

Projet économétrie.

La loi d'Okun

16/12/2016

Université de Bordeaux.

Max FITZPATRICK

M1 Banque Finance

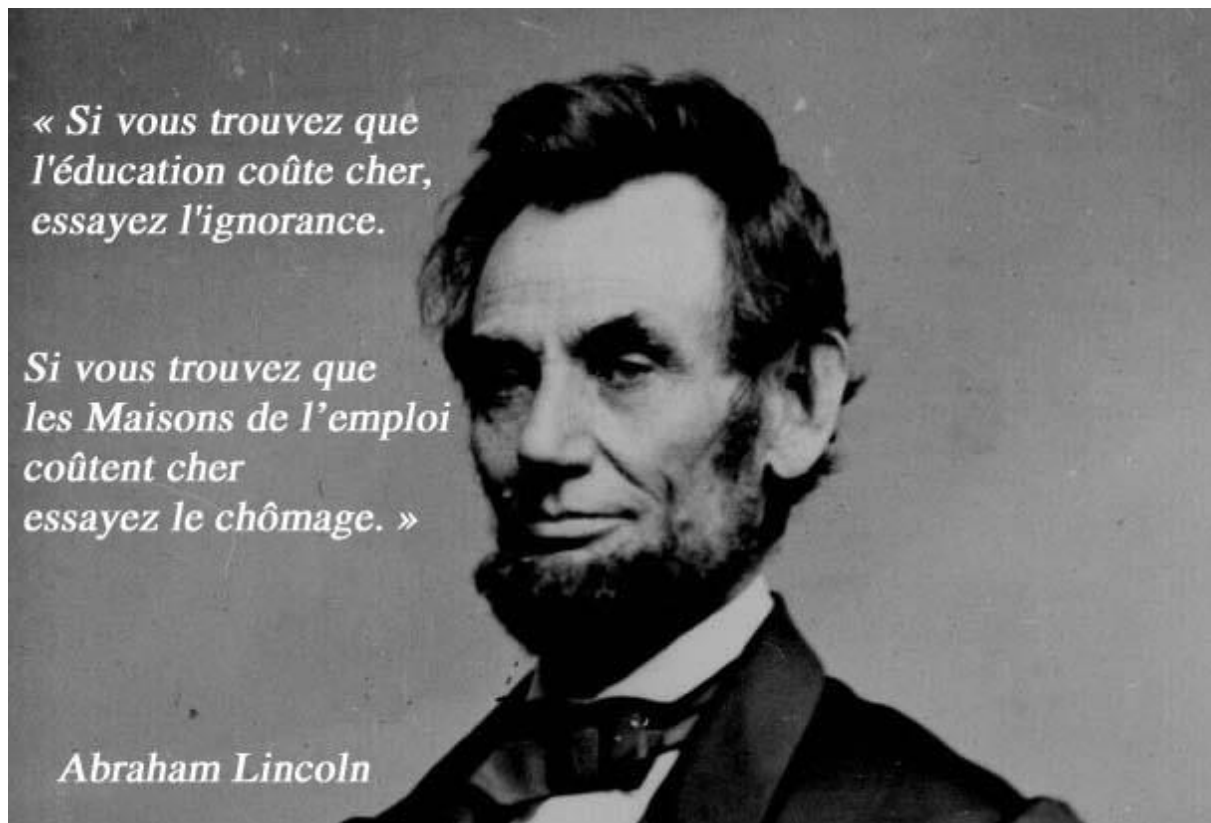
université
de **BORDEAUX**

Table des matières.

Introduction	4
Présentation et histoire de la loi d'Okun.....	5
Qu'est-ce que le chômage : Étymologie, définition, typologie ?	7
La croissance économique : Citation, Étymologie, Théoriciens.	8
Trois grandes études empiriques validant la loi d'Okun.	10
La loi d'Okun et l'exemple français.....	11
Quelles sont les conséquences de la Loi d'Okun ?	11
Le coefficient d'Okun en France.....	12
Présentation des données.....	13
Sources et définitions.....	13
Exploration des données.....	14
Chômage et PIB ensemble	18
I : Stationnarisation des données	20
Stratégie de test Dickey-Fuller	21
Chômage.....	22
PIB	25
La cointégration des variables	27
Identification de sous-périodes (ruptures structurelles)	27
Régression de la période 1 (1948q2 à 1959q3)	29
Table ANOVA.....	30
L'espérance des résidus	34
Normalité des résidus	35
Homoscédasticité des résidus	36
Multicolinéarité	36
Indépendance/autocorrélation.....	36
Goodness of fit.....	38
Conclusion	38
Régression de la période 2 (1959q4 à 2003q3)	39
ANOVA.....	39
Espérance	43
Normalité.....	43
Homoscédasticité.....	44
Multicolinéarité	44
Indépendance/autocorrélation.....	44
Goodness of fit.....	45

Conclusion	46
Régression période 3 (de 2003q4 à 2016q3)	47
ANOVA.....	47
Espérance	49
Normalité.....	49
Homoscédasticité	50
Multicolinéarité	52
Indépendance/Autocorrélation	52
Goodness of fit.....	53
Conclusion	54
Conclusion.....	55
OUVERTURE.....	56

Introduction



"Le libéralisme ne semble pas pour autant, face à cette hébétude de la gauche, recueillir la sympathie massive des citoyens. Appliquée avec une implacable rigueur au cours de la décennie 80 aux Etats-Unis par M. Reagan et au Royaume-Uni par Mme Thatcher, cette doctrine économico-politique a entraîné de trop douloureuses conséquences sociales. Aggravation des inégalités, augmentation du **chômage**, désindustrialisation, dégradation des services publics, délabrement des équipements collectifs..."

Ignacio Ramonet - Le Monde diplomatique, mai 1992

©copyright

http://www.toupie.org/Citations/Liberalisme_2.htm

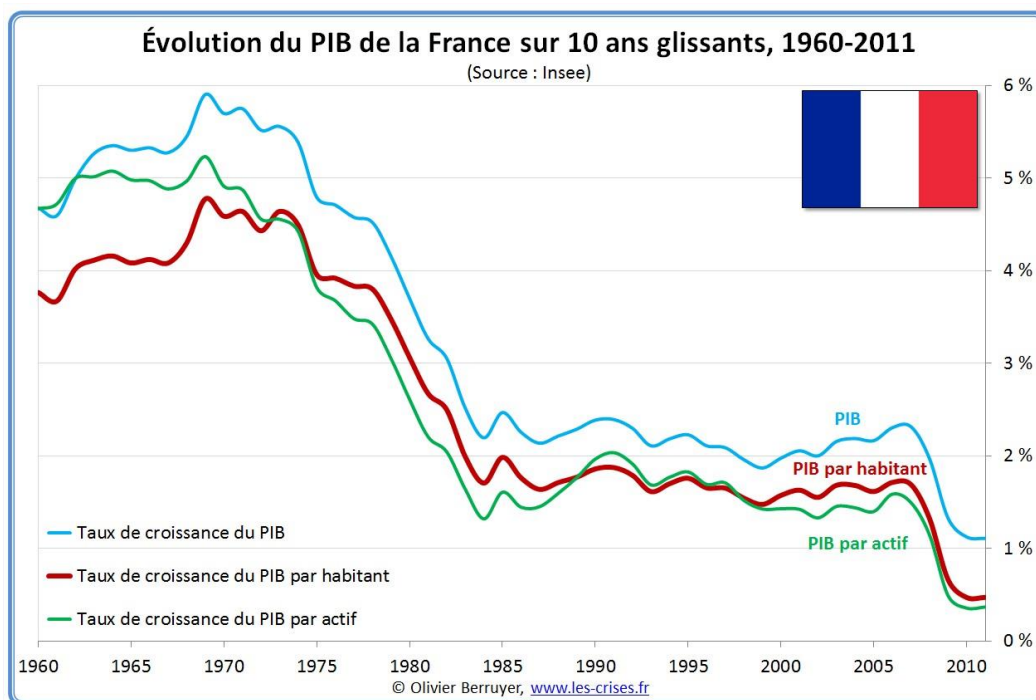
Présentation et histoire de la loi d'Okun

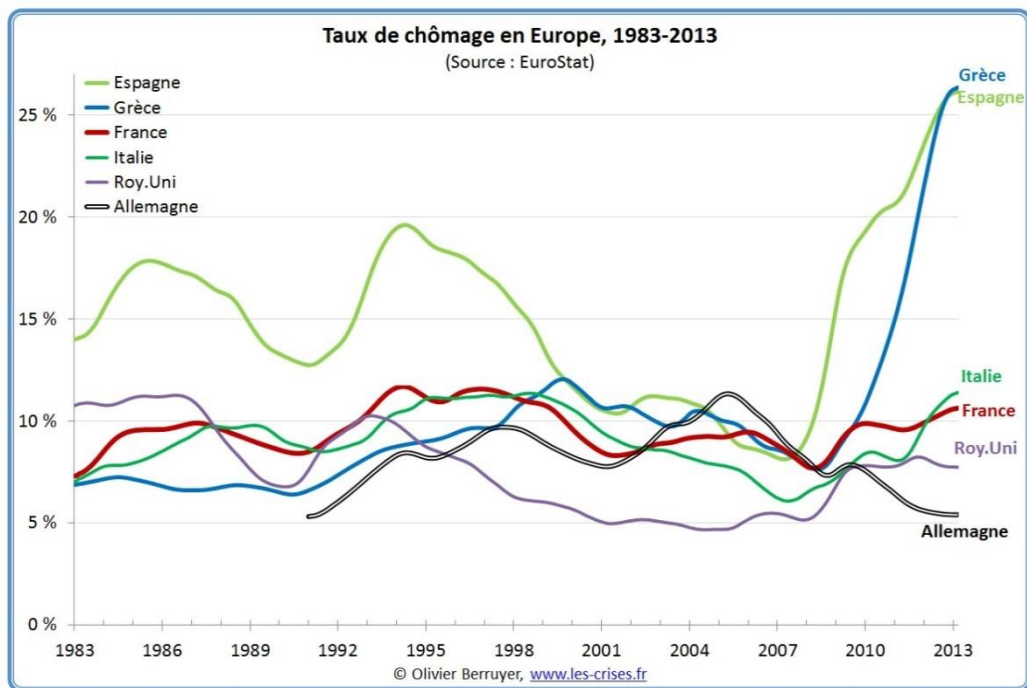
En cette période d'insécurité, d'instabilité financière, d'augmentation du chômage, d'inégalités démographiques, nous en venons à nous demander en quoi les variables clés qui doivent être impactées pourraient provoquer une meilleure sécurité face à une précarisation croissante et à une balkanisation progressive de la société.

C'est pourquoi nous en venons à utiliser une loi fondamentale, stable et durable qui établit une relation entre le taux de croissance et le taux du chômage. C'est en 1962 que l'honorable économiste américain Arthur Malvin Okun, conseiller économique du président Kennedy et professeur à l'Université de Yale établit cette loi d'Okun qui apporta et apportera aux gouvernements un outil puissant permettant d'optimiser et de rentabiliser les politiques économiques, sociales, équitables et démographiques.

Bien que présentant une multitude d'avantages dans la gouvernance économique, sociale et solidaire, cette relation varie en fonction des états, des époques et des dirigeants des différents États qui seront au cœur d'ouvrer pour favoriser le plein-emploi, grâce aux politiques entreprises qui favoriseront la croissance, afin d'obtenir le plein-emploi ou qui œuvreront au contraire pour stimuler les inégalités salariales, démographiques au profit de la compétitivité stratégique.

De plus, si le PIB augmente pour que le chômage diminue, il en ressort également qu'il faut absolument franchir un certain seuil de croissance pour parvenir à réussir à réduire le chômage. Ce seuil de croissance correspond approximativement à la croissance potentielle. Ces seuils peuvent différer en fonction des différents États étudiés, en effet chaque pays possède des caractéristiques structurelles et conjoncturelles qui déterminent leur trajectoire économique ce qui modifie le coefficient d'Okun.





Qu'est-ce que le chômage : Étymologie, définition, typologie ?



Source : <http://www.toupie.org/Dictionnaire/Chomage.htm>

Chômage vient du latin *caumare*, se reposer pendant la chaleur, venant du grec *kauma*, chaleur brûlante.

Le chômage est la situation d'une personne qui, souhaitant travailler et ayant la capacité de le faire, se trouve sans emploi malgré ses recherches. Du point de vue économique, le chômage est interprété comme la résultante d'un déséquilibre entre l'offre et la demande sur le marché du travail.

Il existe de nombreuses formes de chômage :

- Le chômage naturel (ou frictionnel) est dû au fait que chaque jour il y a, pour diverses raisons, de nouveaux demandeurs d'emploi, et qu'il y a toujours un temps de battement entre le début de la recherche de l'emploi et l'entrée dans un nouvel emploi. Un taux de chômage nul ne peut donc être atteint. Le plein-emploi est considéré comme atteint lorsque le taux de chômage se situe aux alentours de 3,5% à 4%.
- Le chômage conjoncturel (ou keynésien) correspond à un ralentissement de l'activité économique provoquant une réduction temporaire des besoins de main-d'œuvre dans l'économie. Les entreprises licencient pour adapter leur capacité de production à la baisse de l'activité économique. Le chômage conjoncturel se résorbe avec le retour de la croissance économique qui nécessite des embauches de la part des entreprises.
- Le chômage structurel est lié à des changements de structures économiques dans un pays, provoquant une inadéquation qualitative entre l'offre et la demande de travail. L'évolution des qualifications dues aux évolutions techniques conduit à rendre inemployable une partie de la population active qui ne trouve plus d'emplois correspondant à ses qualifications.
- Le chômage technique, au sein d'une entreprise, découle de l'impossibilité pour d'autres secteurs d'activité ou d'autres entreprises de lui fournir les éléments nécessaires à la fabrication de ses produits.
- Le chômage partiel peut être provoqué par une baisse d'activité anormale de l'entreprise qui est obligée de réduire les horaires de travail.
- Le chômage saisonnier concerne certaines branches professionnelles dont l'activité varie sensiblement selon les périodes de l'année.

La croissance économique : Citation, Étymologie, Théoriciens.

"La croissance n'a été capable ni de réduire la pauvreté ni de renforcer la cohésion sociale. Un même taux de croissance peut signifier un accroissement ou une réduction des inégalités. Et une croissance illimitée dans un monde fini est une illusion."

Manifeste Utopia – 2008



Source : <http://www.toupie.org/Dictionnaire/Croissance.htm>

Le mot croissance vient du latin *crescere*, croître, grandir.

En économie, la **croissance** désigne l'évolution annuelle, exprimée en pourcentage, du **PIB** (Produit intérieur brut) ou du **PNB** (Produit national brut). Pour éviter le problème dû à l'**augmentation des prix**, la croissance est calculée en "**monnaie constante**" (hors inflation), le PIB étant corrigé de l'augmentation de l'indice des prix. Ceci permet de calculer une croissance en volume.

La formule de calcul, dans le cas du PIB de l'année "n", est la suivante.

$$\text{Croissance} = [\text{PIB}_{(n)} - \text{PIB}_{(n-1)}] / \text{PIB}_{(n-1)}$$

On distingue généralement :

- **La croissance extensive** : augmentation des quantités de facteurs de production (culture de nouvelles terres, ouverture de nouvelles usines). La croissance extensive génère des créations d'emplois.
- **La croissance intensive** : augmentation, par des gains de productivité, de la production à volume de facteurs de production identiques, notamment sans création d'emplois supplémentaires.

Une croissance du PIB n'implique pas nécessairement une élévation du niveau de vie. En effet, si la croissance démographique est plus rapide que la croissance du PIB, le PIB par habitant diminue.

D'une manière plus générale, la croissance correspond, pour une nation, à une **augmentation soutenue et durable** - pendant une période suffisamment longue - de la **production de biens et de services** appréhendée par des indicateurs comme le PIB ou le PNB. Cependant, n'étant qu'une mesure quantitative d'un agrégat économique, la croissance n'est qu'une des composantes du développement qui est une notion plus abstraite et qualitative. Il peut donc y avoir croissance sans développement et inversement du développement sans croissance.

Les dernières décennies ont vu se succéder une série de cycles de croissance soutenue et de récession ou de faible croissance :

- croissance des années 1919-1929
- récession des années 1930
- forte croissance de l'après-guerre : les "30 glorieuses"
- ralentissement après le choc pétrolier de 1973
- fortes croissances des années 1980-1990
- ralentissement de 1992 à 1997 après la première guerre du Golfe
- reprise de la croissance de 1997 à 2001 (attentat du 11 septembre)

Différentes visions de la croissance chez quelques économistes :

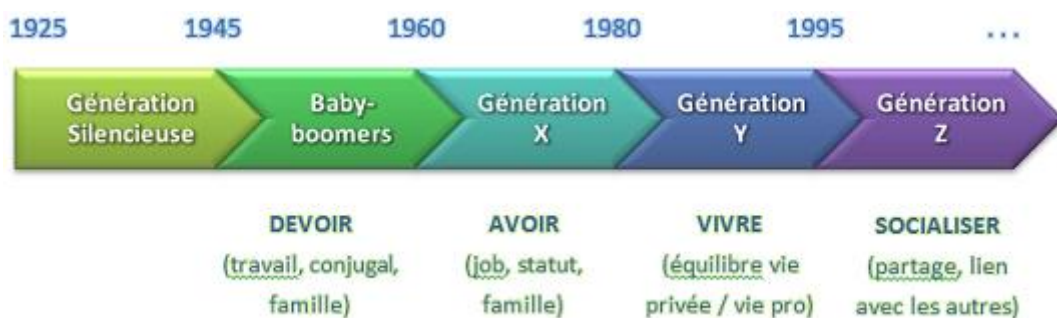
- **Adam Smith** (1723-1790)
Dans "Richesse", il développe les premiers éléments de la théorie de la croissance. Prenant sa source dans la division du travail, la croissance lui apparaît comme illimitée.
- **Thomas Robert Malthus** (1766-1834)
La croissance de la population, plus rapide que la production de la terre, conduit à des famines qui permettent de rétablir, à court terme, le bon rapport entre les deux... jusqu'à ce que l'écart entre population et production de la terre provoque une nouvelle crise.
- **David Ricardo** (1772-1823)
Pour faire face à la croissance de la population, de nouvelles terres doivent être cultivées. Or celles-ci ont un rendement décroissant (les meilleures étant déjà utilisées). Il s'ensuit inéluctablement à long terme un état stationnaire.
- **Karl Marx** (1818-1883)
Pour lui, l'accumulation du capital permet à ce dernier de se substituer au travail. L'augmentation du chômage et la baisse des salaires qui en découlent provoquent une baisse de la consommation et du taux de profit et par conséquent de la croissance.
- **John Maynard Keynes** (1883-1946)
L'économiste britannique insiste sur le rôle de l'Etat qui, par les investissements publics, peut relancer l'économie en jouant sur la demande et favoriser ainsi la croissance.

Trois grandes études empiriques validant la loi d'Okun.

Empiriquement, d'une part, [Sandrine Cazes, Sher Verick et Fares Al Hussami \(2013\)](#) analysent la Grande Récession et ils valident majoritairement la relation d'Okun tout en expliquant le fait que le chômage soit plus sensible par rapport à la croissance économique dans certains pays. De plus, suite à la Grande Récession, il semblerait que le coefficient d'Okun augmente aux États-Unis, au Canada, en Espagne alors qu'il a diminué à court terme dans des pays tels que l'Allemagne et les Pays-Bas.

D'autre part, [Laurence Ball, Daniel Leigh et Prakash Loungani \(2013\)](#) étudièrent la relation d'Okun pour les États-Unis d'Amérique entre 1948 et 2011 et ajoutèrent à leur étude 20 pays de l'OCDE entre 1980 et 2011. Ils affirmèrent alors que la loi d'Okun est l'une des relations les plus fiables et stables en macroéconomie pour la plupart des pays.

[Robert Dixon, G.C. Lim et Jan C. van Ours \(2016\)](#) quant à eux analysèrent les éventuelles asymétries potentielles et pour cela ils s'appuient sur les périodes d'expansion, de récession, et des variables temporelles. Ils utilisèrent pour cela une base de données très riche répertoriant 20 pays de l'OCDE pour la période de 1985 à 2013. Ils ont également l'intelligence de différencier les catégories d'âge, car rappelons-le, les jeunes travailleurs sont plus dans une propension à être dans une situation de précarité et au chômage que la génération des baby-boomers. De plus, ils utilisent les taux de chômage désagrégés selon l'âge et le genre. Ils concluront dès lors que la proportion des travailleurs temporaires a pu jouer un rôle déterminant pour expliquer les variations dans le coefficient d'Okun au cours du temps. Ils affirmeront donc que le coefficient d'Okun est également différent pour la génération Y, la génération des baby-boomers, les travailleurs d'âge intermédiaire et les travailleurs âgés.



©copyright

<http://www.3hcoaching.com/intergenerationnel-2/pas-de-jaloux-parlons-des-autres-generations-silencieuse-baby-boomers-x-et-z/>

La loi d'Okun et l'exemple français

En France à ce jour, il faut nécessairement que la croissance du PIB atteigne 1.9% pour que le chômage baisse, autant dire que nous sommes loin de cela. Mais d'où provient ce taux de 1.9% ?

Pour la période de 1990 à 2007, l'INSEE constate une hausse de la productivité horaire de 1.7% en moyenne par an, de plus nous avons une légère hausse de 0.9% en moyenne par an de la population active, à cela s'ajoute la baisse de la durée du travail de - 0.7% par an.



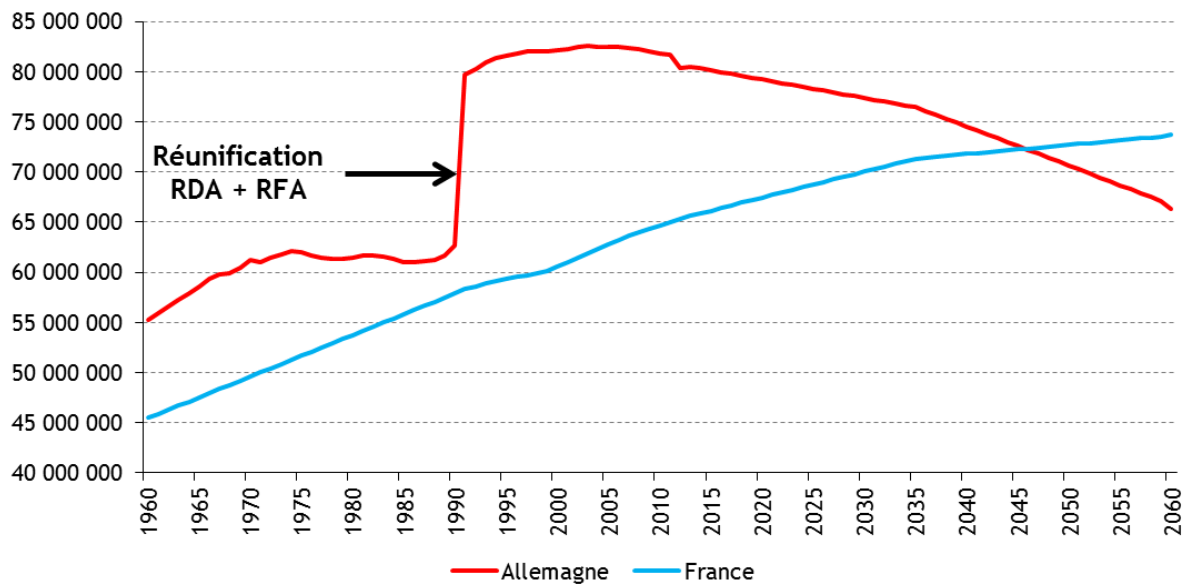
Source <https://www.insee.fr/fr/statistiques/1281283>

Cette baisse de la durée du travail s'explique par la réforme des 35 heures propulsée par Martine Aubry qui fut la ministre chargée des affaires sociales sous le gouvernement de Lionel Jospin. Cette loi fut proposée afin de créer des emplois.

Quelles sont les conséquences de la Loi d'Okun ?

Si nous voulons vraiment que le chômage baisse, nous devons avoir un fort taux de croissance, l'augmentation progressive de la productivité doit être stoppée afin que plus de personnes puissent participer à la production nationale. La réduction du temps de travail a permis de faire participer plus de travailleurs. Nous devons avoir un choc d'innovations afin de créer les emplois de demain. La réduction de la démographie semble être une bonne idée afin d'optimiser les investissements. De plus, il faudrait réussir à attirer et à conserver les élites de notre État-nation afin d'éviter qu'ils s'exilent à l'étranger pour travailler. La loi d'Okun reste méconnue par l'humanité, car s'orienter vers le plein-emploi, c'est également réduire le pouvoir des entreprises et des sociétés cotées. En effet, plus de chômage implique plus de compétition entre les chômeurs ce qui permet d'obtenir un maximum d'agilité et un maximum de productivité pour les possesseurs de capitaux et d'infrastructures de gouvernance. Le pouvoir, c'est le savoir. Le savoir, c'est l'emploi. Une politique malthusienne : réduire la population permet de s'orienter vers le plein-emploi, cependant nous en connaissons très bien les dérives inhumaines que cela peut engendrer. Un régime autoritaire ne peut pas être la solution dans un monde de plus en plus ouvert et connecté et qui s'enrichit des différences des résidents planétaires.

Population en Allemagne et en France



Le coefficient d'Okun en France.



<http://www.andlil.com/la-loi-dokun-6078.html#la-loi-dokun-et-lexemple-francais>

Alors que le coefficient d'Okun pour la période de 1970 à 1989 était de 0.19 en France, il est à ce jour de 0.57 actuellement. Le problème étant que l'évolution de la valeur de ce coefficient signifie que certes les entreprises embauchent plus en cas de hausse de la production, mais elles licencient plus rapidement en cas de difficulté.

Notre étude consistera dès maintenant à utiliser les outils économétriques afin de préciser, d'approfondir et d'étudier en termes mathématiques la loi d'Okun aux États-Unis d'Amérique.

Présentation des données

Dans cette première partie, nous allons présenter les données que nous utilisons dans cette étude. Nous présenterons leurs sources et définitions ainsi qu'une présentation graphique et statistique de ces séries temporelles.

Sources et définitions

Dans cette étude nous utilisons des données qui viennent du Federal Reserve Bank of St. Louis (FRED), leur site internet met de multiples bases de données à disposition. Nous avons utilisé des données trimestrielles du PIB et du taux de chômage de 1948q2 à 2016q3 pour les États-Unis. Nos données expriment le changement en pourcentage d'une période à la suivante ($\Delta\%$).

Le PIB, au sens du FRED, se définit comme étant : la valeur totale de la production interne de biens et services aux États-Unis au cours d'un trimestre par les agents résidents à l'intérieur du territoire national. Cette série temporelle mesure le PIB « réel » et ainsi il est déflaté par l'inflation.

Le chômage, au sens du FRED, se définit comme étant : Le taux de chômage représente le pourcentage de personnes sans emploi parmi la population active. La population active comprend les personnes âgées de plus de 16 résidant actuellement dans un des 50 états américains. La population active exclut les personnes détenues dans une institution judiciaire ou psychiatrique ainsi que les personnes ayant un statut militaire.

Nos données expriment le changement en pourcentage d'une période à la suivante du PIB en valeur nominale (milliards de dollars) et le changement en pourcentage d'une période à la suivante du taux de chômage.

Les sources pour nos données peuvent être consultées aux adresses suivantes :

- PIB : <https://fred.stlouisfed.org/graph/?g=c8fa>
- Chômage : <https://fred.stlouisfed.org/graph/?g=c8fd>

Exploration des données

Dans cette sous-partie, nous allons explorer nos données à travers la statistique descriptive et l'analyse graphique.

Chômage

```
. summarize $ylist, detail
```

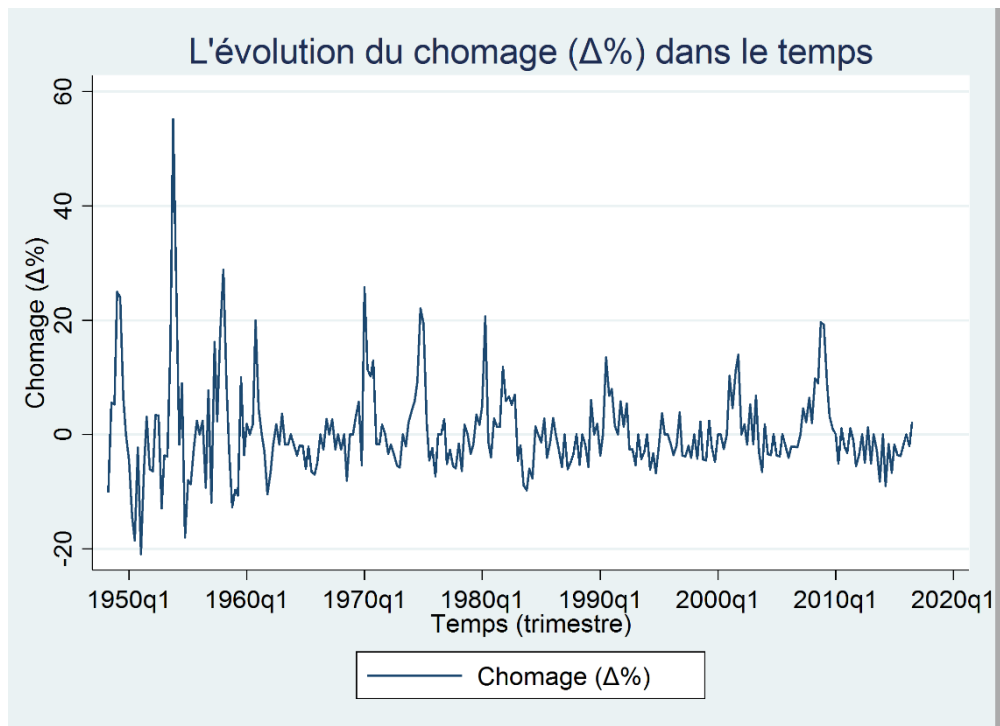
Chomage (Δ%)				
	Percentiles	Smallest		
1%	-18.03279	-20.93023		
5%	-8.95522	-18.51852		
10%	-6.45161	-18.03279	Obs	274
25%	-3.7037	-14.28571	Sum of Wgt.	274
50%	-.54945		Mean	.3694251
		Largest	Std. Dev.	7.937618
75%	2.43902	25.71429		
90%	8.95522	26.66667	Variance	63.00578
95%	16.21622	28.84615	Skewness	2.040738
99%	26.66667	55.17241	Kurtosis	12.39299

En utilisant la commande « summarize » de stata nous pouvons voir un résumé de la variable chômage. Nous pouvons voir que la distribution de cette variable n'est pas normale et que sa moyenne est proche de zéro (0.37). Nous pouvons voir également qu'il y a quelques valeurs extrêmes qui peuvent influencer notre droite de régression. Par exemple, la valeur de 55.17 fait largement plus de 3 écarts types.

```
rc) Stem-and-leaf plot for un (Chomage (Δ%))
^ un rounded to nearest multiple of .1
plot in units of .1

-2** | 09
-1** | 85,80
-1** |
-1** | 43
-1** | 29,27
-1** | 19,07,04,00
-0** | 98,97,93,90,89,87,82,81,80
-0** | 77,73,70,68,67,67,66,65,65,63,62,61,61,60,60
-0** | 59,59,58,57,57,56,56,55,54,54,53,51,51,51,50,49,48,48,46,45, ... (29)
-0** | 38,38,38,38,37,37,37,37,37,36,36,36,36,35,35,34,34,33,33,33, ... (49)
-0** | 19,19,19,19,18,18,18,18,18,18,18,18,18,17,17,17,17,17,16,16, ... (25)
0** | 00,00,00,00,00,00,00,00,00,00,00,00,00,00,00,00,00,00,00,00, ... (59)
0** | 20,20,21,22,22,23,23,24,24,24,26,26,27,28,28,29,29,31,32,33, ... (25)
0** | 41,45,45,47,50,52,53,53,54,56,57,58,59,59
0** | 60,64,65,67,68,68,69,77,79
0** | 89,89,90,92,93,98
1** | 00,02,03,11,14,18
1** | 30,35
1** | 40
1** | 60,62
1** | 82,92,94,97
2** | 00,06
2** | 20
2** | 40,50,57
2** | 67
2** | 88
2** |
```

Lorsque nous enlevons ces valeurs extrêmes, la distribution de la variable chômage ressemble plus à une distribution normale.



Voici un graphique montrant l'évolution de la variable chômage dans le temps. Nous pouvons constater qu'il semble varier autour d'une moyenne nulle et que sa variance reste plus ou moins constante dans le temps. De plus, nous pouvons identifier aucune tendance dans cette série temporelle.

PIB

```
. summarize $xlist, detail
```

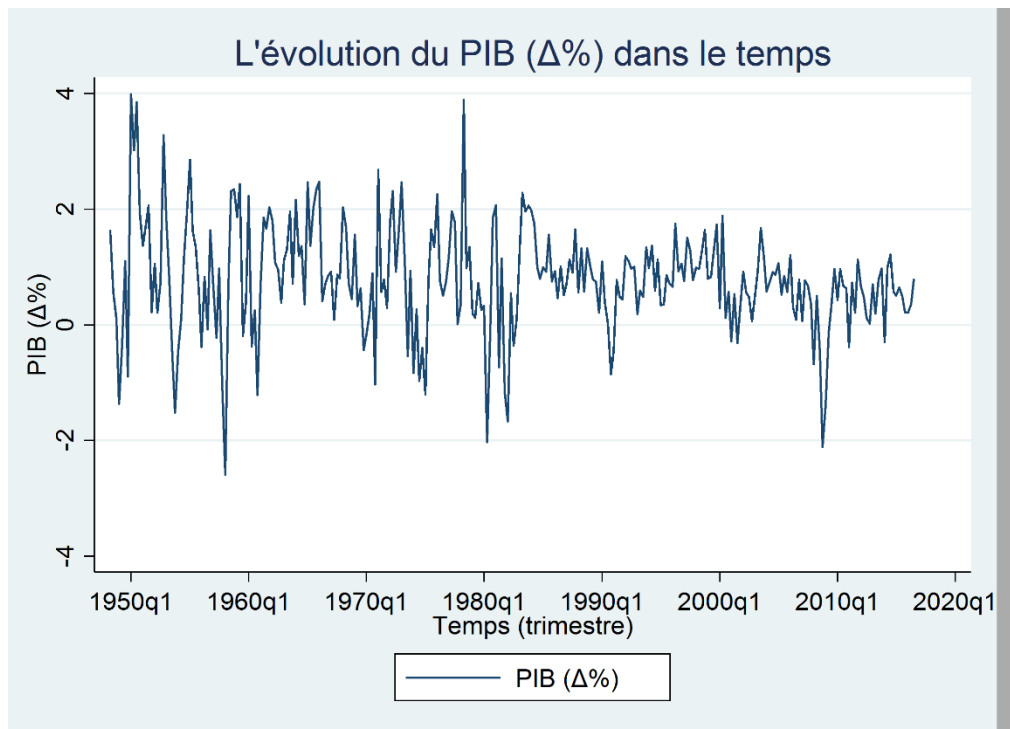
PIB ($\Delta\%$)				
Percentiles		Smallest		
1%	-2.02813	-2.59243		
5%	-.89702	-2.11298		
10%	-.37949	-2.02813	Obs	274
25%	.29882	-1.67277	Sum of Wgt.	274
50%	.76769		Mean	.784294
		Largest	Std. Dev.	.9574772
75%	1.30362	3.2819		
90%	1.95793	3.85408	Variance	.9167625
95%	2.31571	3.88804	Skewness	-.0286005
99%	3.85408	3.98603	Kurtosis	4.477229

Cette table présente les statistiques descriptives pour la variable PIB. Comme pour la variable précédente, le PIB n'a pas une distribution normale, mais il n'a pas de valeurs extrêmes qui pourraient surinfluencer la droite de régression.

```
rc Stem-and-leaf plot for pib (PIB ( $\Delta\%$ ))
^ pib rounded to nearest multiple of .01
  plot in units of .01

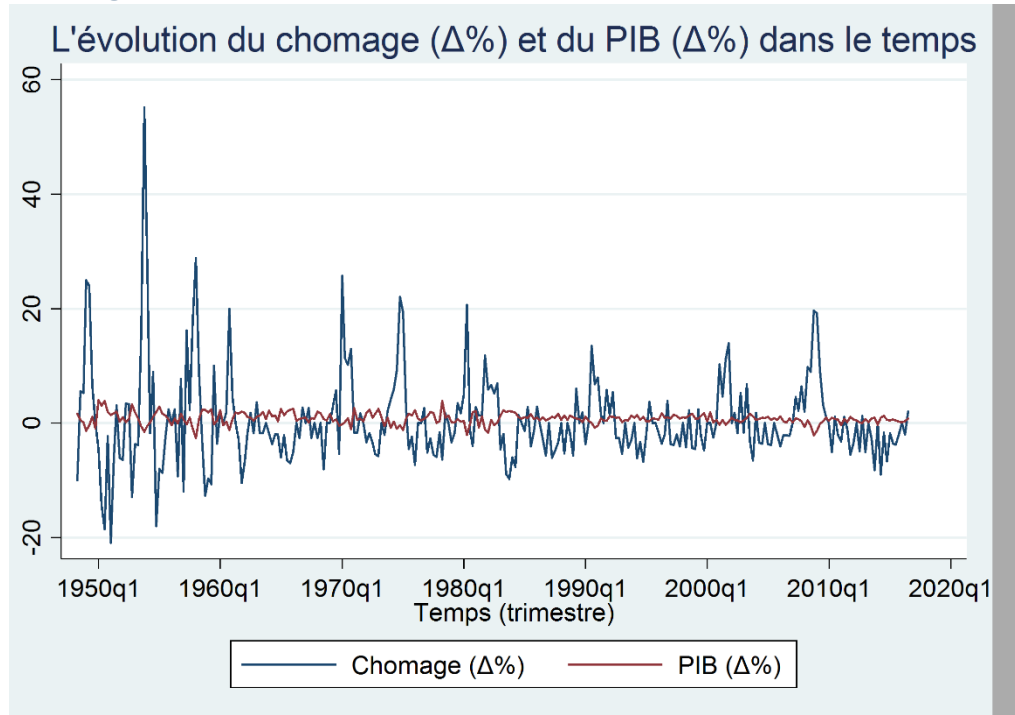
-2** | 59
-2** |
-2** | 11,03
-1** |
-1** | 67
-1** | 52
-1** | 39,37,21,21
-1** | 17,03,03
-0** | 97,90,85,83
-0** | 73,68
-0** | 56,54,48,47,46,44,40
-0** | 39,39,38,36,34,32,30,28,22
-0** | 19,18,15,14,08
0** | 01,02,02,06,06,09,09,10,10,11,12,12,12,18,19,19
0** | 20,21,21,21,21,22,22,25,26,26,28,29,29,30,32,32,33,34,35,35, ... (26)
0** | 41,43,44,45,46,47,48,49,49,49,50,51,51,52,52,52,53,54,55,56, ... (29)
0** | 60,63,63,64,65,66,66,66,67,67,68,70,70,71,71,71,72,72,72,73, ... (36)
0** | 80,80,80,82,82,84,86,86,86,86,89,91,91,92,92,92,92,93,93,93, ... (34)
1** | 00,01,01,06,06,07,08,09,10,10,11,12,13,13,13,14,15,16,17,18,19
1** | 20,22,26,27,30,31,31,32,32,33,34,34,35,36,37,37,37
1** | 51,55,56
1** | 62,63,64,64,65,65,66,67,67,69,73,74,75,76,77,79
1** | 80,85,86,86,86,89,93,95,95,96,96,99
2** | 03,03,03,06,06,07,16
2** | 23,26,28,31,32,34,36
2** | 43,46,47,47
2** | 68
2** | 85
3** | 02
3** | 28
3** |
3** |
3** | 85,89,99
```

Voici la distribution de la variable PIB, nous pouvons constater que sa distribution n'est pas complètement normale, mais qu'elle ressemble.

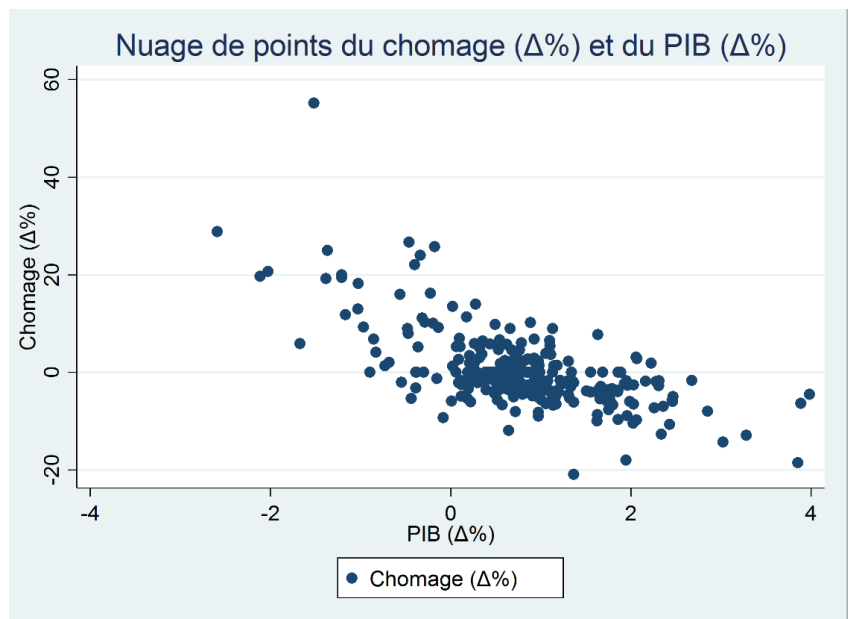


Voici un graphique montrant l'évolution du PIB dans le temps. La série semble varier autour d'une moyenne stable (0.78), sa variance ne croit pas avec le temps, et nous ne pouvons pas identifier une tendance.

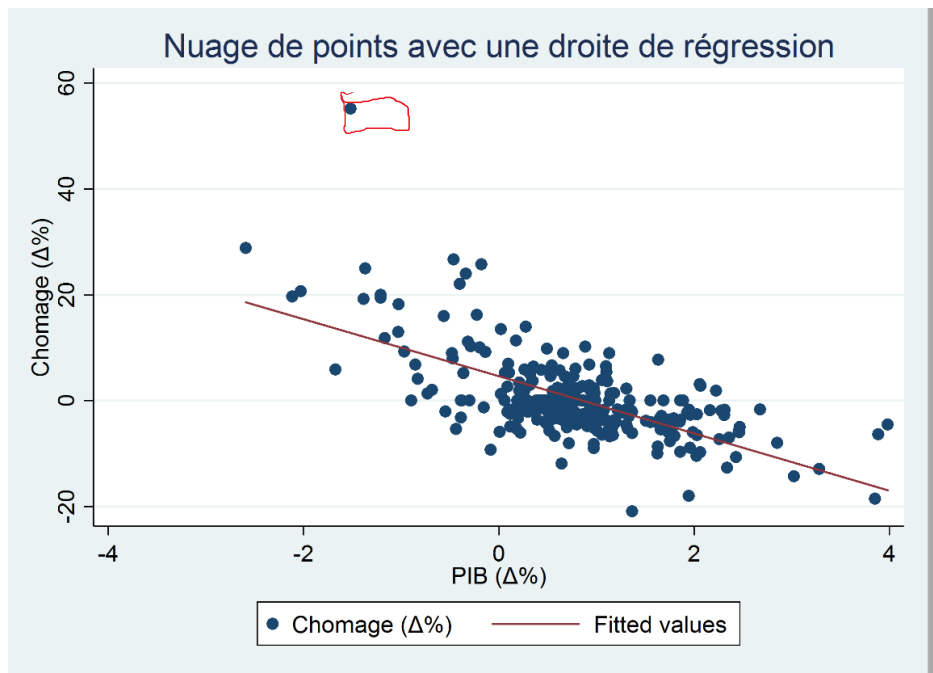
Chômage et PIB ensemble



Dans ce graphique nous voyons l'évolution des deux séries dans le temps. Nous pouvons visuellement identifier une relation inverse entre ces deux variables.



Cette relation inverse soupçonnée est confirmée par le nuage de points. Nous pouvons voir qu'il existe bien une relation inverse entre le PIB($\Delta\%$) et le taux de chômage($\Delta\%$). Un modèle non linéaire (une courbe) semble être la meilleure représentation de cette relation, cependant un modèle linéaire (une droite) serait adéquat.



Voici le nuage de points avec une droite de régression (Chômage~PIB), nous pouvons voir que ce modèle linéaire est une bonne représentation de la relation entre le PIB(Δ%) et le taux de chômage(Δ%). La seule valeur qui semble être problématique est entourée en rouge, c'est la valeur extrême que nous avons identifiée lors de notre exploration de la variable chômage (55.17). Cette valeur extrême à part, notre droite de régression est une excellente représentation de ces deux variables pour la période donnée.

```
. correlate un pib
(obs=274)
```

	un	pib
un	1.0000	
pib	-0.6514	1.0000

Ici, nous confirmons numériquement la relation inverse entre ces deux variables. Ainsi, nous avons une confirmation graphique et statistique de cette relation inverse.

I : Stationnarisation des données

Avant le traitement d'une série chronologique, il convient d'en étudier les caractéristiques stochastiques. Si ces caractéristiques – c'est-à-dire son espérance et sa variance – se trouvent modifiées dans le temps, la série chronologique est considérée comme non stationnaire ; dans le cas d'un processus stochastique invariant, la série temporelle est alors stationnaire. De manière formalisée, le processus stochastique y_t est stationnaire si :

- $E(y_t) = E(y_{t+m}) = \mu \quad \forall t \text{ et } \forall m$, la moyenne est constante et indépendante du temps ;
- $\text{var}(y_t) < \infty \quad \forall t$, la variance est finie et indépendante du temps ;
- $\text{cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k$, la covariance est indépendante du temps.

Une série chronologique est donc stationnaire si elle est la réalisation d'un processus stationnaire. Ceci implique que la série ne comporte ni tendance, ni saisonnalité et plus généralement aucun facteur évoluant avec le temps.

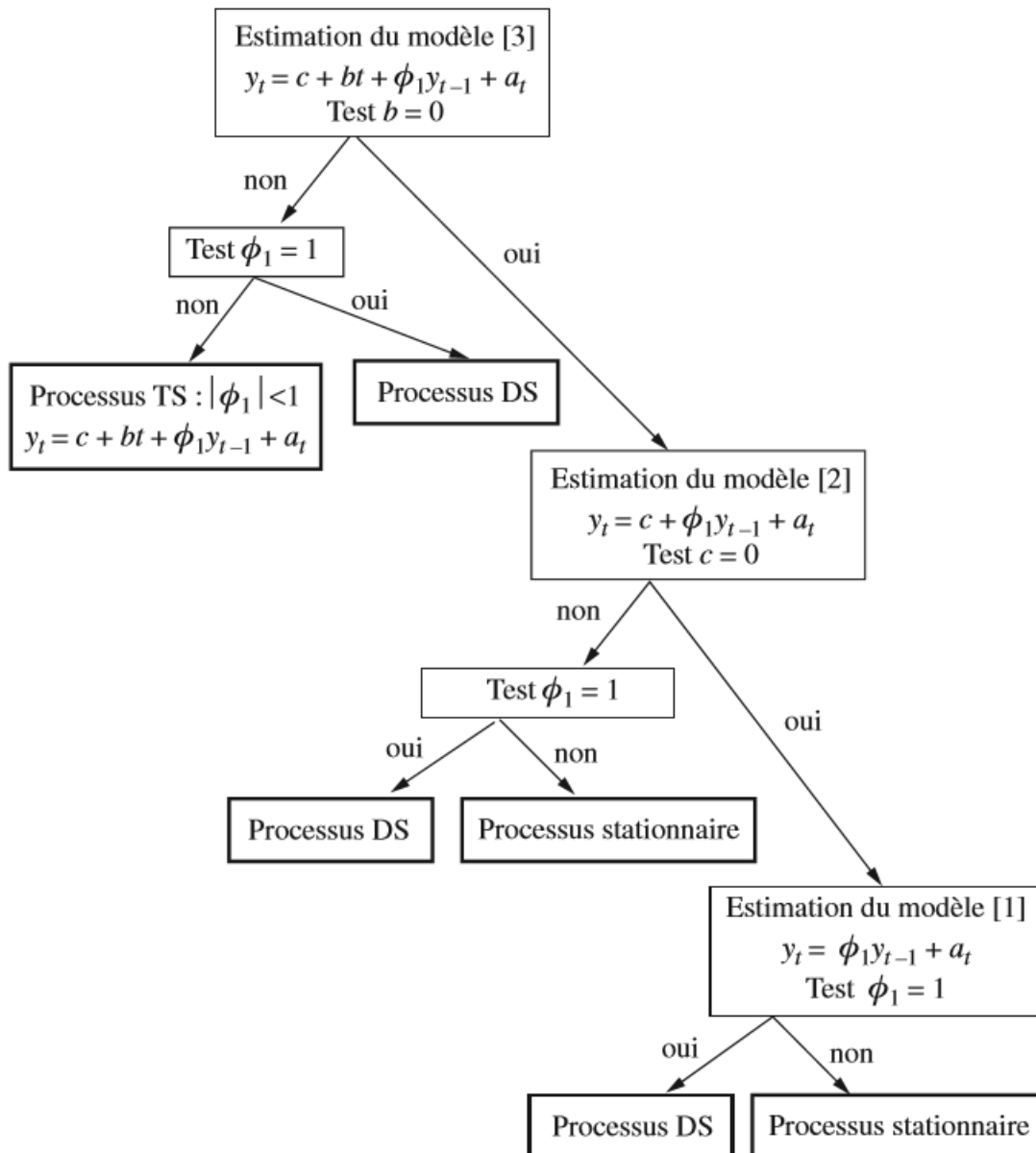
Cette étude de stationnarité s'effectue essentiellement à partir de l'étude des fonctions d'autocorrélation (ou de leur représentation graphique appelée « corrélogramme ») et par tests numériques comme le test Dickey-Fuller.

Pour analyser la non-stationnarité, deux types de processus sont distingués :

- les processus TS (Trend Stationary) qui représentent une non-stationnarité de type déterministe. Pour un processus TS, la bonne méthode de stationnarisation est celle des moindres carrés ordinaires ($y_t \sim t$).
- Les processus DS (Difference Stationary) pour les processus non stationnaires aléatoires. Pour un processus DS, la bonne méthode de stationnarisation est le filtre aux différences premières.

Le test Dickey-Fuller est un test de racine unitaire qui nous permet d'identifier la stationnarité d'une série. La stratégie de test est la suivante : nous testons d'abord le modèle M3 et si la tendance n'est pas significative (< 0.05) nous passons au modèle M2. Dans le cadre du modèle M2 si la constante n'est pas significative nous passons au modèle M1. Dans le modèle M1, nous regardons directement la statistique de test pour déterminer la stationnarité de la série chronologique.

Stratégie de test Dickey-Fuller



[M3] modèle autorégressif avec tendance.

[M2] modèle autorégressif avec constante.

[M1] modèle autorégressif d'ordre 1.

H0 du test Dickey-Fuller est la présence d'une racine unitaire

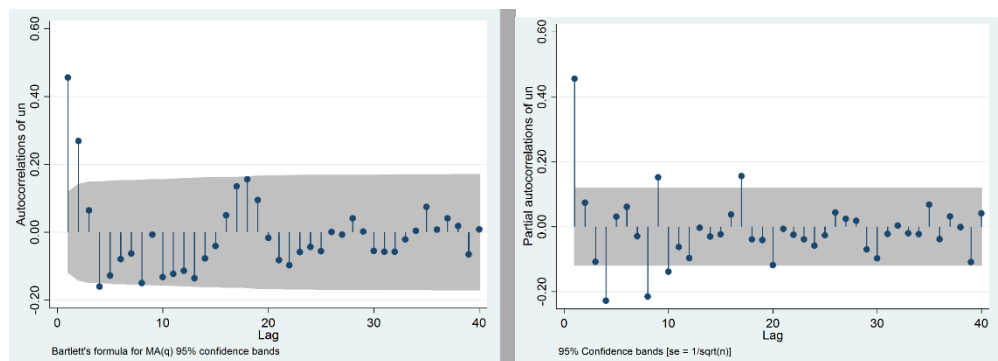
Maintenant nous allons tester la stationnarité de nos séries avec l'aide du corrélogramme et le test Dickey-Fuller.

Chômage

Cette partie concerne nos tests sur notre variable chômage (qui s'appelle « un » dans stata).

```
. corrgram un
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.4557	0.4558	57.529	0.0000						
2	0.2689	0.0739	77.638	0.0000						
3	0.0642	-0.1086	78.789	0.0000						
4	-0.1603	-0.2282	85.983	0.0000						
5	-0.1282	0.0309	90.605	0.0000						
6	-0.0792	0.0610	92.377	0.0000						
7	-0.0634	-0.0297	93.514	0.0000						
8	-0.1501	-0.2154	99.916	0.0000						
9	-0.0073	0.1516	99.931	0.0000						
10	-0.1328	-0.1387	104.98	0.0000						
11	-0.1227	-0.0629	109.31	0.0000						
12	-0.1138	-0.0971	113.05	0.0000						
13	-0.1356	-0.0037	118.38	0.0000						
14	-0.0776	-0.0307	120.13	0.0000						
15	-0.0407	-0.0242	120.62	0.0000						
16	0.0498	0.0374	121.34	0.0000						
17	0.1350	0.1558	126.71	0.0000						
18	0.1559	-0.0391	133.89	0.0000						
19	0.0947	-0.0417	136.55	0.0000						
20	-0.0168	-0.1188	136.63	0.0000						
21	-0.0827	-0.0070	138.68	0.0000						
22	-0.0972	-0.0250	141.51	0.0000						
23	-0.0586	-0.0392	142.55	0.0000						
24	-0.0435	-0.0590	143.12	0.0000						
25	-0.0557	-0.0263	144.06	0.0000						
26	0.0006	0.0430	144.06	0.0000						
27	-0.0073	0.0236	144.08	0.0000						
28	0.0411	0.0175	144.6	0.0000						
29	0.0019	-0.0706	144.6	0.0000						
30	-0.0551	-0.0978	145.54	0.0000						
31	-0.0579	-0.0222	146.59	0.0000						
32	-0.0580	0.0031	147.64	0.0000						
33	-0.0213	-0.0203	147.78	0.0000						
34	0.0040	-0.0235	147.78	0.0000						
35	0.0747	0.0679	149.55	0.0000						
36	0.0079	-0.0386	149.57	0.0000						
37	0.0410	0.0316	150.1	0.0000						
38	0.0181	-0.0015	150.21	0.0000						
39	-0.0658	-0.1090	151.6	0.0000						
40	0.0085	0.0404	151.62	0.0000						



Voici le corrélogramme et les graphiques d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) pour la variable chômage. Nous pouvons voir que l'autocorrélation chute rapidement en zéro, ce qui indique la stationnarité.

Nous allons passer au test Dickey-Fuller, mais nous devons estimer le nombre de lags (retards) à inclure dans le test. Afin de déterminer ceci, nous allons utiliser la commande « varsoc VARIABLE, lutstat », cette commande de stata calcule le nombre de lags à inclure.

```
. varsoc un, lutstat
```

Selection-order criteria (lutstats)
Sample: 1949q2 - 2016q3 Number of obs = 270

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-937.781				61.3168	4.10865	4.10865	4.10865
1	-905.222	65.117	1	0.000	48.5351	3.87488	3.88023	3.88821
2	-904.486	1.4733	1	0.225	48.6299	3.87683	3.88753	3.90349
3	-903.384	2.2024	1	0.138	48.5935	3.87608	3.89214	3.91606
4	-895.808	15.153*	1	0.000	46.2832*	3.82737*	3.84877*	3.88068*

Endogenous: un
Exogenous: _cons

Le test varsoc nous indique qu'il faut inclure 4 lags dans notre test Dickey-Fuller.

Nous commençons notre stratégie de test par le modèle M3, nous essayons de voir si notre série contient une tendance déterministe.

```
. dfuller un, lags(4) trend regress
```

Augmented Dickey-Fuller test for unit root Number of obs = 269

Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
Z(t)	-7.786	-3.989	-3.429

MacKinnon approximate p-value for Z(t) = 0.0000

D.un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
un					
L1.	-.6666241	.0856138	-7.79	0.000	-.8352027 -.4980454
LD.	.0587834	.0772882	0.76	0.448	-.0934017 .2109684
L2D.	.201246	.0722984	2.78	0.006	.0588862 .3436059
L3D.	.2020193	.0685152	2.95	0.003	.0671087 .3369299
L4D.	-.0303732	.0600542	-0.51	0.613	-.1486235 .087877
_trend	-.0012672	.0053176	-0.24	0.812	-.011738 .0092035
_cons	.2925023	.8488843	0.34	0.731	-1.379002 1.964006

Nous sommes dans le cadre du modèle M3. Nous pouvons voir qu'avec une P value de 0.812 la tendance (_trend) n'est pas significative et nous pouvons passer au modèle M2.

```
. dfuller un, lags(4) regress
```

Augmented Dickey-Fuller test for unit root Number of obs = 269

Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
Z(t)	-7.812	-3.458	-2.879

MacKinnon approximate p-value for Z(t) = 0.0000

D.un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
un					
L1.	-.6647266	.0850897	-7.81	0.000	-.8322703 -.4971828
LD.	.0572907	.0768957	0.75	0.457	-.0941188 .2087002
L2D.	.2001062	.0720105	2.78	0.006	.0583157 .3418967
L3D.	.2011259	.0682898	2.95	0.004	.0666615 .3355902
L4D.	-.0308848	.0599081	-0.52	0.607	-.1488453 .0870756
_cons	.1156549	.4114362	0.28	0.779	-.6944731 .925783

Ici nous pouvons voir qu'avec une P value de 0.779 la constante (_cons) n'est pas significative et nous pouvons passer au modèle M1

. dfuller un, lags(4) nocons regress					
Augmented Dickey-Fuller test for unit root			Number of obs = 269		
Test Statistic		1% Critical Value	Interpolated Dickey-Fuller		10% Critical Value
			5% Critical Value		
Z(t)		-7.826	-1.950		-1.620
D.un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
un					
L1.	-.6629802	.0847144	-7.83	0.000	-.8297822 -.4961783
LD.	.0559411	.0766117	0.73	0.466	-.0949065 .2067888
L2D.	.1990531	.0717875	2.77	0.006	.0577043 .340402
L3D.	.2002451	.0680988	2.94	0.004	.0661592 .334331
L4D.	-.0314597	.0597686	-0.53	0.599	-.1491435 .0862241

Ici nous regardons directement la statistique de test qui vaut -7.826 en l'espèce. La statistique de test est inférieure à la valeur tabulée (-1.95 au seuil de 5%) et donc nous pouvons rejeter H0. Ainsi il n'y a pas de racine unitaire et notre série est stationnaire.

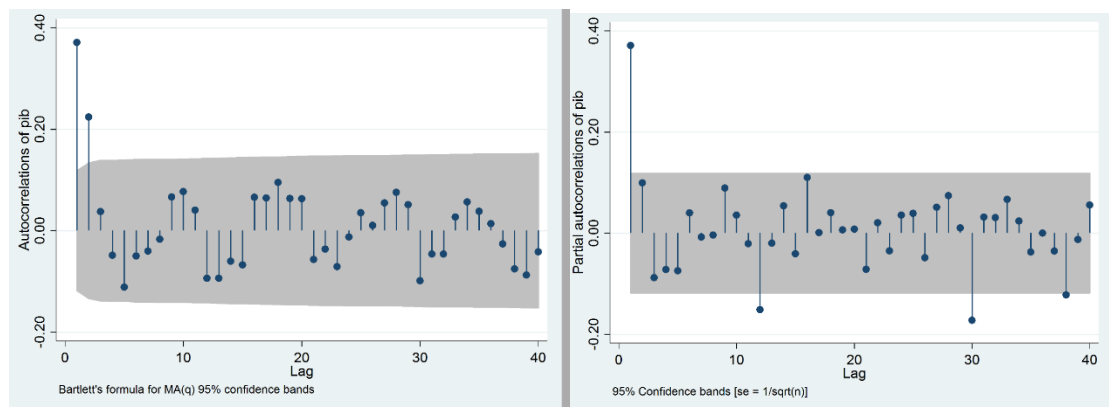
Chômage~I(0)

PIB

Cette partie concerne nos tests sur notre variable PIB (qui s'appelle « PIB » dans stata).

```
. corrgram pib
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.3713	0.3713	38.194	0.0000						
2	0.2241	0.0993	52.151	0.0000						
3	0.0378	-0.0880	52.55	0.0000						
4	-0.0485	-0.0719	53.21	0.0000						
5	-0.1115	-0.0743	56.705	0.0000						
6	-0.0501	0.0400	57.414	0.0000						
7	-0.0403	-0.0076	57.874	0.0000						
8	-0.0168	-0.0038	57.954	0.0000						
9	0.0662	0.0894	59.204	0.0000						
10	0.0772	0.0356	60.911	0.0000						
11	0.0404	-0.0209	61.381	0.0000						
12	-0.0941	-0.1512	63.937	0.0000						
13	-0.0939	-0.0197	66.49	0.0000						
14	-0.0601	0.0540	67.541	0.0000						
15	-0.0674	-0.0408	68.865	0.0000						
16	0.0657	0.1102	70.131	0.0000						
17	0.0644	0.0012	71.353	0.0000						
18	0.0952	0.0404	74.032	0.0000						
19	0.0636	0.0063	75.231	0.0000						
20	0.0630	0.0079	76.411	0.0000						
21	-0.0569	-0.0713	77.379	0.0000						
22	-0.0363	0.0204	77.775	0.0000						
23	-0.0708	-0.0349	79.286	0.0000						
24	-0.0125	0.0355	79.334	0.0000						
25	0.0352	0.0393	79.709	0.0000						
26	0.0102	-0.0488	79.741	0.0000						
27	0.0546	0.0513	80.655	0.0000						
28	0.0759	0.0743	82.424	0.0000						
29	0.0512	0.0102	83.234	0.0000						
30	-0.0987	-0.1722	86.252	0.0000						
31	-0.0461	0.0318	86.915	0.0000						
32	-0.0464	0.0310	87.587	0.0000						
33	0.0271	0.0671	87.817	0.0000						
34	0.0569	0.0239	88.836	0.0000						
35	0.0379	-0.0375	89.291	0.0000						
36	0.0136	0.0004	89.35	0.0000						
37	-0.0263	-0.0356	89.57	0.0000						
38	-0.0753	-0.1223	91.389	0.0000						
39	-0.0878	-0.0127	93.868	0.0000						
40	-0.0418	0.0555	94.432	0.0000						



Voici le corrélogramme et le graphique d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) pour la variable PIB. Nous pouvons voir que l'autocorrélation chute rapidement en zéro, ce qui indique la stationnarité.

```
. varsoc pib, lutstat
```

Selection-order criteria (lutstats)
Sample: 1949q2 - 2016q3 Number of obs = 270

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-369.675				.911983	-.099541	-.099541	-.099541
1	-349.695	39.96*	1	0.000	.79237	-.240134	-.234783*	-.226807*
2	-348.243	2.9044	1	0.088	.789721*	-.243484*	-.23278	-.216829
3	-347.37	1.7469	1	0.186	.790463	-.242547	-.226491	-.202564
4	-346.655	1.4289	1	0.232	.792138	-.240431	-.219024	-.187121

Endogenous: pib
Exogenous: _cons

La commande « lutstat » nous indique qu’il faut inclure 1 lag dans notre test Dickey-Fuller.

```
. dfuller pib, lags(1) trend regress
```

Augmented Dickey-Fuller test for unit root Number of obs = 272

Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
Z (t)	-8.549	-3.989	-3.429

MacKinnon approximate p-value for Z(t) = 0.0000

D.pib	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pib					
L1.	-.5918415	.0692254	-8.55	0.000	-.7281363 -.4555468
LD.	-.0854567	.06077	-1.41	0.161	-.2051041 .0341907
_trend	-.0012877	.0006998	-1.84	0.067	-.0026656 .0000901
_cons	.6398345	.1316037	4.86	0.000	.3807258 .8989432

Nous sommes dans le cadre du modèle M3. Nous pouvons voir qu’avec une P value de 0.067 la tendance (_trend) n’est pas significative et nous pouvons passer au modèle M2

```
. dfuller pib, lags(1) regress
```

Augmented Dickey-Fuller test for unit root Number of obs = 272

Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
Z (t)	-8.312	-3.458	-2.879

MacKinnon approximate p-value for Z(t) = 0.0000

D.pib	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pib					
L1.	-.5645434	.0679163	-8.31	0.000	-.6982584 -.4308284
LD.	-.0993115	.0605686	-1.64	0.102	-.2185603 .0199374
_cons	.4413813	.0757516	5.83	0.000	.2922399 .5905228

Nous sommes dans le cadre du modèle M2. Ici nous pouvons voir qu’avec une P value de 0 la constante (_cons) est significative et nous pouvons donc interpréter la statistique de test. En l’espèce la statistique de test (-8.312) est inférieure à la valeur tabulée au seuil de 5% (-2.879) et ainsi nous rejetons H0. Notre série n’a pas de racine unitaire et elle est donc stationnaire.

PIB~I(0)

Ainsi, nos deux séries sont stationnaires en niveaux.

La cointégration des variables

La cointégration est une propriété statistique des séries temporelles introduite dans l'analyse économique, notamment par Engle et Newbold (1974). En des termes simples, la cointégration permet de détecter la relation de long terme entre deux ou plusieurs séries temporelles. Sa formalisation rigoureuse est due à Granger (1981), Engle et Granger (1987) et Johansen (1991, 1995). Techniquement, la notion de cointégration implique implicitement celle d'intégration.

Normalement nous détectons la cointégration avec le test de Granger en deux étapes. Il faut faire une régression des deux variables en niveaux ($Y_t \sim X_t$) puis on récupère les résidus de cette régression. Si ces résidus sont stationnaires, nous disons que les deux séries sont cointégrées. La cointégration confirme l'existence d'une relation à long terme entre les deux variables, ensuite il faut utiliser un modèle à correction d'erreur pour confirmer l'existence d'une relation à court terme et pour estimer la force de rappel.

En l'espèce, nos deux séries sont stationnaires en niveau ($I(0)$) ce qui exclue la possibilité de la cointégration. Ainsi, nous pouvons estimer les coefficients des paramètres pour notre modèle linéaire avec une régression simple par MCO.

Identification de sous-périodes (ruptures structurelles)

L'idée ici est de voir s'il existe des sous-périodes dans le temps où les coefficients changent. Peut-on considérer le modèle comme étant stable sur la totalité de la période, ou bien doit-on considérer des sous-périodes distinctes d'estimation (changement structurel du modèle) ? La spécification du modèle est la même, mais les valeurs estimées des coefficients pour les échantillons sont différentes.

Nous avons décidé d'utiliser un test de Wald pour identifier nos points de rupture car nous n'avons pas réussi à identifier des points de rupture graphiquement possible à tester avec le test de Chow. Ainsi, nous allons utiliser le test de Wald pour identifier les points de rupture puis nous allons confirmer ces points de ruptures avec la fonction CUSUM et avec le test de Chow. Le test de Wald présente plusieurs avantages pour identifier les points de rupture, il peut les identifier même avec une date inconnue et il est robuste à l'hétéroscédasticité. Le test de Wald s'exécute avec la commande « sbsingle » dans stata.

```
Test for a structural break: Unknown break date

                                Number of obs =      274

Full sample:                    1948q2 - 2016q3
Trimmed sample:                 1958q4 - 2006q2
Estimated break date:          1959q4
Ho: No structural break

+-----+-----+-----+
| Test      | Statistic | p-value |
+-----+-----+-----+
| swald     | 24.6782  | 0.0001  |
+-----+-----+-----+

Exogenous variables:           pib
Coefficients included in test: pib_cons
```

Nous faisons une régression sur toute la période puis on exécute la commande « sbsingle ». Pour ce test, H_0 est l'absence d'une rupture structurelle et ainsi H_a est la présence d'une rupture structurelle.

En l'espèce, nous avons une P value de 0.0001 qui est largement inférieure au seuil de 0.05 et nous pouvons donc rejeter H_0 . Ainsi, nous identifions une rupture structurelle en 1959q4

```

Test for a structural break: Unknown break date

                                Number of obs =          228

Full sample:                    1959q4 - 2016q3
Trimmed sample:                 1968q3 - 2008q1
Estimated break date:          2003q4
Ho: No structural break

Test      Statistic      p-value
-----
swald      15.4910      0.0091

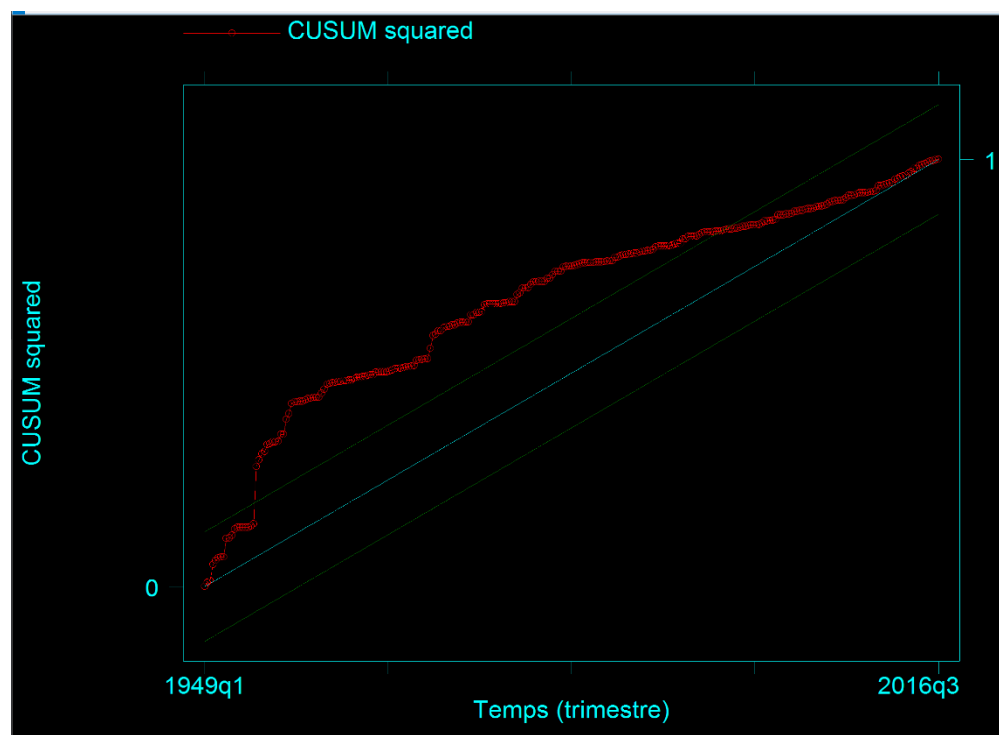
Exogenous variables:          pib
Coefficients included in test: pib _cons

```

Nous avons déjà identifié une sous-période allant de 1948q2 à 1959q3, ainsi nous allons faire une régression sur la 2e période (à partir de 1959q4) pour voir s'il existe une autre rupture structurelle. Le test de Wald nous indique avec une P value de 0.0091 (rejet de H_0) qu'il existe un point de rupture en 2003q4.

Ainsi, nous avons identifié 3 sous-périodes distinctes à modéliser. La spécification du modèle reste inchangée, mais les coefficients sur chaque paramètre estimé changent pour chaque période. Nos trois sous-périodes sont les suivantes :

- Identification d'une période de 1948q2 à 1959q3
- Identification d'une période de 1959q4 à 2003q3
- Identification d'une période de 2003q4 à 2016q3



Nous pouvons voir le CUSUM² confirme plus ou moins les dates de rupture proposées par le test de Wald. Quand la ligne rouge sort des intervalles, celui-ci indique une rupture structurelle éventuelle et donc le changement des coefficients des paramètres.

Nous allons désormais faire un dernier test pour confirmer nos points de rupture que le test Wald a identifié. Nous allons vérifier ces points de rupture avec le test de Chow. Le test de Chow calcule une statistique F que nous comparons ensuite à une table de valeurs critiques en fonction de k et de n.

À titre d'information, voici la formule pour la statistique de Chow :

Statistique de Chow : $F^* = [(scr - (scr1 + scr2)) / (scr1 + scr2)] * n - 2k / k$

Si $F^* > F(k, n, 5\%)$ nous rejetons H_0 et nous confirmons donc la présence d'une rupture structurelle.

Les détails et modalités du calcul de la statistique de Chow sont dans le do-file.

- F^* pour le 1^{er} point de rupture : 10.236869
- F tabulé (2, 112) : 1.850

En l'espèce $F^* > F$ et ainsi nous rejetons H_0 . Donc nous confirmons l'existence d'une rupture structurelle en 1959q3

Les détails et modalités du calcul de la statistique de Chow sont dans le do-file

- F^* pour le 2^{ème} point de rupture : 9.3360766
- F tabulé (2, 135) : 1.817

En l'espèce $F^* > F$ et ainsi nous rejetons H_0 . Donc nous confirmons l'existence d'une rupture structurelle en 2003q4

Régression de la période 1 (1948q2 à 1959q3)

Nous essayons de modéliser le modèle suivant :

$y_t = a_0 + a_1 x_t + \varepsilon_t$ pour $t = 1, \dots, n$

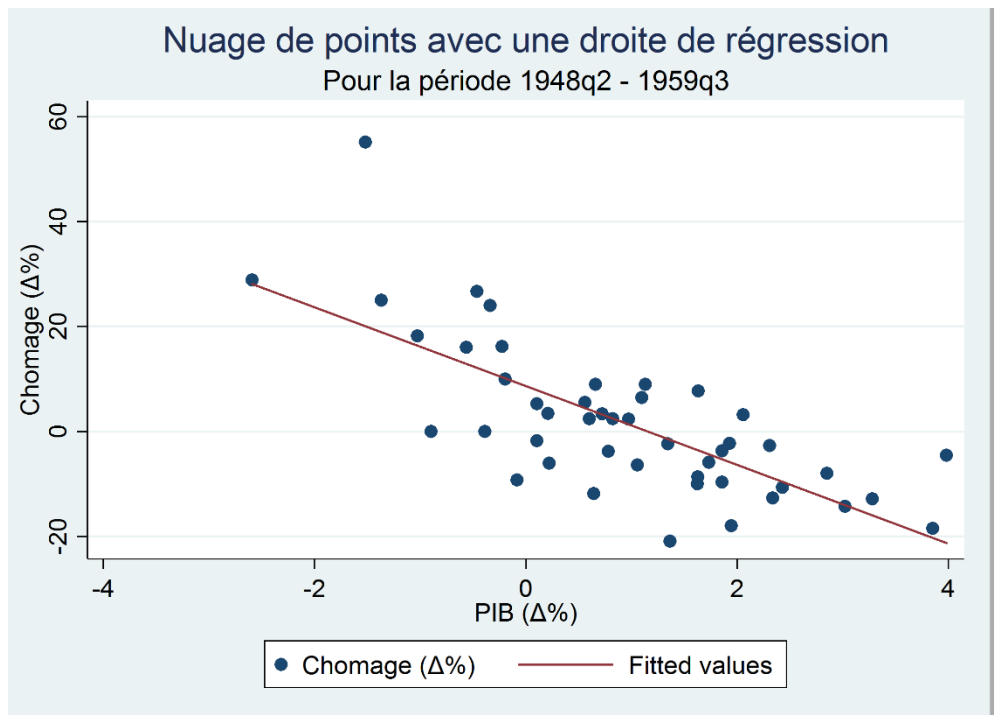
Avec :

- y_t = variable à expliquer au temps t, $\Delta\%$ Chomage en l'espèce ;
- x_t = variable explicative au temps t, $\Delta\%$ PIB en l'espèce ;
- a_0, a_1 = paramètres du modèle, où a_0 représente le terme constant (l'intercepte) et a_1 représente le coefficient sur $\Delta\%$ PIB ;
- ε_t = erreur de spécification (différence entre le modèle vrai et le modèle spécifié), cette erreur est inconnue et restera inconnue ;
- n = nombre d'observations.

Pour que notre modèle soit valide, il doit respecter les hypothèses suivantes :

- H_1 : le modèle est linéaire en x_t (ou en n'importe quelle transformation de x_t).
- H_2 : les valeurs x_t sont observées sans erreur (x_t non aléatoire).

- H3 : $E(\varepsilon_t) = 0$, l'espérance mathématique de l'erreur est nulle : en moyenne le modèle est bien spécifié et donc l'erreur moyenne est nulle.
- H4 : $E(\varepsilon_t^2) = \sigma^2$, la variance de l'erreur est constante: le risque de l'amplitude de l'erreur est le même quelle que soit la période.
- H5 : $E(\varepsilon_t \varepsilon_{t'}) = 0$ si $t \neq t'$, les erreurs sont non corrélées (ou encore indépendantes) : une erreur à l'instant t n'a pas d'influence sur les erreurs suivantes.
- H6 : $\text{Cov}(x_t, \varepsilon_t) = 0$, l'erreur est indépendante de la variable explicative.



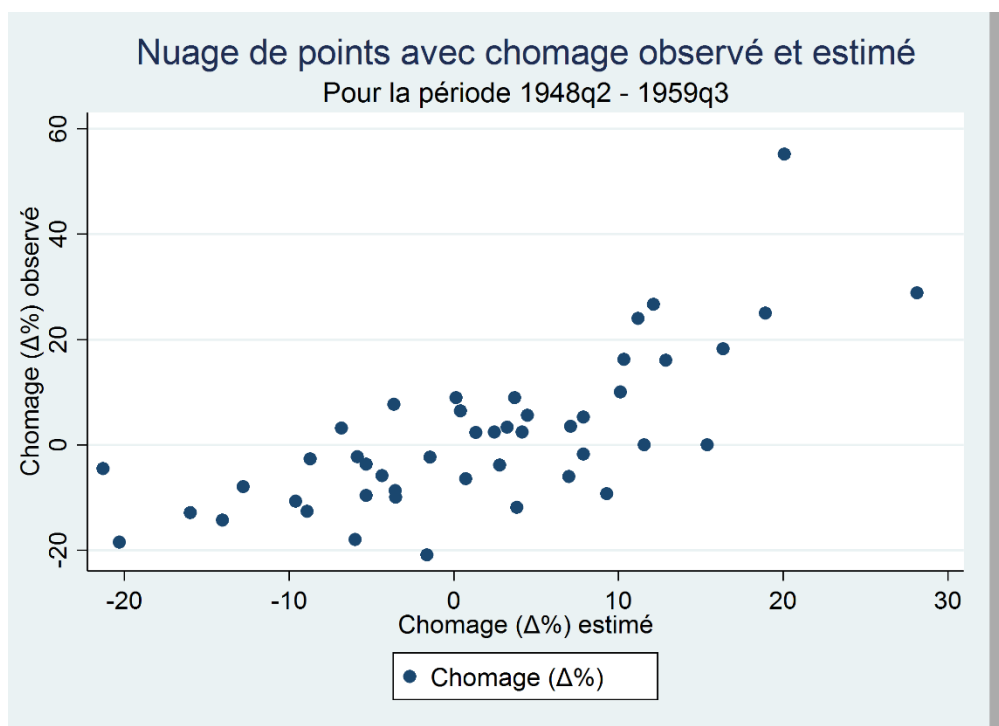
Voici un nuage de points avec chômage(Δ%) en y et PIB(Δ%) en x pour la 1re période, nous pouvons constater qu'il existe une relation négative entre ces deux variables qui est bien représentée par un modèle linéaire.

Table ANOVA

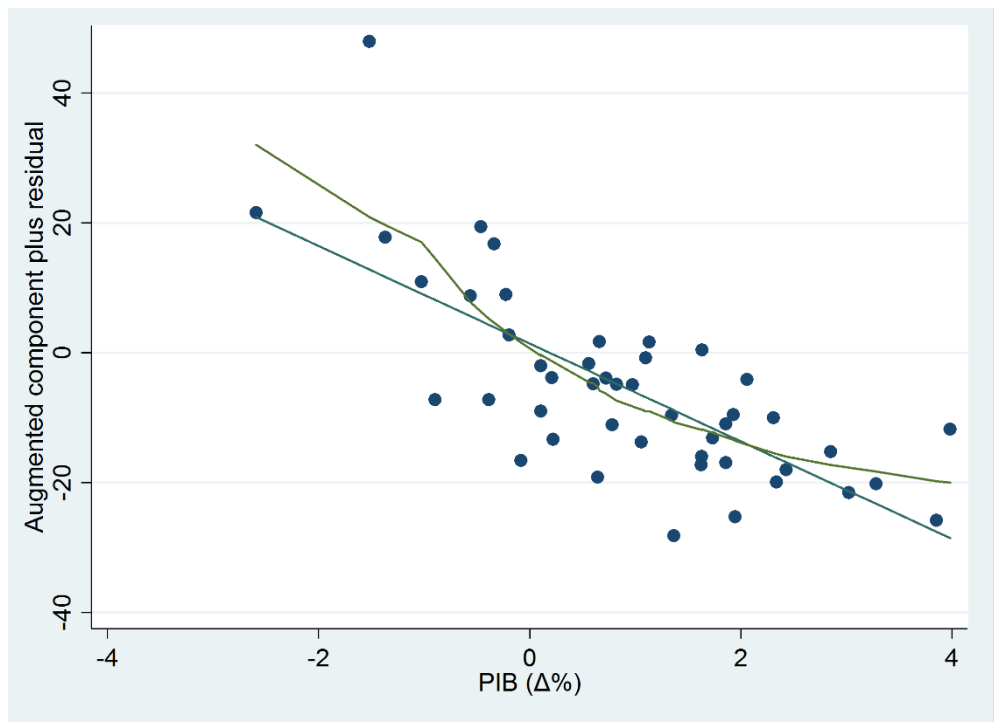
. reg un pib in 1/46						
Source	SS	df	MS	Number of obs	=	46
Model	5083.14322	1	5083.14322	F(1, 44)	=	50.84
Residual	4399.67402	44	99.9925914	Prob > F	=	0.0000
				R-squared	=	0.5360
				Adj R-squared	=	0.5255
Total	9482.81724	45	210.729272	Root MSE	=	9.9996
un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pib	-7.512546	1.053671	-7.13	0.000	-9.63608	-5.389013
_cons	8.669796	1.774911	4.88	0.000	5.092698	12.24689

Voici l'output de notre régression pour la période. Cette table contient plusieurs informations essentielles.

- La statistique F et $\text{prob} > F$ indiquent la significativité statistique du modèle. En l'espèce, la P value de notre statistique F est 0, et ainsi nous pouvons conclure que notre modèle est statistiquement significatif.
- Le R^2 et le R^2 ajusté indiquent la variation de chômage qui est expliquée par notre modèle. En l'espèce, environ 53% de la variation du taux de chômage est expliqué par la variation du PIB.
- Cette table indique également la significativité et les valeurs des coefficients.
- En l'espèce, la variable PIB a une P value de 0 et elle est donc significative. Son coefficient est de -7.51 ce qui confirme la relation inverse entre le chômage et le PIB. De plus, en vue des intervalles de confiance (-9.63 à -5.39) nous pouvons constater que le coefficient est significativement différent de zéro.
- Ce modèle contient également une constante. En l'espèce la constante est significative avec une P value de 0, sa valeur est de 8.67. Cette constante représente l'intercepte. Ainsi, lorsque $x_t=0$, $y_t=k$
- Number of obs indique le nombre d'observations pour la régression. En l'espèce notre « n » vaut 46 et donc la taille de notre échantillon est supérieure à 35 ce qui implique qu'elle est suffisamment grande pour tirer des conclusions.

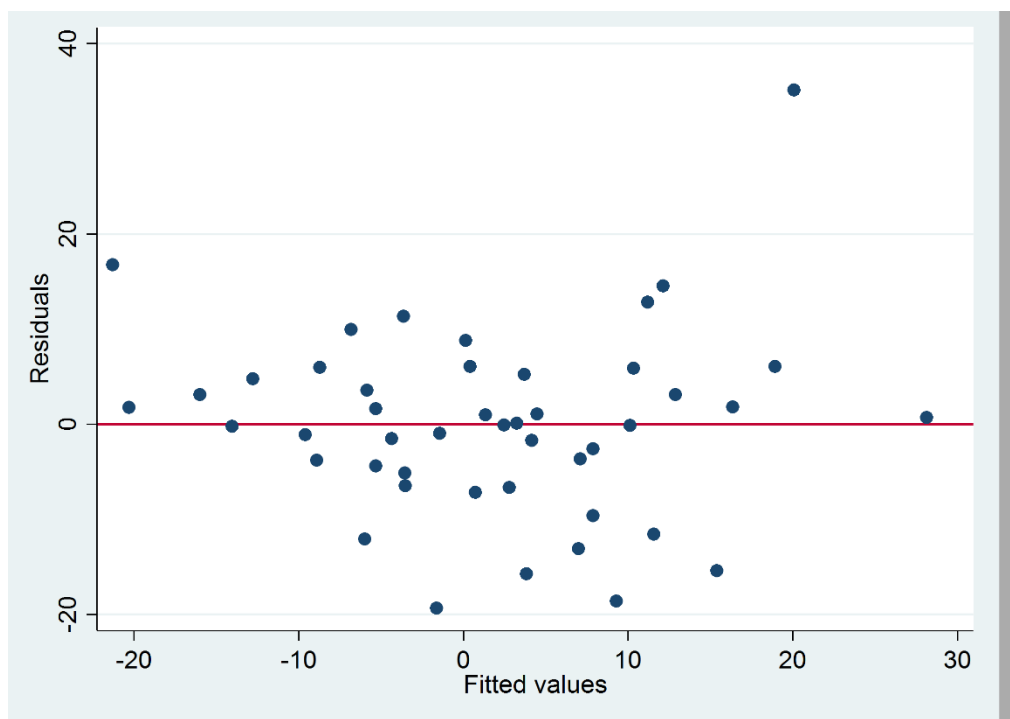


Voici un nuage de points montrant les valeurs observées et les valeurs estimées par notre modèle. Nous pouvons constater que la relation est linéaire et que notre modèle est une bonne représentation de la variation du taux de chômage.



Ce graphique montre la droite de notre régression et la courbe « idéale » pour expliquer la relation entre nos deux variables. Nous pouvons voir qu'un modèle non linéaire serait une meilleure représentation de cette relation, mais que notre modèle linéaire est tout à fait adéquat.

Désormais, nous allons tester l'homoscédasticité, car si les résidus sont hétéroscédastiques il faudra faire une régression « robuste » et réinterpréter la table ANOVA.




```
. estat imtest
Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	4.89	2	0.0866
Skewness	3.75	1	0.0528
Kurtosis	1.11	1	0.2926
Total	9.75	4	0.0448

```
. estat hettest
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of un
chi2(1) = 6.73
Prob > chi2 = 0.0095
```

Voici un graphique qui montre la variance et une table contenant deux tests de variance constante. Nous pouvons constater que la variance n'est pas constante, car elle est plus élevée entre -10 et +10. Ceci est confirmé par les deux tests numériques. H0 pour ces deux tests est l'homoscédasticité, en l'espèce les deux P values sont inférieures à 0.05 et nous rejetons donc H0. Ainsi, nous pouvons confirmer la présence d'hétéroscédasticité. L'hétéroscédasticité viole une des hypothèses de MCO que nous avons cité ultérieurement.

Pour corriger cet effet d'hétéroscédasticité, nous devons faire une régression « robust ». Nous allons voir que suite à cette nouvelle régression nos intervalles de confiance vont s'élargir pour compenser l'hétéroscédasticité.

```
. reg un pib in 1/46, robust
```

```
Linear regression               Number of obs   =          46
                               F(1, 44)          =          32.91
                               Prob > F           =          0.0000
                               R-squared           =          0.5360
                               Root MSE        =          9.9996
```

un	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
pib	-7.512546	1.309629	-5.74	0.000	-10.15193	-4.873163
_cons	8.669796	2.293213	3.78	0.000	4.048128	13.29146

Voici la table ANOVA après notre régression robuste. Nous pouvons constater que le modèle reste statistiquement significatif et que notre R² vaut 0.5360. Nos coefficients restent statistiquement significatifs et malgré l'élargissement des intervalles de confiance ils restent significativement différents de zéro.

L'espérance des résidus

```
. ttest res==0
```

One-sample t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
res	46	2.09e-08	1.457891	9.887898	-2.936343	2.936343

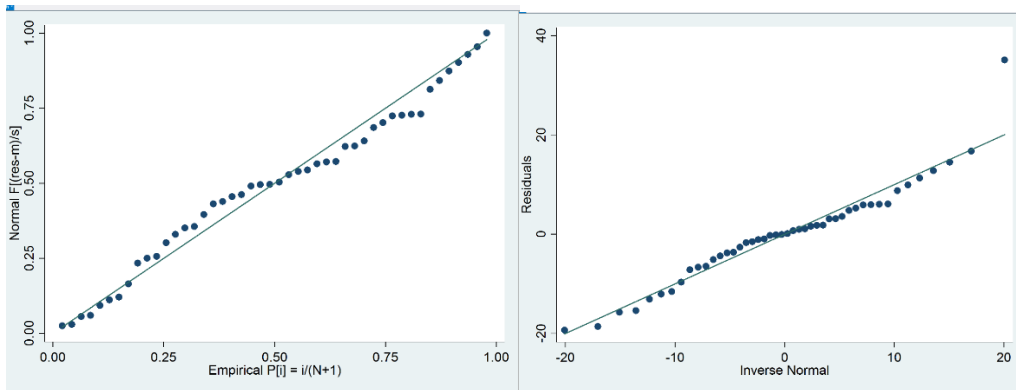
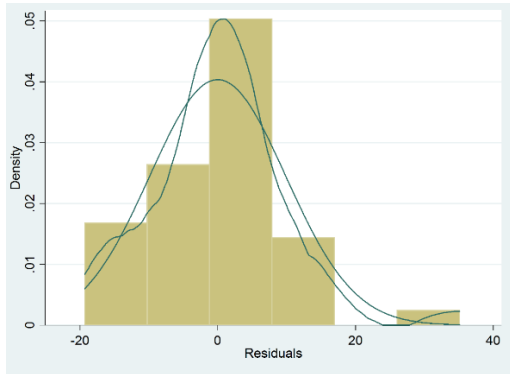
```
mean = mean(res)
Ho: mean = 0
Ha: mean < 0
Pr(T < t) = 0.5000

t = 0.0000
degrees of freedom = 45
Ha: mean != 0
Pr(|T| > |t|) = 1.0000
Ha: mean > 0
Pr(T > t) = 0.5000
```

Une autre hypothèse des MCO à respecter est que l'espérance de résidus doit être zéro. Ici nous avons fait un test de Student pour confirmer cette hypothèse. Nous pouvons voir que H_0 pour ce test est une espérance nulle et que la P Value vaut 1. Ainsi, nous ne rejetons pas H_0 et nous pouvons confirmer que nos résidus ont une espérance nulle.

Normalité des résidus

L'hypothèse de la normalité des résidus n'est pas indispensable afin d'obtenir des estimateurs convergents, mais elle va nous permettre de construire des tests statistiques concernant la validité du modèle estimé. La violation de cette hypothèse implique que les intervalles de confiance des coefficients ne sont pas fiables. De plus, la non-normalité rend difficile la prévision à partir du modèle.



```
. swilk res
```

Variable	Obs	W	V	z	Prob>z
res	46	0.94890	2.251	1.722	0.04256

```
. sktest res
```

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
res	46	0.0527	0.0142	8.42	0.0149

Ici nous utilisons des représentations visuelles et des tests numériques pour tester la normalité de nos résidus. Graphiquement nous pouvons voir que la distribution n'est pas normale. Ceci est confirmé par nos tests numériques. Nous avons utilisé un test Swilk et un test Jarque-Bera, H_0 dans ces deux tests est la normalité des résidus. Nos deux P values sont inférieures à 0.05 et ainsi nous pouvons rejeter H_0 . La distribution de nos résidus n'est pas normale.

La conséquence de cette non-normalité est que nos intervalles de confiance sont inadaptes.

Homoscédasticité des résidus

Une de nos hypothèses est que la variance de l'erreur est constante (homoscédastique). Les conséquences de l'hétéroscédasticité sont identiques à celles de l'autocorrélation des erreurs. En cas d'hétéroscédasticité stata permet d'utiliser une régression « robuste » qui élargit les intervalles de confiance des coefficients. C'est ce que nous avons fait en l'espèce.

Multicolinéarité

Nous pouvons citer trois effets principaux de la multicolinéarité :

- a) augmentation de la variance estimée de certains coefficients lorsque la colinéarité entre les variables explicatives augmente (le t de Student diminue) ;
- b) instabilité des estimations des coefficients des moindres carrés, de faibles fluctuations concernant les données entraînent de fortes variations des valeurs estimées des coefficients ;
- c) en cas de multicolinéarité parfaite, la matrice $X'X$ est singulière (le déterminant est nul), l'estimation des coefficients est alors impossible et leur variance est infinie.

. vif

Variable	VIF	1/VIF
pib	1.00	1.000000
Mean VIF	1.00	

Un VIF inférieur à 10 implique l'absence de colinéarité. En l'espèce, le VIF de notre variable explicative vaut 1 et ainsi nous n'avons pas de multicolinéarité. De plus, nous avons un modèle bivarié et ainsi nous n'avons pas besoin de nous soucier de la multicolinéarité. Nous avons cependant décidé de faire ce test dans un but pédagogique.

Indépendance/autocorrélation

Le test de Durbin et Watson (DW) permet de détecter une autocorrélation des erreurs d'ordre 1 selon la forme : $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ avec $v_t \rightarrow N(0, \sigma^2_v)$

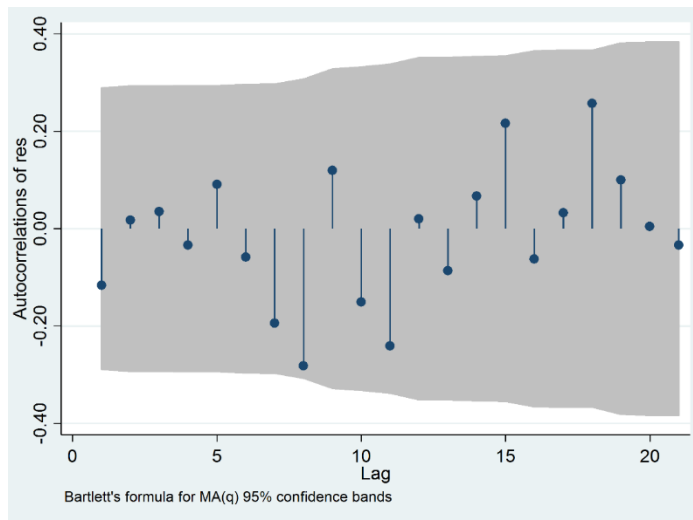
Le test d'hypothèses est le suivant :

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

De par sa construction, cette statistique varie entre 0 et 4 et nous avons $DW = 2$ lorsque $\rho = 0$. Afin de tester l'hypothèse H_0 , Durbin et Watson ont tabulé les valeurs critiques de DW au seuil de 5 % en fonction de la taille de l'échantillon n et du nombre de variables explicatives (k). La lecture de la table permet de déterminer deux valeurs d_1 et d_2 comprises entre 0 et 2 qui délimitent l'espace entre 0 et 4.

Pour éviter de devoir chercher dans des tables, stata propose un test de « durbinalt » qui nous permet de rejeter ou valider l'autocorrélation à l'aide d'une P value. De plus le test « durbinalt » est robuste à la non normalité des résidus.



Voici le graphique ACF de nos résidus, nous l'utilisons pour tester l'autocorrélation des résidus. En l'espèce nous pouvons voir que les résidus ne sont pas autocorrélés, cependant nous allons confirmer cela avec l'aide de plusieurs tests numériques.

```
. dwstat
Durbin-Watson d-statistic( 2, 46) = 2.22357
. estat durbinalt, force
Durbin's alternative test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	0.609	1	0.4351

H0: no serial correlation

Pour le test « durbinalt » H0 est l'absence d'autocorrélation. En l'espèce notre P value vaut 0.4351 et nous ne rejetons donc pas H0. Nos résidus ne sont pas autocorrélés.

```
. correlate res pib in 1/46
(obs=46)
```

	res	pib
res	1.0000	
pib	0.0000	1.0000

Nous pouvons exclure également la corrélation entre nos résidus et notre variable explicative.

Goodness of fit

. ovtest						
Ramsey RESET test using powers of the fitted values of un						
Ho: model has no omitted variables						
F(3, 41) = 2.51						
Prob > F = 0.0720						
. linktest						
Source	SS	df	MS	Number of obs	=	46
Model	5607.72638	2	2803.86319	F(2, 43)	=	31.11
Residual	3875.09086	43	90.1183921	Prob > F	=	0.0000
				R-squared	=	0.5914
				Adj R-squared	=	0.5723
Total	9482.81724	45	210.729272	Root MSE	=	9.4931
un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_hat	.9201107	.1372053	6.71	0.000	.6434099	1.196812
_hatsq	.0223372	.0092583	2.41	0.020	.0036662	.0410083
_cons	-2.397506	1.730119	-1.39	0.173	-5.886623	1.09161

Le test de Ramsey, aussi appelé le test de RESET (Regression Error Specification Test), porte sur la pertinence de la forme fonctionnelle du modèle, telle que :

- une relation fonctionnelle non adaptée (passage aux logarithmes, fonctions inverses...) entre la variable à expliquer et les variables explicatives ;
- l'absence d'une variable explicative dans le modèle ;
- la corrélation entre la variable explicative et le terme d'erreur ;

Si l'hypothèse H0 est acceptée, le modèle est bien linéaire et il n'existe pas de problème de spécification. En l'espèce la P value vaut 0.07 et nous ne rejetons pas H0. Ainsi nous pouvons conclure qu'il existe bien une relation linéaire entre nos deux variables et en plus nous n'avons pas omis des variables explicatives.

Nous avons également fait un « linktest » qui teste la viabilité d'un modèle non linéaire pour expliquer la relation entre nos deux variables. En l'espèce `_hatsq` est significatif au seuil de 5% et nous pouvons donc conclure qu'un modèle non linéaire serait également une bonne représentation de cette relation.

Conclusion

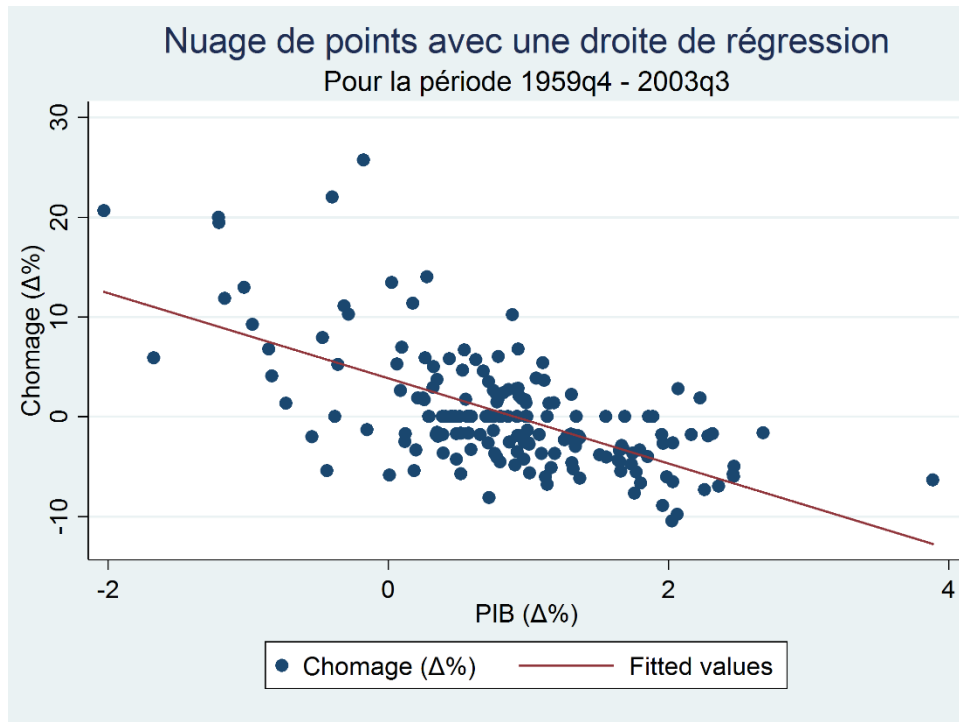
Ainsi, nous pouvons conclure que notre modèle respecte les hypothèses essentielles d'une régression par MCO sauf l'homoscédasticité, mais en passant par une régression robuste nous avons pris cette violation en compte. Nous pouvons constater également qu'il existe bien une relation linéaire et négative entre la variation du chômage et la variation du PIB. De plus, avec un R² de 0.536, notre modèle explique environ 54% de cette variation.

Notre modèle final pour cette période est la suivante :

$$\text{Ch\^omage}\Delta\%t = 8.669796 + (-7.512546 * \text{PIB}\Delta\%t)$$

Régression de la période 2 (1959q4 à 2003q3)

Dans cette partie nous allons modéliser la 2e période



Grâce au graphique de nuage de points, nous pouvons confirmer qu'il existe une relation négative entre le taux de chômage et le PIB dans cette période. Nous pouvons soupçonner que cette période aura de l'hétéroscédasticité, car la variance des résidus semble être décroissante.

ANOVA

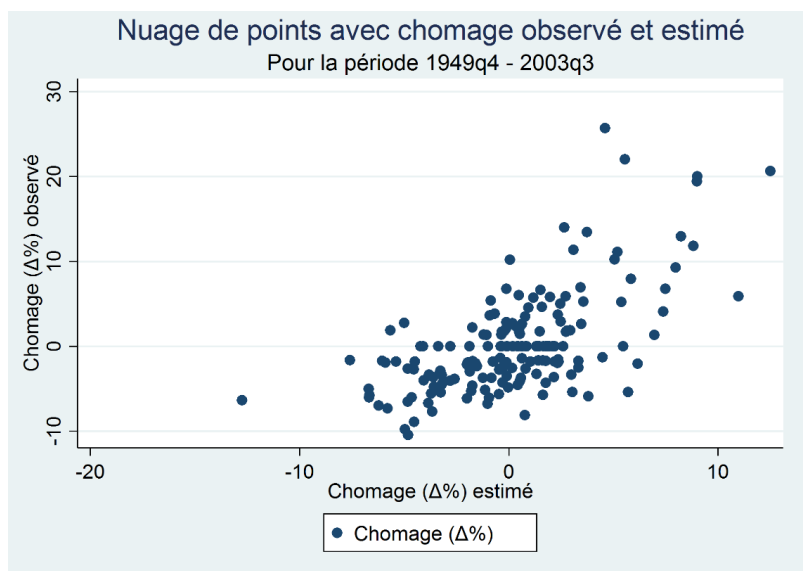
```
. reg un pib in 47/222
```

Source	SS	df	MS	Number of obs	=	176
Model	2411.36457	1	2411.36457	F(1, 174)	=	116.53
Residual	3600.56013	174	20.6928743	Prob > F	=	0.0000
				R-squared	=	0.4011
				Adj R-squared	=	0.3977
Total	6011.92469	175	34.3538554	Root MSE	=	4.5489

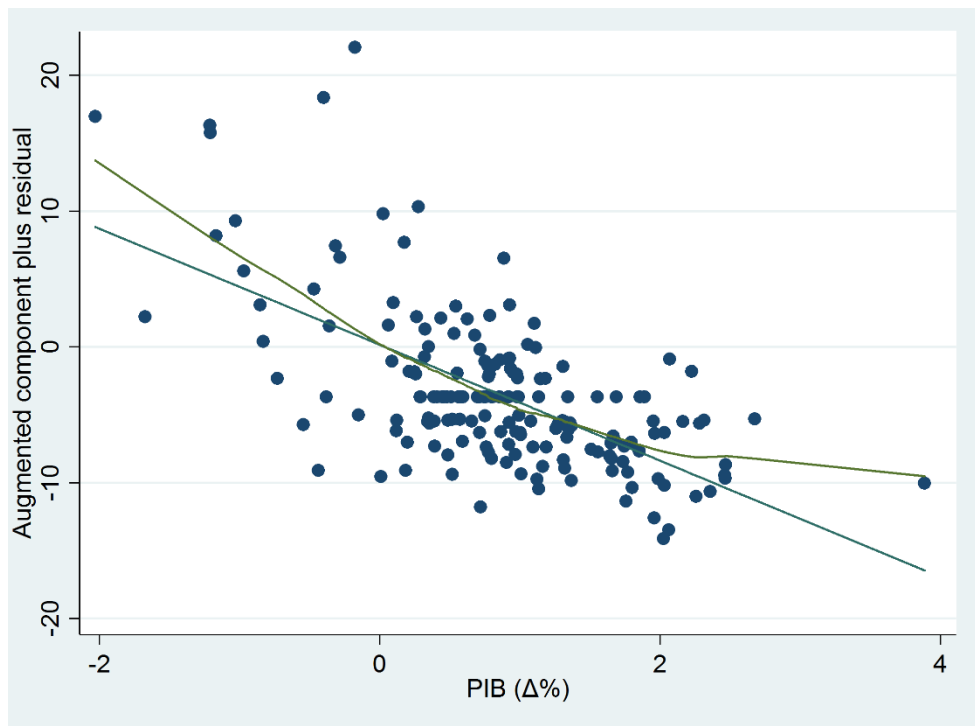
un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pib	-4.268355	.3954025	-10.79	0.000	-5.048757	-3.487952
_cons	3.841276	.4796828	8.01	0.000	2.894531	4.788022

Voici l'output de notre régression pour la période. Cette table contient plusieurs informations essentielles.

- La statistique F et $\text{prob} > F$ indiquent la significativité statistique du modèle. En l'espèce, la P value de notre statistique F est 0, et ainsi nous pouvons conclure que notre modèle est statistiquement significatif.
- Le R^2 et le R^2 ajusté indiquent la variation de chômage qui est expliquée par notre modèle. En l'espèce, 40% de la variation du taux de chômage est expliqué par la variation du PIB et donc une baisse par rapport à la 1re période.
- En l'espèce, la variable PIB à une P value de 0 et elle est donc significative. Son coefficient est de -4.27 ce qui confirme la relation inverse entre le chômage et le PIB. De plus, en vue des intervalles de confiance (-5.05 à -3.49) nous pouvons constater que le coefficient est significativement différent de zéro.
- Ce modèle contient également une constante. En l'espèce la constante est significative avec une P value de 0, sa valeur est de 3.84. Cette constante représente l'intercepte. Ainsi, lorsque $x_t=0$, $y_t=k$
- Number of obs indique le nombre d'observations pour la régression. En l'espèce notre « n » vaut 176 et donc la taille de notre échantillon est supérieure à 35 ce qui implique qu'elle est suffisamment grande pour tirer des conclusions.

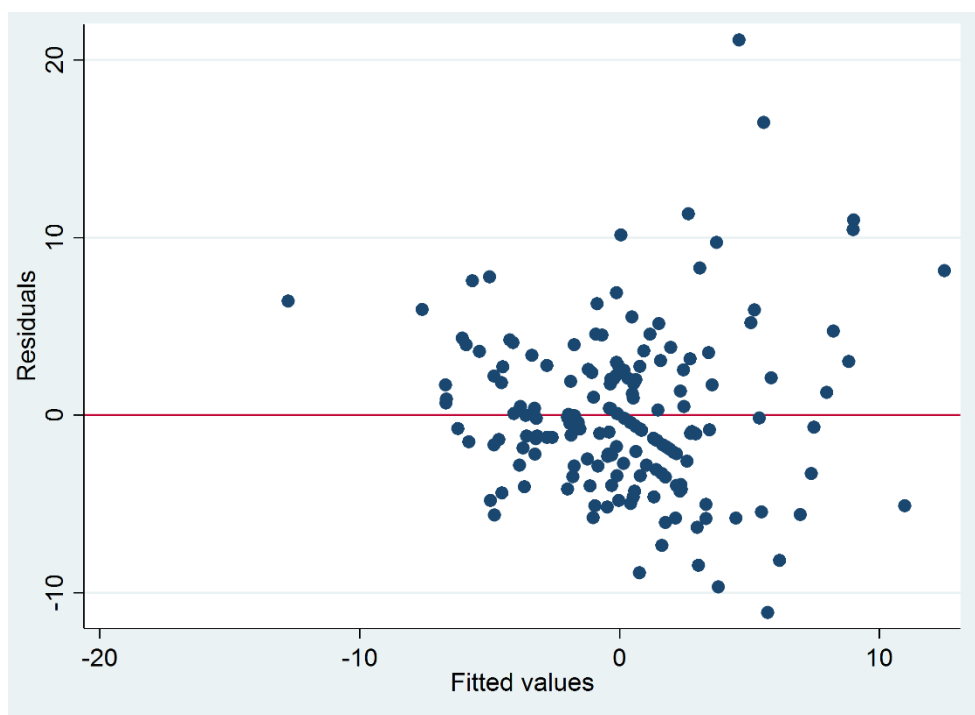


Voici un nuage de points montrant les valeurs observées et les valeurs estimées par notre modèle. Nous pouvons constater que la relation est linéaire et que notre modèle est une bonne représentation de la variation du taux de chômage.



Ce graphique montre la droite de notre régression et la courbe « idéale » pour expliquer la relation entre nos deux variables. Nous pouvons voir qu'un modèle non linéaire serait une meilleure représentation de cette relation, mais que notre modèle linéaire est tout à fait adéquat.

Désormais, nous allons tester l'homoscédasticité, car si les résidus sont hétéroscédastiques il faudra faire une régression « robuste » et réinterpréter la table ANOVA.



```
. estat imtest

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	18.27	2	0.0001
Skewness	7.58	1	0.0059
Kurtosis	1.88	1	0.1699
Total	27.73	4	0.0000

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of un

chi2(1)      =    33.21
Prob > chi2   =    0.0000
```

Voici un graphique qui montre la variance et une table contenant deux tests de variance constante. Nous pouvons constater que la variance n'est pas constante, car elle est croissante. Ceci est confirmé par les deux tests numériques. H_0 pour ces deux tests est l'homoscédasticité, en l'espèce les deux P values sont inférieures à 0.05 et nous rejetons donc H_0 . Ainsi, nous pouvons confirmer la présence d'hétéroscédasticité. L'hétéroscédasticité viole une des hypothèses de MCO que nous avons cité ultérieurement.

Pour corriger cet effet d'hétéroscédasticité, nous devons faire une régression « robust ». Nous allons voir que suite à cette nouvelle régression nos intervalles de confiance vont s'élargir pour compenser l'hétéroscédasticité.

```
. reg un pib in 47/222, robust
```

Linear regression

Number of obs	=	176
F(1, 174)	=	66.92
Prob > F	=	0.0000
R-squared	=	0.4011
Root MSE	=	4.5489

un	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
pib	-4.268355	.5217561	-8.18	0.000	-5.29814 -3.238569
_cons	3.841276	.6744028	5.70	0.000	2.510213 5.172339

Voici la table ANOVA après notre régression robuste. Nous pouvons constater que le modèle reste statistiquement significatif et que notre R^2 vaut 0.4. Nos coefficients restent statistiquement significatifs et malgré l'élargissement des intervalles de confiance ils restent significativement différents de zéro.

Espérance

```
. ttest res==0
```

One-sample t test

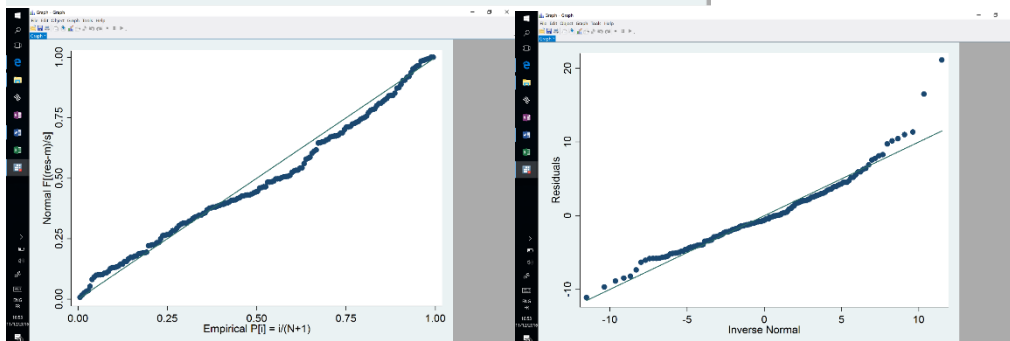
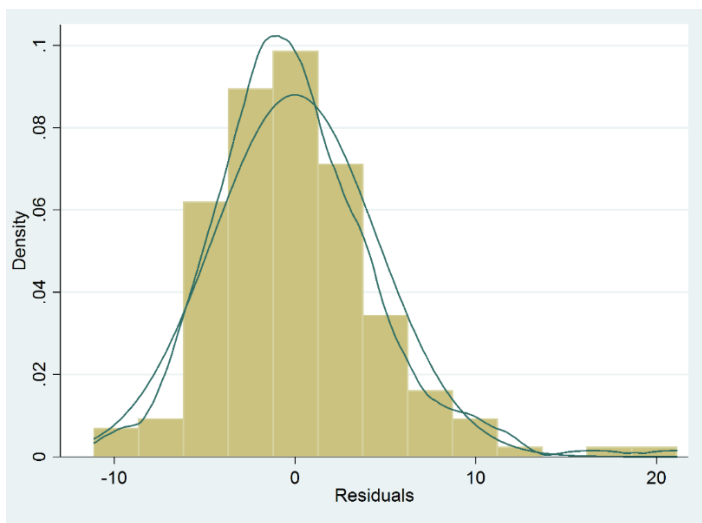
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
res	176	1.30e-08	.3419083	4.535927	-.6747945	.6747945

mean = mean(**res**)
Ho: mean = 0
Ha: mean < 0
Pr(T < t) = 0.5000

t = 0.0000
degrees of freedom = 175
Ha: mean != 0
Pr(|T| > |t|) = 1.0000
Ha: mean > 0
Pr(T > t) = 0.5000

Une autre hypothèse des MCO à respecter est que l'espérance de résidus doit être zéro. Ici nous avons fait un test de student pour confirmer cette hypothèse. Nous pouvons voir que H0 pour ce test est une espérance nulle et que la P Value vaut 1. Ainsi, nous ne rejetons pas H0 et nous pouvons confirmer que nos résidus ont une espérance nulle.

Normalité



```
. swilk res
```

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
res	176	0.95102	6.540	4.293	0.00001

```
. sktest res
```

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
res	176	0.0000	0.0000	30.37	0.0000

Ici nous utilisons des représentations visuelles et des tests numériques pour tester la normalité de nos résidus. Graphiquement nous pouvons voir que la distribution n'est pas normale. Ceci est confirmé par nos tests numériques. Nous avons utilisé un test Swilk et un test Jarque-Bera, H0 dans ces deux tests est la normalité des résidus. Nos deux P values sont inférieures à 0.05 et ainsi nous pouvons rejeter H0. La distribution de nos résidus n'est pas normale.

La conséquence de cette non-normalité est que nos intervalles de confiance sont inadaptés.

Homoscédasticité

Nous avons déjà identifié la présence d'hétéroscédasticité et nous l'avons pris en compte en faisant une régression robuste.

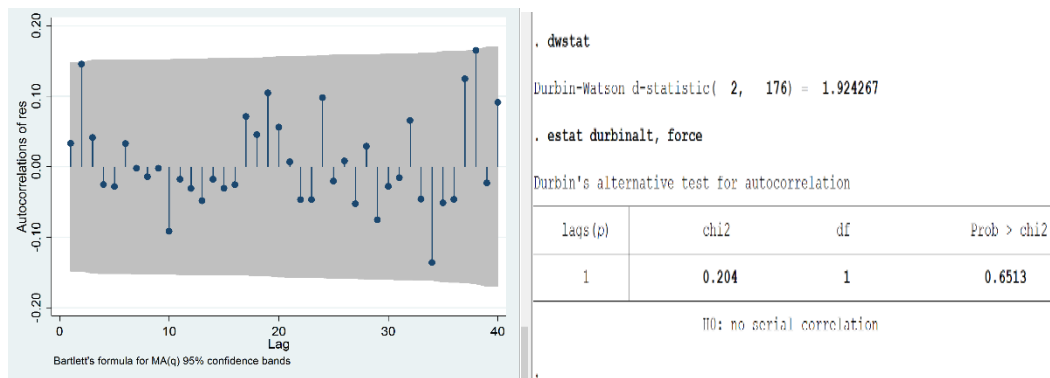
Multicolinéarité

```
. vif
```

Variable	VIF	1/VIF
pib	1.00	1.000000
Mean VIF	1.00	

Un VIF inférieur à 10 implique l'absence de colinéarité. En l'espèce, le VIF de notre variable explicative vaut 1 et ainsi nous n'avons pas de multicolinéarité. De plus, nous avons un modèle bivarié et ainsi nous n'avons pas besoin de nous soucier de la multicolinéarité. Nous avons cependant décidé de faire ce test dans un but pédagogique.

Indépendance/autocorrélation



Voici le graphique ACF de nos résidus, nous l'utilisons pour tester l'autocorrélation des résidus. En l'espèce nous pouvons voir que les résidus ne sont pas autocorrélés, cependant nous allons confirmer cela avec l'aide de plusieurs tests numériques.

Pour le test « durbinalt » H_0 est l'absence d'autocorrélation. En l'espèce notre P value vaut 0.6513 et nous ne rejetons donc pas H_0 . Nos résidus ne sont pas autocorrélés.

```
. correlate res pib in 47/222
(obs=176)
```

	res	pib
res	1.0000	
pib	-0.0000	1.0000

Nous pouvons exclure également la corrélation entre nos résidus et la variable explicative.

Goodness of fit

```
. ovtest
Ramsey RESET test using powers of the fitted values of un
Ho: model has no omitted variables
F(3, 171) = 3.67
Prob > F = 0.0135

. linktest
```

Source	SS	df	MS	Number of obs	=	176
Model	2627.04926	2	1313.52463	F(2, 173)	=	67.13
Residual	3384.87543	173	19.5657539	Prob > F	=	0.0000
Total	6011.92469	175	34.3538554	R-squared	=	0.4370
				Adj R-squared	=	0.4305
				Root MSE	=	4.4233

un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_hat	.9431726	.0916893	10.29	0.000	.7621989 1.124146
_hatsq	.0465419	.0140179	3.32	0.001	.0188738 .07421
_cons	-.6274113	.38376	-1.63	0.104	-1.384866 .1300432

Le test de Ramsey, aussi appelé le test de RESET (Regression Error Specification Test), porte sur la pertinence de la forme fonctionnelle du modèle, telle que :

- une relation fonctionnelle non adaptée (passage aux logarithmes, fonctions inverses...) entre la variable à expliquer et les variables explicatives ;
- l'absence d'une variable explicative dans le modèle ;
- la corrélation entre la variable explicative et le terme d'erreur ;

Si l'hypothèse H_0 est acceptée, le modèle est bien linéaire et il n'existe pas de problème de spécification. En l'espèce la P value vaut 0.01 et nous ne rejetons pas H_0 . Ainsi nous pouvons conclure qu'il n'existe peut-être pas une relation linéaire entre nos deux variables et en plus nous avons omis des variables explicatives.

Nous avons également fait un « linktest » qui teste la viabilité d'un modèle non linéaire pour expliquer la relation entre nos deux variables. En l'espèce $_hatsq$ est significatif au seuil de 5% et nous pouvons donc conclure qu'un modèle non linéaire serait également une bonne représentation de cette relation.

Ainsi, en vue de ces deux tests, nous pouvons conclure qu'un modèle non linéaire serait mieux adapté pour expliquer cette relation et qu'en plus nous devrions chercher plus de variables explicatives qui pourraient expliquer la variation du taux de chômage.

Conclusion

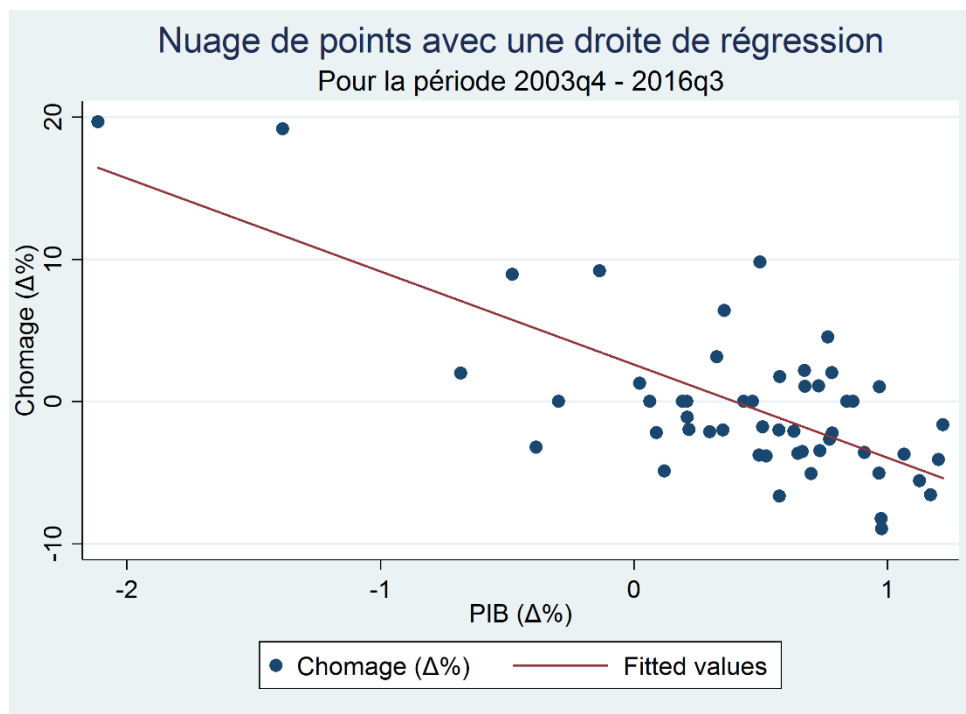
Ainsi, nous pouvons conclure que notre modèle confirme l'existence d'une relation négative entre le taux de chômage et le PIB. Cependant, notre R^2 et notre coefficient sur la variable PIB sont plus petits que dans la 1re période. Cela signifie que le PIB continue à affecter le taux de chômage, mais que son effet est affaibli dans cette période. Le test RESET confirme cette notion en nous indiquant l'omission de variables explicatives dans notre modèle.

Le modèle final pour la 2e période est le suivant :

$$\text{Ch\^omage}\Delta\%t = 3.841276 + (-4.268355 * \text{PIB}\Delta\%t)$$

Régression période 3 (de 2003q4 à 2016q3)

Dans cette partie nous allons modéliser la 3e période



Grâce au graphique de nuage de points, nous pouvons confirmer qu'il existe une relation négative entre le taux de chômage et le PIB dans cette période. Nous pouvons soupçonner que cette période aura de l'homoscédasticité, car la variance des résidus semble être constante autour de la droite de régression.

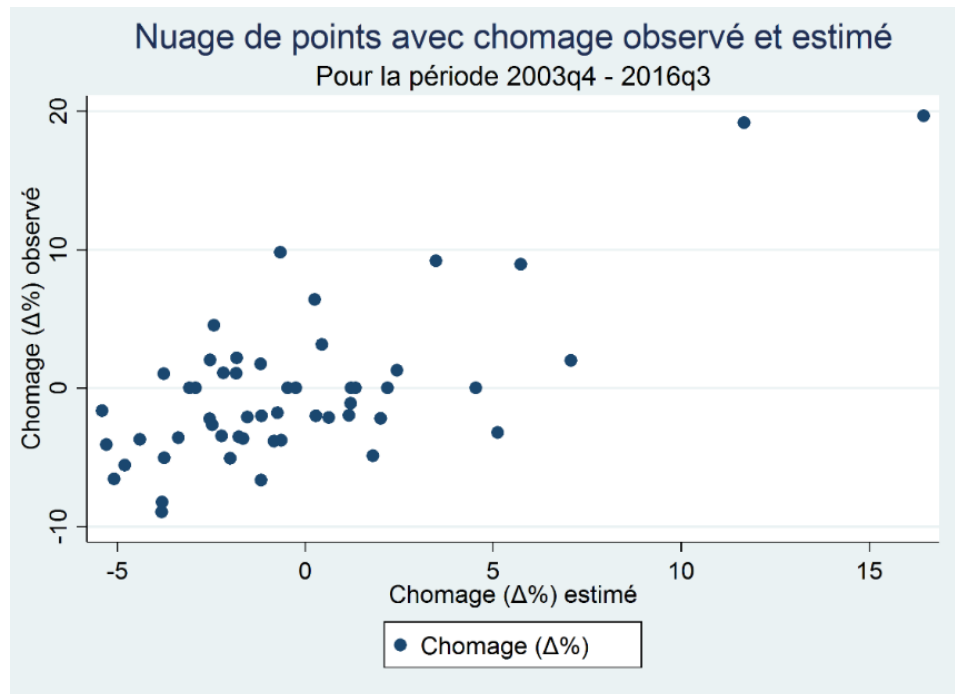
ANOVA

```
. reg un pib if number>222
```

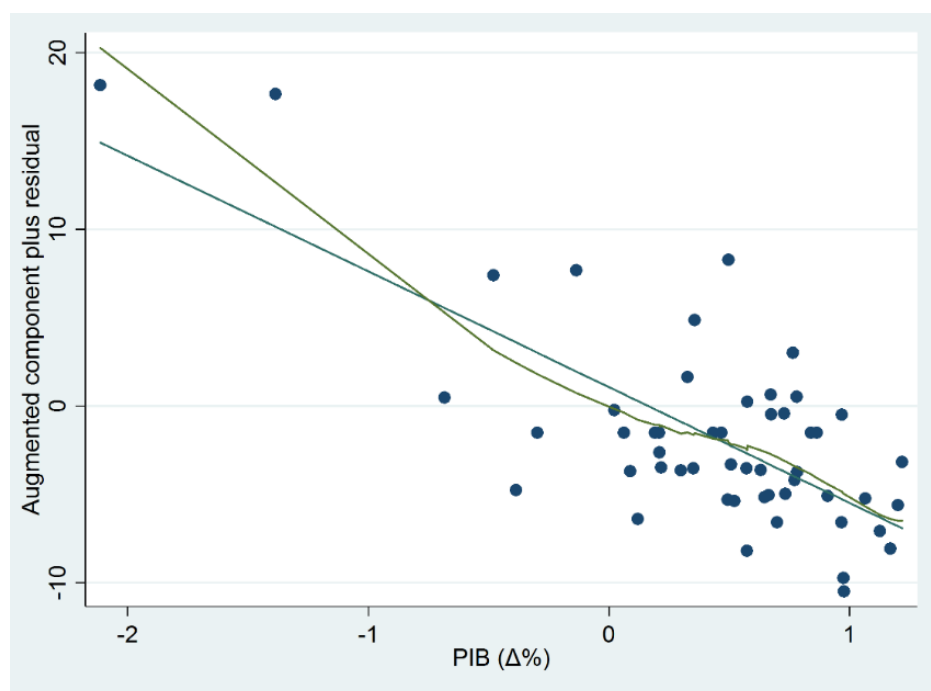
Source	SS	df	MS	Number of obs	=	52
Model	840.758754	1	840.758754	F(1, 50)	=	54.61
Residual	769.745668	50	15.3949134	Prob > F	=	0.0000
Total	1610.50442	51	31.5785181	R-squared	=	0.5220
				Adj R-squared	=	0.5125
				Root MSE	=	3.9236

un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pib	-6.552935	.8867247	-7.39	0.000	-8.333974	-4.771896
_cons	2.593006	.665223	3.90	0.000	1.256867	3.929146

Voici l'output de notre régression. Nous voyons que le R^2 vaut 0.52 (une hausse par rapport à la 2e période) et nous voyons que notre modèle est statistiquement significatif avec une P value de 0 en l'espèce. Les coefficients de nos paramètres sont significatifs et différents de zéro, et la variable PIB a toujours un coefficient négatif.



Voici un nuage de points montrant les valeurs observées et les valeurs estimées par notre modèle. Nous pouvons constater que la relation est linéaire et que notre modèle est une bonne représentation de la variation du taux de chômage.



Ce graphique montre la droite de notre régression et la courbe « idéale » pour expliquer la relation entre nos deux variables. Nous pouvons voir que notre modèle linéaire est tout à fait adéquat.

Espérance

```
. ttest res==0
```

One-sample t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
res	52	1.40e-08	.5387495	3.884978	-1.081585	1.081585

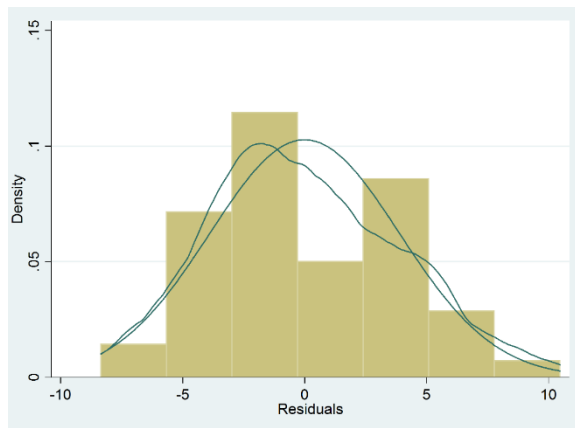
mean = mean(**res**)
Ho: mean = 0
Ha: mean < 0
Pr(T < t) = 0.5000

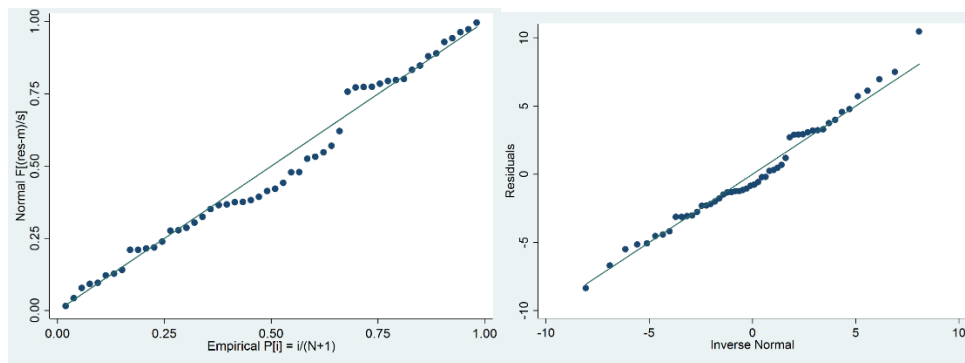
t = 0.0000
degrees of freedom = 51
Ha: mean != 0
Pr(|T| > |t|) = 1.0000

Ha: mean > 0
Pr(T > t) = 0.5000

Une autre hypothèse des MCO à respecter est que l'espérance de résidus doit être zéro. Ici nous avons fait un test de student pour confirmer cette hypothèse. Nous pouvons voir que H0 pour ce test est une espérance nulle et que la P Value vaut 1. Ainsi, nous ne rejetons pas H0 et nous pouvons confirmer que nos résidus ont une espérance nulle.

Normalité





```
. swilk res
```

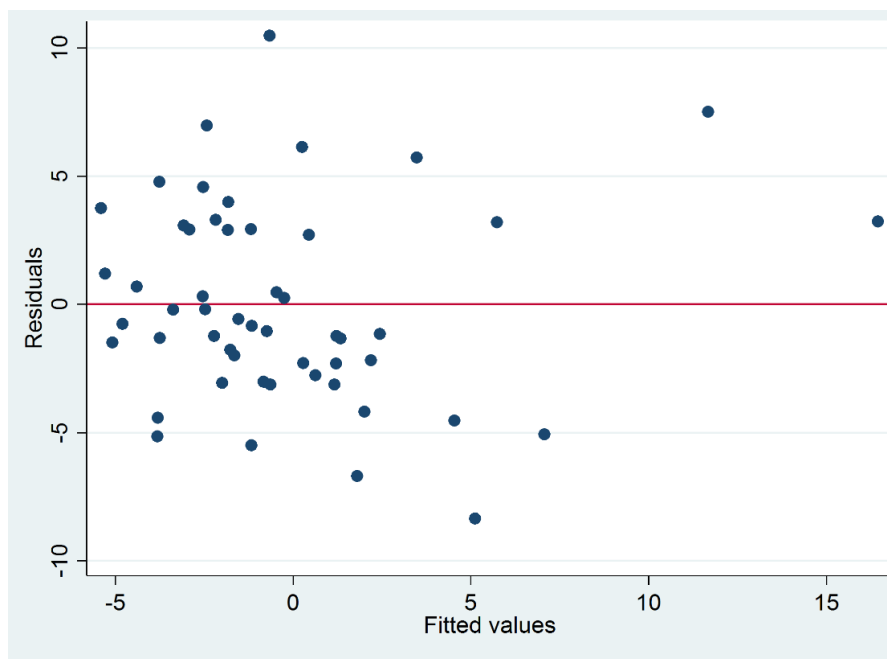
Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
res	52	0.98091	0.926	-0.165	0.56535

```
. sktest res
```

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
res	52	0.2349	0.8295	1.52	0.4666

Ici nous utilisons des représentations visuelles et des tests numériques pour tester la normalité de nos résidus. Graphiquement nous pouvons voir que la distribution est plutôt normale. Ceci est confirmé par nos tests numériques. Nous avons utilisé un test Swilk et un test Jarque-Bera, H_0 dans ces deux tests est la normalité des résidus. Nos deux P values sont supérieures à 0.05 et ainsi nous pouvons accepter H_0 . La distribution de nos résidus est normale.

Homoscédasticité



```
. estat imtest
```

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	4.48	2	0.1063
Skewness	1.44	1	0.2305
Kurtosis	0.06	1	0.8100
Total	5.98	4	0.2007

```
. estat hettest
```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of un

chi2(1) = 3.65
Prob > chi2 = 0.0560

```
. estat archlm
```

LM test for autoregressive conditional heteroskedasticity (ARCH)

lags(p)	chi2	df	Prob > chi2
1	0.233	1	0.6295

H0: no ARCH effects vs. H1: ARCH(p) disturbance

Voici un graphique qui montre la variance et une table contenant deux tests de variance constante. Nous pouvons constater que la variance semble être constante. Ceci est confirmé par les deux tests numériques. H0 pour ces deux tests est l'homoscédasticité, en l'espèce les deux P values sont supérieures à 0.05 et nous acceptons donc H0. Ainsi, nous pouvons confirmer la présence d'homoscédasticité. De plus, nous pouvons exclure la possibilité d'effets ARCH.

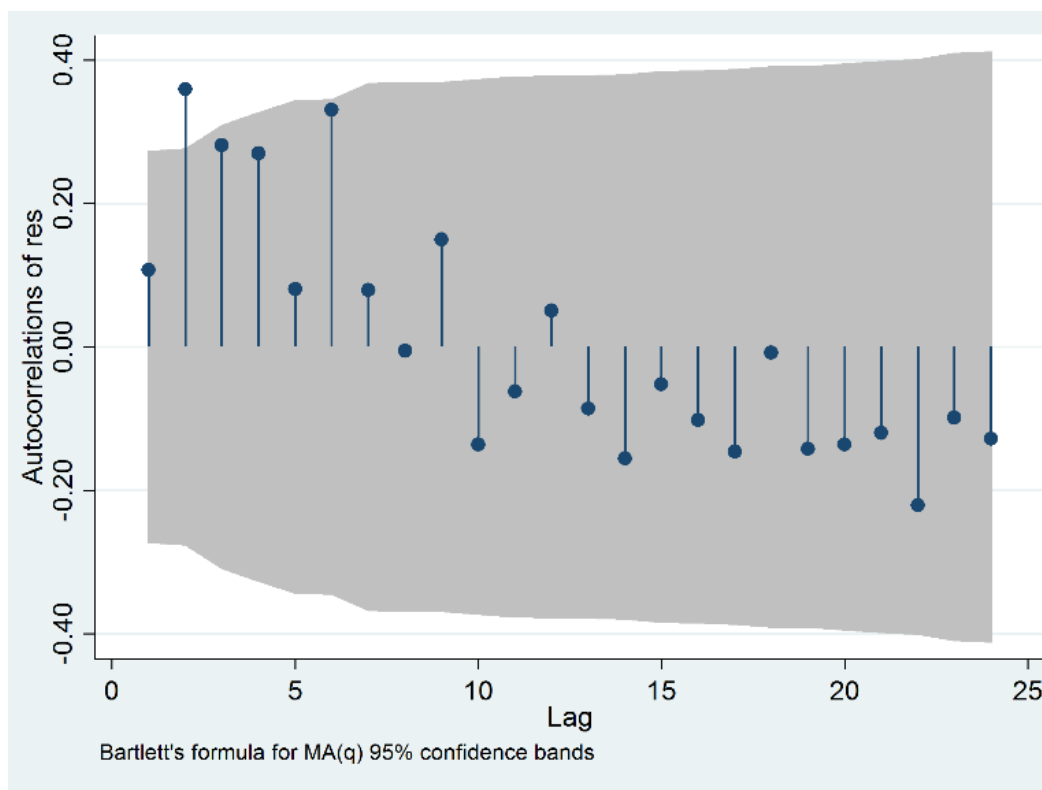
Multicolinéarité

. vif

Variable	VIF	1/VIF
pib	1.00	1.000000
Mean VIF	1.00	

Un VIF inférieur à 10 implique l'absence de colinéarité. En l'espèce, le VIF de notre variable explicative vaut 1 et ainsi nous n'avons pas de multicolinéarité. De plus, nous avons un modèle bivarié et ainsi nous n'avons pas besoin de nous soucier de la multicolinéarité. Nous avons cependant décidé de faire ce test dans un but pédagogique.

Indépendance/Autocorrélation



Voici le graphique ACF de nos résidus, nous l'utilisons pour tester l'autocorrélation des résidus. En l'espèce, nous pouvons voir que les résidus ne sont pas autocorrélés, cependant nous allons confirmer cela avec l'aide de plusieurs tests numériques.

```
. dwstat
```

Durbin-Watson d-statistic(2, 52) = 1.755785

```
. estat bgodfrey
```

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	0.682	1	0.4088

H0: no serial correlation

```
. estat durbinalt
```

Durbin's alternative test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	0.652	1	0.4195

H0: no serial correlation

Pour les tests « durbinalt » et « bgodfrey » H0 est l'absence d'autocorrélation. En l'espèce nos P values sont supérieures à 0.05 et nous ne rejetons donc pas H0. Nos résidus ne sont pas autocorrélés.

```
. correlate res pib if number>222
(obs=52)
```

	res	pib
res	1.0000	
pib	-0.0000	1.0000

Nous pouvons exclure également la corrélation entre nos résidus et la variable explicative.

Goodness of fit

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of un
Ho: model has no omitted variables
F(3, 47) = 2.71
Prob > F = 0.0556

```
. linktest
```

Source	SS	df	MS	Number of obs	=	52
Model	896.840493	2	448.420246	F(2, 49)	=	30.79
Residual	713.663929	49	14.56457	Prob > F	=	0.0000
				R-squared	=	0.5569
				Adj R-squared	=	0.5388
Total	1610.50442	51	31.5785181	Root MSE	=	3.8164

un	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_hat	.7440874	.1852873	4.02	0.000	.3717386 1.116436
_hatsq	.0356647	.0181751	1.96	0.055	-.0008595 .0721888
_cons	-.6388123	.6221142	-1.03	0.310	-1.888998 .611373

Le test de Ramsey, aussi appelé le test de RESET (Regression Error Specification Test), porte sur la pertinence de la forme fonctionnelle du modèle, telle que :

- une relation fonctionnelle non adaptée (passage aux logarithmes, fonctions inverses...) entre la variable à expliquer et les variables explicatives ;
- l'absence d'une variable explicative dans le modèle ;

– la corrélation entre la variable explicative et le terme d'erreur ;

Si l'hypothèse H_0 est acceptée, le modèle est bien linéaire et il n'existe pas de problème de spécification. En l'espèce la P value vaut 0.0556 et nous ne rejetons pas H_0 . Ainsi nous pouvons conclure qu'il existe bien une relation linéaire entre nos deux variables et en plus nous n'avons pas omis des variables explicatives.

Nous avons également fait un « linktest » qui teste la viabilité d'un modèle non linéaire pour expliquer la relation entre nos deux variables. En l'espèce, $_hatsq$ n'est pas significatif au seuil de 5% et nous pouvons donc conclure qu'un modèle linéaire est la meilleure représentation de la relation entre le chômage et le PIB. Cependant, il faut remarquer qu'en augmentant notre seuil de rejet à 6% pour le RESET et le Linktest on aurait rejeté les deux H_0 . Ainsi, si on allait faire une étude plus approfondie il serait pertinent de tester des modèles non linéaires et des modèles avec plusieurs variables explicatives.

Conclusion

Ainsi, nous pouvons conclure que notre modèle respecte toutes les hypothèses essentielles d'une régression par MCO et que nos résidus sont des bruits blancs gaussiens. Nous pouvons constater également qu'il existe bien une relation linéaire et négative entre la variation du chômage et la variation du PIB. De plus, avec un R^2 de 0.52, notre modèle explique environ 52% de cette variation. La période 1 et la période 3 se ressemblent énormément avec un R^2 qui est quasiment identique et un coefficient sur PIBt qui est très similaire. Nous pouvons donc conclure que la loi d'Okun regagne son efficacité après plusieurs années de crise.

Le modèle final pour la 3e période est le suivant :

$$\text{Ch\^omage}\Delta\%t = 2.593006 + (-6.552935 * \text{PIB}\Delta\%t)$$

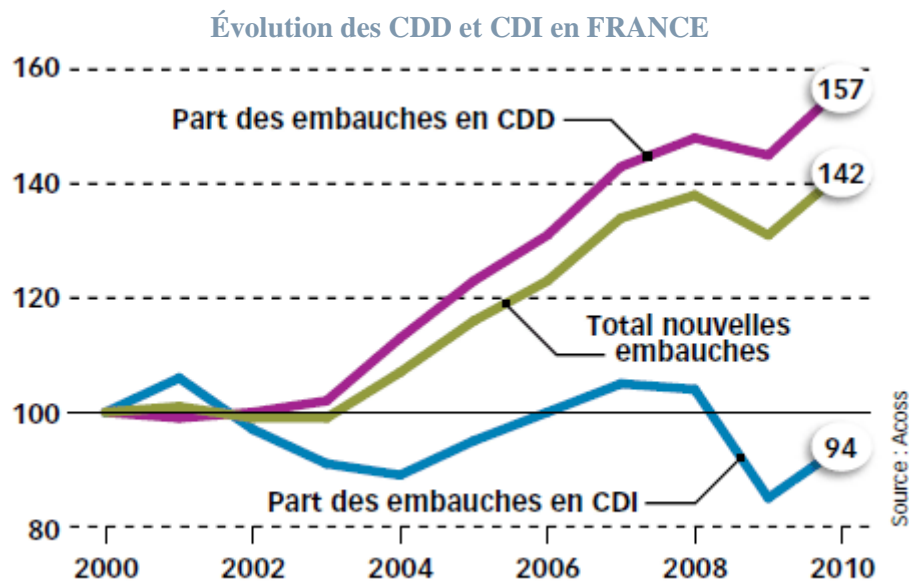
Conclusion

Comme nous pouvons le constater grâce à nos manœuvres sur Stata, nous avons pour la période 1, un R^2 de 0.53 avec un Coefficient de -7.5. Pour la période 2, nous obtenons un R^2 de 0.4 et un coefficient de -4.27 puis pour la période 3 un R^2 de 0.52 et un coefficient de -6.56.

Nous en concluons donc que la théorie est bien validée : lorsque la production augmente, le taux de chômage diminue. La loi d'Okun est bien stable, fiable et durable, et ce même durant les périodes de crise comme nous pouvons le constater durant l'ensemble de notre étude des années 1950 à aujourd'hui.

La période 1 et la période 3 sont analogues, nous obtenons un moment de "faiblesse" pour la théorie pendant la 2e période où le coefficient sur PIBt et le R^2 baissent. Nous pouvons expliquer cela en partie avec les crises économiques et financières, dont le choc pétrolier de 1973. De plus, nous avons un changement radical de trajectoire propulsé par l'émergence des politiques néolibérales qui appliquent la politique des 3D : déréglementation, désintermédiation, décloisonnement.

Si le coefficient et le R^2 sont si élevés à ce jour, nous pouvons l'expliquer à travers l'augmentation des contrats de travail à durée déterminée pour la France. Les entreprises peuvent varier la production plus facilement en fonction de la demande. Cela a alors un impact sur le taux de chômage en fonction de la variation de la production.



Pour les États-Unis, « royaume du libéralisme », les États fédérés peuvent déterminer le cadre légal du contrat de travail. Les contrats peuvent également se faire sous la forme « At Will », c'est-à-dire par accord verbal. Les contrats écrits, les « *Contracts employee* » disposeront de la durée du contrat et de la rémunération. Ainsi, nous avons un maximum de flexibilité aux États-Unis d'Amérique, cela explique donc très bien pourquoi nous obtenons un coefficient et un R^2 élevé.

OUVERTURE

À ce jour les politiques accommodantes localisées en zone euro et aux États-Unis d'Amérique semblent être en difficulté. Ces politiques accommodantes également nommées politiques non conventionnelles ou quantitative easing visent à augmenter l'inflation, à stimuler l'activité économique et à diminuer le chômage, la courbe de Phillips et la loi d'Okun se retrouvent donc au cœur de ces politiques non conventionnelles.

Cependant, d'une part en zone euro, nous voyons apparaître une trappe à liquidité avec comme conséquence des investissements dans l'économie réelle qui se font rares, mais la sphère financière et donc les cours boursiers s'en retrouvent grandement «dopés ». Cela peut s'expliquer à travers la récente réglementation prudentielle Bâle III qui incite les banques à augmenter leurs réserves et à prêter moins à l'économie réelle.

D'autre part, aux États-Unis d'Amérique, la Réserve fédérale et plus particulièrement la présidente Janet Yellen a décidé d'augmenter les taux, cependant des incertitudes sont persistantes ce qui peut se refléter dans les marchés. La réserve fédérale augmentera donc progressivement ses taux pour pallier les incertitudes.

La Courbe de Phillips qui met en relation le taux de chômage et le taux d'inflation se rapproche de la loi d'Okun qui met en relation la production et le taux de chômage. Dès lors, nous en venons à obtenir une relation entre taux d'inflation et production qui permet d'obtenir la courbe d'offre agrégée. C'est pourquoi les politiques actuelles cherchent à stimuler le taux d'inflation en s'appuyant sur la théorie quantitative de la monnaie $MV = PT$ à travers les politiques non conventionnelles et à favoriser une politique d'offre pour produire plus et donc dans un objectif de diminuer le taux le chômage. Cela permettra donc de favoriser le bien-être de la population comme peuvent l'indiquer les différentes écoles qui favorisent l'économie de l'offre par exemple, les néoclassiques. Cependant, cela présente à l'heure actuelle beaucoup de problématique : risque de trappe à liquidité, problèmes de confiance envers les autorités politiques et monétaires, forte contestation en France de la loi El Khomri concernant le remodelage de la réglementation du travail.

Nous verrons donc si l'idéologie régnante prenant parti majoritairement pour les entreprises donc pour la politique de l'offre dans un contexte de mondialisation perdure ou au contraire se retrouve complètement inefficace et non performante.

