



Rapport d'étude économétrique sur l'investissement des administrations publiques aux Etats-Unis de 1947 à 2022

Enseignante : Mme Ghattassi

Etudiant : Hraouli Samir

Date : 26/03/2023

Master Ingénierie Financière et Modélisation

Table des matières

I. Introduction.....	3
II. Transformation de la donnée	3
III. Représentation des investissements des administrations publiques	3
a. Test de stationnarité.....	4
IV. Estimation du modèle ARIMA	5
a. ACF et PACF	5
b. Test de saisonnalité	5
c. Tests de modèles ARIMA.....	6
V. Diagnostic des résidus.....	7
a. Analyse des résidus	7
b. Test Box Ljung	7
c. Test de Breusch Godfrey.....	8
d. Test de Shapiro.....	8
VI. Prédiction de l'investissement des administrations publiques pour 2023.....	9
VII. Conclusion.....	9
VIII. Bibliographie	10

I. Introduction

Ce projet met en lumière les investissements des administrations publiques des Etats-Unis. Les objectifs des investissements publics sont de favoriser la croissance économique en finançant des projets publics tels que des infrastructures publiques. Ce rapport présente une étude économétrique menée sur les investissements des administrations publiques des Etats-Unis depuis 1947 jusqu'à 2022.

II. Transformation de la donnée

Dans un premier temps, un travail de remise en forme de la donnée était nécessaire car les données fournies par la bea n'étaient pas exploitables. En effet, celles-ci étaient présentées tels que chaque trimestre correspondent à une colonne. J'ai donc programmé un script sous Python à l'aide des librairies Pandas pour le traitement de Dataframe et Numpy pour générer une séquence de date. Après quelques lignes de code, j'ai pu générer un dataset, avec des données organisées par ligne, où chaque investissement du gouvernement est assigné à une date, et un trimestre. Le script Python est disponible sur mon github à l'adresse suivante :

[econometrics/projet_semestre2_IFIM_timeseries at main · samhrrl/econometrics \(github.com\)](https://github.com/samhrrl/econometrics)

III. Représentation des investissements des administrations publiques

Après avoir transformé la donnée, nous pouvons faire quelques statistiques descriptives et observer comment évoluent les investissements des administrations publiques aux Etats-Unis. En effet, on constate que celle-ci est croissante, elle ne cesse d'évoluer chaque année.

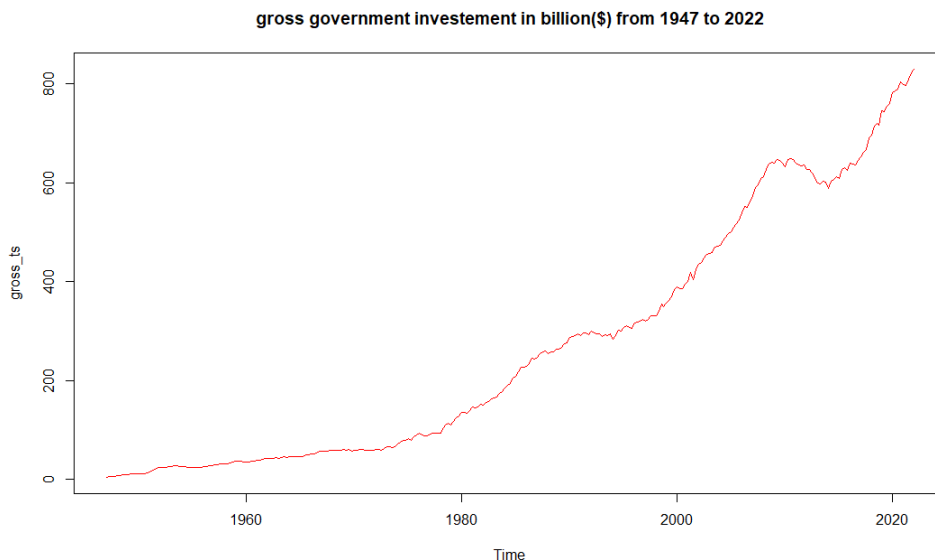


Figure 1 Investissements des administrations publiques aux Etats-Unis de 1947 à 2022

Dans les années 80 et 90, les investissements des administrations publiques ont fortement augmenté notamment dans les secteurs de l'éducation, les infrastructures publiques, les soins santé et la sécurité sociale.

En outre, pendant la période covid, les Etats-Unis ont engagés des dépenses massives pour diminuer les effets économiques et sanitaires pour lutter contre cette crise.

Enfin, les élections présidentielles de 2020 aux Etats-Unis ont remis en cause les dépenses et investissements des administrations publiques. En effet, Joe Biden adopte une politique keynésienne, donc il est prêt à relancer l'économie en stimulant la croissance du pays en créant de l'emploi et en investissant dans des infrastructures publiques.

a. Test de stationnarité

La série présentée montre qu'elle n'est pas stationnaire, nous allons donc la stationnariser en prenant le logarithme de la différence première afin de mener cette étude économétrique.

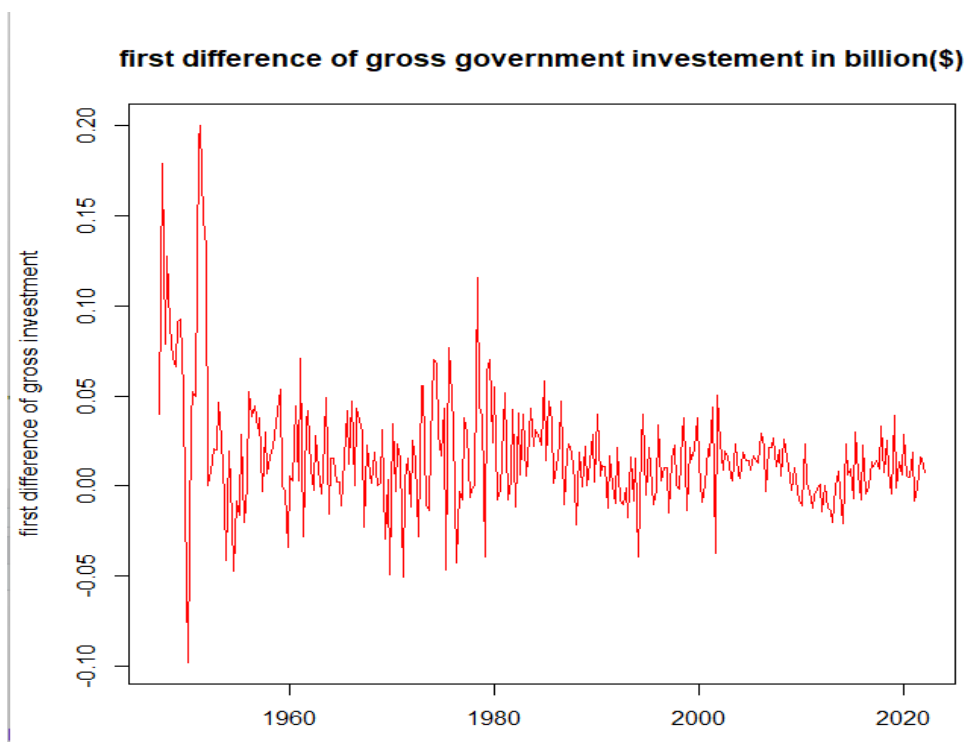


Figure 2 Différence première des investissements des administrations publiques des Etats-Unis

Après une différence première, la série semble stationnaire, nous allons faire un test statistique pour vérifier qu'elle est bien stationnaire.

```
Augmented Dickey-Fuller Test
data: diff_premiere
Dickey-Fuller = -5.9674, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Figure 3 Test de Dickey-Fuller

En réalisant le test de Dickey-Fuller, on constate que la $p\text{-value}=0.01 < 5\%$, donc nous pouvons rejeter l'hypothèse nulle qui affirme que la série n'est pas stationnaire.

IV. Estimation du modèle ARIMA

a. ACF et PACF

Nous avons conclu par un test de Dickey-Fuller que les investissements des administrations publiques des Etats-Unis est stationnaire. Nous allons après-en visualiser la fonction d'auto corrélation et leur fonction d'auto corrélation partiel pour estimer un modèle ARIMA.

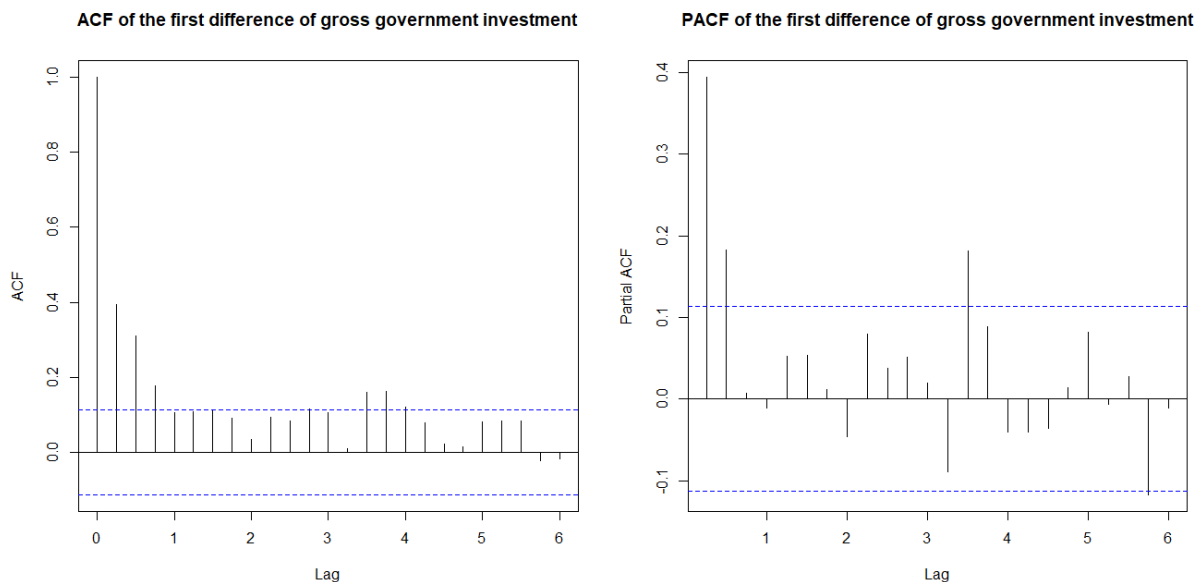


Figure 4 ACF et PACF des investissements des administrations publiques

On constate sur l'ACF, une décroissance forte de nos autos corrélations, on peut donc penser qu'il s'agit d'un modèle autorégressif $AR(p)$. Cependant, on constate un effet de saisonnalité qui se répète régulièrement. Il faudra donc faire un test de saisonnalité pour vérifier s'il y a un aspect saisonnier ou non.

Sur le PACF, on observe un effet sinusoïdale des nos autos corrélations ce qui nous laisse penser qu'il s'agit d'un modèle à moyenne mobile $MA(q)$. De plus, on constate une rupture à l'ordre 2. On peut donc mettre en lumière qu'on aurait un $AR(2)$. Quant à l'ACF, on peut voir une rupture à l'ordre 3 ou 4. On peut donc penser qu'il s'agit d'un $MA(3)$ ou d'un $MA(4)$. Ces spéculations sont établies d'un point de vue conjecturel, nous testerons par la suite, à l'aide de test, quel modèle ARIMA on pourrait modéliser.

b. Test de saisonnalité

J'ai réalisé un test statistique qui permet d'évoquer ou non s'il y a un effet saisonnier. J'ai utilisé la librairie (seastests) sur R pour faire un test de saisonnalité. Le test présente deux hypothèses :

H₀ : la série présente un effet de saisonnalité, le test retourne un booléen **TRUE**

H_a : la série ne présente pas d'effet de saisonnalité, le test retourne un booléen **FALSE**

Après avoir réalisé le test, j'en ai déduit que la série ne présentait pas d'aspect saisonnier car le test m'a retourné FALSE.

c. Tests de modèles ARIMA

Comme la série présente ne présente pas d'aspect saisonnier, j'ai testé plusieurs modèles ARIMA à l'aide de la fonction **estimate** sur R. Pour choisir le modèle, j'ai pris en considération des critères tels que la significativité des t-value, qui devait être supérieur à 1.96 en valeur absolue, la p-value qui devait être significatif au seuil des 10% et le modèle qui minimise le critère d'information AIC. A partir de ces critères, j'ai testé plusieurs modèles, et j'ai constaté lors des premiers tests que la significativité n'était pas toujours correcte, j'ai donc décidé de faire varier mes paramètres p et q en testant des ARIMA(3,0,4), ARIMA(4,0,4), ARIMA(5,0,4), ARIMA(4,0,5) et ARIMA(5,0,3). J'ai conclu que le meilleur modèle était ARIMA (5, 0,3), car mes résultats étaient significatifs sauf pour l'AR(4) où ma t-value = -0.775 < 1.96 et AR(5) = -1.461 < 1.96, Seul, le MA(3) présente un résultat non significatif sur sa p-value, mais j'ai décidé d'accepter le modèle car les autres valeurs étaient acceptables et que ce modèle minimisait également le critère d'information AIC= -1239.748.

ARIMA(5,0,3) model is estimated for variable: diff_première

Conditional-Sum-of-Squares & Maximum Likelihood Estimation

	Estimate	S.E	t.value	p.value	Lag
MU	0.0213	0.0100	2.126	3.44e-02	1
AR 1	1.4638	0.0772	18.972	0.00e+00	1
AR 2	-1.3576	0.1082	-12.552	3.56e-29	2
AR 3	1.0486	0.1201	8.732	2.22e-16	3
AR 4	-0.0828	0.1068	-0.775	4.39e-01	4
AR 5	-0.0955	0.0653	-1.461	1.45e-01	5
MA 1	-1.1741	0.0505	-23.256	2.47e-68	1
MA 2	1.2329	0.0364	33.831	0.00e+00	2
MA 3	-0.9213	0.0546	-16.869	5.29e-45	3

Residual Diagnostics Plots

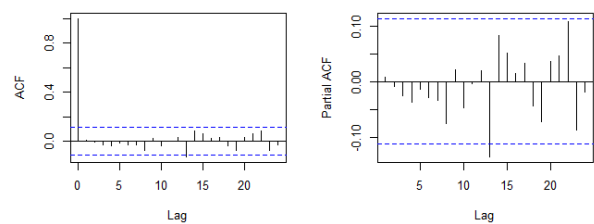


Figure 5 Estimation d'un modèle ARIMA(5,0,3)

Series: gross_ts
ARIMA(0,2,2)(0,0,2)[4]

Coefficients:

	ma1	ma2	sma1	sma2
	-1.1132	0.2969	0.1480	-0.1804
s.e.	0.0593	0.0625	0.0643	0.0566

sigma^2 = 25.27: log likelihood = -906.03
AIC=1822.05 AICc=1822.26 BIC=1840.56

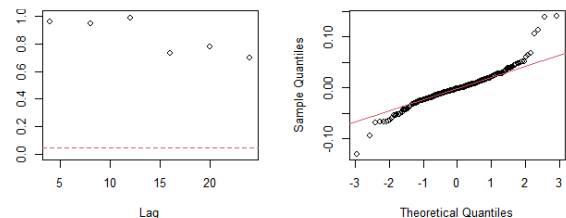


Figure 6 Résidus du modèle ARIMA(5,0,3)

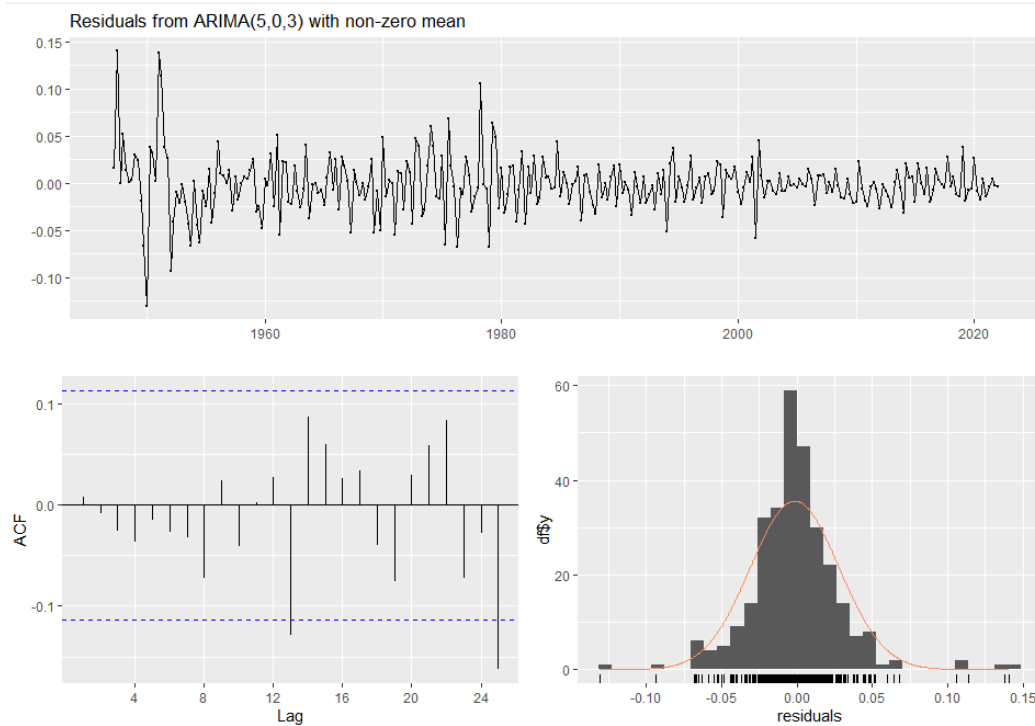
Figure 7 Estimation d'un modèle ARIMA automatique sous R

Enfin, j'ai également interrogé Rstudio pour qu'il me propose un modèle automatique ARIMA. Par conséquent, le modèle proposé est un ARIMA (0,2, 2). J'ai rejeté ce modèle car celui-ci ne contient pas de composante AR. Le modèle fourni par R est un MA pur qui ne contient pas de données observables mais uniquement un ensemble de chocs. De plus, le modèle proposé par R utilise la différence seconde, alors que la série est stationnaire après une différence première. Donc, on perd de l'information.

V. Diagnostic des résidus

Nous allons réaliser un diagnostic des résidus afin de s'assurer que le modèle ARIMA choisi est approprié pour faire des prévisions futures.

a. Analyse des résidus



D'après le graphique, on constate que les résidus se concentrent autour de la moyenne. De plus, on observe un pic conséquent dans les années 45. En effet, pendant la période post seconde guerre mondiale, les Etats-Unis ont massivement dépensé pour leur défense mais aussi pour le soutien apporté à leurs alliés. De plus, après la guerre, le gouvernement a investi dans la reconstruction de l'Europe dans le cadre du plan Marshall et a commencé à investir dans des projets plus complexes tels que la construction d'infrastructure, l'innovation et la recherche scientifique ainsi que l'éducation ; Ceci peut expliquer pourquoi la volatilité est haute pendant cette période.

b. Test Box Ljung

Nous allons ensuite réaliser un **test d'indépendance des résidus**. Ce test est constitué de deux hypothèses :

H₀ : la corrélation entre les résidus est très faible (indépendance)

H_a : la corrélation entre les résidus est significativement différente de 0 (il n'y a pas indépendance)

Box-Ljung test

```
data: fit$residuals
X-squared = 3.6288, df = 10, p-value = 0.9625
```

J'en conclus, avec le test de Box Ljung que la p-value=0.9625, ce résultat étant supérieur à 10%, nous pouvons donc rejeter l'hypothèse H_0 que la corrélation entre les résidus est significativement différente de 0. Donc, les résidus sont indépendants.

c. Test de Breusch Godfrey

Nous allons ensuite faire un test d'homoscédasticité des résidus afin de vérifier que nos résidus sont homoscédastiste, c'est-à-dire que leur variance ne varie pas dans le temps. Nous utiliserons le test de Breusch Godfrey sur R. Ce teste présente deux hypothèses :

H_0 : les résidus présente un cas d'homoscédasticité, la variance est donc constante dans le temps

H_a : les résidus présente un cas d'hétéroscédasticité, la variance varie dans le temps.

Après avoir réalisé le test, nous obtenons les résultats suivants

```
studentized Breusch-Pagan test

data: fit$residuals ~ fitted(fit)^2
BP = 21.789, df = 1, p-value = 3.043e-06
```

Nous constatons que la p-value=3.04x10E(-6), est inférieur au seuil des 10%. Donc, je rejette l'hypothèse nulle H_0 . Ainsi, la série n'est pas un cas d'homoscédasticité, la variance varie dans le temps, il s'agit d'un cas d'hétéroscédasticité.

Celle-ci est rejeté en raison de la haute volatilité de nos résidus notamment le pic conséquent qu'on retrouve pendant la période après guerre.

d. Test de Shapiro

Nous finirons par un test de normalité des résidus pour savoir si les résidus suivent ou non une distribution normale avec un test de Shapiro. On constate après les résultats du test, que la p-value est faible : 6.15x10E (-11), qui est inférieur au seuil des 10%, j'en conclus que les résidus ne suivent pas une distribution normale.

```
shapiro-wilk normality test

data: fit$residuals
W = 0.92719, p-value = 6.157e-11
```


VI. Pr vision de l'investissement des administrations publiques pour 2023

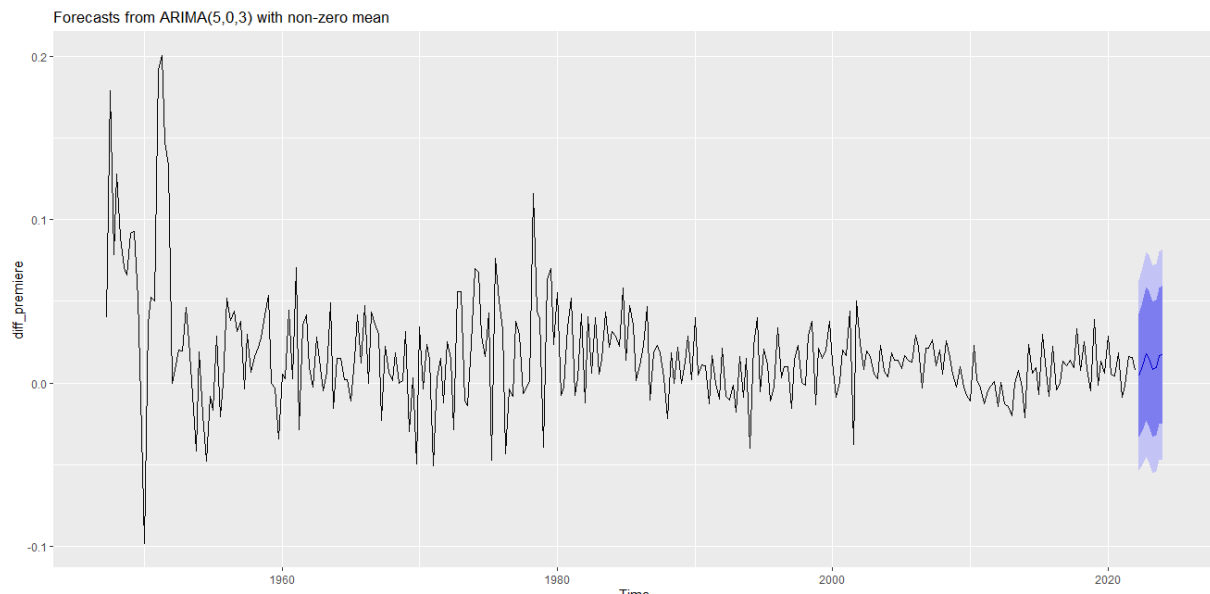


Figure 8 Pr vision des investissements des administrations publiques pour l'ann e 2023

Nous pouvons alors faire des pr visions pour 2023 en utilisant la librairie forecast. On constate que l'investissement des administrations publiques semble augmenter en d but d'ann e et diminuer par la suite avant de recro tre au trimestre 1 de 2024.

Nous pouvons consid rer cette augmentation des investissements car les Etats-Unis sont fortement impliqu  dans le conflit de guerre en Ukraine en leur versant des aides financi res et mat rielles qui s' l vent   50 milliard de dollars.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2022 Q2	0.004678844	-0.03313823	0.04249592	-0.05315738	0.06251507
2022 Q3	0.010494821	-0.02888464	0.04987429	-0.04973087	0.07072052
2022 Q4	0.017958884	-0.02295976	0.05887753	-0.04462078	0.08053855
2023 Q1	0.015308783	-0.02610361	0.05672118	-0.04802601	0.07864357
2023 Q2	0.008405714	-0.03308936	0.04990078	-0.05505552	0.07186695
2023 Q3	0.009537136	-0.03202291	0.05109718	-0.05402346	0.07309774
2023 Q4	0.016613147	-0.02519576	0.05842205	-0.04732805	0.08055435
2024 Q1	0.017703264	-0.02436638	0.05977290	-0.04663670	0.08204323

VII. Conclusion

Nous avons men  cette  tude  conom trique en estimant et en faisant une pr vision des futures valeurs de l'investissement des administrations publiques aux Etats-Unis avec le meilleur mod le ARIMA en tenant compte de la significativit  de nos r sultats ainsi que le mod le minimisant le crit re d'information AIC. Toutefois, le mod le ARIMA ne serait pas le plus adapt  en raison de la haute volatilit  de notre mod le, o  le test d'homosc dasticit  est rejet . Il serait donc pr f rable d'orienter cette  tude en utilisant un mod le GARCH ou ARCH qui est un mod le reposant sur la haute volatilit .

VIII. Bibliographie

[États-Unis – Décision de la FED sur les taux d'intérêt \(investing.com\)](#)

[8.7 ARIMA modelling in R | Forecasting: Principles and Practice \(2nd ed\) \(otexts.com\)](#)

[r - Plot forecast and actual values - Stack Overflow](#)

[Nouveau sursaut de l'inflation en janvier aux États-Unis | Le Devoir](#)

[Le keynésianisme impérialiste de Joe Biden - CONTRETEMPS](#)

[gov_glance-2013-23-fr.pdf \(oecd-ilibrary.org\)](#)