

Analyse des variables du jeu de données

Jules Corbel & Paul Guillotte

12 décembre 2018

Contents

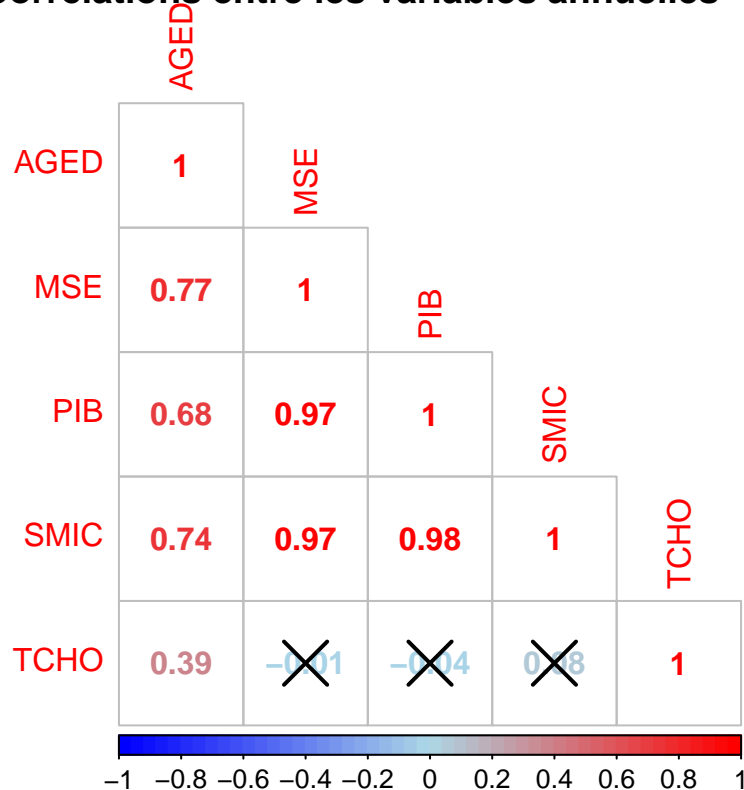
Ensemble des variables annuelles	3
Masse salariale	3
PIB	7
SMIC	10
Taux de chômage	12
Allocations Garde Enfant à Domicile (AGED)	16
Ensemble des variables trimestrielles	19
MSE	19
SMIC	26
Taux de chômage des femmes	30

Il y a deux ensembles de séries de données : les séries annuelles et les séries trimestrielles. Dans un premier temps, nous nous intéressons aux corrélations entre les variables deux à deux dans ces deux ensembles afin de nous faire une première idée du lien qu'il existe entre les variables.

```
annuelle <- read.csv("~/PFE_Time_Series/Data/Data_Annuel.csv", sep=";", dec=",", nrow=30)
trim <- read.csv("~/PFE_Time_Series/Data/Data_Trim.csv", sep=";", dec=",")

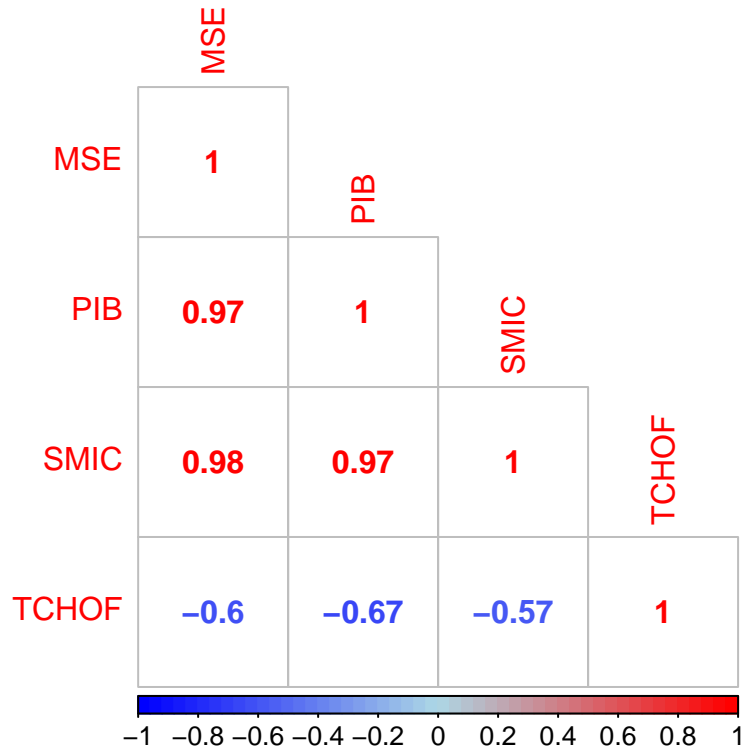
corrplot(cor(annuelle[1:28,-1]), method = "number", type="lower",
  p.mat=cor.mtest(annuelle[1:28,-1], 0.95)[[1]], insig="pch",
  col=colorRampPalette(c("blue", "light blue", "red"))(50), title = "
  Corrélations entre les variables annuelles")
```

Corrélations entre les variables annuelles



```
corrplot(cor(trim[1:109,-1]), method = "number", type="lower",
  p.mat=cor.mtest(trim[1:109,-1], 0.95)[[1]], insig="pch",
  col=colorRampPalette(c("blue", "light blue", "red"))(50), title = "
  Corrélations entre les variables trimestrielles")
```

Corrélations entre les variables trimestrielles



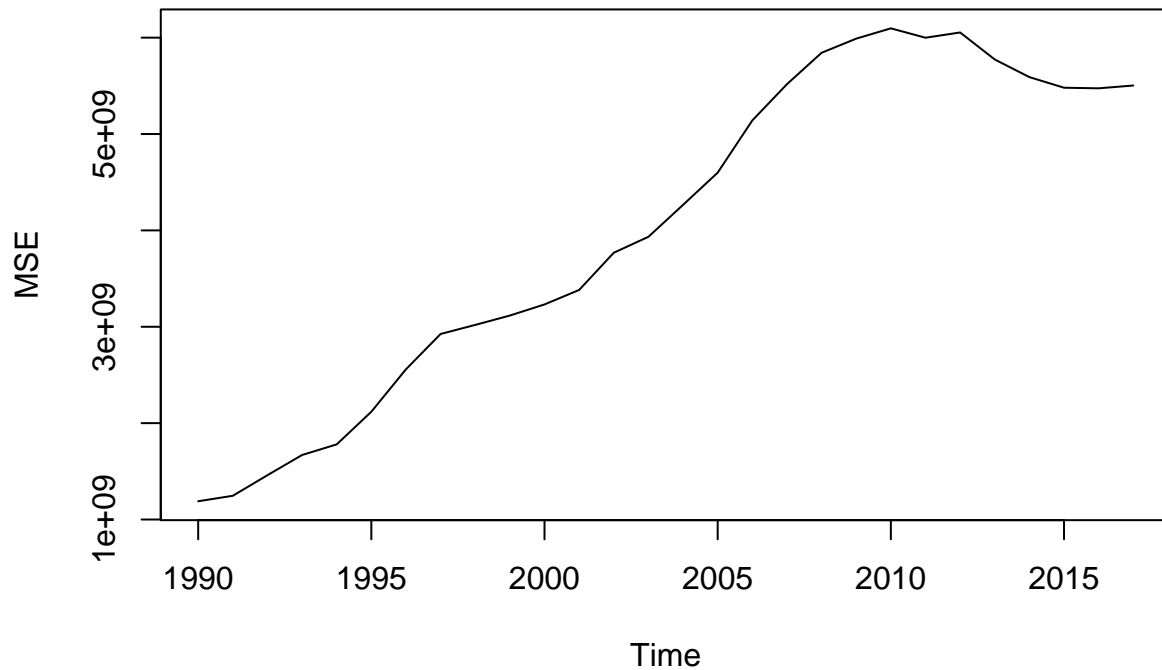
On se rend compte que dans l'ensemble annuel, le taux de chômage est faiblement corrélé avec la variable AGED et non corrélé avec les autres variables. La variable AGED est très liée avec la MSE (variable à expliquer), le PIB et le SMIC. Enfin, les trois variables du trio PIB, MSE et SMIC sont extrêmement corrélées entre elles. Pour ce qui concerne l'ensemble trimestriel, le taux de chômage est corrélé négativement avec toutes les autres variables. Une nouvelle fois, les trois variables du trio PIB, MSE et SMIC sont extrêmement liées entre elles.

Ensemble des variables annuelles

Masse salariale

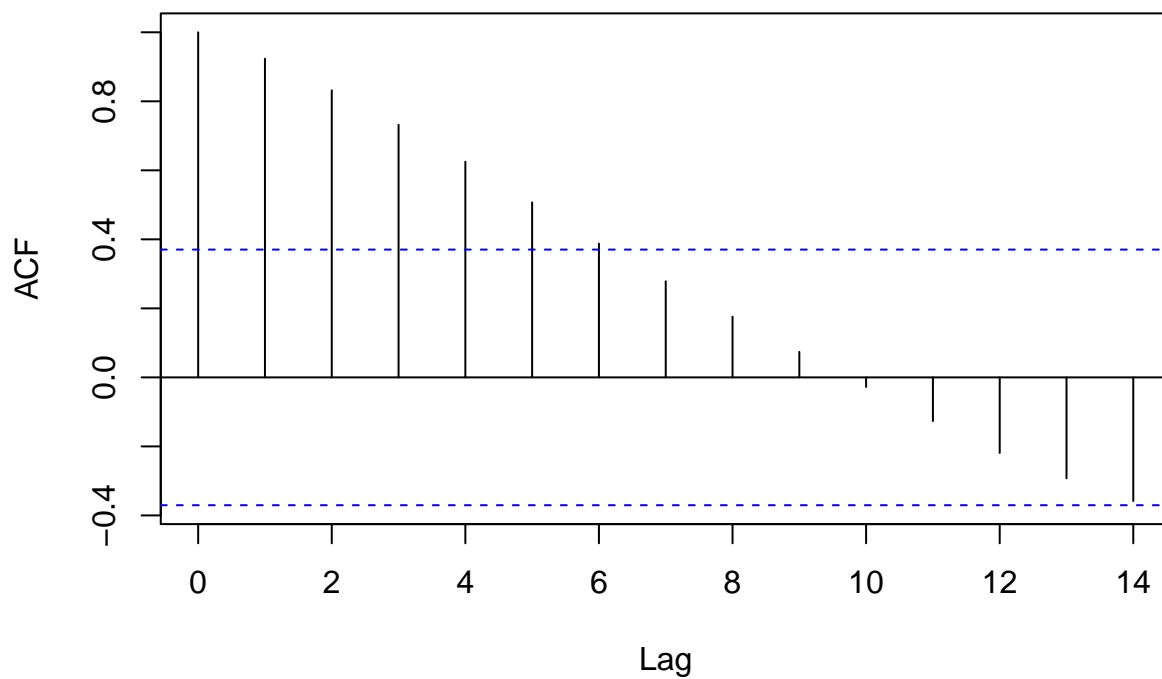
```
MSE <- ts(annuelle$MSE, start = 1990, end = c(2017, 1))
plot(MSE, main="Evolution de la masse salariale annuelle")
```

Evolution de la masse salariale annuelle



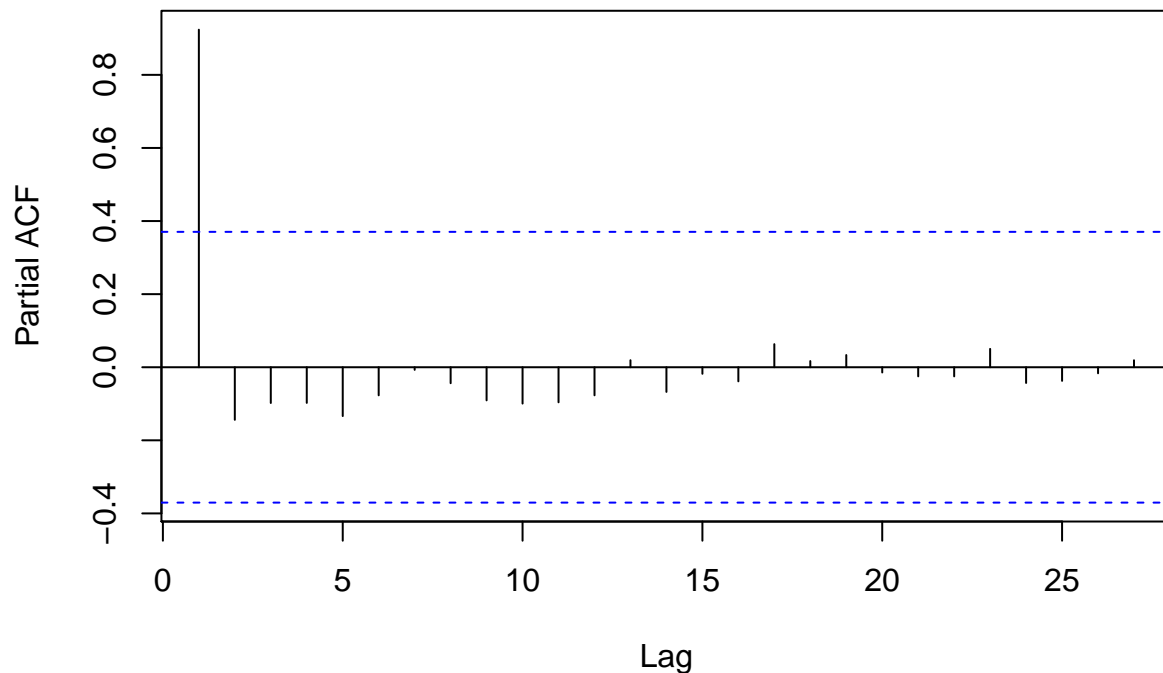
```
acf(MSE, main="Auto-corrélation de la masse salariale annuelle")
```

Auto-corrélation de la masse salariale annuelle



```
pacf(MSE, "Autocorrélation partielle de la masse salariale")
```

Series MSE



```
adfTest(MSE, type='ct')
```

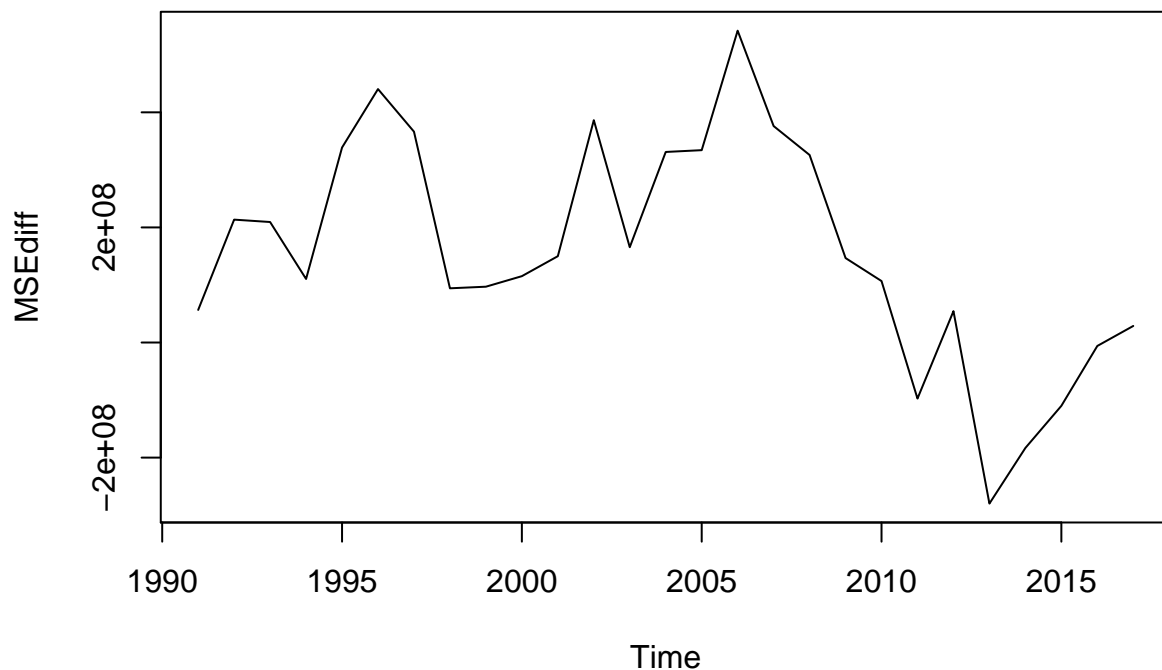
```
##
## Title:
##   Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 1
##   STATISTIC:
##     Dickey-Fuller: -1.1671
##   P VALUE:
##     0.8912
##
## Description:
##   Thu Dec 13 10:18:30 2018 by user:
```

La masse salariale annuelle possède une composante de tendance de 1990 à 2010. La série tend par la suite à stagner et même à s'inverser à partir de 2013. Nous ne voyons pas de saisonnalité pour cette série.

Comme la série comporte une tendance, elle ne correspond pas à la première condition de la stationnarité du second ordre, soit que la série possède une moyenne constante. Cela est confirmé par la fonction ACF qui décroît régulièrement. Nous effectuons également un test de Dickey-Fuller augmenté servant à vérifier si la série est stationnaire ou non. Les hypothèses de ce test sont les suivantes H_0 : la série possède une racine unitaire i.e la série n'est pas stationnaire H_1 : la série est stationnaire La p-value vaut 0.82 ce qui nous confirme que la série n'est pas stationnaire avec un risque de première espèce de 5%.

Afin de la stationnariser, nous effectuons donc dans un premier temps une différenciation d'ordre 1.

```
MSEdiff<-diff(MSE)
plot(MSEdiff)
```



```
kpss.test(MSEdiff)
```

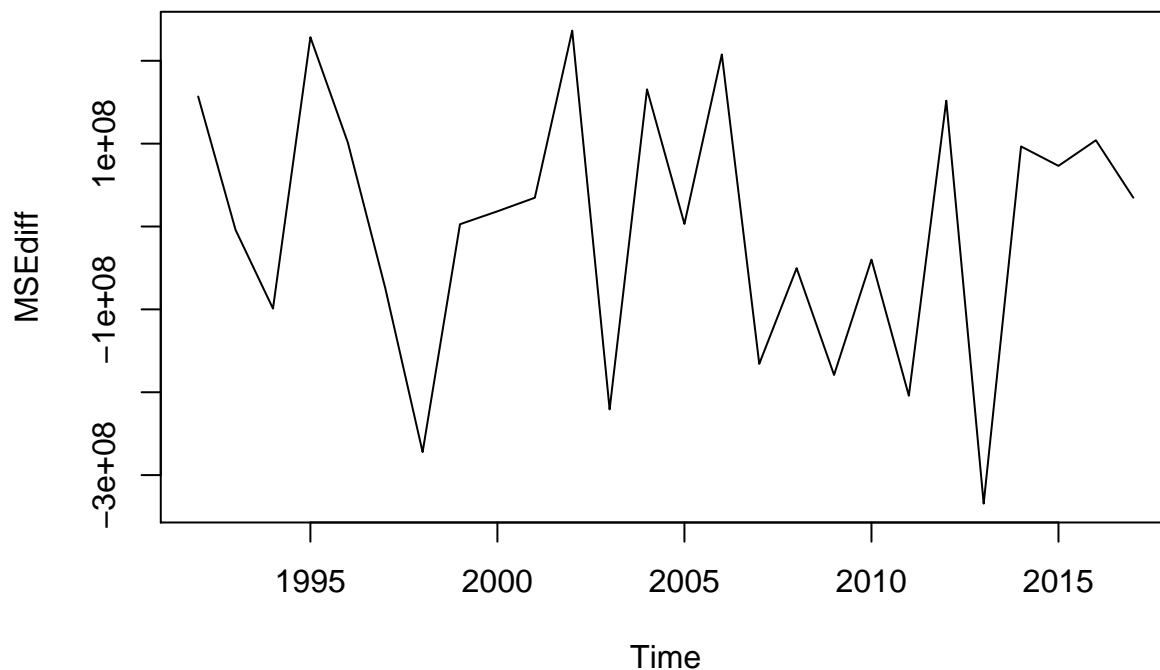
```
##
## KPSS Test for Level Stationarity
##
## data: MSEdiff
## KPSS Level = 0.49629, Truncation lag parameter = 1, p-value =
## 0.0425
```

```
adfTest(MSEdiff, type='c')
```

```
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -1.4882
## P VALUE:
## 0.5042
##
## Description:
## Thu Dec 13 10:18:31 2018 by user:
```

Le graphique ci-dessus ne semble pas représenter une série stationnaire. Cela est confirmé par la réalisation d'un test KPSS, dont l'hypothèse nulle est que la série est stationnaire ainsi que d'un test ADF. Avec un risque de 5%, ces deux tests nous permettent de conclure que la série n'est pas stationnaire. Nous choisissons donc d'effectuer une différenciation d'ordre 2. Cette fois-ci, les tests nous permettent de voir que la série obtenue est stationnaire.

```
MSEdiff<-diff(MSE, differences=2)
plot(MSEdiff)
```



```
kpss.test(MSEdiff)
```

```
## Warning in kpss.test(MSEdiff): p-value greater than printed p-value
```

```
##
```

```
## KPSS Test for Level Stationarity
```

```
##
```

```
## data: MSEdiff
```

```
## KPSS Level = 0.1142, Truncation lag parameter = 1, p-value = 0.1
```

```
adfTest(MSEdiff, type='c')
```

```
##
```

```
## Title:
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## Test Results:
```

```
## PARAMETER:
```

```
## Lag Order: 1
```

```
## STATISTIC:
```

```
## Dickey-Fuller: -3.2508
```

```
## P VALUE:
```

```
## 0.031
```

```
##
```

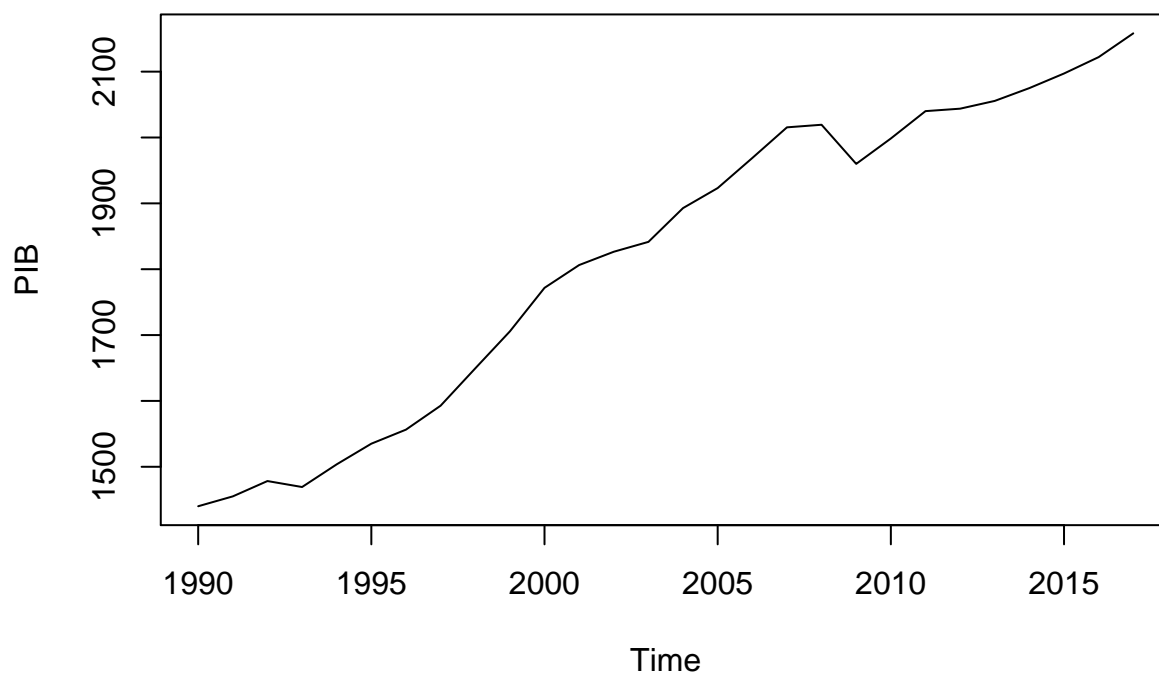
```
## Description:
```

```
## Thu Dec 13 10:18:31 2018 by user:
```

PIB

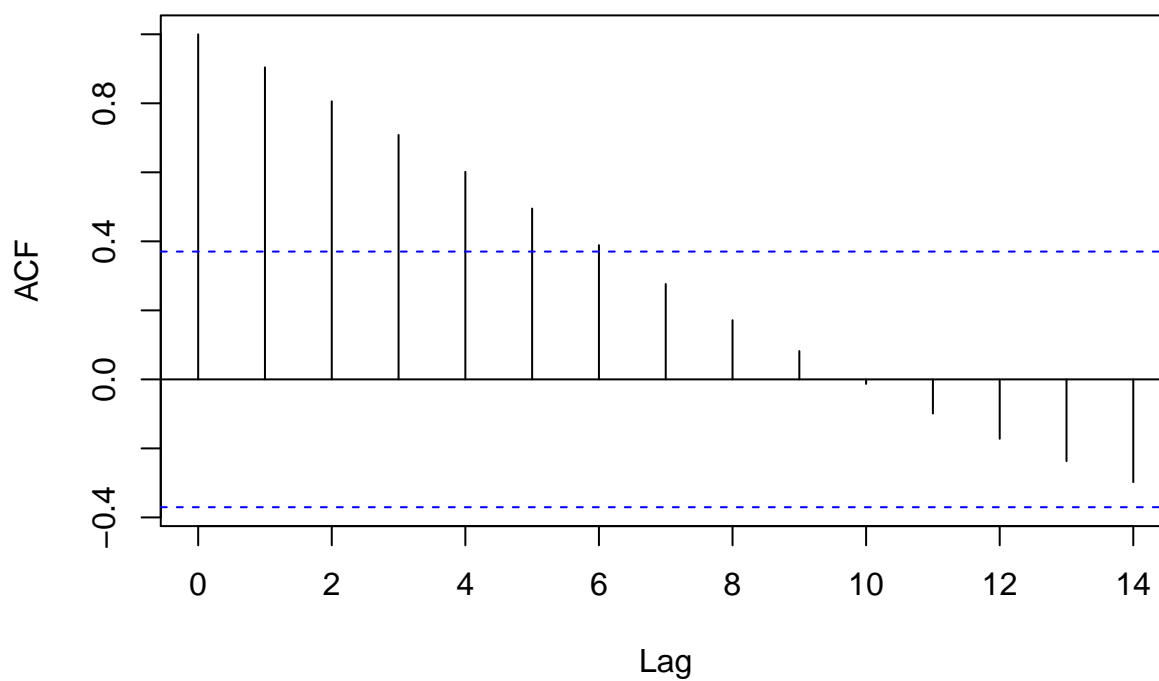
```
PIB <- ts(annuelle$PIB, start = 1990, end = 2017)
plot(PIB, main="Evolution du PIB annuel")
```

Evolution du PIB annuel



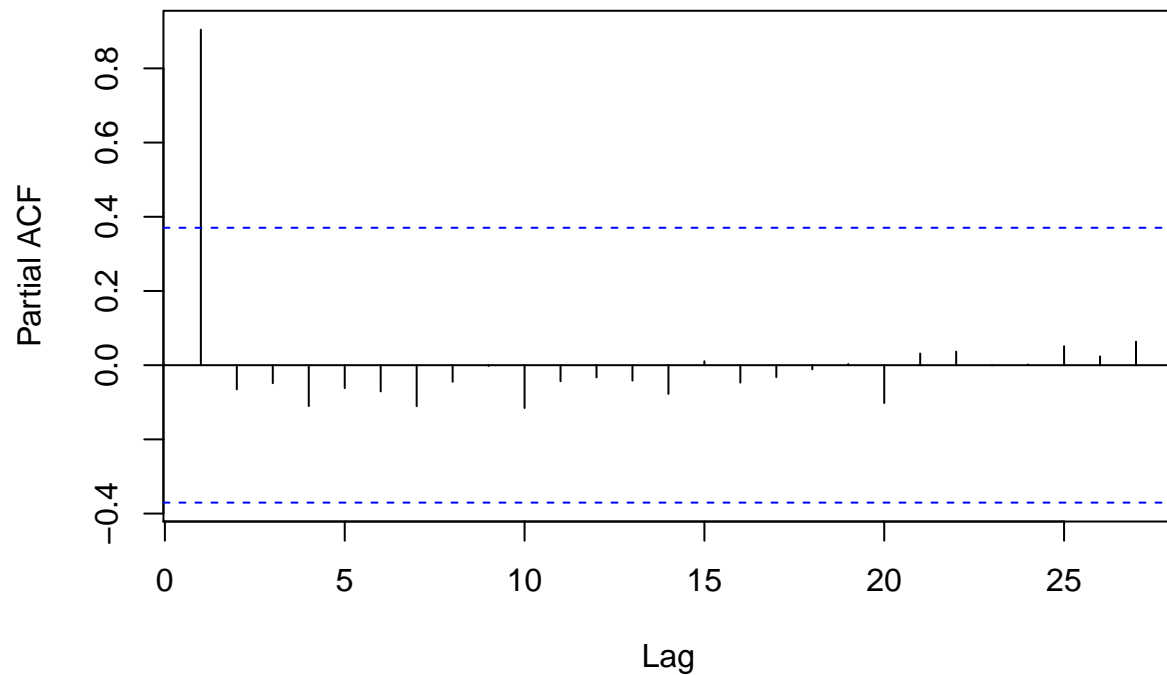
```
acf(PIB, main="Auto-corrélation du PIB annuel")
```

Auto-corrélation du PIB annuel



```
pacf(PIB, "Autocorrélation partielle du PIB annuel")
```


Series PIB

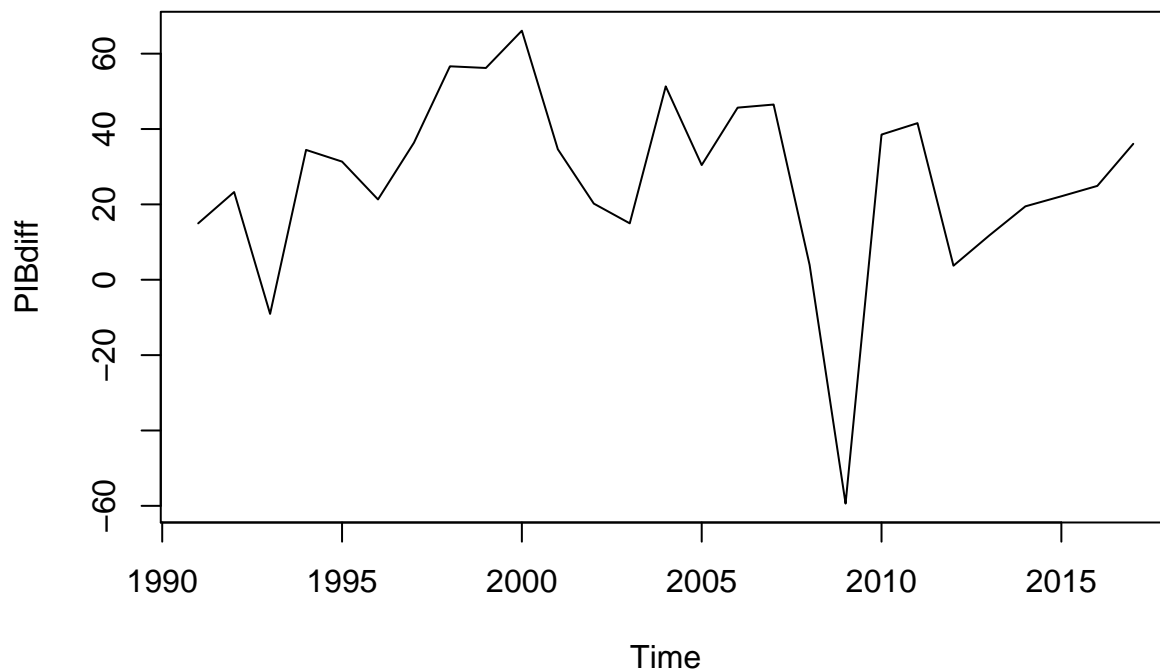


```
adfTest(PIB, type='c')
```

```
##
## Title:
##   Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 1
##   STATISTIC:
##     Dickey-Fuller: -0.9377
##   P VALUE:
##     0.6995
##
## Description:
##   Thu Dec 13 10:18:32 2018 by user:
```

Comme pour la masse salariale, le PIB annuel possède une tendance mais pas de saisonnalité. Cette série n'est donc pas non plus stationnaire. Cependant, il semble cette fois-ci que la tendance soit linéaire, donc une différenciation d'ordre 1 pourrait suffire à stationnariser la série.

```
PIBdiff<-diff(PIB)
plot(PIBdiff)
```



```
kpss.test(PIBdiff)
```

```
## Warning in kpss.test(PIBdiff): p-value greater than printed p-value
```

```
##
```

```
## KPSS Test for Level Stationarity
```

```
##
```

```
## data: PIBdiff
```

```
## KPSS Level = 0.1531, Truncation lag parameter = 1, p-value = 0.1
```

```
adfTest(PIBdiff, type='c')
```

```
##
```

```
## Title:
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## Test Results:
```

```
## PARAMETER:
```

```
## Lag Order: 1
```

```
## STATISTIC:
```

```
## Dickey-Fuller: -3.5859
```

```
## P VALUE:
```

```
## 0.01565
```

```
##
```

```
## Description:
```

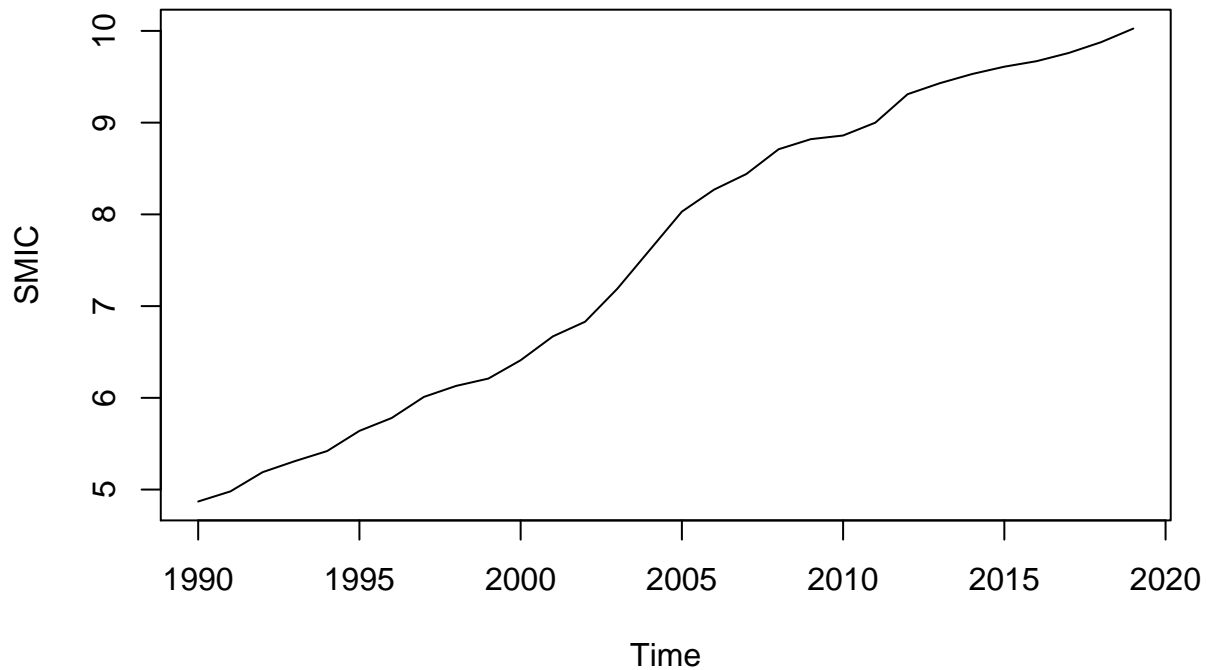
```
## Thu Dec 13 10:18:33 2018 by user:
```

En prenant un risque de première espèce de 5%, le test KPSS et le test ADF nous indiquent que la série différenciée est bien stationnaire.

SMIC

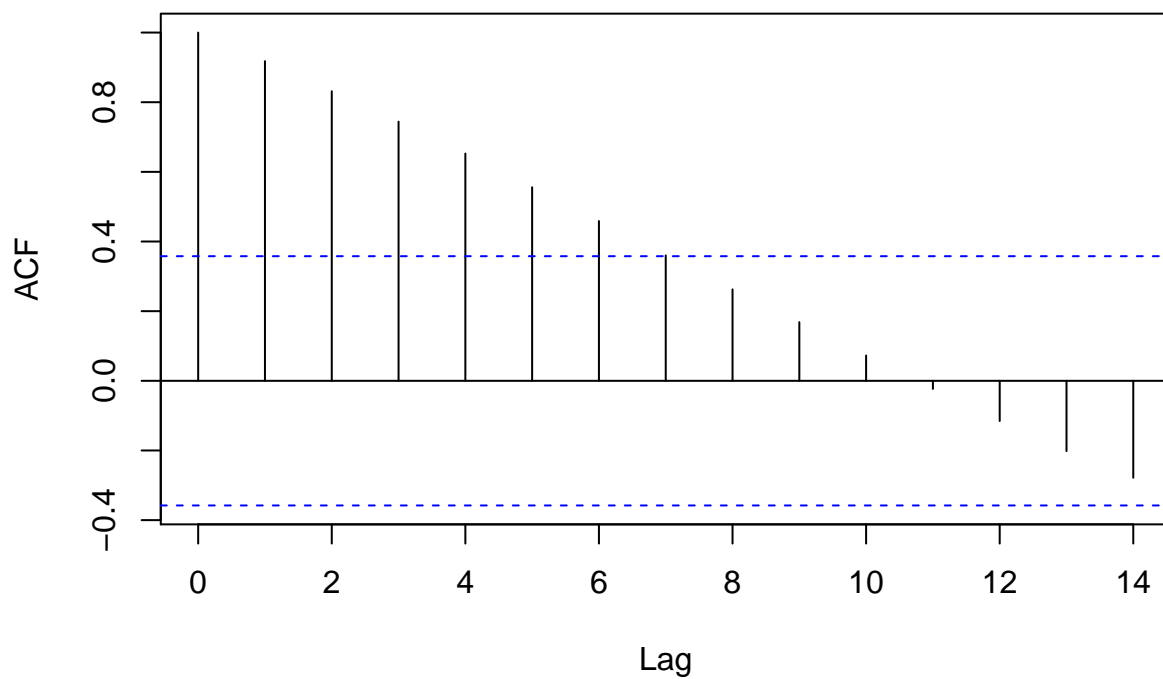
```
SMIC <- ts(annuelle$SMIC, start = 1990, end = 2019)
plot(SMIC, main="Evolution du SMIC annuel")
```

Evolution du SMIC annuel



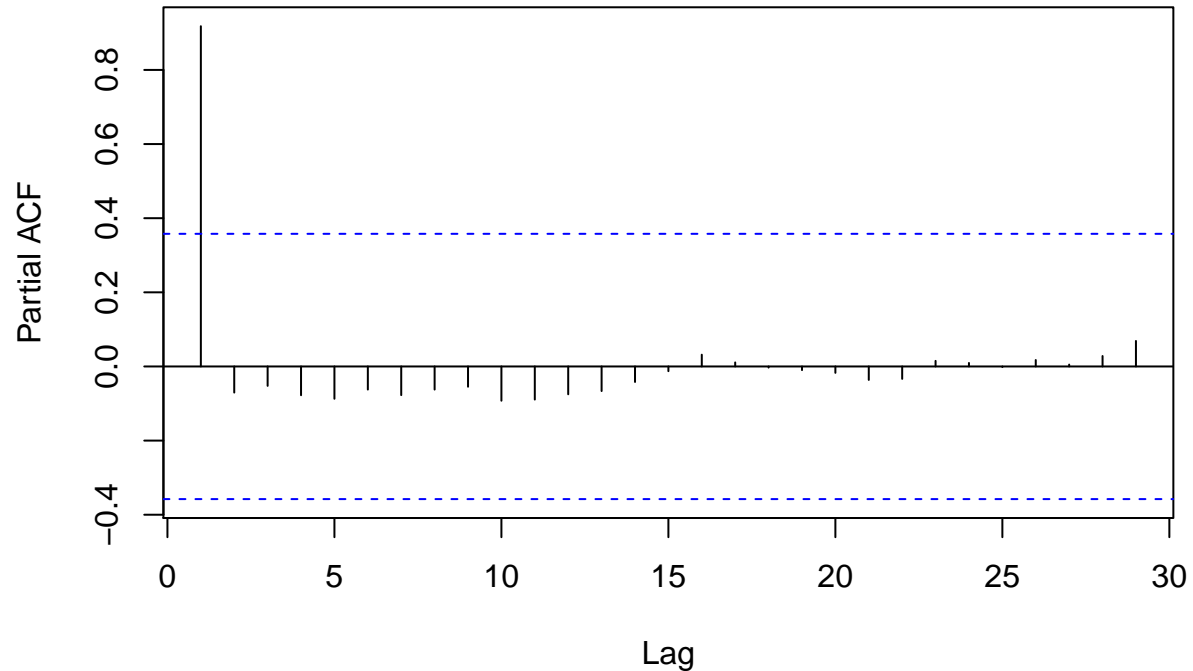
```
acf(SMIC, main="Auto-corrélation du SMIC annuel")
```

Auto-corrélation du SMIC annuel



```
pacf(SMIC, "Autocorrélation partielle du SMIC annuel")
```

Series SMIC



```
adfTest(SMIC, type='c')
```

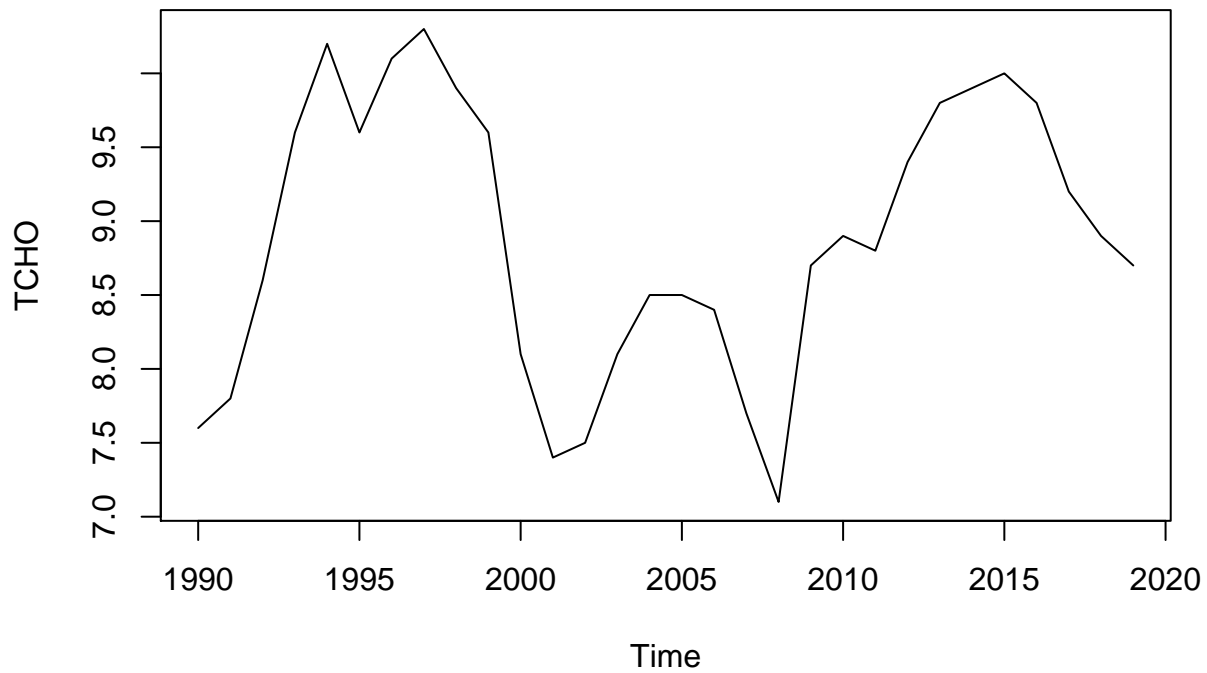
```
##
## Title:
##   Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 1
##   STATISTIC:
##     Dickey-Fuller: -1.0369
##   P VALUE:
##     0.6646
##
## Description:
##   Thu Dec 13 10:18:33 2018 by user:
```

La série SMIC possède la même forme que celle du PIB. Nous avons utilisé la même méthodologie pour la rendre stationnaire.

Taux de chômage

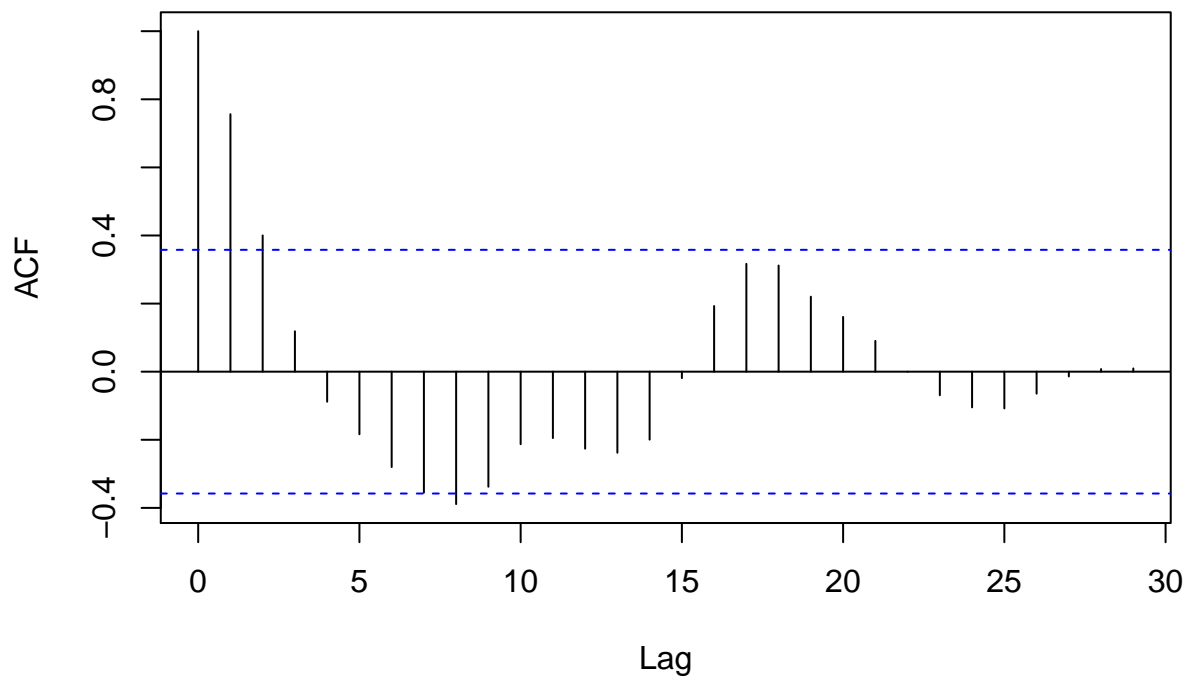
```
TCHO <- ts(annuelle$TCHO, start = 1990, end = 2019)
plot(TCHO, main="Evolution du taux de chômage annuel")
```

Evolution du taux de chômage annuel



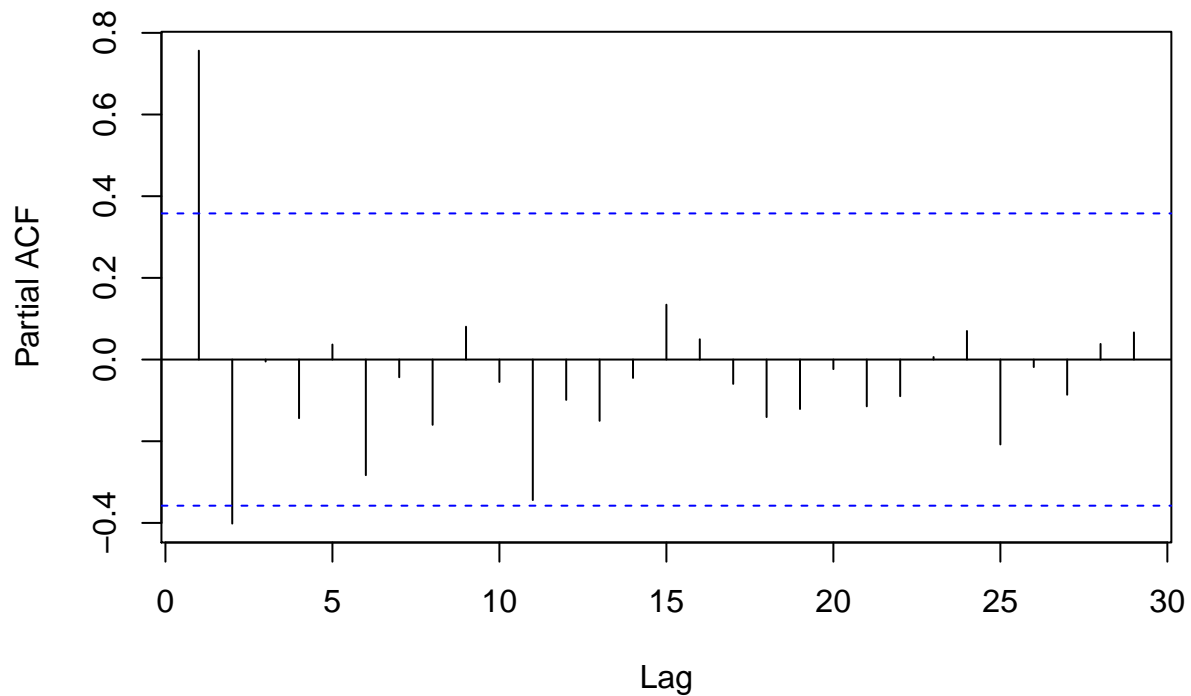
```
acf(TCHO, main="Auto-corrélation du taux de chômage annuel", lag=30)
```

Auto-corrélation du taux de chômage annuel



```
pacf(TCHO, "Autocorrélation partielle du taux de chômage annuel")
```

Series TCHO

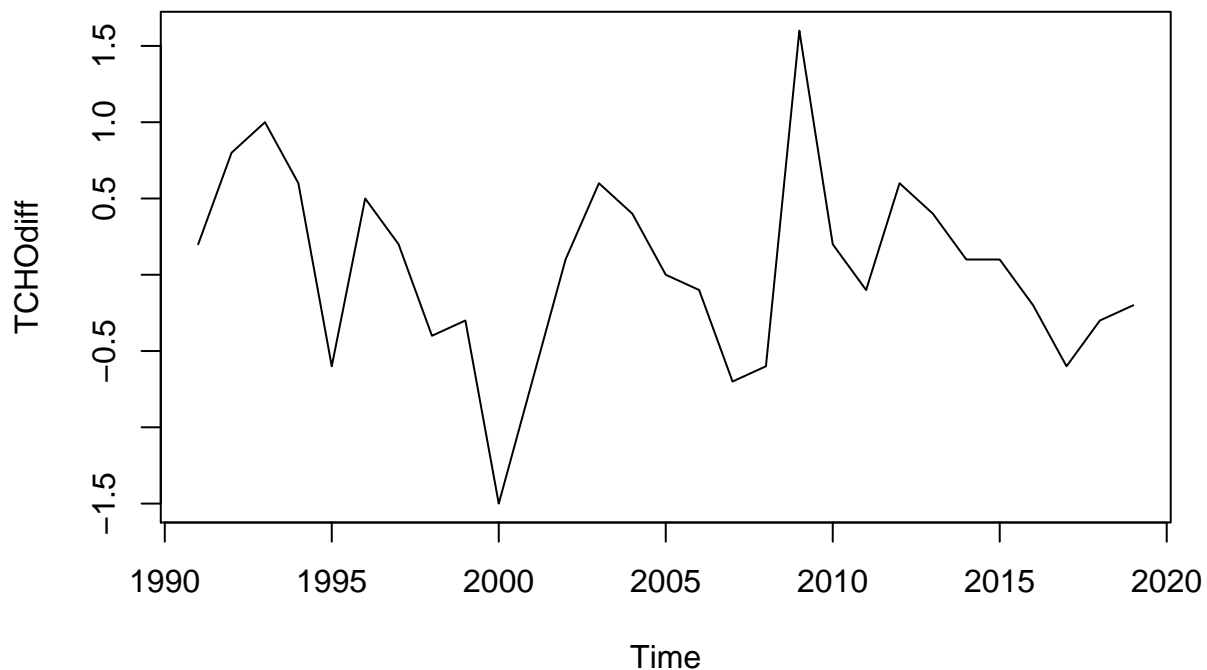


```
adfTest(TCHO, type='c')
```

```
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -2.8737
## P VALUE:
## 0.06582
##
## Description:
## Thu Dec 13 10:18:34 2018 by user:
```

Contrairement aux séries précédentes, le taux de chômage ne possède pas de tendance ou de saisonnalité clairement définie. Cependant, cela ne nous permet pas d'affirmer que la série est stationnaire. Le test ADF nous amène à conserver l'hypothèse de non-stationnarité. Nous construisons donc la série différenciée d'ordre 1.

```
TCHOdif<-diff(TCHO)
plot(TCHOdif)
```



```
kpss.test(TCH0diff)
```

```
## Warning in kpss.test(TCH0diff): p-value greater than printed p-value
```

```
##
```

```
## KPSS Test for Level Stationarity
```

```
##
```

```
## data: TCH0diff
```

```
## KPSS Level = 0.10821, Truncation lag parameter = 1, p-value = 0.1
```

```
adfTest(TCH0diff, type='c')
```

```
##
```

```
## Title:
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## Test Results:
```

```
## PARAMETER:
```

```
## Lag Order: 1
```

```
## STATISTIC:
```

```
## Dickey-Fuller: -3.6025
```

```
## P VALUE:
```

```
## 0.01462
```

```
##
```

```
## Description:
```

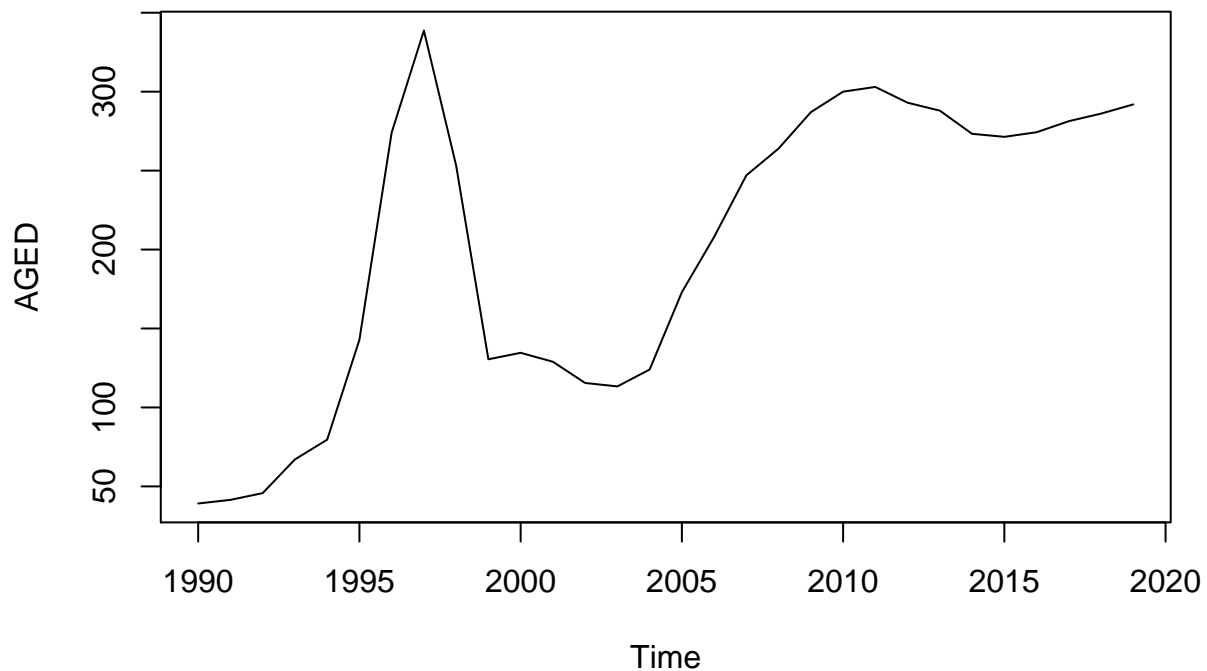
```
## Thu Dec 13 10:18:35 2018 by user:
```

Après la différenciation, la représentation graphique ainsi que les résultats des tests de stationnarité nous font conserver cette série.

Allocations Garde Enfant à Domicile (AGED)

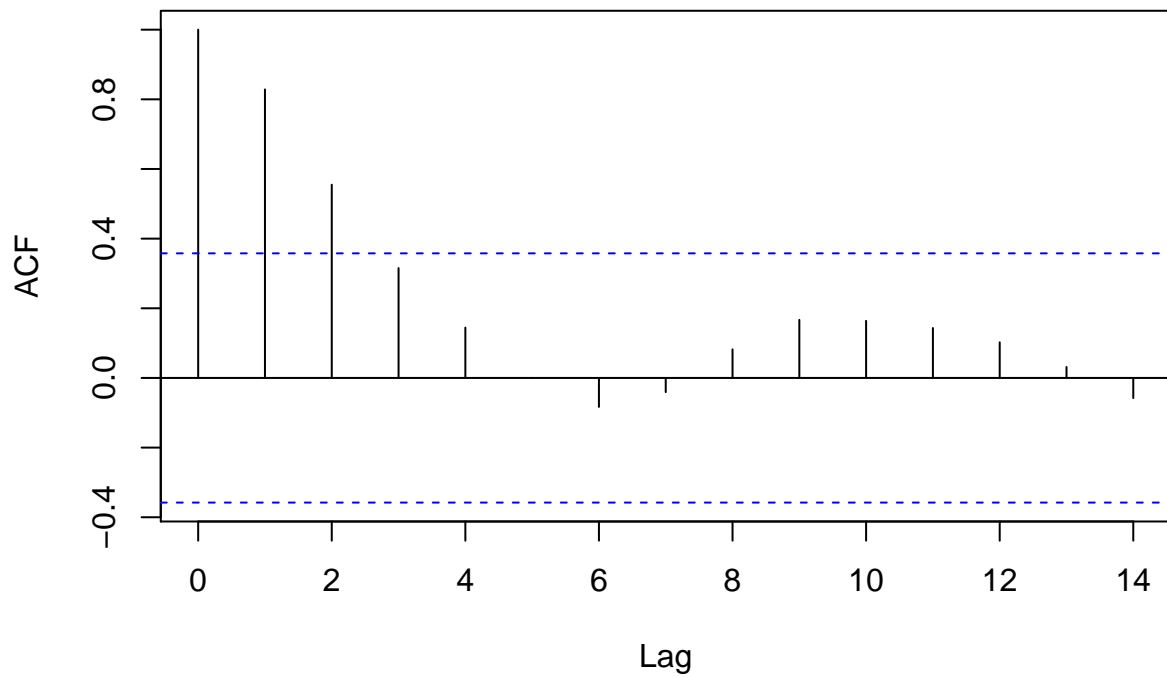
```
AGED <- ts(annuelle$AGED, start = 1990, end = 2019)
plot(AGED, main="Evolution de l'AGED annuelle")
```

Evolution de l'AGED annuelle



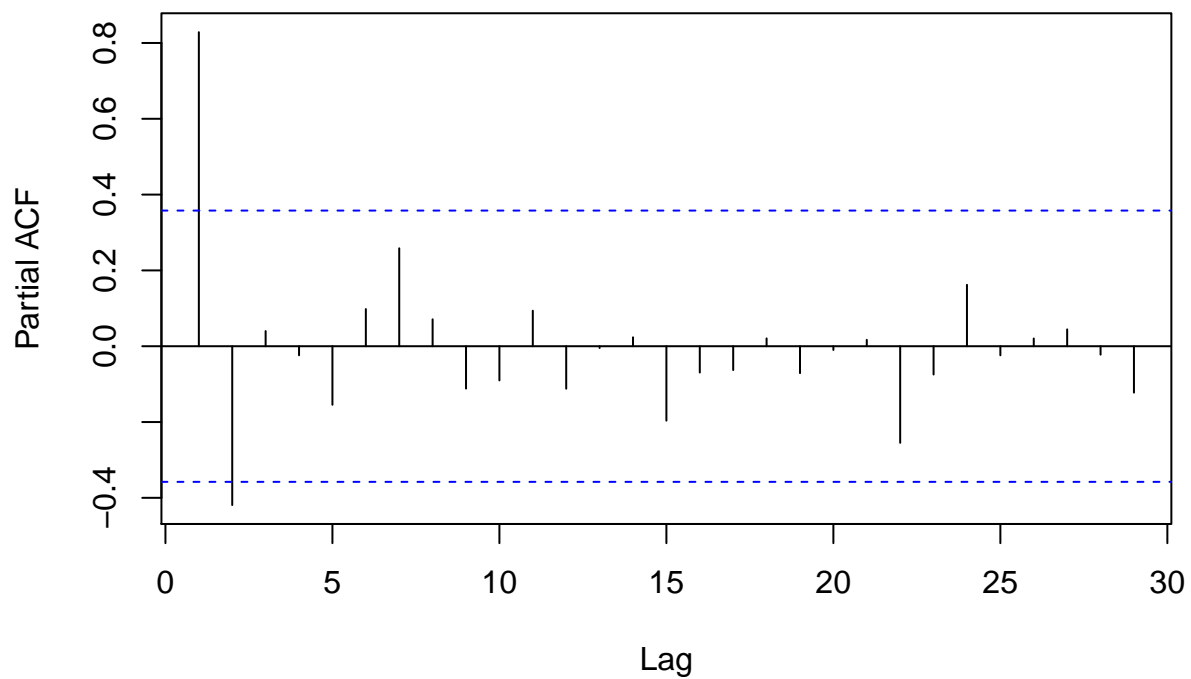
```
acf(AGED, main="Auto-corrélation de l'AGED annuelle")
```


Auto-corrélation de l'AGED annuelle



```
pacf(AGED, "Autocorrélation partielle de l'AGED annuelle")
```

Series AGED



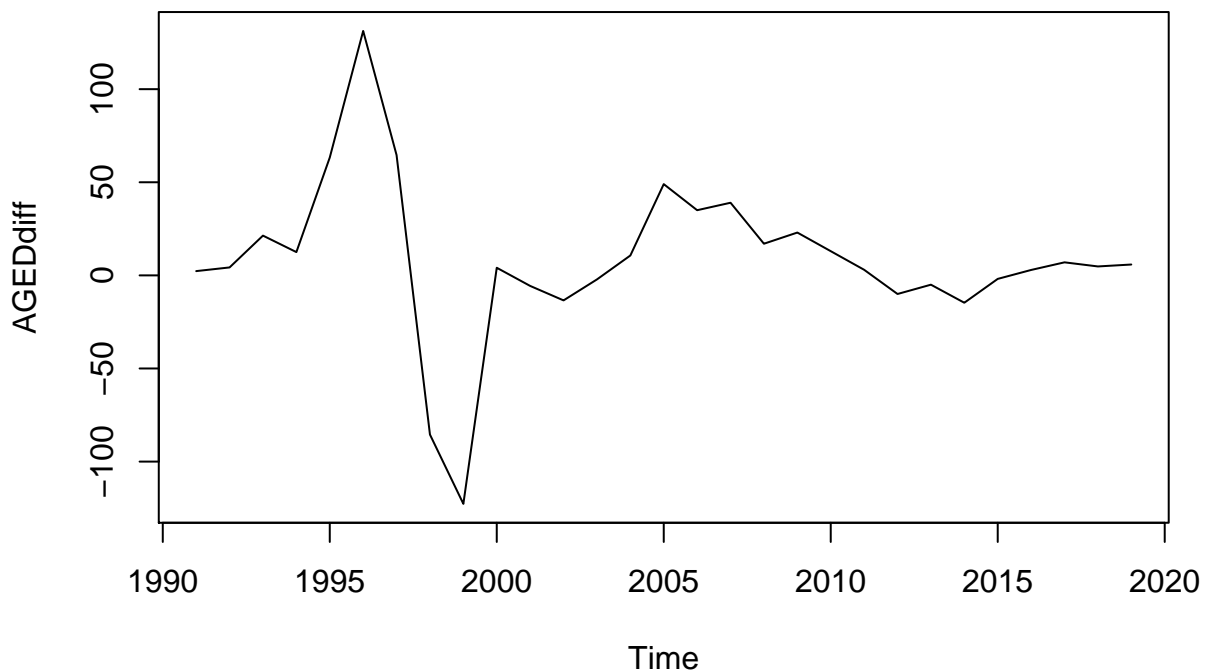
```
adfTest(AGED, type='c')
```

##

```
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -2.717
## P VALUE:
## 0.08738
##
## Description:
## Thu Dec 13 10:18:36 2018 by user:
```

Cette variable correspond aux Allocations des Gardes d'Enfants à Domicile. Comme pour le taux de chômage, il est difficile de visualiser une tendance ou une saisonnalité mais le test ADF nous indique que la série n'est pas stationnaire. La différenciation d'ordre 1 suffit pour stationnariser la série.

```
AGEDdiff<-diff(AGED)
plot(AGEDdiff)
```



```
kpss.test(AGEDdiff)
```

```
## Warning in kpss.test(AGEDdiff): p-value greater than printed p-value
##
## KPSS Test for Level Stationarity
##
## data: AGEDdiff
## KPSS Level = 0.068362, Truncation lag parameter = 1, p-value = 0.1
```

```
adfTest(AGEDdiff, type='c')
```

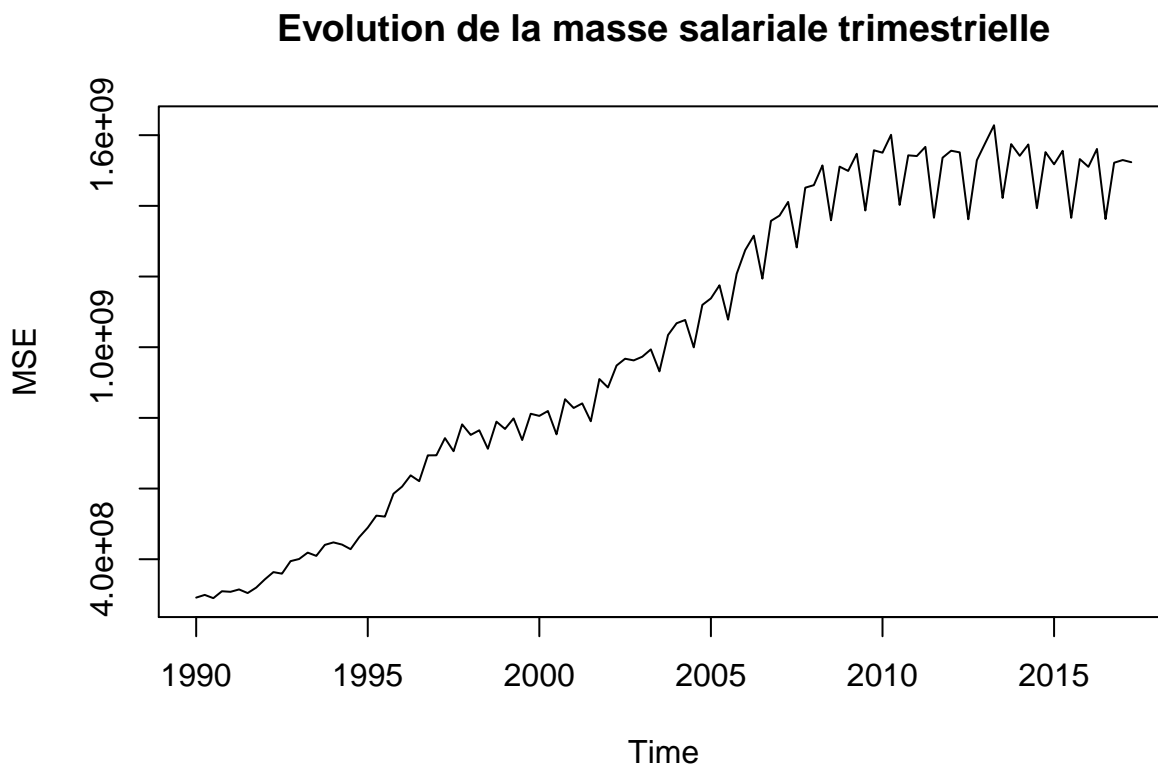
```
## Warning in adfTest(AGEDdiff, type = "c"): p-value smaller than printed p-
## value
```

```
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -4.5957
## P VALUE:
## 0.01
##
## Description:
## Thu Dec 13 10:18:37 2018 by user:
```

Ensemble des variables trimestrielles

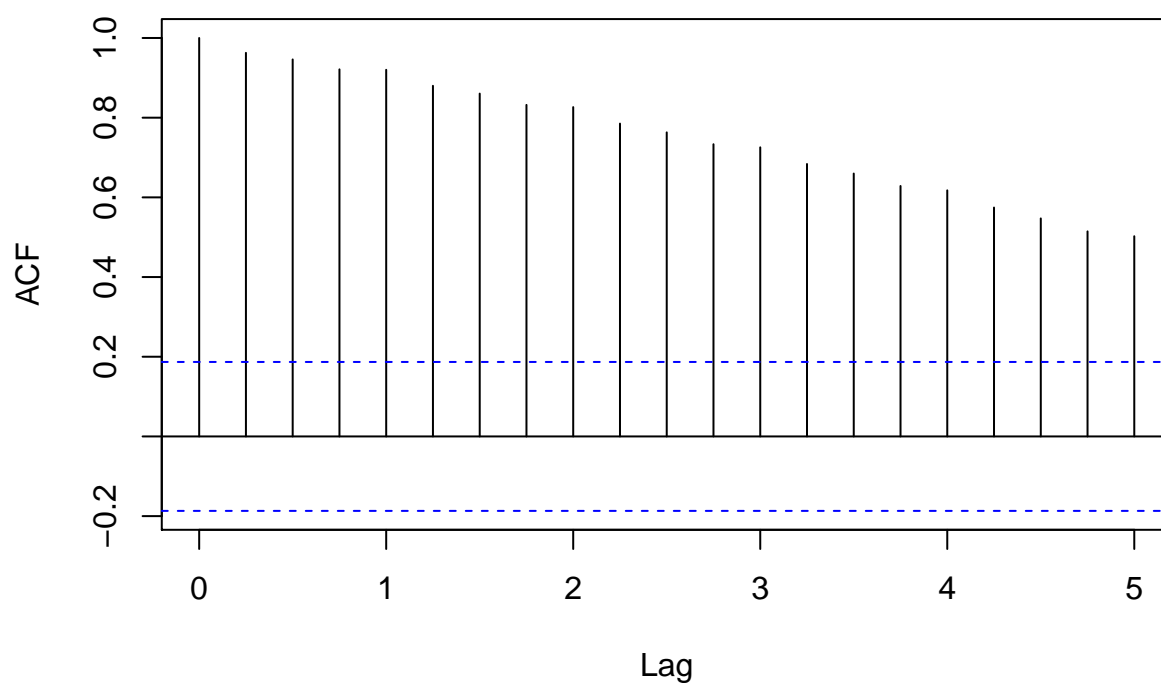
MSE

```
MSE <- ts(trim$MSE, start = 1990, end = c(2017, 2), frequency=4)
plot(MSE, main="Evolution de la masse salariale trimestrielle")
```



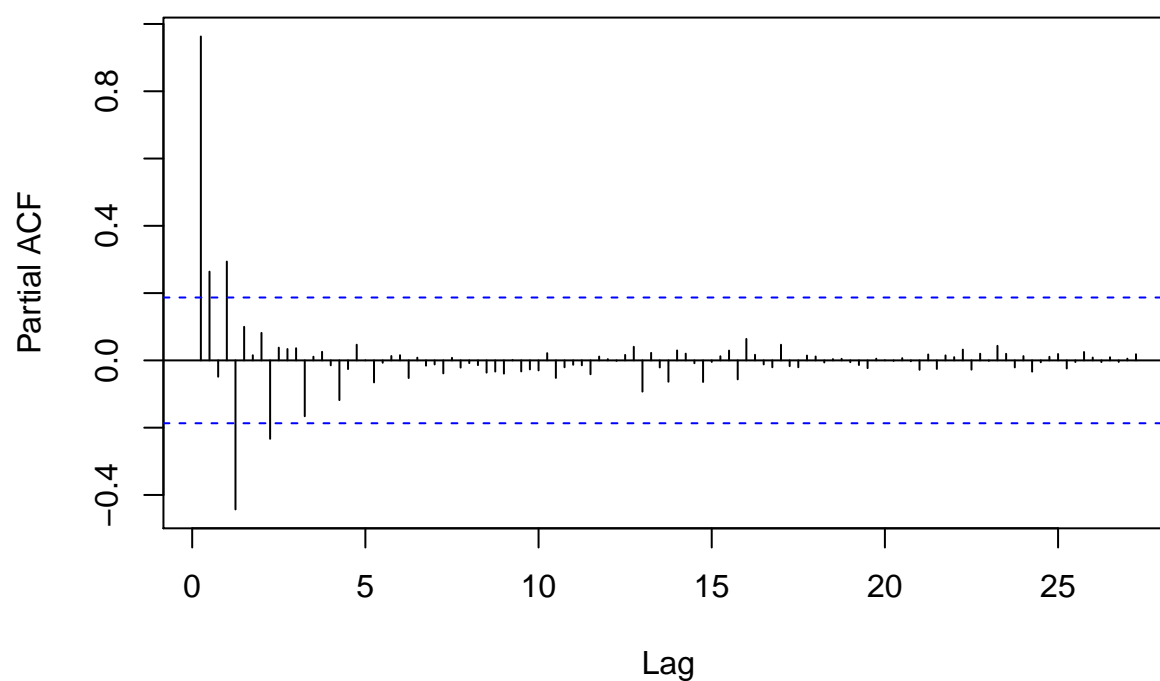
```
acf(MSE, main="Auto-corrélation de la masse salariale trimestrielle")
```

Auto-corrélation de la masse salariale trimestrielle



```
pacf(MSE, "Autocorrélation partielle de la masse trimestrielle")
```

Series MSE



```
adfTest(MSE, type='ct')
```

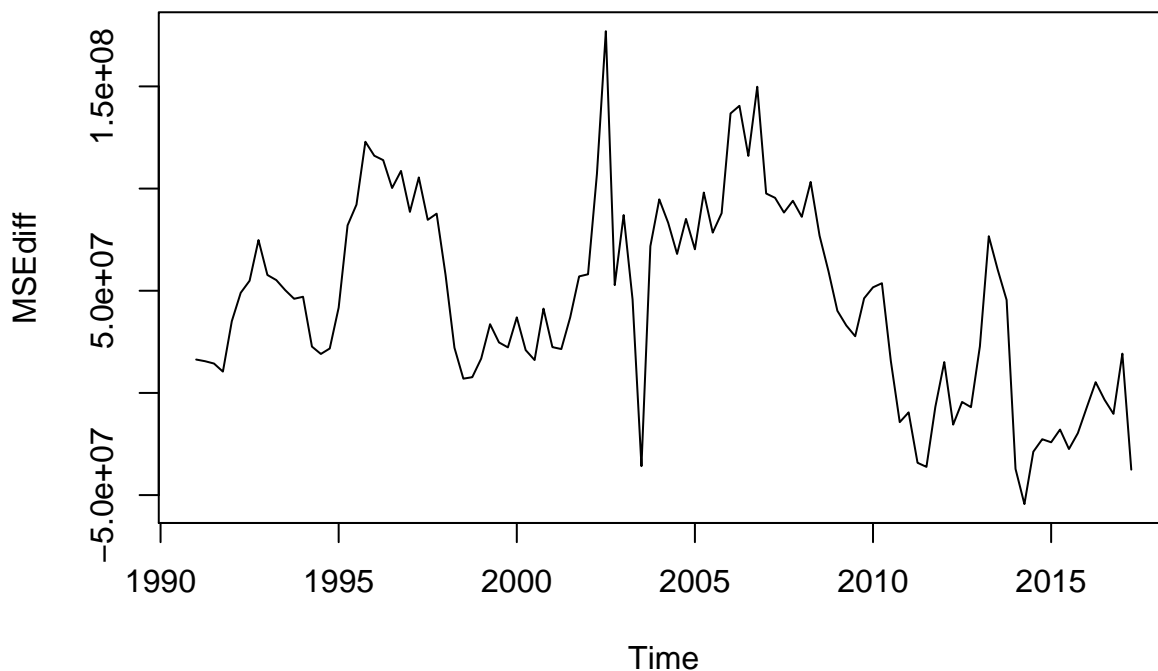
```
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -1.7925
## P VALUE:
## 0.6627
##
## Description:
## Thu Dec 13 10:18:37 2018 by user:
```

La masse salariale trimestrielle possède une composante de tendance de 1990 à 2010. La série tend par la suite à stagner. Nous remarquons également une saisonnalité sur cette série, qui est de plus en plus marquée à mesure que le temps passe.

Comme la série comporte une tendance et une saisonnalité, elle ne correspond pas aux deux premières conditions de la stationnarité du second ordre, soit que la série possède une moyenne et un écart-type constants. Cela est confirmé par la fonction ACF qui décroît régulièrement. Nous effectuons également un test de Dickey-Fuller augmenté servant à vérifier si la série est stationnaire ou non. La p-value est de 0.66 ce qui nous confirme que la série n'est pas stationnaire avec un risque de première espèce de 5%.

Afin de la stationnariser, nous effectuons donc dans un premier temps une différenciation d'ordre 1.

```
MSEdiff<-diff(MSE, lag=4)
plot(MSEdiff)
```



```
kpss.test(MSEdiff)
```

```
## Warning in kpss.test(MSEdiff): p-value smaller than printed p-value
```

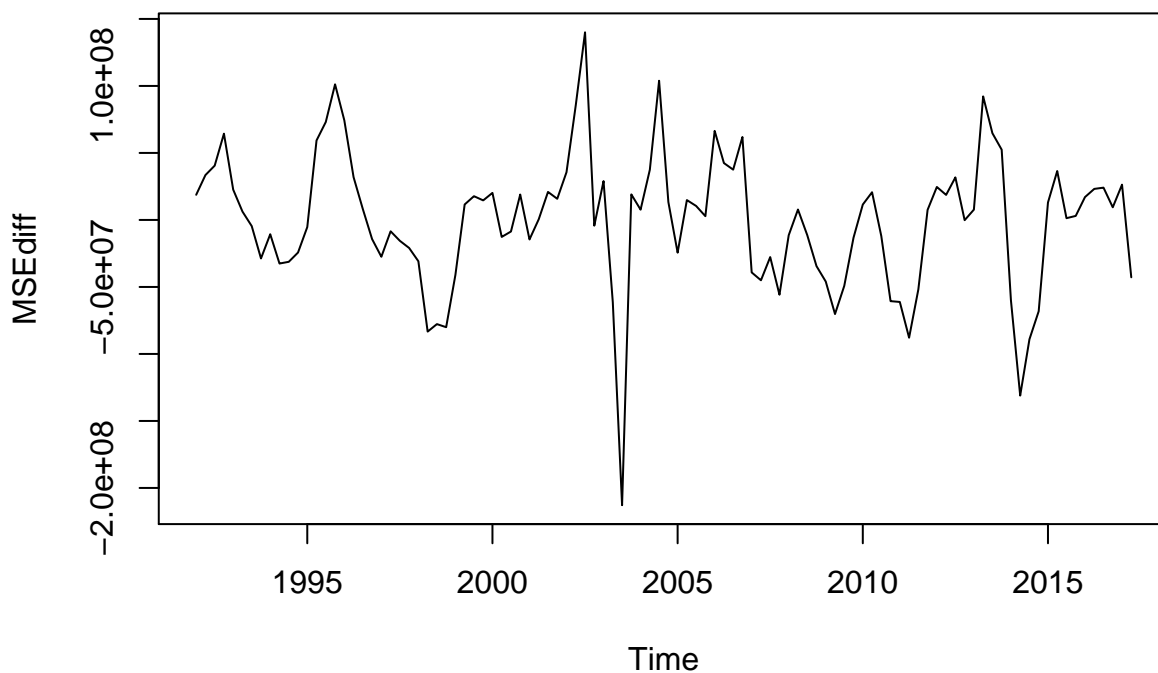
```
##
## KPSS Test for Level Stationarity
##
## data: MSeDiff
## KPSS Level = 0.85062, Truncation lag parameter = 2, p-value = 0.01
```

```
adfTest(MSeDiff, type='c')
```

```
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -2.678
## P VALUE:
## 0.08414
##
## Description:
## Thu Dec 13 10:18:38 2018 by user:
```

Le graphique ci-dessus ne semble pas représenter une série stationnaire. Cela est confirmé par la réalisation d'un test KPSS et d'un test ADF. Avec un risque de 5%, ces deux tests nous permettent de conclure que la série n'est pas stationnaire. Nous choisissons donc d'effectuer une différenciation d'ordre 2. Cette fois-ci, les tests nous permettent de voir que la série obtenue est stationnaire.

```
MSeDiff<-diff(MSE, differences=2, lag=4)
plot(MSeDiff)
```



```
kpss.test(MSeDiff)
```

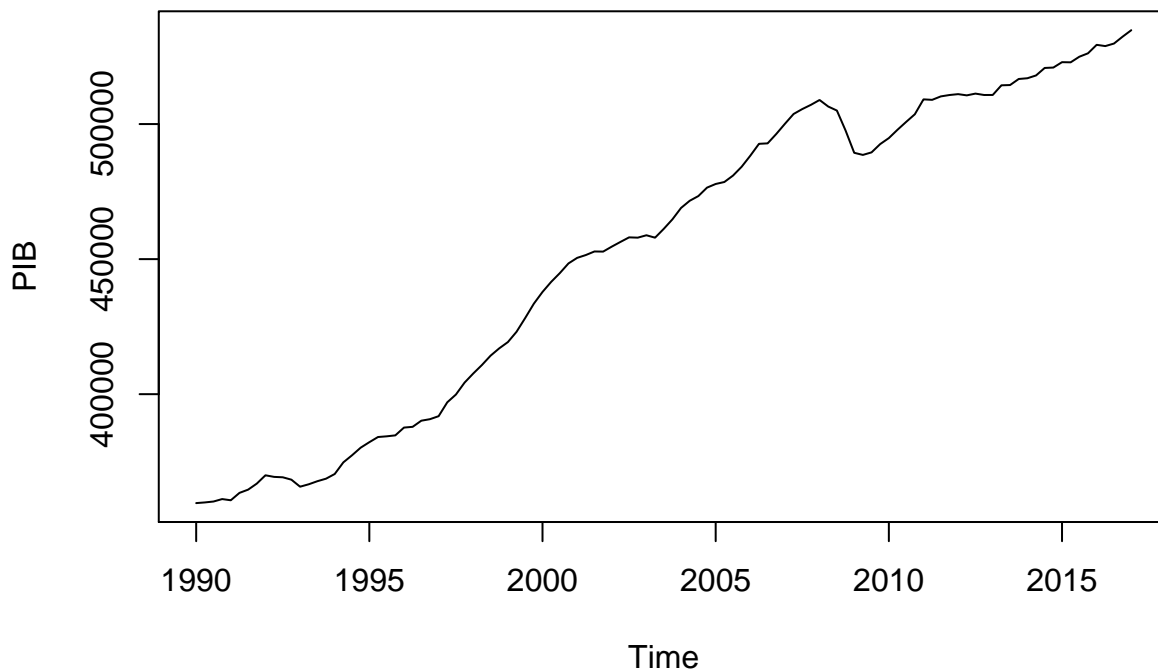
```
## Warning in kpss.test(MSeDiff): p-value greater than printed p-value
```

```
##
## KPSS Test for Level Stationarity
##
## data: MSEdiff
## KPSS Level = 0.11395, Truncation lag parameter = 2, p-value = 0.1
adfTest(MSEdiff, type='c')

## Warning in adfTest(MSEdiff, type = "c"): p-value smaller than printed p-
## value

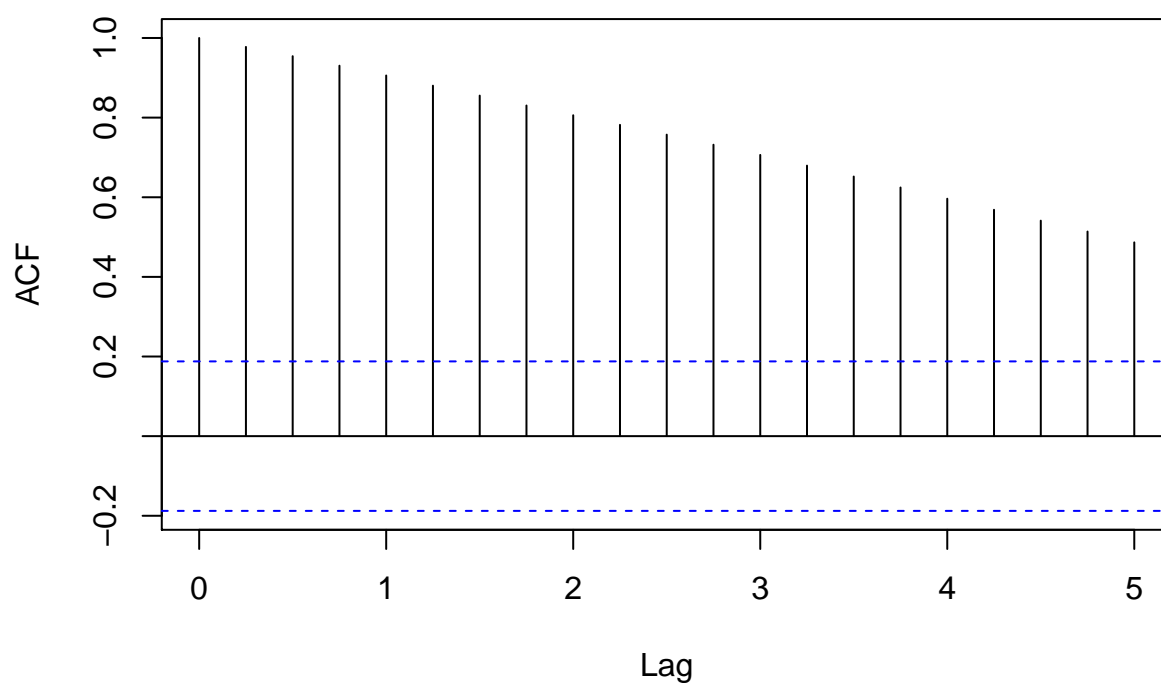
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -5.2989
## P VALUE:
## 0.01
##
## Description:
## Thu Dec 13 10:18:39 2018 by user:
PIB <- ts(trim$PIB, start = 1990, end = c(2017, 1), frequency=4)
plot(PIB, main="Evolution du PIB trimestriel")
```

Evolution du PIB trimestriel



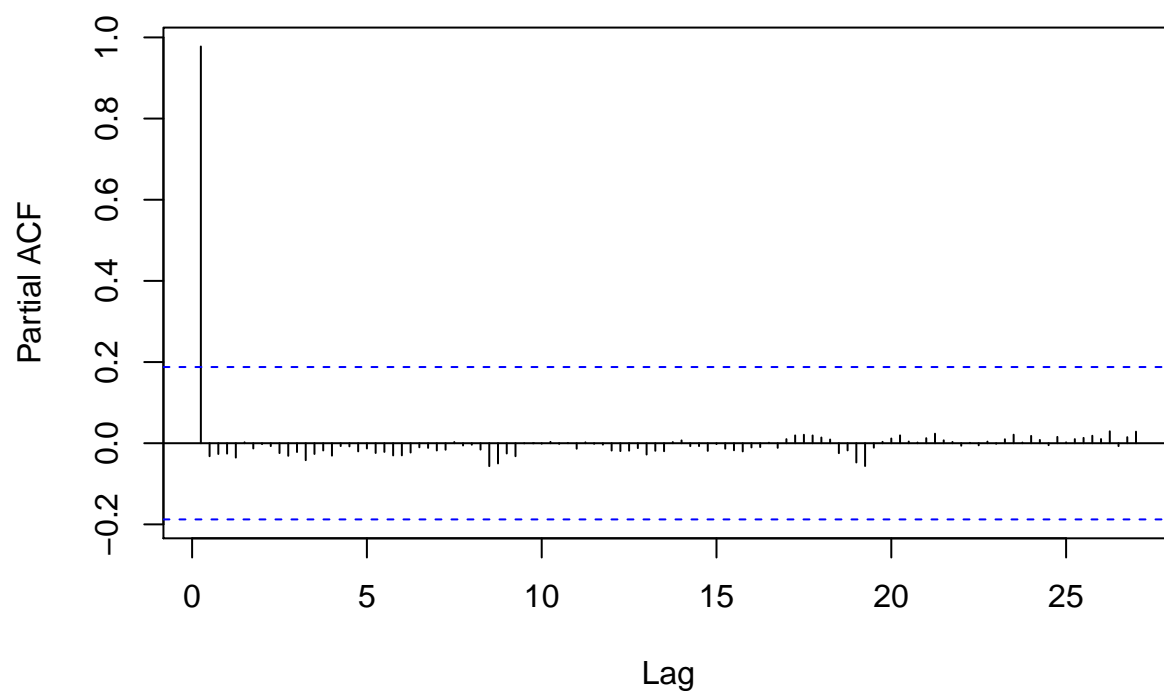
```
acf(PIB, main="Auto-corrélation du PIB trimestriel")
```

Auto-corrélation du PIB trimestriel



```
pacf(PIB, "Autocorrélation partielle du PIB trimestriel")
```

Series PIB

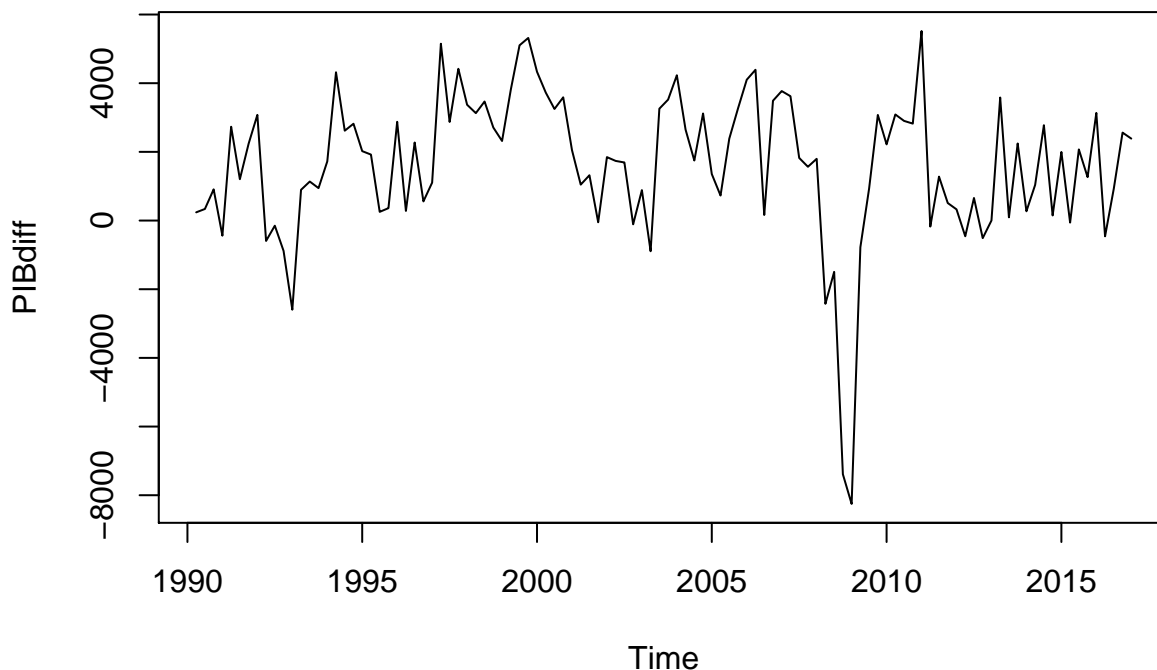



```
adfTest(PIB, type='ct')
```

```
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -1.5079
## P VALUE:
## 0.7808
##
## Description:
## Thu Dec 13 10:18:39 2018 by user:
```

Comme pour la masse salariale, le PIB annuel possède une tendance. Cependant, il ne semble pas posséder de saisonnalité. Cette série n'est donc pas non plus stationnaire. Cependant, il semble cette fois-ci que la tendance soit linéaire, donc une différenciation d'ordre 1 pourrait suffire à stationnariser la série.

```
PIBdiff<-diff(PIB)
plot(PIBdiff)
```



```
kpss.test(PIBdiff)
```

```
## Warning in kpss.test(PIBdiff): p-value greater than printed p-value
##
## KPSS Test for Level Stationarity
##
## data: PIBdiff
## KPSS Level = 0.20736, Truncation lag parameter = 2, p-value = 0.1
```

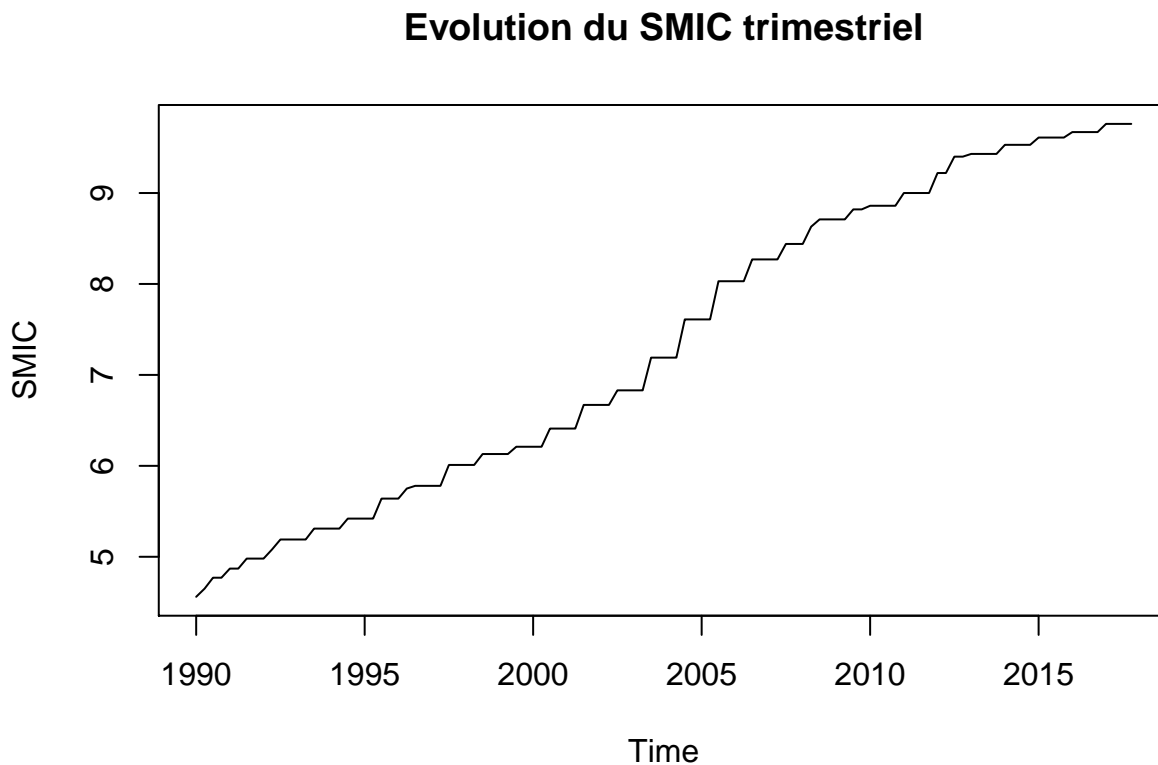
```
adfTest(PIBdiff, type='c')
```

```
## Warning in adfTest(PIBdiff, type = "c"): p-value smaller than printed p-  
## value  
  
##  
## Title:  
## Augmented Dickey-Fuller Test  
##  
## Test Results:  
## PARAMETER:  
## Lag Order: 1  
## STATISTIC:  
## Dickey-Fuller: -4.0967  
## P VALUE:  
## 0.01  
##  
## Description:  
## Thu Dec 13 10:18:40 2018 by user:
```

En prenant un risque de première espèce de 5%, le test KPSS et le test ADF nous indiquent que la série différenciée est bien stationnaire, sans avoir eu à prendre en compte une quelconque saisonnalité.

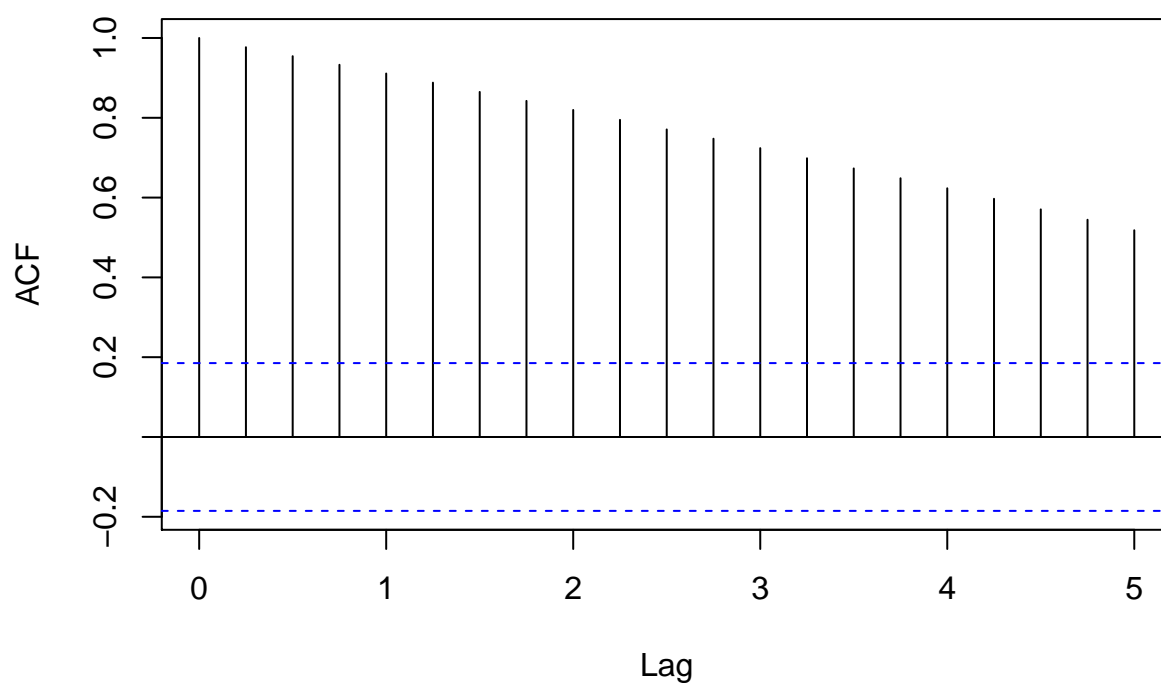
SMIC

```
SMIC <- ts(trim$SMIC, start = c(1990,1), end = c(2017, 4), frequency = 4)  
plot(SMIC, main="Evolution du SMIC trimestriel")
```



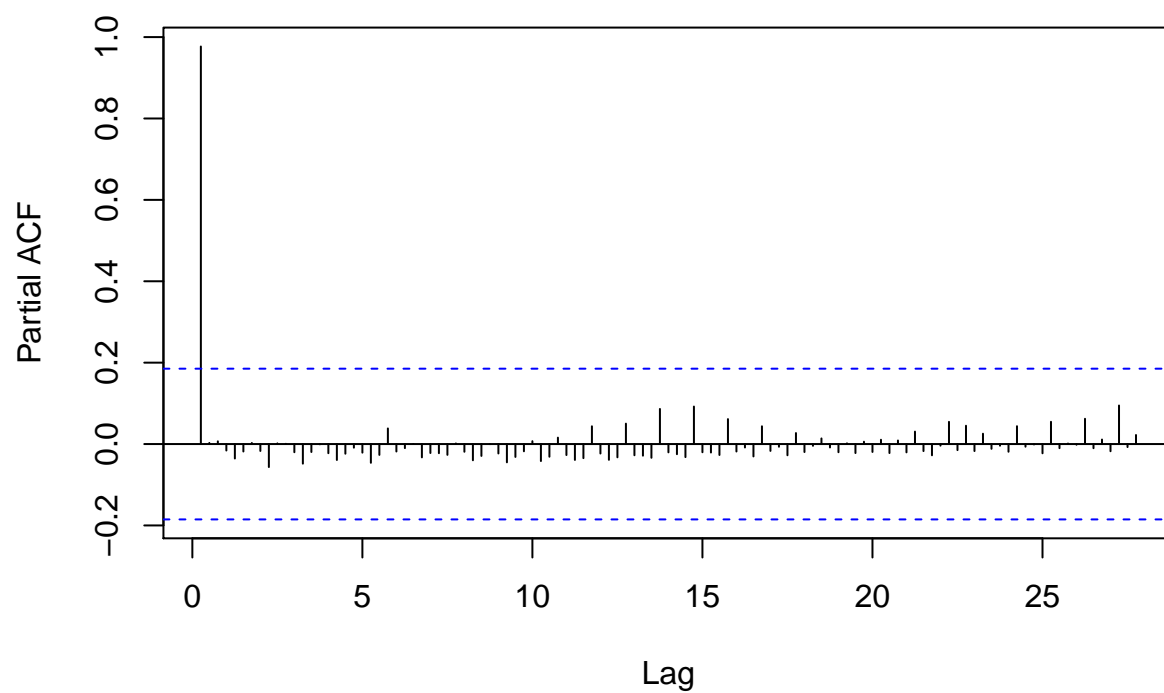
```
acf(SMIC, main="Auto-corrélation du SMIC trimestriel")
```

Auto-corrélation du SMIC trimestriel



```
pacf(SMIC, "Autocorrélation partielle du SMIC trimestriel")
```

Series SMIC



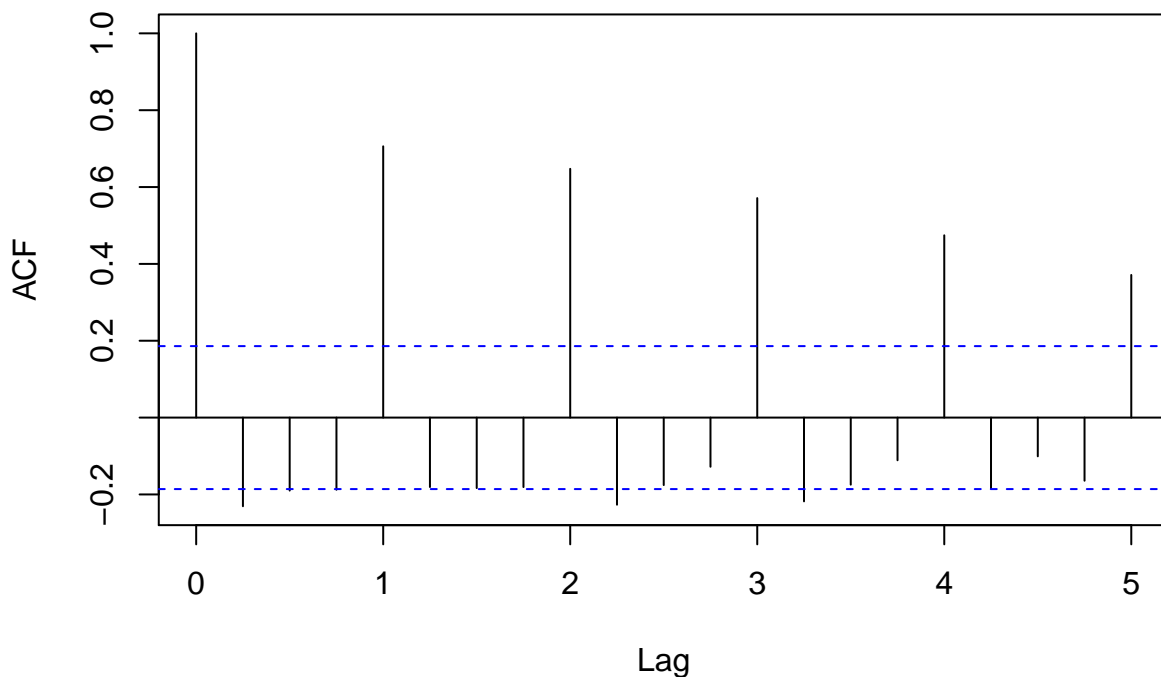
```
adfTest(SMIC, type='ct')
```

```
##  
## Title:  
## Augmented Dickey-Fuller Test  
##  
## Test Results:  
## PARAMETER:  
## Lag Order: 1  
## STATISTIC:  
## Dickey-Fuller: -1.1095  
## P VALUE:  
## 0.9174  
##  
## Description:  
## Thu Dec 13 10:18:40 2018 by user:
```

Au regard de la représentation graphique, on s'aperçoit qu'il y a bien une tendance. Pour la saisonnalité, il est plus difficile de savoir s'il en existe une ou pas, puisque la série semble augmenter seulement à certains temps. Nous allons donc supprimer la tendance dans un premier temps pour définir s'il existe ou non une saisonnalité.

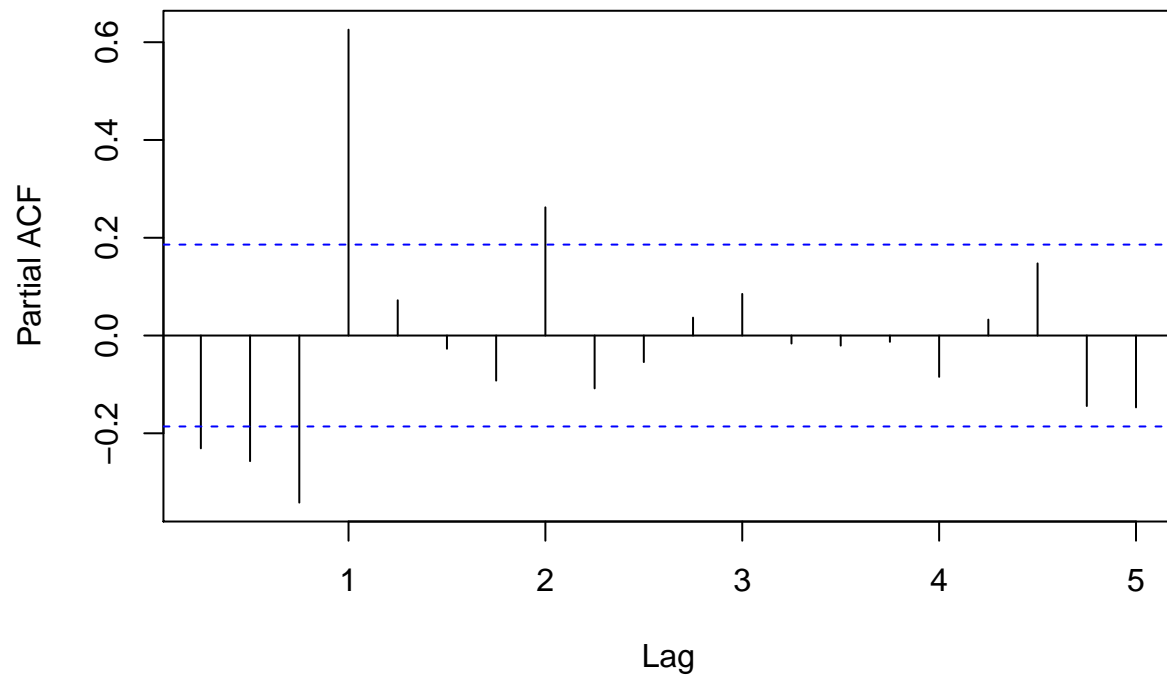
```
SMICdiff<-diff(SMIC)  
acf(SMICdiff)
```

Series SMICdiff



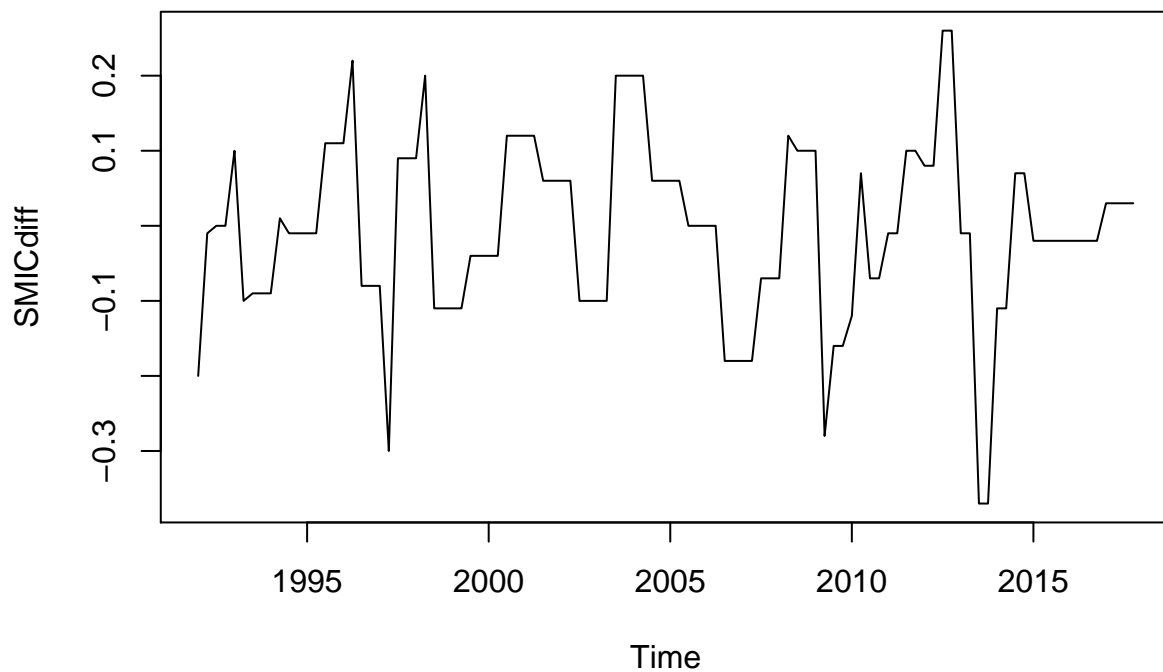
```
pacf(SMICdiff)
```

Series SMICdiff



Après analyse de la fonction ACF, il semblerait qu'il existe bien une saisonnalité. Nous allons donc la supprimer avec la tendance, pour rendre la série stationnaire.

```
SMICdiff<-diff(SMIC, lag=4, differences = 2)
plot(SMICdiff)
```



```
kpss.test(SMICdiff)
```

```
## Warning in kpss.test(SMICdiff): p-value greater than printed p-value
```

```
##
## KPSS Test for Level Stationarity
##
## data: SMICdiff
## KPSS Level = 0.058654, Truncation lag parameter = 2, p-value = 0.1
adfTest(SMICdiff, type='c')

## Warning in adfTest(SMICdiff, type = "c"): p-value smaller than printed p-
## value

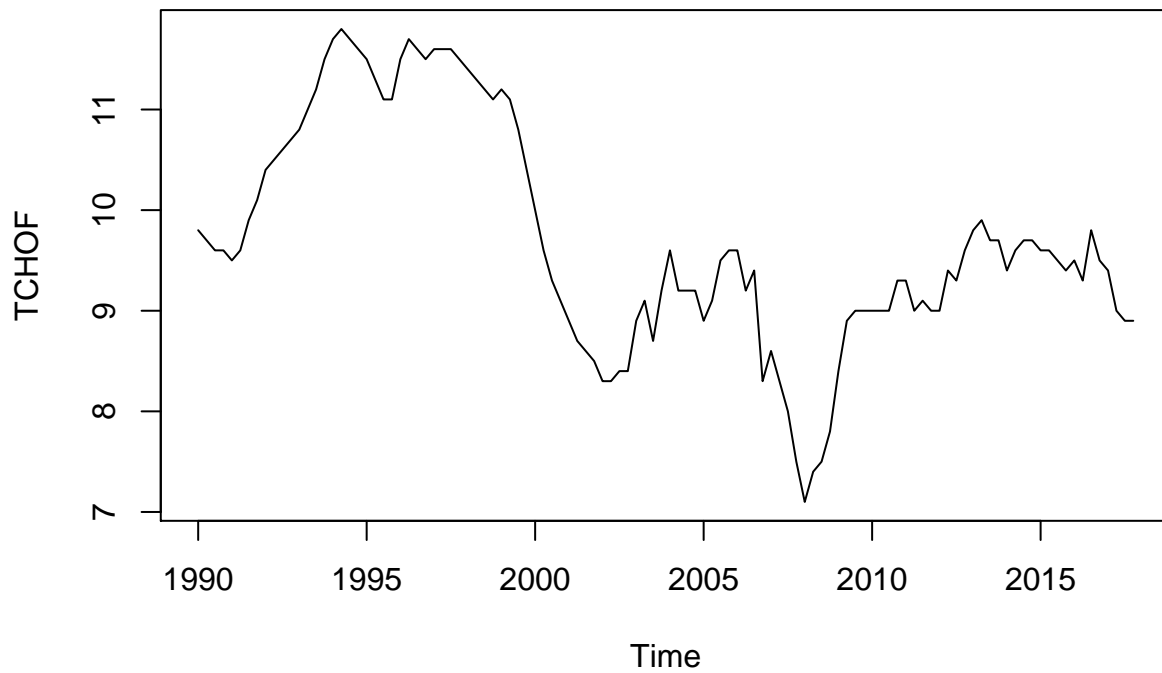
##
## Title:
## Augmented Dickey-Fuller Test
##
## Test Results:
## PARAMETER:
## Lag Order: 1
## STATISTIC:
## Dickey-Fuller: -4.9522
## P VALUE:
## 0.01
##
## Description:
## Thu Dec 13 10:18:42 2018 by user:
```

Les résultats des tests d'hypothèse nous permettent donc de confirmer que cette série différenciée est stationnaire avec un risque de premier espèce à 5%.

Taux de chômage des femmes

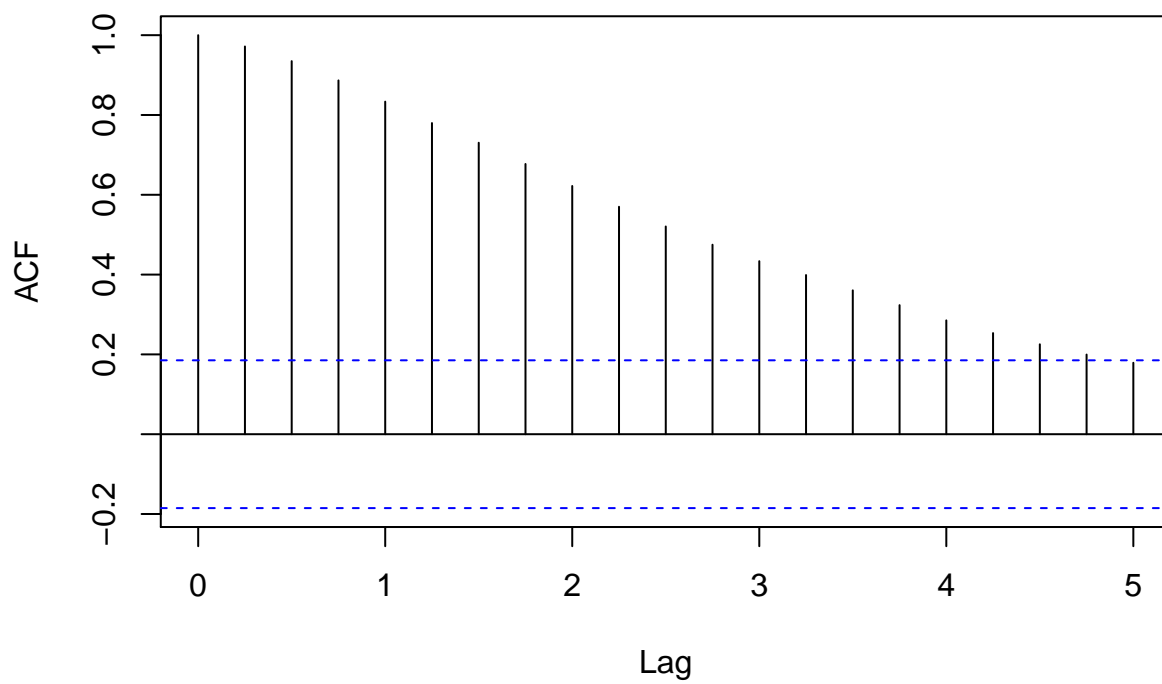
```
TCHOF <- ts(trim$TCHOF, start = c(1990,1), end = c(2017, 4), frequency = 4)
plot(TCHOF, main="Evolution du taux de chômage des femmes trimestriel")
```

Evolution du taux de chômage des femmes trimestriel



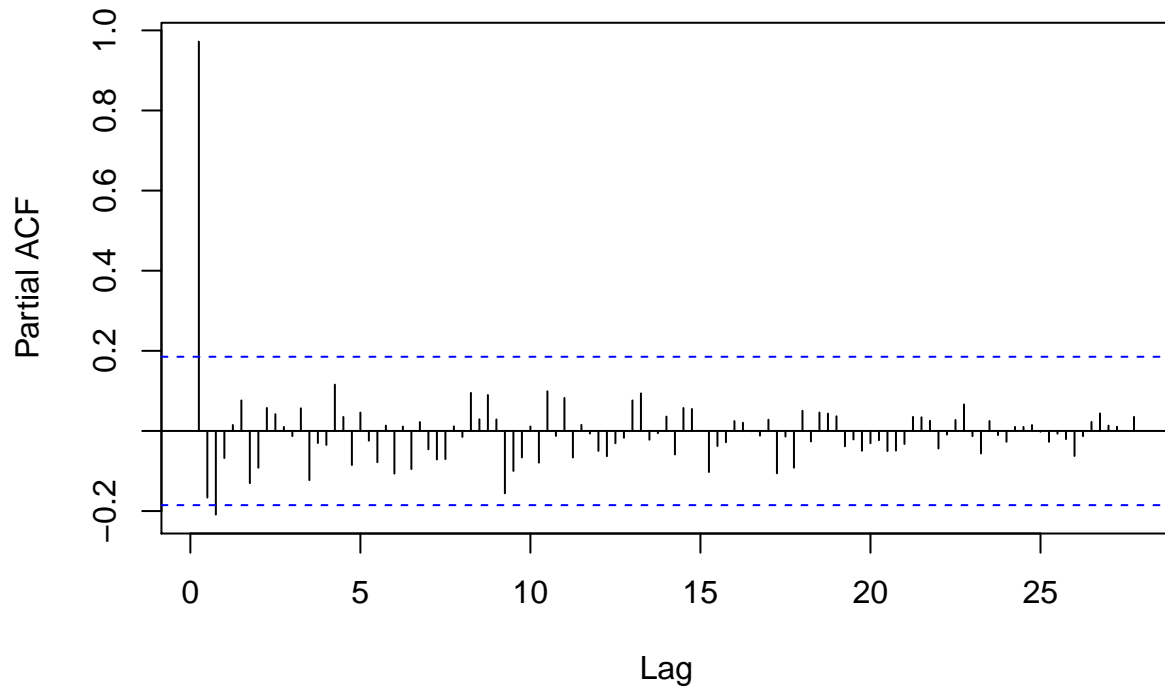
```
acf(TCHOF, main="Auto-corrélation du taux de chômage des femmes trimestriel")
```

Auto-corrélation du taux de chômage des femmes trimestriel



```
pacf(TCHOF, "Autocorrélation partielle du taux de chômage des femmes trimestriel")
```

Series TCHOF

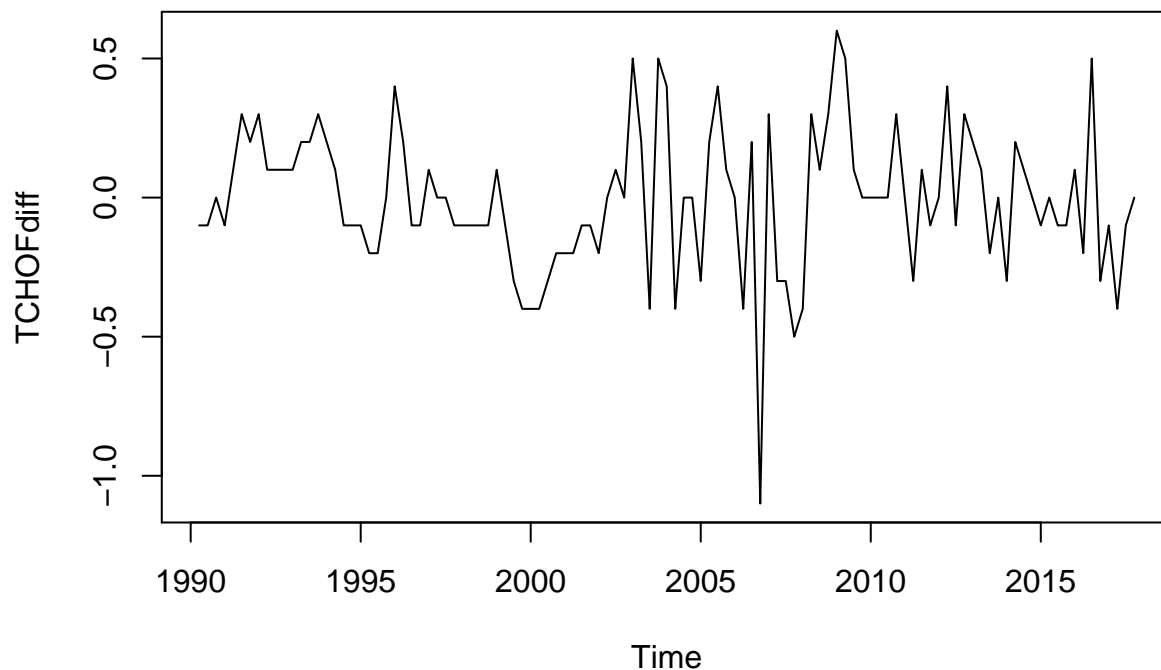


```
adfTest(TCHOF, type='ct')
```

```
##
## Title:
##   Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 1
##   STATISTIC:
##     Dickey-Fuller: -1.903
##   P VALUE:
##     0.6169
##
## Description:
##   Thu Dec 13 10:18:43 2018 by user:
```

Pour cette dernière série qui représente le taux de chômage trimestriel des femmes, il ne semble pas y avoir de saisonnalité. On remarque cependant qu'il y a bien une tendance. Le test ADF nous confirme que la série n'est pas stationnaire. Nous allons donc tenter de la stationnariser.

```
TCHOFdiff<-diff(TCHOF)
plot(TCHOFdiff)
```

```
kpss.test(TCHOEdiff)
```

```
## Warning in kpss.test(TCHOEdiff): p-value greater than printed p-value
```

```
##
```

```
## KPSS Test for Level Stationarity
```

```
##
```

```
## data: TCHOEdiff
```

```
## KPSS Level = 0.11393, Truncation lag parameter = 2, p-value = 0.1
```

```
adfTest(TCHOEdiff, type='c')
```

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

```
##
```

```
## Title:
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## Test Results:
```

```
## PARAMETER:
```

```
## Lag Order: 1
```

```
## STATISTIC:
```

```
## Dickey-Fuller: -5.3716
```

```
## P VALUE:
```

```
## 0.01
```

```
##
```

```
## Description:
```

```
## Thu Dec 13 10:18:44 2018 by user:
```

Après une différenciation d'ordre 1, les tests ADF et KPSS nous amènent à conclure que la série est stationnaire.