

Projet Économétrie II: Investissements Directs Etrangers et croissance économique française

Clothilde Dano et Alexandre Antoine

2023-11-30

Sommaire

Introduction

Partie 1: MODÉLISATION UNIVARIÉE

1. Choix des séries temporelles
2. Représentation graphiques et autocorrélogrammes simples et partiels
 - 2.1 PIB
 - 2.2 IDE
3. Stratégie de test de racine unitaire, test KPSS et stationnarisation
 - 3.1 PIB
 - 3.2 IDE
4. Identification du modèle ARMA et validation
 - 4.1 Identification
 - 4.2 Validation
5. Prévisions

Partie 2: MODÉLISATION MULTIVARIÉE

1. Estimation du modèle VAR avec un nombre de retards optimal
2. Relation de causalité entre les variables
 - 2.1 Test d'absence de causalité du PIB sur les IDE
 - 2.2 Test d'absence de causalité des IDE sur le PIB
3. Analyse impulsion-réponse des chocs
 - 3.1 La méthode des prévisions récursives avec le VAR
 - 3.2 La méthode des projection locales
 - 3.3 La décomposition de la variable
4. Test de cointégration de Johansen

Introduction

Dans ce projet d'économétrie, nous plongeons dans l'étude de deux séries temporelles cruciales pour l'économie française : les Investissements Directs Étrangers (IDE) envoyés par la France et le taux de croissance du Produit Intérieur Brut (PIB). L'objectif principal de cette analyse est d'explorer la dynamique de ces deux variables et d'identifier toute relation économique potentielle qui pourrait les lier.

L'histoire économique montre que la croissance robuste de la France a souvent coïncidé avec des périodes où les entreprises françaises ont massivement investi à l'étranger. Par exemple, pendant les périodes de croissance économique soutenue en France, les entreprises ont souvent augmenté leurs investissements dans des secteurs clés à l'international, renforçant ainsi leur présence mondiale.

Inversement, des fluctuations dans la croissance économique nationale ont également eu un impact sur les IDE sortants. Les périodes de ralentissement économique en France ont parfois conduit à une réduction des IDE sortants, car les entreprises se concentraient davantage sur la consolidation de leurs opérations nationales pour faire face à des conditions économiques plus difficiles.

Par ailleurs, des réorientations stratégiques de la part des entreprises françaises, telles que des investissements dans de nouveaux marchés ou des restructurations sectorielles, ont influencé la croissance économique interne en France. Ces mouvements ont eu des implications sur la compétitivité et l'innovation des entreprises, affectant ainsi leur contribution à l'économie nationale. On peut également penser à certaines politiques d'investissement massifs, qui ont permis de relancer l'économie française lors de périodes de récession.

En synthèse, cette analyse économétrique vise à explorer la relation dynamique entre la croissance économique de la France et les IDE sortants du pays. Comprendre cette interrelation bidirectionnelle est crucial pour orienter les politiques économiques et d'investissement afin de stimuler une croissance économique durable et d'encourager les entreprises françaises à prospérer tant sur le marché national qu'international.

Partie 1 : Modélisation Univariée

1. Choix des séries temporelles

Dans notre modèle, nous avons décidé d'explorer la causalité qui existe entre le taux de croissance du PIB de la France d'un trimestre à l'autre et les IDE de la France calculés en % du PIB.

Le taux de croissance du Produit Interieur Brut (PIB) d'un trimestre à l'autre évalue la variation de la production de biens et de services d'un pays sur une période spécifique. Il représente l'augmentation ou la diminution de la richesse créée par l'économie pendant ce laps de temps.

Le PIB trimestriel que nous avons choisi mesure l'activité économique en termes de valeur monétaire ajustée pour l'inflation, permettant ainsi une comparaison plus précise des performances économiques sur le long terme. Cette mesure est corrigée des variations saisonnières (CVS) et exprimée en pourcentage pour mieux évaluer la croissance économique du pays. Les données utilisées pour calculer cette mesure sont fondées sur le Système de Comptabilité Nationale de 2008 (SCN 2008). Le pays étudié est ici la France, et la période concernée est ici 1999-2021.

Les investissements directs étrangers (IDE) sont des investissements réalisés par une entreprise résidente d'une économie (l'investisseur direct) dans le but d'acquérir un intérêt durable dans une entreprise opérant dans une autre économie (l'entreprise d'investissement direct). Cet intérêt durable implique une relation à long terme entre l'investisseur direct et l'entreprise d'investissement direct, avec une influence substantielle sur la gestion de cette dernière.

Ici, afin de neutraliser l'impact des différences de taille des économies entre les pays déclarants, nous avons choisi des données exprimées en pourcentage du PIB. Nos chiffres traduisent les IDE de la France exprimés en pourcentage du PIB de la France, envoyés dans le reste du monde de 1999 à 2021, les données sont trimestrielles. Notre base est issue de la section "Investissements étrangers" de la balance des paiements d'Eurostat.

Étudier la causalité entre les investissements directs étrangers de la France, exprimés en pourcentage du PIB, et le taux de croissance du pays revêt un intérêt crucial pour plusieurs raisons. Une croissance économique robuste stimule les IDE sortants en renforçant la compétitivité des entreprises sur les marchés mondiaux, en favorisant l'innovation et le développement technologique, ainsi qu'en diversifiant les activités économiques à l'échelle internationale. En parallèle, les IDE sortants contribuent à la croissance économique nationale en créant des emplois, en stimulant les exportations et en soutenant la compétitivité des entreprises sur le marché mondial.

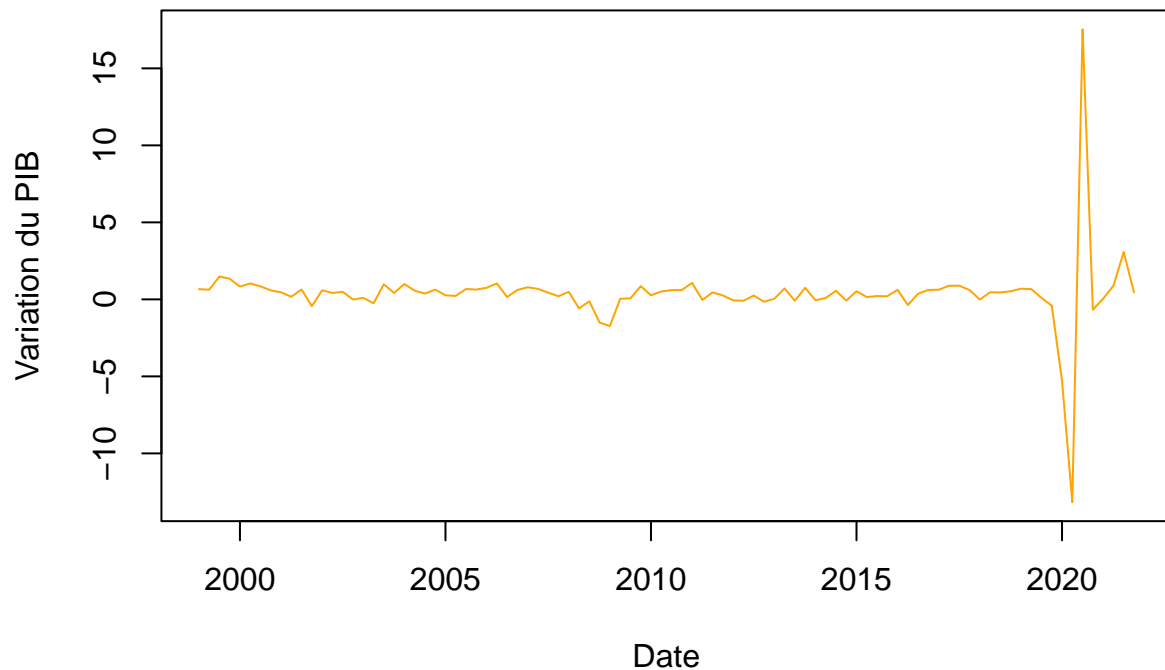
Par ailleurs, dans notre modèle économétrique, nous avons décidé de logariser les IDE de la France calculés en pourcentage du PIB. En effet, cette logarisation présente plusieurs avantages significatifs. Tout d'abord, la transformation logarithmique peut être bénéfique pour corriger d'éventuelles asymétries ou écarts importants dans les données des IDE, offrant ainsi une meilleure stabilité aux données en minimisant les variations extrêmes. Ensuite, elle permet de mieux représenter une éventuelle relation non linéaire entre les IDE et la croissance économique, en rendant les effets proportionnels plutôt qu'absolus. Finalement, en logarisant les IDE, cette approche facilite l'interprétation des effets en termes de pourcentage de changement, offrant une perspective plus intuitive et significative pour comprendre la relation.

2. Graphiques et Autocorrélogrammes

Représentation graphique des séries :

Nous utilisons le logiciel R afin d'étudier les séries temporelles que nous avons choisies.

```
plot(PIB,xlab="Date",ylab="Variation du PIB", col="orange")
```

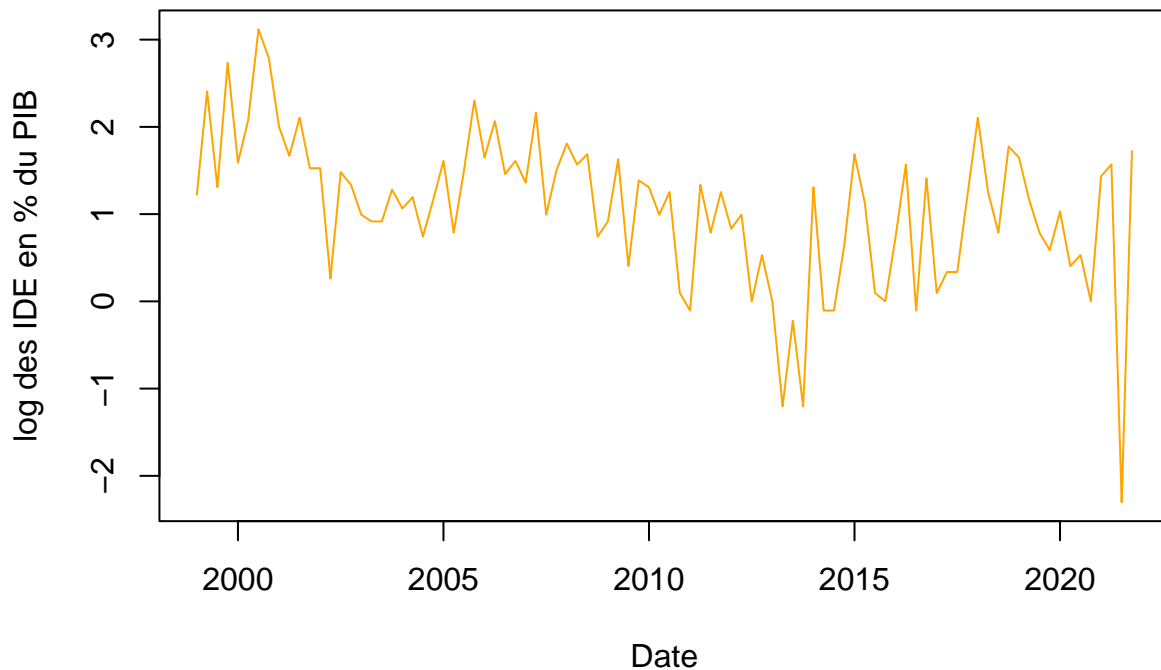


La figure ci-dessus représente l'évolution du taux de croissance du PIB de la France entre le premier trimestre de 1999 et le dernier trimestre de 2021, exprimé en % de la période précédente. Il nous montre que le PIB a évolué de manière stable de 1999 à 2021. Avec une tendance à la hausse en général, le taux de croissance se trouve généralement au-dessus de 0 malgré quelques périodes de baisse. On observe cependant quelques variations sur la courbe. En 2008, on voit que la croissance du PIB français a chuté, ceci peut s'expliquer par la crise financière mondiale qui empiriquement a fait chuter le PIB français de 3% en 2008. Le PIB a ensuite retrouvé sa croissance, mais il a connu une nouvelle baisse en 2020 en raison de la pandémie de COVID-19.

En conclusion, l'évolution du PIB de la France sur la période 1999-2021 est globalement positive, avec une tendance à la hausse. Cependant, la croissance du PIB a été ponctuée par quelques crises économiques.

On remarque par ailleurs que les variations sont similaires au cours du temps, on peut donc penser que la série est stationnaire.

```
plot(logIDE,xlab="Date",ylab="log des IDE en % du PIB", col= "orange")
```



Le graphique ci-joint représente l'évolution des investissements directs étrangers (IDE) sortants de la France entre 1999 et 2021.

L'analyse économique de ce graphique permet de tirer plusieurs conclusions. Tout d'abord, les IDE sortants sont un indicateur important de l'internationalisation des économies. En France, ils ont connu une forte croissance jusqu'en 2007, avant de se stabiliser à un niveau inférieur. Cette évolution est le reflet de plusieurs facteurs, notamment la mondialisation, l'ouverture des marchés au début des années 2000 a en effet fait augmenter les IDE vers le reste du monde; les effets tardifs de la crise financière de 2008, on voit en effet une tendance à la baisse après 2008, et la crise sanitaire de 2020, on remarque le pic de baisse des IDE en 2020/2021, ceci s'explique par la fermeture des économies et la hausse des investissements internes nécessaires durant cette période.

Finalement, au vu du graphique, la série présente une tendance à long-terme, et son écart-type varie, on peut donc penser que les IDE de la France logarisés n'est pas une série stationnaire.

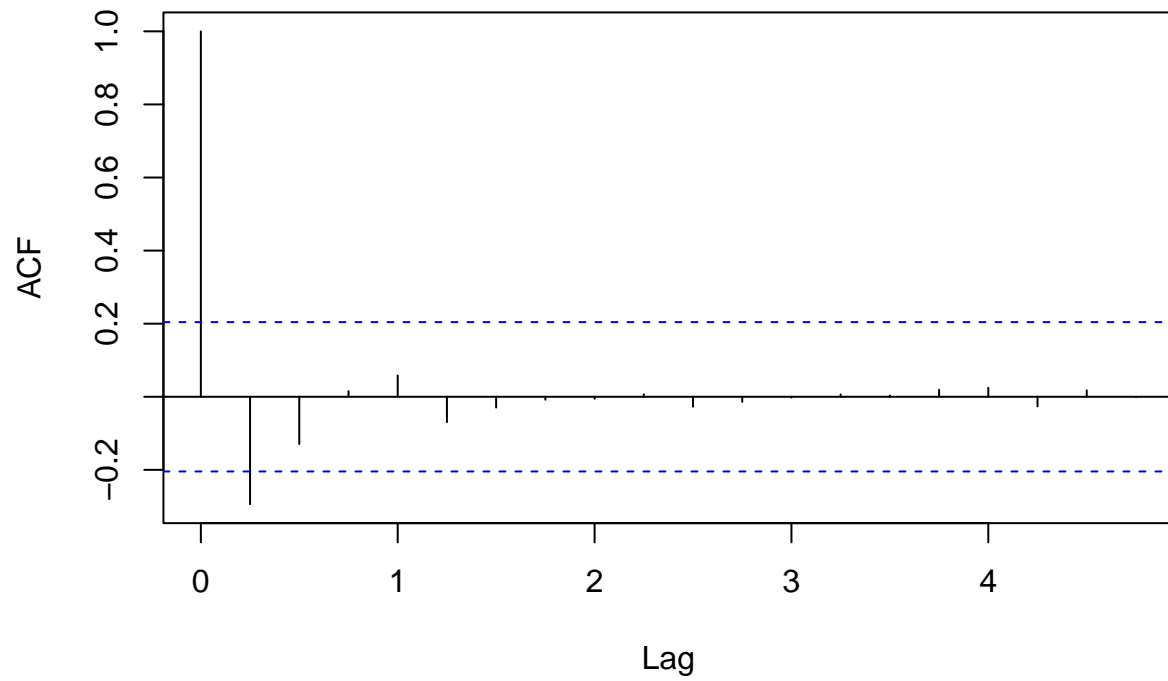
L'autocorrélation se réfère à la corrélation d'une variable avec elle-même à différents moments dans le temps. En d'autres termes, il s'agit de la corrélation entre les valeurs successives d'une même variable dans une série temporelle. Si une variable est autocorrélée, cela signifie qu'il existe une corrélation statistiquement significative entre les valeurs passées et/ou présentes de cette variable.

On regarde l'autocorrélation à l'aide des autocorrélogrammes simples et partiels pour chaque variable.

Autocorrélogrammes simples :

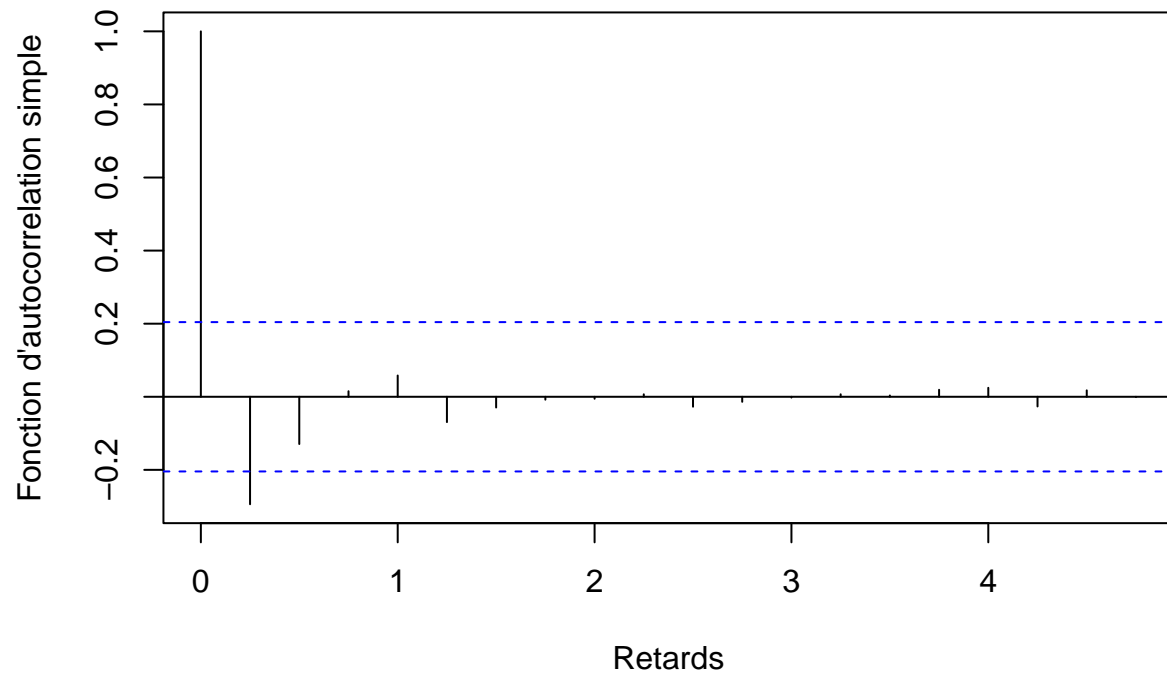
```
acfPIB<-acf(PIB)
```

Series PIB



```
plot(acfPIB,xlab="Retards",ylab="Fonction d'autocorrelation simple",main="ACF du taux de croissance du PIB",ylim=c(-0.2,1.0))
```

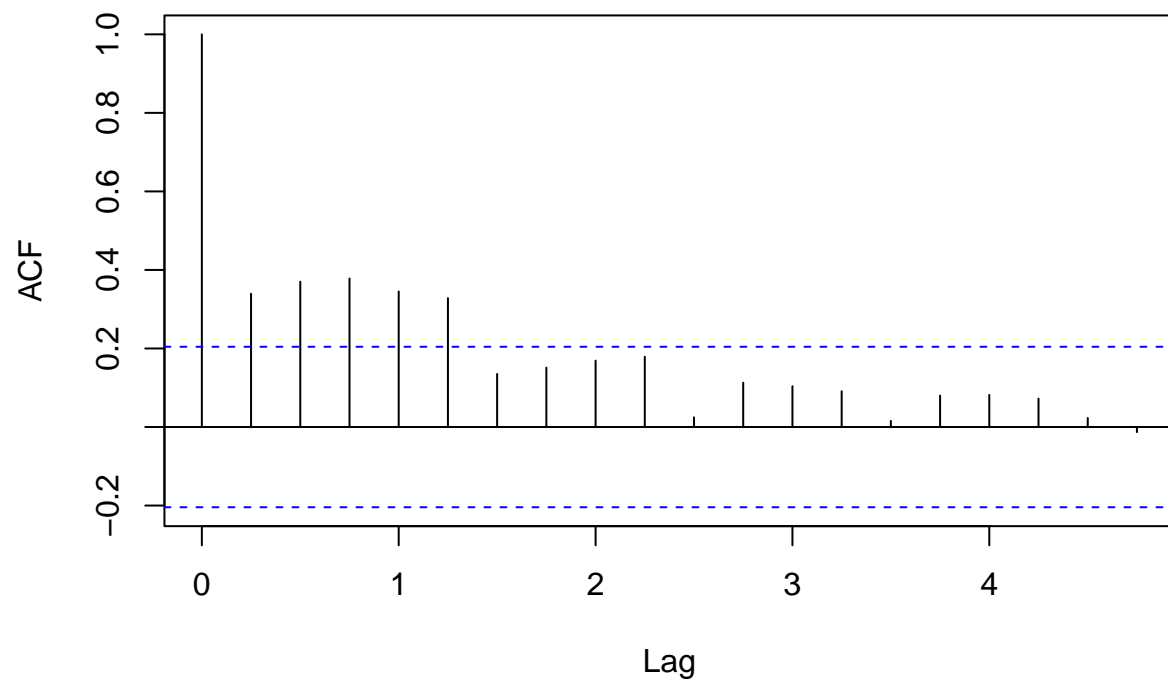
ACF du taux de croissance du PIB



On voit que la fonction d'autocorrélation simple tend vers 0 au fur et à mesure que les retards augmentent. La série pourrait ne pas être stationnaire.

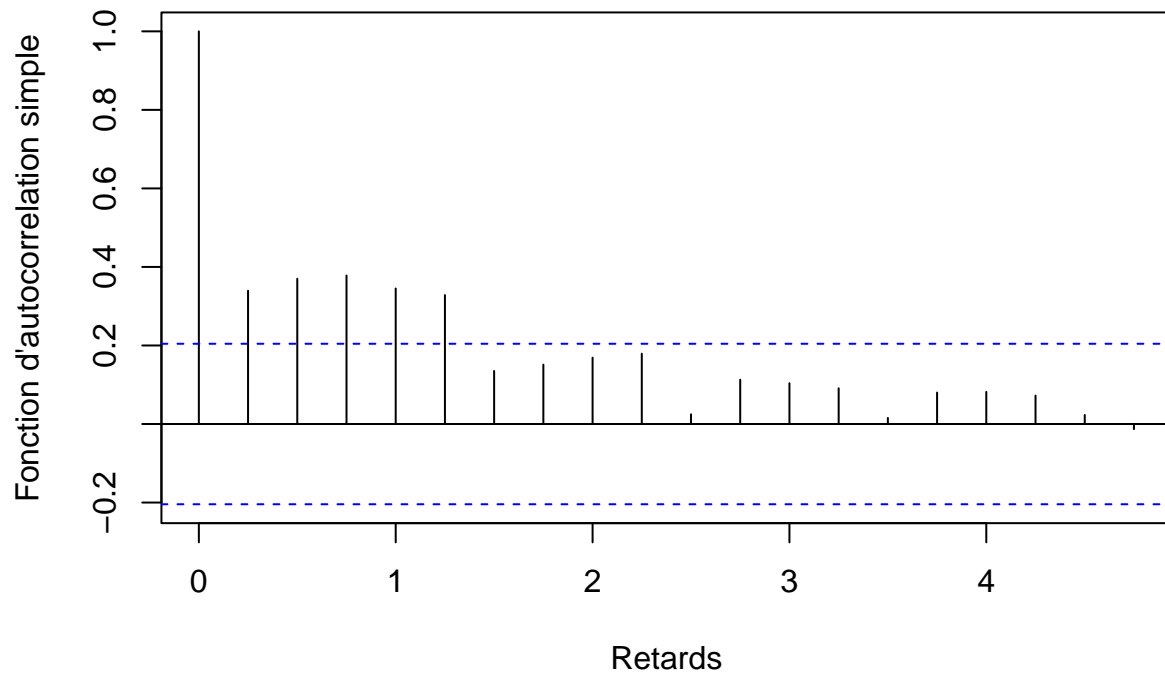
```
acflogIDE <- acf(logIDE)
```


Series logIDE



```
plot(acflogIDE,xlab="Retards",ylab="Fonction d'autocorrelation simple",main="ACF du log des IDE")
```

ACF du log des IDE



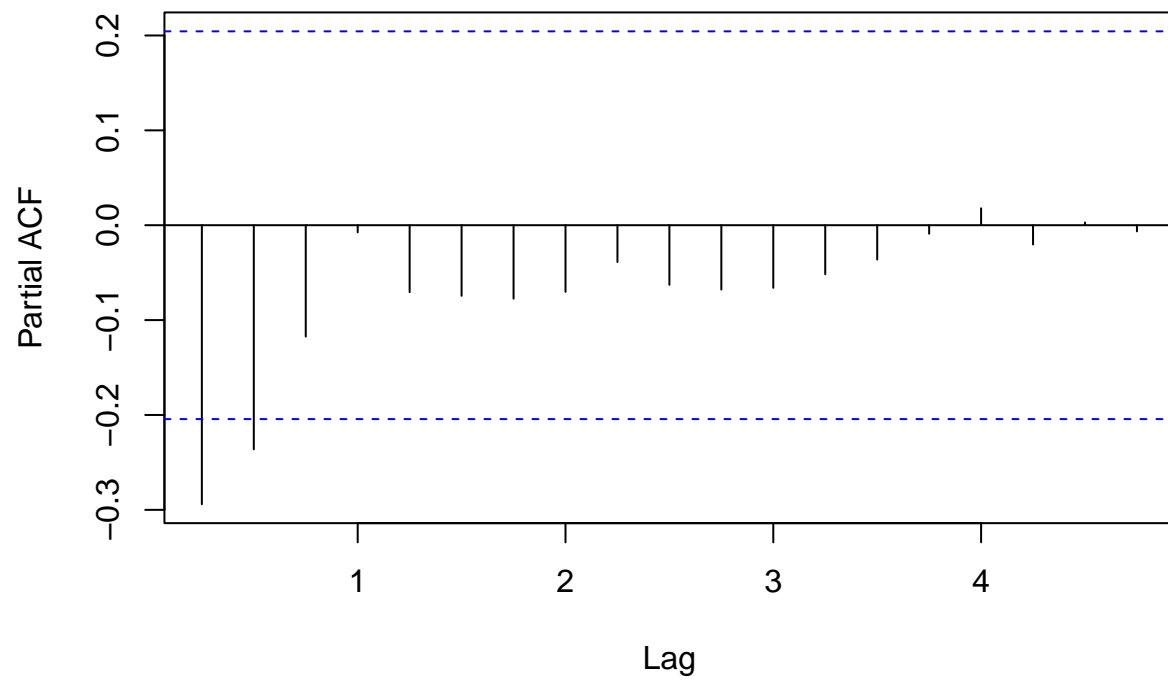
On voit que la fonction d'autocorrélation simple décroît avec h , et tend vers 0, ce qui est révélateur d'une série non stationnaire. Ce qui nous semble être cohérent avec les hypothèses mises en lumière lors de la lecture graphique.

Autocorrélogrammes partiels :

L'autocorrélogramme partiel retire l'influence des variables intermédiaires.

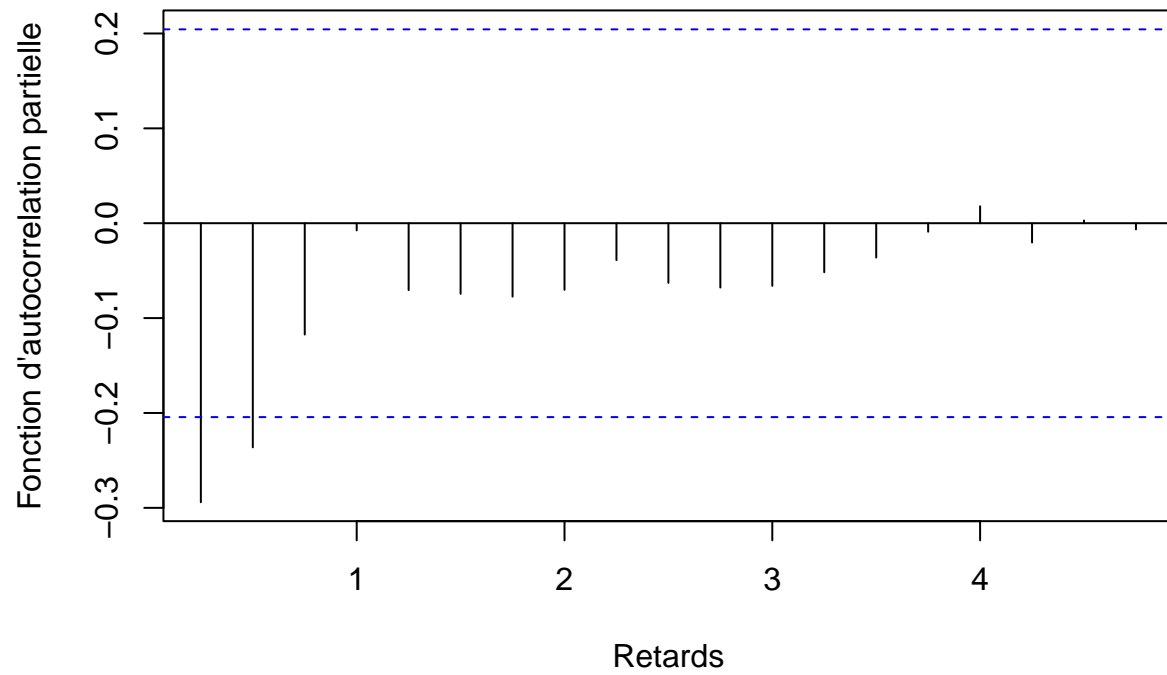
```
pacfPIB <- pacf(PIB)
```

Series PIB



```
plot(pacfPIB,xlab="Retards",ylab="Fonction d'autocorrelation partielle",main="PACF du PIB")
```

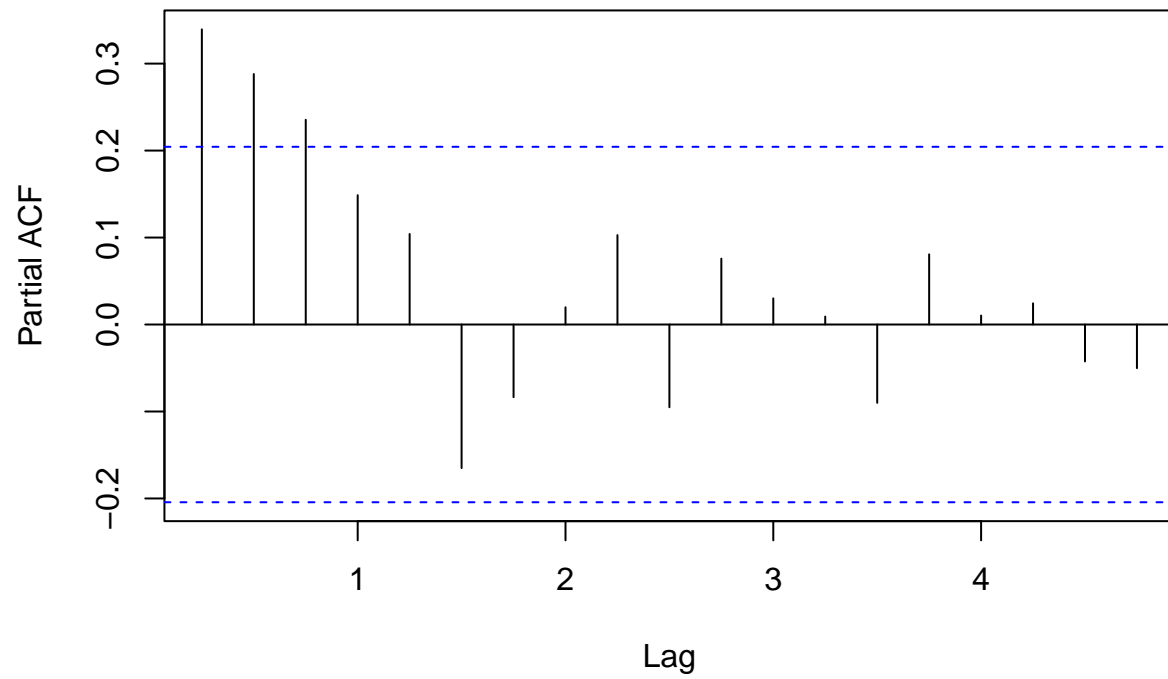
PACF du PIB



Au delà du premier retard, l'autocorrélation partielle n'est plus significative (comprise entre les barres en pointillées du graphe), ce qui laisse présager une série non stationnaire.

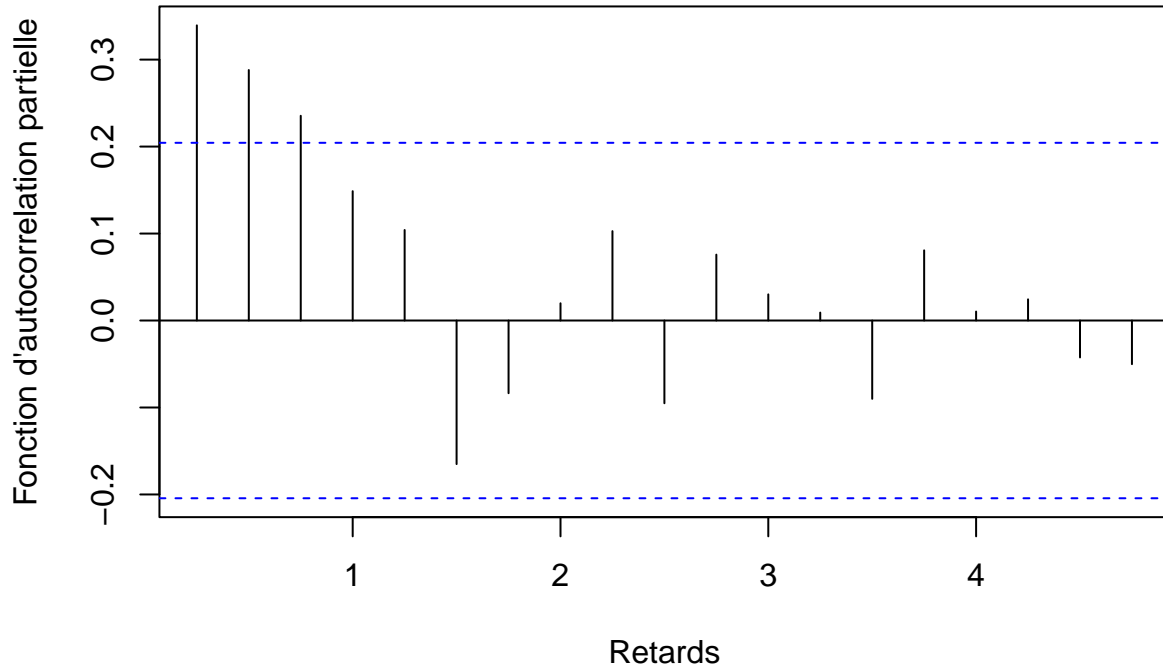
```
pacflogIDE <- pacf(logIDE)
```

Series logIDE



```
plot(pacflogIDE,xlab="Retards",ylab="Fonction d'autocorrelation partielle",main="PACF du log des IDE")
```

PACF du log des IDE



Au delà du premier retard, l'autocorrélation partielle n'est plus significative, ce qui laisse également présager une série non stationnaire.

Pour pouvoir ensuite effectuer des prévisions sur ces séries, il est essentiel de travailler sur des séries stationnaires. Or, les autocorrélogrammes simples et partiels ne sont pas à eux-seuls des indicateurs suffisants de stationnarité, d'où l'importance d'effectuer des tests sur ces séries, afin de conclure sur leur stationnarité. C'est l'objet de la prochaine section.

3. Stratégie de test de racine unitaire et méthode de stationnarisation

a. L'analyse pour le PIB

a.1 - Test de Dickey Fuller sur le PIB

Avec les applications aux deux modèles qui suivent, les hypothèses du test ADF (Augmented Dickey Fuller) sont les suivantes:

$$H_0 : \phi = 1 \Rightarrow \text{racine unitaire contre } H_1 : \phi < 1 \Rightarrow \text{pas de racine unitaire}$$

La statistique de test du test de Dickey Fuller est définie telle que :

$$\hat{t}_\rho = \frac{\hat{\rho}}{\hat{\sigma}_\rho}$$

La règle de décision consiste à rejeter H_0 si $\hat{t}_\rho < c$. Ici, c représente la valeur critique, obtenue à partir des tables de Dickey-Fuller (modèle 3, 92 observations) qui sont comme suit :

α	1%	5%	10%
M3	- 4.1	- 3.47	- 3.17.
M2	- 3.55	- 2.91	- 2.59
M1	- 2.61	- 1.95	- 1.61

Le nombre de retards optimal est choisi en fonction de la minimisation du critère AIC, avec un retard maximal égal à 4, étant donné que cette série est trimestrielle.

Le modèle 3 avec la tendance déterministe Le modèle sur lequel nous travaillons s'écrit ici:

$$\Delta X_t = c + bt + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
library(urca)
adfPIB1<-ur.df(PIB,type=c("trend"), lag=10, selectlags = c('AIC'))
summary(adfPIB1)
```

```
##
## #####
## # 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
## -15.8051  -0.2545   0.1317   0.4554  10.6858
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.441163   0.652011   0.677   0.5007
## z.lag.1      -1.630398   0.178996  -9.109 7.41e-14 ***
## tt           0.001099   0.011578   0.095   0.9246
## z.diff.lag    0.252621   0.111231   2.271   0.0259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.436 on 77 degrees of freedom
## Multiple R-squared:  0.6727, Adjusted R-squared:  0.66
## F-statistic: 52.76 on 3 and 77 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -9.1085 27.6649 41.4951
##
## 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 statistique du test ADF vaut -9.1085, elle est inférieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10% (lus sur la ligne tau3). On rejette donc l'hypothèse nulle de racine unitaire.

La tendance déterministe est-elle significative ?

Comme on a rejeté l'hypothèse nulle, on utilise un test de student pour analyser la significativité de la tendance déterministe

$$H_0 : b = 0 \text{ contre } H_1 : b \neq 0$$

La statistique de test de Student pour un échantillon est donnée par :

$$\hat{t}_b = \frac{\hat{b}}{\hat{\sigma}_b}$$

La t-stat de la tendance déterministe est égale à 0.095. La valeur critique qui figure sur la table de Student est égale à :

$$|\hat{t}_b| = 2.86$$

(prise pour un risque de première espèce de 5%) donc on a

$$\hat{t}_b < VC.$$

L'hypothèse nulle n'est donc pas rejetée, c'est-à-dire la tendance déterministe n'est pas significative, donc on l'enlève. La conclusion est cohérente avec la p-valeur du test, qui vaut 0.9246, bien supérieure aux seuils de rejet de 1%, 5%, et 10%.

Le modèle 2 avec la constante Nous travaillons ici avec le modèle:

$$\Delta X_t = c + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
adfPIB2<-ur.df(PIB,type=c("drift"), lag=4, selectlags = c('AIC'))
summary(adfPIB2)
```

```
##
## #####
## # 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
## -15.7764  -0.2711   0.1385   0.4102  10.7831
##
## Coefficients:
```



```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.5342    0.2570   2.079  0.0407 *
## z.lag.1      -1.6196    0.1712  -9.461   7e-15 ***
## z.diff.lag    0.2467    0.1065   2.317  0.0229 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.34 on 84 degrees of freedom
## Multiple R-squared:  0.6707, Adjusted R-squared:  0.6629
## F-statistic: 85.56 on 2 and 84 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -9.4612 44.7573
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

La statistique du test ADF vaut -9.4612, elle est inférieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10% (lus sur la ligne tau3). On rejette donc l'hypothèse nulle de racine unitaire.

La constante est-elle significative ?

$$H_0 : c = 0 \text{ contre } H_1 : c \neq 0$$

On voit que sa p-valeur vaut 0.0407, donc on rejette H_0 au seuil 5%, et on conclut que la constante est significative.

On en conclut que nous travaillons avec un processus de la forme: $I(0) + C$

Pour confirmer ces résultats, nous effectuons un test KPSS, dans le but ici de minimiser le risque de seconde espèce, alors que le test ADF minimisait le risque de première espèce.

a.2 - Test KPSS sur le PIB

Avec les applications aux deux modèles qui suivent, les hypothèses du test KPSS (Kwiatkowski-Phillips-Schmidt-Shin) sont les suivantes:

$$H_0 : \phi < 1 \Rightarrow \text{pas de racine unitaire contre } H_1 : \phi = 1 \Rightarrow \text{racine unitaire}$$

La statistique de test est donnée par :

$$LM = \frac{1}{T} \sum_{t=1}^T \frac{S^2}{s^2} \cdot \frac{1}{T^2} \cdot t$$

La règle de décision est de rejeter H_0 si $LM > c$, où c est la valeur critique.

Test KPSS avec la constante uniquement : On travaille sur:

$$\Delta X_t = c + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
kpss_PIB1 <- ur.kpss(y=PIB, type="tau", lags="short" )
summary(kpss_PIB1)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: tau with 3 lags.
##
## Value of test-statistic is: 0.0447
##
## Critical value for a significance level of:
##          10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
```

On trouve une statistique de test égale à $LM = 0.0447$, donc inférieure aux seuils de rejet indiqués lors du test. D'après la règle de décision pour le test KPSS, on ne peut pas rejeter H_0 , ce qui est cohérent avec les résultats obtenus avec les tests ADF.

On effectue à nouveau le test avec l'ajout d'une tendance déterministe.

Test KPSS avec la constante et la tendance déterministe On travaille ici avec:

$$\Delta X_t = c + bt + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
kpss_PIB2 <- ur.kpss(y=PIB, type="mu", lags="short" )
summary(kpss_PIB2)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0629
##
## Critical value for a significance level of:
##          10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

On a donc une statistique de test, qui vaut ici 0.0629, qui est inférieure aux valeurs critiques indiquées, donc on ne peut pas rejeter H_0 , ce qui est cohérent avec les résultats obtenus avec les tests ADF.

Le processus est donc bien $I(0) + C$. C'est un processus stationnaire.

b. L'analyse pour le log des IDE

b.1 - Test de Dickey Fuller augmenté sur le log des IDE

On applique ici exactement la même stratégie. Les hypothèses sont:

$$H_0 : \phi = 1 \Rightarrow \text{racine unitaire contre } H_1 : \phi < 1 \Rightarrow \text{pas de racine unitaire}$$

La statistique de test de Dickey Fuller est définie comme suit :

$$\hat{t}_\rho = \frac{\hat{\rho}}{\hat{\sigma}_\rho}$$

Les valeurs critiques du test de Dickey Fuller pour les 3 modèles sont obtenues à partir des tables de Dickey-Fuller (modèle 3, 92 observations) elles sont comme suit :

α	1%	5%	10%
M3	- 4.1	- 3.47	- 3.17.
M2	- 3.55	- 2.91	- 2.59
M1	- 2.61	- 1.95	- 1.61

Le nombre de retards optimal est à nouveau choisi en fonction de la minimisation du critère AIC, avec un retard maximal égal à 4 de nouveau, étant donné que cette série est elle aussi trimestrielle.

Le modèle 3 avec la tendance déterministe Le modèle sur lequel nous travaillons s'écrit ici:

$$\Delta X_t = c + bt + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
library(urca)
adflogIDE3<-ur.df(logIDE,type=c("trend"), lag=4, selectlags = c('AIC'))
summary(adflogIDE3)
```

```
##
## #####
## # 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
## -2.8751 -0.3881 -0.0231  0.4223  1.5603
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.852925   0.321550   2.653  0.00959 **
## z.lag.1      -0.514925   0.154382  -3.335  0.00128 **
## tt           -0.007275   0.003899  -1.866  0.06565 .
## z.diff.lag1  -0.394293   0.145374  -2.712  0.00814 **
## z.diff.lag2  -0.232677   0.124070  -1.875  0.06430 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7435 on 82 degrees of freedom
```

```
## Multiple R-squared:  0.4518, Adjusted R-squared:  0.4251
## F-statistic: 16.9 on 4 and 82 DF,  p-value: 3.86e-10
##
##
## Value of test-statistic is: -3.3354 3.7961 5.566
##
## 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 statistique du test ADF vaut -3.3354, elle est supérieure aux seuils de rejets pour les risques de première espèce de 1% et de 5% (lus sur la ligne tau3). On ne peut donc pas rejeter l'hypothèse nulle de racine unitaire.

Pour la significativité de la tendance déterministe, les hypothèses de test sont celles de Dickey Fuller comme l'on a pas rejeté l'hypothèse nulle, alors :

$$H_0 : b = 0 \text{ contre } H_1 : b \neq 0$$

On va regarder la valeur critique de Dickey Fuller au seuil de 5%. La valeur critique vaut -3.15, que l'on compare à la t-stat associée à la tendance déterministe en valeur absolue : -1.866. $1.866 < 3.45$, donc on ne rejette pas H_0 , la tendance déterministe n'est pas significative, on passe donc au modèle 2, avec constante uniquement.

Le modèle 2 avec la constante uniquement On travaille ici avec:

$$\Delta X_t = c + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
adflogIDE2<-ur.df(logIDE,type=c("drift"), lag=4, selectlags = c('AIC'))
summary(adflogIDE2)
```

```
##
## #####
## # 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
## -3.08513 -0.40989 -0.00407  0.43765  1.53776
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2741     0.1610   1.702 0.092540 .
## z.lag.1       -0.3017     0.1310  -2.303 0.023834 *
```

```
## z.diff.lag1 -0.5858      0.1462  -4.005 0.000135 ***
## z.diff.lag2 -0.4198      0.1501  -2.797 0.006429 **
## z.diff.lag3 -0.1765      0.1247  -1.415 0.160933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7501 on 82 degrees of freedom
## Multiple R-squared:  0.4422, Adjusted R-squared:  0.4149
## F-statistic: 16.25 on 4 and 82 DF,  p-value: 7.737e-10
##
##
## Value of test-statistic is: -2.3026 2.8195
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

La statistique du test ADF vaut -2.3026, elle est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10% (lus sur la ligne tau3). On ne peut donc pas rejeter l'hypothèse nulle de racine unitaire.

Pour la significativité de la constante, les hypothèses de test sont les suivantes:

$$H_0 : c = 0 \text{ contre } H_1 : c \neq 0$$

Comme nous n'avons pas rejeté l'hypothèse nulle de racine unitaire dans le cadre du modèle 2, on teste la significativité de la constante à l'aide de la statistique de Dickey Fuller.

Pour la significativité de la constante, on va donc regarder la valeur critique de Dickey Fuller au seuil de 5% qui vaut -2.58 que l'on compare en valeur absolue à la t-stat de la constante qui vaut 1.702. On a $1.702 < 2.58$, donc on ne rejette pas H_0 , la constante n'est pas significative, on passe donc au modèle 1.

Le modèle 1 sans la constante ni la tendance déterministe On travaille ici avec:

$$\Delta X_t = \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
adflogIDE1<-ur.df(logIDE,type=c("none"), lag=4, selectlags = c('AIC'))
summary(adflogIDE1)
```

```
##
## #####
## # 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
## -2.9235 -0.4192  0.1086  0.4528  1.6831
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.10191    0.06613  -1.541 0.127177
## z.diff.lag1  -0.78354    0.11876  -6.598 3.82e-09 ***
## z.diff.lag2  -0.59004    0.14577  -4.048 0.000117 ***
## z.diff.lag3  -0.35038    0.14947  -2.344 0.021493 *
## z.diff.lag4  -0.17873    0.12462  -1.434 0.155303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7538 on 82 degrees of freedom
## Multiple R-squared:  0.4366, Adjusted R-squared:  0.4022
## F-statistic: 12.71 on 5 and 82 DF,  p-value: 3.792e-09
##
##
## Value of test-statistic is: -1.541
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

La statistique du test ADF vaut -1.541, elle est supérieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10% (lus sur la ligne tau3). On ne peut donc pas rejeter l'hypothèse nulle de racine unitaire.

On en conclut que notre processus est DS intégré d'ordre 1, ou I(1). Lorsqu'on dit qu'un processus est un processus DS intégré d'ordre 1, cela signifie que la série temporelle doit être différenciée une fois pour devenir stationnaire, donc qu'elle ne l'est pas à l'origine.

Vérifions ces résultats à l'aide du test KPSS.

Test KPSS avec la constante uniquement : On rappelle les hypothèses du test:

$$H_0 : \phi < 1 \Rightarrow \text{pas de racine unitaire contre } H_1 : \phi = 1 \Rightarrow \text{racine unitaire}$$

Le modèle est ici:

$$\Delta X_t = c + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
kpss_logIDE1 <- ur.kpss(y=logIDE, type="tau", lags="short" )
summary(kpss_logIDE1)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: tau with 3 lags.
##
## Value of test-statistic is: 0.1537
##
```

```
## Critical value for a significance level of:
##           10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
```

La statistique du test KPSS vaut 0.1537, elle est supérieure aux valeurs de rejet pour les risques de première espèce de 5% et 10%. On ne peut donc pas rejeter l'hypothèse nulle de stationarité. On effectue à nouveau le test avec l'ajout d'une tendance déterministe.

Test KPSS avec la constante et la tendance déterministe Le modèle est le suivant:

$$\Delta X_t = c + bt + \phi X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t$$

```
kpss_logIDE2 <- ur.kpss(y=logIDE, type="mu", lags="short" )
summary(kpss_logIDE2)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 1.0668
##
## 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 vaut ici 1.0668. Elle est supérieure aux valeurs de rejet pour les risques de première espèce de 1%, 5% et 10%. On ne peut donc pas rejeter l'hypothèse nulle de stationarité.

Comme trouvé précédemment, le log des IDE est donc bien une variable non stationnaire.

Il est donc nécessaire de la différencier afin de la rendre stationnaire.

c. Différenciation de la variable non stationnaire : le log des IDE

Faisons la différence première de la série du log des IDE.

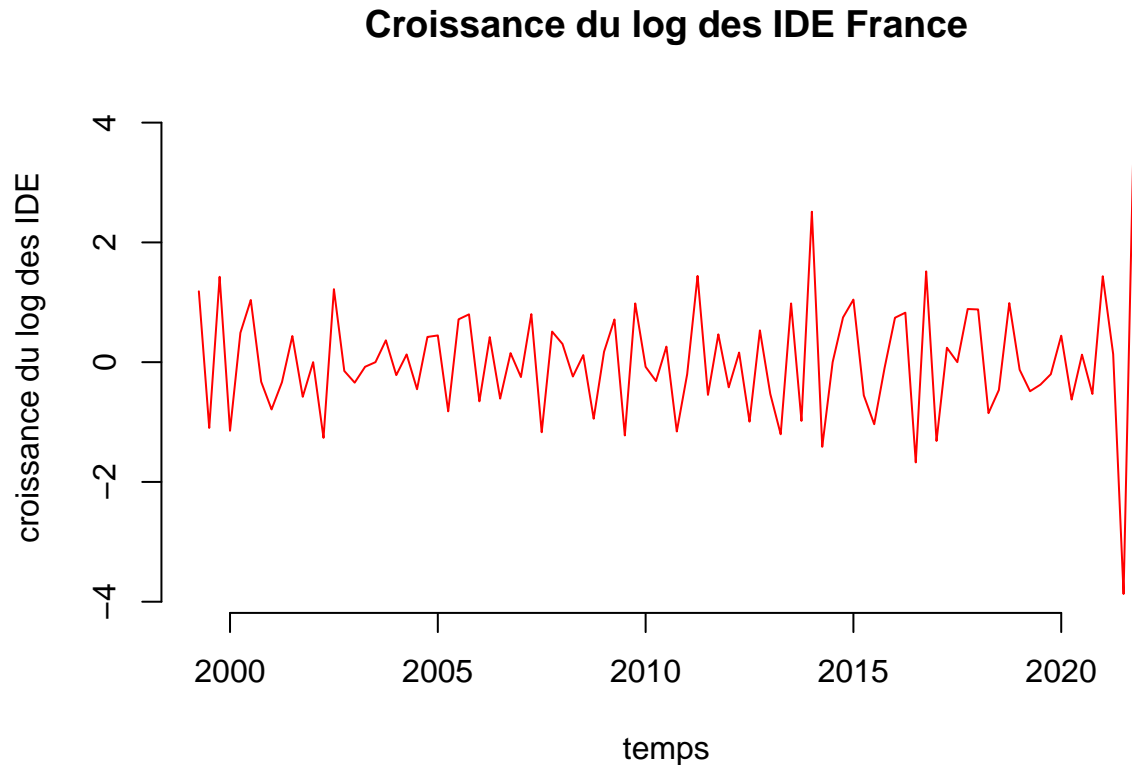
La différence première, obtenue à l'aide de la fonction `diff()` en R, représente la variation entre chaque valeur successive dans une série temporelle. Elle est utilisée pour analyser la tendance de croissance ou de décroissance dans les séries temporelles.

En appliquant la différence première au logarithme des IDE, on obtient la variation relative ou le taux de croissance instantané des IDE entre les périodes de temps successives.

```
DIFFlogIDE <-diff(logIDE)
```

Représentation graphique du taux de croissance du log des IDE de la France

```
plot.ts(DIFFlogIDE,main="Croissance du log des IDE France",frame=F,
        xlab='temps',ylab='croissance du log des IDE',col="red")
```



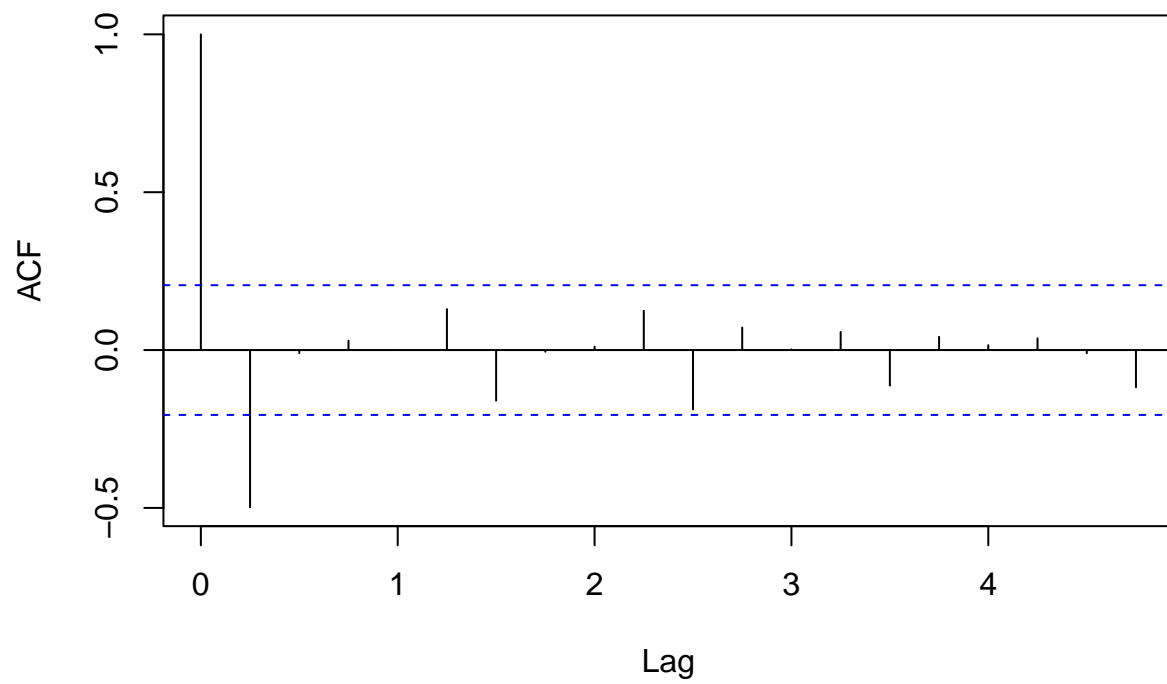
Au vu du graphique, on peut maintenant supposer que la série est stationnaire. Pour le confirmer, on réapplique à notre série différenciée les tests de racine unitaire de Dickey Fuller et les tests KPSS.

On représente maintenant les autocorrélogrammes de la série différenciée :

Autocorrélations simples du taux de croissance du log des IDE de la France

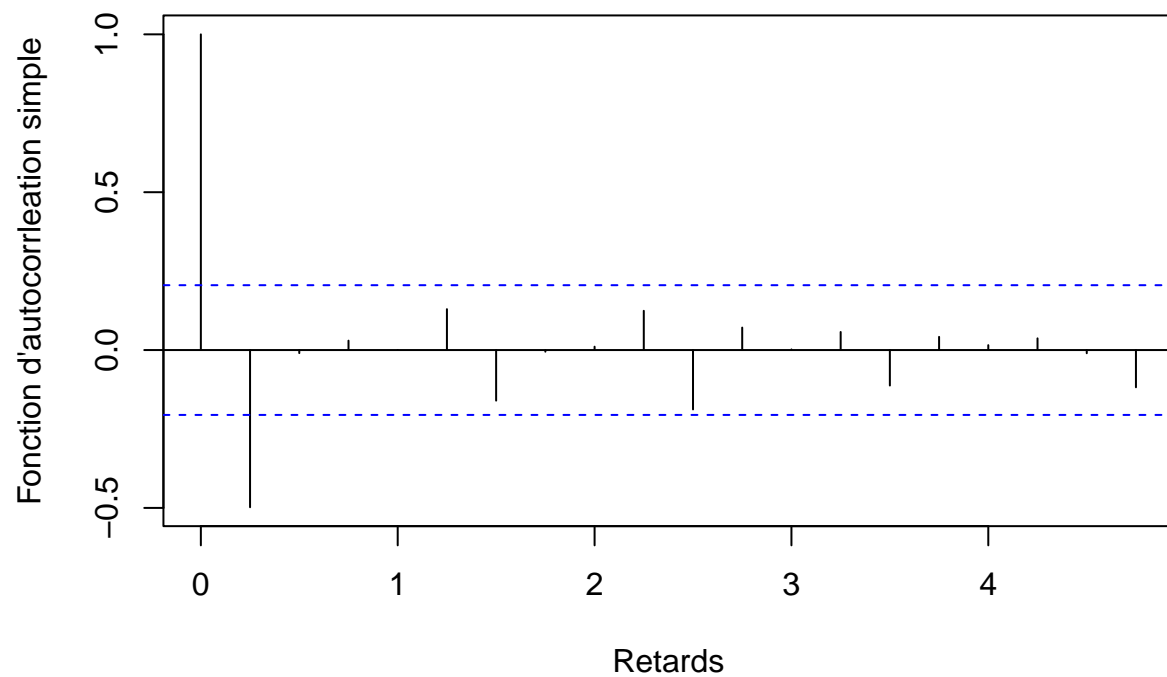
```
acfDIFFlogIDE<-acf(DIFFlogIDE)
```


Series DIFFlogIDE



```
plot(acfDIFFlogIDE,xlab="Retards",ylab="Fonction d'autocorrleation simple",main="ACF du log des IDE dif
```

ACF du log des IDE difference

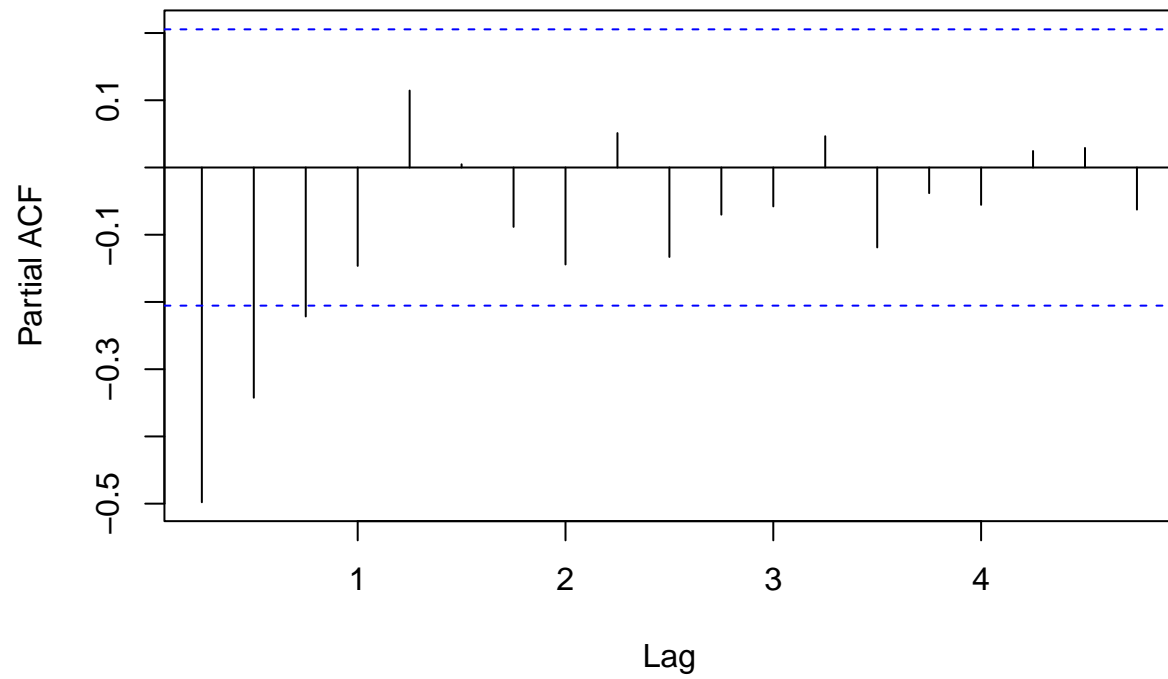


On observe que les autocorrélations ne sont plus significatives à partir du premier retard.

Autocorrélations partielles du taux de croissance du log des IDE de la France

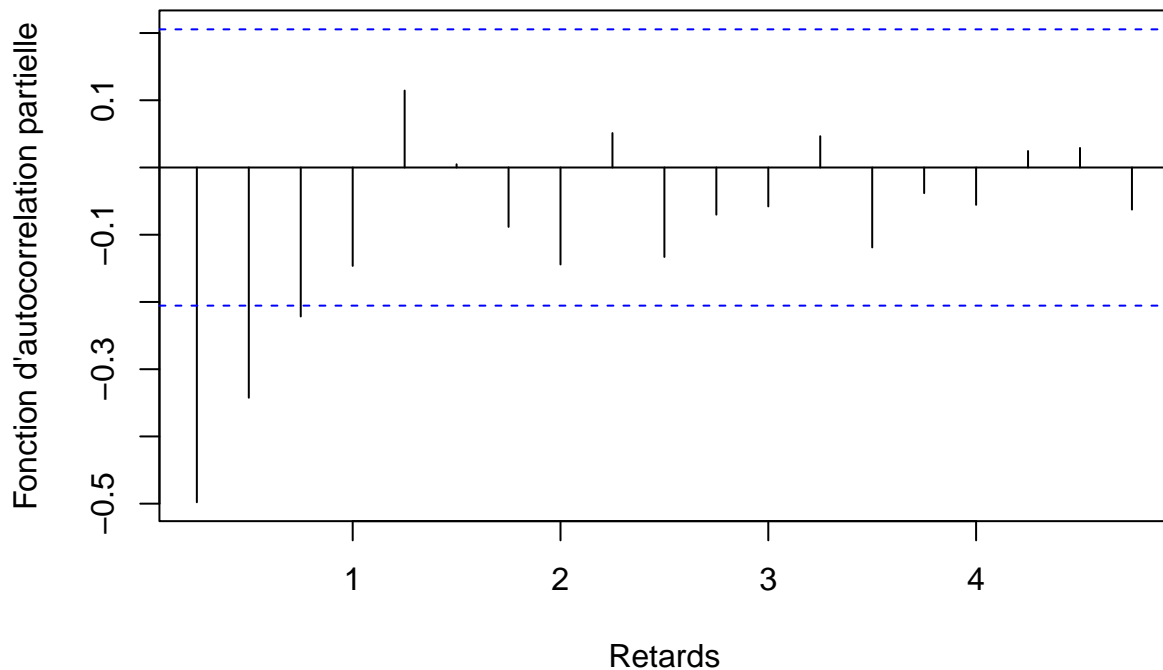
```
pacfDIFFlogIDE<-pacf(DIFFlogIDE)
```

Series DIFFlogIDE



```
plot(pacfDIFFlogIDE,xlab="Retards",ylab="Fonction d'autocorrelation partielle",main="PACF du log des IDE")
```

PACF du log des IDE differenceie



De la même manière, les autocorrélations partielles ne sont plus significatives à partir du premier retard.

Application du test de Dickey Fuller sur notre série différenciée :

Modèle avec tendance déterministe et constante: Le nouveau modèle général se présente sous cette forme:

$$\Delta X_t = c + bt + Yt$$

On teste la stationarité du modèle avec tendance déterministe et constante. Le nombre de retards maximum est fixé à 4 (toujours car les données sont trimestrielles) et le nombre de retards optimal est déterminé par la minimisation du critère AIC. On rappelle les hypothèses de test :

$$H_0 : \phi = 1 \Rightarrow \text{racine unitaire contre } H_1 : \phi < 1 \Rightarrow \text{pas de racine unitaire}$$

```
library(urca)
adfDIFFlogIDE3<-ur.df(DIFFlogIDE,type=c("trend"), lag=4, selectlags = c('AIC'))
summary(adfDIFFlogIDE3)
```

```
##
## #####
## # 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
## -2.97900 -0.47929  0.02905  0.42502  1.78663
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0993943  0.1802579  -0.551  0.582896
## z.lag.1      -3.0991825  0.4174534  -7.424  1.08e-10 ***
## tt           0.0008945  0.0033613   0.266  0.790843
## z.diff.lag1  1.2423573  0.3512222   3.537  0.000676 ***
## z.diff.lag2  0.5893548  0.2491289   2.366  0.020418 *
## z.diff.lag3  0.2037441  0.1283453   1.587  0.116351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7711 on 80 degrees of freedom
## Multiple R-squared:  0.7905, Adjusted R-squared:  0.7774
## F-statistic: 60.36 on 5 and 80 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -7.424 18.3722 27.5581
##
## 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 statistique du test ADF vaut -7.424, elle est inférieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10% (lus sur la ligne tau3). On rejette donc l'hypothèse nulle de racine unitaire.

La tendance déterministe est-elle significative ?

$$H_0 : b = 0 \text{ contre } H_1 : b \neq 0$$

La t-stat de la tendance déterministe est égale à 0.266. La valeur critique qui figure sur la table de Student est égale à 2.86 pour un risque de première espèce de 5%, l'hypothèse nulle n'est donc pas rejetée, c'est-à-dire la tendance déterministe n'est pas significative, donc on l'enlève.

On regarde maintenant le modèle 2 avec la constante uniquement.

Modèle avec la constante uniquement Ici on travaille avec:

$$\Delta X_t = c + Y_t$$

```
adfDIFFlogIDE2<-ur.df(DIFFlogIDE,type=c("drift"), lag=4, selectlags = c('AIC'))
summary(adfDIFFlogIDE2)
```

```
##
```

```
## #####
## # 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
## -2.94393 -0.47705  0.01892  0.40486  1.79681
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.05689    0.08306  -0.685 0.495336
## z.lag.1      -3.09513    0.41478  -7.462 8.53e-11 ***
## z.diff.lag1   1.23759    0.34875   3.549 0.000648 ***
## z.diff.lag2   0.58664    0.24749   2.370 0.020147 *
## z.diff.lag3   0.20280    0.12756   1.590 0.115756
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7666 on 81 degrees of freedom
## Multiple R-squared:  0.7903, Adjusted R-squared:  0.7799
## F-statistic: 76.31 on 4 and 81 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -7.4622 27.8423
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

La statistique du test ADF vaut -7.4622, elle est inférieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10% (lus sur la ligne tau2). On rejette donc l'hypothèse nulle de racine unitaire.

Est-ce que la constante est significative ?

$$H_0 : c = 0 \text{ contre } H_1 : c \neq 0$$

La t-stat de la constante est égale à -0.685. Comme l'on a rejeté l'hypothèse de racine unitaire dans le cadre du modèle 2, on teste la significativité de la constante à l'aide de la table de Student.

Le seuil de rejet qui figure sur la table est égal à 2.86 pour un risque de première espèce de 5% sur un test bilatéral, donc on ne rejette pas l'hypothèse nulle. La constante n'est donc pas significative.

On teste donc maintenant le modèle 1 sans constante ni tendance déterministe.

Modèle sans la constante et sans la tendance déterministe Le modèle est:

$$\Delta X_t = Y_t$$

```
adfDIFFlogIDE3<-ur.df(DIFFlogIDE,type=c("none"), lag=4, selectlags = c('AIC'))
summary(adfDIFFlogIDE3)
```

```
##
## #####
## # 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
## -3.00895 -0.53841 -0.03457  0.35009  1.74994
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1         -3.0700     0.4118  -7.455 8.31e-11 ***
## z.diff.lag1      1.2188     0.3465   3.517 0.000714 ***
## z.diff.lag2      0.5748     0.2461   2.336 0.021946 *
## z.diff.lag3      0.1972     0.1269   1.554 0.123934
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7642 on 82 degrees of freedom
## Multiple R-squared:  0.7892, Adjusted R-squared:  0.7789
## F-statistic: 76.75 on 4 and 82 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -7.4549
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

La statistique du test ADF vaut -7.4549, elle est inférieure aux seuils de rejets pour les risques de première espèce de 1%, 5% et 10% (lus sur la ligne tau1). On rejette donc l'hypothèse nulle de racine unitaire.

On en conclut que notre nouveau processus Y_t différencié est $I(0)$. C'est un processus stationnaire sans dérive.

La différence première a bien permis de stationnariser notre série, on vérifie tout de même la stationnarité du nouveau processus autour d'une constante avec un test KPSS.

Test de stationarité KPSS sur notre nouveau processus

On rappelle les hypothèses du test KPSS:

$$H_0 : \phi < 1 \Rightarrow \text{pas de racine unitaire contre } H_1 : \phi = 1 \Rightarrow \text{racine unitaire}$$

Modèle avec constante Soit :

$$\Delta X_t = c + Y_t$$

```
kpss_DIFFlogIDE1 <- ur.kpss(y=DIFFlogIDE, type="tau", lags="short" )
summary(kpss_DIFFlogIDE1)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: tau with 3 lags.
##
## Value of test-statistic is: 0.037
##
## Critical value for a significance level of:
##          10pct  5pct  2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
```

On trouve une statistique de test égale à 0.037, elle est inférieure aux seuils de rejet pour les risques de première espèce de 1%, 5% et 10%. On ne rejette donc pas l'hypothèse nulle de stationarité.

Modèle avec constante et tendance déterministe : Soit :

$$\Delta X_t = c + bt + Y_t$$

```
kpss_DIFFlogIDE2 <- ur.kpss(y=DIFFlogIDE, type="mu", lags="short" )
summary(kpss_DIFFlogIDE2)
```

```
##
## #####
## # KPSS Unit Root Test #
## #####
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.0457
##
## Critical value for a significance level of:
##          10pct  5pct  2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

On trouve une statistique de test égale à 0.0457, elle est inférieure aux seuils de rejet pour les risques de première espèce de 1%, 5% et 10%. On ne rejette donc pas l'hypothèse nulle de stationarité.

Les tests KPSS confirment bien les tests de Dickey Fuller. Notre série différenciée est dorénavant stationnaire. La stationnarité avait d'ailleurs été remarquée dans l'observation du graphe de la série différenciée du log des IDE. Notre processus initial X_t est $I(1)$, donc DS sans dérive.

4. Identification et validation du modèle ARMA

Identification du modèle

Nous choisissons ici d'étudier la série sur le PIB, puisque l'on a conclu dans la section précédente qu'elle était $I(0)+C$, donc stationnaire. Nous rappelons que nous travaillons avec 92 observations. Dans la suite, nous retirerons les 8 observations des périodes 2020-2021, les données n'étant pas représentatives à cause de la crise du COVID, et nous retirerons aussi les 3 dernières observations de l'année 2019, dans le but de comparer les prévisions faites sur les horizons 1 à 3 avec les valeurs observées. Après observation des autocorrélogrammes réalisés plus tôt sur la série, nous remarquons que la fonction d'autocorrélation simple diminue vers 0, avec seulement le premier terme significatif. L'étude de la fonction de l'autocorrélation partielle traduit une décroissance brusque vers zéro avec uniquement la première autocorrélation significativement différente de zéro. Suite à notre première analyse, nous pouvons supposer que nous pouvons identifier notre processus comme un processus AR (1).

Nous allons effectuer la méthode de la recherche de grille pour déterminer la meilleure configuration des paramètres du modèle ARMA en fonction des critères d'information associés, le but étant de minimiser les critères d'information, ici les critères AIC et BIC. Le code associé à cette méthode est le suivant:

```
# Initialisation des valeurs
p_values <- 1:5
q_values <- 1:5
best_aic <- Inf
best_bic <- Inf
best_order_aic <- c(0, 0)
best_order_bic <- c(0, 0)

# Boucle pour la selection de modele
for (p in p_values) {
  for (q in q_values) {
    arma_model <- arima(PIB[1:81], order = c(p, 0, q))
    current_aic <- AIC(arma_model)
    current_bic <- BIC(arma_model)
    if (current_aic < best_aic) {
      best_aic <- current_aic
      best_order_aic <- c(p, 0, q)
    }
    if (current_bic < best_bic) {
      best_bic <- current_bic
      best_order_bic <- c(p, 0, q)
    }
  }
}

#Affichage des resultats
cat("Meilleure combinaison d'ordres pour AIC (p, d, q):", best_order_aic, "\n")

## Meilleure combinaison d'ordres pour AIC (p, d, q): 3 0 2

cat("Meilleur AIC:", best_aic, "\n")

## Meilleur AIC: 103.5714
```

```
cat("Meilleure combinaison d'ordres pour BIC (p, d, q):", best_order_bic, "\n")
```

```
## Meilleure combinaison d'ordres pour BIC (p, d, q): 1 0 2
```

```
cat("Meilleur BIC:", best_bic, "\n")
```

```
## Meilleur BIC: 117.2106
```

On obtient que le modèle minimisant le critère d'information AIC est l'ARMA(3,2), avec un AIC qui vaut 103,57.

Concernant le critère BIC, c'est l'ARMA(1,2) qui minimise le critère, avec un BIC qui vaut 117,21.

Au vu des résultats observés sur les différents autocorrélogrammes, il nous semble plus pertinent de commencer à travailler un modèle ARMA(1,2).

Validation du modèle

Nous déterminons donc la validité du modèle suivant:

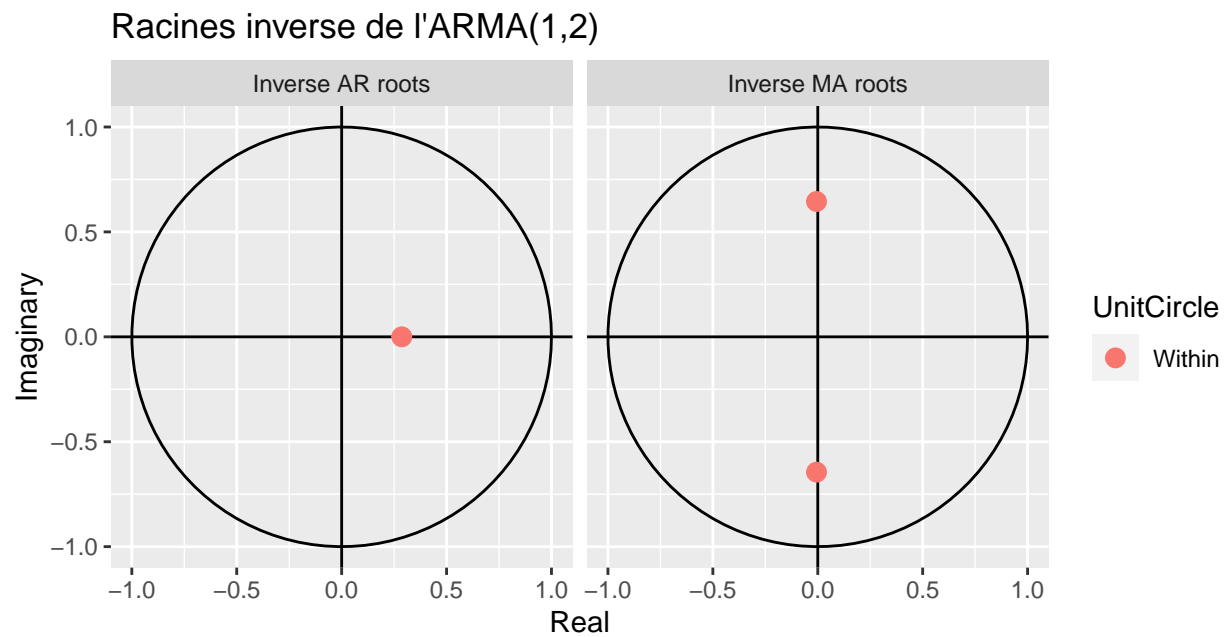
$$X_t = c + \phi X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

```
library(forecast)
library(TTR)
library(ggplot2)
ARMA12 <- arima(PIB[1:81], order = c(1, 0, 2))
summary(ARMA12)
```

Test de stationnarité et inversibilité du modèle

```
##
## Call:
## arima(x = PIB[1:81], order = c(1, 0, 2))
##
## Coefficients:
##          ar1      ma1      ma2  intercept
##          0.2867  0.0121  0.4160      0.3783
## s.e.      0.2291  0.1996  0.1637      0.0955
##
## sigma^2 estimated as 0.1885:  log likelihood = -47.62,  aic = 105.24
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.0002600285 0.4341439 0.3367752 227.8594 286.7731 0.8084158
##              ACF1
## Training set 0.009063318
```

```
# Tracé des racines inverses
autoplot(ARMA12, main = "Racines inverse de l'ARMA(1,2)")
```



Les inverses des racines sont bien contenues à l'intérieur du cercle unitaire, preuve de la stationnarité et de l'inversibilité du processus.

Tests de significativité des coefficients

$$H_0 : \phi_i = 0 \text{ contre } H_1 : \phi_i \neq 0$$

```
library(lmtest)
coeftest(ARMA12)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1      0.286666  0.229138  1.2511  0.21091
## ma1      0.012148  0.199599  0.0609  0.95147
## ma2      0.415989  0.163656  2.5419  0.01103 *
## intercept 0.378268  0.095506  3.9607 7.474e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La stat de test suit une loi normale centrée réduite sous H_o puisque l'on observe 81 observations. Les seuils critiques sont 2.57 pour un risque de première espèce égal à 1%, et 1.96 pour un risque de première espèce égal à 5%, d'après la table de la loi normale.

La constante: La stat de test vaut 3.96 pour la constante, donc est supérieure aux seuils critiques à 1% et 5%, on rejette donc H_o , et on conclut que la constante est significative. Nous la conservons dans le modèle.

Le coefficient AR: Sa stat de test vaut 1.25, donc on ne peut pas rejeter H_o , autrement dit le coefficient n'est pas significatif et on peut l'enlever. Cette conclusion est cohérente avec la p-valeur, ici supérieure à alpha pour tous les niveaux.

Le modèle s'inscrit désormais sous la forme d'un MA(2) avec constante:

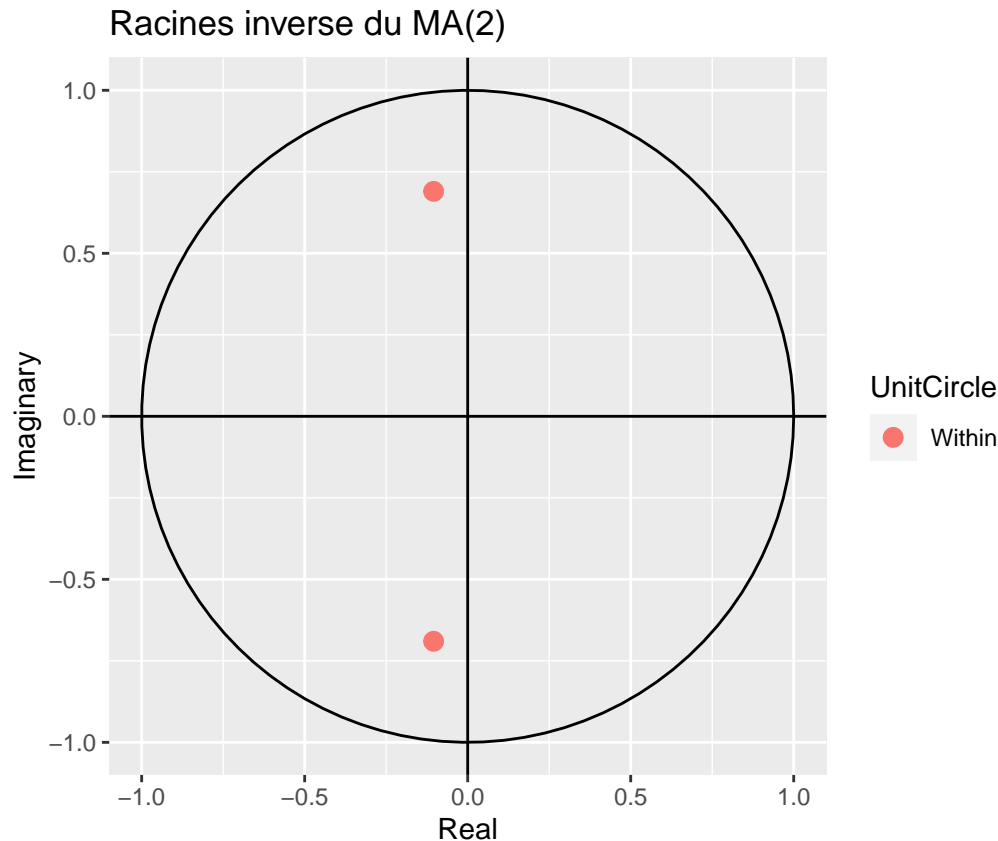
$$X_t = c + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

On procède à une nouvelle estimation du modèle:

```
MA2<-Arima(PIB[1:81],order=c(0,0,2))
coeftest(MA2)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1      0.208955   0.094139  2.2196 0.0264434 *
## ma2      0.487036   0.133411  3.6506 0.0002616 ***
## intercept 0.375975   0.082133  4.5777 4.702e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
autoplot(MA2,main="Racines inverse du MA(2)")
```



Les coefficients sont ici tous significatifs, donc on garde ce modèle. Ces conclusions sont bien cohérentes avec les p-valeurs obtenues.

À nouveau, le processus est bien stationnaire et inversible.

On note ici qu'initialement le premier coefficient du MA n'était pas non plus significatif. En choisissant d'enlever celui-ci, on obtenait un ARMA(1,1) avec tous ses coefficients significatifs. Or, on voit que c'est le modèle MA(2) qui minimise les deux critères d'information AIC et BIC, c'est donc avec celui-ci que nous travaillerons par la suite.

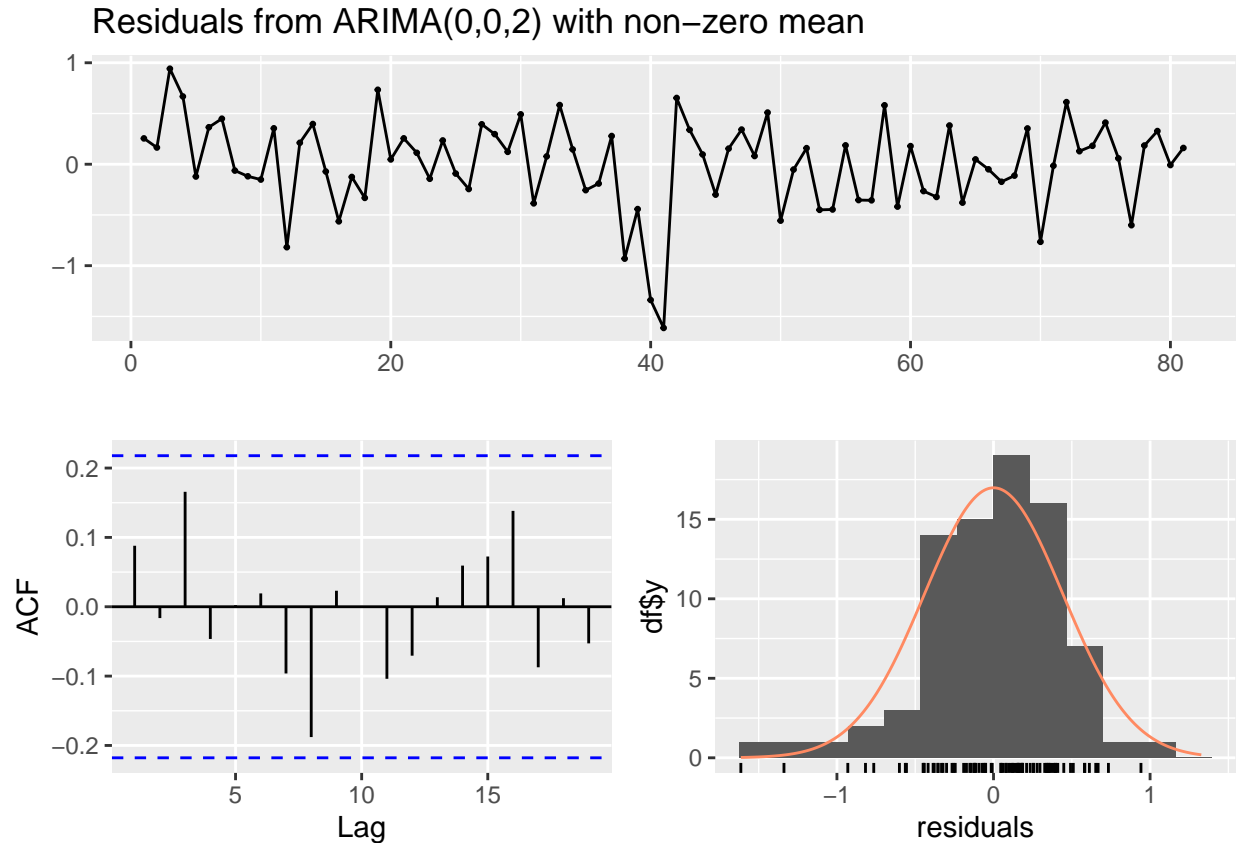
Tests sur les résidus

Test d'absence d'autocorrélation de Ljung-Box Les hypothèses du test sont les suivantes:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_H = 0 \text{ (absence d'autocorrélation) contre } H_1 : \overline{H}_0 \text{ (hypothèse alternative)}$$

On décide du rejet ou non-rejet de H_0 en regardant les p-values associées aux statistiques de Ljung-Box :

```
#Test de Ljung-Box
library(forecast)
LB<-checkresiduals(MA2,lag = 10)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,2) with non-zero mean
## Q* = 7.4042, df = 8, p-value = 0.4937
##
## Model df: 2.   Total lags used: 10
```

Avec 10 retards, l'autocorrélogramme montre que toutes les autocorrélations sont situées dans les bornes de l'intervalle de confiance à 95% : aucune des autocorrélations n'est significativement différente de 0. Donc on ne rejette pas l'hypothèse nulle d'absence d'autocorrélation des résidus.

On pouvait également vérifier ce résultat avec la statistique du test de Ljung-Box: elle est égale à 7.4042.

Sous H_0 , la stat suit une loi du Chi-2 avec $10-2=8$ degrés de liberté, car on a utilisé ici 10 retards et notre modèle possède 2 paramètres non constants. Les seuils critiques sont donc tirés de la loi Chi-2(8), et valent:

- 13.36 pour un risque de première espèce de 10%
- 15.51 pour un risque de première espèce de 5%
- 20.09 pour un risque de première espèce de 1%.

Notre stat est donc inférieure à ces seuils critiques. On ne rejette donc pas l'hypothèse nulle d'absence d'autocorrélation des résidus.

Test d'absence d'effets ARCH de Engle On teste ici l'homoscédasticité des résidus, i.e s' ils ont tous la même variance.

Les hypothèses sont les suivantes:

$$H_0 : \gamma_1 = \dots = \gamma_4 = 0 \Rightarrow \text{pas d'effet ARCH contre } H_1 : \gamma_1 \neq 0 \text{ ou } \dots \gamma_4 \neq 0 \Rightarrow \text{effet ARCH}$$

Nous choisissons ici 5 retards pour le test.

```
#Test de Engle
install.packages("FinTS")
library(FinTS)
ArchTest(MA2$residuals,lags=5,demean = FALSE)

##
## ARCH LM-test; Null hypothesis: no ARCH effects
##
## data: MA2$residuals
## Chi-squared = 15.721, df = 5, p-value = 0.007686
```

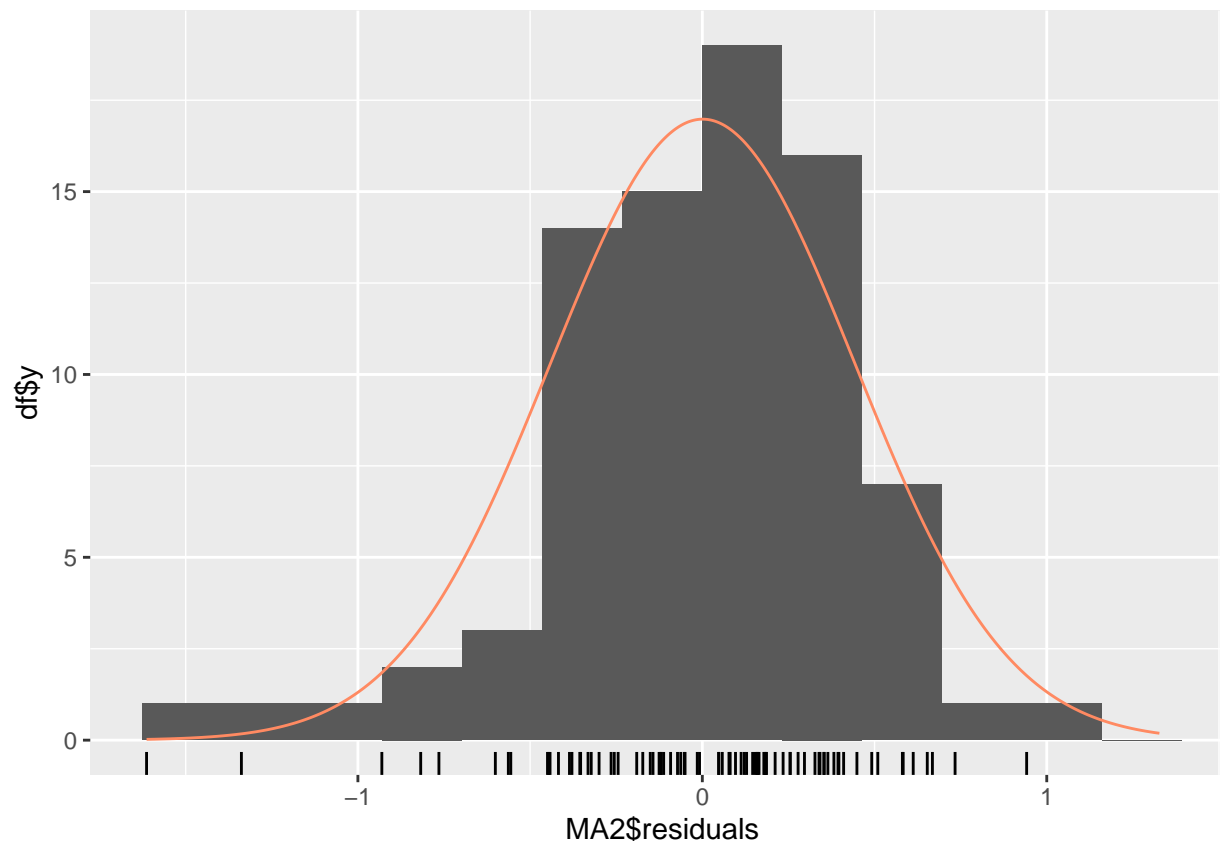
On obtient une stat de test qui vaut 15.721, et qui sous H_0 suit une loi du Chi-2(5), puisque nous avons choisi ici 5 retards. En lisant la table du Chi-2(5), on obtient les seuils critiques suivants:

- 9.24 pour un risque de première espèce de 10%
- 11.07 pour un risque de première espèce de 5%
- 15.09 pour un risque de première espèce de 1%.

On rejette donc toujours l'hypothèse nulle d'absence d'effet ARCH.

Test de normalité de Jarque-Bera Ce test permet de vérifier la normalité des résidus.

```
gghistogram(MA2$residuals, add.normal = TRUE)
```



```
install.packages("moments")
library(moments)
sprintf("Coefficient d'asymetrie des residus : %f",skewness(MA2$residuals))
```

```
## [1] "Coefficient d'asymetrie des residus : -0.917361"
```

```
sprintf("Coefficient d'aplatissement : %f",kurtosis(MA2$residuals))
```

```
## [1] "Coefficient d'aplatissement : 4.761215"
```

```
library(tseries)
jarque.bera.test(MA2$residuals)
```

```
##
## Jarque Bera Test
##
## data: MA2$residuals
## X-squared = 21.83, df = 2, p-value = 1.819e-05
```

On trouve un coefficient d'asymétrie des résidus égal à -0.917361 et un coefficient d'aplatissement égal à 4.761215.

Les hypothèses du test sont les suivantes::

$$H_0 : S(X) = 0 \text{ et } K(X) = 3 \text{ contre } H_1 : S(X) \neq 0 \text{ ou } K(X) \neq 3$$

S et K désignent respectivement le coefficient d'asymétrie et le coefficient d'aplatissement, et valent respectivement 0 et 3 pour une loi normale centrée réduite.

La statistique du test vaut 21.83, et suit sous H_0 une loi du Chi-2(2). Avec sa table, on trouve des seuils critiques qui valent:

- 4.61 pour un risque de première espèce de 10%
- 5.99 pour un risque de première espèce de 5%
- 9.21 pour un risque de première espèce de 1%.

Donc on rejette l'hypothèse nulle, c'est-à-dire que les résidus ne suivent pas une loi normale. On aurait pu également le voir avec la p-valeur qui vaut ici 1.819e-05, qui est bien inférieure aux alpha.

5. Prévisions du modèle sur les horizons 1 à 3

Puisque nous travaillons avec un MA(2), nous aurons ici besoin des deux derniers résidus, en t-1 et t-2, pour estimer en t. On rappelle que notre modèle s'écrit:

$$X_t = c + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Commençons par estimer les coefficients du modèle.

```
MA2<-Arima(PIB[1:81],order=c(0,0,2))
coeftest(MA2)

##
## z test of coefficients:
##
##          Estimate Std. Error z value  Pr(>|z|)
## ma1         0.208955   0.094139  2.2196 0.0264434 *
## ma2         0.487036   0.133411  3.6506 0.0002616 ***
## intercept   0.375975   0.082133  4.5777 4.702e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Donc notre modèle s'écrit:

$$X_t = 0.375975 + \varepsilon_t - 0.208955\varepsilon_{t-1} - 0.487036\varepsilon_{t-2}$$

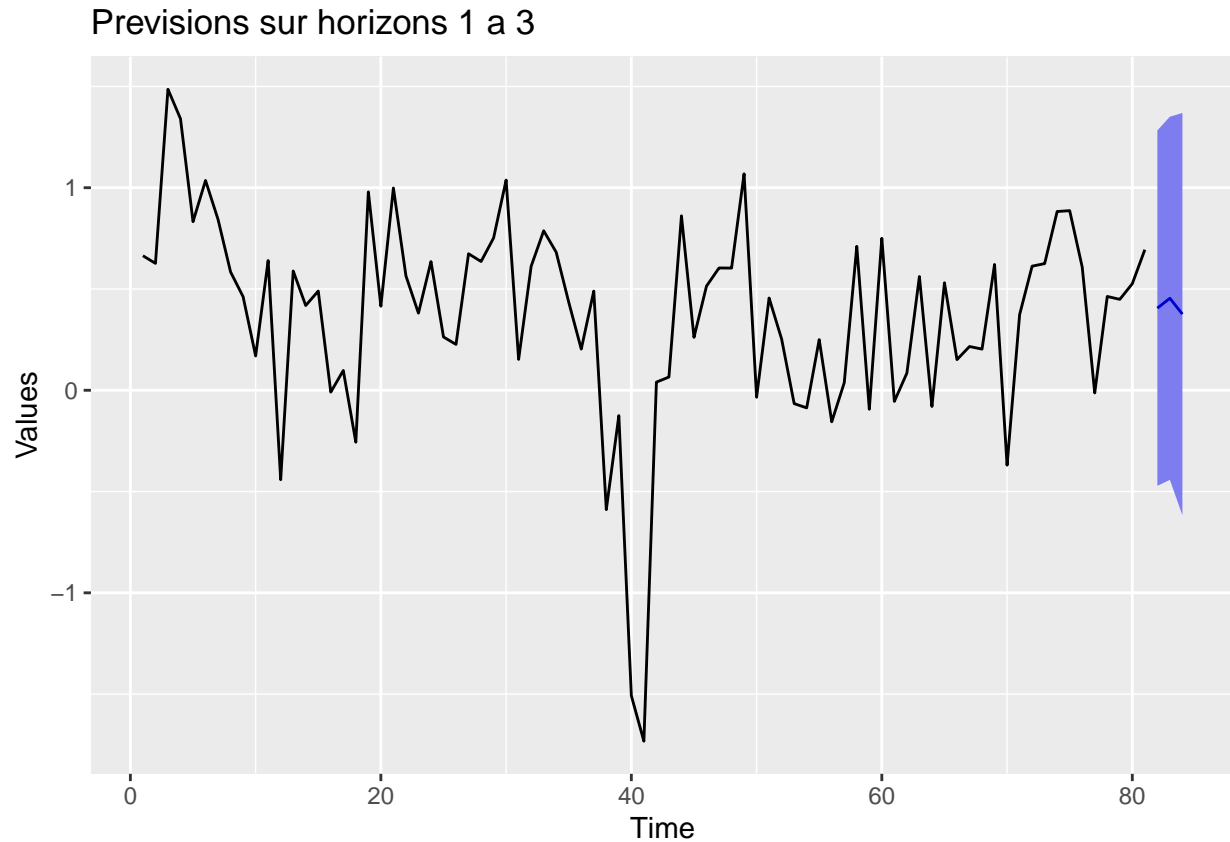
La valeur indiquée par R au niveau "intercept" est normalement celle de l'espérance, mais ici est directement celle de la constante puisque l'espérance de la variable est égale à la constante dans le cadre d'un modèle MA.

On calcule les prévisions sur la série modélisée, sur les horizons 1 à 3, c'est-à-dire les trois derniers trimestres de 2019, avec un intervalle de prévision de niveau 95%.

```
library(forecast)
previsions_MA2<-forecast(MA2,h=3,level=95)
previsions_MA2
```

```
##      Point Forecast      Lo 95      Hi 95
## 82      0.4055598 -0.4718643 1.282984
## 83      0.4541419 -0.4422327 1.350516
## 84      0.3759752 -0.6170527 1.369003
```

```
autoplot(previsions_MA2,ylab="Values",main="Previsions sur horizons 1 a 3")
```



Prévision pour 2019 Q2: h=1

Nous avons ici besoin des erreurs de 2019 Q1 et 2018 Q4.

$$X_t = 0.375975 + \varepsilon_t - 0.208955\varepsilon_{t-1} - 0.487036\varepsilon_{t-2}$$

L'erreur en 2018_Q4 vaut $0.52611517 - 0.5316451 = -0.00552993$. L'erreur en 2019_Q1 vaut $0.69404909 - 0.5239436 = 0.1701055$

Donc ici $\text{PIB}_{2019Q2} = 0.375975 - 0.208955 * 0.1701055 - 0.487036 * (-0.00552993) = 0.3431239$

Prévision pour 2019 Q3: h=2

Nous avons ici besoin de l'erreur en 2019 Q2. L'erreur en 2019_Q2 vaut $0.66797527 - 0.4055598 = 0.2624155$

Donc ici $\text{PIB}_{2019Q3} = 0.375975 - 0.208955 * 0.2624155 - 0.487036 * 0.1701055 = 0.2382945$

Prévision pour 2019 Q4: h=3

Nous avons ici besoin de l'erreur en 2019 Q3 L'erreur en 2019_Q3 vaut $0.10054614 - 0.4541419 = -0.3535958$

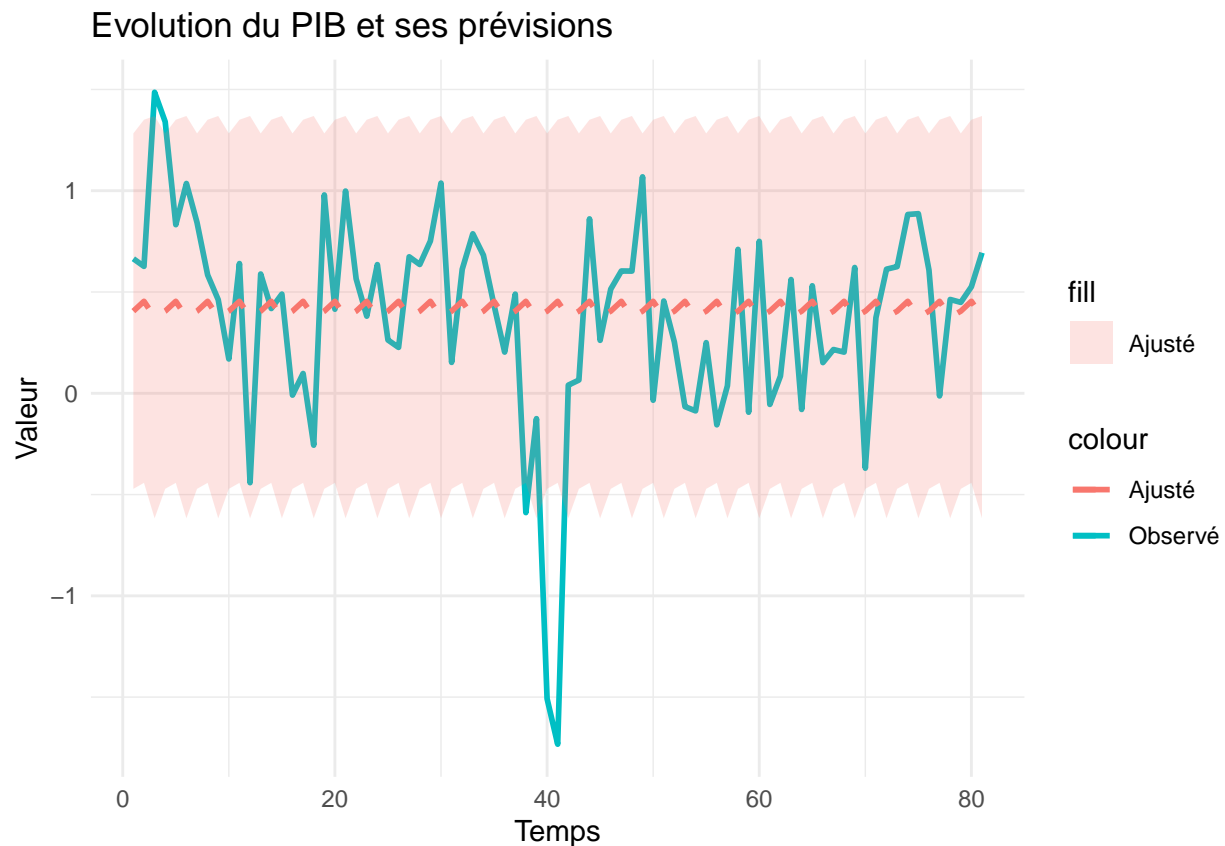
Donc ici $\text{PIB}_{2019Q4} = 0.375975 - 0.208955 * (-0.3535958) - 0.487036 * 0.2624155 = 0.3220548$

```

PIB_fusion <- data.frame(
  Time = time(PIB[1:81]),
  Observed = as.numeric(PIB[1:81]),
  Forecasted = as.numeric(previsions_MA2$mean),
  Upper = as.numeric(previsions_MA2$upper),
  Lower = as.numeric(previsions_MA2$lower)
)
library(reshape2)
data_melt=melt(as.data.frame(PIB_fusion),id.vars = 1)

ggplot(PIB_fusion, aes(x = Time)) +
  geom_line(aes(y = Observed, color = "Observé"), linewidth = 1) +
  geom_line(aes(y = Forecasted, color = "Ajusté"), linetype = "dashed", linewidth = 1) +
  geom_ribbon(aes(ymin = Lower, ymax = Upper, fill = "Ajusté"), alpha = 0.2) +
  labs(title = "Evolution du PIB et ses prévisions",
       x = "Temps",
       y = "Valeur") +
  theme_minimal()

```



Ainsi, nous sommes plutôt satisfaits de notre modèle. Les prévisions effectuées sont assez proches des données observées. Les différences peuvent s'expliquer de plusieurs manières. Bien sûr, notre modèle MA(2) n'est pas parfait, et cela explique certaines différences. Cependant, on peut aussi évoquer certains événements inhérents à la période 2019, qui ont pu réduire anormalement la croissance du PIB, notamment la crise des Gilets Jaunes, certaines réformes économiques et fiscales, mais également de manière plus générale, un ralentissement économique a été observé l'année 2019 à l'échelle mondiale, ralentissement auquel la France n'a pas échappé.

Partie 2 : Modélisation multivariée

6. Estimation du modèle VAR

Représentation des données

On utilise un modèle VAR (Vecteur Auto-Régressif) pour étudier les relations entre le taux de croissance du PIB de la France d'un trimestre à un autre, et le log des IDE différencié une fois. Nous cherchons ici à modéliser ces deux séries grâce aux informations sur leurs valeurs passées. La stationnarité des deux processus a été vérifiée au préalable dans la première partie.

On commence par représenter sur un même graphe l'évolution de ces deux variables.

```
install.packages("reshape2")

library(ggplot2)

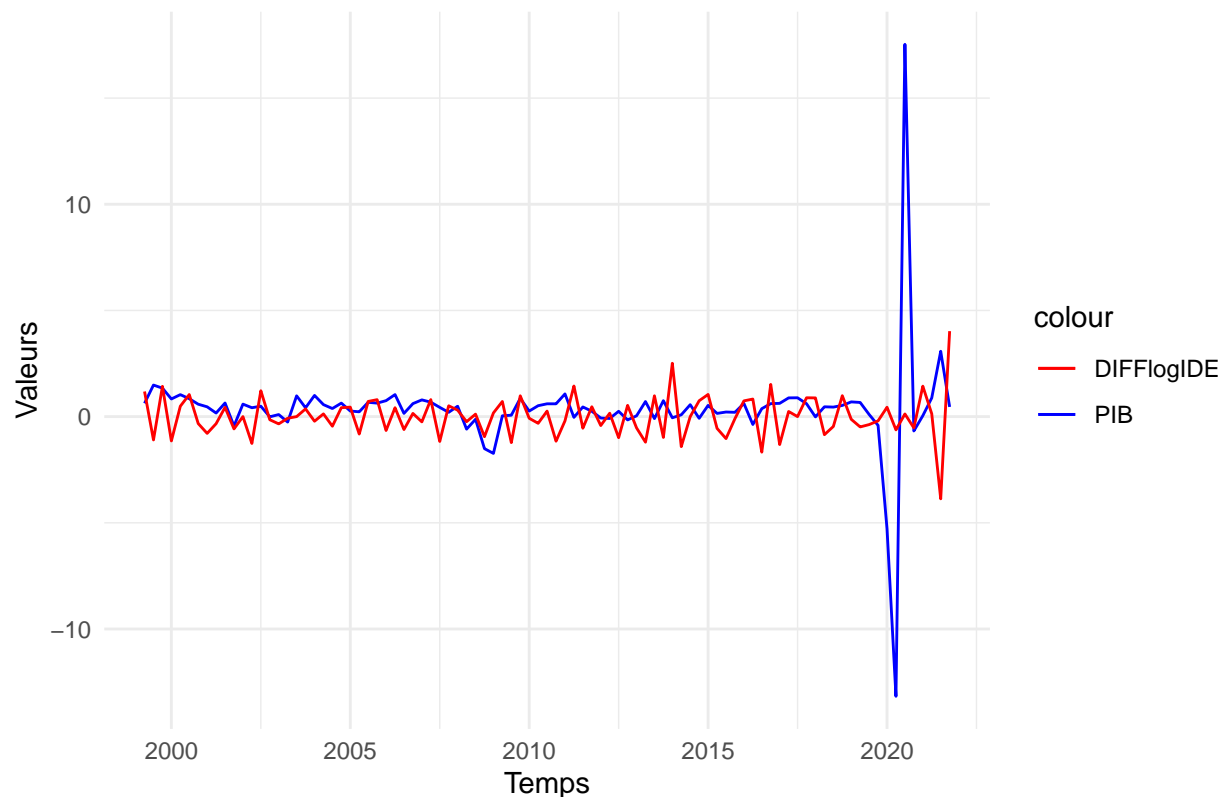
library(reshape2)

# Utilisation de ts.intersect pour obtenir les séries temporelles
endogen <- ts.intersect(PIB, DIFFlogIDE)

# Conversion des séries temporelles en dataframes
endogen_df <- data.frame(
  temps = time(endogen),
  PIB = as.numeric(endogen[, 'PIB']),
  DIFFlogIDE = as.numeric(endogen[, 'DIFFlogIDE'])
)

# Création du graphique avec ggplot en utilisant les dataframes
ggplot(data = endogen_df, aes(x = temps)) +
  geom_line(aes(y = PIB, color = "PIB")) +
  geom_line(aes(y = DIFFlogIDE, color = "DIFFlogIDE")) +
  labs(title = "Evolution de PIB et DIFFlogIDE en fonction du temps", x = "Temps", y = "Valeurs") +
  scale_color_manual(values = c(PIB = "blue", DIFFlogIDE = "red")) +
  theme_minimal()
```

Evolution de PIB et DIFFlogIDE en fonction du temps



En raison de la spécificité de la période Covid, nous choisissons de nous limiter dans la suite de cette étude à la période avant Covid : c'est à dire du 2ème trimestre de 1999 au 4ème trimestre de 2019.

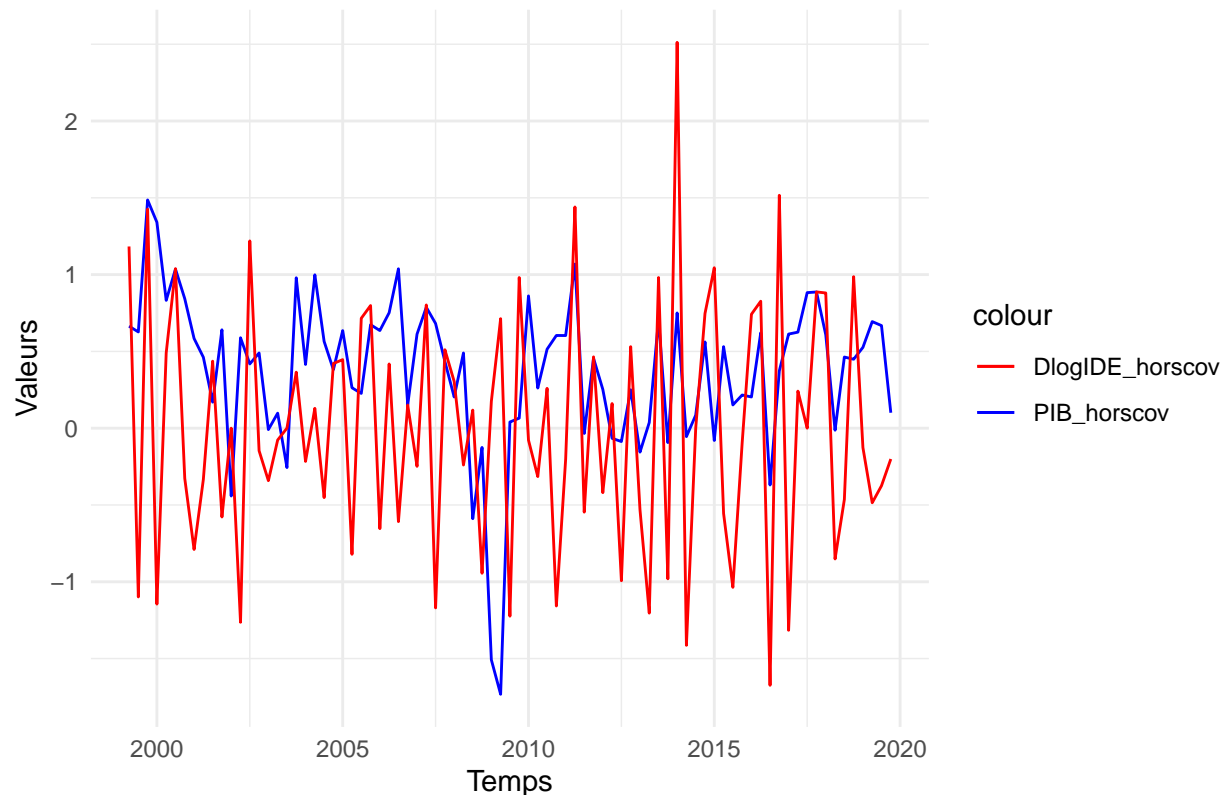
```
PIB_horscov <- ts(PIB[1:83],frequency = 4,start = c(1999,2))
DlogIDE_horscov <-ts(DIFFlogIDE[1:83],frequency = 4,start = c(1999,2))

endogen_horscov <- ts.intersect(PIB_horscov, DlogIDE_horscov)

# Conversion des séries temporelles en dataframes
endogen_horscov_df <- data.frame(
  temps = time(endogen_horscov),
  PIB_horscov = as.numeric(endogen_horscov[, 'PIB_horscov']),
  DlogIDE_horscov = as.numeric(endogen_horscov[, 'DlogIDE_horscov'])
)

# Création du graphique avec ggplot en utilisant les dataframes
ggplot(data = endogen_horscov_df, aes(x = temps)) +
  geom_line(aes(y = PIB_horscov, color = "PIB_horscov")) +
  geom_line(aes(y = DlogIDE_horscov, color = "DlogIDE_horscov")) +
  labs(title = "Evolution de PIB et DIFFlogIDE en fonction du temps, sans la période Covid", x = "Temps") +
  scale_color_manual(values = c(PIB_horscov = "blue", DlogIDE_horscov = "red")) +
  theme_minimal()
```

Evolution de PIB et DIFFlogIDE en fonction du temps, sans la période Covid



Choix des retards

On peut maintenant estimer le modèle VAR correspondant avec le nombre de retards optimal.

On commence par déterminer le nombre de retards par l'application de critères d'information multivariés. On doit spécifier, le lag.max, c'est à dire le nombre maximal de retards; et le type, c'est à dire la forme de la composante déterministe. On considère généralement au maximum 4 retards sur des données trimestrielles. On considère ici une constante comme composante déterministe, cela semble être en accord avec les résultats trouvés sur l'estimation du modèle dans le cas univarié. En effet, on a rejeté l'hypothèse selon laquelle il existait une tendance déterministe significative dans notre modèle, donc on estime un modèle VAR avec une constante.

```
pselect <-VARselect(y=endogen_horscov, lag.max = 4, type="const")
```

```
#Affichage des valeurs des critères des sélection
pselect$criteria
```

##	1	2	3	4
## AIC(n)	-2.2166608	-2.279061	-2.2553145	-2.2166111
## HQ(n)	-2.1445640	-2.158900	-2.0870888	-2.0003210
## SC(n)	-2.0367027	-1.979131	-1.8354123	-1.6767369
## FPE(n)	0.1089804	0.102415	0.1049383	0.1091945

```
#Nombres de retards optimal selon les critères de sélection
pselect$selection
```

```
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      2      1      2
```

Le retard optimal est celui qui minimise les critères. Ces critères sont des mesures de la qualité du modèle, visant à trouver le bon équilibre entre la complexité du modèle et son ajustement aux données. Les retards pour lesquels ces critères sont les plus faibles sont généralement considérés comme les meilleurs parce qu'ils représentent des modèles qui sont à la fois relativement simples et qui ajustent bien les données observées.

On choisit donc $p=2$ sur la base des critères HQ, AIC et FPE.

Estimation du modèle VAR

Equation du PIB Analysons les résultats obtenus avec la commande `summary` effectuée sur le modèle VAR:

- le coefficient de `PIB_horscov.l1` est significatif et positif pour un risque de première espèce de 1%
- le coefficient de `DlogIDE_horscov.l1` n'est pas significatif et négatif
- le coefficient de `PIB_horscov.l2` est significatif et positif pour un risque de première espèce de 5%
- le coefficient de `DlogIDE_horscov.l2` n'est pas significatif et négatif
- la constante est significative pour un risque de première espèce de 5%
- le R^2 ajusté vaut 20%

Equation des IDE

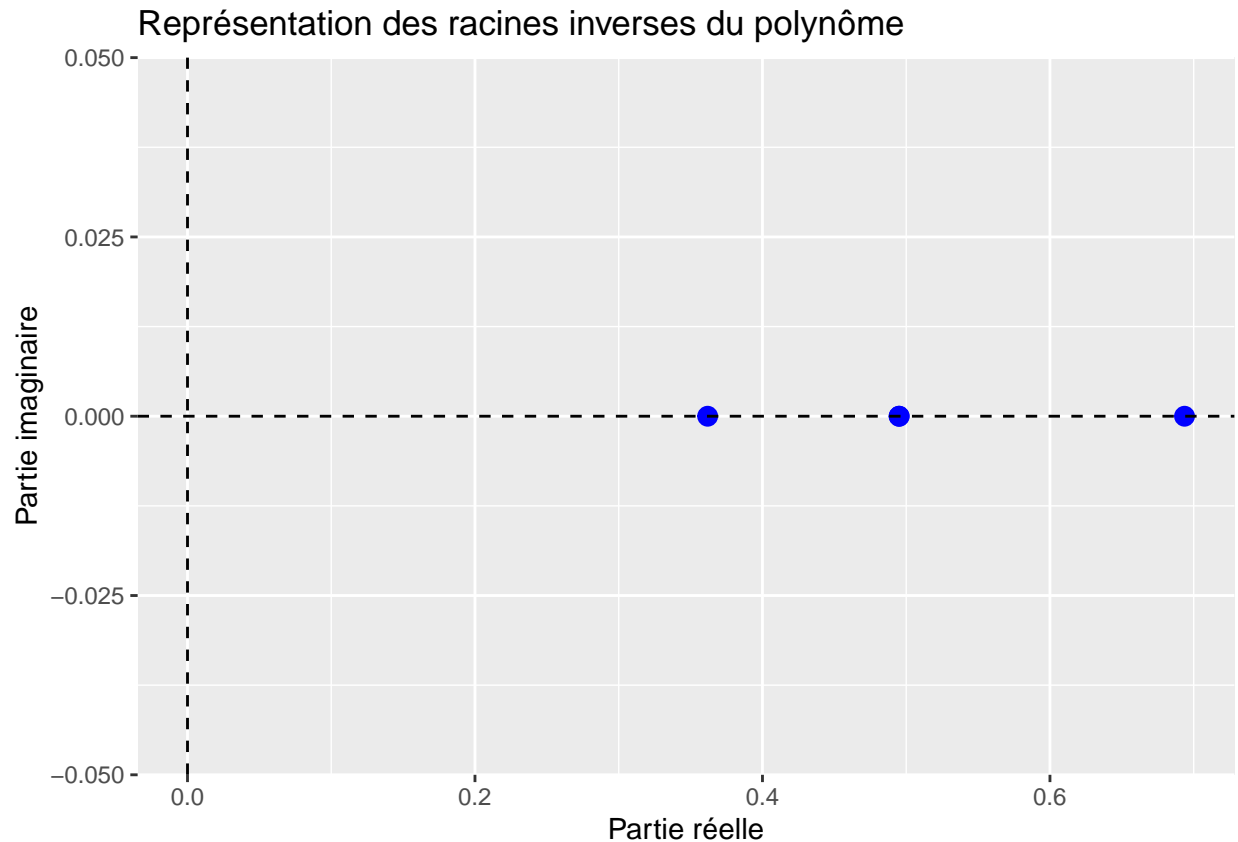
- le coefficient de `PIB_horscov.l1` n'est pas significatif et positif
- le coefficient de `logIDE_diff_horscov.l1` est significatif et négatif pour un risque de première espèce de 1%
- le coefficient de `PIB_horscov.l2` n'est pas significatif et positif
- le coefficient de `logIDE_diff_horscov.l2` est significatif et négatif pour un risque de première espèce de 5%
- la constante n'est pas significative
- le R^2 ajusté vaut 25%

Commentaire: Le coefficient de détermination R^2 ajusté pour le PIB est de 20%, et de 25% pour les log des IDE différenciés. Le pouvoir explicatif de la régression est donc meilleur pour la deuxième série. En ce qui concerne la significativité des coefficients, nous trouvons que les coefficients sont significatifs pour les retards du PIB dans l'équation du PIB, et significatifs pour les retards des IDE dans l'équation des IDE. On peut alors penser que les variables ne sont pas significativement liées.

On vérifie ensuite que le VAR est bien stationnaire, en se posant la question suivante : le module des inverses des racines du polynôme caractéristique appartient-il au disque unité complexe ? On le vérifie avec la commande `roots` de R.

```
racine_inverse <- roots(VAR2)
#On représente le résultat graphiquement
racines <- as.complex(racine_inverse)
racines_df <- data.frame(Re = Re(racines), Im = Im(racines))
library(ggplot2)
```

```
ggplot(racines_df, aes(x = Re, y = Im)) +
  geom_point(size = 3, color = "blue", shape = 19) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "black") +
  geom_vline(xintercept = 0, linetype = "dashed", color = "black") +
  labs(x = "Partie réelle", y = "Partie imaginaire", title = "Représentation des racines inverses du polynôme")
```



Les inverses des racines du polynôme caractéristique sont de module strictement inférieur à 1 : 0.6936 0.4951 0.3619. La condition de stationnarité est donc satisfaite.

On vérifie maintenant que les résidus ne sont pas autocorrélés. On choisit de tester l'autocorrélation sur 10 retards, donc on choisit `lags.pt=10`.

```
serial.test(VAR2,lags.pt=10,type = "PT.adjusted")
```

```
##
## Portmanteau Test (adjusted)
##
## data: Residuals of VAR object VAR2
## Chi-squared = 36.251, df = 32, p-value = 0.2769
```

La p-valeur vaut 0.2769, est donc supérieure aux niveaux de significativité 1,5 et 10%. On ne rejette donc pas l'hypothèse nulle d'absence d'autocorrélation des résidus, c'est-à-dire que les résidus du modèle ne sont pas significativement corrélés.

On détermine ensuite les coefficients associé à notre modèle.


```
A<-Acoef(VAR2)
A
```

```
## [[1]]
##                PIB_horscov.l1 DlogIDE_horscov.l1
## PIB_horscov          0.30363440        -0.03422718
## DlogIDE_horscov       0.02715135        -0.60942078
##
## [[2]]
##                PIB_horscov.l2 DlogIDE_horscov.l2
## PIB_horscov          0.2822241        -0.06810277
## DlogIDE_horscov       0.1282344        -0.24892016
```

Nous travaillons donc avec le modèle suivant:

$$\begin{cases} PIB_t = 0.30363 PIB_{t-1} + 0.28222 PIB_{t-2} - 0.03423 IDE_{t-1} - 0.06810 IDE_{t-2} + e_{1t} \\ IDE_t = -0.60942 IDE_{t-1} - 0.24892 IDE_{t-2} + 0.02715 PIB_{t-1} + 0.12823 PIB_{t-2} + e_{2t} \end{cases}$$

Nous pouvons écrire ce système sous forme matricielle :

$$\begin{bmatrix} PIB_t \\ IDE_t \end{bmatrix} = \begin{bmatrix} 0.30363 & 0.28222 \\ 0.02715 & 0.12823 \end{bmatrix} \begin{bmatrix} PIB_{t-1} \\ PIB_{t-2} \end{bmatrix} + \begin{bmatrix} -0.03423 & -0.06810 \\ 0.02715 & 0.12823 \end{bmatrix} \begin{bmatrix} IDE_{t-1} \\ IDE_{t-2} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

7. Relations de causalité entre les variables

L'intérêt dans cette partie est d'évaluer la causalité ou non d'une variable sur l'autre, en utilisant la causalité au sens de Granger, que l'on rappelle ici: soient 2 variables X_t et Y_t . On dit que X_t cause Y_t au sens de Granger si la connaissance des valeurs passées de X_t améliore la prévision de Y_t .

Test d'absence de causalité du PIB sur les IDE

Les hypothèses du test sont les suivantes:

H_0 : absence de causalité, soit les coefficients associés à PIB dans IDE sont nuls; contre H_1 : causalité

```
causality(VAR2,cause="PIB_horscov")
```

```
## $Granger
##
## Granger causality H0: PIB_horscov do not Granger-cause DlogIDE_horscov
##
## data: VAR object VAR2
## F-Test = 0.41639, df1 = 2, df2 = 152, p-value = 0.6602
##
##
## $Instant
##
## H0: No instantaneous causality between: PIB_horscov and DlogIDE_horscov
##
## data: VAR object VAR2
## Chi-squared = 2.7303, df = 1, p-value = 0.09846
```

On ne rejette donc pas l'hypothèse nulle d'absence de causalité. On peut le voir soit par lecture des valeurs critiques indiquées par la table de Fisher(2,152), car la statistique de test est toujours inférieure à ces valeurs, ou simplement en voyant que la p-valeur, qui vaut 0.6602 est bien supérieure aux seuils 1,5 et 10%. Ainsi, le taux de croissance du PIB d'un trimestre sur l'autre ne semble pas impacter les IDE.

Test d'absence de causalité des IDE sur le PIB

Les hypothèses de test sont les suivantes:

H_0 : absence de causalite, soit les coefficients associes a IDE dans PIB sont nuls; contre H_1 : causalite

```
causality(VAR2,cause="DlogIDE_horscov")
```

```
## $Granger
##
##   Granger causality H0: DlogIDE_horscov do not Granger-cause PIB_horscov
##
## data:   VAR object VAR2
## F-Test = 0.43833, df1 = 2, df2 = 152, p-value = 0.6459
##
##
## $Instant
##
##   H0: No instantaneous causality between: DlogIDE_horscov and PIB_horscov
##
## data:   VAR object VAR2
## Chi-squared = 2.7303, df = 1, p-value = 0.09846
```

On ne rejette donc pas l'hypothèse nulle d'absence de causalité. On peut le voir soit par lecture des valeurs critiques indiquées par la table de Fisher(2,152), car la statistique de test est toujours inférieure à ces valeurs, ou simplement en voyant que la p-valeur, qui vaut 0.6459 est bien supérieure aux seuils 1,5 et 10%. Ainsi, le log des IDE (différencié une fois) en part du PIB ne semble pas impacter le taux de variation du PIB.

8. Analyse impulsion réponse des chocs

Nous passons maintenant à l'analyse des fonctions de réponse. La fonction de réponse d'un VAR retrace l'effet d'un choc en t sur l'une des variables, sur les variables présentes et futures de cette variable et sur les autres variables du VAR.

On compare à cette fin le profil prévu de la série à un horizon h suite à un choc survenu en t, à celui en l'absence de choc.

D'après la propriété de la fonction de réponse, si le VAR est stationnaire, les fonctions de réponse au choc doivent tendre vers 0 à mesure que l'on s'éloigne de la date du choc. Cela signifie que leur effet est seulement transitoire. Etant donné que l'on a montré que notre VAR est stationnaire, nous devons donc trouver que ces fonctions tendent vers 0 au cours du temps.

Pour calculer ces fonction de réponse, nous allons présenter deux méthodes : le calcul des prévisions de façon recursive à partir du VAR estimé, puis la méthode des projections locales de Jorda.

La méthode des VAR

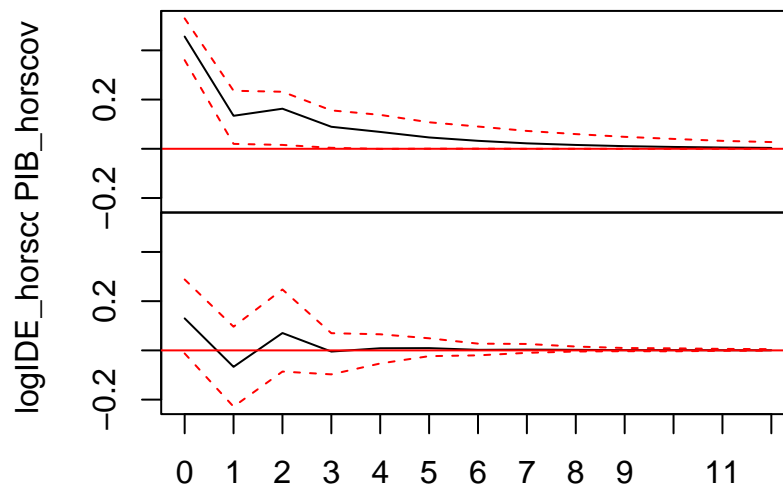
Le calcul des réponses est ici fondé sur des prévisions récursives avec le modèle VAR estimé. Les chocs structurels sont obtenus avec une décomposition de Cholesky, pour créer des chocs orthogonaux (non corrélés). Pour cela, nous devons choisir l'ordre de causalité des variables. Economiquement, et avec ce que l'on a observé à la question précédente, la potentielle causalité du log des IDE sur le taux de croissance du PIB est plus forte que l'inverse. Trouver arguments économiques. On place donc le log des IDE différencié de la France en premier.

Nous avons ajouté un intervalle de confiance de 95% pour déterminer la significativité des réponses.

Finalement, nous choisissons d'observer les impulsions réponses sur 12 périodes futures suite au choc.

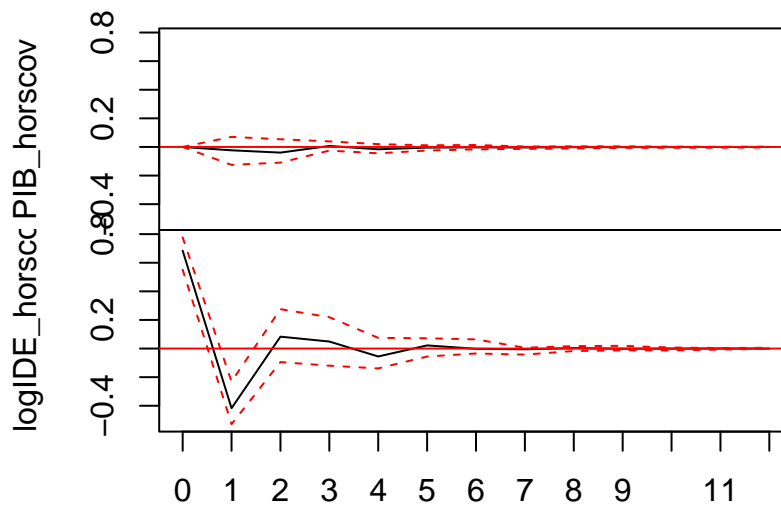
```
endogen_horscov <-ts.intersect(DlogIDE_horscov,PIB_horscov)
irf<-irf(VAR2,n.ahead=12, ortho = TRUE,ci=0.95)
plot(irf)
```

Orthogonal Impulse Response from PIB_horscov



95 % Bootstrap CI, 100 runs

Orthogonal Impulse Response from DlogIDE_horscov



95 % Bootstrap CI, 100 runs

La première figure retrace les réponses du taux de croissance du PIB de la France (premier graphe) et du log des IDE différencié de la France (second graphe) à un choc sur le taux de croissance du PIB de la France.

La seconde figure retrace les réponses du taux de croissance du PIB de la France (premier graphe) et du log des IDE différencié de la France (second graphe) à un choc sur le log des IDE différencié de la France.

Remarque : Les quatre fonctions convergent vers zéro, ce qui est conforme avec la stationnarité du VAR. En effet, la stationnarité s'interprète comme une non-persistance des chocs, la série finit par retourner vers sa moyenne.

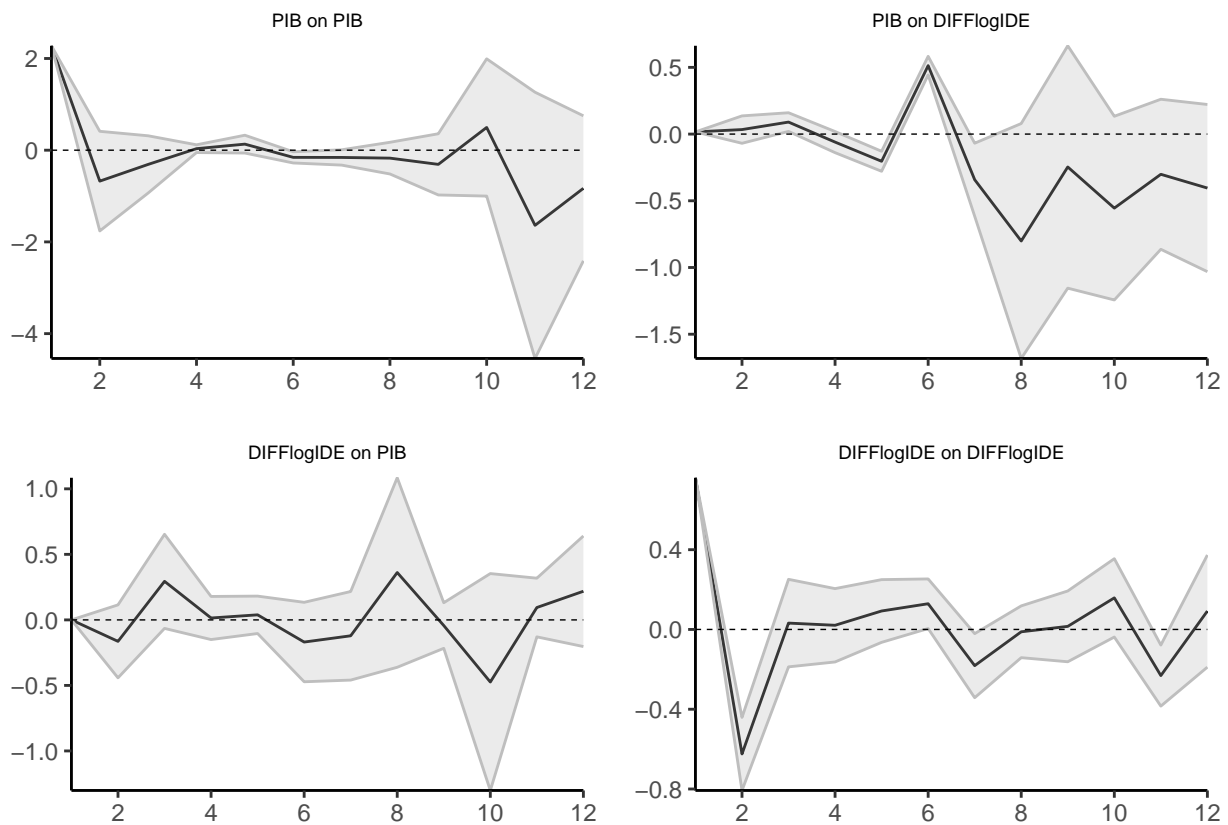
Effet d'un choc sur le taux de croissance du PIB de la France: L'effet est positif et immédiat sur le log des IDE différencié, puis on observe un effet négatif, puis de nouveau positif jusqu'à tendre vers 0. Cela ne semble pas forcément en accord avec les résultats trouvés lors de l'application des tests de Granger, mais cela ne semble pas absurde d'un point de vue macroéconomique. En effet, les investissements globaux peuvent s'exprimer directement en fonction du PIB avec peu de facteurs externes supplémentaires, donc les IDE, qui représentent une part des investissements, peuvent être sensibles à un choc sur le PIB.

Effet d'un choc sur le log des IDE différencié : On remarque qu'un tel choc a très peu d'effet sur le taux de croissance du PIB de la France, il subit une très très faible baisse suite au choc mais l'effet n'est pas immédiat ni significatif. A nouveau, d'un point de vue macroéconomique, ce résultat ne semble pas absurde dans la mesure où le PIB est une variable qui dépend d'énormément de facteurs, les IDE n'étant pas forcément un facteur majeur comparé aux autres.

La méthode des projections locales

On présente ensuite le calcul des réponses, fondé sur la méthode des projections locales de Jorda. Les chocs structurels sont obtenus avec une décomposition de Cholesky.

```
install.packages("collapse")
install.packages("lpirfs")
library(collapse)
library(lpirfs)
irf_LP<-lp_lin(endog_data= as.data.frame(endogen),
               lags_endog_lin = NaN,
               lags_criterion="BIC",
               max_lags = 4,
               trend = 0,
               shock_type = 0,
               confint = 1.96,
               hor = 12,
               adjust_se=TRUE)
plot(irf_LP)
```



Les résultats obtenus par la méthode des projections locales sont en accord avec ceux obtenus avec la méthode des VAR. Les graphiques associés à cette méthode sont selon nous encore plus lisibles que ceux obtenus avec la méthode des VAR.

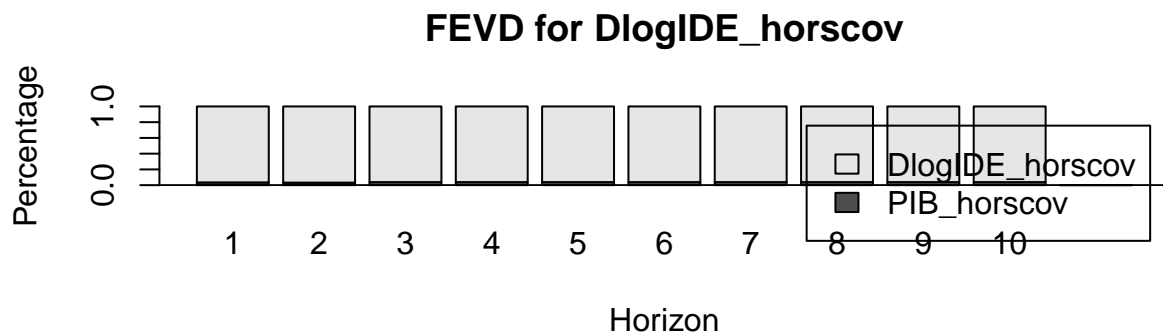
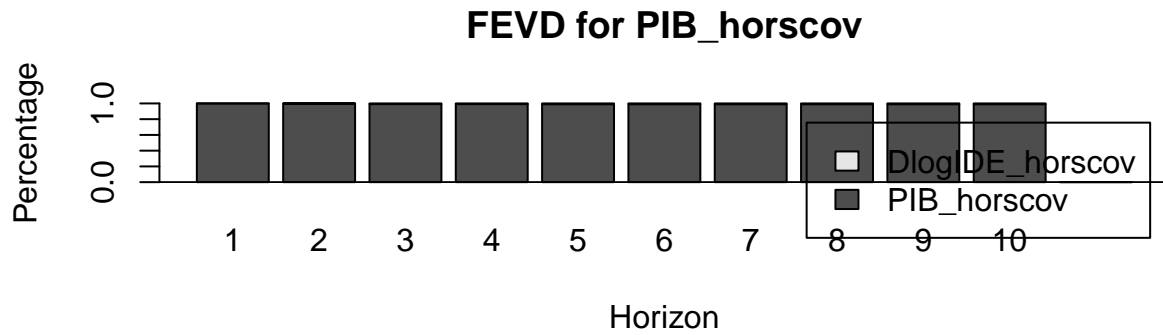
Analyse de la décomposition de la variance d'erreur

Cette analyse va permettre de confirmer les choix effectués dans l'ordre des variables dans le calcul des fonctions de réponse.

```
dec_var <- fevd(VAR2,n.ahead=10)
print(dec_var)
```

```
## $PIB_horscov
##      PIB_horscov DlogIDE_horscov
## [1,]  1.0000000    0.000000000
## [2,]  0.9975806    0.002419429
## [3,]  0.9917432    0.008256833
## [4,]  0.9918113    0.008188672
## [5,]  0.9909639    0.009036091
## [6,]  0.9909625    0.009037480
## [7,]  0.9909659    0.009034105
## [8,]  0.9909353    0.009064703
## [9,]  0.9909338    0.009066167
## [10,] 0.9909318    0.009068201
##
## $DlogIDE_horscov
##      PIB_horscov DlogIDE_horscov
## [1,]  0.03488307    0.9651169
## [2,]  0.03224912    0.9677509
## [3,]  0.03904723    0.9609528
## [4,]  0.03894021    0.9610598
## [5,]  0.03886962    0.9611304
## [6,]  0.03896234    0.9610377
## [7,]  0.03896930    0.9610307
## [8,]  0.03898154    0.9610185
## [9,]  0.03898847    0.9610115
## [10,] 0.03899006    0.9610099
```

```
plot(dec_var)
```



Les résultats obtenus par décomposition de la variance sont cohérents avec ceux trouvés lors des analyses de causalité entre les deux variables. On remarque tout de même bien que notre choix d'ordre effectué lors du calcul des fonctions de réponse est cohérent, dans le sens où les chocs sur le PIB n'affectent presque pas du tout la variance d'erreur des IDE, alors qu'un choc sur les IDE affecte (en comparaison) plus la variance d'erreur du PIB.

9. Test de cointégration de Johansen

Le test de cointégration de Johansen est généralement applicable lorsque toutes les séries temporelles impliquées dans l'analyse sont intégrées de même ordre, c'est-à-dire qu'elles sont toutes $I(d)$ avec le même ordre d'intégration d . Ce n'est pas le cas ici, puisque la série du PIB est $I(0)$ alors que la série du log des IDE est $I(1)$.

Conclusion générale

Lors de ce projet, nous avons pu effectuer sur nos séries choisies les méthodes vues dans ce cours d'Econométrie II. Dans une première partie, la dynamique de nos séries a pu être analysée, et avec un modèle linéaire nous avons réussi à effectuer des prévisions qui nous semblent fiables sur notre série du taux de variation semestriel du PIB. En seconde partie, la mise en relation des deux séries à l'aide d'une modélisation multivariée a permis d'affirmer ou d'infirmer la présence de corrélations entre ces deux séries. Nous en avons conclu que nos deux séries étaient faiblement corrélées, au regard des tests de causalité de Granger et des fonctions d'impulsion-réponse.

En effet, la relation entre les investissements directs étrangers en France et le taux de croissance du PIB peut être moins directe que ce que l'on pourrait penser, en raison de divers facteurs et mécanismes économiques.

Premièrement, il existe un décalage temporel entre ces deux variables. Les effets des IDE sur le PIB peuvent ne pas être immédiats. Les investissements nécessitent souvent du temps pour être mis en œuvre et commencer à contribuer à la croissance économique. Ainsi, la corrélation peut ne pas être instantanée.

De plus, il est important de regarder la nature des investissements. Les IDE peuvent être orientés vers des secteurs spécifiques qui ont des cycles économiques différents. Par exemple, les investissements dans des industries de haute technologie peuvent avoir des effets sur la croissance à long terme, tandis que d'autres investissements peuvent avoir des impacts plus immédiats.

Egalement, dans ce projet l'analyse est basée sur deux variables, mais il est évident qu'il existe une multitude de variables supplémentaires pouvant être corrélatées plus significativement à nos variables étudiées. On peut les voir comme des facteurs externes. En effet, la corrélation entre les IDE et le PIB peut être influencée par des facteurs externes tels que les conditions économiques mondiales, les politiques commerciales, les taux de change, etc. Les variations dans ces facteurs peuvent contribuer à des mouvements indépendants du PIB et des IDE. Parmi ces facteurs, on peut citer notamment les politiques gouvernementales françaises effectuées sur la période étudiée, qui ont concerné notamment les incitations fiscales, les réformes structurelles, et la réglementation des investissements.

En résumé, bien que les IDE et le taux de croissance du PIB soient liés dans leur contribution globale à l'économie, la nature complexe de ces interactions, combinée à des facteurs externes et des délais dans la réalisation des effets, peut conduire à une corrélation moins directe que ce que l'on pourrait anticiper intuitivement.