Revue Française d'Economie et de Gestion
ISSN : 2728- 0128
Volume 3 : Numéro 7

Modélisation du PIB de la France de 1949 à 2014 par la méthode de Box-Jenkins

Modeling of France's GDP from 1949 to 2014 using the Box-Jenkins method

MOUHOUMED ELMI Yacin
Doctorant en Économétrie-Statistique
Laboratoire de Recherche en Économie de Saint-Louis (LARES)
Université Gaston Berger
Gestionnaire et Suivi-évaluateur des projets et Spécialiste en Marketing digital
Ministère de l'Économie et des Finances, chargé de l'Industrie
République de Djibouti
yacinmouhoumed@hotmail.fr

Date de soumission : 11/04/2022

Date d'acceptation : 25/06/2022

Pour citer cet article :

MOUHOUMED ELMI.Y (2022) « Modélisation du PIB de la France de 1949 à 2014 par la méthode de Box Jenkins », Revue Française d'Economie et de Gestion « Volume 3 : Numéro 7 » pp : 49 -70 .

Résumé :

L'article essaie avec une approche claire et intelligible de modéliser par la méthode de Box-Jenkins le PIB de la France en fonction du temps pour bien comprendre son évolution et de faire une prévision dans le futur.

D'après l'analyse déductive, la valeur du PIB français varie entre 3,199 milliards d'euros et 539,400 milliards d'euros soit une étendue de 536,201 milliards d'euros. Il faut noter aussi que le PIB varie progressivement dans le temps à l'exception de la période de 2008 à 2010 qui correspond à la crise bancaire et financière de 2008.

En conclusion, c'est le modèle d'ARIMA qui permet de mieux modéliser le PIB de la France par la méthode de Box-Jenkins avec une bonne prévision.

Mots clés :

« Modélisation économétrique ; Statistique et analyse d'une série temporelle ; Produit Intérieur Brut ; Méthode Box-Jenkins »

Abstract:

The article tries with a clear and intelligible approach to model by Box-Jenkins method the GDP of France according to time to understand its evolution and to make a forecast in the future.

According to the deductive analysis, the value of the French GDP varies between 3.199 billion euros and 539.400 billion euros with a range of 536.201 billion euros. It should also be noted that GDP varies gradually over time with the exception of the period from 2008 to 2010 which corresponds to the banking and financial crisis of 2008.

In conclusion, it is ARIMA model that makes it possible to better model the GDP of France by the Box-Jenkins method with a good forecast.

Keywords:

"Econometric modeling; Statistics and analysis of a time series; Gross Domestic Product; Box-Jenkins method"

Introduction

L'indicateur économique dominant est actuellement, et depuis longtemps, le Produit Intérieur Brut (PIB). Il mesure le niveau de production d'un pays. C'est sa variation qui permet de mesurer le taux de croissance économique. Le PIB est apparu en France en même temps que la Comptabilité Nationale (CN) en 1945.

En plus de mesurer la croissance économique et d'être l'indicateur de base des comparaisons internationales, il devait aussi permettre de mesurer, avec le PIB par habitant, l'évolution du bien-être économique des individus.

Selon les chiffres publiés, vendredi 26 décembre 2014, par le Centre of Economics and Business Research (CEBR), un institut statistique indépendant basé à Londres, la France n'est plus la cinquième économie du monde ni la deuxième économie européenne, positions pour lesquelles elle est désormais devancée par le Royaume-Uni. Les chiffres du PIB 2014 sont de 2.828 milliards de dollars pour le Royaume-Uni et de 2.827 pour la France.Il se situe bien en-dessous de la marge d'erreur, à 1 petit milliard alors que l'an dernier, l'écart entre les deux économies se situait à 200 milliards de dollars environ en faveur de la France,

Pour cela, il est intéressant de modéliser le PIB de la France en fonction du temps pour bien comprendre son évolution et de faire une prévision dans le futur.

Dans le présent travail, nous allons modéliser le Produit Intérieur Brut français à partir de ses mesures trimestrielles de 1949 jusqu'au 2014 par l'INSEE en utilisant le logiciel R comme application statistique.

Après une définition détaillée du Produit Intérieur Brut, on va présenter l'origine des données avant de modéliser le PIB français par la méthode de Box-Jenkins. Ensuite on fera le test de ARCH pour voir s'il y'a l'hétéroscédasticité des erreurs pour pouvoir appliquer la modélisation GARCH avant de conclure le travail.

1. Méthodologie et analyse descriptive

1.1. Description des données

La série que nous allons modéliser par la méthode de Box-Jenkins est le PIB de la France de 1949 à 2014. Notre série PIB est calculée tous les trimestres chaque année en France avec une périodicité trimestrielle (série trimestrielle) et nous avons téléchargé notre série d'étude sur le site web de l'Insee de France.

Nous allons laisser les 10 dernières valeurs de notre série pour évaluer la qualité de la prévision plus tard.

1.2. Analyse descriptive de la série

Il s'agit d'un tableau qui regroupe les statistiques sommaires de la série PIB du premier trimestre 1949 jusqu'au dernier trimestre de 2014.

En moyenne, la France a enregistré un PIB qui est égale à 192,300 milliards d'euros du premier trimestre 1949 jusqu'au dernier trimestre de 2014. Depuis le premier trimestre de 1949, la valeur du PIB français varie entre 3,199 milliards d'euros et 539,400 milliards d'euros soit une étendue de 536,201 milliards d'euros.

Tableau N°1 : Statistiques sommaires du PIB (1949-2014) en milliards d'euros

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.199	19.270	138.400	192.300	339.900	539.400

Source : INSEE, Auteur

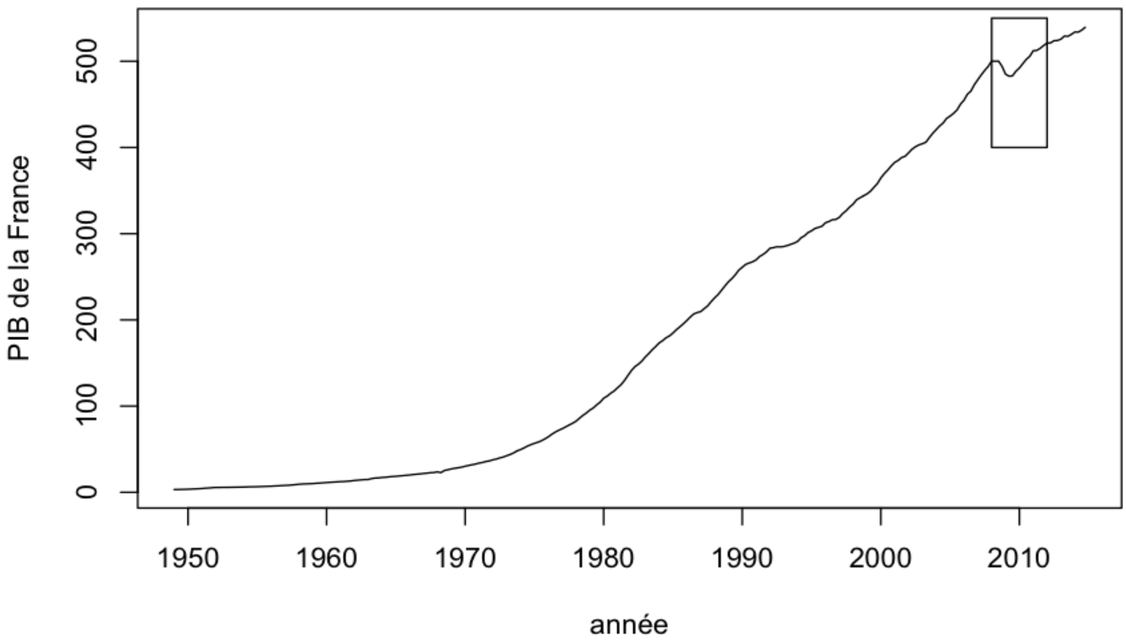
1.3. Tendance de la série

Il s'agit d'un chronogramme qui représente le PIB de France en fonction du temps (figure N°1). Ce graphique permet de déterminer l'évolution du PIB de 1949 à 2014 en France. Nous allons s'intéresser à la tendance de cette série pour décrire l'évolution du PIB. La tendance du PIB connaît une évolution lente et continue de 1949 à 1970.

Cette évolution lente peut être expliquée par la situation d'après-guerre (la période qui suit immédiatement la Seconde Guerre Mondiale et qui dure jusqu'aux débuts de la guerre froide et la mort de Staline) qui est fréquemment marquée par le manque de nourriture, de produits de tous types et des conditions de vies mauvaises dans le monde. Mais à partir de 1970 le PIB accroit fortement avant de diminuer en 2008 puis se redresse à partir de 2010. Cette diminution est du probablement à la crise bancaire et financière de l'automne 2008.

Pour la France, cette crise de 2008 était un tremblement de terre qui vient bouleverser cette tendance. Alors qu'elle a mieux résisté par rapport à ces voisins à l'onde de choc liée à la faillite de Lehman Brothes grâce à son Etat-providence.

Figure N°1 : Chronogramme du PIB



Source : INSEE, Auteur

2. Tests de stationnarité et de saisonnalité

2.1. Test de stationnarité

Il convient d'effectuer un test permettant de détecter l'existence d'une saisonnalité avant d'effectuer sur série temporelle une correction de variations saisonnières. Le test le plus communément utilisé est celui de Fisher par analyse de la variance du facteur mensuel (ou trimestriel) par rapport à la variance totale. Ce test se présente comme suit :

- Calcul de la somme U* des carrés du modèle avec tendance simple
- Calcul de la somme U** des carrés du modèle avec tendance et saisonnalité.

Calcul de la valeur du Fisher empirique :

$$\boldsymbol{F^{*}} = \frac{\boldsymbol{U^{*} - U^{**}}}{\boldsymbol{K - P}} \Big/ \frac{\boldsymbol{U^{**}}}{\boldsymbol{T - K}}$$

La valeur du F* empirique est à comparer à la valeur du F théorique donné par la table de la loi de Fisher-Snedecor aux degrés de liberté K-P et T-K, avec : K, le nombre de paramètres indépendants à estimer dans le cadre du modèle avec tendance et saisonnalité, soit K = 2+12-1 (car les 12 coefficients saisonniers sont liés entre eux mais dans notre cas on a que 4 coefficients saisonniers). P, le nombre de paramètres à estimer dans le cadre du modèle à tendance seule, P=2. T est le nombre d'observations.

Les hypothèses du test de Fisher :

Hypothèse nulle : F* > F alors la série est saisonnière.

Hypothèse alternative : F* ≤ F alors la série n'est pas saisonnière.

Pour une chronique relativement longue (4 ans en données mensuelles), F peut être approximé par 2 ; cela évite une lecture systématique de la table pour un risque d'erreur faible.

Nous ferons les calculs de ce test avec le logiciel Excel qu'on va mettre en annexe le tableau de calcul des différentes composantes de la série (tendance, saisonnalité et les différents sommes U* et U** des carrés) et on va déterminer la tendance de la série par la méthode de moindre carrée ordinale (MCO). Comme on a une série longue de 264 observations alors on prendra F=2 pour un risque d'erreur de 5%.

Après le calcul effectué sous Excel, On a : F*= 0,107902191 et F_table= 2,63945467

Comme la valeur de la Fisher empirique (F*) est inférieur à celle de la Fisher théorique (F_table) alors on conclut que la série du PIB n'est pas saisonnière.

2.2. Test de saisonnalité

Une des grandes questions dans l'étude de séries temporelles (ou chronologiques) est de savoir si celles-ci suivent un processus stationnaire. On entend par là le fait que la structure du processus sous-jacent supposé évolue ou non avec le temps. Si la structure reste la même, le processus est dit alors stationnaire.

- Stationnarité forte : Soit un processus temporel à valeurs réelles et en temps discret Z_1, Z_2, …, Z_t .Il est dit stationnaire au sens fort si pour toute fonction f mesurable ont même loi.

$$f(Z_1, Z_2, ..., Z_t) \ et \ f(Z_{1+k}, Z_{2+k}, ..., Z_{t+k})$$

- Stationnarité faible : Soit un processus temporel à valeurs réelles et en temps discret Z_1, Z_2, …, Z_t. Il est dit stationnaire au sens faible (ou « de second ordre », ou « en covariance ») si ses propriétés statistiques ne varient pas dans le temps (espérance, variance, auto-corrélation).

  - $E[Z_i] = \mu \qquad \forall i = 1...t$
  - $Var[Z_i] = \sigma^2 \neq \infty \qquad \forall i = 1...t$
  - $Cov[Z_i, Z_{i-k}] = f(k) = \rho_k \qquad \forall i = 1...t, \quad \forall k = 1...t$

Après avoir tester la saisonnalité de notre série d'étude, on va tester maintenant si notre série est stationnaire ou pas et on va la rendre stationnaire si elle est non stationnaire. Pour cela on va effectuer le test de Dickey-Fuller augmenté.

Les hypothèses du test de Dickey-Fuller sont :

Hypothèse nulle : La série comporte une racine unitaire (la série n'est pas stationnaire),

Hypothèse alternative : La série ne comporte pas de racine unitaire (la série est stationnaire).

Décision : si la statistique du test (la Dickey-Fuller augmenté) < la valeur critique alors on accepte l'hypothèse alternative : la série ne comporte pas de racine unitaire (la série est stationnaire).

Figure N°2 : Test de Dickey-Fuller augmenté

```
###############################################
# 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
-7.1328 -0.1202  0.0832  0.6249  4.1457

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1    0.0020879  0.0005128   4.072 6.27e-05 ***
z.diff.lag 0.7509133  0.0436245  17.213  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.311 on 250 degrees of freedom
Multiple R-squared:  0.7976,  Adjusted R-squared:  0.796
F-statistic: 492.6 on 2 and 250 DF,  p-value: < 2.2e-16


Value of test-statistic is: 4.0715

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Source : INSEE, Auteur

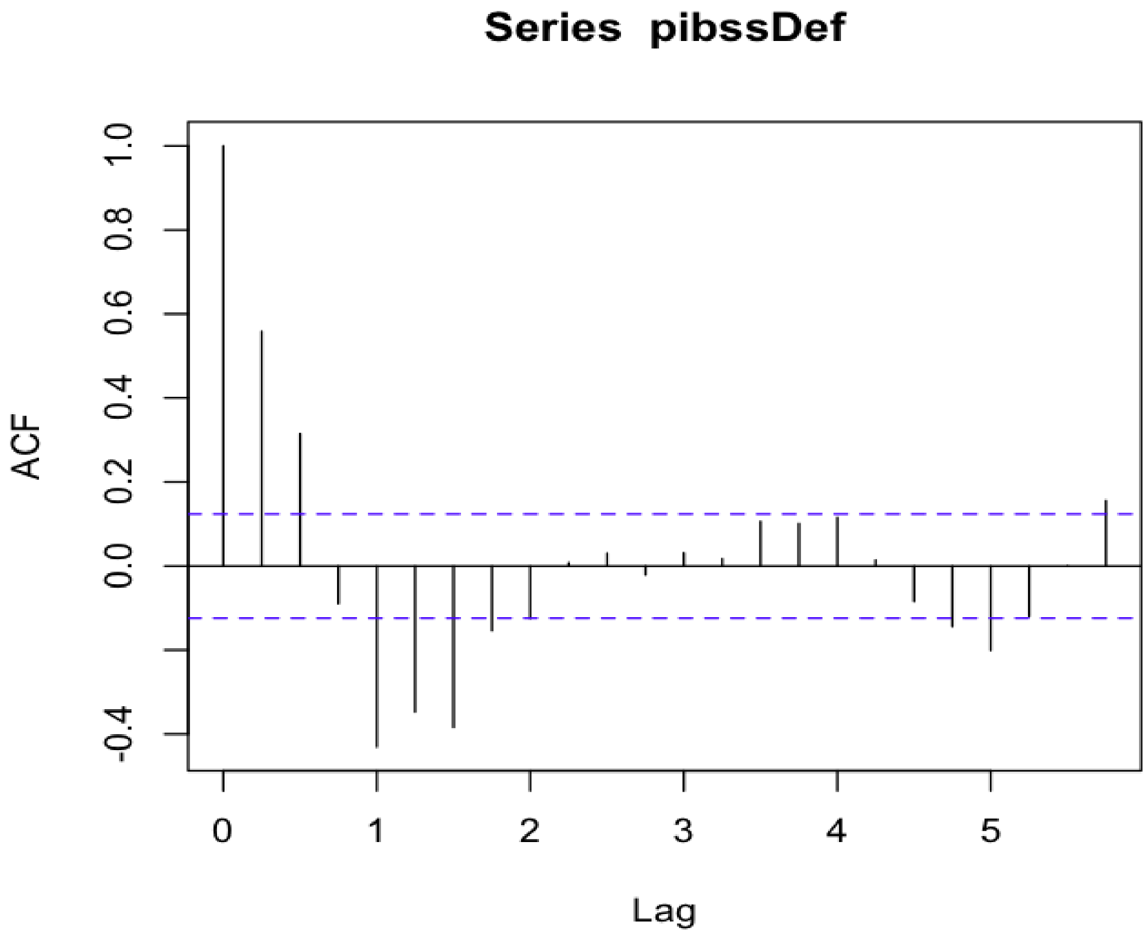
Interprétation du résultat du test et Décision : La ligne Value of test-statisticis: contient la valeur de la statistique du test. La statistique du test de Dickey-Fuller augmenté est égale à 4,0715 et la valeur critique du test pour un risque de 5% est -1,95. Au risque de 5% comme la statistique du test de Dickey-Fuller augmenté est supérieure à la valeur critique du test alors on accepte l'hypothèse nulle. Donc la série comporte une racine unitaire (la série n'est pas stationnaire). Mais avant de modéliser cette série on va la rendre stationnaire par la méthode de la différence première.

3. Modélisation Box et Jenkins

3.1. Identification de l'ARMA

- Détermination de l'ordre de moyenne mobile q par la fonction d'autocorrélation :

Figure N°3 : Le corrélogramme de la fonction d'autocorrélation du PIB



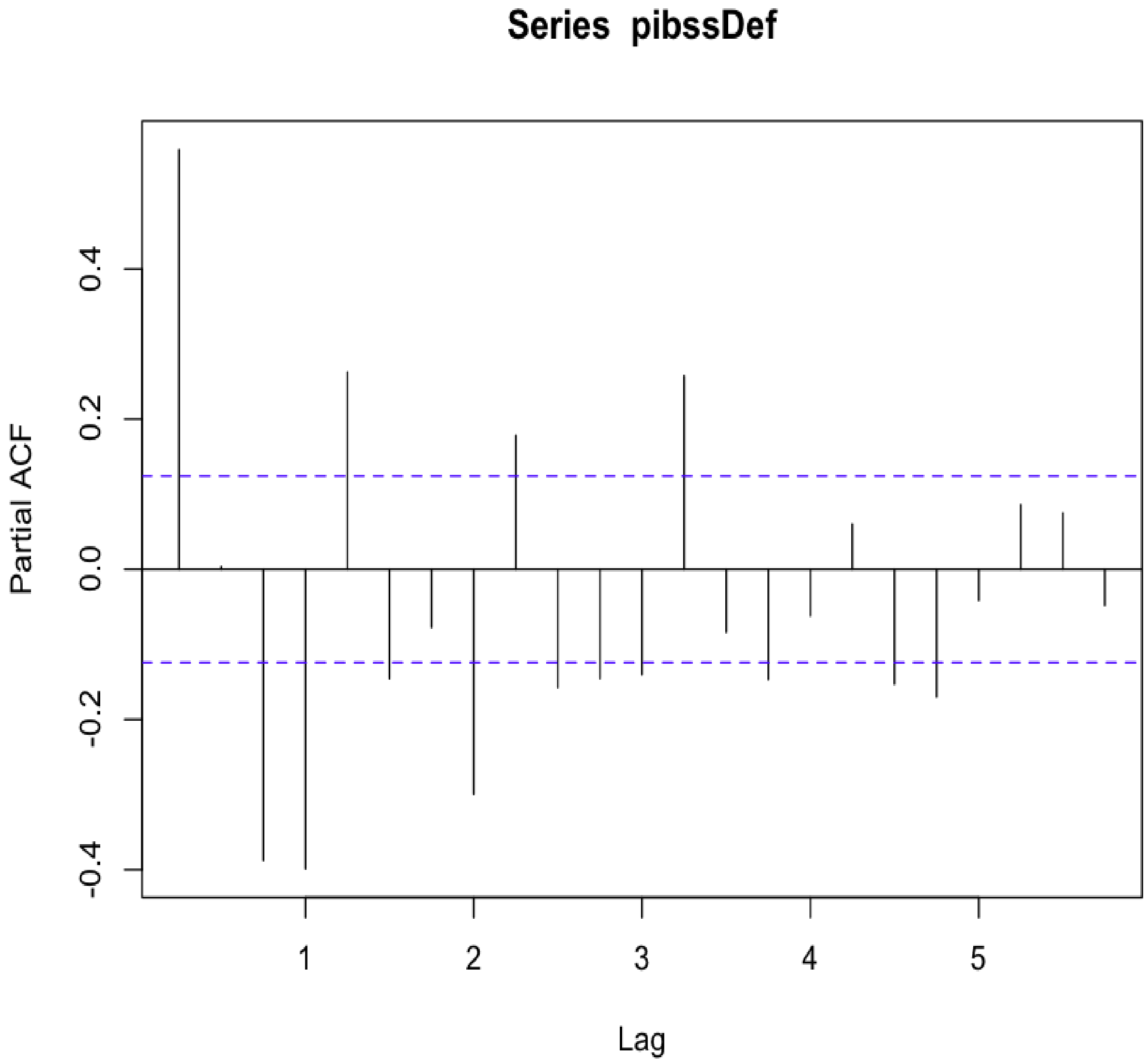
Source : INSEE, Auteur

Il s'agit du corrélogramme de la fonction d'autocorrélation du PIB après avoir enlevé la saisonnalité avant de la rendre stationnaire. La fonction d'autocorrélation est la fonction notée ρk qui mesure la corrélation de la série avec elle-même décalée de k périodes. Cette fonction permet de déterminer l'ordre de moyenne mobile q en regardant les barres qui sortent de la bande en bleu mais dans le cas pratique on ne prend pas les ordres éloignés.

Dons notre cas, on prend les ordres 0, 1, et 2 comme ordre de moyenne mobile q (q= {0, 1, 2}).

- Détermination de l'ordre d'autorégressif p par la fonction d'autocorrélation partielle :

Figure N°4 : Le corrélogramme partiel de la fonction d'autocorrélation partielle du PIB



Source : INSEE, Auteur

Ce graphique représente le corrélogramme de la fonction d'autocorrélation partielle de la série. La fonction d'autocorrélation partielle mesure la corrélation entre Yt et Yt-k, l'influence des autres variables (Yt-1, Yt-2, …, Yt-k+1) ayant été retirée. Cette fonction permet de déterminer l'ordre d'autorégressif p en regardant les barres qui sortent de la bande (l'intervalle) en bleu mais dans le cas pratique on ne prend pas les ordres éloignés. D'après ce corrélogramme partiel les ordres d'autorégressif possible de p sont 0, 2 et 3 (p= {0, 2, 3}).

- Les modèles d'ARIMA possibles sont : (0, 1) ; (0, 2) ; (2, 0) ; (2, 1) ; (2, 2) ; (3, 0) ; (3, 1) ; (3, 2).

3.2. Estimations et test de significativité des paramètres

- **Pour le modèle (0, 1) :**

Figure N°5 : Estimation et test de significativité des coefficients du modèle01

```
> model01<-arima(pib1,order=c(0,1,1))
> model01

Call:
arima(x = pib1, order = c(0, 1, 1))

Coefficients:
         ma1
      0.6718
s.e.  0.0403

sigma^2 estimated as 4.249:  log likelihood = -542.29,  aic = 1088.59
> library(caschrono)
> t_stat(model01)
            ma1
t.stat 16.68718
p.val   0.00000
```

Source : INSEE, Auteur

Les hypothèses :

Hypothèse nulle : q=0 (le modèle suit un ARMA (p, q-1))

Hypothèse alternative : q≠0 (le modèle suit un ARMA (p, q))

Décision : Comme la p_value est inférieur au risque de 5% donc on rejette de l'hypothèse nulle (on accepte l'hypothèse alternative). Alors le modèle (0, 1) suit un ARMA (0, 1). D'après ce test de significativité, **on conclut que le modèle (0, 1) est retenu.**

- **Pour le modèle (0, 2) :**

Figure N°6 : Estimation et test de significativité des coefficients du modèle02

```
> model02<-arima(pib1,order=c(0,1,2))
> model02

Call:
arima(x = pib1, order = c(0, 1, 2))

Coefficients:
         ma1     ma2
      0.7584  0.6332
s.e.  0.0618  0.0385

sigma^2 estimated as 2.651:  log likelihood = -482.94,  aic = 971.87
> library(caschrono)
> t_stat(model02)
            ma1      ma2
t.stat 12.27661 16.46661
p.val   0.00000  0.00000
```

Source : INSEE, Auteur

Les hypothèses :

Hypothèse nulle : q=0 (le modèle suit un ARMA (p, q-1))

Hypothèse alternative : q≠0 (le modèle suit un ARMA (p, q))

Décision : Comme la p_value (0,000) de ma2 est inférieur au risque de 5% donc on rejette de l'hypothèse nulle (on accepte l'hypothèse alternative). Alors le modèle (0, 2) suit un ARMA (0, 2). D'après ce test de significativité, **on conclut que le modèle (0, 2) est lui aussi retenu.**

- **Pour le modèle (2, 0) :**

Figure N°7 : Estimation et test de significativité des coefficients du modèle20

```
> model20<-arima(pib1,order=c(2,1,0))
> model20

Call:
arima(x = pib1, order = c(2, 1, 0))

Coefficients:
         ar1     ar2
      0.6578  0.2526
s.e.  0.0609  0.0610

sigma^2 estimated as 1.695:  log likelihood = -426.52,  aic = 859.05
> library(caschrono)
> t_stat(model20)
            ar1      ar2
t.stat 10.79577 4.138821
p.val   0.00000 0.000035
```

Source : INSEE, Auteur

Les hypothèses :

Hypothèse nulle : p=0 (le modèle suit un ARMA (p-1, q))

Hypothèse alternative : p≠0 (le modèle suit un ARMA (p, q))

Décision : Comme la p_value (0,000035) de ar2 est inférieur au risque de 5% donc on rejette de l'hypothèse nulle (on accepte l'hypothèse alternative). Alors le modèle (2, 0) suit un ARMA (2, 0). D'après ce test de significativité, **on conclut que le modèle (2, 0) est aussi retenu.**

- **Pour le modèle (2, 1) :**

Figure N°8 : Estimation et test de significativité des coefficients du modèle21

```
> model21<-arima(pib1,order=c(2,1,1))
> model21

Call:
arima(x = pib1, order = c(2, 1, 1))

Coefficients:
         ar1     ar2     ma1
      0.3644  0.5137  0.3071
s.e.  0.1260  0.1119  0.1341

sigma^2 estimated as 1.672:  log likelihood = -424.83,  aic = 857.66
> library(caschrono)
> t_stat(model21)
            ar1      ar2      ma1
t.stat 2.892296 4.590042 2.289781
p.val  0.003824 0.000004 0.022034
```

Source : INSEE, Auteur

Les hypothèses :

H nulle : p=0 (le modèle suit un ARMA (p-1, q)) / q=0 (le modèle suit un ARMA (p, q-1)

H alternative : p≠0 (le modèle suit un ARMA (p, q)) / q≠0 (le modèle suit un ARMA (p, q)

Décision : Comme la p_value (0,000004) de ar2 et celle (0,0220) de ma1 sont toutes les deux inferieurs au risque de 5% donc on rejette de l'hypothèse nulle (on accepte l'hypothèse alternative). Alors le modèle (2, 1) suit un ARMA (2, 1). D'après ce test de significativité, **on conclut que le modèle (2, 1) est aussi retenu.**

- **Pour le modèle (2, 2) :**

Figure N°9 : Estimation et test de significativité des coefficients du modèle22

```
> model22<-arima(pib1,order=c(2,1,2))
> model22

Call:
arima(x = pib1, order = c(2, 1, 2))

Coefficients:
         ar1     ar2     ma1     ma2
      0.4707  0.3769  0.2508  0.1840
s.e.  0.1450  0.1376  0.1433  0.0917

sigma^2 estimated as 1.649:  log likelihood = -423.1,  aic = 856.21
> library(caschrono)
> t_stat(model22)
            ar1      ar2      ma1      ma2
t.stat 3.246238 2.739950 1.750454 2.007001
p.val  0.001169 0.006145 0.080040 0.044750
```

Source : INSEE, Auteur

Les hypothèses :

H nulle : p=0 (le modèle suit un ARMA (p-1, q)) / q=0 (le modèle suit un ARMA (p, q-1)

H alternative : p≠0 (le modèle suit un ARMA (p, q)) / q≠0 (le modèle suit un ARMA (p, q)

Décision : Comme la p_value (0,0061) de ar2 et celle (0,045) de ma2 sont toutes les deux inferieurs au risque de 5% donc on rejette de l'hypothèse nulle (on accepte l'hypothèse alternative). Alors le modèle (2, 2) suit un ARMA (2, 2). D'après ce test de significativité, **on conclut que le modèle (2, 2) est lui aussi retenu.**

➢ **Pour le modèle (3, 0) :**

Figure N°10 : Estimation et test de significativité des coefficients du modèle30

```
> model30<-arima(pib1,order=c(3,1,0))
> model30

Call:
arima(x = pib1, order = c(3, 1, 0))

Coefficients:
         ar1     ar2      ar3
      0.6919  0.3393  -0.1317
s.e.  0.0625  0.0731   0.0626

sigma^2 estimated as 1.665:  log likelihood = -424.33,  aic = 856.66
> library(caschrono)
> t_stat(model30)
            ar1      ar2       ar3
t.stat 11.07182 4.638704 -2.105056
p.val   0.00000 0.000004  0.035286
```

Source : INSEE, Auteur

Les hypothèses :

Hypothèse nulle : p=0 (le modèle suit un ARMA (p-1, q))

Hypothèse alternative : p≠0 (le modèle suit un ARMA (p, q)).

Décision : Comme la p_value (0,035) de ar3 est inferieurs au risque de 5% donc on rejette de l'hypothèse nulle (on accepte l'hypothèse alternative). Alors le modèle (3, 0) suit un ARMA (3, 0). D'après ce test de significativité, **on conclut que le modèle (3, 0) est lui aussi retenu.**

➢ **Pour le modèle (3, 1) :**

Figure N°11 : Estimation et test de significativité des coefficients du modèle31

```
> model31<-arima(pib1,order=c(3,1,1))
> model31

Call:
arima(x = pib1, order = c(3, 1, 1))

Coefficients:
         ar1     ar2      ar3     ma1
      0.5833  0.4131  -0.1070  0.1115
s.e.  0.2207  0.1524   0.0844  0.2148

sigma^2 estimated as 1.663:  log likelihood = -424.2,  aic = 858.4
> library(caschrono)
> t_stat(model31)
            ar1      ar2       ar3      ma1
t.stat 2.642991 2.710941 -1.267041 0.518870
p.val  0.008218 0.006709  0.205141 0.603851
```

Source : INSEE, Auteur

Les hypothèses :

H nulle : p=0 (le modèle suit un ARMA (p-1, q)) / q=0 (le modèle suit un ARMA (p, q-1)

H alternative : p≠0 (le modèle suit un ARMA (p, q)) / q≠0 (le modèle suit un ARMA (p, q)

Décision : Comme la p_value (0,205) de ar3 et celle (0,6039) de ma1 sont toutes les deux supérieurs au risque de 5% donc on accepte de l'hypothèse nulle (le modèle suit un ARMA (p-1, q-1)). Alors le modèle (3, 1) suit un ARMA (2, 1). D'après ce test de significativité, **on conclut que le modèle (3, 1) n'est pas retenu.**

- **Pour le modèle (3, 2) :**

Figure N°12 : Estimation et test de significativité des coefficients du modèle32

```
> model32<-arima(pib1,order=c(3,1,2))
> model32

Call:
arima(x = pib1, order = c(3, 1, 2))

Coefficients:
          ar1     ar2     ar3     ma1     ma2
      -0.7912  0.6640  0.8261  1.5237  0.7254
s.e.   0.0774  0.0523  0.0565  0.1050  0.0810

sigma^2 estimated as 1.559:  log likelihood = -416.28,  aic = 844.56
> library(caschrono)
> t_stat(model32)
             ar1      ar2      ar3      ma1      ma2
t.stat -10.21827 12.70637 14.62223 14.51072 8.957805
p.val    0.00000  0.00000  0.00000  0.00000 0.000000
```

Source : INSEE, Auteur

Les hypothèses :

H nulle : p=0 (le modèle suit un ARMA (p-1, q)) / q=0 (le modèle suit un ARMA (p, q-1)

H alternative : p≠0 (le modèle suit un ARMA (p, q)) / q≠0 (le modèle suit un ARMA (p, q)

Décision : Comme la p_value (0,000) de ar3 et celle (0,000) de ma2 sont toutes les deux inferieurs au risque de 5% donc on rejette de l'hypothèse nulle (on accepte l'hypothèse alternative). Alors le modèle (3 2) suit un ARMA (3, 2). D'après ce test de significativité, **on conclut que le modèle (3, 2) est lui aussi retenu.**

3.3. Validation des modèles

La validation des modèles estimés se fait à l'aide du test de significativité des paramètres qui doivent être significativement différents de zéro et d'une analyse sur les résidus estimés permettant de tester les hypothèses sur le processus d'innovation (εt). L'estimateur de l'erreur εt doit suivre un processus de bruit blanc (BB). Si plusieurs modèles sont validés, l'étape de validation se poursuit par la comparaison des critères (AIC, BIC, HQ).

- D'après le test de significativité des coefficients précédemment effectué, les modèles valide sont :(0, 1) ; (0, 2) ; (2, 0) ; (2, 1) ; (2, 2) ; (3, 0) et (3, 2). Maintenant on va procéder à une analyse sur les résidus des modèles valides en faisant le test de Box Pierce sur les résidus des modèles valides.

Le test de Box Pierce est un test statistique qui permet de savoir si une d'un groupe de autocorrélations d'une série de temps sont différents de zéro. Au lieu de tester aléatoirement à chaque décalage distinct, il teste le caractère aléatoire "global" basé sur un certain nombre de retard, et est donc un test de valise.

Le test de Box Pierce est couramment utilisé dans autorégressif moyenne mobile intégrée (ARIMA) modélisation. Notez qu'il est appliqué aux résidus d'un modèle ajusté ARIMA, pas la série originale, et dans de telles applications étant en fait l'hypothèse testée est que les résidus du modèle ARIMA n'ont pas d'autocorrélation.

Les hypothèses de ce test sont définies comme suit :

H_0 : Les données (les résidus) sont indépendamment distribuées (les résidus suivent un bruit blanc (BB)) : c'est à dire les corrélations dans la population d'où l'échantillon est prélevé sont 0, de sorte que toute corrélations dans les données observées résultent de caractère aléatoire du processus d'échantillonnage.

H_a : Les données (les résidus) ne sont pas réparties de façon indépendante ; ils présentent une corrélation sérielle (les résidus ne suivent pas un bruit blanc (BB)).

- **Pour le modèle (0, 1) :**

Figure N°13 : Test de Box-Perce sur les résidus du modèle01

```
> Box.test(model01$residuals, lag=10, type="Box-Pierce")

	Box-Pierce test

data:  model01$residuals
X-squared = 190.1264, df = 10, p-value < 2.2e-16
```

Source : INSEE, Auteur

Décision : Comme le p_value du test (2,2e-16) de Box Pierce est inférieur au risque de 5% donc on accepte l'hypothèse nulle. Alors on conclut que les résidus sont indépendamment distribués. Autrement dit **les résidus du modèle (0, 1) suivent un bruit blanc**.

❖ **Pour le modèle (0, 2) :**

Figure N°14 : Test de Box-Perce sur les résidus du modèle02

```
> Box.test(model02$residuals, lag=30, type="Box-Pierce")

	Box-Pierce test

data:  model02$residuals
X-squared = 190.4938, df = 30, p-value < 2.2e-16
```

Source : INSEE, Auteur

Décision : Comme le p_value du test (2,2e-16) de Box Pierce est inférieur au risque de 5% donc on accepte l'hypothèse nulle. Alors on conclut que les résidus sont indépendamment distribués. Autrement dit **les résidus du modèle (0, 2) suivent un bruit blanc**.

❖ **Pour le modèle (2, 0) :**

Figure N°15 : Test de Box-Perce sur les résidus du modèle20

```
> Box.test(model20$residuals, lag=30, type="Box-Pierce")

	Box-Pierce test

data:  model20$residuals
X-squared = 77.075, df = 30, p-value = 5.158e-06
```

Source : INSEE, Auteur

Décision : Comme le p_value du test (5,158e-06) de Box Pierce est inférieur au risque de 5% donc on accepte l'hypothèse nulle. Alors on conclut que les résidus sont indépendamment distribués. Autrement dit **les résidus du modèle (2, 0) suivent un bruit blanc**.

❖ **Pour le modèle (2, 1) :**

Figure N°16 : Test de Box-Perce sur les résidus du modèle21

```
> Box.test(model21$residuals, lag=30, type="Box-Pierce")

        Box-Pierce test

data:  model21$residuals
X-squared = 70.6332, df = 30, p-value = 3.992e-05
```

Source : INSEE, Auteur

Décision : Comme le p_value du test (3,992e-05) de Box Pierce est inférieur au risque de 5% donc on accepte l'hypothèse nulle. Alors on conclut que les résidus sont indépendamment distribués. Autrement dit **les résidus du modèle (2, 1) suivent un bruit blanc**.

- **Pour le modèle (2, 2) :**

Figure N°17 : Test de Box-Perce sur les résidus du modèle22

```
> Box.test(model22$residuals, lag=30, type="Box-Pierce")

        Box-Pierce test

data:  model22$residuals
X-squared = 63.0526, df = 30, p-value = 0.0003871
```

Source : INSEE, Auteur

Décision : Comme le p_value du test (0,0003871) de Box Pierce est inférieur au risque de 5% donc on accepte l'hypothèse nulle. Alors on conclut que les résidus sont indépendamment distribués. Autrement dit **les résidus du modèle (2, 2) suivent un bruit blanc**.

- **Pour le modèle (3, 0) :**

Figure N°18 : Test de Box-Perce sur les résidus du modèle30

```
> Box.test(model30$residuals, lag=30, type="Box-Pierce")

        Box-Pierce test

data:  model30$residuals
X-squared = 68.2064, df = 30, p-value = 8.405e-05
```

Source : INSEE, Auteur

Décision : Comme le p_value du test (8,405e-05) de Box Pierce est inférieur au risque de 5% donc on accepte l'hypothèse nulle. Alors on conclut que les résidus sont indépendamment distribués. Autrement dit **les résidus du modèle (3, 0) suivent un bruit blanc**.

- **Pour le modèle (3, 2) :**

Figure N°19 : Test de Box-Perce sur les résidus du modèle32

```
> Box.test(model32$residuals, lag=30, type="Box-Pierce")

        Box-Pierce test

data:  model32$residuals
X-squared = 50.6012, df = 30, p-value = 0.01073
```

Source : INSEE, Auteur

Décision : Comme le p_value du test (0,01073) de Box Pierce est inférieur au risque de 5% donc on accepte l'hypothèse nulle. Alors on conclut que les résidus sont indépendamment distribués. Autrement dit **les résidus du modèle (3, 2) suivent un bruit blanc**.

- D'après le test de Box-Pierce sur les résidus des modèles valides, on conclut que tous les résidus des modèles restants suivent un bruit blanc (BB). Pour cela on va poursuivre la validation par la comparaison des critères d'information d'Akaike (AIC).

Le critère d'information d'Akaike, (en anglais Akaike information criterion ou AIC) est une mesure de la qualité d'un modèle statistique proposée par Hirotugu Akaike en 1973. Lorsque l'on estime un modèle statistique, il est possible d'augmenter la vraisemblance du modèle en ajoutant un paramètre. Le critère d'information d'Akaike, tout comme le critère d'information bayésien, permet de pénaliser les modèles en fonction du nombre de paramètres afin de satisfaire le critère de parcimonie. On choisit alors le modèle avec le critère d'information d'Akaike le plus faible pour faire la prévision.

- Comparaison des critères d'information d'Akaike :

Figure N°20 : Comparaison des critères d'information d'Akaike des modèles restants

```
> # Pour le Modèle (0, 1)
> AIC(model01)
[1] 1088.586
> # Pour le Modèle (0, 2)
> AIC(model02)
[1] 971.8723
> # Pour le Modèle (2, 0)
> AIC(model20)
[1] 859.0475
> # Pour le Modèle (2, 1)
> AIC(model21)
[1] 857.6566
> # Pour le Modèle (2, 2)
> AIC(model22)
[1] 856.2087
> # Pour le Modèle (3, 0)
> AIC(model30)
[1] 856.6559
> # Pour le Modèle (3, 2)
> AIC(model32)
[1] 844.5615
```

Source : INSEE, Auteur

- D'après la comparaison des critères d'information d'Akaike, le modèle qui possède le critère d'information d'Akaike le plus faible est le modèle (3, 2) (model32). Alors on va prendre le modèle (3, 2) pour faire la prévision du PIB.

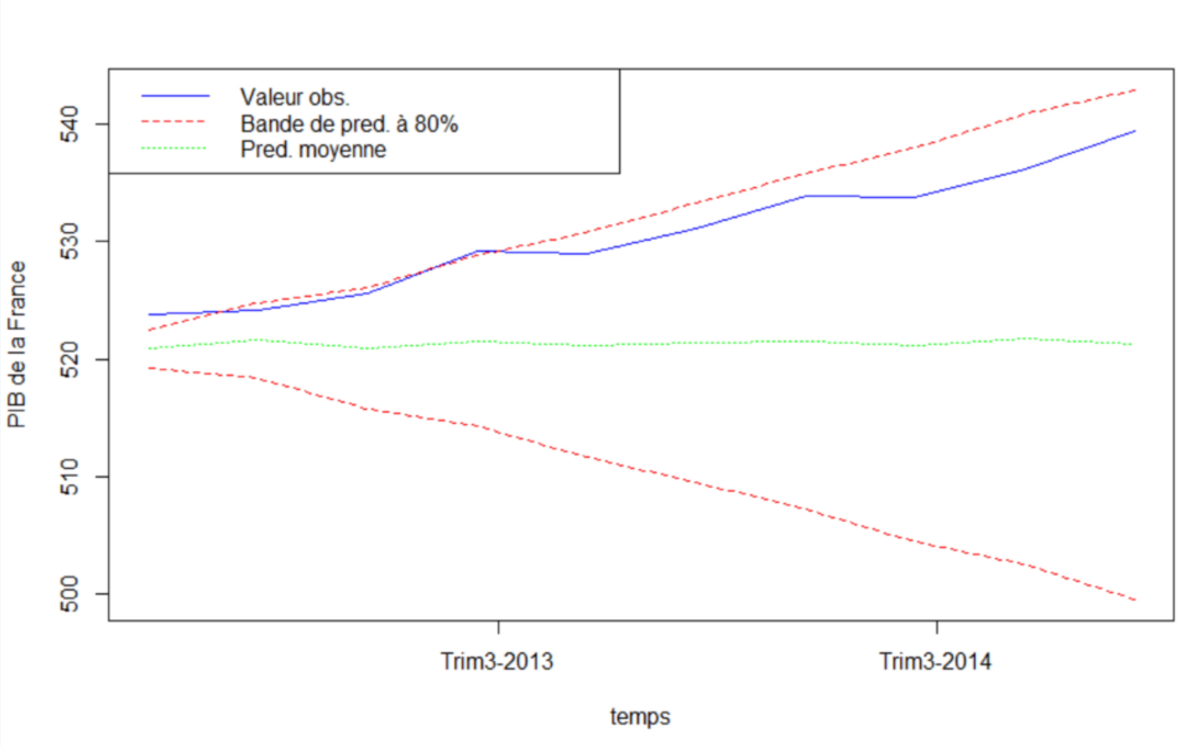
3.4. Prévision

Pour la prévision, on va appuyer surle modèle (3, 2) car c'est le modèle qui a le plus petit critère d'information d'Akaike.La série s'arrêtant au deuxième trimestre de 2012, la prévision pour chaque trimestre jusqu'au quatrième trimestre de 2014 s'obtient par forecast() avec un horizon de 10 trimestres et un niveau de confiance de 80%.

```
> prevision_PIB=forecast(model32,h=10,level=80)
```

Si nous examinons la structure de la sortie par « str (prevision_PIB) », nous observons que la commande « forecast() » produit une liste de 10 éléments dont certains sont eux-mêmes des listes. L'élément « mean » de type Time-Series est la prévision ponctuelle, c'est-à-dire la moyenne conditionnelle au passé arrêté au deuxième trimestre de 2012 et les limites de la bande de prédiction sont données par lower et upper. Superposons maintenant la réalisation et la prédiction du PIB :

Figure N°21 : La prévision pour la période du 3ème trimestre 2012 au 4ème trimestre 2014



Source : INSEE, Auteur

Nous notons (fig. 20) que la prévision du PIB est constamment inferieur la réalisation du PIB et surtout, que la bande de prévision à 80% contient presque toute la série. Il est manifeste que le modèle (3, 2) convient bien partir du 3ième trimestres 2012.

3.5. Test de de ARCH

Maintenant, nous allons faire le test de ARCH pour voir s'il y a hétéroscédasticité des erreurs (la variance des erreurs bouge) pour pouvoir appliquer une modélisation GARCH à notre série (PIB).

Nous parlons d'hétéroscédasticité lorsque les variances des variables examinées sont différentes. La notion d'hétéroscédasticité s'oppose à celle d'homoscédasticité, qui correspond au cas où la variance de l'erreur des variables est constante. Le cas d'homoscédasticité, nous avons $Var(\varepsilon_i) = \sigma^2\ \forall i$, tandis que dans le cas hétéroscédasticité nous avons $Var(\varepsilon_i) = \sigma_i^2$, où σ_i^2 peut être différent de σ_j^2, pour $i \neq j$.

❖ Test de ARCH :

Le test de ARCH appelé aussi test de multiplicateur de Lagrange permet de tester l'hypothèse nulle d'homoscédasticité contre l'hypothèse alternative d'une composante ARCH dans la série. Le test est fondé sur le test de Fisher classique. Dans la pratique, on travaille TR^2 avec T le nombre d'observation et R^2 le pouvoir prédictive du modèle.

Figure N°22 : Test d'ARCH sur les résidus du modèle32

```
> library(FinTS)
> ArchTest(res,50)

        ARCH LM-test; Null hypothesis: no ARCH effects

data:  res
Chi-squared = 131.5426, df = 50, p-value = 2.935e-09
```

Source : INSEE, Auteur

Décision : Comme le p_value qui est 2,935e-09 est inférieur au risque de 5% donc on accepte l'hypothèse nulle d'homoscédasticité des erreurs.

Alors on peut conclure que la variance des erreurs ne bouge pas et que ce n'est pas la pleine d'appliquer une modélisation GARCH à notre série d'étude (PIB).

Conclusion

Pour achever ce travail, on peut conclure que c 'est le modèle d'ARIMA (3, 1, 2) qui permet de mieux modéliser le PIB de la France par la méthode de Box-Jenkins.

D'après le test de ARCH, comme la variance des erreurs du modèle d'ARIMA (3, 1, 2) reste constante (homoscédasticité) alors on n'a pas besoin de faire une modélisation GARCH sur le PIB français. Car la modélisation de Box-Jenkins nous suffit de bien expliquer l'évolution du PIB de la France du premier trimestre 1949 jusqu'au dernier trimestre 2014.

Enfin, le modèle d'ARIMA (3, 1, 2) nous donne une bonne compréhension et une bonne prévision du Produit Intérieur Brut français car le PIB observé du troisième trimestre 2012 jusqu'au dernier trimestre 2014 se trouve bien dans la bande de prédiction à 80%.

BIBLIOGRAPHIE

1. **Publication :**

- Cours de séries temporelles linéaires ENSAE 2ème année Note sur les tests de racine unité par Catherine DOZ, Avril 2004

ISSN : 2728- 0128
Volume 3 : Numéro 7

- Document de Travail, analyse et prévision des séries temporelles par la méthode de box & Jenkins par la direction de la prévision et des études économiques en décembre 2007.

2. **Livre :**

- Prévision des ventes par régis bourbonnais
- Séries temporelles avec R Méthodes et cas de Yves Aragon
- Documentation en ligne du logiciel R :

https://stat.ethz.ch/R-manual/R-patched/library/stats/html/box.test.html

3. **Site web :**

- Le site de l'Insee : http://www.insee.fr/fr/
- Le site de l'atlantico : http://www.atlantico.fr
- Le site de Wikipédia : http://en.wikipedia.org/
- Help Center :

https://help.xlstat.com/tutorial-guides/time-series-analysis

https://help.xlstat.com/getting-started/data-analysis