



Projet

MODS202 - Econométrie

Présenter par :

Elie NAKAD

Hussein CHAHROUR

Novembre 2022

Superviseur :

Pr. Patrick WAELEBROECK

Table of Contents

1	Partie I – Régression	6
1.1	Question 1	6
1.2	Question 2	6
1.2.1	L'ensemble des femmes	6
1.2.2	Les femmes dont le salaire du mari est supérieur à la médiane de l'échantillon.....	7
1.2.3	Les femmes dont le salaire du mari est inférieur à la médiane de l'échantillon	7
1.3	Question 3	8
1.4	Question 4	8
1.5	Question 5	9
1.6	Question 6	10
1.7	Question 7	10
1.8	Question 8	12
1.9	Question 9	12
1.10	Question 10	13
1.11	Question 11	13
1.12	Question 12	13
1.13	Question 13	14
1.14	Question 14	15
1.15	Question 15	16
1.16	Question 16	16
2	Partie II – Séries temporelles.....	17
2.1	Question 1	17
2.2	Question 2	17
2.3	Question 3	18
2.4	Question 4	18
2.5	Question 5	19
2.6	Question 6	19
2.7	Question 7	20
2.8	Question 8	21
2.9	Question 9	21

2.10 Question 10 23

2.11 Question 11 23

Table of Figures

Figure 1: Donnée avec "wage" strictement positive	6
Figure 2: Descriptive du salaire, âge et éducation de l'ensemble des femmes	6
Figure 3: Descriptive du salaire, âge et éducation des femmes dont le salaire du mari est supérieur à la médiane	7
Figure 4: Descriptive du salaire, âge et éducation des femmes dont le salaire du mari est inférieur la médiane	7
Figure 5: Histogramme de la variable wage	8
Figure 6: Histogramme du log de 'wage'	8
Figure 7: Corrélation entre éducation du père et celle de la mère	8
Figure 8: Nuage de points entre salaire et niveau d'éducation du père	9
Figure 9: Corrélation entre salaire et niveau d'éducation du père	9
Figure 10: Corrélation entre salaire et niveau d'expérience	9
Figure 11: Nuage de points entre salaire et niveau d'expérience	9
Figure 12: Nuage de points entre salaire et niveau d'éducation du père	10
Figure 13: Corrélation entre salaire et niveau d'éducation du père	10
Figure 14: Résultats de la régression	11
Figure 15: Histogramme des résidus	11
Figure 16: Histogramme des résidus (q8)	12
Figure 17: Résultats de la régression (q7)	12
Figure 18: Test d'hypothèse pour seuil de 1%, 5% et 10%	12
Figure 19: Résultats à 5%	13
Figure 20: Statistique de Fisher	13
Figure 21: Salaire en fonction de l'éducation	14
Figure 22: Nuage de points salaire en fonction de l'éducation	14
Figure 23: Salaire en fonction d'années d'expérience	14
Figure 24: Nuage de points salaire en fonction d'années d'expérience	14
Figure 25: Résultats de la régression (q13)	15
Figure 26: Résultats de la régression (q14)	15
Figure 27: p-valeurs et f-stat	16
Figure 28: p-valeurs et f-stat (q16)	16
Figure 29: Information sur les données	17
Figure 30: Le taux inflation en fonction du temps	17
Figure 31: L'autocorrélogramme et l'autocorrélogramme partiel d'inflation	18
Figure 32: AIC et BIC	19
Figure 33: Taux de chômage en fonction de l'inflation	19
Figure 34: Courbe de Philips	19
Figure 35: Résultats de la regression (q6)	20
Figure 36: Durbin Watson et resultats de la regression(q7)	20
Figure 37: Résultats de la régression du modèle corrigé	21
Figure 38: Résultats de la régression sur les erreurs du modèle corrigé	21

Figure 39: Résultats de la régression sur la première fenêtre de l'échantillon	21
Figure 40: Evolution des taux d'inflation et taux de chômage sur la première fenêtre de l'échantillon...	22
Figure 41: Résultats de la régression sur la deuxième fenêtre de l'échantillon	22
Figure 42: Evolution des taux d'inflation et taux de chômage sur la deuxième fenêtre de l'échantillon..	22
Figure 43: Tableau des valeurs.....	23
Figure 44: Résultats de la régression avec toutes les variables	23
Figure 45: Résultats de la régression sans la variable de chômage	23
Figure 46: Délais distribués.....	24
Figure 47: Délais d'inflation et de chômage	24

1 Partie I – Régression

1.1 Question 1

- Lire le fichier `mroz.txt`. Ne sélectionner que les observations pour lesquelles la variable **wage** est strictement positive.

	0	1	2	3	4	5	6	7	8	9	...	12	13	14	15	16	17	18	19	20	21
0	1	1610	1	0	32	12	3.3540	2.65	2708	34	...	16310	0.7215	12	7	5.0	0	14	10.910060	1.210154	196
1	1	1656	0	2	30	12	1.3889	2.65	2310	30	...	21800	0.6615	7	7	11.0	1	5	19.499980	.3285121	25
2	1	1980	1	3	35	12	4.5455	4.04	3072	40	...	21040	0.6915	12	7	5.0	0	15	12.039910	1.514138	225
3	1	456	0	3	34	12	1.0965	3.25	1920	53	...	7300	0.7815	7	7	5.0	0	6	6.799996	.0921233	36
4	1	1568	1	2	31	14	4.5918	3.60	2000	32	...	27300	0.6215	12	14	9.5	1	7	20.100060	1.524272	49
...
423	1	680	0	5	36	10	2.3118	0.00	3430	43	...	19772	0.7215	7	7	7.5	0	2	18.199980	.8380265	4
424	1	2450	0	1	40	12	5.3061	6.50	2008	40	...	35641	0.6215	7	7	5.0	1	21	22.641060	1.668857	441
425	1	2144	0	2	43	13	5.8675	0.00	2140	43	...	34220	0.5815	7	7	7.5	1	22	21.640080	1.769429	484
426	1	1760	0	1	33	12	3.4091	3.21	3380	34	...	30000	0.5815	12	16	11.0	1	14	23.999980	1.226448	196
427	1	490	0	1	30	12	4.0816	2.46	2430	33	...	18000	0.6915	12	12	7.5	1	7	16.000020	1.406489	49

428 rows × 22 columns

Figure 1: Donnée avec "wage" strictement positive

La colonne 6 représente la variable wage. On commence par nettoyer en enlevant les lignes ayant comme valeur «.» pour la variable wage puis changer le type à « float » pour faciliter la comparaison. On obtient 428 observations, avec 22 variables chacune (représenté par les colonnes).

1.2 Question 2

- Faire les statistiques descriptives du salaire, de l'âge et de l'éducation pour :

1.2.1 L'ensemble des femmes

Le salaire moyen des femmes est de 4.17 (par heure). L'âge moyen est 41.97 ans. Le nombre d'années d'éducation moyen est de 12.66 années.

	wage	age	educ
count	428.000000	428.000000	428.000000
mean	4.177682	41.971963	12.658879
std	3.310282	7.721084	2.285376
min	0.128200	30.000000	5.000000
25%	2.262600	35.000000	12.000000
50%	3.481900	42.000000	12.000000
75%	4.970750	47.250000	14.000000

Figure 2: Descriptive du salaire, âge et éducation de l'ensemble des femmes

1.2.2 Les femmes dont le salaire du mari est supérieur à la médiane de l'échantillon

	wage	age	educ
count	214.000000	214.000000	214.000000
mean	4.896822	42.275701	13.242991
std	4.041606	7.388843	2.359045
min	0.161600	30.000000	5.000000
25%	2.513850	36.000000	12.000000
50%	3.846400	43.000000	12.000000
75%	5.854125	48.000000	16.000000
max	25.000000	59.000000	17.000000

Figure 3: Descriptive du salaire, âge et éducation des femmes dont le salaire du mari est supérieur à la médiane

Le salaire moyen des femmes dont le mari a un salaire plus que la médiane est de 4.9. L'âge moyen est 42.27 ans, et le nombre d'années d'éducation moyen est de 13.24 années.

1.2.3 Les femmes dont le salaire du mari est inférieur à la médiane de l'échantillon

	wage	age	educ
count	214.000000	214.000000	214.000000
mean	3.458541	41.668224	12.074766
std	2.143274	8.045482	2.054200
min	0.128200	30.000000	6.000000
25%	2.117275	35.000000	12.000000
50%	2.971800	41.000000	12.000000
75%	4.393800	47.000000	12.000000
max	18.267000	60.000000	17.000000

Figure 4: Descriptive du salaire, âge et éducation des femmes dont le salaire du mari est inférieur la médiane

Le salaire moyen des femmes dont le mari a un salaire moins que la médiane est de 3.46. L'âge moyen est 41.67 ans, et le nombre d'années d'éducation moyen est de 12.07 années.

- On peut remarquer que les femmes, dont le mari gagne plus que la médiane, gagnent en moyenne plus que dans le cas contraire.
- D'autre part, l'écart-type (std) est plus élevé pour les femmes dont le mari gagne plus que la médiane.
- On peut supposer l'existence de deux sous-catégories :
 - Les femmes ayant bénéficié de moins d'éducation et disposant d'un salaire inférieur

- Les femmes plus jeunes, et dont le mari est probablement plus jeune, qui gagnent moins que la médiane.

1.3 Question 3

- Faire l'histogramme de la variable wage. Calculer le log de wage et faire l'histogramme.

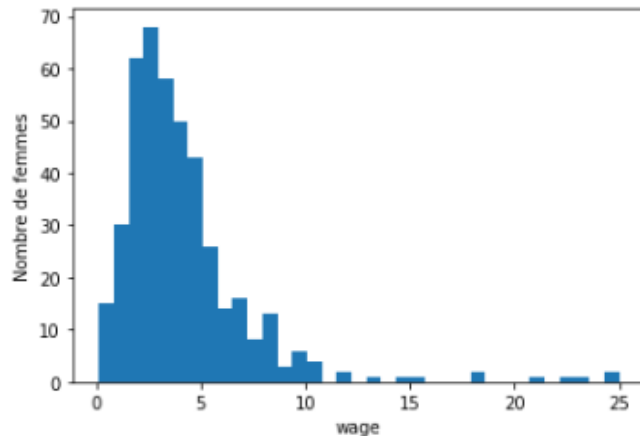


Figure 5: Histogramme de la variable wage

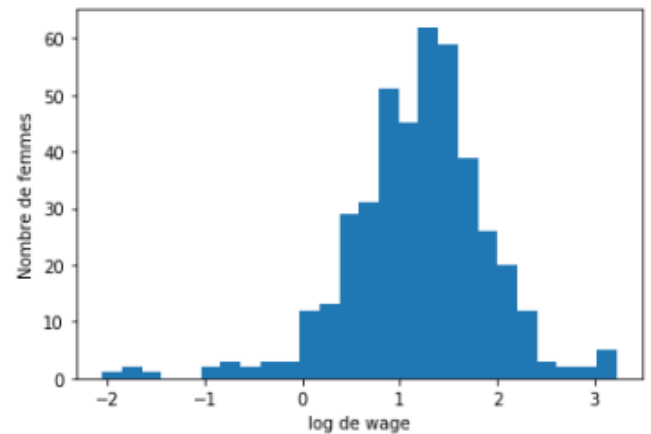


Figure 6: Histogramme du log de 'wage'

- Comparez les deux histogrammes et commentez.

On remarque que le deuxième histogramme est plus proche de la fonction normale que le premier. L'application de log sur les salaires permet de centrer les valeurs autour de la valeur de $1.4 \simeq \log(4)$ et ainsi de résoudre certains problèmes de la variable « wage ».

1.4 Question 4

- Calculer les corrélations motheduc et fatheduc.

Corrélation entre éducation du père et celle de la mère est : 0.5540632184311678

- Commentez.

La corrélation est de 0.55 (55%) entre l'éducation de la mère et celle du père. (P-value presque nulle)

Cela est logique puisque les deux personnes au sein d'un couple appartiennent souvent à une classe sociale similaire et bénéficient des mêmes possibilités d'accès à l'éducation.

- Il y a-t-il un problème de multicollinéarité si l'on utilise ces variables comme explicatives ?

Puisque la Corrélation est supérieure à 0.5, on ne doit pas les utiliser tous les deux comme variables explicatives et créera un problème de multicollinéarité. Prendre les deux variables en considération peut surpondérer l'information sur l'éducation des parents.

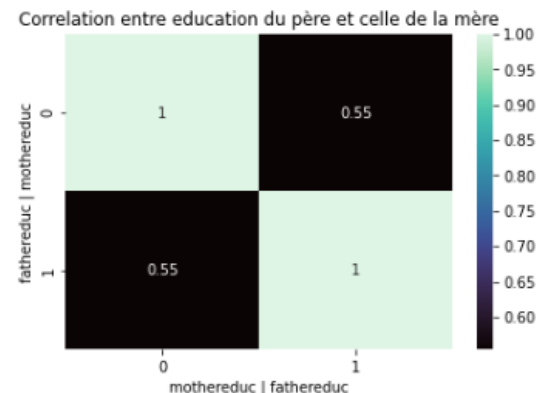


Figure 7: Corrélation entre éducation du père et celle de la mère

1.5 Question 5

- Faites un graphique en nuage de point entre wage et educ, wage et exper, wage et fatheduc. Commentez. S'agit-il d'un effet "toute chose étant égale par ailleurs ?"

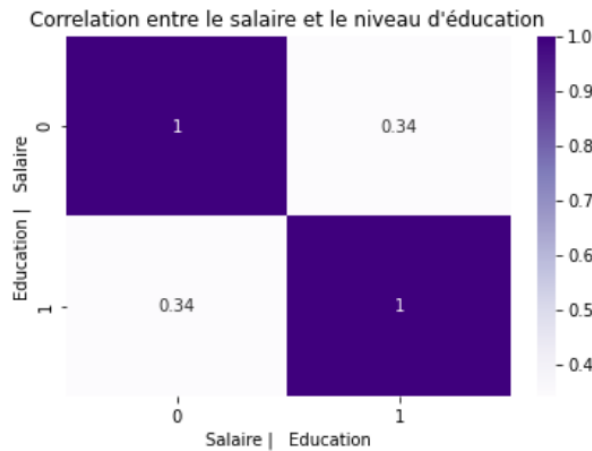


Figure 8: Corrélation entre salaire et niveau d'éducation du père

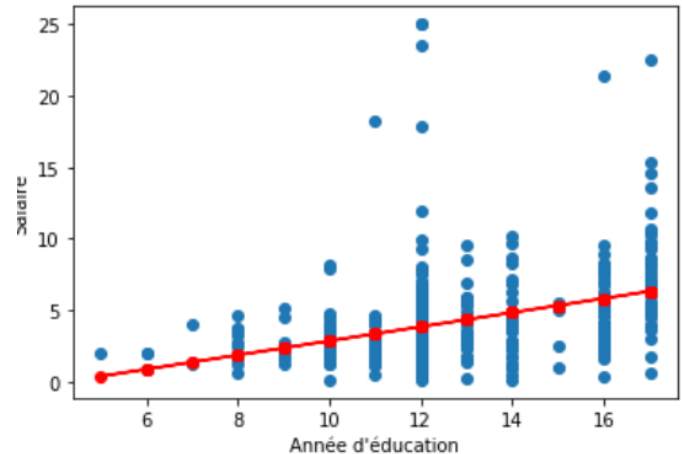


Figure 9: Nuage de points entre salaire et niveau d'éducation du père

- L'intercorrélation entre la variable de salaire et l'éducation est $0.34 \leq 0.5$. De ce fait, les deux variables ne sont pas corrélées. Le salaire augmente avec les années d'éducation. Par contre, on peut avoir un problème d'hétéroscédacité puisque la variance est non uniforme (des pics sont présents).
- Il ne s'agit pas d'un effet "toute chose étant égale par ailleurs", car pour chaque donnée, le reste des variables ne sont pas constantes.

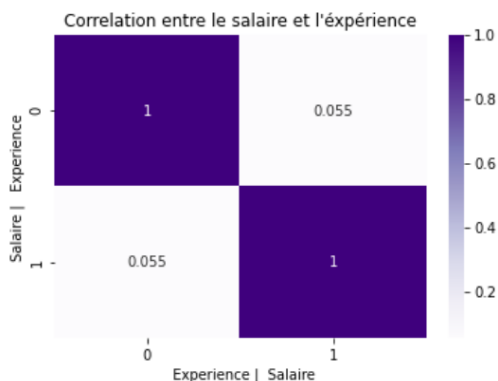


Figure 10: Corrélation entre salaire et niveau d'expérience

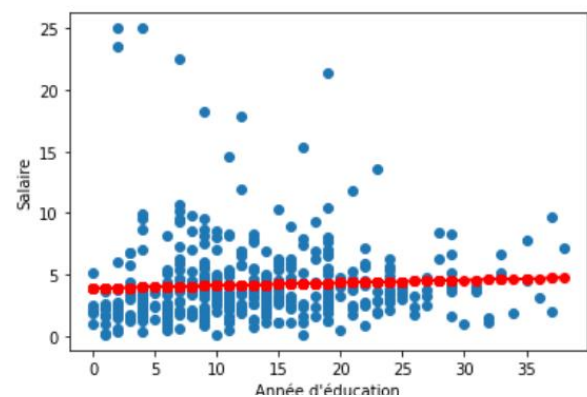


Figure 11: Nuage de points entre salaire et niveau d'expérience

- L'intercorrélation entre la variable de salaire et l'expérience est $0.05 \leq 0.5$ moins signifiant que le cas précédent. De ce fait, les deux variables sont faiblement corrélées. Le salaire ne dépend pas des années d'expérience.
- Il existe un biais par rapport au nombre d'années d'éducation en amont.

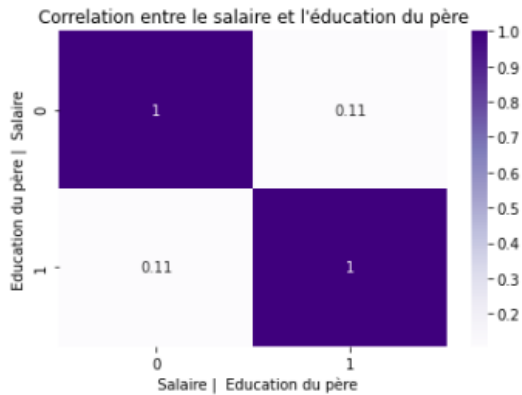


Figure 12: Corrélation entre salaire et niveau d'éducation du père

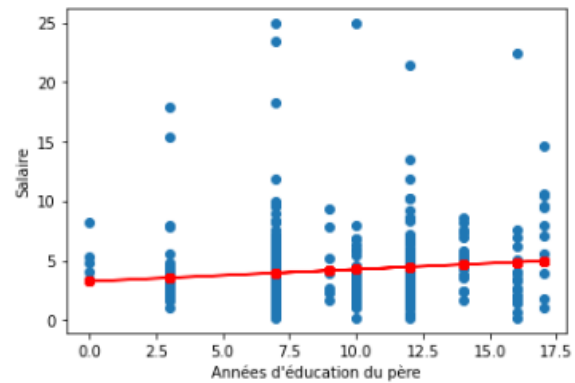


Figure 13: Nuage de points entre salaire et niveau d'éducation du père

- L'intercorrélation entre la variable de salaire et l'éducation du père est $0.11 \leq 0.5$. De ce fait, les deux variables sont faiblement corrélées aussi. Le salaire dépend légèrement de l'éducation du père. Cette influence reste désormais plus faible que celle de l'éducation propre de la personne.
- Des années d'éducation supplémentaires du père semblent apporter un meilleur salaire aux enfants. Cependant, la variance ne semble pas uniforme selon le nombre d'années d'études du père.

1.6 Question 6

- Quelle est l'hypothèse fondamentale qui garantit des estimateurs non biaisés ?

L'hypothèse de normalité des résidus garantit l'obtention du meilleur estimateur linéaire non biaisé (Si $E(x|u) = 0$, alors l'estimateur est non biaisé). Ainsi, les résidus sont centrés en zéro, de variance constante à travers le temps (iid). On parle alors d'homoscédasticité.

- Expliquer le biais de variable omise.

Une variable qui n'est pas incluse dans le modèle se trouve sans doute dans le résidu. On estime un modèle en prenant certaines variables, mais ils en existent d'autres, que l'on ne peut pas contrôler par manque de données. Une variable omise viole l'hypothèse de normalité des résidus, car l'effet des variables omises se retrouve en partie dans les résidus. On peut donc avoir un problème d'hétéroscédasticité.

1.7 Question 7

- Faire la régression de wage en utilisant les variables explicatives une constante, city, educ, exper, nwifeinc, kidslt6, kidsgt6. Commentez l'histogramme des résidus.
 - On a : $wage = c + \beta_1 city + \beta_2 educ + \beta_3 exper + \beta_4 nwifeinc + \beta_5 kidslt6 + \beta_6 kidsgt6$
 - D'après la figure 14 des résultats de la régression, seule l'éducation apparaît comme variable significative quand on utilise une constante
 - D'après l'histogramme des résidus (figure 15), la distribution des erreurs n'est pas normale. Pourtant, les erreurs sont condensées auprès de 0. Par contre, un nombre non

négligeable d'observations présente une erreur non nulle et aussi importante (atteint 20 dans certains cas).

OLS Regression Results						
=====						
Dep. Variable:	6	R-squared:	0.127			
Model:	OLS	Adj. R-squared:	0.115			
Method:	Least Squares	F-statistic:	10.23			
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	1.41e-10			
Time:	21:12:17	Log-Likelihood:	-1090.0			
No. Observations:	428	AIC:	2194.			
Df Residuals:	421	BIC:	2222.			
Df Model:	6					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-2.4035	0.963	-2.495	0.013	-4.297	-0.510
x1	0.3698	0.327	1.132	0.258	-0.272	1.012
x2	0.4600	0.070	6.546	0.000	0.322	0.598
x3	0.0238	0.021	1.141	0.255	-0.017	0.065
x4	0.0152	0.015	0.984	0.326	-0.015	0.046
x5	0.0362	0.397	0.091	0.927	-0.744	0.816
x6	-0.0619	0.125	-0.494	0.622	-0.308	0.185
=====						
Omnibus:	345.825	Durbin-Watson:	2.056			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6499.375			
Skew:	3.389	Prob(JB):	0.00			
Kurtosis:	20.847	Cond. No.	178.			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 14: Résultats de la régression

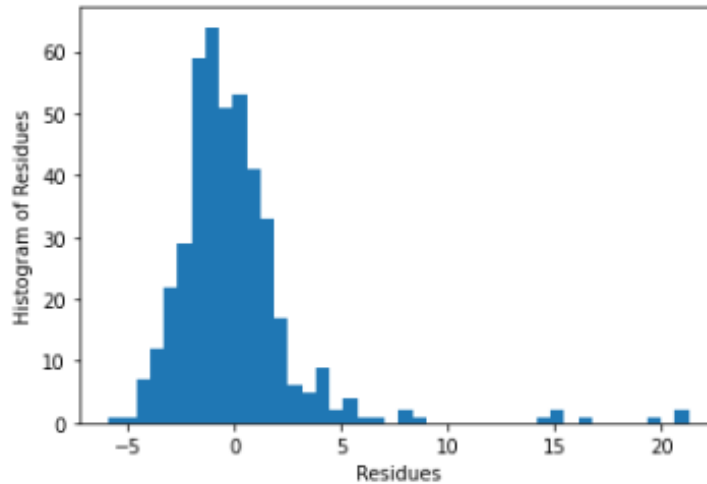


Figure 15: Histogramme des résidus

1.8 Question 8

- Faire la régression de *lwage* sur une constante, *city*, *educ*, *exper*, *nwifeinc*, *kidslt6*, *kidsgt6*. Comparer l'histogramme obtenu à celui de la question 7.

OLS Regression Results

```

=====
Dep. Variable:          6      R-squared:          0.156
Model:                 OLS      Adj. R-squared:     0.144
Method:                Least Squares      F-statistic:    12.92
Date:                  Mon, 22 Nov 2021      Prob (F-statistic): 2.00e-13
Time:                  01:51:07      Log-Likelihood:   -431.92
No. Observations:      428      AIC:              877.8
Df Residuals:          421      BIC:              906.3
Df Model:              6
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-0.3990	0.207	-1.927	0.055	-0.806	0.008
x1	0.0353	0.070	0.503	0.616	-0.103	0.173
x2	0.1022	0.015	6.771	0.000	0.073	0.132
x3	0.0155	0.004	3.452	0.001	0.007	0.024
x4	0.0049	0.003	1.466	0.143	-0.002	0.011
x5	-0.0453	0.085	-0.531	0.596	-0.213	0.122
x6	-0.0117	0.027	-0.434	0.664	-0.065	0.041

```

=====
Omnibus:              79.542      Durbin-Watson:      1.979
Prob(Omnibus):        0.000      Jarque-Bera (JB):    287.193
Skew:                 -0.795      Prob(JB):            4.33e-63
Kurtosis:             6.685      Cond. No.            178.
=====

```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 16 : Résultats de la régression (q7)

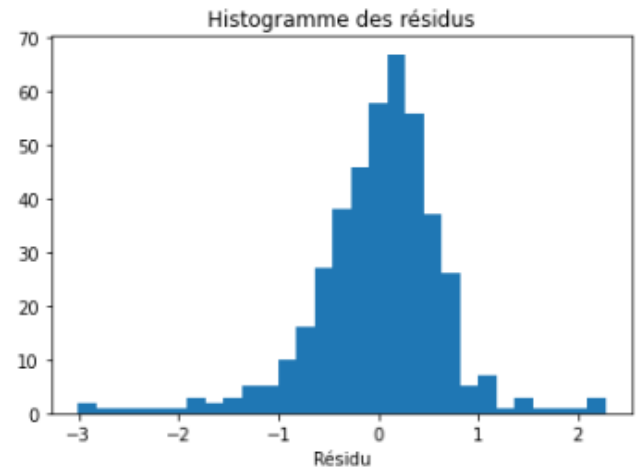


Figure 17 : Histogramme des résidus (q8)

- On a : $\log(\text{wage}) = c + \beta_1 \text{city} + \beta_2 \text{educ} + \beta_3 \text{exper} + \beta_4 \text{nwifeinc} + \beta_5 \text{kidslt6} + \beta_6 \text{kidsgt6}$
- La distribution des erreurs est quasi normale. Les erreurs sont centrées auprès de 0. Par contre, on a moins d'étalement par rapport à 0, la variance de l'erreur est presque constante ce qui est différent à l'histogramme précédent. On a donc une « Homoscédasticité » approximative, ce qui nous amènera à adopter cette transformation dans la suite.
- Les variables éducation et expérience apparaissent alors comme significatives.

1.9 Question 9

- Tester l'hypothèse de non significativité de *nwifeinc* avec un seuil de significativité de 1%, 5% et 10% (test alternatif des deux côtés).

```

threshold 1 %
critical threshold 2.3352378938061142
p value= 0.14340791214517742
student value 1.465951741624178
0.01 < 0.1434 we accept the hypothesis

threshold 5 %
critical threshold 1.6484810571255268
p value= 0.14340791214517742
student value 1.465951741624178
0.05 < 0.1434 we accept the hypothesis

threshold 10 %
critical threshold 1.2835656801791953
p value= 0.14340791214517742
student value 1.465951741624178
0.1 < 0.1434 we accept the hypothesis

```

Figure 18 : Test d'hypothèse pour seuil de 1%, 5% et 10%

- Le test d'hypothèse qu'on a est le suivant : $H_0 : \beta_4 = 0$ et $H_1 : \beta_4 \neq 0$
- Commentez les p-values.
 - La p value associé à la variable nwifeinc étant large, on accepte l'hypothèse de la non significativité de cette variable dans tous les cas.

1.10 Question 10

- Tester l'hypothèse que le coefficient associé à nwifeinc est égal à 0.01 avec un seuil de significativité de 5% (test à alternatif des deux côtés)
 - Le test d'hypothèse est : $H_0 : \beta_4 = 0.01$ et $H_1 : \beta_4 \neq 0.01$
 - On utilise : $t_{nwifeinc} = \frac{nwifeinc - 0.01}{Se(nwifeinc)}$

t value -1.5363887846744726
p-value 0.12519423515110772

Figure 19: Résultats à 5%

- On que 2.5% est inferieur à p= 12.5% --> on accepte l'hypothèse nulle H_0 .

1.11 Question 11

- Tester l'hypothèse jointe que le coefficient de nwifeinc est égal à 0.01 et que celui de city est égal à 0.05.

- Le test d'hypothèse est :

$$H_0 : \beta_4 = 0.01 \text{ et } \beta_1 = 0.05$$

$$H_1 : \beta_4 \neq 0.01 \text{ et } \beta_1 \neq 0.05$$

- La statistique de Fisher est calculée par :

$$F = \frac{\frac{SSR_{constrained} - SSR_{unconstrained}}{q}}{\frac{SSR_{unconstrained}}{(n - k - 1)}}$$

- On obtient avec q=2:

SSR0 = 188.589973
SSR1 = 189.787874
F = 1.337070
columns in X0 7
p-value for fisher: 0.263727

We get p-value > 5% thus we accept the null hypothesis

Figure 20: Statistique de Fisher

1.12 Question 12

- Faites une représentation graphique de la manière dont le salaire augmente avec l'éducation et l'expérience professionnelle. Commentez

Augmentation du salaire avec le nombre d'année d'éducation

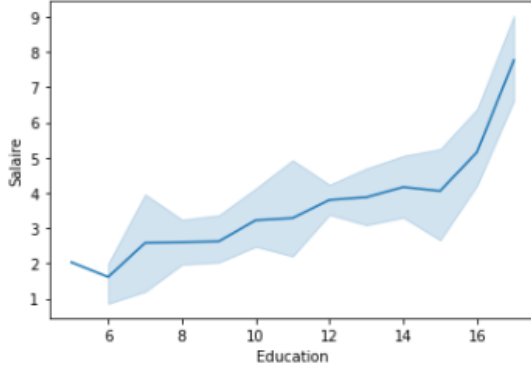


Figure 21: Salaire en fonction de l'éducation

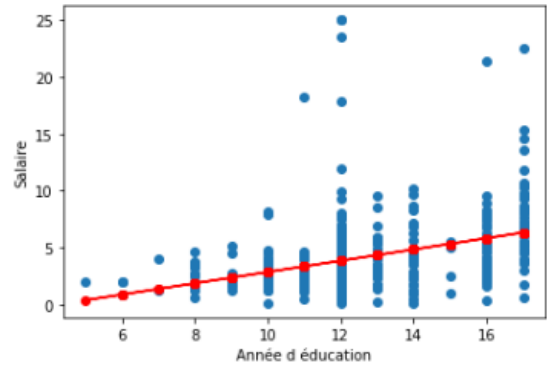


Figure 22: Nuage de points salaire en fonction de l'éducation

Augmentation du salaire avec le nombre d'année d'expérience

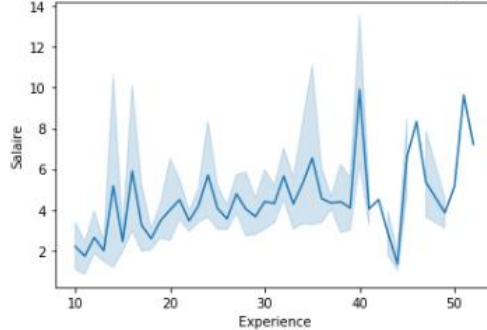


Figure 23: Salaire en fonction d'années d'expérience

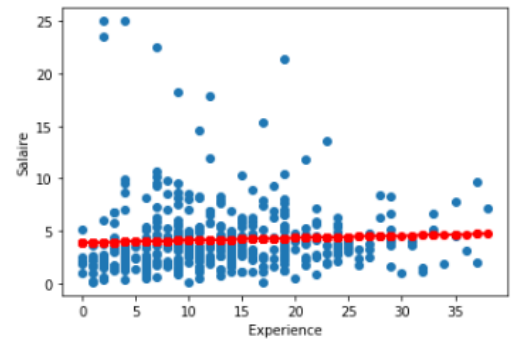


Figure 24: Nuage de points salaire en fonction d'années d'expérience

- Le salaire augmente avec le niveau d'éducation d'après les figures 21 et 22. Le salaire ne dépend pas réellement de l'expérience d'après les figures 23 et 24.
- De plus, si nous fixons une valeur pour l'une des variables explicatives, nous pouvons voir que la régression est en relation positive avec l'autre. Par exemple, fixer une valeur pour le nombre d'années de l'éducation, nous observons que plus le nombre d'années d'expérience est élevé, plus le salaire.

1.13 Question 13

- Tester l'égalité des coefficients associés aux variables *kidslt6* et *kidsgt6*. Interprétez.
 - Pour tester si les coefficients de *kidslt6* et *kidsgt6* sont égaux, on fait une régression de y sur $\text{const, city, educ, exper, nwfeinc, } 10 \text{ kidslt6, kidslt6+kidsgt6}$.
 - On pose : $y = \log(\text{wage}) = c + \beta_1 \text{city} + \beta_2 \text{educ} + \beta_3 \text{exper} + \beta_4 \text{nwfeinc} + \beta_5 \text{kidslt6} + \beta_6 \text{kidsgt6}$
 - On peut écrire : $\theta = \beta_6 - \beta_5$ alors $\theta + \beta_5 = \beta_6$ donc :

$$\log(\text{wage}) = c + \beta_1 \text{city} + \beta_2 \text{educ} + \beta_3 \text{exper} + \beta_4 \text{nwfeinc} + \beta_5 (\text{kidslt6} + \text{kidsgt6}) + \theta \text{kidsgt6}$$

- On teste : $H_0 : \theta = 0$ et $H_1 : \theta \neq 0$

```

=====
                        OLS Regression Results
=====
Dep. Variable:          6      R-squared:          0.156
Model:                  OLS      Adj. R-squared:       0.144
Method:                 Least Squares      F-statistic:       12.92
Date:                   Sat, 26 Nov 2022      Prob (F-statistic):    2.00e-13
Time:                   18:27:32      Log-Likelihood:      -431.92
No. Observations:      428      AIC:                877.8
Df Residuals:          421      BIC:                906.3
Df Model:              6
Covariance Type:       nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const         -0.3990      0.207      -1.927      0.055      -0.806      0.008
x1             0.0353      0.070       0.503      0.616      -0.103      0.173
x2             0.1022      0.015       6.771      0.000      0.073      0.132
x3             0.0155      0.004       3.452      0.001      0.007      0.024
x4             0.0049      0.003       1.466      0.143      -0.002      0.011
x5            -0.0336      0.090      -0.372      0.710      -0.211      0.144
x6            -0.0117      0.027      -0.434      0.664      -0.065      0.041
=====
Omnibus:              79.542      Durbin-Watson:       1.979
Prob(Omnibus):        0.000      Jarque-Bera (JB):    287.193
Skew:                 -0.795      Prob(JB):            4.33e-63
Kurtosis:             6.685      Cond. No.            178.
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
We get p-value: 0.710268, > 0.05 we accept the hypothesis

```

Figure 25: Résultats de la régression (q13)

- On obtient p-value : 0.710268, > 0.05 on accepte l'hypothèse.

1.14 Question 14

- En utilisant le modèle de la question 7, faire le test d'hétéroscédasticité de forme linéaire en donnant la p-valeur. Déterminer la ou les sources d'hétéroscédasticité et corriger avec les méthodes vues en cours. Comparer les écarts-types des coefficients estimés avec ceux obtenus à la question 7. Commenter.
 - Soit l'hypothèse H_0 : donnés à des sources d'homoscédasticité
 H_1 : donnés à des sources d'hétéroscédasticité
 - On suppose : $u^2 = \delta_0 + \delta_1 x_1 + \dots \delta_k x_k + v$
 - Alors : $H_0: \delta_1 = \delta_2 = \dots = \delta_k = 0$

```

=====
                        OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.022
Model:                  OLS      Adj. R-squared:       0.008
Method:                 Least Squares      F-statistic:       1.593
Date:                   Sat, 26 Nov 2022      Prob (F-statistic):    0.148
Time:                   18:27:32      Log-Likelihood:      -2207.4
No. Observations:      428      AIC:                4429.
Df Residuals:          421      BIC:                4457.
Df Model:              6
Covariance Type:       nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const         1.4856      13.111       0.113      0.910      -24.285      27.256
x1             5.9644      4.444       1.342      0.180      -2.770      14.699
x2             0.8077      0.956       0.845      0.399      -1.072      2.687
x3            -0.5341      0.284      -1.880      0.061      -1.093      0.024
x4             0.0435      0.211       0.206      0.837      -0.371      0.458
x5             4.9573      5.402       0.918      0.359      -5.661      15.575
x6            -0.4018      1.706      -0.236      0.814      -3.756      2.952
=====
Omnibus:              638.793      Durbin-Watson:       2.029
Prob(Omnibus):        0.000      Jarque-Bera (JB):    96122.227
Skew:                 8.127      Prob(JB):            0.00
Kurtosis:             74.595      Cond. No.            178.
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Figure 26: Résultats de la régression (q14)

- P-value est égale à 14.8% donc on accepte H0 sous 5%. (On ne peut pas déduire d'hétéroscédasticité) L'erreur dépend principalement de la première et la cinquième variable explicatives donc l'hétéroscédasticité provient des variables city et kidslt6.

1.15 Question 15

- *Tester le changement de structure de la question 8 entre les femmes qui ont plus de 43 ans et les autres : test sur l'ensemble des coefficients. Commentez et donnez les p-valeurs.*

- Le test d'hypothèse est : H_0 : Il y a changement de structure
 H_1 : Il n'y a pas de changement de structure

- Le test statistique est calculée (pour deux groupes) :

$$F_{Chow} = \frac{SSR_c - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \times \frac{n_1 + n_2 - 2k}{k}$$

- On obtient : $n_1=211$, $n_2=217$ et $k=7$

f-stat: 0.8260377871071876
p-value: 0.5660361961109925

Figure 27: p-valeurs et f-stat

- P-value = 0.566 donc à 5% donc on en peut donc pas rejeter l'hypothèse nulle. Il n'y a donc pas de changement de structure entre les deux groupes d'individus considérés.

1.16 Question 16

- *Refaire la question 15 en supposant que seuls les rendements de l'éducation et de l'expérience professionnelle changent selon l'âge de la femme. Formuler l'hypothèse H0 et tester-la. Donnez la p-valeur.*

- Le test d'hypothèse est : H_0 : Il n'y a pas de changement concernant l'éducation et l'expérience professionnelle lorsqu'on sépare les données en fonction de l'âge
- On prend : $y = \log(wage) = c + \beta_1 educXeduc + \beta_2 experXexper$
- On obtient :

f_stat is 2.045890086920155 and p_value is 0.10681389474947503

Figure 28: p-valeurs et f-stat (q16)

- La p-valeur du test de Chow est 0.106, on ne peut donc pas rejeter l'hypothèse nulle. Il n'y a donc pas de changement de structure entre les deux groupes d'individus considérés sur l'éducation et l'expérience.

2 Partie II – Séries temporelles

2.1 Question 1

- Importer les données du fichier *quarterly.xls* (corriger le problème éventuel d'observations manquantes)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 212 entries, 0 to 211
Data columns (total 19 columns):
#   Column      Non-Null Count  Dtype
---  -
0   DATE         212 non-null   object
1   FFR          212 non-null   float64
2   Tbill        212 non-null   float64
3   Tbl1yr       212 non-null   float64
4   r5           212 non-null   float64
5   r10          212 non-null   float64
6   PPINSA       212 non-null   float64
7   Finished     212 non-null   float64
8   CPI          212 non-null   float64
9   CPICORE      212 non-null   float64
10  M1NSA        212 non-null   float64
11  M2SA         212 non-null   float64
12  M2NSA        212 non-null   float64
13  Unemp        212 non-null   float64
14  IndProd      212 non-null   float64
15  RGDP         212 non-null   float64
16  Potent       212 non-null   float64
17  Deflator     212 non-null   float64
18  Curr         212 non-null   float64
dtypes: float64(18), object(1)
memory usage: 31.6+ KB
```

Figure 29: Information sur les données

- On remarque que pour les 212 entrées, aucune valeur est nulle (manquantes).

2.2 Question 2

- Calculer inf , le taux d'inflation à partir de la variable CPI .

Par définition le taux d'inflation est donné par : $Inf_i = 100 \frac{CPI_{i+1} - CPI_i}{CPI_i}$

- Faire un graphique dans le temps de inf . Commentez.

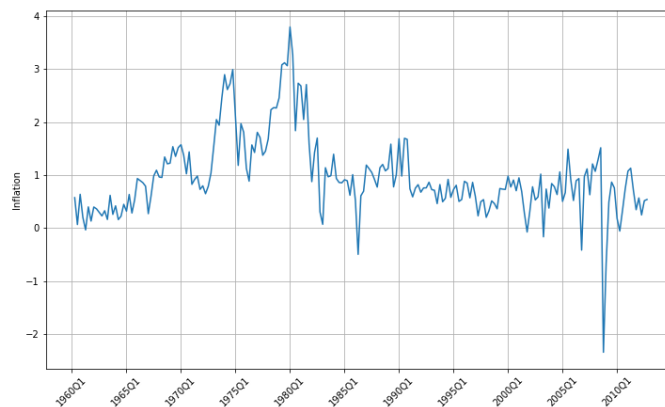


Figure 30: Le taux inflation en fonction du temps

- L'inflation est une série temporelle non stationnaire. Elle était élevée début des années 1980, puis a chuté au cours des années. On remarque une déflation sur la période de crise financière de 2008. Le régime d'inflation semblait plus instable et fluctue énormément avant les années 90, puis contrôlé entre 1990 et 2008, avant la crise financière.

2.3 Question 3

- *Interpréter l'autocorrélogramme et l'autocorrélogrammes partiels de inf.*

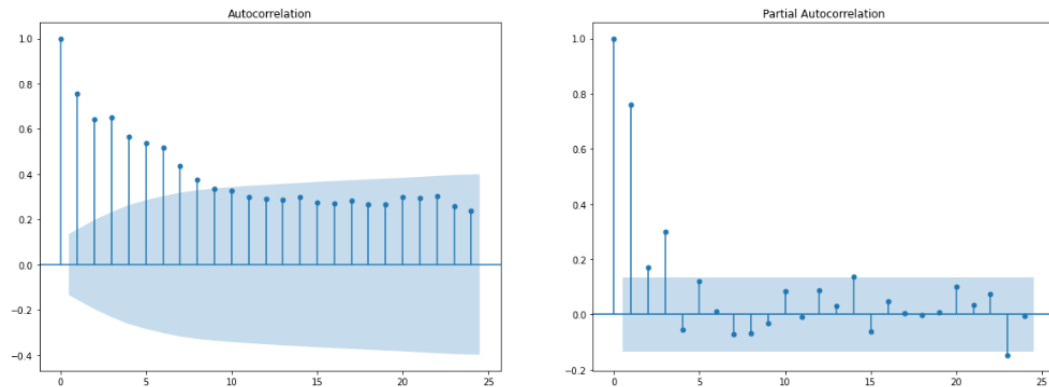


Figure 31: L'autocorrélogramme et l'autocorrélogramme partiel d'inflation

- L'autocorrélogramme indique que l'autocorrélation diminue avec le temps comme dans un processus ARMA(p,q)
 - L'autocorrélogramme partiel oscille autour de 0 comme dans un processus de type Moving Average MA(1)
 - On peut alors supposer que la série n'est pas stationnaire (il y a une influence antérieure pour la détermination des valeurs actuelles)
- *Quelle est la différence entre ces deux graphiques ?*
 - L'autocorrélogramme donne l'influence d'une série à un temps $t-k$ dans le passé sur la valeur de la série au temps t , indépendamment du reste des observations.
 - L'autocorrélogramme partiel réalise la régression de toutes les valeurs présentes sur toutes les valeurs passées jusqu'au temps $t-k$.
 - Ainsi, on identifie les effets joints des différentes années.

2.4 Question 4

- Quelle est la différence entre la stationnarité et l'ergodicité ? Pourquoi a-t-on besoin de ces deux conditions
 - La stationnarité : Si $x(1) \dots x(n)$ et $x(1+h) \dots x(n+h)$ ont la même distribution donc le décalage n'affecte pas la loi associée. Ainsi les propriétés des séries temporelles restent les mêmes quel que soit la période du temps. La stationnarité est indispensable pour faire de l'estimation précise des paramètres d'un processus.

- L'ergodicité : Les observations se répètent dans le temps. Un processus ergodique est un processus stochastique pour lequel les statistiques peuvent être approchées par l'étude d'une seule réalisation suffisamment longue. Un processus non ergodique change de façon erratique à un rythme inconsistant. L'ergodicité est une hypothèse fondamentale pour l'étude des séries temporelles puisque les processus ne sont connus qu'à travers une de leurs réalisations. Il est alors possible de faire des estimations de paramètres (moyenne, variance...).
- Nous avons besoin de ces deux critères pour satisfaire les conditions de gauss-markov. (Appliquer la Loi des grands nombres et le Théorème de limite centrale).
- Autrement dit, on s'assure que la série ne diverge pas de son espérance.

2.5 Question 5

- Proposer une modélisation $AR(p)$ de π_t , en utilisant tous les outils vus au cours.

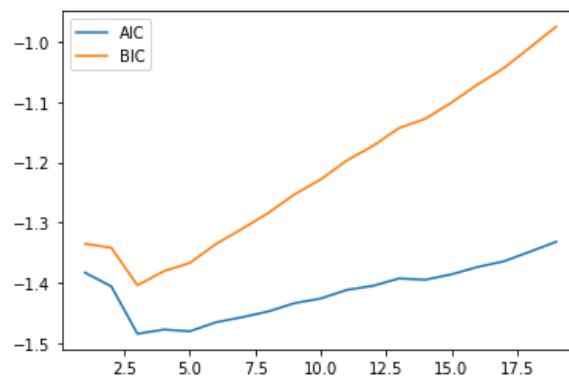


Figure 32: AIC et BIC

- Les critères AIC et BIC sont minimaux pour un nombre de lag=3. Un modèle $AR(3)$ semble être la meilleure modélisation de l'inflation.

2.6 Question 6

- Estimer le modèle de la courbe de Philips qui explique le taux de chômage ($Unemp$) en fonction du taux d'inflation courant et une constante.

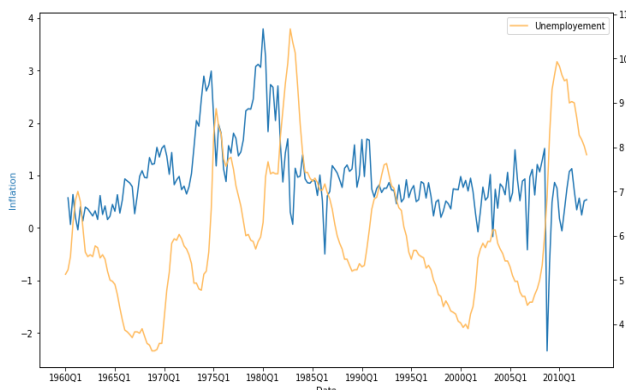


Figure 33: Taux de chômage en fonction de l'inflation

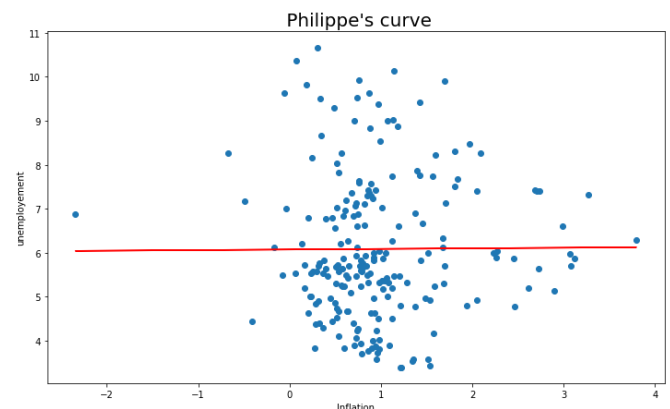


Figure 34: Courbe de Philips

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Unemp      R-squared:          0.000
Model:                  OLS        Adj. R-squared:     -0.005
Method:                 Least Squares  F-statistic:       0.008229
Date:                   Thu, 08 Dec 2022  Prob (F-statistic): 0.928
Time:                   16:04:24    Log-Likelihood:    -400.28
No. Observations:       211        AIC:              804.6
Df Residuals:           209        BIC:              811.3
Df Model:                1
Covariance Type:        nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
const          6.0735         0.182     33.388     0.000      5.715      6.432
INF            0.0134         0.148      0.091     0.928     -0.279      0.305
=====
Omnibus:             13.851   Durbin-Watson:       0.044
Prob(Omnibus):        0.001   Jarque-Bera (JB):     15.331
Skew:                 0.659   Prob(JB):             0.000469
Kurtosis:             2.935   Cond. No.             3.00
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Figure 35: Résultats de la regression (q6)

- Une augmentation de 1% de l'inflation entraîne une augmentation de 0.0232% du chômage.
- Il ne semble pas que la variable CPI soit significative à 5% étant donné que la p-value est supérieure à ce seuil.

2.7 Question 7

- Tester l'autocorrélation des erreurs.
 - Pour tester l'autocorrélation des erreurs, on peut appliquer le test de Durbin Watson sur les résidus.

```

Durbin Watson: 0.04411869153477157
-----

=====
                        OLS Regression Results
=====
Dep. Variable:          y      R-squared (uncentered): 0.956
Model:                  OLS    Adj. R-squared (uncentered): 0.956
Method:                 Least Squares  F-statistic:         4565.
Date:                   Fri, 09 Dec 2022  Prob (F-statistic): 5.77e-144
Time:                   10:43:47    Log-Likelihood:     -70.233
No. Observations:       210        AIC:               142.5
Df Residuals:           209        BIC:               145.8
Df Model:                1
Covariance Type:        nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
x1            0.9800         0.015     67.566     0.000      0.951      1.009
=====
Omnibus:             78.877   Durbin-Watson:       0.666
Prob(Omnibus):        0.000   Jarque-Bera (JB):     244.919
Skew:                 1.582   Prob(JB):             6.56e-54
Kurtosis:             7.240   Cond. No.             1.00
=====

Notes:
[1] R2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Figure 36: Durbin Watson et resultats de la regression(q7)

- Si le test d'autocorrélation de Durbin Watson vaut environ 2, il n'y a pas d'autocorrélation, s'il vaut proche de 0, il y a une autocorrélation positive, et s'il vaut environ 4, l'autocorrélation est négative.

- Ici, la valeur du test de DW est proche de 0. On en conclut que l'autocorrélation doit être positive.
- Si l'on regarde l'autocorrélogramme, on confirme ce constat.
- De plus p-value=0, on confirme ce constat de nouveau.

2.8 Question 8

- Corriger l'autocorrélation des erreurs par la méthode vue en cours.

OLS Regression Results						
Dep. Variable:	Unemp		R-squared:	0.000		
Model:	OLS		Adj. R-squared:	-0.005		
Method:	Least Squares		F-statistic:	0.01777		
Date:	Fri, 09 Dec 2022		Prob (F-statistic):	0.894		
Time:	11:01:03		Log-likelihood:	422.77		
No. Observations:	210		AIC:	-841.5		
Df Residuals:	208		BIC:	-834.9		
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.1214	0.004	33.233	0.000	0.114	0.129
INF	0.0198	0.148	0.133	0.894	-0.272	0.312
Omnibus:	14.332	Durbin-Watson:	0.044			
Prob(Omnibus):	0.001	Jarque-Bera (JB):	15.911			
Skew:	0.674	Prob(JB):	0.000351			
Kurtosis:	2.965	Cond. No.	66.2			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified

Figure 37: Résultats de la régression du modèle corrigé

OLS Regression Results					
=====					
Dep. Variable:	y	R-squared (uncentered):	0.956		
Model:	OLS	Adj. R-squared (uncentered):	0.956		
Method:	Least Squares	F-statistic:	4506.		
Date:	Fri, 09 Dec 2022	Prob (F-statistic):	6.34e-143		
Time:	11:01:03	Log-Likelihood:	746.51		
No. Observations:	209	AIC:	-1491.		
Df Residuals:	208	BIC:	-1488.		
Df Model:	1				
Covariance Type:	nonrobust				
=====					
	coef	std err	t	P> t	[0.025 0.975]

x1	0.9805	0.015	67.124	0.000	0.952 1.009
=====					
Omnibus:	77.833	Durbin-Watson:	0.667		
Prob(Omnibus):	0.000	Jarque-Bera (JB):	238.762		
Skew:	1.571	Prob(JB):	1.42e-52		
Kurtosis:	7.188	Cond. No.	1.00		

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
 [2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 38: Résultats de la régression sur les erreurs du modèle corrigé

- Il n'y a plus de corrélation entre $u(t)$ et $u(t-1)$.
- Nous pouvons donc utiliser ces valeurs corrigées pour la suite.

2.9 Question 9

- Tester la stabilité de la relation chômage-inflation sur deux sous-périodes de taille identique (test de changement de structure avant et après la moitié de la période d'observation).

OLS Regression Results						
=====						
Dep. Variable:	Unemp	R-squared:	0.007			
Model:	OLS	Adj. R-squared:	-0.003			
Method:	Least Squares	F-statistic:	0.7061			
Date:	Fri, 09 Dec 2022	Prob (F-statistic):	0.403			
Time:	11:01:03	Log-Likelihood:	-204.20			
No. Observations:	105	AIC:	412.4			
Df Residuals:	103	BIC:	417.7			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	5.9353	0.290	20.477	0.000	5.360	6.510
INF	0.1611	0.192	0.840	0.403	-0.219	0.541
=====						
Omnibus:	3.872	Durbin-Watson:	0.059			
Prob(Omnibus):	0.144	Jarque-Bera (JB):	3.769			
Skew:	0.461	Prob(JB):	0.152			
Kurtosis:	2.890	Cond. No.	3.49			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 39: Résultats de la régression sur la première fenêtre de l'échantillon

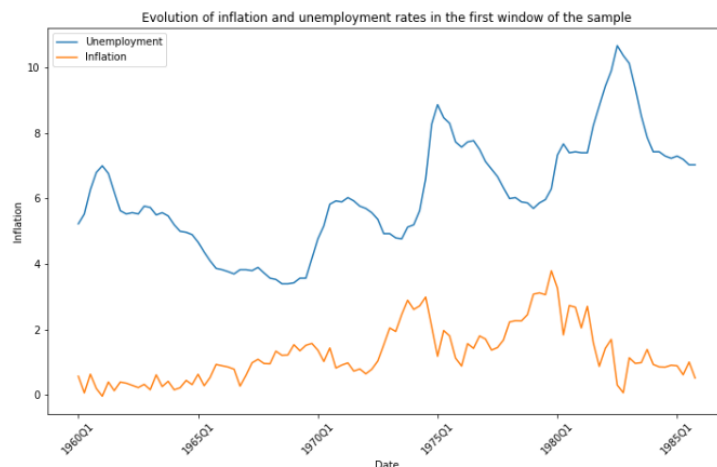


Figure 40: Evolution des taux d'inflation et du taux de chômage sur la première fenêtre de l'échantillon

OLS Regression Results						
Dep. Variable:	Unemp	R-squared:	0.031			
Model:	OLS	Adj. R-squared:	0.021			
Method:	Least Squares	F-statistic:	3.285			
Date:	Fri, 09 Dec 2022	Prob (F-statistic):	0.0728			
Time:	11:01:03	Log-Likelihood:	-193.37			
No. Observations:	106	AIC:	390.7			
Df Residuals:	104	BIC:	396.1			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	6.4223	0.258	24.925	0.000	5.911	6.933
INF	-0.5434	0.300	-1.812	0.073	-1.138	0.051
Omnibus:	12.693	Durbin-Watson:		0.082		
Prob(Omnibus):	0.002	Jarque-Bera (JB):		14.137		
Skew:	0.892	Prob(JB):		0.000851		
Kurtosis:	3.122	Cond. No.		3.24		

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 41: Résultats de la régression sur la deuxième fenêtre de l'échantillon

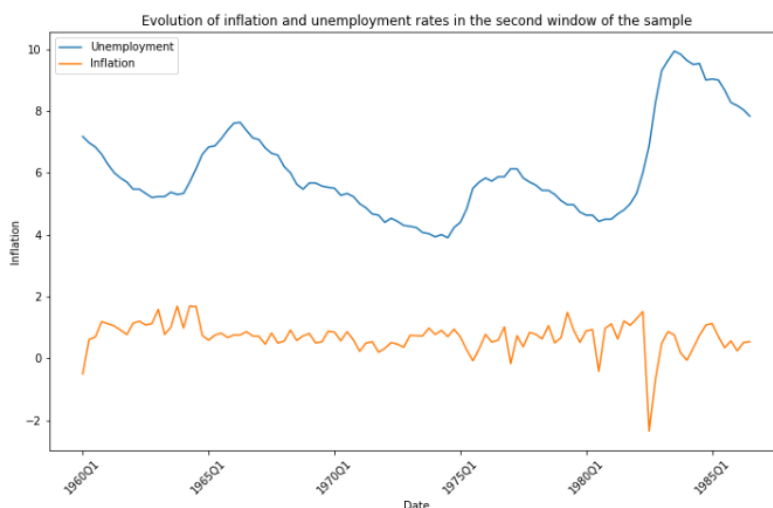


Figure 42: Evolution des taux d'inflation et du taux de chômage sur la deuxième fenêtre de l'échantillon

- Le test de changement de structure est significatif. On accepte donc l'hypothèse de changement de structure entre les deux sous-groupes.

2.10 Question 10

- Estimer la courbe de Philips en supprimant l'inflation courante des variables explicatives mais en ajoutant les délais d'ordre 1, 2, 3 et 4 de l'inflation et du chômage. Faire le test de Granger de non causalité de l'inflation sur le chômage. Donnez la p-valeur.

	const	INF_1	INF_2	INF_3	INF_4	Unemp_1	Unemp_2	Unemp_3	Unemp_4
4	1.0	0.201072	0.638012	0.067590	0.574907	6.8	6.27	5.53	5.23
5	1.0	-0.033523	0.201072	0.638012	0.067590	7.0	6.80	6.27	5.53

Figure 43: Tableau des valeurs

$$\text{inf}_t = \beta_0 + \beta_{\text{inf}_1} \text{inf}_{t-1} + \beta_{\text{inf}_2} \text{inf}_{t-2} + \beta_{\text{inf}_3} \text{inf}_{t-3} + \beta_{\text{inf}_4} \text{inf}_{t-4} + \beta_{\text{Unemp}_1} \text{Unemp}_{t-1} + \beta_{\text{Unemp}_2} \text{Unemp}_{t-2} + \beta_{\text{Unemp}_3} \text{Unemp}_{t-3} + \beta_{\text{Unemp}_4} \text{Unemp}_{t-4}$$

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.979			
Model:	OLS	Adj. R-squared:	0.978			
Method:	Least Squares	F-statistic:	1143.			
Date:	Fri, 09 Dec 2022	Prob (F-statistic):	3.10e-161			
Time:	11:59:19	Log-Likelihood:	4.5455			
No. Observations:	207	AIC:	8.909			
Df Residuals:	198	BIC:	38.90			
Df Model:	8					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
const	0.1454	0.072	2.007	0.046	0.003	0.288
INF_1	0.0275	0.038	0.722	0.471	-0.048	0.102
INF_2	-0.0204	0.041	-0.493	0.623	-0.102	0.061
INF_3	0.0705	0.040	1.748	0.082	-0.009	0.150
INF_4	0.0174	0.038	0.457	0.648	-0.058	0.092
Unemp_1	1.5931	0.071	22.379	0.000	1.453	1.733
Unemp_2	-0.6476	0.134	-4.837	0.000	-0.912	-0.384
Unemp_3	0.0253	0.135	0.187	0.852	-0.241	0.292
Unemp_4	-0.0101	0.070	-0.143	0.886	-0.148	0.128
=====						
Omnibus:	28.701	Durbin-Watson:	1.997			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	67.462			
Skew:	0.618	Prob(JB):	2.24e-15			
Kurtosis:	5.509	Cond. No.	143.			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 44: Résultats de la régression avec toutes les variables

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.977			
Model:	OLS	Adj. R-squared:	0.977			
Method:	Least Squares	F-statistic:	2165			
Date:	Fri, 09 Dec 2022	Prob (F-statistic):	1.36e-164			
Time:	11:59:19	Log-likelihood:	-2.9991			
No. Observations:	207	AIC:	16.00			
Df Residuals:	202	BIC:	32.66			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.2157	0.071	3.036	0.003	0.076	0.356
Unemp_1	1.6459	0.070	23.393	0.000	1.507	1.785
Unemp_2	-0.6975	0.135	-5.159	0.000	-0.964	-0.431
Unemp_3	0.0238	0.135	0.177	0.860	-0.241	0.289
Unemp_4	-0.0078	0.070	-0.112	0.911	-0.145	0.130
=====						
Omnibus:	39.770	Durbin-Watson:	1.996			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	83.656			
Skew:	0.908	Prob(JB):	6.83e-19			
Kurtosis:	5.530	Cond. No.	138.			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 45: Résultats de la régression sans la variable de chômage

- La p-valeur du test de Fisher est $0.004999 \approx 0.05$, donc on rejette l'hypothèse nulle de non significativité globale des coefficients. On peut ainsi dire qu'il y a une relation de causalité du chômage sur l'inflation.

2.11 Question 11

- Représentez graphiquement les délais distribués et commentez. Calculer l'impact à long de terme de l'inflation sur le chômage.

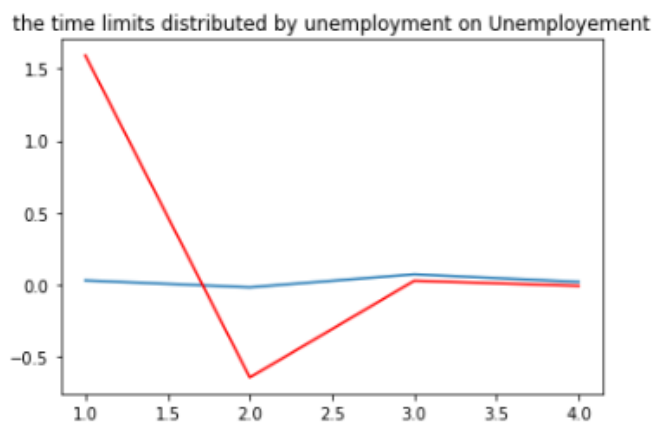


Figure 46: Délais distribués

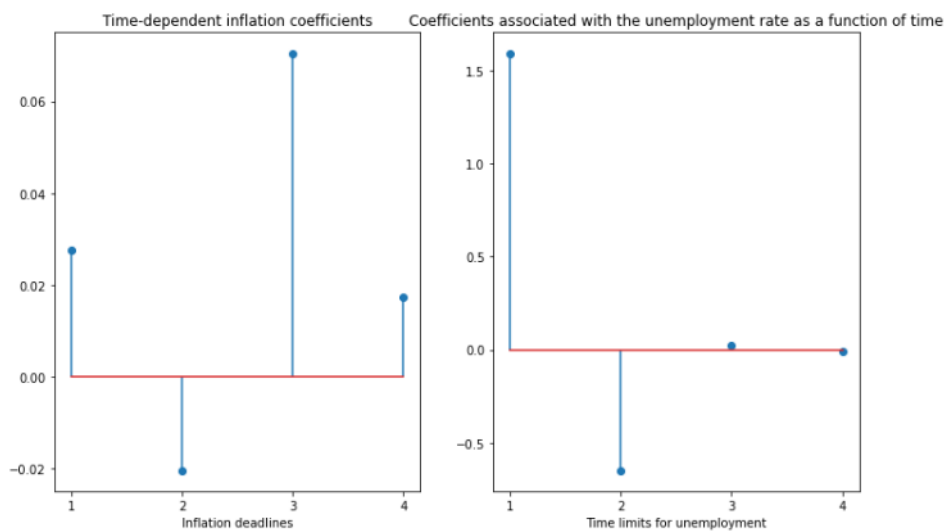


Figure 47: Délais d'inflation et de chômage

- Somme des coefficients associés à l'inflation : 0.095
- Les délais inflation suivent la même tendance que l'autocorrélogramme d'inflation. Ceux associés au taux de chômage suivent un modèle AR à coefficients négatifs.