Package vars

Paul GUILLOTTE & Jules CORBEL

01/02/2019

Abstract

bbbbb

Contents

Introduction

1 Visualisation des séries

1.1 Rappel sur la stationnarité du second ordre

Avant de commencer à analyser les séries, nous rappelons des bases sur des notions dont nous aurons besoin par la suite.

Dans de nombreux modèles de séries temporelles, la série en entrée doit satisfaire une hypothèse de stationnarité. Les conditions de la stationnarité du second ordre.

$E[y_t] = \mu \forall t = 1...T$

$Var[y_t] = \sigma^2 \neq \infty \forall t = 1...t Var[y_t] = \sigma^2 \neq \infty \forall i = 1...T$

$Cov[y_i, Z_{i-k}] = f(k) \forall i = 1...t, \forall k = 1...t$

Nous nous intéressons dans cette partie aux différentes séries trimestrielles à notre disposition. Dans un premier temps, nous nous intéressons aux corrélations entre les variables deux à deux afin de nous faire une première idée du lien qu'il existe entre les variables.

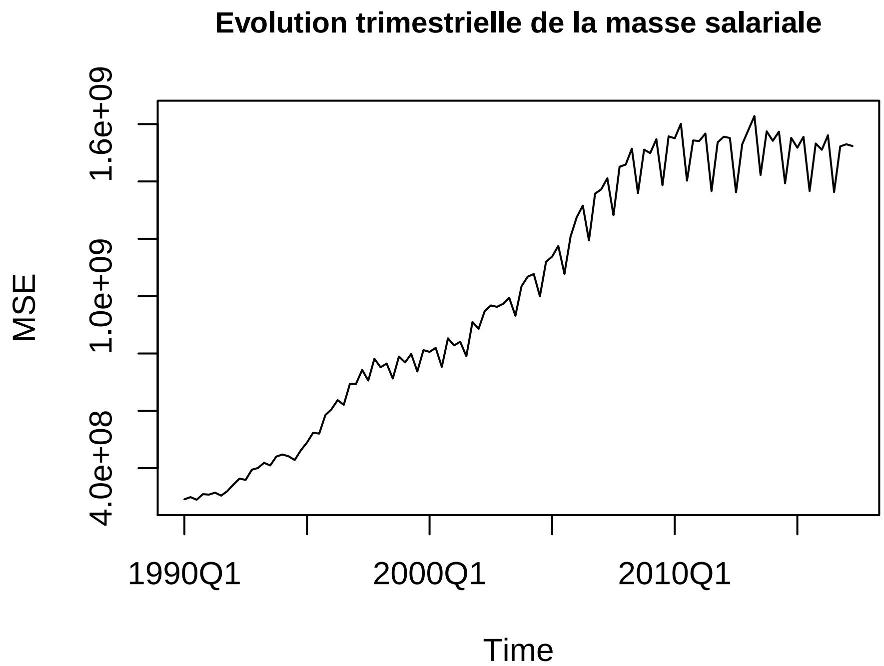

Figure 1:

1.2 Masse salariale

```
MSE <- ts(trim$MSE, start = 1990, end = c(2017, 2), frequency=4)
plot(MSE, main="Evolution trimestrielle de la masse salariale", xaxt="n", cex.main=0.9)
axis(side=1, at=seq(1990,2015,5), labels=c("1990Q1", "1995Q1", "2000Q1", "2005Q1",
                                           "2010Q1", "2015Q1"))
```

```
par(mfrow=c(1,2), cex.main=0.8)
acf(MSE, main="Auto-corrélation de la
    masse salariale trimestrielle", lag.max=20)
pacf(MSE, main="Autocorrélation partielle
     de la masse salariale trimestrielle", lag.max=20)
```

```
kpss.test(MSE)
```

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

```
##
##  KPSS Test for Level Stationarity
##
## data:  MSE
## KPSS Level = 3.6772, Truncation lag parameter = 2, p-value = 0.01
```

```
adf.test(MSE)
```

```
## Warning in adf.test(MSE): p-value greater than printed p-value
```

```
##
```

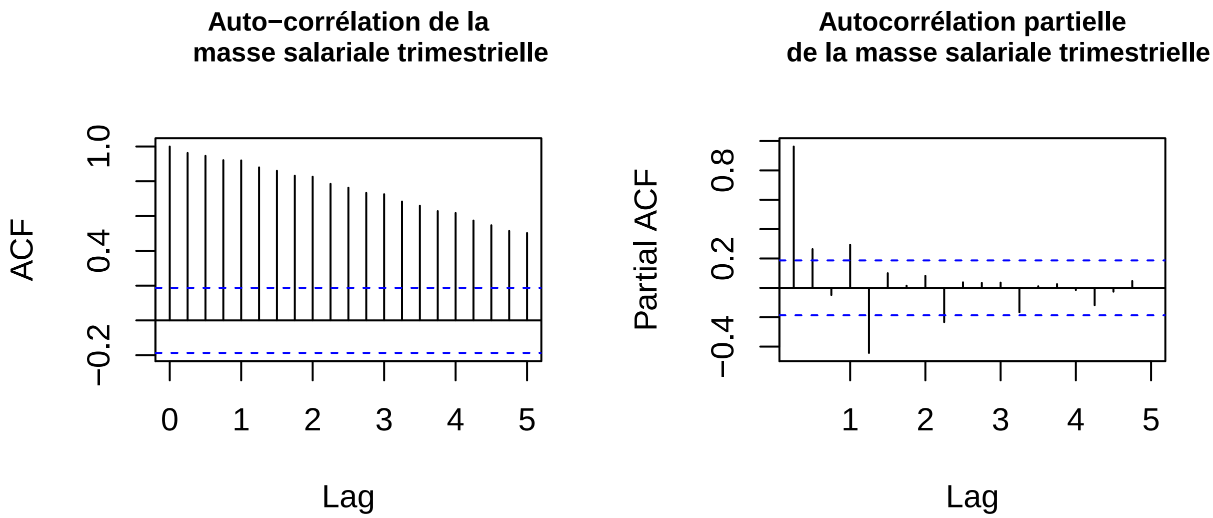

Figure 2:

```
##  Augmented Dickey-Fuller Test
##
## data:  MSE
## Dickey-Fuller = -0.20821, Lag order = 4, p-value = 0.99
## alternative hypothesis: stationary
```

La masse salariale trimestrielle, représentée en Figure 1 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 Figure 2, qui nous montre fonction ACF qui décroît régulièrement. Nous effectuons également un test de KPSS (test de stationnarité) servant à vérifier si la série est stationnaire ou non (sous l'hypothèse H_0 la série est stationnaire, et sous l'hypothèse H_1 elle ne l'est pas). La série est dite stationnaire si ses propriétés statistiques (espérance, variance et auto-corrélation) sont fixes au cours du temps. La p-value est de 0.01 ce qui nous confirme que la série n'est pas stationnaire avec un risque de première espèce de 5%. Nous mettons également en place un test de racines unitaires, le test de Dickey Fuller augmenté. Son hypothèse nulle est que la série a été générée par un processus présentant une racine unitaire, et donc que la série n'est pas stationnaire. Ici, avec un risque de premier espèce à 5%, on conserve l'hypothèse nulle est on conclut, à l'aide des deux tests effectués, que la série n'est pas stationnaire.

1.3 PIB

```
PIB <- ts(trim$PIB, start = 1990, end = c(2017, 1), frequency=4)
plot(PIB, main="Evolution trimestrielle du PIB",xaxt="n", cex.main=0.9)
axis(side=1, at=seq(1990,2015,5), labels=c("1990Q1", "1995Q1", "2000Q1", "2005Q1", "2010Q1",
```

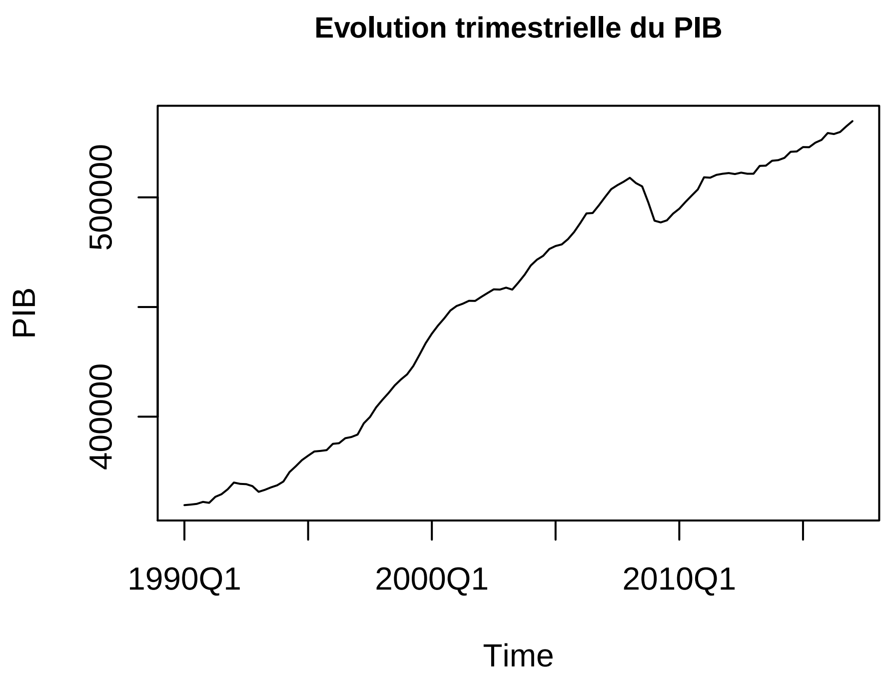

Figure 3:

```
par(mfrow=c(1,2), cex.main=0.8)
acf(PIB, main="Auto-corrélation
    du PIB trimestriel", lag.max=40)
pacf(PIB, main="Autocorrélation partielle
     du PIB trimestriel", lag.max=40)
```

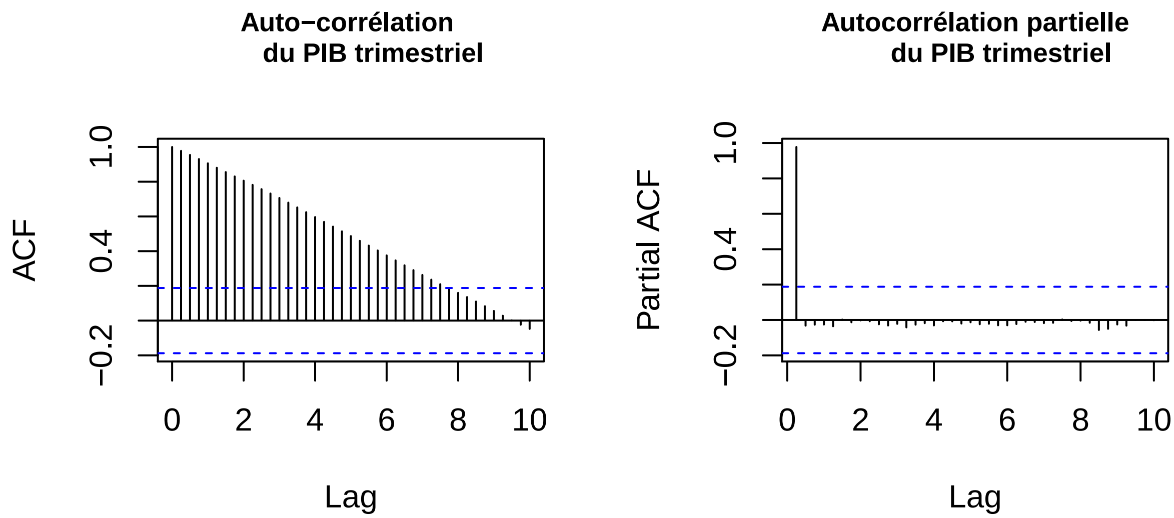

Figure 4:

```
par(mfrow=c(1,1))
kpss.test(PIB)
```

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

```
##
##  KPSS Test for Level Stationarity
##
## data:  PIB
## KPSS Level = 3.6473, Truncation lag parameter = 2, p-value = 0.01
```

```
adf.test(PIB)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  PIB
## Dickey-Fuller = -1.3274, Lag order = 4, p-value = 0.8557
## alternative hypothesis: stationary
```

La Figure 3 nous montre le PIB trimestriel qui, comme pour la masse salarialepossède une tendance. Cependant, il ne semble pas posséder de saisonnalité. Cette série ne semble donc pas non plus stationnaire. Nous effectuons à nouveau un test de KPSS. La p-value est de 0.01 ce qui nous confirme que la série n'est pas stationnaire avec un risque de première espèce de 5%. Même conclusion au regard du test augmenté de Dickey Fuller.

1.4 SMIC

```
SMIC <- ts(trim$SMIC, start = c(1990,1), end = c(2017, 4), frequency = 4)
plot(SMIC, main="Evolution trimestrielle du SMIC", xaxt="n", cex.main=0.9)
axis(side=1, at=seq(1990,2015,5), labels=c("1990Q1", "1995Q1", "2000Q1", "2005Q1", "2010Q1",
```

```
par(mfrow=c(1,2), cex.main=0.8)
acf(SMIC, main="Auto-corrélation du
    SMIC trimestriel", lag.max=20)
pacf(SMIC, main="Autocorrélation partielle
     du SMIC trimestriel", lag.max=20)
```

```
par(mfrow=c(1,1))
kpss.test(SMIC)
```

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

```
##
##  KPSS Test for Level Stationarity
##
## data:  SMIC
## KPSS Level = 3.8382, Truncation lag parameter = 2, p-value = 0.01
```

```
adf.test(SMIC)
```

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

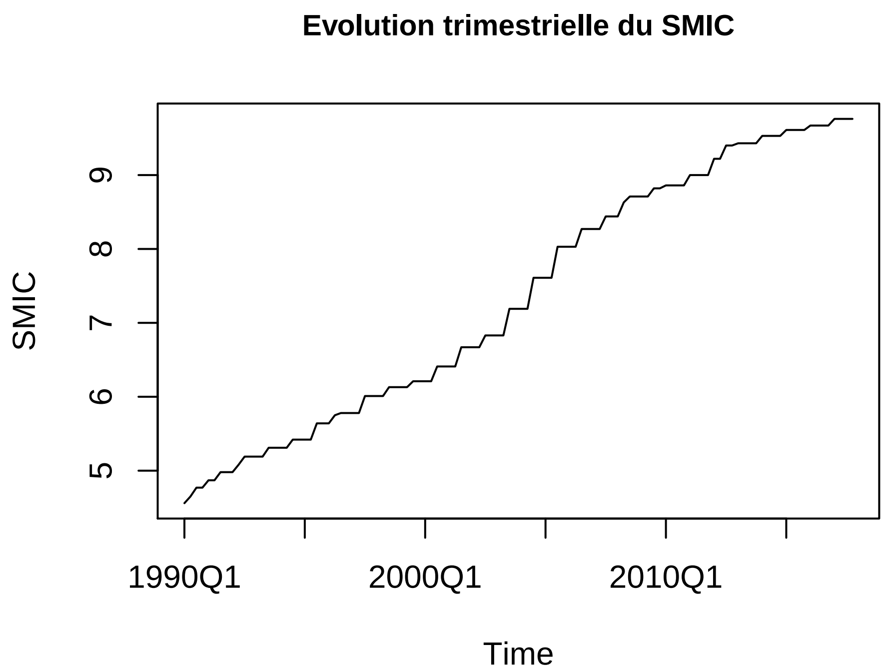

Figure 5:

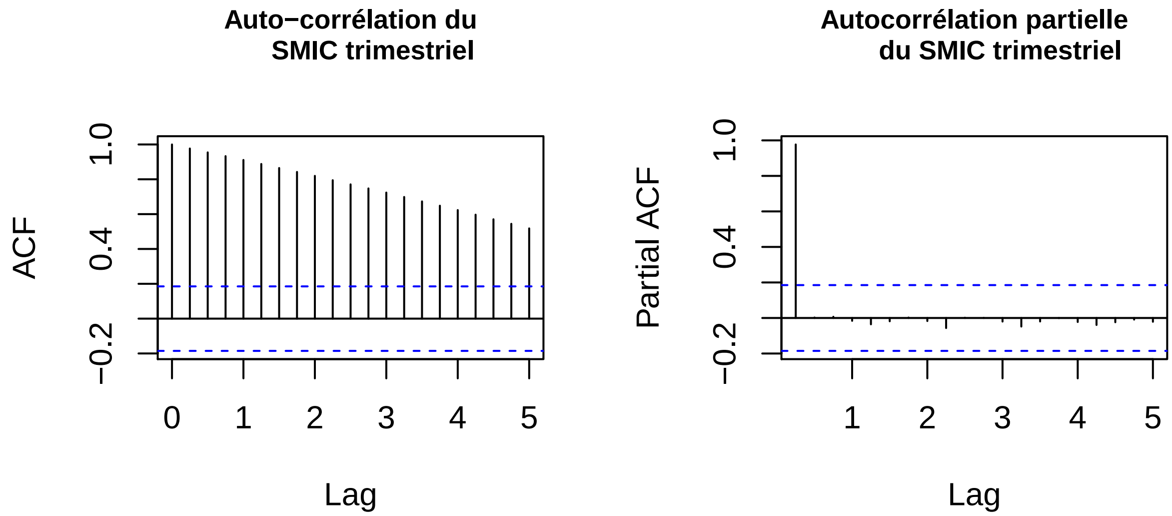

Figure 6:

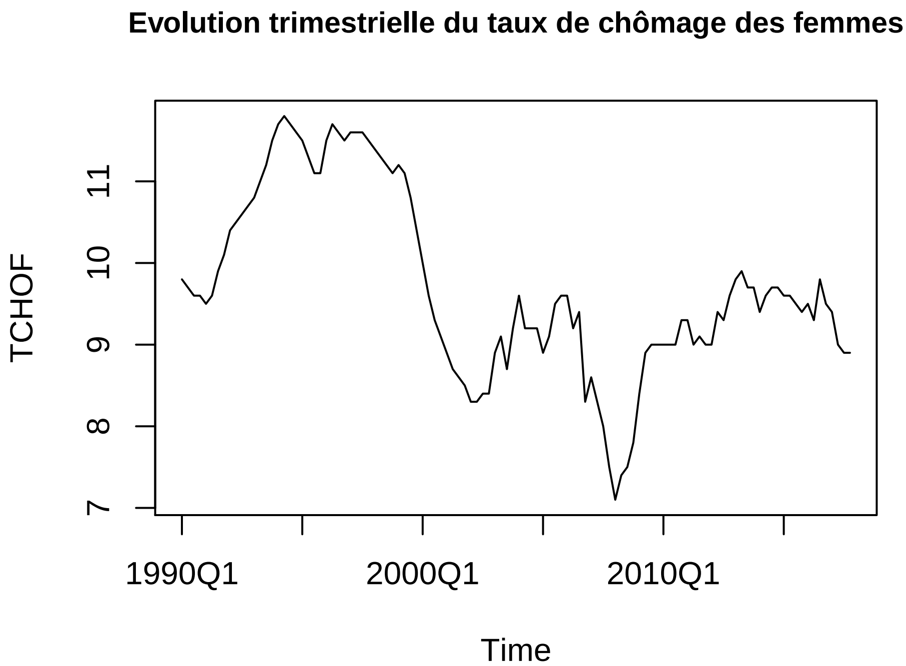

Figure 7:

```
##
## data:  SMIC
## Dickey-Fuller = -1.4174, Lag order = 4, p-value = 0.8184
## alternative hypothesis: stationary
```

Au regard de la Figure 5, 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. Les tests de KPSS et de Dickey Fuller augmenté nous confirment que la série n'est pas stationnaire.

1.5 Taux de chômage des femmes

```
TCHOF <- ts(trim$TCHOF, start = c(1990,1), end = c(2017, 4), frequency = 4)
plot(TCHOF, main="Evolution trimestrielle du taux de chômage des femmes", xaxt="n", cex.main=
axis(side=1, at=seq(1990,2015,5), labels=c("1990Q1", "1995Q1", "2000Q1", "2005Q1", "2010Q1",
```

```
par(mfrow=c(1,2), cex.main=0.8)
acf(TCHOF, main="Auto-corrélation du taux de
    chômage des femmes trimestriel", lag.max=20)
pacf(TCHOF, main="Autocorrélation partielle du
     taux de chômage des femmes trimestriel", lag.max=20)
```

```
par(mfrow=c(1,1))
kpss.test(TCHOF)
```

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

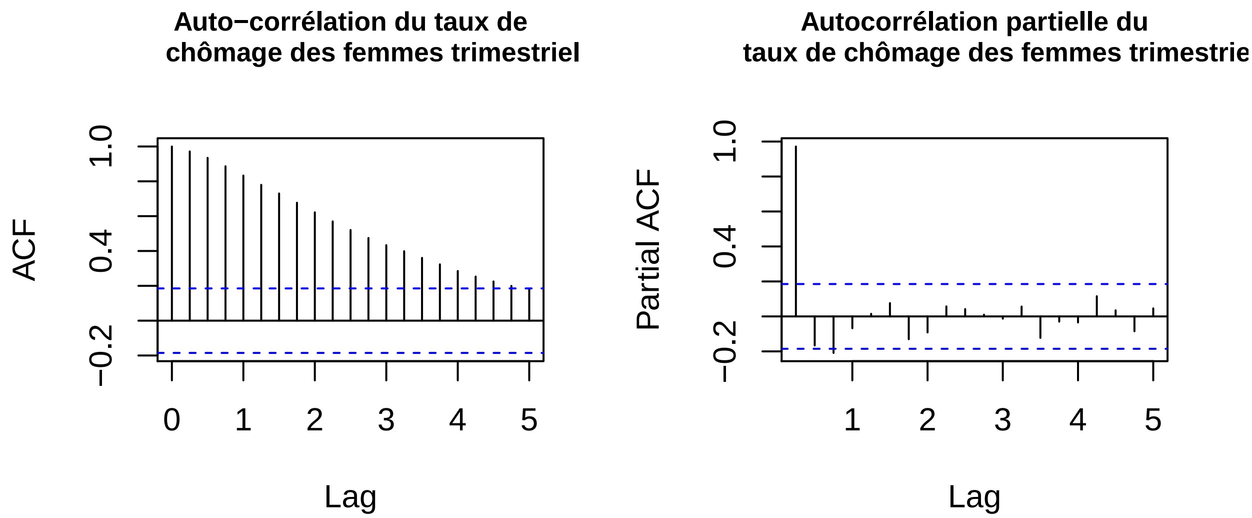

Figure 8:

```
##
##  KPSS Test for Level Stationarity
##
## data:  TCHOF
## KPSS Level = 1.6407, Truncation lag parameter = 2, p-value = 0.01
```

```
adf.test(TCHOF)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  TCHOF
## Dickey-Fuller = -2.5838, Lag order = 4, p-value = 0.3344
## alternative hypothesis: stationary
```

Pour cette dernière série (Figure 7) 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, au regard de la Figure 8. En regardant la série de plus près, on s'aperçoit que la tendance semble être "par morceaux" : d'abord une hausse de 1990 à 1996, puis elle décroît jusqu'en 2002, avant d'augmenter à nouveau jusqu'en 2007, de chuter jusqu'en 2010. Si la série ne possède pas une tendance uniforme sur toute la durée étudiée, elle semble donc bien posséder une tendance par morceaux. Les tests KPSS et de Dickey Fuller augmenté nous confirment que la série n'est pas stationnaire, avec un risque de première espèce de 5%.

1.6 Calcul des corrélations

```
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")
```

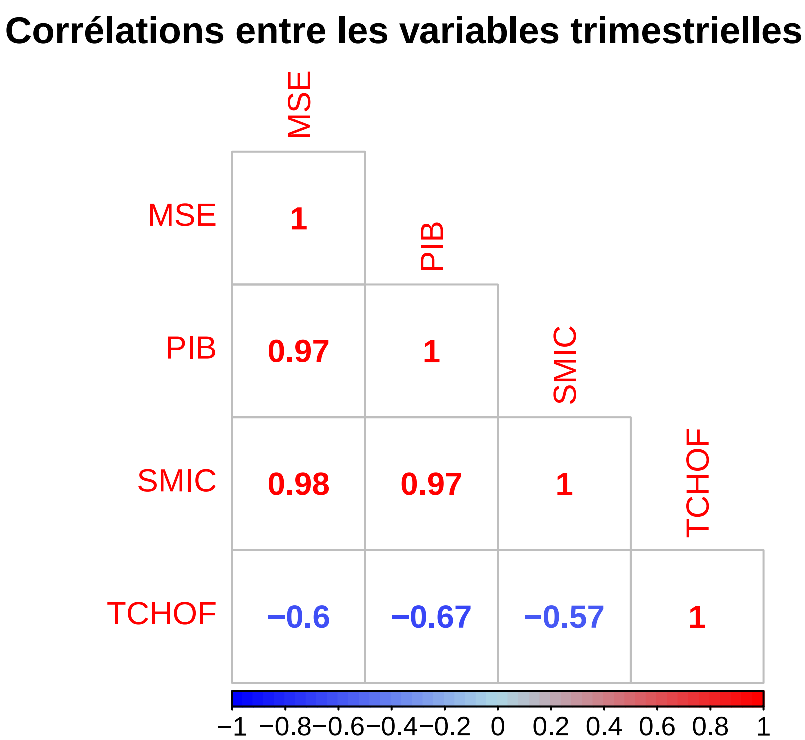

Figure 9:

```
corr <- cor.mtest(trim[1:109,-1], 0.95)[[1]]
rownames(corr) <- c("MSE","PIB","SMIC","TCHOF")
colnames(corr) <- c("MSE","PIB","SMIC","TCHOF")
corr
```

```
##                MSE          PIB         SMIC        TCHOF
## MSE   0.000000e+00 3.851955e-69 1.436967e-74 3.321841e-12
## PIB   3.851955e-69 0.000000e+00 1.898200e-71 2.387179e-15
## SMIC  1.436967e-74 1.898200e-71 0.000000e+00 1.377731e-10
## TCHOF 3.321841e-12 2.387179e-15 1.377731e-10 0.000000e+00
```

Nous affichons la matrice des corrélations des différentes variables en Figure 9. On se rend compte que le taux de chômage des femmes est corrélé négativement avec toutes les autres variables. Le trio de variables PIB, masse salariale et SMIC sont extrêmement liées entre elles. En regardant le tableau des p-values associées au test de Student (H0 : La corrélation entre les deux variables est nulle), on s'aperçoit que toutes les variables prises deux à deux présentent une corrélation.

2 Modélisation individuelle

Une fois que nous avons analysé le comportement des différentes séries temporelles à notre disposition, nous souhaitons les modéliser afin de prédire les valeurs futures de ces différentes séries. En effet, si nous voulons prédire la MSE pour des valeurs futures, nous aurons également besoin des valeurs associées pour les variables explicatives, qui ne seront peut-être pas à notre disposition. Nous avons utilisé à la fois des modèles basés sur un lissage exponentiel et des processus ARMA.

2.1 Découpage des séries

Pour chacune des séries, nous allons créer un échantillon d'apprentissage, qui nous permettra de construire les différents modèles, ainsi qu'un échantillon de test, qui nous permettra de comparer les prédictions des modèles construits avec des vraies valeurs. L'échantillon d'apprentissage sera composé de toutes les valeurs du premier trimestre 1990 jusqu'au 4e trimestre 2015, tandis que celui de test comprendra toutes les valeurs à partir du 1er trimestre 2016.

```
MSETrain <- window(MSE, start=1990, end=c(2015,4))
MSETest <- window(MSE, start=2016)
PIBTrain <- window(PIB, start=1990, end=c(2015,4))
PIBTest <- window(PIB, start=2016)
SMICTrain <- window(SMIC, start=1990, end=c(2015,4))
SMICTest <- window(SMIC, start=2016)
TCHOFTrain <- window(TCHOF, start=1990, end=c(2015,4))
TCHOFTest <- window(TCHOF, start=2016)
```

2.2 Lissage exponentiel

###Définition Le lissage exponentiel permet de prédire les valeurs d'une série temporelle en lissant successivement les données à partir d'une valeur initiale. Plus les observations sont éloignées dans le passé, moins leur poids est important lors du calcul. Pour une série stationnaire, la formule de calcul d'une valeur est la suivante : $s_t = \alpha y_t + (1 - \alpha)s_{t-1}$, le paramètre α étant le facteur de lissage. Le nom de cette méthode est un lissage exponentiel **simple**. Afin de modéliser les séries possédant une tendance, nous introduisons un paramètre β permettant de la prendre en compte, la méthode étant appelée lissage exponentiel **double**. Enfin, Holt et Winters ont également modifié la méthode pour qu'elle puisse modéliser les séries comportant une saisonnalité en introduisant un paramètre γ. Ils ont donné leur nom à cette méthode, qui est donc un lissage exponentiel de **Holt-Winters**.

Dans notre cas, nous ne calculons pas nous-mêmes α β et γ Ces paramètres sont déterminés automatiquement par la fonction *ets* du package **forecast** de façon à optimiser la qualité de la prédiction. Cette fonction permet également de choisir la valeur initiale permettant d'obtenir le meilleur modèle.

2.3 Résultats obtenus

Nous résumons dans le