

# Caractérisation du taux de chômage des Etats-Unis

Yannick Le Pen

2023-07-20

## Importation et des données à partir de DB.nomics

On importe les données trimestrielles du taux de chômage aux Etats-Unis du 01/01/1994 au 23/01/01 (117 observations). Les données sont importées du site db.nomics. Les données proviennent du “Bureau of Labor Statistics”. On leur importe en utilisant leur identifiant ids.

```
rm(list = ls())
library(data.table)
library(rdbnomics)

df<-rdb(ids = "BLS/lm/LNS13327708Q")
#class(df)
#str(df)
```

On crée un dataframe contenant uniquement les dates et la série du taux de chômage.

```
df_us<-df[,c("period","value")]
colnames(df_us)<-c('date','taux')
head(df_us)
```

```
##           date  taux
## 1: 1994-01-01   7.9
## 2: 1994-04-01   7.4
## 3: 1994-07-01   7.3
## 4: 1994-10-01   6.8
## 5: 1995-01-01   6.6
## 6: 1995-04-01   6.8
```

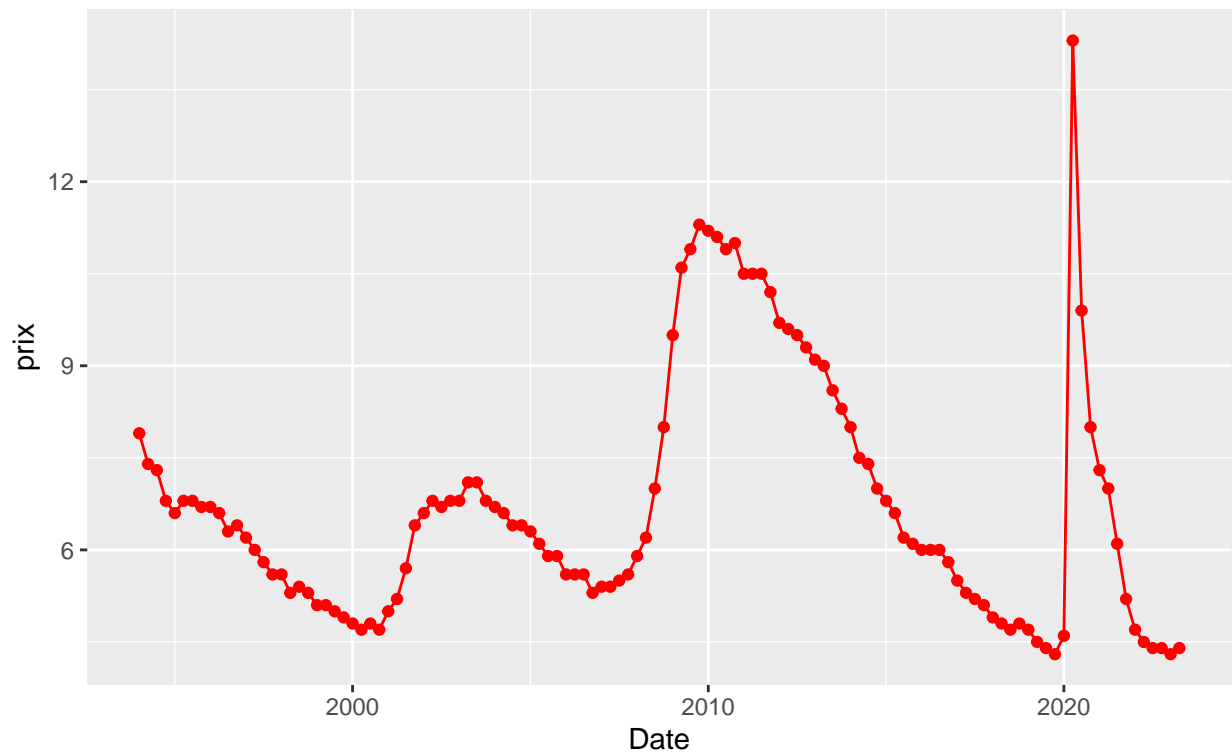
## Représentation graphique

Le taux de chômage aux Etats-Unis fluctue autour de 6% du début de l'échantillon jusqu'à l'année 2008 à partir de laquelle il augmente très rapidement jusqu'au niveau de 11%, sous l'effet de la crise des subprimes. On observe aussi l'épisode de la crise du Covid avec une hausse instantanée du taux de chômage autour de 14% suivie d'un retour rapide à des niveaux très bas. La période du covid peut avoir un impact sur les résultats des tests statistiques. Au vu du graphique, il est assez difficile de trancher sur la stationnarité du taux de chômage.

```
library(ggplot2)
p_us<-ggplot(data=df_us,aes(x=date,y=taux))+geom_point(color='red')+geom_line(color='red')+xlab('Date')+
p_us+labs(y="prix",title = "Taux de chômage aux Etats-Unis",subtitle = "1994/01/01 à 2023/01/01")
```

## Taux de chômage aux Etats-Unis

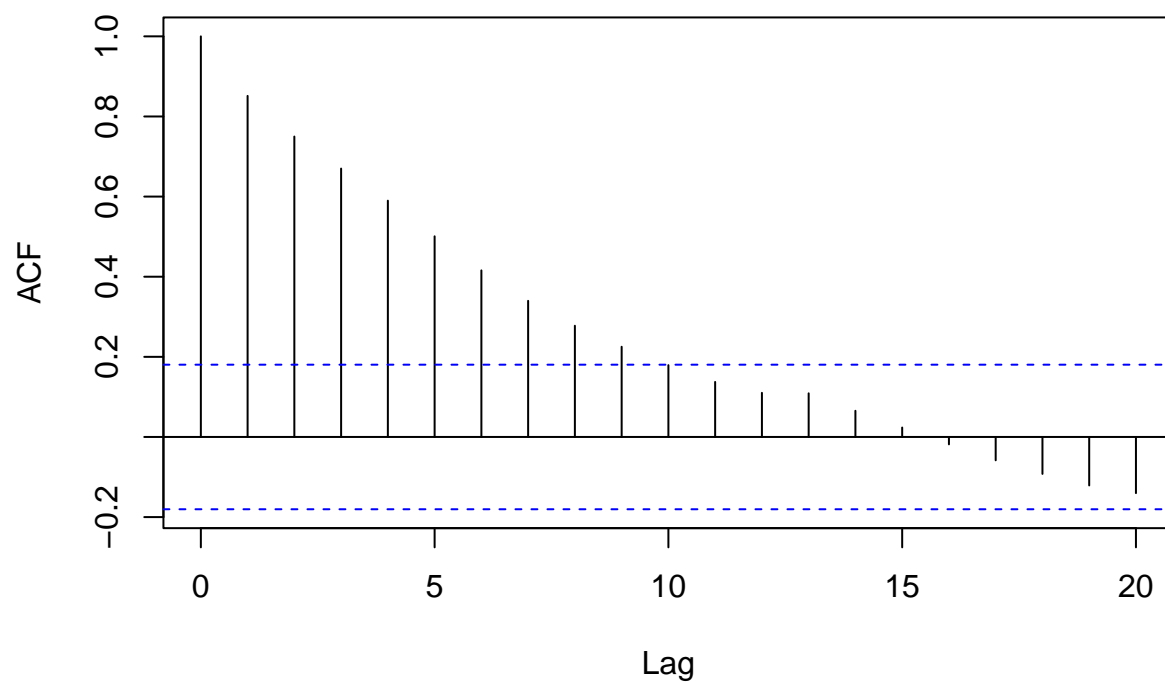
1994/01/01 à 2023/01/01



### Autocorrélogramme du taux de chômage

```
acf_us_rate<-acf(x=df_us$taux,main='Autocorrélogramme du taux de chômage des Etats-Unis')
```

## Autocorrélogramme du taux de chômage des Etats-Unis



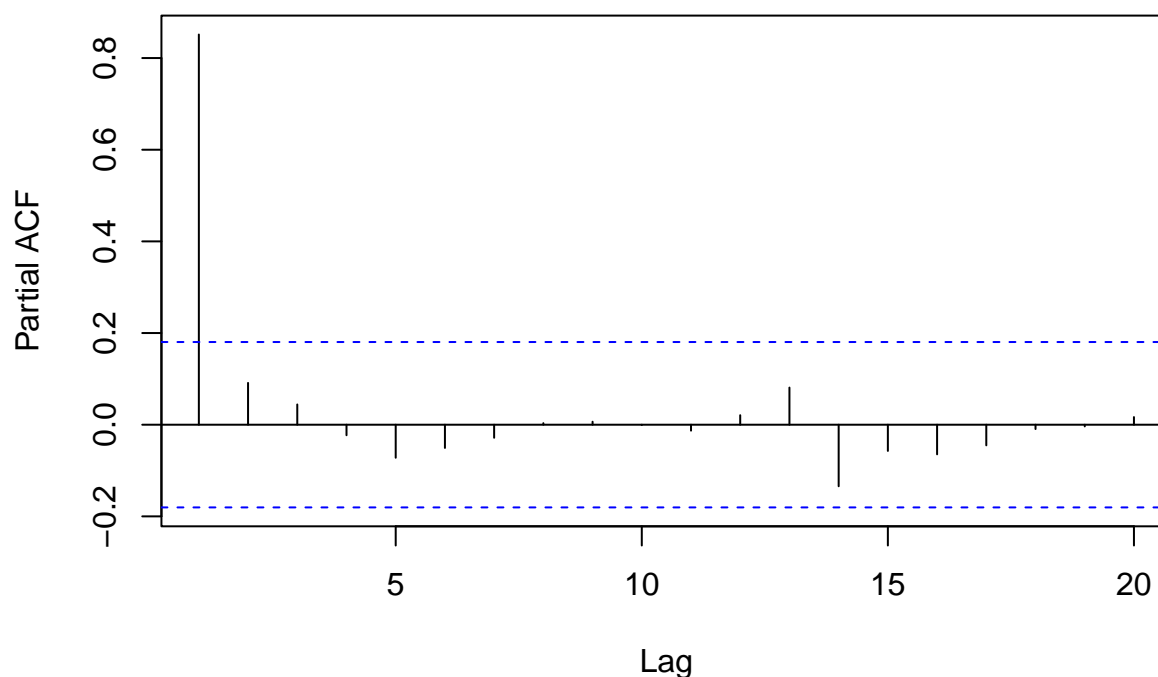
```
acf_us_rate$acf
```

```
## , , 1
##
##      [,1]
## [1,] 1.00000000
## [2,] 0.85138711
## [3,] 0.74995655
## [4,] 0.66998853
## [5,] 0.58982443
## [6,] 0.50089258
## [7,] 0.41587082
## [8,] 0.33983021
## [9,] 0.27763319
## [10,] 0.22519609
## [11,] 0.17924791
## [12,] 0.13738377
## [13,] 0.11019914
## [14,] 0.10899009
## [15,] 0.06539666
## [16,] 0.02354248
## [17,] -0.01846834
## [18,] -0.05830022
## [19,] -0.09222182
## [20,] -0.12108968
## [21,] -0.14026802
```

On observe que l'autocorrélation est supérieure à 0.5 jusqu'à l'ordre 4, ce qui est plutôt élevé. Elle n'est plus significative, pour un risque de première espèce de 5%, à partir de l'ordre 11. L'autocorrélogramme du taux de chômage serait compatible avec celui d'un processus stationnaire.

```
pacf(x=df_us$taux,main='Autocorrélogramme partiel du taux de chômage des Etats-Unis')
```

## Autocorrélogramme partiel du taux de chômage des Etats-Unis



Aucune autocorrélation partielle n'est significativement différente de 0. L'autocorrélogramme et l'autocorrélogramme partiel nous suggère que le taux de chômage pourrait être modélisé par un processus AR(p).

## Test de Dickey-Fuller augmenté

### Test ADF avec constante et tendance déterministe

On applique le test ADF avec une tendance déterministe. Le nombre de retards maximum est fixé à 8 et le nombre de retards optimal est déterminé par la minimisation du critère AIC.

```
library(urca)
adf1<-ur.df(y = df_us$taux,type=c("trend"), lag=8, selectlags = c('AIC'))
summary(adf1)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
```

```
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4246 -0.3135 -0.1312  0.0857  9.5244
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.8247025  0.4032732   2.045  0.0434 *
## z.lag.1      -0.1206021  0.0525214  -2.296  0.0236 *
## tt          -0.0006008  0.0032668  -0.184  0.8544
## z.diff.lag   -0.1043086  0.0979169  -1.065  0.2892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.067 on 105 degrees of freedom
## Multiple R-squared:  0.07579,    Adjusted R-squared:  0.04939
## F-statistic:  2.87 on 3 and 105 DF,  p-value: 0.03993
##
##
## Value of test-statistic is: -2.2962 1.8372 2.7264
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

La statistique du test ADF est égale à  $t_{ADF} = -2.296$ . Les seuils de rejet figurent sur la ligne tau3. La statistique de test

est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10%. On ne peut pas rejeter l'hypothèse nulle de racine unitaire.

Le t-stat de la tendance déterministe est égal à  $t_b = -0.184$ . Le seuil de rejet qui figure la table de est égal à 3.14 pour un risque de première espèce de 5%. on a donc : on ne peut rejeter l'hypothèse nulle que la tendance déterministe n'est pas significative.

## Test avec constante uniquement

```
adf2<-ur.df(y = df_us$taux,type=c("drift"), lag=8, selectlags = c('AIC'))
summary(adf2)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -2.4540 -0.3013 -0.1324  0.1007  9.4967
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.79360    0.36444   2.178  0.0317 *
## z.lag.1      -0.12160    0.05200  -2.339  0.0212 *
## z.diff.lag   -0.10316    0.09727  -1.061  0.2913
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.062 on 106 degrees of freedom
## Multiple R-squared:  0.0755, Adjusted R-squared:  0.05805
## F-statistic: 4.328 on 2 and 106 DF,  p-value: 0.0156
##
##
## Value of test-statistic is: -2.3385 2.764
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

La statistique du test ADF est égale à  $t_{ADF} = -2.296$ . Les seuils de rejet figurent sur la ligne tau3. La statistique de test

$t_{ADF}$  est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10%. On ne peut pas rejeter l'hypothèse nulle de racine unitaire.

Le t-stat de la tendance déterministe est égal à  $t_{\hat{c}} = 2.170$ . Le seuil de rejet qui figure la table de  $t_{\hat{c}}$  est égal à 2.86 pour un risque de première espèce de 5%. on a donc  $|t_{\hat{c}}| = 2.170 < 2.86$  : on ne peut rejeter l'hypothèse nulle que la constante n'est pas significative.

## Test sans constante ni tendance déterministe

```
adf3<-ur.df(y = df_us$taux,type=c("none"), lag=8, selectlags = c('AIC'))
summary(adf3)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -2.6527 -0.1550 -0.0434  0.1199  9.8075
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.01287    0.01477  -0.871  0.3854
```

```
## z.diff.lag -0.16116    0.09518  -1.693   0.0933 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.08 on 107 degrees of freedom
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.01645
## F-statistic: 1.912 on 2 and 107 DF,  p-value: 0.1528
##
##
## Value of test-statistic is: -0.8715
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

La statistique du test ADF est égale à  $t_{ADF} = -0.871$ . Les seuils de rejet figurent sur la ligne tau3. La statistique de test

$t_{ADF}$  est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10%. On ne peut pas rejeter l'hypothèse nulle de racine unitaire.

## Conclusion

Les résultats des tests ADF nous amènent à conclure que la série du taux de chômage est une série DS sans tendance ni constante. On dit aussi qu'elle est un processus DS sans dérive (drift).

## Test de stationnarité de KPSS¶

On applique le test de stationnarité de KPSS.

### Stationnarité autour d'une tendance déterministe

```
kpss_tau<-ur.kpss(y=df_us$taux,type="tau",lags="short")
summary(kpss_tau)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: tau with 4 lags.
##
## Value of test-statistic is: 0.2405
##
## Critical value for a significance level of:
##      10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
```

### Stationnarité autour d'une constante

On présente les résultats du test avec une constante. La représentation graphique de la série et les conclusions du test ADF montrent que la taux de chômage ne contient pas de tendance déterministe.

```
kpss_mu<-ur.kpss(y=df_us$taux,type="mu",lags="short")
summary(kpss_mu)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: mu with 4 lags.
##
## Value of test-statistic is: 0.2549
##
## Critical value for a significance level of:
##          10pct  5pct  2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

La statistique du test KPSS avec une constante est égale à  $LM_{KPSS} = 0.2578$ . Elle est inférieure au seuil de rejet à 5% :  $0.2578 < 0.436$  (et aussi à celle à 1%). On ne peut donc pas rejeter l'hypothèse de stationnarité du taux de chômage autour d'une constante.

## Conclusion générale sur les test de racine unitaire et de stationnarité

On voit que les conclusions des tests ADF et KPSS sont contradictoires en ce qui concerne la stationnarité du taux de chômage. Des études empiriques sur le taux de chômage aux Etats-Unis conclut à la stationnarité du taux de chômage (Nelson and Plosser, 1982, Perron, 1988, Xiao and Phillips, 1997). On reviendra sur cette question lors l'étude de la modélisation du taux de chômage par un modèle ARMA (chapitre III).

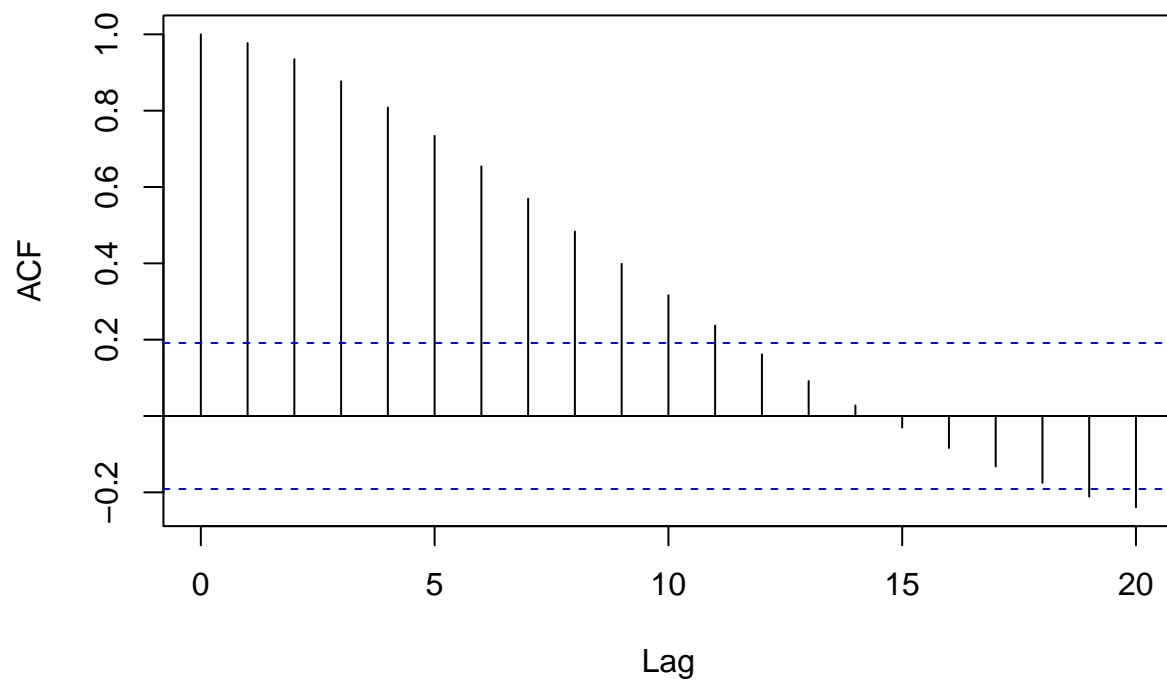
## Tests sans la période Covid

On exclu les observations postérieures au premier semestre de l'année 2020. Les résultats des tests de racine unitaire et de stationnarité ne sont modifiés. On peut dire aussi qu'ils sont robustes à cette modification de la période de test.

```
df_us_20Q1=df_us[df_us$date<"2020-04-01",]
acf_us_urate<-acf(x=df_us_20Q1$taux,main='Autocorrélogramme du taux de chômage des Etats-Unis')
```

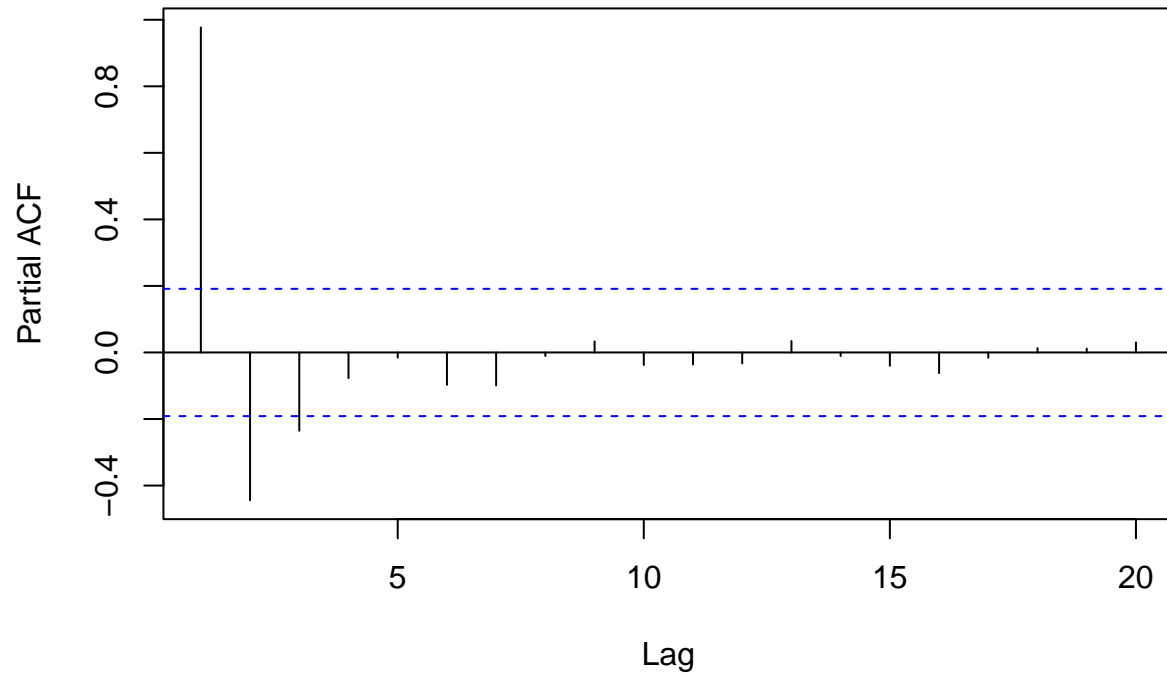


## Autocorrélogramme du taux de chômage des Etats-Unis



```
pacf(x=df_us_20Q1$taux,main='Autocorrélogramme partiel du taux de chômage des Etats-Unis')
```

## Autocorrélogramme partiel du taux de chômage des Etats-Unis



### Test de Dickey-Fuller augmenté

**Avec constante et tendance déterministe** On applique le test ADF avec une tendance déterministe. Le nombre de retards maximum est fixé à 8 et le nombre de retards optimal est déterminé par la minimisation du critère AIC.

```
adf1<-ur.df(y = df_us_20Q1$taux,type=c("trend"), lag=8, selectlags = c('AIC'))
summary(adf1)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58235 -0.15906 -0.01244  0.12640  0.78606
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1950389  0.0911865   2.139  0.03512 *
```

```
## z.lag.1      -0.0312594  0.0125379  -2.493  0.01447 *
## tt          0.0002514  0.0008452   0.297  0.76679
## z.diff.lag1  0.5296286  0.1006194   5.264  9.35e-07 ***
## z.diff.lag2  0.2803233  0.1031839   2.717  0.00789 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2219 on 91 degrees of freedom
## Multiple R-squared:  0.5374, Adjusted R-squared:  0.5171
## F-statistic: 26.43 on 4 and 91 DF,  p-value: 1.484e-14
##
##
## Value of test-statistic is: -2.4932 2.0941 3.1336
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

La statistique du test ADF est égale à  $t_{ADF} = -2.493$ . Les seuils de rejet figurent sur la ligne tau3. La statistique de test

$t_{ADF}$  est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10%. On ne peut pas rejeter l'hypothèse nulle de racine unitaire.

Le t-stat de la tendance déterministe est égal à  $t_b = 0.297$ . Le seuil de rejet qui figure la table de  $t_b$  est égal à 3.14 pour un risque de première espèce de 5%. on a donc  $|t_b| = 0.297 < 3.14$  : on ne peut rejeter l'hypothèse nulle que la tendance déterministe n'est pas significative.

```
adf2<-ur.df(y = df_us_20Q1$taux,type=c("drift"), lag=8, selectlags = c('AIC'))
summary(adf2)
```

### Avec constante

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.57754 -0.16127 -0.01475  0.12106  0.79018
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.20389    0.08577   2.377  0.01951 *
## z.lag.1       -0.03048    0.01220  -2.498  0.01426 *
## z.diff.lag1    0.52786    0.09994   5.282  8.53e-07 ***
## z.diff.lag2    0.27742    0.10221   2.714  0.00793 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2208 on 92 degrees of freedom
## Multiple R-squared:  0.537, Adjusted R-squared:  0.5219
## F-statistic: 35.57 on 3 and 92 DF,  p-value: 2.365e-15
##
##
## Value of test-statistic is: -2.4981 3.1279
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

La statistique du test ADF est égale à  $t_{ADF} = -2.498$ . Les seuils de rejet figurent sur la ligne tau3. La statistique de test

$t_{ADF}$  est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10%. On ne peut pas rejeter l'hypothèse nulle de racine unitaire.

Le t-stat de la tendance déterministe est égal à  $t_{\hat{c}} = 2.377$ . Le seuil de rejet qui figure la table de  $t_{\hat{c}}$  est égal à 2.86 pour un risque de première espèce de 5%. on a donc  $|t_{\hat{c}}| = 2.377 < 2.86$  : on ne peut rejeter l'hypothèse nulle que la constante n'est pas significative.

```
adf3<-ur.df(y = df_us_20Q1$taux,type=c("none"), lag=8, selectlags = c('AIC'))
summary(adf3)
```

#### Sans constante ni tendance déterministe

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60273 -0.11937  0.01128  0.13966  0.80186
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.002501   0.003297  -0.759   0.4501
## z.diff.lag1   0.538428   0.102311   5.263 9.07e-07 ***
## z.diff.lag2   0.224645   0.102236   2.197  0.0305 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2262 on 93 degrees of freedom
## Multiple R-squared:  0.5109, Adjusted R-squared:  0.4951
## F-statistic: 32.38 on 3 and 93 DF,  p-value: 2.027e-14
```

```
##
##
## Value of test-statistic is: -0.7585
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

La statistique du test ADF est égale à  $t_{ADF} = -0.759$ . Les seuils de rejet figurent sur la ligne tau3. La statistique de test

$t_{ADF}$  est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10%. On ne peut pas rejeter l'hypothèse nulle de racine unitaire.

## Test de stationnarité de KPSS

**Test de stationnarité autour d'une constante** On applique le test de stationnarité de KPSS. On présente les résultats du test avec une constante. La représentation graphique de la série et les conclusions du test ADF montrent que la taux de chômage ne contient pas de tendance déterministe. On ne rejette pas l'hypothèse nulle de stationnarité autour d'une constante.

```
kpss_mu<-ur.kpss(y=df_us_20Q1$taux,type="mu",lags="short")
summary(kpss_mu)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: mu with 4 lags.
##
## Value of test-statistic is: 0.306
##
## Critical value for a significance level of:
##      10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```