Correction

La correction ci-dessous se veut extensive. Nous n'attendions pas forcément à ce que vous produisez quelque chose d'aussi détaillé. En revanche, la justification (par l'usage de l'intuition ou des mathématiques) reste primordiale pour répondre aux différentes questions posées.
Bonne lecture ! ☺

1 Exercice 1 : Des Enchères de Bégonias (12 points)

Selon la théorie économique (et plus particulièrement la théorie des jeux), le prix d'un bien qui est mis aux enchères doit augmenter lorsque le nombre d'enchérisseurs augmente.

En effet, plus le nombre d'enchérisseurs est élevé, plus il y a de concurrence sur le bien et donc le prix doit être plus élevé. Néanmoins, une régression simple du prix sur le nombre d'enchérisseurs peut conduire à des estimations biaisées et non convergentes s'il existe de l'endogeneité : l'objet qui est mis en enchère pourrait avoir des caractéristiques inobservées qui à la fois attirent les enchérisseurs et augmentent les prix.

Dans ce problème, nous utilisons une régression à variable instrumentale pour traiter de l'endogeneité afin d'étudier l'effet du nombre d'enchérisseurs sur les prix.

Les données que nous utiliserons viennent d'enchères hollandaises de bégonias. 45% des flux internationaux de fleurs et des plantes transitent par ces enchères sur ce marché. Chaque matin, avant 6h, les agriculteurs préparent leurs bégonias pour cette enchère et la vente commence à 6h30. **Le déroulement des enchères (ou l'ordre de passage des agriculteurs) est aléatoire** : **chaque agriculteur passe l'un après l'autre et ce, indépendamment des caractéristiques des fleurs.**

Nous utiliserons une base de données qui contient des informations sur deux types de bégonias. Les variables incluses dans cette base sont :

Table 1 – Tableau descriptif des variables

Nom de la variable	Description
Price	Prix payé à l'enchère
Bidders	Nombre d'enchérisseurs présents à l'enchère
time	Secondes après 6h30 quand a eu lieu l'enchère
red	=1 si le bégonia est rouge
yellow	=1 si le bégonia est jaune
monday	=1 si l'enchère a eu lieu lundi
tuesday	=1 si l'enchère a eu lieu mardi
wednesday	=1 si l'enchère a eu lieu mercredi
thursday	=1 si l'enchère a eu lieu jeudi

On suppose (pour simplifier) qu'il n'existe que des bégonias rouges et jaunes et qu'il n'y a pas d'enchères les vendredi, samedi et dimanche.

Dans cet exercice, nous cherchons à expliquer le prix des bégonias présentés par un agriculteur dans une enchère en fonction du nombre d'enchérisseurs, du type de bégonia et du jour de l'enchère. On dispose de n=79 observations d'enchères. La régression que nous estimons est la suivante :

$$ln(price)_i = \beta_0 + \beta_1 ln(bidders)_i + \beta_2 Red_i + \beta_3 Tuesday_i + \beta_4 Wednesday_i + \beta_5 Thursday + \varepsilon_i \quad (1)$$

1. (a) Expliquez soigneusement comment s'interprète le coefficient β_1 et le coefficient β_2

   Le coefficient β_1 s'interprète comme l'elasticité partielle prix-enchérisseurs.
   - Lorsque le nombre d'enchérisseurs augmente de 1%, le prix des fleurs augmente de β_1%, toutes choses égales par ailleurs.

   Le coefficient β_2 quant à lui s'interprète comme l'impact de la couleur des bégonias sur le logarithme du prix, toutes choses égales par ailleurs.
   - Lorsque les bégonias sont rouges, le prix augmente de $100 * \beta_2$ % ceteris paribus sic stantibus.

   (b) Quel est, selon vous, le signe attendu de β_1 ?

   Selon la théorie économique, on s'attend à ce que le signe de β_1 soit positif. Plus il y a d'enchérisseurs, plus le prix du bien doit être élevé.

2. Pourquoi les variables $yellow_i$ et $monday_i$ ne sont pas présentes dans la régression ?

   Ces deux variables ne sont pas présentes dans la régression en raison d'un problème de multicolinéarité : comme la constante est présente dans la régression, elle serait une combinaison linéaire entre red et yellow lorsqu'on introduit yellow ou une combinaison linéaire entre tuesday, wednesday, thursday et monday lorsqu'on introduit monday. Partant, le modèle n'est plus estimable. Pour la culture générale, on parle de **"Dummy Trap"**.

3. On présente la régression par moindres carrés de l'équation (1) sur stata. Le tableau de sortie est le suivant :

Source	SS	df	MS
Model	10.671743	5	2.13434859
Residual	22.8677329	73	.313256614
Total	33.5394758	78	.42999328

Number of observations	=	79
F(5.73)	=	6.81
Prob >F	=	0.0000
R-squared	=	0.3182
Adj R-squared	=	0.2715
Root MSE	=	.55969

lnprice	Coef.	Std. Err.	t	P> $\|t\|$	[95% Conf.	Interval]
lnbidders	1.479757	.3227842	4.58	0.000	.8364492	2.123065
red	.1154484	.1302029	0.89	0.378	-.1440455	.3749423
tuesday	.3197937	.1723074	1.86	0.067	-.0236144	.6632018
wednesday	-.1143352	.1588379	-0.72	0.474	-.4308987	.2022283
thursday	-.231193	.2040158	-1.13	0.261	-.6377959	.1754099
_cons	-.9303385	1.290439	-0.72	0.473	-3.502179	1.641502

TABLE 2 – Régression MCO

   Commenter le tableau présenté ci-dessus (R^2, $\overline{R^2}$, significativité globale, significativité des coefficients). Ces résultats sont-ils cohérents avec ce que prédit la théorie économique ?

   Le $R^2 = 1 - \frac{SCR}{SCT}$ de la régression est de 32%. En revanche, comme la valeur du R^2 peut être artificiellement gonflée par le nombre de variables explicatives, il vaut mieux s'intéresser au $\overline{R^2}$ =

$1 - \frac{SCR/(n-k-1)}{SCT/(n-1)}$. Ici, le modèle semble prédire 27% de la variation du logarithme du prix.

Le modèle est globalement significatif selon la statistique de Fisher ($H_0 : \beta_0 = \beta_1 \ldots = \beta_5 = 0$ vs H_1 : au moins un des coefficients est différent de 0). Comme la Prob<F est inférieure à 1%, on rejette H_0 au seuil de 1%. Il existe au moins un coefficient significativement différent de 0 dans cette régression.

On constate enfin que la seule variable significative à 1% dans ce modèle est lnbidders (p>t inférieure à 1%). Le signe correspond bien à la théorie : une augmentation du nombre d'enchérisseurs est correlée positivement au prix, toutes choses égales par ailleurs. La variable Tuesday est également significative à 10% ($P > |t|$ inférieur à 10%).

4. On s'intéresse aux résidus de l'estimation par moindres carrés. On trace graphiquement les distributions des résidus avec la commande *kdensity*. La distribution correspond à la figure (1).

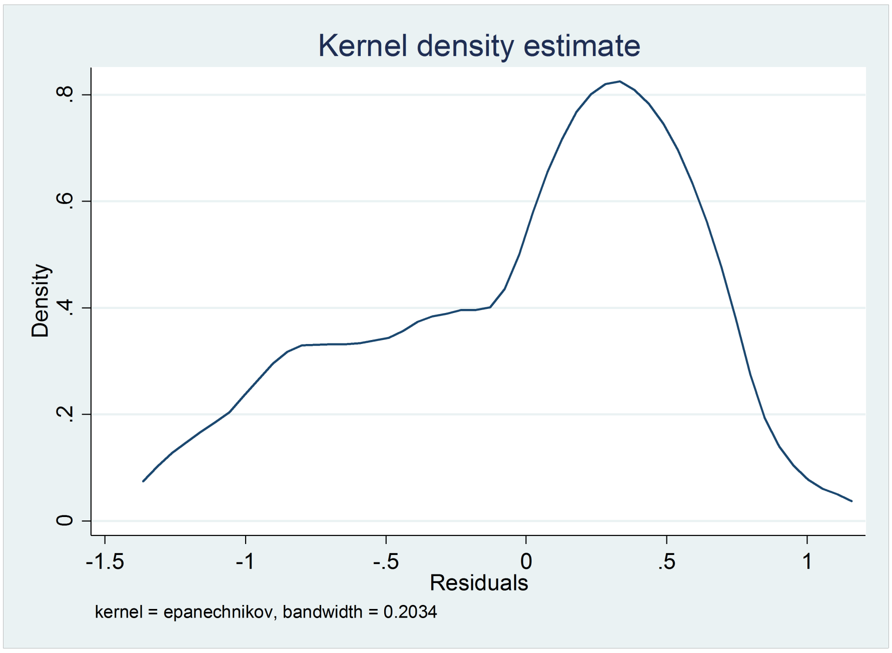


FIGURE 1 – Distribution des résidus par densité

(a) Commentez le graphique

On constate que les résidus sont distribués autour de la valeur 0. En effet, le programme de minimisation des erreurs du programme des moindres carrés conduit mécaniquement à ce que la somme des résidus soit nulle. En revanche, on constate que la distribution est très asymétrique : la skewness (coefficient d'asymétrie) est négative, le mode est légèrement décalé vers la droite et le kurtosis (coefficient d'aplatissement) est faible. Il faudrait ainsi effectuer le test de Jarque-Bera (H_0 : les données suivent une distribution normale contre H_1 : les données ne suivent pas une distribution normale) pour tester la normalité de la distribution.

(b) Comment doit-être la loi des résidus ? Pourquoi cette hypothèse est-elle cruciale ?

L'hypothèse de normalité des résidus est une hypothèse cruciale afin de réaliser tous les tests statistiques et la construction des intervalles de confiance. En effet, en l'absence de cette normalité, la loi de probabilité de $\frac{\hat{\beta}_i - \beta}{\sigma \beta_i}$ n'est plus connue. On ne peut plus partant conclure quoi que ce soit sur la significativité des coefficients estimés.

Dans la suite de cet exercice, nous ferons comme si les résidus "se comportent bien".

5. Les bégonias ont des caractéristiques inobservées (comme leur taille) qui peuvent augmenter le nombre d'enchérisseurs. Supposez un modèle simplifié avec une seule variable explicative :

$$ln(price)_i = \beta_0 + \beta_1 ln(bidders)_i + \varepsilon_i \tag{2}$$

On suppose ainsi que $cov\Big(taille_i, ln(bidders)_i\Big) > 0$ et que $\varepsilon_i = taille_i + \nu_i$

Comment cette variable inobservée influe-t-elle sur le coefficient β_1 ? (Justifiez théoriquement votre réponse)

La formule du coefficient estimé $\hat{\beta}_1$ s'écrit de la manière suivante :

$$\begin{aligned}
\hat{\beta}_1 = \frac{cov\Big(ln(bidders)_i, ln(price)_i\Big)}{V\Big(ln(bidders)_i\Big)} &= \frac{cov\Big(ln(bidders)_i, \beta_0 + \beta_1 ln(bidders)_i + \varepsilon_i\Big)}{V\Big(ln(bidders)_i\Big)} \\
&= 0 + \beta_1 \frac{V\Big(ln(bidders)_i\Big)}{V\Big(ln(bidders)_i\Big)} + \frac{cov\Big(ln(bidders)_i; \varepsilon_i\Big)}{V\Big(ln(bidders)_i\Big)} \\
&= \beta_1 + \frac{cov\Big(ln(bidders)_i; taille + \mu_i\Big)}{V\Big(ln(bidders)_i\Big)} \\
&= \beta_1 + \frac{cov\Big(ln(bidders)_i; taille_i\Big)}{V\Big(ln(bidders)_i\Big)} > \beta_1 \blacksquare
\end{aligned}$$

car $cov\Big(ln(bidders)_i; taille_i\Big) > 0$ et $V\Big(ln(bidders)_i\Big) \geq 0$. On en déduit ainsi que $\hat{\beta}_1$ sera surestimé si la variable $taille_i$ est inobservée.

6. Pour traiter ce problème de variable omise, nous utiliserons une régression à variable instrumentale. L'instrument que nous utilisons ici est $time_i$ le temps de l'enchère. Pourquoi pensons-nous qu'il s'agit d'une bonne variable instrumentale ?

Time pourrait être une bonne variable instrumentale pour régler le problème d'endogeneité de la variable lnbidders.

— En effet, on peut penser que les enchérisseurs sont plus nombreux à mesure que l'enchère se déroule (car certains peuvent arriver en retard et que 6h30, c'est quand même relativement tôt...). Ainsi, on aurait $cov(time, lnbidders) > 0$.

— D'autre part, l'énoncé atteste que le déroulement des enchères est indépendant des caractéristiques des fleurs et donc que $cov(time, \varepsilon) = 0$. C'est pour ces deux raisons que nous pouvons penser que time est une bonne variable instrumentale pour lnbidders.

Nous instrumentons donc la variable endogène $ln(bidders)_i$ par $time_i$ et les variables explicatives du modèle (1) :

$$ln(bidders)_i = \alpha_0 + \alpha_1 time_i + \alpha_2 red_i + \alpha_3 tuesday_i + \alpha_4 wednesday_i + \alpha_5 thursday_i + e_i \quad (3)$$

Source	SS	df	MS
Model	.65534547	5	.131069094
Residual	2.82140507	73	.038649384
Total	3.47675054	78	.044573725

Number of obs	=	79
F(5,73)	=	3.39
Prob >F	=	0.0082
R-squared	=	0.1885
Adj R-squared	=	0.1329
Root MSE	=	.19659

lnbidders	Coef.	Std. Err.	t	P	[95% Conf.	Interval]
time	.0004569	.0002087	2.19	0.032	.0000409	.000873
red	-.0869847	.0451667	-1.93	0.058	-.1770018	.0030324
tuesday	.0332964	.0627371	0.53	0.597	-.0917385	.1583313
wednesday	.1040419	.0624083	1.67	0.100	-.0203378	.2284215
thursday	-.0944148	.0760507	-1.24	0.218	-.2459837	.057154
_cons	3.765482	.1067965	35.26	0.000	3.552637	3.978327

TABLE 3 – Première étape des 2MCO/2SLS

7. $Time_i$ est-elle une variable significative ? Si oui, à quel seuil ?

Time est significative à 5%. Le signe est bien celui que nous attendions. On pourrait donc penser à juste titre que le biais induit par les variables inobservées est bien positif.

La régression par variable instrumentale (2MCO) donne le tableau suivant sur Stata :

Source	SS	df	MS
Model	9.00972936	5	1.80194587
Residual	24.5297465	73	.336023924
Total	33.5394758	78	.42999328

Number of obs	=	79
F(5,73)	=	2.49
Prob >F	=	0.0385
R-squared	=	0.2686
Adj R-squared	=	0.2185
Root MSE	=	.57968

lnprice	Coef.	Std. Err.	t	P> $\|t\|$	[95% Conf.	Interval]
lnbidders	1.03659	.5307879	1.95	0.055	-.0196779	2.092857
red	.0586522	.1676926	0,35	0.728	-.2755587	.3928632
tuesday	.3176616	.1784984	1,78	0.079	-.0380852	.6734083
wednesday	-.0831461	.1733764	-0,48	0.633	-.4286848	.2623926
thursday	-.3526324	.3001199	-1,17	0.244	-.9507704	.2455057
_cons	2.030436	5.365384	0,38	0.706	-8.662759	12.72363
Instrumented	lnbidders					
Instruments	time	red	tuesday	wednesday	thursday	

TABLE 4 – Régression par variable instrumentale

8. Commentez le tableau de régression.

   La régression suivante correspond à la régression DMCO lorsque la variable lnbidders est instrumentée par time et par toutes les autres variables explicatives du modèle. Rappelez vous qu'il faut au moins autant, si ce n'est plus de variables instrumentales que de variables explicatives, pour que le modèle soit juste ou sur-identifié.

   Le R^2 et le R^2 ajusté ont diminué drastiquement (32% → 19% pour le R^2 et 27% → 13% pour $\overline{R^2}$) car l'instrumentation fait perdre une partie de l'information à la régression.

   Il faut surtout constater que même si le coefficient de lnbidders a diminué en magnitude (1.48 →1.03) comme nous l'avions prévu en vertu du biais d'endogeneité, ce coefficient n'est significatif qu'à 10%.

   Dernier point, mais non des moindres. Le modèle est globalement significatif à 5% (Prob>F =0.03).Notre régression a donc considérablement perdu en pouvoir explicatif. De tels résultats nous amènent à envisager deux pistes de réflexion : est-ce que time est une bonne variable instrumentale ? A priori, oui car elle respecte la condition $cov(time, \varepsilon) = 0$ et $cov(time, lnbidders) > 0$. Est-ce que nous disposons assez d'observations pour conclure à la non significativité de lnbidders ? En effet, on sait que la variable instrumentale a des bonnes propriétés lorsque $N \to \infty$. Ainsi, la non-significativité de lnbidders et la non significativité globale de la régression pourraient être dues au manque d'observations dans notre échantillon.

9. Répondez aux questions suivantes :

   (a) Quel test faudrait-il utiliser pour tester l'endogeneité ? Expliquez le principe du test et écrivez les hypothèses à tester.

   Le principe du test d'Hausman repose sur le fait que si il y a de l'endogeneité, β_{MCO} est biaisé et non convergent et β_{IV} est également biaisé mais convergent.

   Si la différence $\beta_{MCO} - \beta_{IV}$ ne converge pas vers 0, alors on peut supputer qu'il y a bien un problème d'endogénéité. Ecrivons les hypothèses, la statistique de test et la zone de rejet :

   H_0 : écart non significatif entre les deux estimations
   H_1 : écart significatif entre les deux estimations

   $$H = (\hat{\beta_{IV}} - \hat{\beta_{MCO}})'\left(V(\hat{\beta_{IV}}) - V(\hat{\beta_{MCO}})\right)^{-1}(\hat{\beta_{IV}} - \hat{\beta_{MCO}}) \to \chi^2(k) \tag{4}$$

   Si $H > \chi^2(k)$, alors on rejette H_0 et il faut appliquer les DMCO.

   (b) Ecrivez les commandes Stata pour ce test.

   reg lnprice lnbidders red tuesday wednesday thursday

   store estimates mco

   ivreg lnprice (lnbidders= time red tuesday wednesday thursday) red tuesday wednesday thursday

   hausman . mco, constant sigmamore

Exercice 2 : Hakuna Matata (8 points)

2

Nous disposons de données hebdomadaires sur le nombre de recherches Google associées à l'expression « hakuna matata ». Il s'agit des recherches effectuées en France pendant la période janvier 2010-décembre 2013.

Figure 2 – Simba

1. Le graphique ci-dessous présente l'évolution du nombre de recherches effectuées chaque semaine entre 2010 et 2013 (ie de la semaine 1 à la semaine 209). Cette série vous semble-t-elle stationnaire ? Pourquoi ?

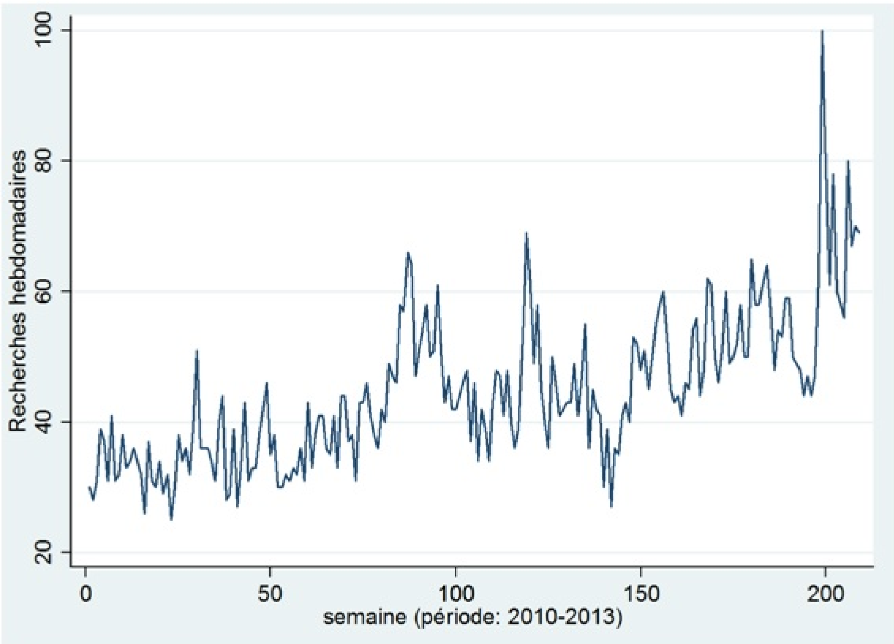


Figure 3 – Recherches associées à l'expression "hakuna matatata"

Le nombre moyen de recherches Google associées à l'expression "hakuna matata" n'est pas constant. Il augmente avec le temps. La variance dépend également du temps. Par conséquent, la série ne semble pas suivre un processus stationnaire.

2. Rappelez les conditions nécessaires pour qu'un processus soit stationnaire au second degré (stationnaire au sens faible).

Un processus est stationnaire au second degré si :

- sa moyenne est constante et indépendante du temps : $E(Y_t) = m, \forall t$
- sa variance est finie et indépendante du temps : $E(Y_t^2) = Var(Y_t) < \infty, \forall t$
- sa covariance est indépendante du temps : $Cov(Y_t, Y_{t+h}) = E[(Y_t - \mu)(Y_{t+h} - \mu)] = \gamma_h$

3. Quelle commande Stata peut-on utiliser pour étudier l'autocorrélation simple et partielle de notre série ? Ecrivez la commande pour la série « hakunamatata » pour 20 retards.

   Le corrélogramme, le graphique des fonctions d'autocorrélation simple et partielle, est obtenu avec la commande : **corrgram hakunamatata, lags(20)**

   Par ailleurs, les représentations graphiques de l'auto-corrélation et de l'auto-corrélation partielle peuvent s'obtenir avec les commandes :
   - auto-corrélation : **ac hakunamatata, lags(20)**
   - auto-corrélation partielle : **pac hakunamatata, lags(20)**

4. Voici les résultats obtenus suite à la commande précédente.

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.7462	0.7628	118.05	0.0000		
2	0.6271	0.1871	201.83	0.0000		
3	0.5973	0.2249	278.22	0.0000		
4	0.5317	0.0582	339.04	0.0000		
5	0.4794	0.0233	388.72	0.0000		
6	0.4672	0.1043	436.15	0.0000		
7	0.5141	0.2344	493.85	0.0000		
8	0.4645	0.0084	541.18	0.0000		
9	0.4300	0.0292	581.96	0.0000		
10	0.3869	-0.0367	615.14	0.0000		
11	0.3320	0.0303	639.69	0.0000		
12	0.3226	0.0827	662.98	0.0000		
13	0.3102	0.0242	684.63	0.0000		
14	0.2893	-0.0558	703.56	0.0000		
15	0.3483	0.2686	731.14	0.0000		
16	0.3618	0.0843	761.04	0.0000		
17	0.3546	0.0711	789.92	0.0000		
18	0.3500	0.0242	818.2	0.0000		
19	0.3536	0.0503	847.22	0.0000		
20	0.3109	-0.0871	869.78	0.0000		

   Etudiez formellement la stationnarité de cette série (mettre le test d'hypothèses, préciser la loi suivie par la statistique Q de ce tableau, conclure quant à la stationnarité de la série).

   On constate que les coefficients d'autocorrélation simple sont assez élevés et ils diminuent lentement, ce qui nous amènerait à penser que la série n'est pas stationnaire.

   Afin d'étudier formellement la stationnarité de cette série, on procède à un test de Ljung-Box :

   $H0 : \rho_1 = \rho_2 = ... = \rho_h = 0 \ (stationnarite)$
   $H1 : il\ existe\ au\ moins\ un\ \rho_h\ significativement\ different\ de\ 0\ (non-stationnarité)$

   La statistique Q de ce tableau est la statistique de Ljung-Box (LB) :

$$Q = T(T+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{T-k} \rightarrow \chi^2(h) \tag{5}$$

   Si $Q > \chi^2(h)$, on rejette H0, au seuil retenu.

   Ici, p-value < 0.01 pour tous les retards retenus. Par conséquent, on rejette l'hypothèse nulle au seuil de 1%. Donc la série "hakuna matata" n'est pas stationnaire.

5. Nous souhaitons caractériser plus précisément notre série. Notamment, nous aimerions savoir si elle possède une tendance (trend).
   (a) Présentez les trois modèles associés au test de Dickey-Fuller augmenté.

   On met en place un test de Dickey-Fuller augmenté afin d'étudier plus précisément la non-stationnarité de la série. Ce test prend en compte l'autocorrélation des erreurs via retards sur variable endogène. En effet, le test permet de détecter l'existence d'une tendance et de déterminer la bonne manière pour stationnariser la série. Les 3 modèles sont présentés ci-dessous :

$$\Delta X_t = \phi X_{t-1} + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t \quad \text{(Sans constante, sans tendance)} \tag{6}$$

$$\Delta X_t = \phi X_{t-1} + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \beta + \varepsilon_t \quad \text{(Avec constante, sans tendance)} \tag{7}$$

$$\Delta X_t = \phi X_{t-1} + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \gamma + \delta t + \varepsilon_t \quad \text{(Avec constante et tendance)} \tag{8}$$

   (b) Quel modèle estime-t-on ci-dessous ?

   Ici, on estime le troisième modèle, avec constante et tendance, pour 14 retards. Stata introduit par défaut une constante et l'option intégrant la tendance (trend) apparaît aussi dans la ligne de commande.

```
. dfuller hakunamatata, lags(14) regress trend

Augmented Dickey-Fuller test for unit root         Number of obs   =       194
```

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.629	-4.009	-3.437	-3.137

MacKinnon approximate p-value for Z(t) = 0.2667

D.hakunama~a	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hakunamatata						
L1.	-.3643229	.1385742	-2.63	0.009	-.637793	-.0908527
LD.	-.1240115	.1403473	-0.88	0.378	-.4009809	.1529579
L2D.	-.1103393	.1375089	-0.80	0.423	-.3817072	.1610286
L3D.	.0143334	.1349722	0.11	0.916	-.2520284	.2806952
L4D.	.0021677	.1308355	0.02	0.987	-.2560305	.2603659
L5D.	-.0605429	.1273028	-0.48	0.635	-.3117697	.1906838
L6D.	-.1233079	.1240863	-0.99	0.322	-.368187	.1215712
L7D.	.1097885	.1219342	0.90	0.369	-.1308434	.3504203
L8D.	.0415903	.1193663	0.35	0.728	-.1939739	.2771545
L9D.	.079586	.1142684	0.70	0.487	-.1459178	.3050898
L10D.	.0129439	.1103395	0.12	0.907	-.2048064	.2306942
L11D.	-.0151398	.1090037	-0.14	0.890	-.2302539	.1999743
L12D.	-.0094267	.1065585	-0.09	0.930	-.2197153	.2008618
L13D.	-.0008166	.0981329	-0.01	0.993	-.1944776	.1928444
L14D.	-.2153728	.0863309	-2.49	0.014	-.385743	-.0450025
_trend	.0509347	.0191354	2.66	0.008	.0131717	.0886977
_cons	11.05022	4.37503	2.53	0.012	2.416288	19.68416

   (c) Commentez les résultats obtenus, en retenant le seuil de 5%. Comment peut-on caractériser notre série ? Justifiez.

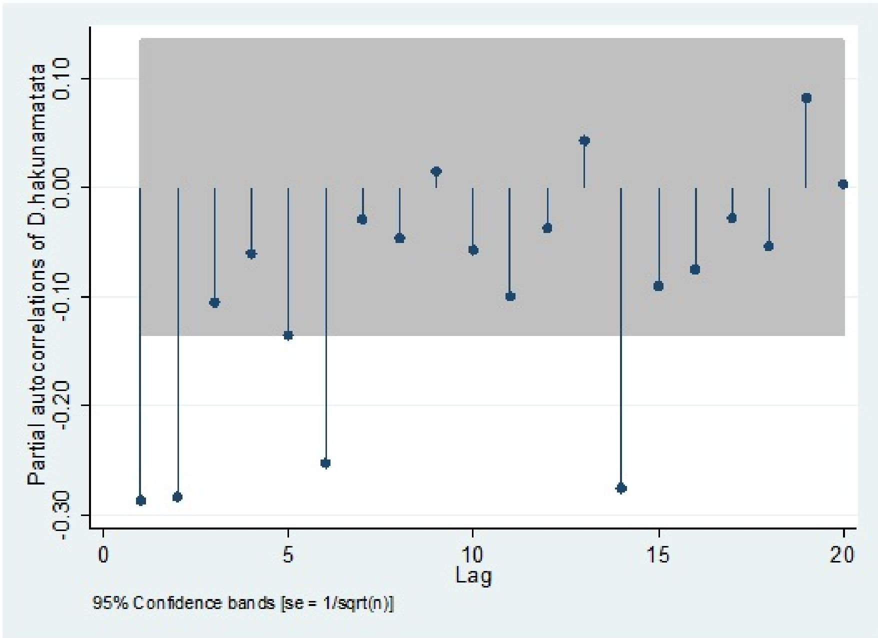
On teste l'hypothèse $H_0 : \phi = 0$ (présence d'une racine unitaire) contre $H_1 : \phi \neq 0$ (pas de racine unitaire)

Lorsque $z_\phi > z_{DF}$, on accepte H_0 .

Ici, $z_\phi = -2.629 > z_{DF}(1\%)$.

On accepte H_0, l'hypothèse de présence d'une racine unitaire, donc la série n'est pas stationnaire. Il ne s'agit pas d'un processus TS.

(d) En vous basant sur le graphique d'autocorrélation partielle de la série différenciée, expliquez pourquoi on a retenu 14 lags pour le test effectué précédemment.



Ici, on a retenu 14 retards pour deux raisons principales.

— Dans un premier temps, le graphique d'autocorrélation partielle de la série differenciée semble indiquer l'existence d'une autocorrélation partielle significative pour le 14eme retard. Le 14ème retard sort de l'intervalle de confiance, donc il est significativement différent de 0.

— Notre inquiétude se confirme puisque le test de Dickey Fuller augmenté montre également que le coefficient associé au 14eme retard est significatif. (p>|t| inférieure à 1%.)