

Économétrie des séries temporelles : projet 1

Arno Barrabès - M1 MAS-IREF

4 avril 2022

Contents

Introduction	2
Étapes préliminaires	2
Rappels sur la stationnarité et les racines unitaires	2
Visualisation	3
Autocorrélation	3
Tests de racines unitaires	5
Dickey-Fuller (DF)	5
Dickey-Fuller augmenté (ADF)	6
Zivot et Andrews (ZA)	9
Lee et Strazicich (LS)	11
Prévisions	12
Série différenciée	12
Modèle ARMA(p, q)	17
EACF	17
Auto.arima	17
Fonction personnelle	18
Bruits blancs	19
Test deJarque-Bera	19
Test de Student	19
Test de Ljung-Box	20
Test d'Engle	21
Prévisions et visualisation	21

Introduction

Le but de ce projet est de travailler sur la série temporelle du taux d'intérêt des obligations de long terme en France, de 1960 à 2021, et de proposer une prédiction annuelle sur 3 ans de sa valeur future.

Il est important d'avoir une idée précise du niveau des taux d'intérêts, en particulier ceux d'un état, pour répondre au besoin d'anticipation qu'éprouvent les investisseurs. En effet, avoir une bonne prédiction de la valeur future du taux permet une meilleure connaissance du marché et permet donc d'appliquer une meilleure stratégie d'investissement.

Pour ce faire, nous avons récupéré sur le site de l'OCDE les valeurs annuelles du taux d'intérêt de long terme sur les 62 dernières années et nous effectuerons une analyse économétrique de cette série temporelle : suite à différents tests de racines unitaires et à une différenciation de la série pour la rendre stationnaire, nous finirons par construire un modèle le plus exact possible qui fournira la valeur estimée du taux d'intérêt de long terme pour chacune des 3 prochaines années.

Étapes préliminaires

Rappels sur la stationnarité et les racines unitaires

Tout le long de ce projet nous prendrons un risque de première espèce $\alpha = 5\%$. Le risque de première espèce d'un test statistique est le risque de rejeter l'hypothèse nulle alors qu'elle est vraie.

Un processus stochastique est stationnaire si son espérance ne varie pas dans le temps, sa variance est définie et l'autocorrélation ne dépend pas du temps. Par exemple, $X_{t,t} \in Z$ est un processus stochastique (une suite de variables aléatoires réelles fonction du temps) stationnaire au second ordre si :

- Son espérance est indépendante du temps : $\forall t \in Z, E(X_t) = \mu$, constante,
- Sa variance est finie : $\forall t \in Z, Var(X_t) < \infty$,
- Son autocovariance au retard h est indépendante de t : $\forall t, h \in Z, Cov(X_t, X_{t+h}) = E[(X_t - m)(X_{t+h} - m)] = \gamma(h)$.

Ainsi, il faut s'assurer de la stationnarité de notre série avant toute prédiction.

Pour cela, nous effectuerons les 4 tests de racine unitaire suivants :

- **Dickey-Fuller (DF)**
- **Dickey-Fuller Augmenté (ADF)**
- **Zivot-Andrews (ZA)**
- **Lee et Strazicich (LS).**

Ces tests nous permettent de déterminer si le PGD (processus générateur de données) est stationnaire, DS ou TS. Dans le cas où le PGD n'est pas stationnaire (TS ou DS), il faudra rendre notre série stationnaire pour pouvoir faire de la prévision.

Processus TS: $X_t = \mu + \delta \text{tendance}_t + u_t$ avec $X_t = \mu + \delta \text{tendance}_t$ la partie déterministe et u_t la partie aléatoire, processus ARMA stationnaire.

Caractéristiques d'une série TS :

- Son espérance dépend du temps : $E(X_t) = \mu + \delta \text{tendance}_t$.
- Sa variance ne dépend pas du temps : $Var(X_t) = \sigma_u^2$.

Processus DS : $X_t = \delta + X_{t-1} + u_t$ avec u_t un bruit blanc (BB) et δ une constante (la dérive).

Caractéristiques d'une série DS :

- Son espérance dépend du temps : $E(X_t|X_0) = X_0 + \delta \text{tendance}_t$.

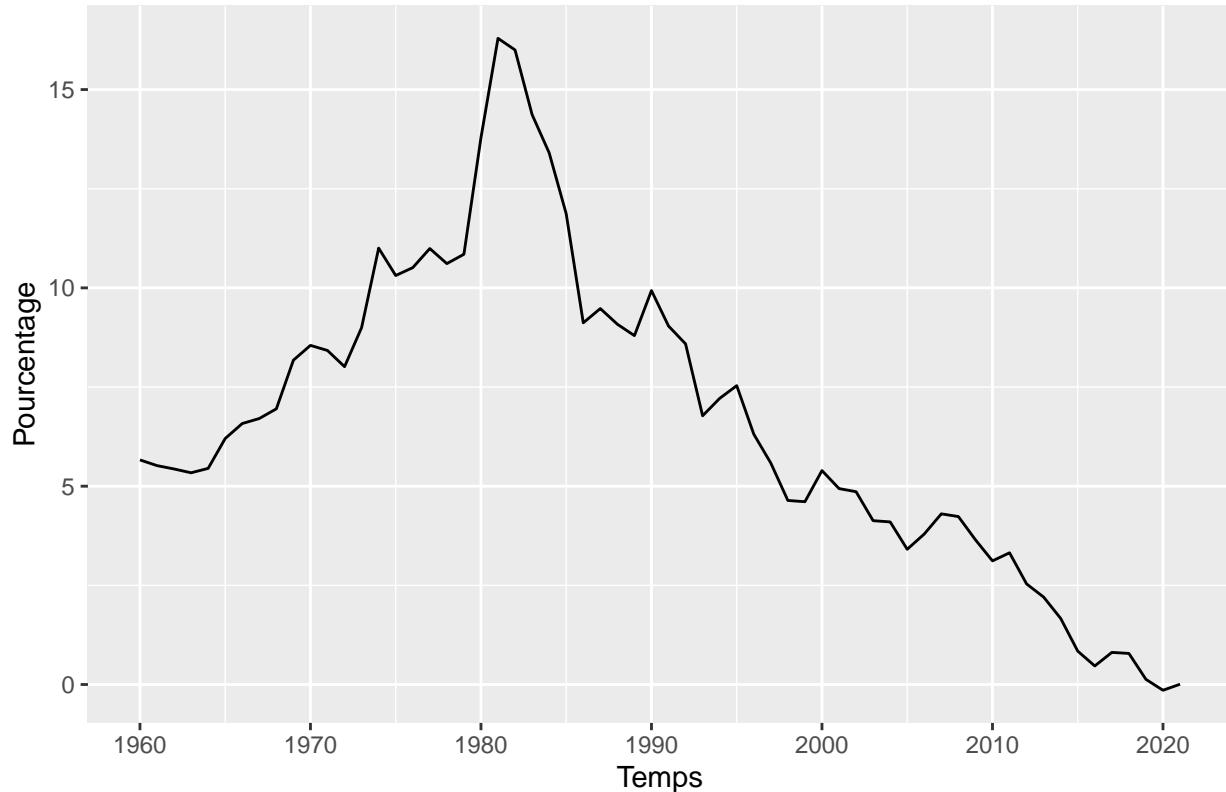
- Sa variance dépend du temps : $Var(X_t|X_0) = \text{tendance}_t \sigma_u^2$.

Pour rendre une série DS stationnaire, on la différencie. On obtient alors : $X_t^* = \Delta X_t - X_{t-1}$ avec X_t^* stationnaire.

Visualisation

Une fois les données chargées et préparées (sélection de la date initiale et du pays), notre premier réflexe est de visualiser la série sous forme de chronogramme. Le chronogramme obtenu est le suivant :

Évolution du taux d'intérêt de long terme en France de 1960 à 2021



On identifie deux phases : une première de 1960 à 1981 où le taux d'intérêt de long terme est globalement croissant, une seconde de 1982 à 2021 où il décroît globalement.

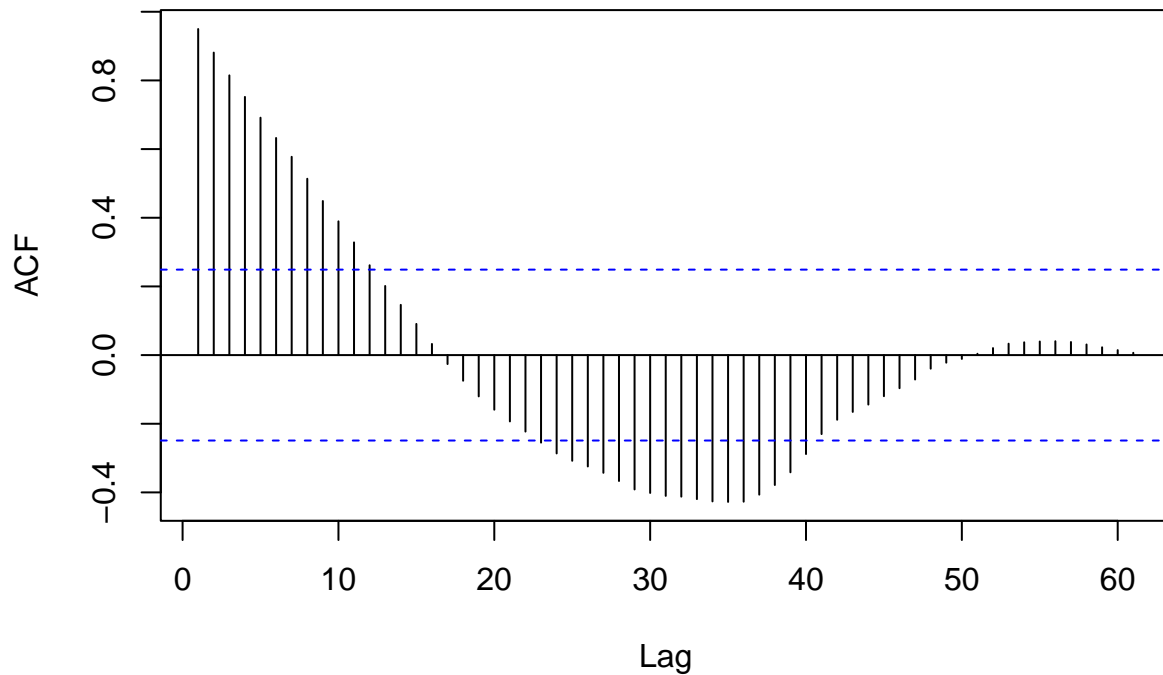
La première phase correspond à la période de fort développement économique des Trente Glorieuses, marquée cependant légèrement par le premier choc pétrolier en 1973, la seconde lorsque la politique monétaire déflationniste est mise en œuvre, peu après le second choc pétrolier en 1979. Contrairement à mon camarade et son taux de court terme, le taux d'intérêt de long terme est moins sensible aux "petits" événements historiques et économiques qu'est par exemple la bulle immobilière.

On remarque une tendance globale décroissante sur la période. Notre série ne semble donc pas stationnaire. Sur les 4 dernières années, la valeur du taux d'intérêt semble se stabiliser autour de 0%, nous pouvons donc nous attendre à prédire 3 valeurs très proches de 0%.

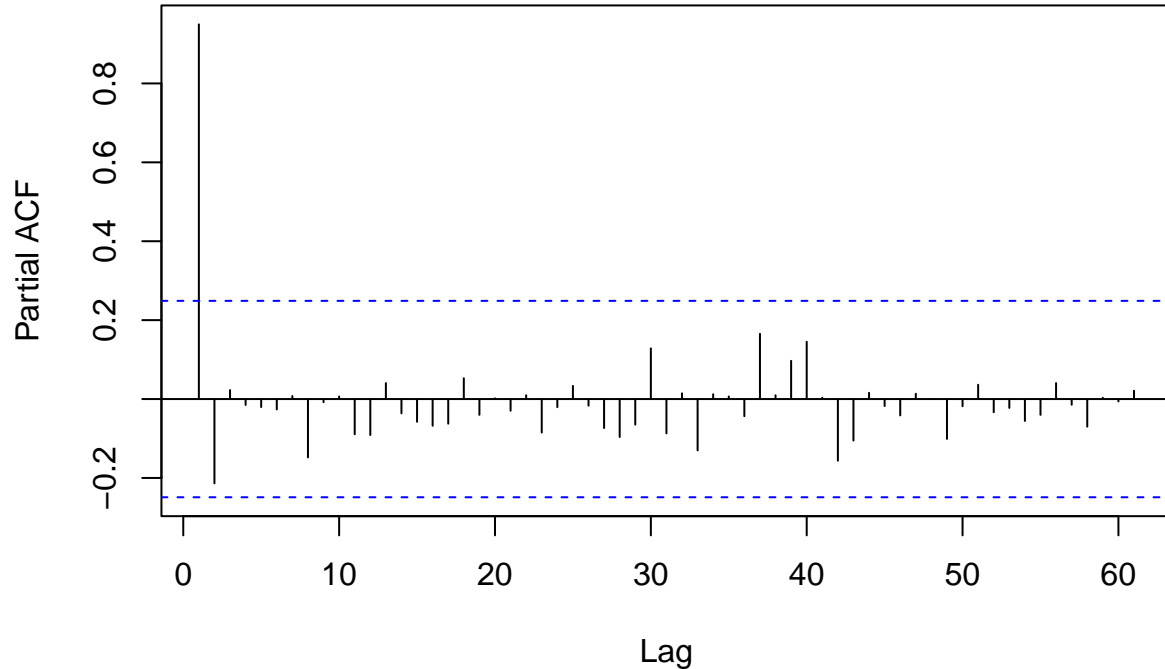
Autocorrélation

Intéressons-nous aux autocorrélations totale et partielle de notre série :

Autocorrélation de la série intiale



Autocorrélation de la série intiale



Avec la fonction d'autocorrélation totale (ACF), nous remarquons qu'il y a plusieurs ρ qui sont significatifs, car ils sont compris dans l'intervalle de confiance à 95%. Néanmoins, l'autocorrélation à l'ordre 1 est très élevée, car $\rho(1) \approx 1$. La fonction d'autocorrélation partielle (PACF), qui permet de mesurer l'autocorrélation pour un décalage k indépendamment des autocorrélations pour les décalages inférieurs, révèle que seul $a(1)$ est significatif : les autocorrélations indiquées par l'ACF sont en fait conséquences de l'autocorrélation existant à

l'ordre 1.

Ainsi, on doit prendre en compte l'autocorrélation dans notre régression.

Tests de racines unitaires

Déterminons la présence de racines unitaires à l'aide des différents tests évoqués plus tôt.

Dickey-Fuller (DF)

Il y a trois versions du test DF :

- La spécification “trend” : avec drift et trend $\Delta y_t = \beta_0 + \beta_1 t + \delta * y_{t-1} + \mu_t$.
- La spécification “drift” : avec drift mais sans trend $\Delta y_t = \beta_0 + \delta * y_{t-1} + \mu_t$.
- La spécification “none” : sans drift ni trend $\Delta y_t = \delta * y_{t-1} + \mu_t$.

Nous commencerons par tester la spécification “trend”, la plus complète.

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4369 -0.5040 -0.1278  0.4416  3.0570
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.983465   0.470020   2.092   0.0408 *
## z.lag.1      -0.063020   0.037389  -1.685   0.0973 .
## tt           -0.021010   0.008226  -2.554   0.0133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.89 on 58 degrees of freedom
## Multiple R-squared:  0.1014, Adjusted R-squared:  0.07038
## F-statistic: 3.271 on 2 and 58 DF,  p-value: 0.04508
##
##
## Value of test-statistic is: -1.6855 2.4014 3.2711
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

On pose :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{cases} .$$

Nous comparons la p-valeur associée au coefficient β_1 au seuil critique de 5%. La p-valeur est égale à $0.0133 < 5\%$ donc on rejette H_0 et β_1 est significatif.

La spécification trend est la bonne et c'est celle que l'on gardera tout le long de cette première partie.

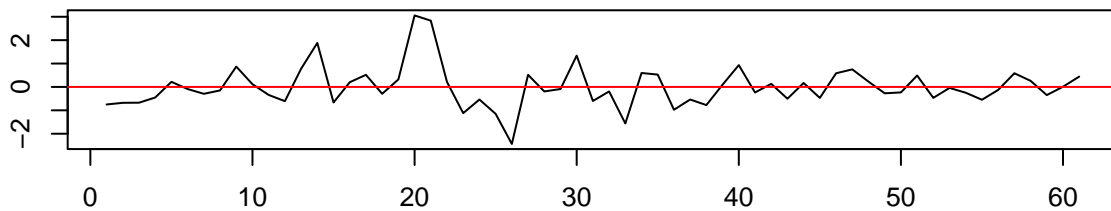
On pose ensuite le test sur le coefficient de période de retard :

$$\begin{cases} H_0 : (\rho - 1) = 0 \\ H_a : (\rho - 1) \neq 0 \end{cases} .$$

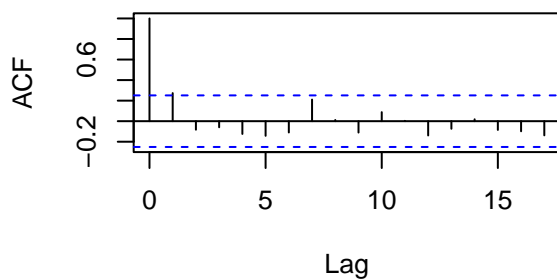
Règle de décision : si la statistique t calculée pour $(\rho - 1)$ est inférieure à la valeur critique tabulée à l'intersection de la ligne tau3 et de la colonne 5%, on rejette H_0 . Dans notre cas, la valeur de la statistique vaut $-1.685 > -3.45$ donc on ne rejette pas H_0 et il y a présence de racines unitaires.

La conclusion du test DF est que le processus qui a généré notre série est DS. Néanmoins, on ne peut valider les résultats obtenus pour le test DF seulement s'il n'y a pas d'autocorrélation dans les résidus ; vérifions-le.

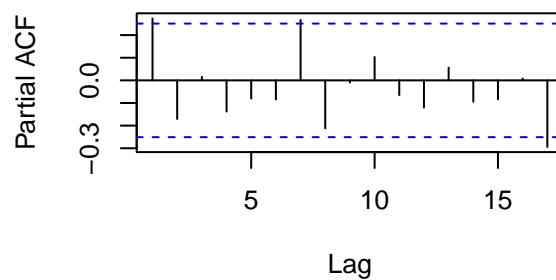
Residuals



Autocorrelations of Residuals



Partial Autocorrelations of Residuals



On constate que les résidus sont autocorrélés. La conclusion du test DF est donc potentiellement erronée et nous n'allons pas la prendre en compte. Afin de prendre en considération cette autocorrélation, nous allons effectuer cette fois-ci le test de Dickey-Fuller augmenté sur notre série.

Dickey-Fuller augmenté (ADF)

Le test ADF correspond à un test DF avec des variables explicatives supplémentaires représentant la variable retardée jusqu'à l'ordre P , le nombre de retards optimal pour tenir compte de l'autocorrélation dans les aléas.

Pour notre spécification “trend”, nous obtenons :

$$\Delta X_t = (\rho - 1)X_{t-1} + \beta_0 + \beta_1 t + \epsilon_t + \sum_{p=1}^P \gamma_p \Delta X_{t-p},$$

où P se calcule à l’aide de p_{\max} fourni par la formule de Schwert : $12 * (\frac{T}{100})^{0.25}$.

Dans notre cas, la formule donne :

$$p_{\max} = \lfloor 12 \times (\frac{62}{100})^{\frac{1}{4}} \rfloor = 10.$$

Nous minimisons ensuite le critère d’information de Ng et Perron à l’aide du MAIC(p), calculé pour les valeurs de p allant de 0 à p_{\max} .

```
## Augmented DF test
##                                     ADF test
## t-test statistic:                 -2.5126480
## p-value:                         0.3210708
## Max lag of the diff. dependent variable: 0.0000000
##
## Call:
## dynlm(formula = formula(model), start = obs.1, end = obs.T)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2976 -0.4965 -0.1235  0.4577  3.0056
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.78284    1.06442   2.614  0.01191 *
## trnd         -0.05065    0.01790  -2.830  0.00678 **
## L(y, 1)      -0.15718    0.06256  -2.513  0.32107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9232 on 48 degrees of freedom
## Multiple R-squared:  0.1431, Adjusted R-squared:  0.1074
## F-statistic:      NA on NA and NA DF, p-value: NA
```

“Max lag of the diff.dependent variable” correspond au nombre de variables explicatives à ajouter pour prendre en compte l’autocorrélation. Ici, ce nombre est 0. On doit donc employer le critère BIC.

```
## Augmented DF test
##                                     ADF test
## t-test statistic:                 -2.8748684
## p-value:                         0.1790089
## Max lag of the diff. dependent variable: 1.0000000
##
## Call:
## dynlm(formula = formula(model), start = obs.1, end = obs.T)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.79272 -0.53872  0.05149  0.33343  2.86602
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.93395    1.01951   2.878  0.00601 **
## trnd         -0.05045    0.01711  -2.949  0.00496 **
## L(y, 1)      -0.17299    0.06017  -2.875  0.17901
## L(d(y), 1)   0.30814    0.13103   2.352  0.02293 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8825 on 47 degrees of freedom
## Multiple R-squared:  0.2334, Adjusted R-squared:  0.1844
## F-statistic:    NA on NA and NA DF,  p-value: NA
```

Dans ce cas, “Max lag of the diff.dependent variable” = 1. On teste alors :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{cases}.$$

Règle de décision : si la p-valeur est inférieure à 5%, on rejette H_0 , et si β_1 est significatif alors on teste :

$$\begin{cases} H_0 : (\rho - 1) = 0 \\ H_a : |\rho| < 1 \end{cases}.$$

Règle de décision : si la statistique t calculée pour $(\rho - 1)$ est inférieure à la valeur critique tabulée à l’intersection de la ligne tau3 et de la colonne 5%, on rejette H_0 .

```
##
## #####
## # 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
## -1.95054 -0.52269 -0.09543  0.43014  3.01521
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.112391   0.477870   2.328  0.0236 *
## z.lag.1      -0.078212   0.037158  -2.105  0.0398 *
## tt          -0.020553   0.008432  -2.437  0.0180 *
## z.diff.lag    0.290637   0.124713   2.330  0.0234 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8588 on 56 degrees of freedom
## Multiple R-squared:  0.192, Adjusted R-squared:  0.1488
## F-statistic: 4.437 on 3 and 56 DF,  p-value: 0.007229
```



```
##
##
## Value of test-statistic is: -2.1048 2.2732 3.2575
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

La p-valeur de tt est significative donc on rejette H_0 dans notre premier test et la spécification trend est correcte. De plus, la statistique t de $z.diff.lag$ vaut $2.330 > 1.6$ donc γ_1 est significatif et il faut bien ajouter une variable explicative. Enfin, la statistique t de $z.lag.1$ vaut -2.105 et est supérieure à la valeur tabulée -3.45 , donc on ne rejette pas H_0 dans notre second test et il y a présence de racines unitaires et le PGD est DS.

Cependant, les résultats des tests DF et ADF ne sont valides que s'il n'y a pas de changements structurels dans les données. Or, comme on a pu le voir dans le chronogramme, la fin des Trente Glorieuses, le deuxième choc pétrolier et la mise en place d'une politique déflationniste a grandement impacté la valeur du taux d'intérêt de long terme. Ainsi, nous allons appliquer le test de Zivot et Andrews afin d'être sûr de la présence d'un tel changement structurel dans notre série.

Zivot et Andrews (ZA)

Pour modéliser un changement structurel, il convient de créer quelques variables. Soient T_B la date du changement structurel, $DU_t = \beta_0$ la variable modélisant un changement dans le niveau de la partie déterministe de la série et $DT_t = \beta_1$ la variable modélisant un changement dans la pente de la partie déterministe de la série.

Il existe 2 spécifications :

- La spécification “both”, qui prend en compte le niveau de la série et son taux de croissance :

$$\Delta X_t = \beta_0 + \beta_1.tendance_t + (\rho - 1).X_{t-1} + \delta_1.DU_t(T_B) + \delta_2.DT_t(T_B) + \sum_{p=1}^P \gamma_p \Delta X_{t-p} + \epsilon_t.$$

- La spécification “crash”, qui prend seulement le niveau de la série en compte :

$$\Delta X_t = \beta_0 + \beta_1.tendance_t + (\rho - 1).X_{t-1} + \delta_1.DU_t(T_B) + \sum_{p=1}^P \gamma_p \Delta X_{t-p} + \epsilon_t.$$

On teste ici :

$$\begin{cases} H_0 : \rho = 1, \text{ processus DS avec } y_t = y_{t-1} + \epsilon_t \text{ ou } y_t = y_{t-1} + \epsilon_t + \beta_0 \\ H_a : |\rho| < 1, \text{ processus TS avec } y_t = \beta_0 + \beta_1 t + \rho \cdot y_{t-1} + \epsilon_t \end{cases}.$$

Sous H_0 , on a DS sans changement structurel alors que sous H_a on a TS avec un unique changement structurel qui survient à une date déterminée par la procédure de test, notée TB.

Nous commencerons par tester la spécification “both”, la plus complète.

```
##
## #####
## # Zivot-Andrews Unit Root Test #
## #####
##
##
## Call:
```

```
## lm(formula = testmat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26901 -0.42066 -0.00948  0.42769  1.25734
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.18928    1.52449   2.092 0.043550 *
## y.l1          0.12304    0.14302   0.860 0.395325
## trend         0.28484    0.10190   2.795 0.008264 **
## y.dl1         0.69006    0.12712   5.428 4.04e-06 ***
## y.dl2         0.32020    0.14873   2.153 0.038109 *
## y.dl3         0.37111    0.13425   2.764 0.008938 **
## y.dl4         0.24064    0.13215   1.821 0.076937 .
## y.dl5         0.33258    0.13067   2.545 0.015351 *
## y.dl6         0.01627    0.11929   0.136 0.892242
## y.dl7         0.38168    0.11537   3.308 0.002137 **
## y.dl8         0.10329    0.12935   0.799 0.429771
## y.dl9         0.06970    0.12325   0.566 0.575200
## y.dl10        0.30018    0.11776   2.549 0.015203 *
## du            2.80316    0.72820   3.849 0.000467 ***
## dt           -0.55816    0.12911  -4.323 0.000116 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6787 on 36 degrees of freedom
## (11 observations effacées parce que manquantes)
## Multiple R-squared:  0.9825, Adjusted R-squared:  0.9757
## F-statistic: 144.3 on 14 and 36 DF, p-value: < 2.2e-16
##
##
## Teststatistic: -6.1316
## Critical values: 0.01= -5.57 0.05= -5.08 0.1= -4.82
##
## Potential break point at position: 20
```

On commence par la significativité du coefficient associé à dt :

$$\begin{cases} H_0 : \delta_2 = 0 \\ H_a : \delta_2 \neq 0 \end{cases}.$$

La p-valeur de dt vaut $0.000116 < 5\%$ donc on rejette H_0 et δ_2 est significatif.

On regarde ensuite la significativité du coefficient associé à du :

$$\begin{cases} H_0 : \delta_1 = 0 \\ H_a : \delta_1 \neq 0 \end{cases}.$$

La p-valeur de du vaut $0.000467 < 5\%$ donc on rejette H_0 et δ_1 est significatif.

Par conséquent, la spécification both du test ZA est la bonne spécification pour notre série.

De plus, la t-valeur de γ_{10} vaut $2.54 > 1.6$ donc γ_{10} est significatif, donc tous les coefficients introduits sont pertinents.

Effectuons enfin le test ZA :

$$\begin{cases} H_0 : \text{La série est DS sans changement structurel} \\ H_a : \text{La série n'est pas DS sans changement structurel} \end{cases}$$

La statistique du test vaut -6.1316 , ce qui est inférieur à la valeur critique -5.08 à 5%. Ainsi, on rejette H_0 et notre série n'est pas DS sans changement structurel.

Cette conclusion peut sembler contradictoire avec le résultat des tests DF et ADF, mais il faut faire attention à la négation logique de H_a ; en effet nous ne concluons pas que notre série n'est pas DS mais qu'elle n'est pas DS sans changement structurel. Il est donc possible qu'elle soit DS avec changement structurel.

Lee et Strazicich (LS)

Pour le savoir, il convient de faire le test LS. On part du modèle (17) détaillé dans le cours et, étant donné que la spécification retenue suite au test ZA était “both”, nous allons utiliser la spécification “break” (au détriment de “crash”).

On pose donc :

$$\begin{cases} H_0 : y_t = \mu_0 + d_1 B_{1t} + d_2 B_{2t} + d_3 D_{1t} + d_4 D_{2t} + y_{t-1} + v_{1t} \\ H_a : y_t = \mu_1 + \gamma \cdot \text{trend}_t + d_1 D_{1t} + d_2 D_{2t} + d_3 DT_{1t} + d_4 DT_{2t} + v_{2t} \end{cases}$$

Comme précisé dans le cours, nous posons 2 conditions. Premièrement, afin de garantir l'efficacité du test en taille et en puissance, nous supposons l'existence d'une seule date de rupture dans notre série. Deuxièmement, vu qu'il s'agit d'un faible nombre de données, nous utilisons la méthode “bootstrap”.

On a alors :

$$\begin{cases} H_0 : \text{Le PGD qui a généré la série a une racine unitaire} \\ H_a : \text{Le PGD qui a généré la série n'a pas de racine unitaire} \end{cases}$$

```
## [[1]]
## [1] -2.747822
##
## [1] "First possible structural break at position: 18"
## [1] "The location of the first break - lambda_1: 0.3 , with the number of total observations: 62"
## Critical values - Break model:
##      lambda    1%    5%   10%
## [1,]    0.1 -5.11 -4.50 -4.21
## [2,]    0.2 -5.07 -4.47 -4.20
## [3,]    0.3 -5.15 -4.45 -4.18
## [4,]    0.4 -5.05 -4.50 -4.18
## [5,]    0.5 -5.11 -4.51 -4.17
## [1] "Number of lags used: 5"
## Runtime:
## Time difference of 0.002671214 mins
```

La valeur de la statistique est -2.75 . Le λ estimé est 0.3, pour une valeur critique au seuil de risque 5% de $-4.45 < -2.75$. Donc on ne rejette pas H_0 et il y a présence de racine unitaire. De plus, $T_{B1} = 1960 + 18 - 1 = 1977$, donc l'année de rupture est 1977.

La conclusion du test LS avec bootstrap est que le processus qui a généré notre série est DS avec une date de rupture dans la constante et dans la tendance en 1977. Donc nous gardons la conclusion que la série est DS avec un changement structurel.

Pour simple confirmation, nous utilisons la fonction `adt.test` du package `aTSA`, qui permet de tester directement la stationnarité d'une série temporelle. L'hypothèse nulle de ce test est H_0 : La série n'est pas stationnaire, versus l'hypothèse alternative H_a : La série est stationnaire.

```
##  
## Augmented Dickey-Fuller Test  
##  
## data:  taux_fr_lt  
## Dickey-Fuller = -2.2535, Lag order = 3, p-value = 0.4723  
## alternative hypothesis: stationary
```

La p-valeur de ce test vaut $0.4723 > 5\%$. Nous pouvons donc accepter l'hypothèse nulle et conclure que notre série est bien non stationnaire.

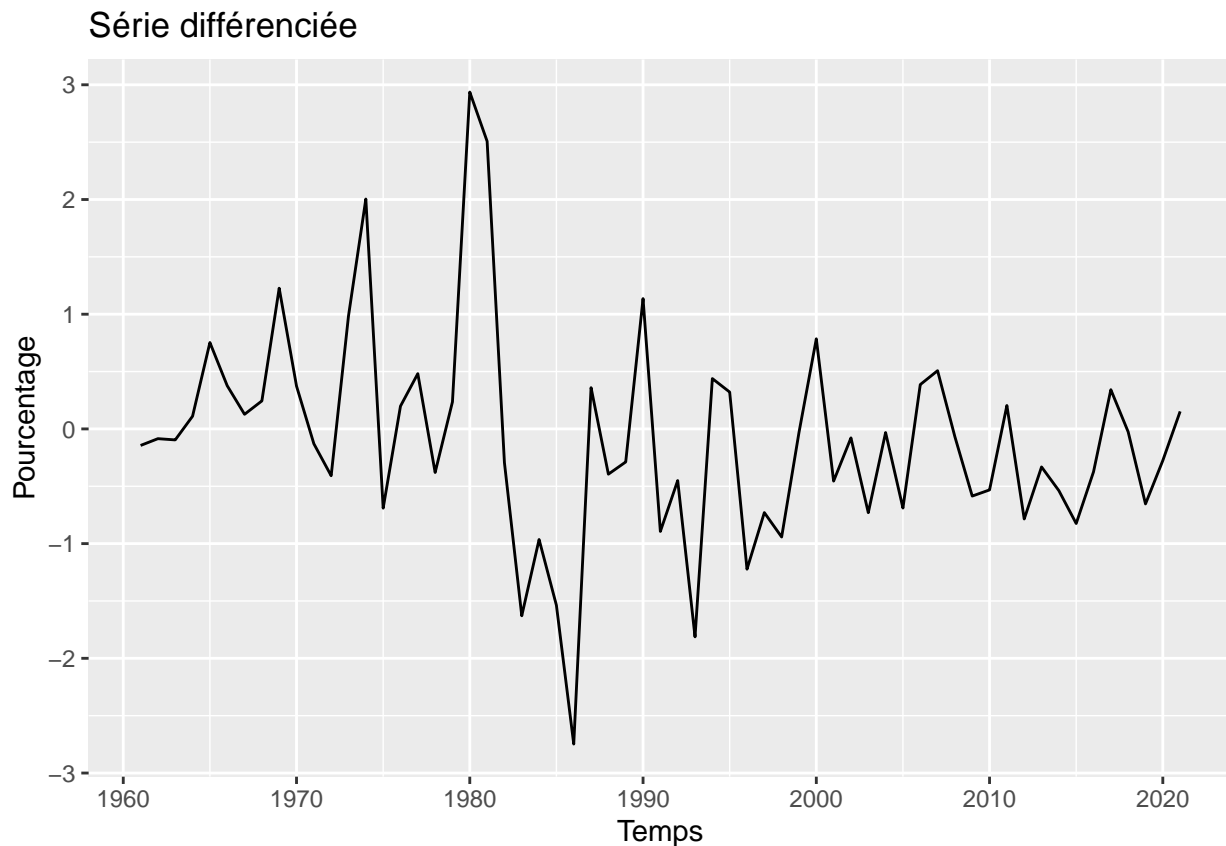
Prévisions

Comme vu au début du projet, nous allons dériver notre série afin de la rendre stationnaire : $X_t^* = \Delta X_t - X_{t-1}$, avec X_t^* stationnaire.

Série différenciée

```
d_taux_fr_lt<-diff(taux_fr_lt)
```

Visualisons le chronogramme de la série différenciée :



Notre série différenciée semble stationnaire, oscillant autour d'une valeur proche de 0. Assurons-nous en par un test DF. On rédigera moins que lors des premiers tests DF vu qu'on a déjà détaillé plus tôt.

La spécification “trend” donne :

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32234 -0.48825 -0.07754  0.40300  2.82955
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.225529   0.234695   0.961    0.341
## z.lag.1      -0.732172   0.127925  -5.723 4.06e-07 ***
## tt           -0.009556   0.006815  -1.402    0.166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8843 on 57 degrees of freedom
## Multiple R-squared:  0.365, Adjusted R-squared:  0.3427
## F-statistic: 16.38 on 2 and 57 DF,  p-value: 2.396e-06
##
##
## Value of test-statistic is: -5.7235 10.9207 16.3802
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

On teste :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{cases} .$$

La p-valeur vaut $0.166 > 5\%$, donc on ne rejette pas H_0 , la spécification “trend” du test DF n’est pas la bonne spécification pour notre série.

La spécification “drift” donne :

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

```
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20428 -0.52287 -0.01661  0.43613  2.92432
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.06153    0.11574  -0.532   0.597
## z.lag.1      -0.68671    0.12478  -5.504 8.83e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8916 on 58 degrees of freedom
## Multiple R-squared:  0.3431, Adjusted R-squared:  0.3317
## F-statistic: 30.29 on 1 and 58 DF,  p-value: 8.827e-07
##
##
## Value of test-statistic is: -5.5036 15.1455
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

On teste :

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_a : \beta_0 \neq 0 \end{cases} .$$

La p-valeur vaut $0.597 > 5\%$, donc on ne rejette pas H_0 , la spécification “drift” du test DF n’est pas la bonne spécification pour notre série.

La spécification “none” donne :

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.25517 -0.58401 -0.07443  0.37669  2.86115
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.6798    0.1233  -5.512 8.22e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8862 on 59 degrees of freedom
## Multiple R-squared:  0.3399, Adjusted R-squared:  0.3287
## F-statistic: 30.38 on 1 and 59 DF,  p-value: 8.222e-07
##
##
## Value of test-statistic is: -5.5116
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

On teste :

$$\begin{cases} H_0 : (\rho - 1) = 0 \\ H_a : (\rho - 1) \neq 0 \end{cases} .$$

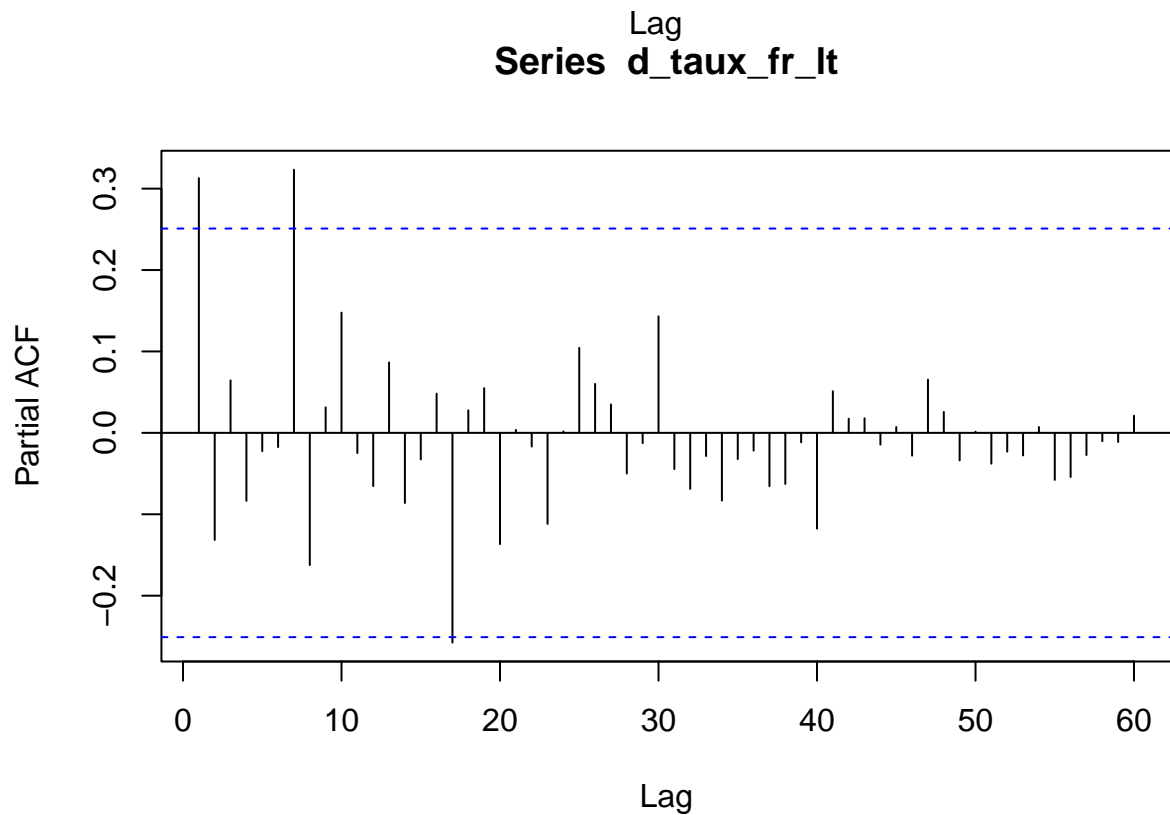
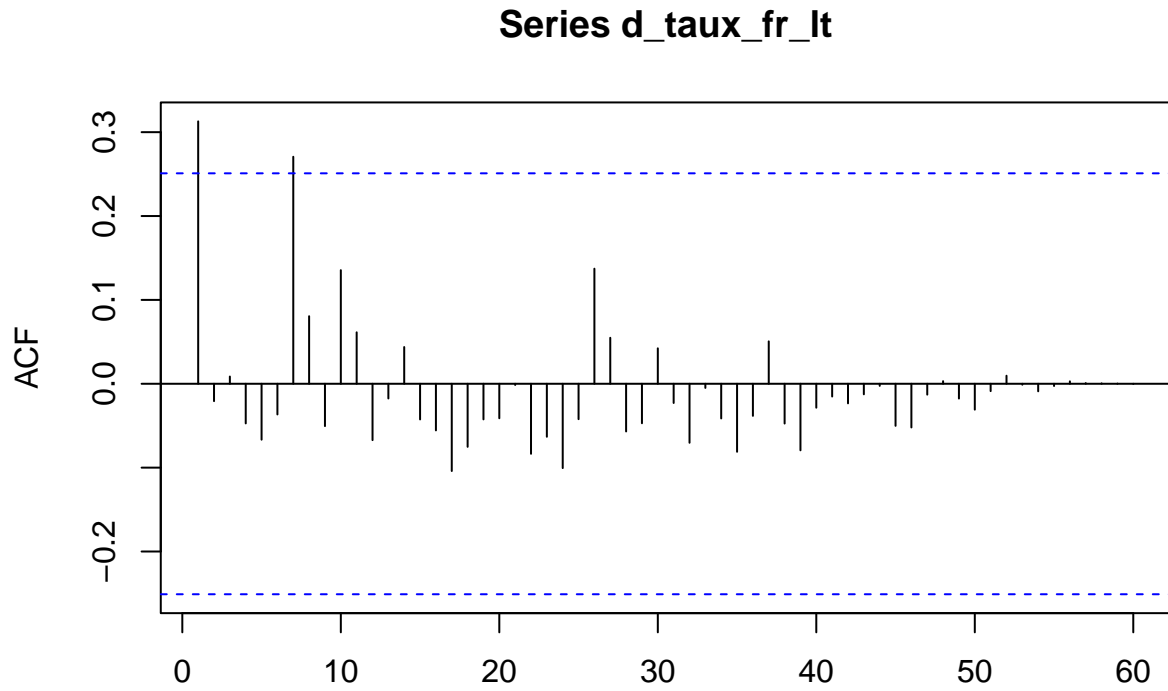
La t-valeur vaut -5.51 et est inférieure à la valeur tabulée -1.95 donc on rejette H_0 et $(\rho - 1)$ est différent de 0, c'est-à-dire que la série différenciée est stationnaire.

Comme tout à l'heure, on peut aussi utiliser la fonction `adf.test` :

```
##
## Augmented Dickey-Fuller Test
##
## data:  d_taux_fr_lt
## Dickey-Fuller = -4.0938, Lag order = 3, p-value = 0.0115
## alternative hypothesis: stationary
```

La p-valeur de ce test vaut $0.0115 < 5\%$. Nous pouvons donc rejeter l'hypothèse nulle et conclure que notre série différenciée est bien stationnaire.

Étudions les corrélogrammes :



On note la présence d'autocorrélation suite à l'application des fonctions ACF et PACF.

L'étape suivante est de déterminer le meilleur modèle $\text{ARMA}(p, q)$ pour notre série.

Modèle ARMA(p, q)

Nous emploierons 3 méthodes différentes.

EACF

Commençons par utiliser la technique de l'EACF :

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o o o o
## 1 x o o o o o x o o o o o o o
## 2 x o o o o o x o o o o o o o
## 3 x x o o o o o o o o o o o
## 4 o o o o o o o o o o o o o
## 5 x o o o o o x o o o o o o o
## 6 o o o o x x x o o o o o o o
## 7 x o x o o o o o o o o o o
```

L'EACF de la série différenciée nous donne comme couple (p, q) : $(0, 6)$. Vérifions que le modèle est bien pertinent en analysant la significativité des coefficients :

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1  0.574707   0.119256  4.8191 1.442e-06 ***
## ma2  0.208587   0.165957  1.2569 0.208800
## ma3 -0.198630   0.152125 -1.3057 0.191656
## ma4  0.053419   0.119232  0.4480 0.654136
## ma5  0.128255   0.165997  0.7726 0.439739
## ma6 -0.377706   0.126558 -2.9845 0.002841 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Les coefficients de $ma(2)$, $ma(3)$, $ma(4)$ et $ma(5)$ ne sont pas significatifs car leur p-valeur est supérieure à 5% ; ce modèle n'est pas pertinent.

Auto.arima

Une autre façon de calculer les coefficients p et q optimaux est d'utiliser la fonction `auto.arima()` du package `forecast` de R :

```
## Series: d_taux_fr_lt
## ARIMA(0,0,1) with zero mean
##
## Coefficients:
##      ma1
##      0.3878
## s.e.  0.1296
##
## sigma^2 = 0.7547: log likelihood = -77.55
## AIC=159.1   AICc=159.31   BIC=163.32
```

La fonction nous renvoie comme couple optimal $(p, q) = (0, 1)$.

```
##
## z test of coefficients:
```

```
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1  0.38777      0.12961  2.9919 0.002772 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On constate cette fois-ci que le coefficient associé à $ma(1)$ est significatif. On conserve donc ce modèle que l'on nomme *reg2* pour la suite.

Fonction personnelle

Pour cette méthode, on a codé à l'aide de quelques boucles “for” imbriquées un programme retournant les 2 meilleurs modèles au sens de la minimisation du BIC vu en cours. Lançons-le :

```
BIC_mini = 1000
BIC_mini2 = 1001
pmin = 0
qmin = 0
pmin2 = 0
qmin2 = 0

for (q in 0:11){
  for (p in 0:11){
    reg_bis <- Arima(taux_fr_lt, order = c(p,1,q),include.mean = FALSE)
    if (BIC(reg_bis) < BIC_mini) {
      BIC_mini2 = BIC_mini
      qmin2 = qmin
      pmin2 = pmin
      BIC_mini = BIC(reg_bis)
      pmin = p
      qmin = q
    }
    if (BIC(reg_bis) < BIC_mini2 & BIC(reg_bis) > BIC_mini ) {
      BIC_mini2 = BIC(reg_bis)
      qmin2 = q
      pmin2 = p
    }
  }
}
```

Affichons les 2 couples optimaux :

```
print(BIC_mini)
```

```
## [1] 163.322
```

```
print(pmin)
```

```
## [1] 0
```

```
print(qmin)
```

```
## [1] 1
```

```
print(BIC_mini2)
```

```
## [1] 164.6911
```

```
print(pmin2)
```

```
## [1] 1
```

```
print(qmin2)
```

```
## [1] 0
```

Nous obtenons comme premier couple optimal (0,1), soit le modèle *reg2* obtenu avec la fonction *auto.arima*, et le second couple (1,0).

```
##
```

```
## z test of coefficients:
```

```
##
```

```
##      Estimate Std. Error z value Pr(>|z|)
```

```
## ar1  0.31519    0.12012  2.6239 0.008692 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On constate que le coefficient associé à *ar*(1) est significatif. On conserve donc ce modèle que l'on nomme *reg3* pour la suite.

Bruits blancs

Maintenant que nous avons 2 modèles pertinents, il faut, avant de faire de la prévision avec, s'assurer que leurs aléas sont des bruits blancs. Rappelons qu'un bruit blanc (BB) est une séquence de variables aléatoires i.i.d. de moyenne non nulle ($\mathbb{E}[\epsilon] = 0$), de variance constante (homoscédasticité) et non autocorrélées.

Dans la série de tests qui suit, nous estimons les aléas par les résidus, mais les conclusions seront faites sur les aléas.

Test deJarque-Bera

Ce test permet de savoir si les résidus sont normalement distribués. On teste :

$$\begin{cases} H_0 : \epsilon \sim \mathcal{N} \\ H_a : \epsilon \not\sim \mathcal{N} \end{cases} .$$

```
##
```

```
## Jarque Bera Test
```

```
##
```

```
## data: reg2$res
```

```
## X-squared = 4.1623, df = 2, p-value = 0.1248
```

```
##
```

```
## Jarque Bera Test
```

```
##
```

```
## data: reg3$res
```

```
## X-squared = 5.9319, df = 2, p-value = 0.05151
```

Pour nos 2 modèles retenus, la p-valeur est supérieure à 5% donc on ne rejette pas H_0 et les résidus sont normalement distribués (ou gaussiens).

Test de Student

Ce test permet de savoir si l'espérance des résidus vaut 0. On teste :

$$\begin{cases} H_0 : \mathbb{E}(\epsilon) = 0 \\ H_a : \mathbb{E}(\epsilon) \neq 0 \end{cases} .$$

```
##
## One Sample t-test
##
## data: reg2$res
## t = -0.59465, df = 60, p-value = 0.5543
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.2877957 0.1558949
## sample estimates:
## mean of x
## -0.06595039

##
## One Sample t-test
##
## data: reg3$res
## t = -0.55756, df = 60, p-value = 0.5792
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.2871225 0.1619487
## sample estimates:
## mean of x
## -0.06258688
```

Pour nos 2 modèles retenus, la p-valeur est supérieure à 5% donc on ne rejette pas H_0 et l'espérance des résidus vaut 0.

Test de Ljung-Box

Ce test permet de savoir s'il existe de l'autocorrélation sur les résidus centrés et réduits de notre modèle. On teste :

$$\begin{cases} H_0 : \text{Absence d'autocorrélation} \\ H_a : \text{Présence d'autocorrélation} \end{cases}$$

```
reg2_stand <- ((reg2$res - mean(reg2$res)) / sd(reg2$res))
Box.test(reg2_stand, type = "Ljung-Box")
```

```
##
## Box-Ljung test
##
## data: reg2_stand
## X-squared = 0.022289, df = 1, p-value = 0.8813
```

```
reg3_stand <- ((reg3$res - mean(reg3$res)) / sd(reg3$res))
Box.test(reg3_stand, type = "Ljung-Box")
```

```
##
## Box-Ljung test
##
## data: reg3_stand
## X-squared = 0.10084, df = 1, p-value = 0.7508
```

Pour nos 2 modèles retenus, la p-valeur est supérieure à 5% donc on ne rejette pas H_0 et il n'y a pas d'autocorrélation sur les résidus centrés et réduits.

Test d'Engle

Ce test permet de détecter d'éventuels clusters de volatilité dans les résidus. On teste :

$$\begin{cases} H_0 : \text{Absence de clusters de volatilité} \\ H_a : \text{Présence de clusters de volatilité} \end{cases}$$

```
##
## ARCH LM-test; Null hypothesis: no ARCH effects
##
## data: reg2_stand
## Chi-squared = 20.737, df = 12, p-value = 0.05436
##
## ARCH LM-test; Null hypothesis: no ARCH effects
##
## data: reg3_stand
## Chi-squared = 20.53, df = 12, p-value = 0.0577
```

Pour nos 2 modèles retenus, la p-valeur est supérieure à 5% donc on ne rejette pas H_0 et il n'y a pas de clusters de volatilité les résidus centrés et réduits.

Prévisions et visualisation

Il ne nous reste plus qu'à utiliser nos 2 modèles pertinents *reg2* et *reg3* pour prédire la valeur du taux d'intérêt de long terme des obligations en France en 2022, 2023 et 2024.

Les modèles nous donnent comme valeurs :

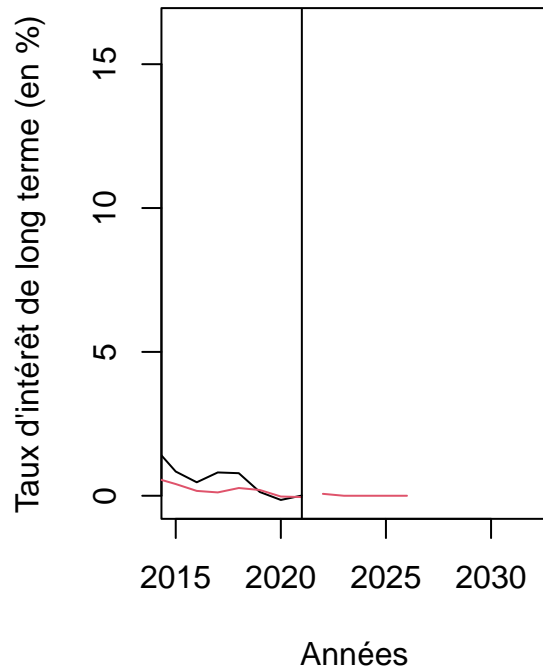
##	Point Forecast	Lo 95	Hi 95
## 2022	0.02064644	-1.682079	1.723372
## 2023	0.00000000	-1.826259	1.826259
## 2024	0.00000000	-1.826259	1.826259

##	Point Forecast	Lo 95	Hi 95
## 2022	0.0023219068	-1.720442	1.725086
## 2023	0.0007318440	-1.805580	1.807044
## 2024	0.0002306706	-1.814172	1.814633

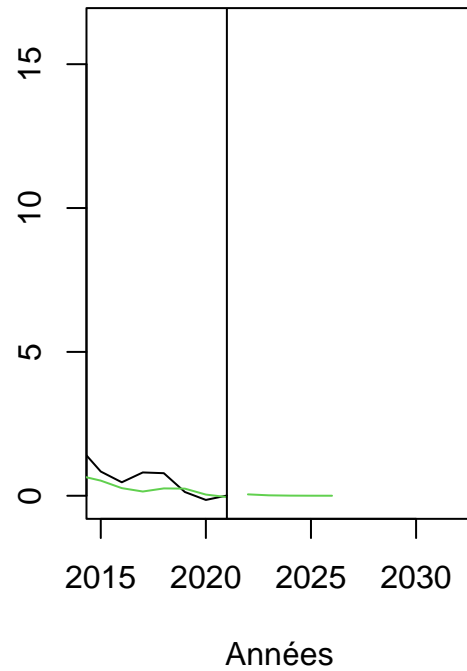
Ces résultats confirment notre hypothèse énoncée en tout début de projet : le taux d'intérêt de long terme prévu pour les 3 prochaines années est quasiment (si ce n'est parfaitement) nul.

On peut présenter les résultats sous forme de graphiques :

Prévision par reg2



Prévision par reg3



Ou encore :

Prévision sur trois ans

