2(2) \tag{5}$$

On a alors grâce aux propriétés de la loi du χ^2 **une région de confiance de niveau α sur les valeurs futures** (X_{T+1}, X_{T+2}) (en notant $q_{1-\alpha}^{\chi^2}$ le quantile d'ordre $1-\alpha$ de cette loi) :

$$\boxed{\begin{aligned} &\forall \alpha \in [0;1], \quad P(\, (X - \widehat{X})' \boldsymbol{\Sigma}^{-1} (X - \widehat{X}) \leq q_{1-\alpha}^{\chi^2} \,) = 1 - \alpha \\ &\Longrightarrow \boldsymbol{IC(\alpha)} = \{x \in \mathbb{R}^2 \,, (x - \widehat{X})' \boldsymbol{\Sigma}^{-1} (x - \widehat{X}) \leq q_{1-\alpha}^{\chi^2}\} \end{aligned}} \tag{6}$$

D'un point de vue univarié, d'après 4, on a :

$$\begin{cases} X_{T+1} - \widehat{X_{T+1|T}} \sim \mathcal{N}(0,\, \sigma_\varepsilon^2) \\ X_{T+2} - \widehat{X_{T+2|T}} \sim \mathcal{N}(0,\, \sigma_\varepsilon^2 (1 + (\phi_1 + \psi_1)^2)\,) \end{cases}$$

On obtient donc les intervalles de confiance suivants pour le niveau $\alpha\%$ (en notant q_x le quantile d'ordre x de la loi normale centrée réduite) :

$$\boxed{\begin{aligned} IC_{100-\alpha\%}(X_{T+1}) &= [\; \widehat{X_{T+1|T}} \pm \sigma_\varepsilon \times q_{1-\alpha/2} \;] \\ IC_{100-\alpha\%}(X_{T+2}) &= [\; \widehat{X_{T+2|T}} \pm \sigma_\varepsilon \sqrt{1 + (\phi_1 + \psi_1)^2} \times q_{1-\alpha/2} \;] \end{aligned}} \tag{7}$$

Q.7

On a utilisé un certain nombre d'hypothèse pour trouver 6, qui sont :

- Le modèle est correct ;
- Les coefficients trouvés partie 2 sont exactement ceux du modèle ;
- Les résidus sont gaussiens $\varepsilon_t \sim \mathcal{N}(0,\, \sigma_\varepsilon^2)$;
- variance connue $\sigma_\varepsilon^2 > 0$

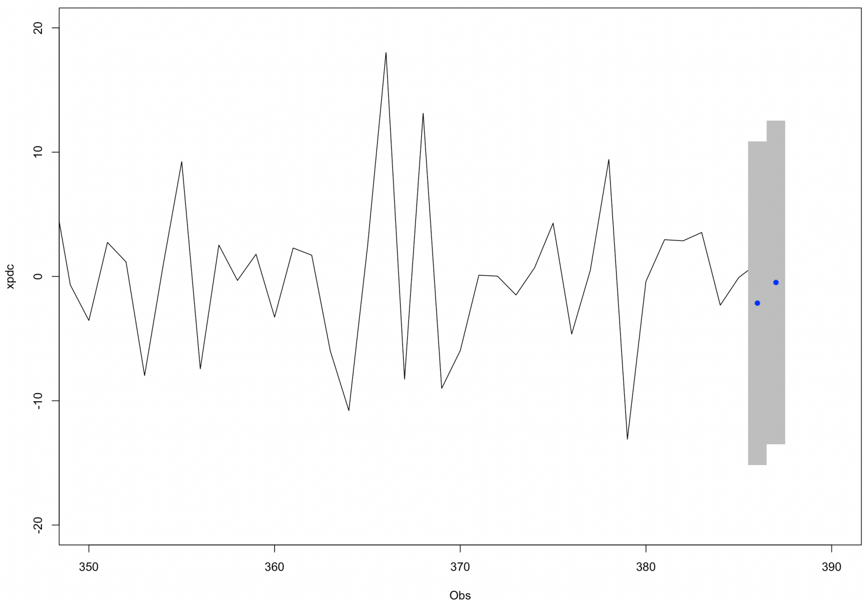
L'hypothèse (3) sur les résidus elle semble être vérifiée lorsqu'on s'intéresse aux résidus du modèle $ARIMA(2,0,2)$ que nous avons choisi. On peut tester l'hypothèse de normalité des résidus en réalisant un **QQ-plot** pour ces derniers. Les points semblent bien s'aligner sur une droite, les résidus semblent donc bien suivre une loi normale [fig.20]. De plus, on a déjà vu que l'hypothèse d'indépendance des résidus n'était pas rejetée lorsque nous avons effectué un test de *Portemanteau* [fig.17]. Ne pas connaître la variance des résidus, contraire à l'hypothèse (4), implique une estimation de cette quantité ce qui à pour conséquence de rendre plus larges bornes de la région de confiance (bruit dans l'estimation).

Q.8

Représentation graphique

Nous pouvons représenter graphiquement les régions de confiance au seuil de 95% trouvées avec l'équation 7 (niveau $\alpha = 5\%$). On trouvera sur la figure en gris les régions de confiance et les points bleus correspondent aux prédictions.

FIGURE 5 – Prédictions à horizon 1 et 2 de X_T



Q.9

Question ouverte

On dit que Y_{T+1} est utile pour prédire X_{T+1} à la date T si la série (Y_t) *cause instantanément* (X_t) i.e. si Y_t cause instantanément X_t au sens de Granger.

Si on considère le processus $Z_t = (X_t, Y_t)'$ et qu'on estime un modèle $VAR(p)$ on peut effectuer un test de Wald et tester l'hypothèse de causalité instantanée entre X_t et Y_t.

Annexes graphiques

A Q.2

Figure 6 – Série brute initiale

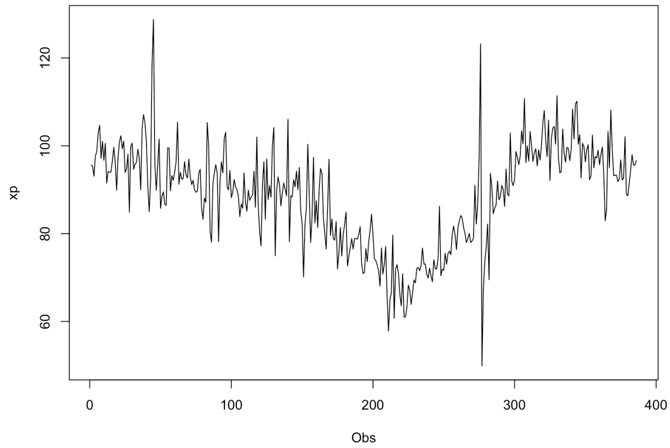


Figure 7 – Série différenciée

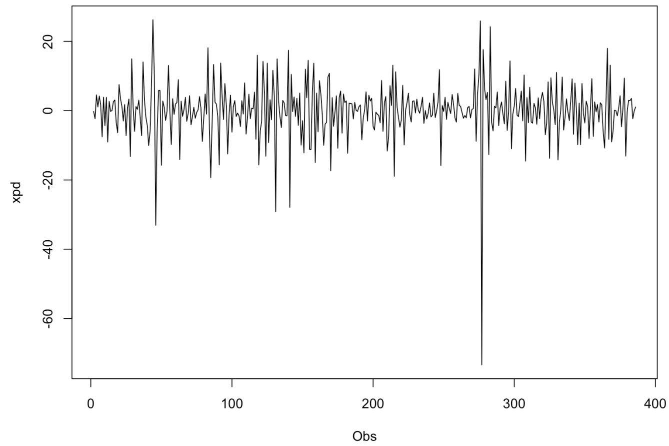


Figure 8 – Autocorrélogramme total

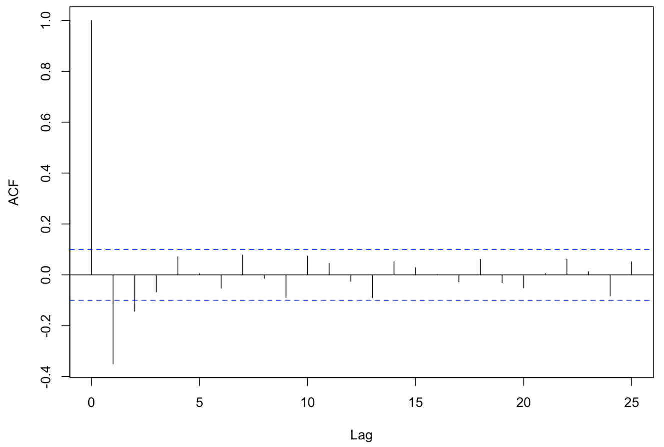
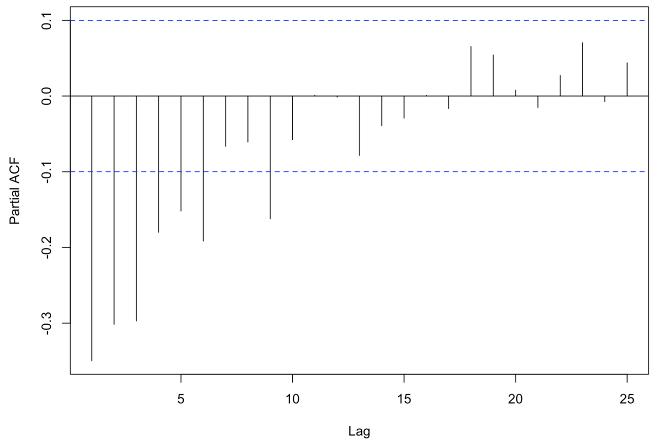


Figure 9 – Autocorrélogramme partiel



Rappel : les lignes bleues sur les corrélogrammes représentent les intervalles de confiance à 95% sous l'hypothèse nulle i.e. nullité jointe des autocorrélations.

Figure 10 – Test Dickey-Fuller augmenté

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 0
  STATISTIC:
    Dickey-Fuller: -28.1527
  P VALUE:
    0.01
```

Figure 11 – Test Perron-Phillips

```
> stationary.test(xpd,method="pp",type="Z_tau")
Phillips-Perron Unit Root Test
alternative: stationary

Type 1: no drift no trend
 lag Z_tau p.value
   5   -40    0.01
-----
 Type 2: with drift no trend
 lag Z_tau p.value
   5 -39.9    0.01
-----
 Type 3: with drift and trend
 lag Z_tau p.value
   5 -39.9    0.01
---------------
Note: p-value = 0.01 means p.value <= 0.01
```

B Q.4

Figure 12 – Autocorrélogramme total pour X_t

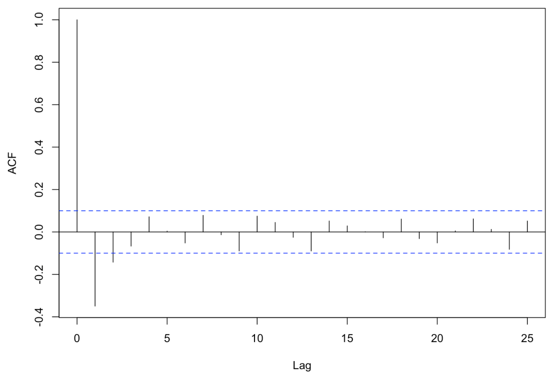
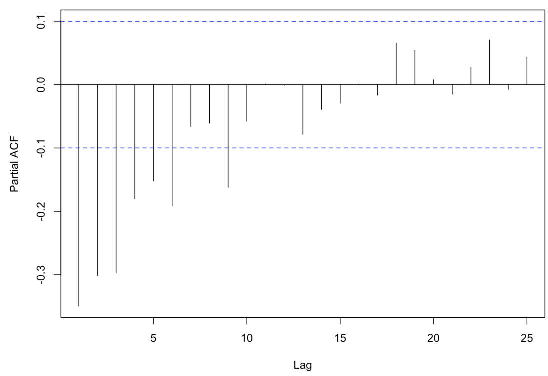


Figure 13 – Autocorrélogramme partiel pour X_t



<u>Notes sur le test de *Portemanteau* :</u> au lieu de tester la nullité des autocorrélations pour chaque lag on considère une statistique globale. L'hypothèse nulle est alors que les résidus sont i.i.d. ; cette hypothèse est testée sur H lags et sous $(H0)$ la statistique de test Q suit alors asymptotiquement (en le nombre d'observations n) une loi du χ^2. Avec la statistique de Box-Pierce en dépit d'un grand nombre d'observation sa distribution reste éloignée de celle d'une χ^2, on préférera donc utiliser la statistique de Ljung-Box dont la distribution est plus proche d'une χ^2 même pour de petits échantillons.

Figure 14 – Test de Ljung-Box du modèle $ARIMA(9,0,2)$ pour 12 lags

```
	Box-Ljung test

data:  test92$residuals
X-squared = 0.29787, df = 1, p-value = 0.5852
```

Figure 15 – Test de Ljung-Box du modèle $ARIMA(9,0,2)$ pour différents lags

```
> TestAutocorr(xpdc,9,0,2)
lag= 12 pvalue= 0.5852201
lag= 13 pvalue= 0.7292566
lag= 14 pvalue= 0.5244197
lag= 15 pvalue= 0.4643361
lag= 16 pvalue= 0.3856995
lag= 17 pvalue= 0.4829412
lag= 18 pvalue= 0.4685035
lag= 19 pvalue= 0.5722092
lag= 20 pvalue= 0.6604265
=> Modèle valide
```

Figure 16 – Test de significativité du modèle $ARIMA(9,0,2)$

```
> TestSignificatif(xpdc,9,0,2)
2/ Test de nullité des coefficients des des ordres les plus élevés :
                 ar9        ma2
coef_est -0.13127645 -0.1115692
se_est    0.06445323  0.3877521
tstat     2.03677074  0.2877333
pval      0.04167302  0.7735509

Significativité au seuil de 95%
  ar9   ma2
 TRUE FALSE
=> Modèle rejeté : au moins l'un des coefficients des ordres les plus élevés n'est pas signifcatif au seuil de 95%
```

Figure 17 – Exemple : *arimafit* pour un ARIMA$(p = 2, 0, q = 2)$

```
> arimafit(xpdc,2,0,2) #V
Try ARIMA( 2 , 0 , 2 )

1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
lag= 5 pvalue= 0.1432252
lag= 6 pvalue= 0.2748197
lag= 7 pvalue= 0.3125677
lag= 8 pvalue= 0.4408755
lag= 9 pvalue= 0.3063634
lag= 10 pvalue= 0.3215626
lag= 11 pvalue= 0.3841944
lag= 12 pvalue= 0.4452545
lag= 13 pvalue= 0.3322722
lag= 14 pvalue= 0.3646585
lag= 15 pvalue= 0.3824745
lag= 16 pvalue= 0.4420231
lag= 17 pvalue= 0.5231237
lag= 18 pvalue= 0.5049354
lag= 19 pvalue= 0.5623255
lag= 20 pvalue= 0.5751821
=> Modèle valide

2/ Test de nullité des coefficients des des ordres les plus élevés :
                   ar2          ma2
coef_est -2.385687e-01 7.768956e-01
se_est    5.959678e-02 1.265533e-01
tstat     4.003047e+00 6.138883e+00
pval      6.253197e-05 8.310368e-10

Significativité au seuil de 95%
 ar2  ma2
TRUE TRUE
=> Modèle bien ajusté
```

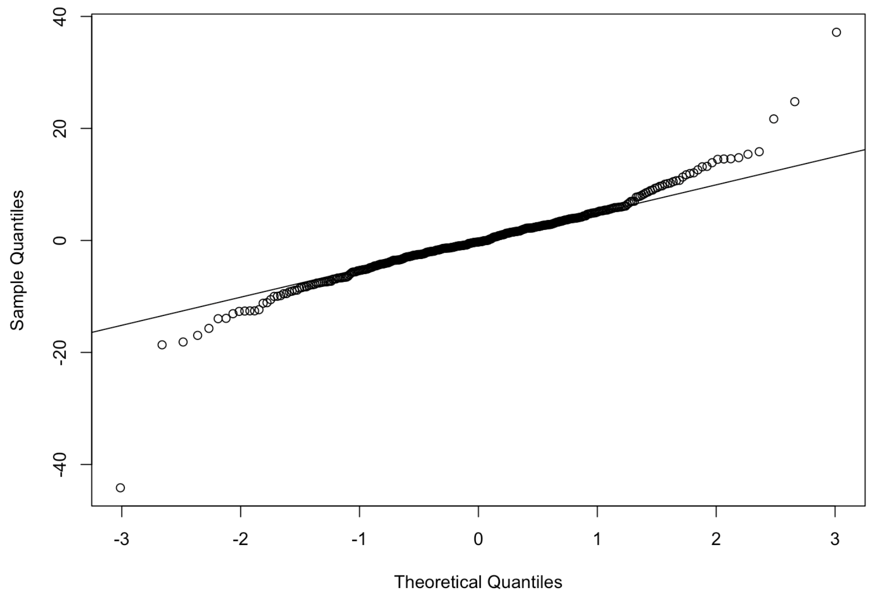
FIGURE 18 – Choix de modèle par critère d'informations

```
> minInfo()
ar2ma2 | AIC : 2562.248 | BIC : 2582.014
ar2ma1 | AIC : 2564.625 | BIC : 2580.438

=> Choix AIC : ar2ma2
=> Choix BIC : ar2ma1
```

FIGURE 19 – Calcul des R^2 ajustés

```
> r2_aj(ar2ma1)
R^2 ajusté :  0.3414002

> r2_aj(ar2ma2)
R^2 ajusté :  0.3474559
```

FIGURE 20 – QQ-plot pour les résidus du modèle $ARIMA(2,0,2)$



Annexe : code commenté

Ce notebook regroupe l'ensemble du code commenté pour le projet de séries temporelles. Se repporter aux questions correspondantes dans le rapport plus haut pour avoir l'analyse et l'interprétation des résultats. Les pages 9 à 19 rapportent les résultats de l'ensemble des tests pour les sous-modèles, ces pages peuvent donc être sautées lors de la lecture.

```
#Packages requis
require(fUnitRoots)
```

```
## Loading required package: fUnitRoots

## Loading required package: timeDate

## Loading required package: timeSeries

## Loading required package: fBasics
```

```
require(aTSA)
```

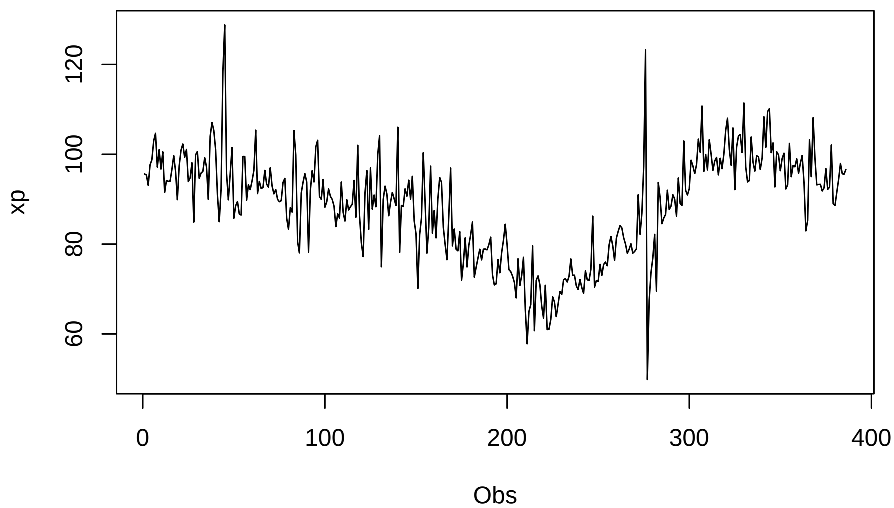
```
## Loading required package: aTSA

##
## Attaching package: 'aTSA'

## The following object is masked from 'package:graphics':
##
##     identify
```

```
require(astsa)
```

```
## Loading required package: astsa

##
## Attaching package: 'astsa'

## The following object is masked from 'package:fBasics':
##
##     nyse
```

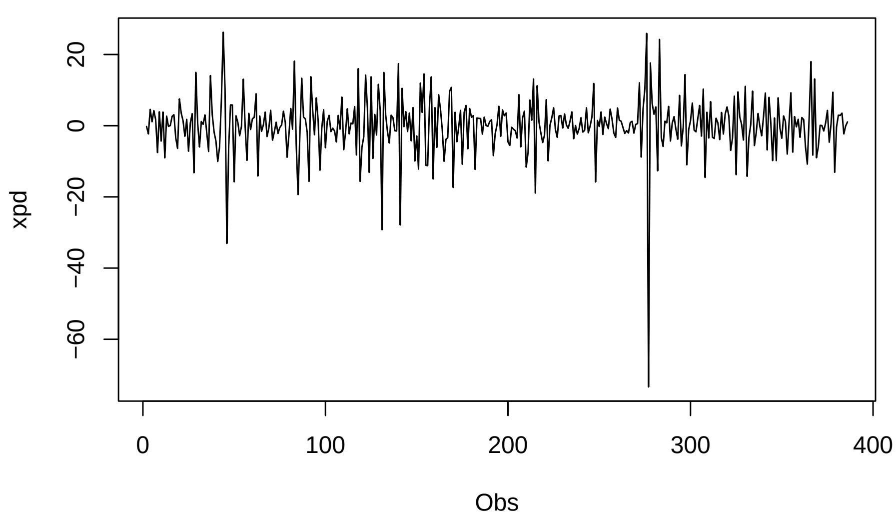
```
#Importations des données
data=read.csv('valeurs_mensuelles.csv',sep=";")
xp.source <- ts(data[[2]]) #données dans la 2e colonne
L <- length(xp.source)
xp <- ts(as.numeric(xp.source[4:L])) #4 premiers éléments h-s
```

```r
plot(xp,xlab="Obs") #Représentation de la série
```

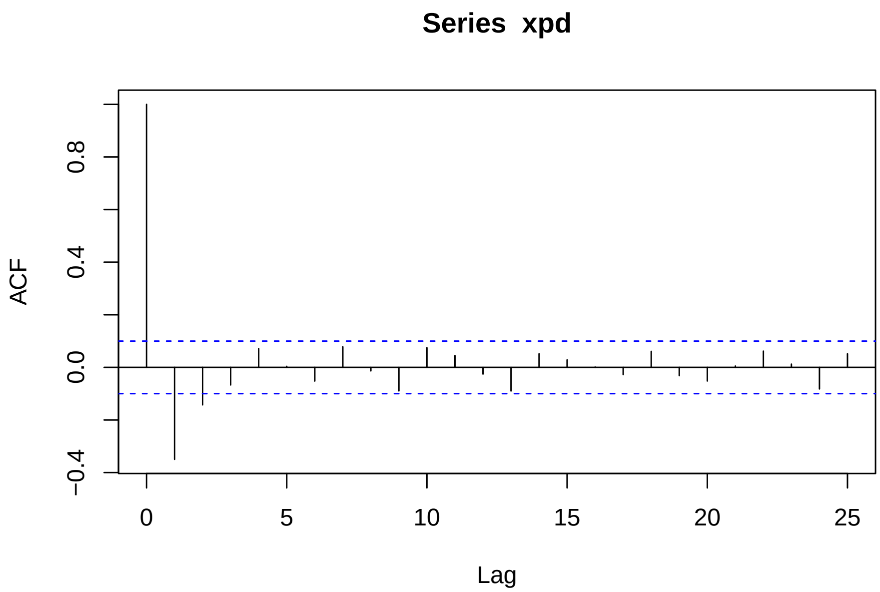


Q.2 & 3

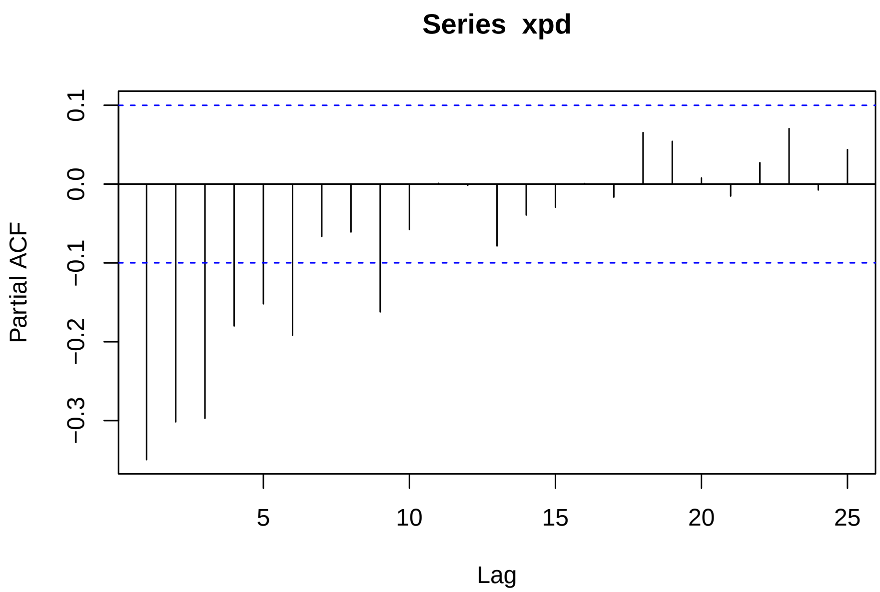
```r
#La série n'a pas l'air stationnaire, on la différencie une fois
xpd <- diff(xp,1)
plot(xpd,xlab="Obs")
```




```r
acf(xpd) #Fonction d'autocorrélation totale
```




```
pacf(xpd) #Fonction d'autocorrélation partielle
```



Test de stationnarité

```
# Test de racine unité
adfTest(xpd,lags=0,type="c") #ADF
```

```
## Warning in adfTest(xpd, lags = 0, type = "c"): p-value smaller than printed p-
## value
```

```
##
## Title:
```

```
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 0
##   STATISTIC:
##     Dickey-Fuller: -28.1527
##   P VALUE:
##     0.01
##
## Description:
##  Tue May  3 16:22:54 2022 by user:
```

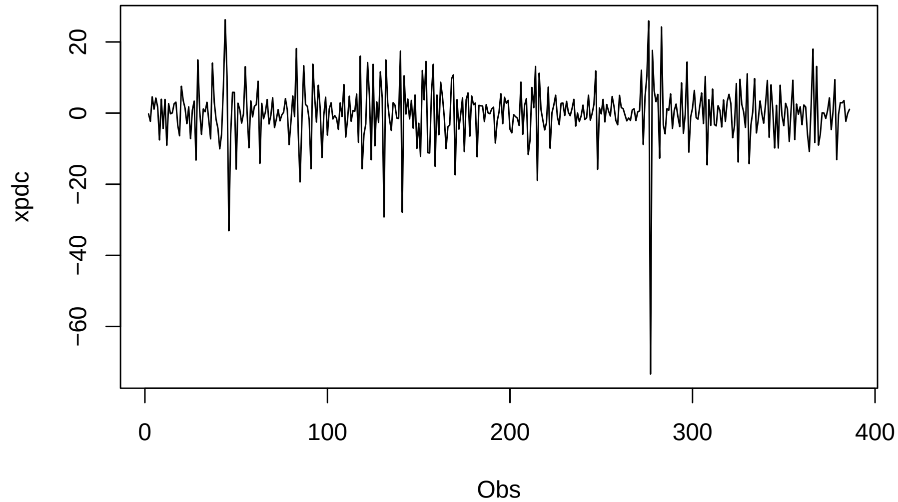
```
stationary.test(xpd,method="pp",type="Z_tau")
```

```
## Phillips-Perron Unit Root Test
## alternative: stationary
##
## Type 1: no drift no trend
##  lag Z_tau p.value
##    5   -40    0.01
## -----
##  Type 2: with drift no trend
##  lag Z_tau p.value
##    5 -39.9    0.01
## -----
##  Type 3: with drift and trend
##  lag Z_tau p.value
##    5 -39.9    0.01
## ---------------
## Note: p-value = 0.01 means p.value <= 0.01
```

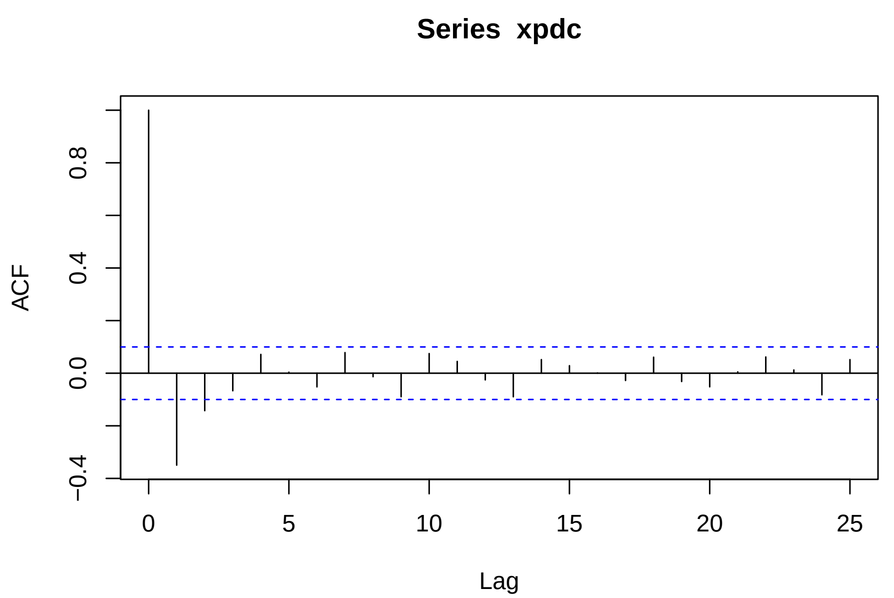
```
#p-value proche de 0 : on rejette l'hypothèse nulle de racine unité
```

Recherche des ordres p et q

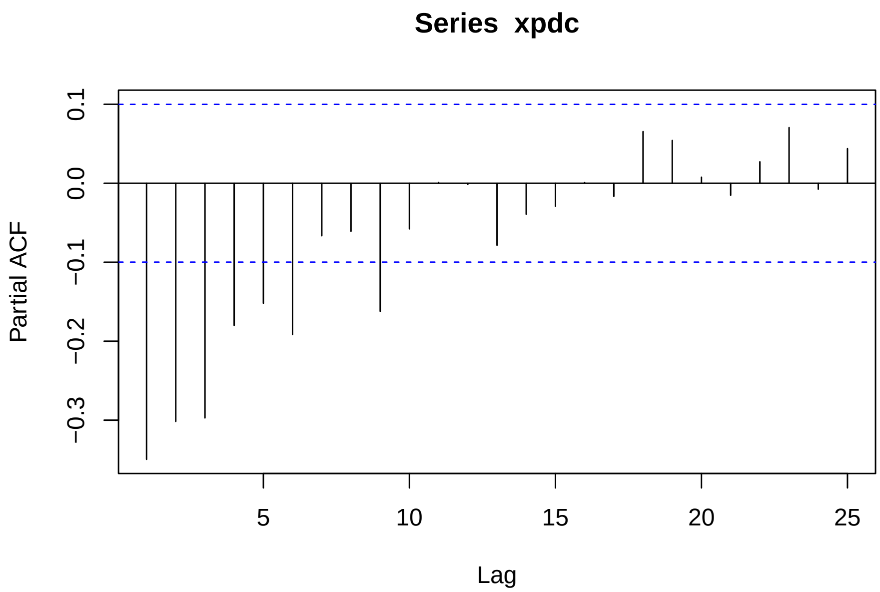
```
xpdc <- xpd - mean(xpd) #on centre la série
plot(xpdc,xlab="Obs")
```




```
acf(xpdc) #q=2
```




```
pacf(xpdc) #p=9
```




```
#Si xpdc suit un ARMA elle suit au plus un ARMA(p=9,q=2)
test92 <- arima(xpdc,order=c(9,0,2),include.mean=FALSE)
```

Commentaires : on doit vérifier que le modèle est valide ie que les résidus ne sont pas autocorrélés. On effectue un test de portemanteau avec statistique de Ljung-Box. On vérifie ensuite que le modèle est bien ajusté en étudiant la significativité à 95% des coefficients des ordres les plus élevés.

Q.4 1.Test d'autocorrélation des résidus

```
Box.test(test92$residuals,lag=12,type="Ljung-Box",fitdf=11)
```

```
##
##  Box-Ljung test
##
## data:  test92$residuals
## X-squared = 0.29787, df = 1, p-value = 0.5852
```

```
# On corrige les degrés de libertés du nombre de régresseurs!p+q=11,
#on ne peut effectuer le test pour lag<12
```

```
# On crée une fonction pour tester la validité des différents modèles
# (H0) : nullité jointe des autocorrélations des résidus

TestAutocorr<-function(var,p,d,q){
  autocorL<-list()
  fitdf<-p+q
  serie<-arima(var,order=c(p,d,q),include.mean=FALSE)
  rej<-0
  for (i in c(1:20)){
    if (i<=fitdf) {NA}
```

```r
    else {
      cat("lag=",i,"pvalue=",Box.test(residuals(serie),lag=i,
                                      type="Ljung-Box",
                                      fitdf = fitdf)$p.value,"\n")
      rej<-rej+(Box.test(residuals(serie),lag=i,fitdf = fitdf)$p.value<0.05)}
  }
  if (rej>0) cat("=> Modèle non valide : l'absence d'autocorrélation
                    des résidus est rejetée à 95%.\n")
  else cat("=> Modèle valide\n")
}
TestAutocorr(xpdc,9,0,2)
```

```
## lag= 12 pvalue= 0.5852201
## lag= 13 pvalue= 0.7292566
## lag= 14 pvalue= 0.5244197
## lag= 15 pvalue= 0.4643361
## lag= 16 pvalue= 0.3856995
## lag= 17 pvalue= 0.4829412
## lag= 18 pvalue= 0.4685035
## lag= 19 pvalue= 0.5722092
## lag= 20 pvalue= 0.6604265
## => Modèle valide
```

```r
#le modèle est valide car l'hypothèse nulle n'est jamais rejetée
#à 95% (pvalue<0.05) jusqu'à 20 retards
```

2.Test de significativité

```r
#On étudie la significativité des coefficients pour les ordres les plus élevés:
TestSignificatif <- function(serie,p,d,q){
  var<-arima(serie,order=c(p,d,q),include.mean=FALSE)
  coef<-var$coef
  se<-sqrt(diag(var$var.coef))
  t<-abs(coef/se)
  t2<-t>1.96  #on rejette l'hypothèse nulle du test de Student ?
  sig<-c(t2[p],t2[p+q])
    #On regarde si les coefficients des plus grands ordres sont significatifs
  res<-t2[p]+t2[p+q]

  #Résultats d'intérêt
  coef_est<-c(coef[p],coef[p+q])
  se_est<-c(se[p],se[p+q])
  tstat<-c(t[p],t[p+q])
  pval <- (1-pnorm(abs(tstat)))*2

  cat("2/ Test de nullité des coefficients des des ordres les plus élevés :\n")
  print(rbind(coef_est,se_est,tstat,pval))
  cat(" \n")
  cat("Significativité au seuil de 95% \n")
  print(sig)
  if (res!=2){cat("=> Modèle rejeté : au moins l'un des coefficients des ordres
                  les plus élevés n'est pas signififcatif au seuil de 95%\n")}
```

```r
  else{cat("=> Modèle bien ajusté\n")}
}
TestSignificatif(xpdc,9,0,2)
```

```
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                   ar9        ma2
## coef_est -0.13127645 -0.1115692
## se_est    0.06445323  0.3877521
## tstat     2.03677074  0.2877333
## pval      0.04167302  0.7735509
##
## Significativité au seuil de 95%
##   ar9   ma2
##  TRUE FALSE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

On définit donc une fonction **arimafit** qui permet d'étudier les sous-modèles.

```r
arimafit<-function(serie,p,d,q){
  cat("Try ARIMA(",p,",",d,",",q,")\n")
  cat("\n")
  cat("1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :\n")
  TestAutocorr(serie,p,d,q)
  cat(" \n")
  TestSignificatif(serie,p,d,q)
}
```

```r
#Test des sous-modèles
arimafit(xpdc,9,0,2)
```

```
## Try ARIMA( 9 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 12 pvalue= 0.5852201
## lag= 13 pvalue= 0.7292566
## lag= 14 pvalue= 0.5244197
## lag= 15 pvalue= 0.4643361
## lag= 16 pvalue= 0.3856995
## lag= 17 pvalue= 0.4829412
## lag= 18 pvalue= 0.4685035
## lag= 19 pvalue= 0.5722092
## lag= 20 pvalue= 0.6604265
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                   ar9        ma2
## coef_est -0.13127645 -0.1115692
## se_est    0.06445323  0.3877521
## tstat     2.03677074  0.2877333
## pval      0.04167302  0.7735509
##
```

```
## Significativité au seuil de 95%
##   ar9   ma2
##  TRUE FALSE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,8,0,2)
```

```
## Try ARIMA( 8 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 11 pvalue= 0.09323723
## lag= 12 pvalue= 0.1874852
## lag= 13 pvalue= 0.1462906
## lag= 14 pvalue= 0.2299154
## lag= 15 pvalue= 0.3028068
## lag= 16 pvalue= 0.3775144
## lag= 17 pvalue= 0.4892671
## lag= 18 pvalue= 0.507351
## lag= 19 pvalue= 0.5905584
## lag= 20 pvalue= 0.6537415
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                   ar8          ma2
## coef_est -0.10051430 8.070041e-01
## se_est    0.06785243 1.168885e-01
## tstat     1.48136610 6.904052e+00
## pval      0.13850904 5.053957e-12
##
## Significativité au seuil de 95%
##   ar8   ma2
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,7,0,2)
```

```
## Try ARIMA( 7 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 10 pvalue= 0.04866708
## lag= 11 pvalue= 0.0919076
## lag= 12 pvalue= 0.1886805
## lag= 13 pvalue= 0.2508765
## lag= 14 pvalue= 0.2460833
## lag= 15 pvalue= 0.2219089
## lag= 16 pvalue= 0.2208462
## lag= 17 pvalue= 0.2830484
## lag= 18 pvalue= 0.2640337
## lag= 19 pvalue= 0.3382754
## lag= 20 pvalue= 0.4018944
## => Modèle valide
```

```
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                 ar7        ma2
## coef_est 0.03451329 -0.1803102
## se_est   0.06905891  0.3619194
## tstat    0.49976588  0.4982056
## pval     0.61723994  0.6183391
##
## Significativité au seuil de 95%
##   ar7   ma2
## FALSE FALSE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,6,0,2)
```

```
## Try ARIMA( 6 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 9 pvalue= 0.1573839
## lag= 10 pvalue= 0.1247692
## lag= 11 pvalue= 0.1685021
## lag= 12 pvalue= 0.2818565
## lag= 13 pvalue= 0.3348548
## lag= 14 pvalue= 0.3179623
## lag= 15 pvalue= 0.2817
## lag= 16 pvalue= 0.2777397
## lag= 17 pvalue= 0.341627
## lag= 18 pvalue= 0.3143429
## lag= 19 pvalue= 0.389326
## lag= 20 pvalue= 0.4486929
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar6        ma2
## coef_est -0.09066988 -0.2933455
## se_est    0.05881503  0.3027701
## tstat     1.54161059  0.9688721
## pval      0.12316825  0.3326090
##
## Significativité au seuil de 95%
##   ar6   ma2
## FALSE FALSE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,5,0,2)
```

```
## Try ARIMA( 5 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 8 pvalue= 0.1353459
## lag= 9 pvalue= 0.1611489
```

```
## lag= 10 pvalue= 0.1697906
## lag= 11 pvalue= 0.2411754
## lag= 12 pvalue= 0.3257572
## lag= 13 pvalue= 0.2585838
## lag= 14 pvalue= 0.3090361
## lag= 15 pvalue= 0.3521921
## lag= 16 pvalue= 0.4312897
## lag= 17 pvalue= 0.5254826
## lag= 18 pvalue= 0.5594907
## lag= 19 pvalue= 0.5833898
## lag= 20 pvalue= 0.5456781
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar5         ma2
## coef_est -0.04060856  0.88208675
## se_est    0.05718140  0.04046756
## tstat     0.71017068 21.79737804
## pval      0.47759830  0.00000000
##
## Significativité au seuil de 95%
##   ar5   ma2
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                   les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,4,0,2)
```

```
## Try ARIMA( 4 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 7 pvalue= 0.08505846
## lag= 8 pvalue= 0.2268833
## lag= 9 pvalue= 0.2728448
## lag= 10 pvalue= 0.1923538
## lag= 11 pvalue= 0.2035898
## lag= 12 pvalue= 0.2981164
## lag= 13 pvalue= 0.3288473
## lag= 14 pvalue= 0.2828082
## lag= 15 pvalue= 0.2385584
## lag= 16 pvalue= 0.2529927
## lag= 17 pvalue= 0.3110568
## lag= 18 pvalue= 0.2627287
## lag= 19 pvalue= 0.32753
## lag= 20 pvalue= 0.3707625
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar4        ma2
## coef_est -0.02552058 -0.3203678
## se_est    0.16343955  1.0641320
## tstat     0.15614691  0.3010602
## pval      0.87591722  0.7633686
##
```

```
## Significativité au seuil de 95%
##   ar4   ma2
## FALSE FALSE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,3,0,2)
```

```
## Try ARIMA( 3 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 6 pvalue= 0.1689453
## lag= 7 pvalue= 0.2034945
## lag= 8 pvalue= 0.3611974
## lag= 9 pvalue= 0.3447224
## lag= 10 pvalue= 0.3153408
## lag= 11 pvalue= 0.3844916
## lag= 12 pvalue= 0.4713922
## lag= 13 pvalue= 0.3601506
## lag= 14 pvalue= 0.4068794
## lag= 15 pvalue= 0.4525685
## lag= 16 pvalue= 0.5203119
## lag= 17 pvalue= 0.6050019
## lag= 18 pvalue= 0.6353903
## lag= 19 pvalue= 0.6475155
## lag= 20 pvalue= 0.6118133
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                 ar3         ma2
## coef_est 0.09152807  0.88407962
## se_est   0.05469213  0.03345167
## tstat    1.67351432 26.42856783
## pval     0.09422609  0.00000000
##
## Significativité au seuil de 95%
##   ar3   ma2
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,2,0,2) #V
```

```
## Try ARIMA( 2 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 5 pvalue= 0.1432252
## lag= 6 pvalue= 0.2748197
## lag= 7 pvalue= 0.3125677
## lag= 8 pvalue= 0.4408755
## lag= 9 pvalue= 0.3063634
## lag= 10 pvalue= 0.3215626
## lag= 11 pvalue= 0.3841944
```

```
## lag= 12 pvalue= 0.4452545
## lag= 13 pvalue= 0.3322722
## lag= 14 pvalue= 0.3646585
## lag= 15 pvalue= 0.3824745
## lag= 16 pvalue= 0.4420231
## lag= 17 pvalue= 0.5231237
## lag= 18 pvalue= 0.5049354
## lag= 19 pvalue= 0.5623255
## lag= 20 pvalue= 0.5751821
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                   ar2          ma2
## coef_est -2.385687e-01 7.768956e-01
## se_est    5.959678e-02 1.265533e-01
## tstat     4.003047e+00 6.138883e+00
## pval      6.253197e-05 8.310368e-10
##
## Significativité au seuil de 95%
##  ar2  ma2
## TRUE TRUE
## => Modèle bien ajusté
```

```
arimafit(xpdc,1,0,2)
```

```
## Try ARIMA( 1 , 0 , 2 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 4 pvalue= 0.0345022
## lag= 5 pvalue= 0.1049022
## lag= 6 pvalue= 0.2047588
## lag= 7 pvalue= 0.15937
## lag= 8 pvalue= 0.2511404
## lag= 9 pvalue= 0.3043682
## lag= 10 pvalue= 0.1813799
## lag= 11 pvalue= 0.1605287
## lag= 12 pvalue= 0.2248902
## lag= 13 pvalue= 0.227911
## lag= 14 pvalue= 0.2027189
## lag= 15 pvalue= 0.1693728
## lag= 16 pvalue= 0.1686662
## lag= 17 pvalue= 0.2082152
## lag= 18 pvalue= 0.178651
## lag= 19 pvalue= 0.2236316
## lag= 20 pvalue= 0.2560769
## => Modèle non valide : l'absence d'autocorrélation
##                         des résidus est rejetée à 95%.
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                 ar1         ma2
## coef_est -0.1432916 -0.27479657
## se_est    0.2066225  0.16181079
## tstat     0.6934949  1.69825861
## pval      0.4879990  0.08945896
```

```
##
## Significativité au seuil de 95%
##   ar1   ma2
## FALSE FALSE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                   les plus élevés n'est pas signififcatif au seuil de 95%
```

```r
arimafit(xpdc,9,0,1)
```

```
## Try ARIMA( 9 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 11 pvalue= 0.5961159
## lag= 12 pvalue= 0.8078979
## lag= 13 pvalue= 0.8901603
## lag= 14 pvalue= 0.6559216
## lag= 15 pvalue= 0.5765549
## lag= 16 pvalue= 0.4844039
## lag= 17 pvalue= 0.5758309
## lag= 18 pvalue= 0.5677895
## lag= 19 pvalue= 0.6615761
## lag= 20 pvalue= 0.735524
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar9        ma1
## coef_est -0.13883361 -0.3599022
## se_est    0.06533146  0.3388757
## tstat     2.12506517  1.0620479
## pval      0.03358118  0.2882139
##
## Significativité au seuil de 95%
##   ar9   ma1
##  TRUE FALSE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                   les plus élevés n'est pas signififcatif au seuil de 95%
```

```r
arimafit(xpdc,8,0,1)
```

```
## Try ARIMA( 8 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 10 pvalue= 0.06083528
## lag= 11 pvalue= 0.1365409
## lag= 12 pvalue= 0.2606386
## lag= 13 pvalue= 0.3274695
## lag= 14 pvalue= 0.3156024
## lag= 15 pvalue= 0.2732039
## lag= 16 pvalue= 0.2733268
## lag= 17 pvalue= 0.3437463
## lag= 18 pvalue= 0.3125587
## lag= 19 pvalue= 0.3930759
## lag= 20 pvalue= 0.4600261
```

```
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar8           ma1
## coef_est -0.05993369 -7.293044e-01
## se_est    0.05620716  1.081463e-01
## tstat     1.06629981  6.743683e+00
## pval      0.28628814  1.544209e-11
##
## Significativité au seuil de 95%
##   ar8   ma1
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,7,0,1)
```

```
## Try ARIMA( 7 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 9 pvalue= 0.2068413
## lag= 10 pvalue= 0.1360321
## lag= 11 pvalue= 0.1746102
## lag= 12 pvalue= 0.2906804
## lag= 13 pvalue= 0.3496475
## lag= 14 pvalue= 0.3308067
## lag= 15 pvalue= 0.2960845
## lag= 16 pvalue= 0.2881226
## lag= 17 pvalue= 0.3530672
## lag= 18 pvalue= 0.3281422
## lag= 19 pvalue= 0.4048827
## lag= 20 pvalue= 0.4673127
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                 ar7         ma1
## coef_est 0.04234466 -0.77904884
## se_est   0.05943787  0.08229544
## tstat    0.71241893  9.46648875
## pval     0.47620540  0.00000000
##
## Significativité au seuil de 95%
##   ar7   ma1
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,6,0,1)
```

```
## Try ARIMA( 6 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 8 pvalue= 0.3384825
```

```
## lag= 9 pvalue= 0.2582384
## lag= 10 pvalue= 0.2045034
## lag= 11 pvalue= 0.2349639
## lag= 12 pvalue= 0.3506468
## lag= 13 pvalue= 0.3907003
## lag= 14 pvalue= 0.3623109
## lag= 15 pvalue= 0.3187136
## lag= 16 pvalue= 0.3152607
## lag= 17 pvalue= 0.3783784
## lag= 18 pvalue= 0.3433515
## lag= 19 pvalue= 0.41649
## lag= 20 pvalue= 0.4702508
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar6         ma1
## coef_est -0.06349989 -0.74696589
## se_est    0.05974350  0.08086264
## tstat     1.06287535  9.23746636
## pval      0.28783848  0.00000000
##
## Significativité au seuil de 95%
##   ar6   ma1
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                   les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,5,0,1)
```

```
## Try ARIMA( 5 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 7 pvalue= 0.1371269
## lag= 8 pvalue= 0.3188309
## lag= 9 pvalue= 0.3394436
## lag= 10 pvalue= 0.245746
## lag= 11 pvalue= 0.2662432
## lag= 12 pvalue= 0.3755392
## lag= 13 pvalue= 0.411762
## lag= 14 pvalue= 0.3722287
## lag= 15 pvalue= 0.3169379
## lag= 16 pvalue= 0.3260815
## lag= 17 pvalue= 0.3897247
## lag= 18 pvalue= 0.3418867
## lag= 19 pvalue= 0.4135635
## lag= 20 pvalue= 0.4651095
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar5         ma1
## coef_est -0.05485478 -0.78849057
## se_est    0.05735908  0.05618503
## tstat     0.95633993 14.03381969
## pval      0.33890052  0.00000000
```

```
##
## Significativité au seuil de 95%
##   ar5   ma1
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,4,0,1)
```

```
## Try ARIMA( 4 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 6 pvalue= 0.2178886
## lag= 7 pvalue= 0.2190911
## lag= 8 pvalue= 0.3857786
## lag= 9 pvalue= 0.4050561
## lag= 10 pvalue= 0.2885708
## lag= 11 pvalue= 0.2900312
## lag= 12 pvalue= 0.3921434
## lag= 13 pvalue= 0.4166621
## lag= 14 pvalue= 0.3561757
## lag= 15 pvalue= 0.3026233
## lag= 16 pvalue= 0.3165941
## lag= 17 pvalue= 0.3780263
## lag= 18 pvalue= 0.3200969
## lag= 19 pvalue= 0.3880036
## lag= 20 pvalue= 0.4302827
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                 ar4         ma1
## coef_est 0.00444119 -0.81270166
## se_est   0.05799383  0.04412482
## tstat    0.07658039 18.41824387
## pval     0.93895736  0.00000000
##
## Significativité au seuil de 95%
##   ar4   ma1
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                    les plus élevés n'est pas signififcatif au seuil de 95%
```

```
arimafit(xpdc,3,0,1)
```

```
## Try ARIMA( 3 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 5 pvalue= 0.4798891
## lag= 6 pvalue= 0.4509925
## lag= 7 pvalue= 0.3818217
## lag= 8 pvalue= 0.5469756
## lag= 9 pvalue= 0.5417134
## lag= 10 pvalue= 0.3994607
```

```
## lag= 11 pvalue= 0.3904518
## lag= 12 pvalue= 0.4945608
## lag= 13 pvalue= 0.5124961
## lag= 14 pvalue= 0.4414424
## lag= 15 pvalue= 0.3788209
## lag= 16 pvalue= 0.3911332
## lag= 17 pvalue= 0.4540213
## lag= 18 pvalue= 0.3881403
## lag= 19 pvalue= 0.4577669
## lag= 20 pvalue= 0.4987703
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar3         ma1
## coef_est -0.09502221 -0.81107776
## se_est    0.05675731  0.03901438
## tstat     1.67418441 20.78919808
## pval      0.09409436  0.00000000
##
## Significativité au seuil de 95%
##   ar3   ma1
## FALSE  TRUE
## => Modèle rejeté : au moins l'un des coefficients des ordres
##                   les plus élevés n'est pas signififcatif au seuil de 95%
```

```r
arimafit(xpdc,2,0,1) #V
```

```
## Try ARIMA( 2 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 4 pvalue= 0.1055642
## lag= 5 pvalue= 0.2686774
## lag= 6 pvalue= 0.4080259
## lag= 7 pvalue= 0.3459798
## lag= 8 pvalue= 0.4822843
## lag= 9 pvalue= 0.538425
## lag= 10 pvalue= 0.3714183
## lag= 11 pvalue= 0.368074
## lag= 12 pvalue= 0.4652445
## lag= 13 pvalue= 0.4727374
## lag= 14 pvalue= 0.4234879
## lag= 15 pvalue= 0.3756504
## lag= 16 pvalue= 0.355401
## lag= 17 pvalue= 0.4116334
## lag= 18 pvalue= 0.3650654
## lag= 19 pvalue= 0.4275402
## lag= 20 pvalue= 0.4739505
## => Modèle valide
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar2         ma1
## coef_est -0.12880707 -0.83886754
## se_est    0.05465091  0.02958735
## tstat     2.35690641 28.35223552
```

```
## pval      0.01842789  0.00000000
##
## Significativité au seuil de 95%
##  ar2  ma1
## TRUE TRUE
## => Modèle bien ajusté
```

```
arimafit(xpdc,1,0,1)
```

```
## Try ARIMA( 1 , 0 , 1 )
##
## 1/ Test d'absence absence d'autocorrélation des résidus [Ljung-Box] :
## lag= 3 pvalue= 0.0128621
## lag= 4 pvalue= 0.03264646
## lag= 5 pvalue= 0.07430691
## lag= 6 pvalue= 0.1362592
## lag= 7 pvalue= 0.09895471
## lag= 8 pvalue= 0.1580118
## lag= 9 pvalue= 0.1911572
## lag= 10 pvalue= 0.1102369
## lag= 11 pvalue= 0.08642385
## lag= 12 pvalue= 0.1252245
## lag= 13 pvalue= 0.1262013
## lag= 14 pvalue= 0.1138645
## lag= 15 pvalue= 0.08829608
## lag= 16 pvalue= 0.09389067
## lag= 17 pvalue= 0.1191042
## lag= 18 pvalue= 0.1009573
## lag= 19 pvalue= 0.1310933
## lag= 20 pvalue= 0.1498639
## => Modèle non valide : l'absence d'autocorrélation
##                       des résidus est rejetée à 95%.
##
## 2/ Test de nullité des coefficients des des ordres les plus élevés :
##                  ar1         ma1
## coef_est 0.162700599 -0.86485771
## se_est   0.056646934  0.02326619
## tstat    2.872187223 37.17230359
## pval     0.004076414  0.00000000
##
## Significativité au seuil de 95%
##  ar1  ma1
## TRUE TRUE
## => Modèle bien ajusté
```

Seul deux modèles sont bien ajustés et valides.

```
ar2ma2<-arima(xpdc,order=c(2,0,2),include.mean=FALSE)
ar2ma1<-arima(xpdc,order=c(2,0,1),include.mean=FALSE)
```

On implémente un critère d'information pour choisir un modèle.

```r
#Critères d'informations pour choisir le modèle.
minInfo<-function(){
  l<-c("ar2ma2","ar2ma1")
  aic_choix<-NA
  bic_choix<-NA
  minAIC<-Inf
  minBIC<-Inf
  for (mod in l){
    cat(mod,"|","AIC :",AIC(get(mod)),"|","BIC :",BIC(get(mod)),"\n")
    if (AIC(get(mod))<minAIC){
      aic_choix<-mod
      minAIC<-AIC(get(mod))
    } else NA
    if (BIC(get(mod))<minBIC){
      bic_choix<-mod
      minBIC<-BIC(get(mod))
    } else NA
  }
  cat("\n")
  cat("=> Choix AIC :",aic_choix,"\n")
  cat("=> Choix BIC :",bic_choix,"\n")
}
minInfo()
```

```
## ar2ma2 | AIC : 2562.248 | BIC : 2582.014
## ar2ma1 | AIC : 2564.625 | BIC : 2580.438
##
## => Choix AIC : ar2ma2
## => Choix BIC : ar2ma1
```

Malheureusement les critères AIC et BIC ne s'accordent pas sur le modèel à choisir. On décide alors de garder celui qui donne la meilleure prévision sur l'échantillon.

```r
#calcul du R^2 ajusté
r2_aj<-function(model){
  p=model$arma[1]
  q=model$arma[2]
  ss_res <- sum(model$residuals^2) #somme des résidus au carré
  ss_tot <- sum(xpdc[-c(1:max(p,q))]^2)
            #somme des observations de l’´echantillon au carré
  n <- model$nobs-max(p,q) #taille de l’échantillon
  adj_r2 <- 1-(ss_res/(n-p-q-1))/(ss_tot/(n-1)) #r2 ajusté
  cat("R^2 ajusté : ",adj_r2,"\n")
  cat("\n")
}
r2_aj(ar2ma1)
```

```
## R^2 ajusté :  0.3414002
```

```r
r2_aj(ar2ma2)
```

```
## R^2 ajusté :  0.3474559
```

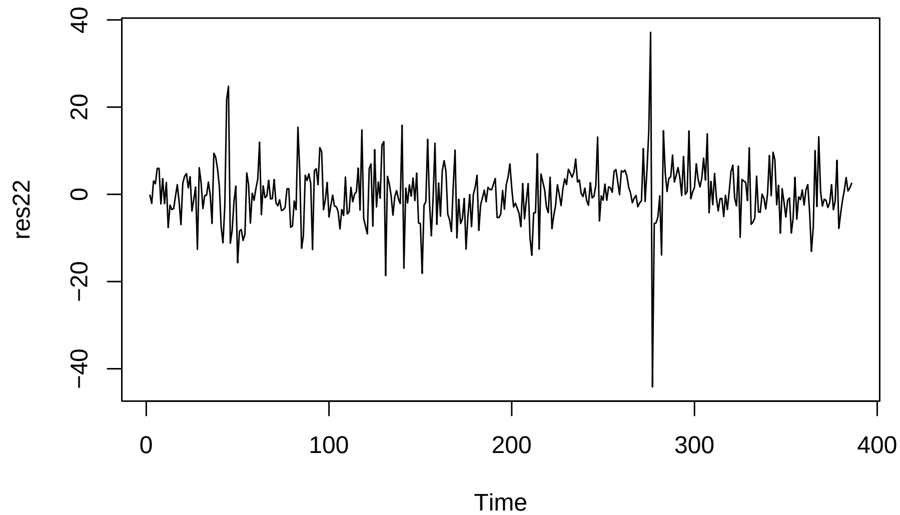
On choisit le modèle avec le R^2 ajusté le plus élevé : choix du modèle ar2ma2.

```r
verifit<-function(serie,p,d,q){
  var<-arima(serie,order=c(p,d,q),include.mean=FALSE)
  coef_est<-var$coef
  se_est<-sqrt(diag(var$var.coef))
  tstat<-abs(coef_est/se_est)
  pvalue<-c()
  #sign<-(tstat>1.96)==1  #on rejette l'hypothèse nulle du test de Student
  for (coef in tstat){
    pvalue<-append(pvalue,pval <- (1-pnorm(abs(coef)))*2)
    #calcul de la pvalue pour le test de student (test bilatéral)
  }
  cat("Modèle estimé \n")
  cat("\n")
  print(rbind(coef_est,se_est,tstat,pvalue))
}
verifit(xpdc,2,0,2) #Tous les coefficients sont significatifs au seuil de 95%
```

Q.5

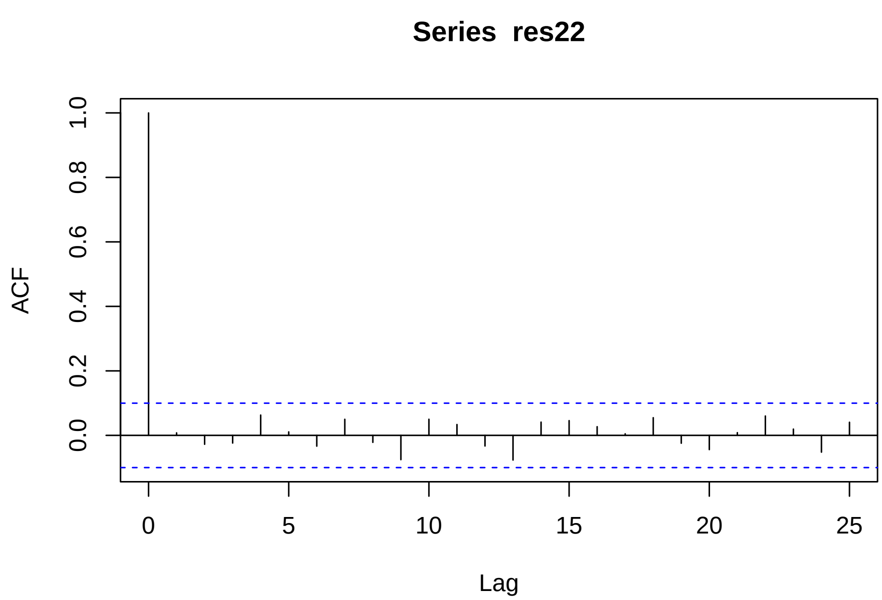
```
## Modèle estimé
##
##                   ar1           ar2        ma1          ma2
## coef_est 1.022184e+00 -2.385687e-01 -1.7346691 7.768956e-01
## se_est   1.519028e-01  5.959678e-02  0.1527955 1.265533e-01
## tstat    6.729198e+00  4.003047e+00 11.3528795 6.138883e+00
## pvalue   1.706013e-11  6.253197e-05  0.0000000 8.310368e-10
```

```r
#résidus du modèle estimé
res22<-ar2ma2$residuals
plot(res22)
```

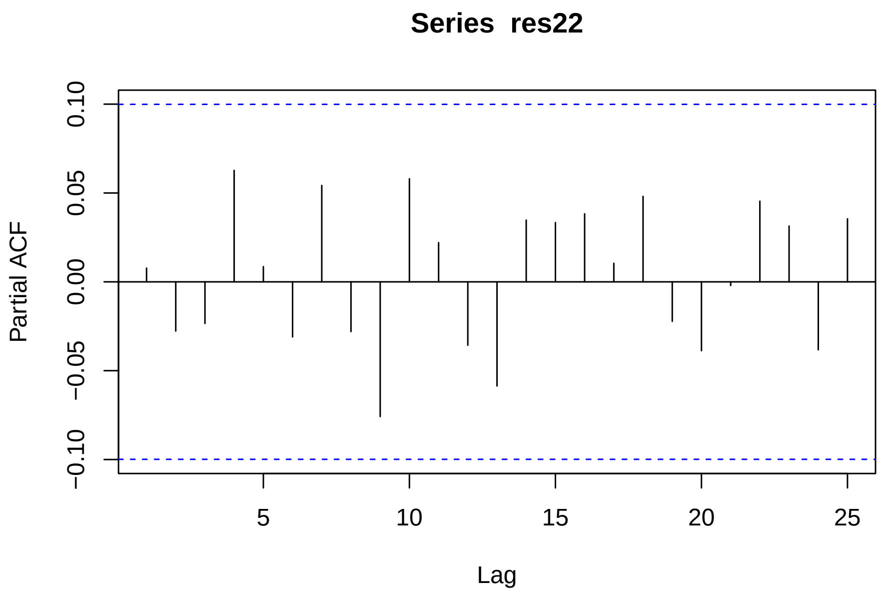


Q.6 & 7

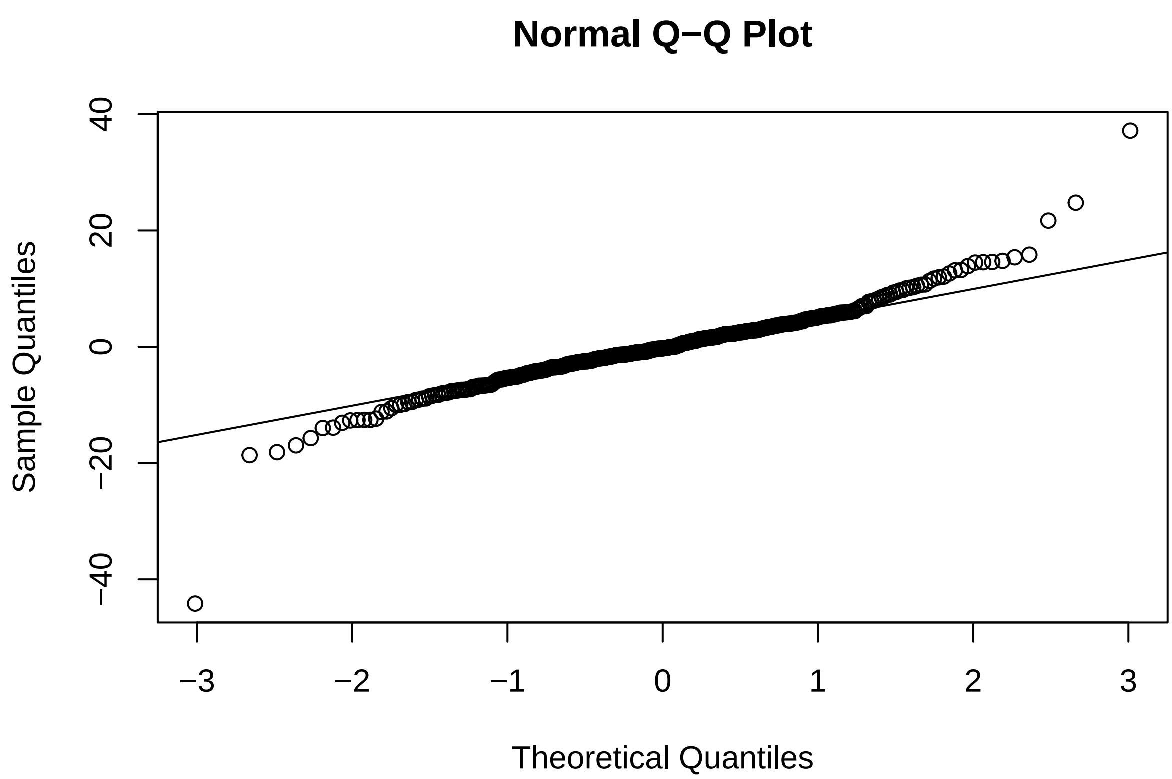
```
acf(res22) #autocorrélations totales
```




```
pacf(res22) #autocorrélations partielles
```




```
#Test de normalité des résidus
qqnorm(res22)
qqline(res22)
```



Q.8 Question ouverte :

```
#Prédictions et représentation graphique
pred<-predict(ar2ma2,2)$pred
eps<-mean((res22)^2)-mean(res22)^2

plot(xpd,xlim=c(350,390),ylim=c(-20,20),xlab="Obs")

#IC à 95% sur X(T+1)
rect(xleft = 385.5,xright = 386.5,
     ybottom = as.numeric(pred[1])-sqrt(eps)*qnorm(1-0.05/2,mean=0,sd=1),
     ytop = as.numeric(pred[1])+sqrt(eps)*qnorm(1-0.05/2,mean=0,sd=1),col="gray"
     ,border=NA )
points(x=386,y=as.numeric(pred[1]),col="blue",pch=16)

#IC à 95% sur X(T+2)
rect(xleft = 386.5,xright = 387.5,
     ybottom = as.numeric(pred[2])-sqrt(eps)*qnorm(1-0.05/2,mean=0,sd=1),
     ytop = as.numeric(pred[2])+sqrt(eps)*qnorm(1-0.05/2,mean=0,sd=1),col="gray"
     ,border=NA )
points(x=387,y=as.numeric(pred[2]),col="blue",pch=16)
```