HAL
archives-ouvertes.fr

Prévision par l'approche méthodologique de Box et Jenkins : Cas d'une Série Non Saisonnière et Non Stationnaire du type TS (Pratique sur EViews et Stata)

Jonas Kibala Kuma

▶ **To cite this version:**

Jonas Kibala Kuma. Prévision par l'approche méthodologique de Box et Jenkins : Cas d'une Série Non Saisonnière et Non Stationnaire du type TS (Pratique sur EViews et Stata). Licence. Congo-Kinshasa. 2018. <cel-01771600>

HAL Id: cel-01771600
https://hal.archives-ouvertes.fr/cel-01771600

Submitted on 19 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kinshasa, Avril 2018

Prévision par l'approche méthodologique de Box et Jenkins : Cas d'une Série Non Saisonnière et Non Stationnaire du type TS (*Pratique sur EViews et Stata*)

Par

Jonas KIBALA KUMA

(DEA-PTC Economie/Unikin en cours)

--

Centre de Recherches Economiques et Quantitatives
(CREQ)

***

« Rien n'est trop tôt, ni trop tard, tout est à sa place ».

Avril 2018

Copyright © Jonas Kibala_Avril_2018 - Tous droits réservés.

Modélisation de l'évolution des recettes de ventes d'une société commerciale à Kinshasa par l'approche de Box et Jenkins[1]

PLAN

I. Analyse exploratoire des données

I.1. Stationnarité et structure de la série

a) Tests préliminaires (informels)

- *Représentation graphique*
- *Comparaison des différents moments statistiques (Ecart-type et Moyenne)*
- *Corrélogramme*

b) Tests formels (Test de ADF et PP)

I.2. Stationnarisation de la série

- *Transformations de la série*
- *Plot de la série transformée*

II. Identification du modèle optimal (provisoire)

III. Estimation du modèle provisoire retenu

IV. Validation du modèle estimé

- *Test de bruit blanc (Test portemanteau)*
- *Conditions de stationnarité*
- *Test de Student des paramètres*
- *Test de linéarité*
- *Performance du modèle estimé*
- *Test de Normalité*
- *Test d'absence d'autocorrélation des erreurs*

V. Prévision

[1] Nous nous servons du logiciel EViews 5 et Stata 9.

PART 1 : PRATIQUE SUR EVIEWS

I. Analyse exploratoire des données

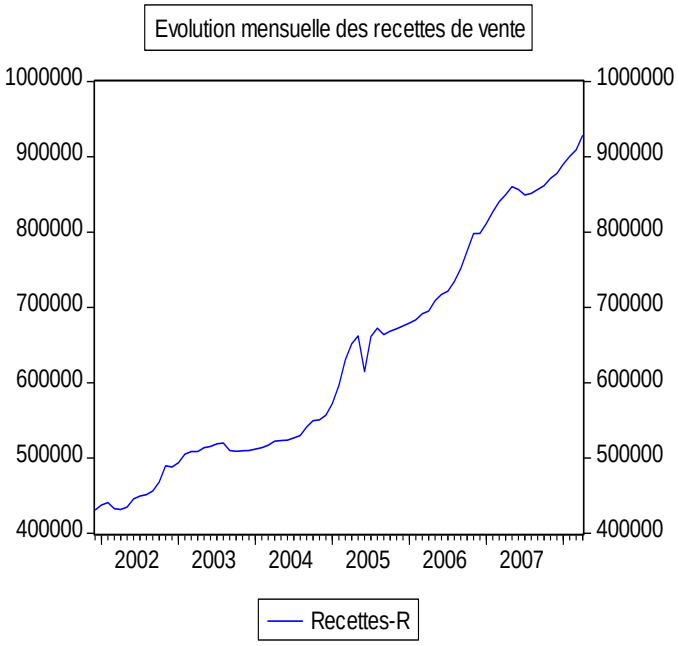
I.1. Stationnarité et structure de la série

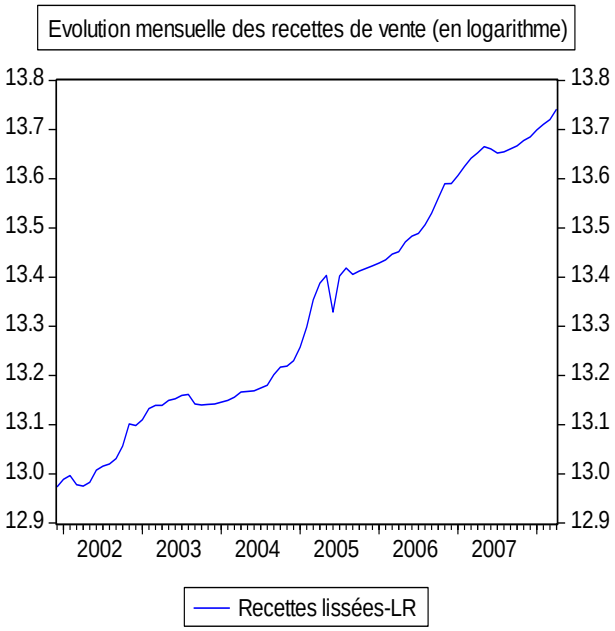
a) Tests préliminaires (informels)

- *Représentation graphique* :

Sur EViews, faire :

```
create m 2001:12 2008:04
data R
plot R
genr LR=log(R)
plot LR
```





Note : Ce graphique (dans tous les cas : série brute et lissée) montre que notre série « R » est affectée d'une tendance[1] – faut-il en savoir encore le type (déterministe/TS ou Stochastique/DS ??) – et semble non saisonnière.

- *Comparaison des différents moments statistiques (Ecart-type et Moyenne)*

Après avoir regroupé les données en classes[2] (suivant les années d'observations : Cfr « Table de Buys-Ballot » en annexe) et calculer, pour chaque classe/année, les moments statistiques d'ordre 1 et 2 (soient la moyenne/MOY et l'écart-type/ECT, respectivement), régresser l'écart-type sur la moyenne (soit la relation : $\mathbf{ECT_t = \alpha_0 + \alpha_1 MOY_t + e_t}$) et tester les hypothèses suivantes :

[1] *La série brute laisse présager une tendance non linéaire, et la série lissée renseigne plutôt sur une tendance haussière.*
[2] *Les observations peuvent être regroupées en 7 à 12 classes (discontinues) à intervalle constante.*

H_0 : $\alpha_1 = 0$: *la série semble stationnaire* $\left(prob > 5\% , |t_c| < |t_t| , \hat{\alpha}_1 < 2\sigma_{\hat{\alpha}_1}\right)$
H_1 :$\alpha_1 \neq 0$: *la série semble non stationnaire* $\left(prob < 5\% , |t_c| > |t_t| , \hat{\alpha}_1 > 2\sigma_{\hat{\alpha}_1}\right)$

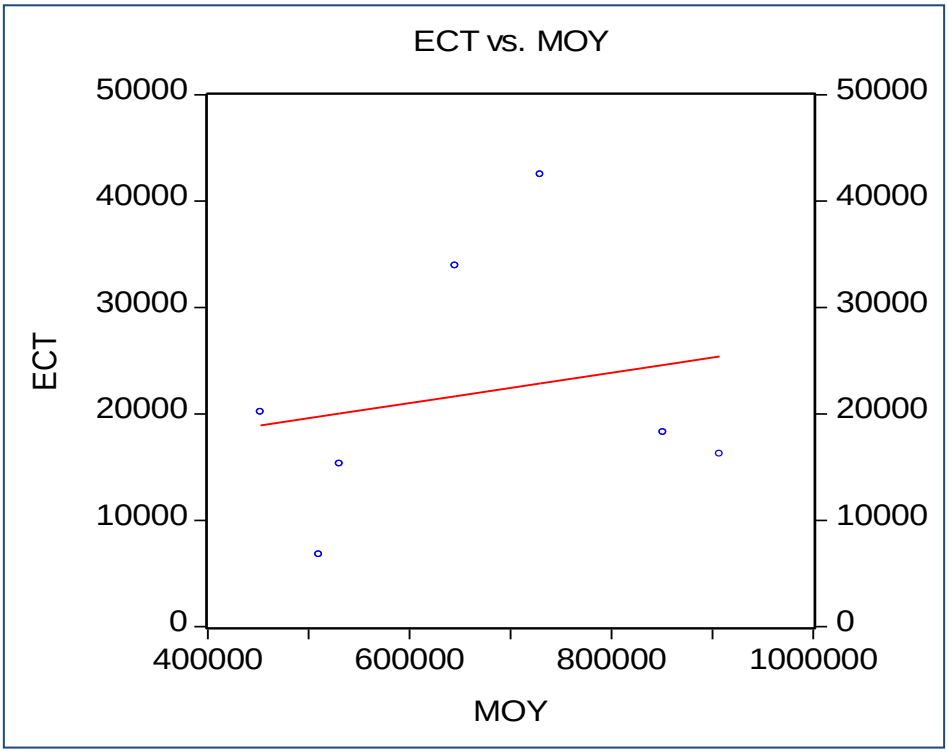
_____Sur EViews, faire : *ls ECT c MOY*

Dependent Variable: ECT
Method: Least Squares
Sample: 2001 2007
Included observations: 7

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MOY	0.032046	0.006852	4.676761	0.0034

La constante a parue non significative et a été élimée, ce qui nous amène à accepter H1 : la série « R » semble non stationnaire.

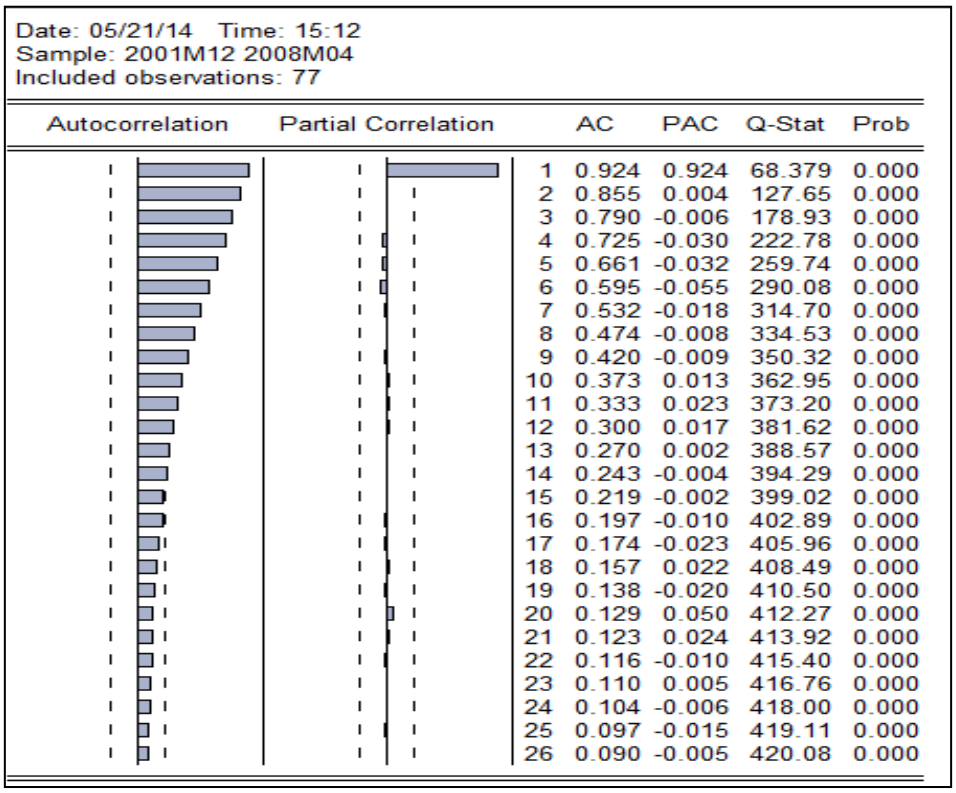
_____Graphique : Sur EViews, taper : Show MOY ECT, et suivre : View/Graph/Scatter/Scatter with Regression…→Ok :



Ce dernier graphique du couple (ECT, MOY)) nous fait constater que les écarts-types affichent une dispersion aléatoire (non proportionnelle) au tour d'une tendance haussière, ce qui est caractéristique d'une série non stationnaire.

Corrélogramme

Ici, il est question d'analyser les fonctions d'autocorrélations simple (AC) et partielle (PAC) dans le souci d'identifier les coefficients d'autocorrélation significatifs (gage pour un test DF/ADF réussi) et se faire une idée sur la stationnarité ou pas de la série sous-étude (soit « *LR* »). Sur EViews, faire : *Show LR→View/Correlogram…→Cocher « Level » et lags to include : 26 (soit : T/3=77/3=25.7).*



Date: 05/21/14 Time: 15:12
Sample: 2001M12 2008M04
Included observations: 77

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.924	0.924	68.379	0.000
		2	0.855	0.004	127.65	0.000
		3	0.790	-0.006	178.93	0.000
		4	0.725	-0.030	222.78	0.000
		5	0.661	-0.032	259.74	0.000
		6	0.595	-0.055	290.08	0.000
		7	0.532	-0.018	314.70	0.000
		8	0.474	-0.008	334.53	0.000
		9	0.420	-0.009	350.32	0.000
		10	0.373	0.013	362.95	0.000
		11	0.333	0.023	373.20	0.000
		12	0.300	0.017	381.62	0.000
		13	0.270	0.002	388.57	0.000
		14	0.243	-0.004	394.29	0.000
		15	0.219	-0.002	399.02	0.000
		16	0.197	-0.010	402.89	0.000
		17	0.174	-0.023	405.96	0.000
		18	0.157	0.022	408.49	0.000
		19	0.138	-0.020	410.50	0.000
		20	0.129	0.050	412.27	0.000
		21	0.123	0.024	413.92	0.000
		22	0.116	-0.010	415.40	0.000
		23	0.110	0.005	416.76	0.000
		24	0.104	-0.006	418.00	0.000
		25	0.097	-0.015	419.11	0.000
		26	0.090	-0.005	420.08	0.000

Deux éléments nous fondent à croire que, partant du corrélogramme ci-dessus, notre série « R » semble non Stationne (ce qui renforce notre présomption), à savoir :

- La *décroissance lente et linéaire* des coefficients d'autocorrélations simples (AC) ;
- *L'autocorrélation sérielle* lisible sur tous les coefficients d'autocorrélation statistiquement différents de zéro (prob < 5%).

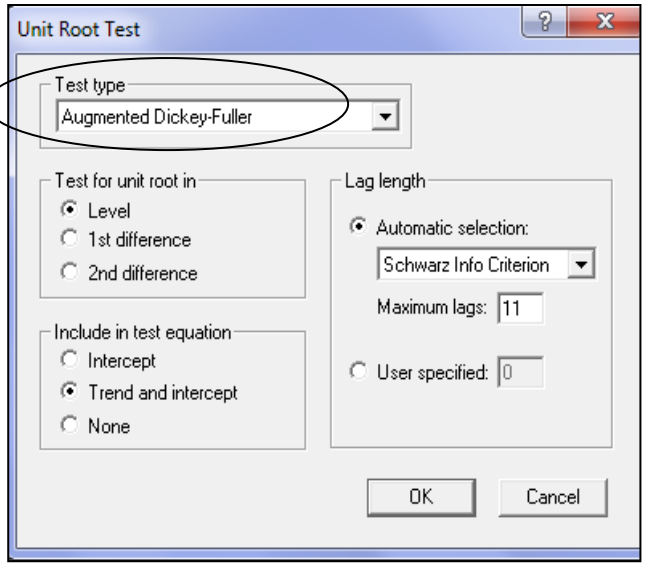
Note : Les tests informels aux quels nous avons fait recours dans les lignes précédentes s'accordent à présumer l'existence d'une racine unitaire au sein de la série « R ». Reste à le vérifier à travers des tests formels (ADF et Philips-Perron) et répondre à la question du type de non stationnarité.

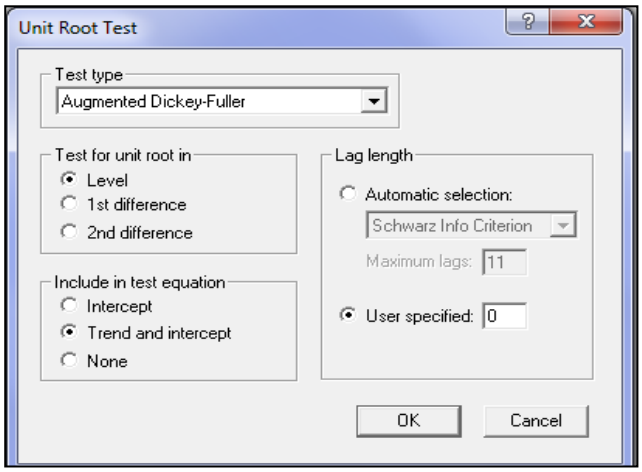
b) *Tests formels (Test de ADF et Philips-Perron)*

Ces tests nous aideront à valider la présomption de non stationnarité sur la série « R » et en même temps nous indiqueront sur la méthode de stationnarisation appropriée, étant donné la structure de la série étudiée.

- **Test ADF (Augmented Dickey-Fuller)**

_____Sur EViews, faire : Show LR→View/Unit Root Test...→(les boîtes de dialogue ci-dessous complètent la procédure) ou taper : *uroot LR*





Null Hypothesis: LR has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.085924	0.5450
Test critical values:	1% level	-4.083355	
	5% level	-3.470032	
	10% level	-3.161982	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LR)
Method: Least Squares
Date: 05/21/14 Time: 15:22
Sample (adjusted): 2002M01 2008M04
Included observations: 76 after adjustments

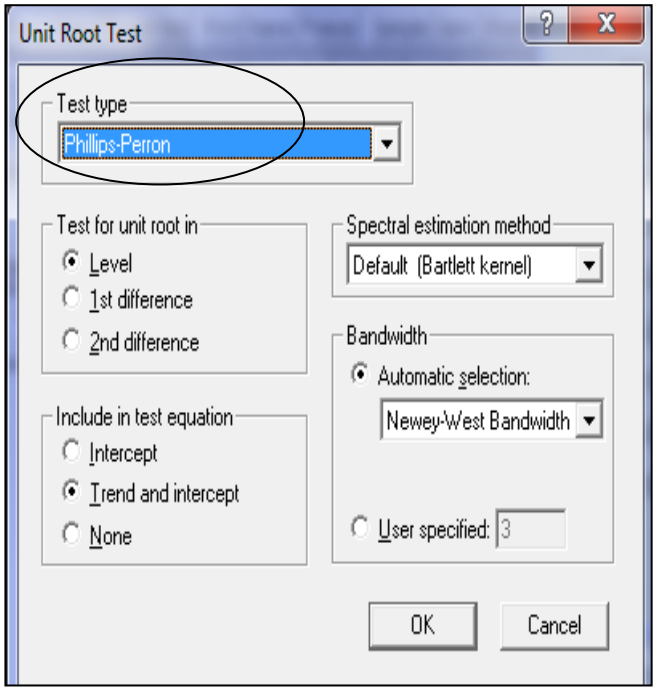
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LR(-1)	-0.104671	0.050180	-2.085924	0.0405
C	1.360214	0.648367	2.097906	0.0394
@TREND(2001M12)	0.001155	0.000533	2.168718	0.0334

R-squared	0.061667	Mean dependent var	0.010105
Adjusted R-squared	0.035959	S.D. dependent var	0.017496
S.E. of regression	0.017179	Akaike info criterion	-5.251602
Sum squared resid	0.021543	Schwarz criterion	-5.159600
Log likelihood	202.5609	F-statistic	2.398750
Durbin-Watson stat	1.899753	Prob(F-statistic)	0.097958

_____**Note** : Au regard des résultats du test ***ADF***, **nous confirmons la non stationnarité de la série « R » et concluons qu'elle est un TS/Trend Stationnairy**[1]. En effet, |ADF|< |McKinnon| ; Trend significatif et |coefficient. $LR(-1)$| $\in]0,1[$ et est significatif. Autrement dit, cette série est victime d'une tendance déterministe et sa stationnarisation nécessite *l'écart à la tendance* comme méthode appropriée. Même résultat avec ***Philips-Perron (PP)***.

[1] *En d'autres termes, l'on dira : notre série est générée par un processus non stationnaire du type TS.*

- **Test PP (Philips-Perron)**



Null Hypothesis: LR has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 2 (Newey-West using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-2.120027	0.5263
Test critical values:	1% level	-4.083355	
	5% level	-3.470032	
	10% level	-3.161982	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000283
HAC corrected variance (Bartlett kernel)	0.000295

Phillips-Perron Test Equation
Dependent Variable: D(LR)
Method: Least Squares
Date: 05/21/14 Time: 16:50
Sample (adjusted): 2002M01 2008M04
Included observations: 76 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LR(-1)	-0.104671	0.050180	-2.085924	0.0405
C	1.360214	0.648367	2.097906	0.0394
@TREND(2001M12)	0.001155	0.000533	2.168718	0.0334

R-squared	0.061667	Mean dependent var	0.010105
Adjusted R-squared	0.035959	S.D. dependent var	0.017496
S.E. of regression	0.017179	Akaike info criterion	-5.251602
Sum squared resid	0.021543	Schwarz criterion	-5.159600
Log likelihood	202.5609	F-statistic	2.398750
Durbin-Watson stat	1.899753	Prob(F-statistic)	0.097958

I.2. Stationnarisation de la série

- Transformations de la série : le recourt à l'écart à la tendance est la procédure de stationnarisation appropriée pour « R », l'avons-nous dit. Pour ce faire : (i) régresser LR sur le temps et, (ii) récupérer le résidu de cette estimation qui correspond à notre série « LR » stationnarisée/transformée (soit « W »). Sur EViews, faire :

Genr T=@trend
LS LR C T
Genr W=resid

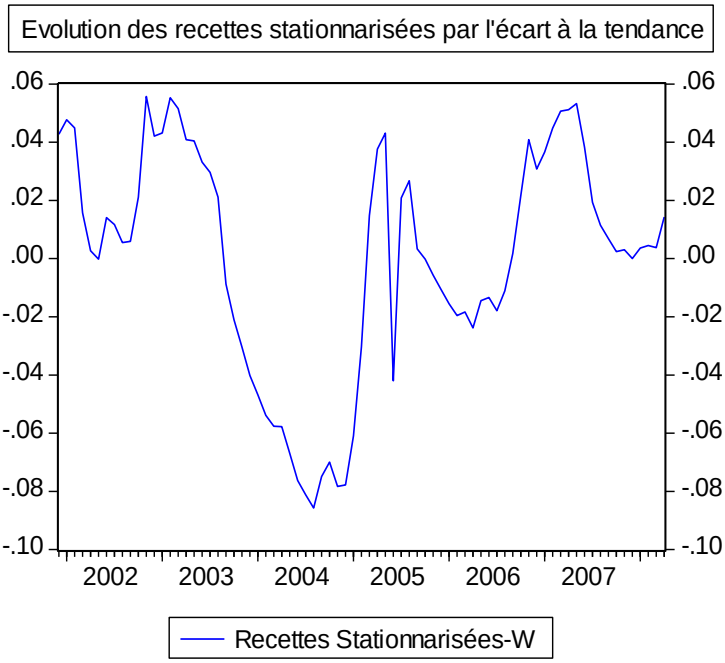
$$LRt = \hat{c} + \hat{d} * T + Wt$$
$$Wt = LRt - \hat{c} - \hat{d} * T = LRt - \widehat{LRt}$$

Dependent Variable: LR
Method: Least Squares
Date: 05/21/14 Time: 15:25
Sample: 2001M12 2008M04
Included observations: 77

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.93071	0.008931	1447.844	0.0000
T	0.010479	0.000203	51.65391	0.0000

R-squared	0.972659	Mean dependent var	13.32892
Adjusted R-squared	0.972294	S.D. dependent var	0.237711
S.E. of regression	0.039567	Akaike info criterion	-3.596015
Sum squared resid	0.117416	Schwarz criterion	-3.535137
Log likelihood	140.4466	F-statistic	2668.127
Durbin-Watson stat	0.195627	Prob(F-statistic)	0.000000

- Plot de la série transformée : W est stationnaire (en tendance surtout).



- Stationnarité de la série transformée « W » : Test d'ADF (lag : sélection Automatique →Soit « 0 ») :

Null Hypothesis: W has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.112286	0.0341
Test critical values:	1% level	-2.595745	
	5% level	-1.945139	
	10% level	-1.613983	

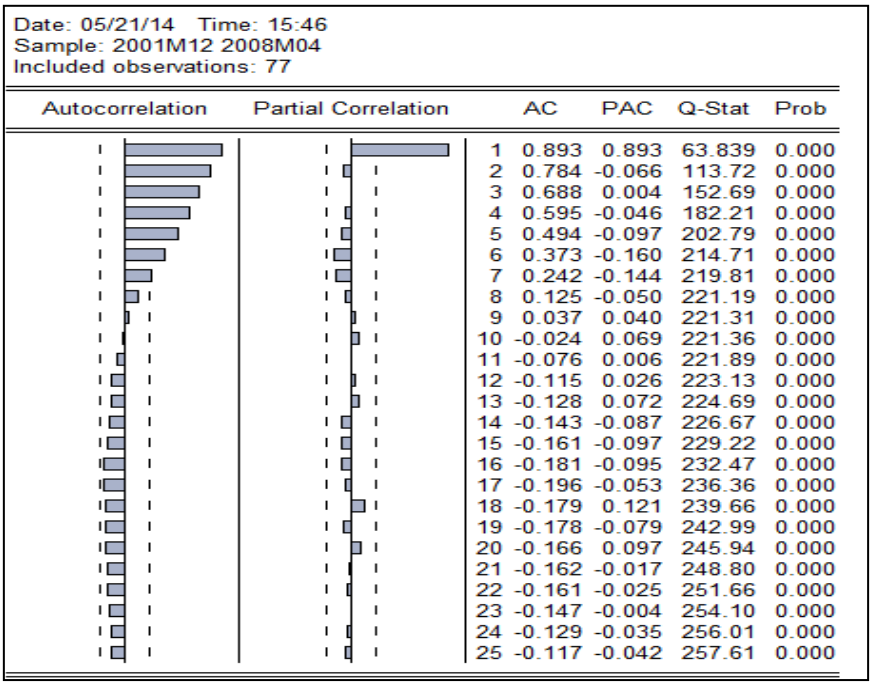
*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(W)
Method: Least Squares
Date: 05/21/14 Time: 19:41
Sample (adjusted): 2002M01 2008M04
Included observations: 76 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
W(-1)	-0.104898	0.049661	-2.112286	0.0380

_____**Note** : Notre série W transformée par l'écart à la tendance est stationnaire sans trend, ni dérive (la tendance et la constante ont parues statistiquement non significatives, d'où leur extraction).

II. Identification du modèle optimal (provisoire) : ARIMA(1,0,0)



Date: 05/21/14 Time: 15:46
Sample: 2001M12 2008M04
Included observations: 77

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.893	0.893	63.839	0.000
		2	0.784	-0.066	113.72	0.000
		3	0.688	0.004	152.69	0.000
		4	0.595	-0.046	182.21	0.000
		5	0.494	-0.097	202.79	0.000
		6	0.373	-0.160	214.71	0.000
		7	0.242	-0.144	219.81	0.000
		8	0.125	-0.050	221.19	0.000
		9	0.037	0.040	221.31	0.000
		10	-0.024	0.069	221.36	0.000
		11	-0.076	0.006	221.89	0.000
		12	-0.115	0.026	223.13	0.000
		13	-0.128	0.072	224.69	0.000
		14	-0.143	-0.087	226.67	0.000
		15	-0.161	-0.097	229.22	0.000
		16	-0.181	-0.095	232.47	0.000
		17	-0.196	-0.053	236.36	0.000
		18	-0.179	0.121	239.66	0.000
		19	-0.178	-0.079	242.99	0.000
		20	-0.166	0.097	245.94	0.000
		21	-0.162	-0.017	248.80	0.000
		22	-0.161	-0.025	251.66	0.000
		23	-0.147	-0.004	254.10	0.000
		24	-0.129	-0.035	256.01	0.000
		25	-0.117	-0.042	257.61	0.000

Note : Après avoir produit le corrélogramme relatif à « W » (série stationnaire), le modèle provisoire identifié dans la famille ARIMA par principe de parcimonie et d'ergodicité – susceptible de reproduire au mieux le mode opératoire de la série « R » – est un ***ARIMA(1,0,0)*** *ou* ***AR(1)***. Ce dernier a paru optimal/adéquat dans le sens où il présente les valeurs AIC et SIC les plus minimales, comparativement à d'autres processus candidats tels : ARIMA(1,0,1) ; ARMA(2,1) ; ARMA(1,2) ; ARMA(1,3) ; MA(1) ; MA(2) ou MA(3). Reste à valider le modèle provisoire après estimation, pour confirmer/juger de son optimalité.

III. Estimation du modèle provisoire retenu : taper : ls W AR(1)

Dependent Variable: W
Method: Least Squares
Date: 05/21/14 Time: 15:41
Sample (adjusted): 2002M01 2008M04
Included observations: 76 after adjustments
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.895102	0.049661	18.02433	0.0000

R-squared	0.812403	Mean dependent var	-0.000562
Adjusted R-squared	0.812403	S.D. dependent var	0.039254
S.E. of regression	0.017002	Akaike info criterion	-5.297909
Sum squared resid	0.021680	Schwarz criterion	-5.267241
Log likelihood	202.3205	Durbin-Watson stat	1.887347

Inverted AR Roots	.90

Le modèle AR(1) théorique et estimé s'écrivent :

$$W_t = \emptyset_1 AR(1) + v_t \quad ou \quad W_t = \emptyset_1 W_{t-1} + v_t \,, avec\ v_t \sim iid(0, \sigma^2)$$

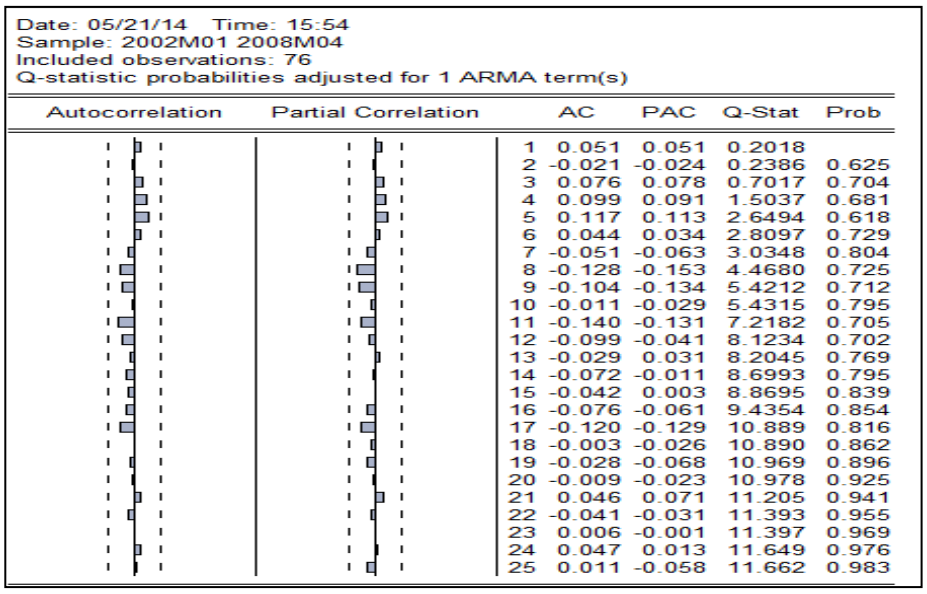
$$\widehat{\mathbf{W_t}} = \mathbf{0.895102} * \mathbf{W_{t-1}}$$
$$\mathbf{(t\text{-}stat) \qquad (18.02)}$$

Note : pour ce genre des modèles, la constante est souvent non significative (c'est le cas avec notre estimation). Ainsi, il convient de l'extraire. Du reste, notre modèle donne bonne mine (toutefois, attendons le verdict de tests diagnostics).

IV. Validation du modèle estimé

Cette étape est d'autant plus cruciale qu'elle sanctionne l'identification du modèle ; laquelle identification, si elle n'est pas bien faite, peut conduire à répéter le cycle « *identification-estimation-validation* » autant de fois jusqu'à ce que le modèle soit valide (*d'où nécessité de bien identifier le modèle adéquat*). Ci-dessous quelques tests de nature à juger de la qualité/optimalité du modèle estimé :

- Test de bruit blanc/Test portemanteau (Q-Q plot simple) : Sur EViews, dans l'output de l'estimation, suivre : *View/Residual Tests/Correlogram-Q-statistics→lags : 25→ok :*



	AC	PAC	Q-Stat	Prob
1	0.051	0.051	0.2018	
2	-0.021	-0.024	0.2386	0.625
3	0.076	0.078	0.7017	0.704
4	0.099	0.091	1.5037	0.681
5	0.117	0.113	2.6494	0.618
6	0.044	0.034	2.8097	0.729
7	-0.051	-0.063	3.0348	0.804
8	-0.128	-0.153	4.4680	0.725
9	-0.104	-0.134	5.4212	0.712
10	-0.011	-0.029	5.4315	0.795
11	-0.140	-0.131	7.2182	0.705
12	-0.099	-0.041	8.1234	0.702
13	-0.029	0.031	8.2045	0.769
14	-0.072	-0.011	8.6993	0.795
15	-0.042	0.003	8.8695	0.839
16	-0.076	-0.061	9.4354	0.854
17	-0.120	-0.129	10.889	0.816
18	-0.003	-0.026	10.890	0.862
19	-0.028	-0.068	10.969	0.896
20	-0.009	-0.023	10.978	0.925
21	0.046	0.071	11.205	0.941
22	-0.041	-0.031	11.393	0.955
23	0.006	-0.001	11.397	0.969
24	0.047	0.013	11.649	0.976
25	0.011	-0.058	11.662	0.983

Les hypothèses du test sont :

H0 : les résidus sont des bruits blancs (prob-ki^{ème} lag>5%)
H1 : les résidus ne sont pas des bruits blancs (prob-ki^{ème} lag<5%)

Note : Partant du corrélogramme ci-dessus, nous acceptons H0 (les résidus du modèle estimé sont des bruits blancs : sont-ils gaussiens[1]??).

[1] *Cfr Test de normalité des erreurs (où nous répondons à cette question).*

- Condition de stationnarité

Soit notre modèle théorique : $W_t = \emptyset_1 W_{t-1} + v_t$ ... [a] . Tester les conditions de stationnarité sur le modèle (a) estimé revient à vérifier que les racines caractéristiques du polynôme de retard associé à ce modèle – soit le polynôme : $(1 - \emptyset_1 L) = 0$ – sont **toutes** supérieures à 1 en valeur absolue (càd : $|L| > 1$) et que le paramètre « $\emptyset_1$ » estimé soit, en valeur absolue, inférieur à l'unité (compris dans le cercle unité du plan complexe. Càd : $|\widehat{\emptyset}_1| < 1$). Ainsi, écrivons notre polynôme retard d'ordre 1 (*degré du lag optimal : L^1*) et calculons la racine caractéristique associée « L » comme suit $(\emptyset(L) = 0, avec\ |L_j| > 1)$:

$1 - 0.895102 * L = 0 \rightarrow L = -1/-0.895102 \rightarrow \mathbf{L = 1.1172 > 1}$ ***(avec*** $|\widehat{\emptyset}_1| < 1$***).***

_____Note : Notre modèle estimé respecte les conditions de stationnarité (*le modèle est convergent ; d'où, la série est stationnaire*).

- Test de Student des paramètres (significativité statistique)

Le paramètre « $\emptyset_1$ » estimé est statistiquement significatif (prob-t student<5%|tc|>|tt|).

- Test de linéarité : Dans l'output, suivre : *View/Stability Test/Ramsey RESET Test…→Fitted : 2→ok.*

Ramsey RESET Test:			
F-statistic	2.893861	Probability	0.061737
Log likelihood ratio	5.798628	Probability	0.055061

Dependent Variable: W
Method: Least Squares
Sample: 2002M01 2008M04
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FITTED^2	-5.584528	3.017266	-1.850857	0.0682
FITTED^3	44.56470	46.09178	0.966869	0.3368
AR(1)	0.856806	0.062917	13.61807	0.0000

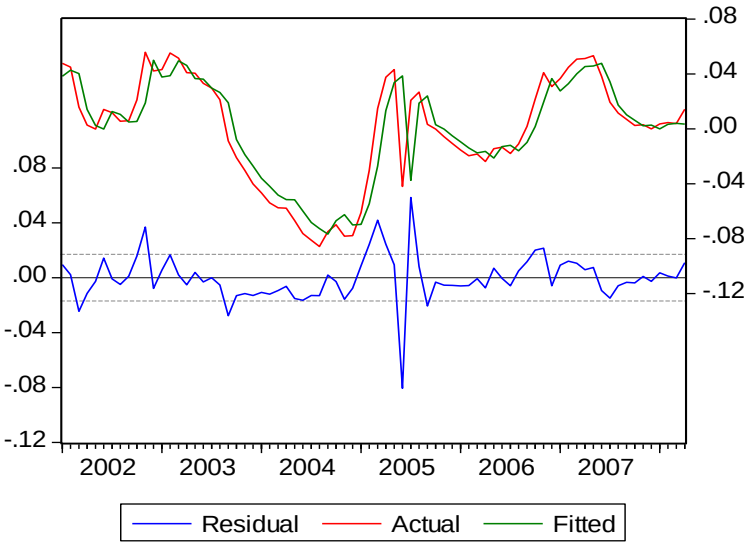
Les hypothèses du test sont :

H0 : le modèle est linéaire/bien spécifié (prob.F-stat > 5%)
H1 : le modèle n'est pas linéaire/mal spécifié (prob.F-stat < 5%)

_____*Note* : Nous acceptons H0 : notre modèle estimé est bien spécifié (linéaire).

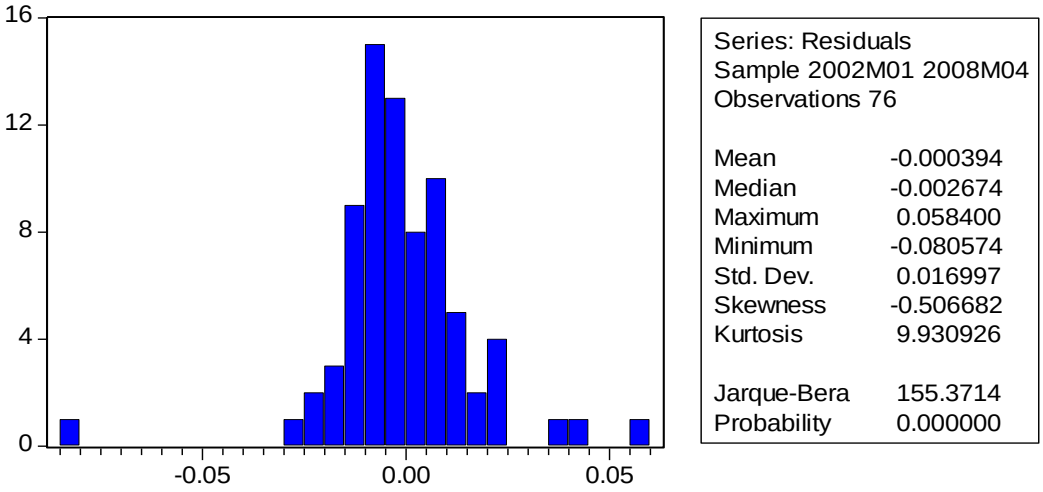
- Performance du modèle estimé

Nous la lisons en comparant les valeurs prévues calculées aux données réelles dans l'échantillon (l'on parle mieux de la « *prévision in casting* »). Pour ce faire, dans l'output de l'estimation du modèle ARIMA(1,0,0), cliquer sur : *Actual, Fitted, Residual/Actual, Fitted, Residual Graph*[1] *:*



Note : Visiblement, l'ajustement est bon à quelques écarts près.

- Test de Normalité : Dans l'output, suivre : *View/Residual Tests/Histogram - Normality Test*



Les hypothèses du test sont :

H0 : les erreurs sont des bruits blancs Gaussiens (prob-JB > 5%)
H1 : les erreurs ne sont pas des bruits blancs Gaussiens (prob-JB < 5%)

Note : Nous rejetons H0 : les résidus du modèle estimé ne sont pas des bruits blancs *Gaussiens (ne sont pas normalement distribués).*

[1] *Cfr Prévision pour les détails de calcul (valeurs ajustées et résidus).*

- Test d'absence d'autocorrélation des erreurs

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.125995	Probability	0.881810
Obs*R-squared	0.220257	Probability	0.895719

Dependent Variable: RESID
Method: Least Squares
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.008699	0.061419	-0.141635	0.8878
RESID(-1)	0.061232	0.131077	0.467143	0.6418
RESID(-2)	-0.014633	0.129588	-0.112921	0.9104

Les hypothèses du test sont :

H0 : Absence d'autocorrélation des erreurs (prob-BG > 5%)
H1 : Présence d'autocorrélation des erreurs (prob-BG < 5%)

Note : Nous acceptons H0 : les erreurs ne sont pas auto-corrélées.

V. *Prévision (h = 3 mois)*

a) *Calculs préliminaires*

$$Wt = LRt - \hat{c} - \hat{d} * T$$

$$LRt = Wt + \hat{c} + \hat{d} * T$$

$$LR_{t+h} = W_{t+h} + \hat{c} + \hat{d} * (T + h)$$

Avec :

$$\widehat{W}_t = \widehat{\emptyset}_1 \mathrm{W}_{\mathrm{t-1}}$$

$$\widehat{W}_{t+h} = \widehat{\emptyset}_1 \mathrm{W}_{\mathrm{t+h-1}}$$

Enfin, trouvons :

$$\boldsymbol{R_{t+h} = e^{(LR_{t+h})}}$$

b) *Prévision (ponctuelle pour h=3 mois)*

- Pour h=1 (moi de « *Mai* » : 05/2008) :

$$\widehat{W}_{t+1} = \widehat{\emptyset}_1 \mathrm{W_t} \rightarrow \boldsymbol{\widehat{W}_{08:05}} = 0.895102 * \mathrm{W}_{08:04} = 0.895 * 0.0000094 = 0.00000841 = 8{,}41\text{e-}06$$

$$LR_{t+1} = W_{t+1} + \hat{c} + \hat{d} * (T+1) \rightarrow \boldsymbol{LR_{08:05}}$$
$$= W_{08:05} + 12.93071 + 0.010479 * (76+1)$$
$$= 0.00000841 + 13.737593 \rightarrow \boldsymbol{LR_{08:05}} = 13.73760141$$

$$R_{t+1} = e^{(LR_{t+1})} \rightarrow \boldsymbol{R_{08:05}} = e^{(LR_{08:05})} = e^{(13.73760141)} = \mathbf{925\,048.4653}$$

- Pour h=2 (moi de « *Juin* » : 06/2008) :

$$\widehat{W}_{t+2} = \widehat{\emptyset}_1 \mathrm{W}_{t+1} \rightarrow \widehat{\boldsymbol{W}}_{\mathbf{08:06}} = 0.895102 * \mathrm{W}_{08:05} = 0.895 * 0.00000841 =$$
$$0.00000753 = 7{,}53\text{e-}06$$

$$LR_{t+2} = W_{t+2} + \hat{c} + \hat{d} * (T+2) \rightarrow \boldsymbol{LR_{08:06}}$$
$$= W_{08:06} + 12.93071 + 0.010479 * (76+2)$$
$$= 0.00000753 + 13.748072 \rightarrow \boldsymbol{LR_{08:06}} = 13.74807953$$

$$R_{t+2} = e^{(LR_{t+2})} \rightarrow \boldsymbol{R_{08:06}} = e^{(LR_{08:06})} = e^{(13.74807953)} = \mathbf{934\,792.1929}$$

- Pour h= 3 (moi de « *Juillet* » : 07/2008) :

$$\widehat{W}_{t+3} = \widehat{\emptyset}_1 \mathrm{W}_{t+2} \rightarrow \widehat{\boldsymbol{W}}_{\mathbf{08:07}} = 0.895102 * \mathrm{W}_{08:06} = 0.895 * 0.00000753 =$$
$$0.00000674 = 6{,}74\text{e-}06$$

$$LR_{t+3} = W_{t+3} + \hat{c} + \hat{d} * (T+3) \rightarrow \boldsymbol{LR_{08:07}}$$
$$= W_{08:07} + 12.93071 + 0.010479 * (76+3)$$
$$= 0.00000674 + 13.758551 \rightarrow \boldsymbol{LR_{08:06}} = 13.75855774$$

$$R_{t+3} = e^{(LR_{t+3})} \rightarrow \boldsymbol{R_{08:07}} = e^{(LR_{08:07})} = e^{(13.75855774)} = \mathbf{944\,638.6383}$$

_____Tableau des valeurs réelles, ajustées, et prévues

Période	***h***	$t+h$	***WP***	LR_t	R_t	***Nature***
03/2008	-1	75	1,05e-05	13.72043255	909 302	*Valeurs Réelles*
04/2008	0	76	9,40e-06	13.74142259	928 590	
05/2008	*1*	*77*	*8,41e-06*	*13.73760141*	***925 048***	*Valeurs prévues*
06/2008	*2*	*78*	*7,53e-06*	*13.74807953*	***934 792***	
07/2008	*3*	*79*	*6,74e-06*	*13.75855774*	***944 639***	

_____Sur EViews, pour obtenir « wp » (w prévu/prédit) :

Dans le workfile, double cliquer sur « Range »→En date : 2008M09→Ok→Yes
Ls W AR(1)
Dans l'output des résultats, cliquer sur «Forecast» →Forecast name : WP→Ok
Show WP

PART 2 : PRATIQUE SUR STATA

I. Analyse exploratoire des données[1]

I.1. Stationnarité et structure de la série

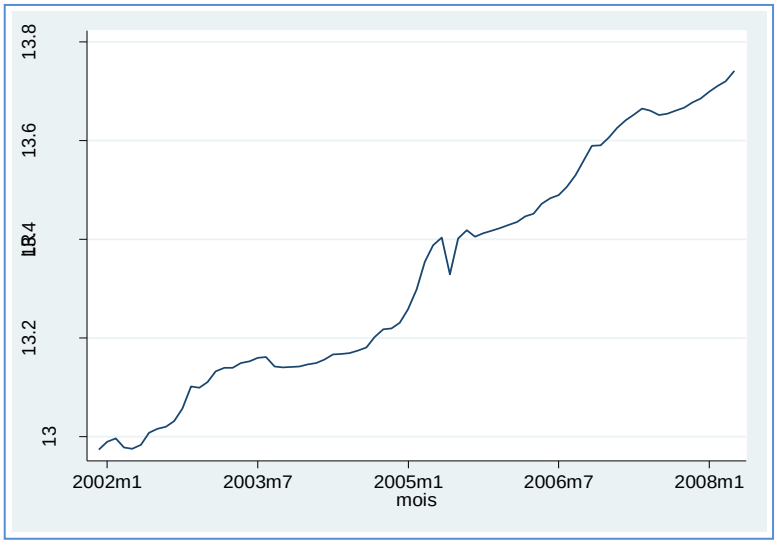
a) Tests préliminaires (informels)

- *Représentation graphique*

Déclaration de la nature temporelle (mensuelle) de nos données à Stata :

- Calcul : $2001 - 1960 = 41\ ans \times 12 = 492 + 11 = 503$ (valeur de base ou code numérique du début de la variable temps)
- Créer une variable temporelle « mois », avec 503 comme valeur de base/début[2] (les valeurs sont : 503 ; 504 ; 505 ; 506 ; 507 ;…).
- Sur Stata, taper : `format var1 %tm`

	R	var1
1	430814	2001m12
2	437540	2002m1
3	440909	2002m2
4	432740	2002m3
5	431640	2002m4
6	434916	2002m5
7	445830	2002m6
8	449446	2002m7
9	451379	2002m8
10	456330	2002m9



- Après avoir renommé la variable temporelle « var1 » (elle est nommée « mois »), déclarer/préciser la dimension temporelle à Stata (le message ci-dessous doit apparaître) : `tsset mois`

```
time variable:  mois, 2001m12 to 2008m4
```

- Générer la variable en logarithme : `gen LR=log(R)`
- Obtenir le graphique en tapant : `twoway (tsline LR)`

- *Comparaison des différents moments statistiques (Ecart-type et Moyenne)*

Commande Stata : *`reg ECT MOY, nocons`*

[1] Dans « Part 2 », nous nous limitons à lister les syntaxes et à présenter les outputs sur Stata. Pour les commentaires, Cfr Pratique sur EViews (Part 1).
[2] En fait, 503 = code numérique du mois de décembre de l'année 2001 dans Stata.

- *Corrélogramme*

_____Taper : `corrgram LR, lags(26)`

```
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1         0.9619   1.0026   74.056  0.0000           |-------           |--------
2         0.9253  -0.0036    143.5  0.0000           |-------           |
3         0.8889   0.0661   208.45  0.0000           |-------           |
4         0.8500  -0.0427   268.66  0.0000           |------            |
5         0.8102  -0.0645   324.12  0.0000           |------            |
6         0.7694  -0.1043   374.84  0.0000           |------            |
7         0.7300  -0.0306   421.15  0.0000           |-----             |
8         0.6905   0.0705   463.18  0.0000           |-----             |
9         0.6510   0.1686    501.1  0.0000           |-----             |-
10        0.6117   0.1698   535.07  0.0000           |----              |-
11        0.5722   0.0798   565.25  0.0000           |----              |
12        0.5352   0.1931   592.05  0.0000           |----              |-
13        0.4979   0.1048   615.62  0.0000           |---               |
14        0.4612   0.0271   636.15  0.0000           |---               |
15        0.4267   0.0395   654.01  0.0000           |---               |
16        0.3932   0.0079   669.43  0.0000           |---               |
17        0.3602   0.0971   682.58  0.0000           |--                |
18        0.3272   0.2117   693.62  0.0000           |--                |-
19        0.2955   0.0333   702.78  0.0000           |--                |
20        0.2656   0.0443   710.31  0.0000           |--                |
21        0.2365  -0.0175   716.39  0.0000           |-                 |
22        0.2057  -0.1074   721.07  0.0000           |-                 |
23        0.1740  -0.0261   724.48  0.0000           |-                 |
24        0.1422   0.0542    726.8  0.0000           |-                 |
25        0.1107   0.0130   728.23  0.0000           |                  |
26        0.0781   0.0347   728.96  0.0000           |                  |
```

b) Tests formels (Test d'ADF et PP)

- Test ADF : `dfuller LR, trend regress lags(0)`

```
Dickey-Fuller test for unit root                    Number of obs   =        76

                                 ---------- Interpolated Dickey-Fuller ---------
                 Test            1% Critical       5% Critical      10% Critical
               Statistic            Value             Value             Value
-------------------------------------------------------------------------------
 Z(t)            -2.086             -4.093            -3.474            -3.164
-------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5540

-------------------------------------------------------------------------------
D.LR            |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+--------------------------------------------------------------
          LR    |
          L1.   |  -.1046721     .05018    -2.09   0.040    -.2046807   -.0046634
       _trend   |   .0011553   .0005327     2.17   0.033     .0000936    .002217
        _cons   |   1.360223   .6483696     2.10   0.039     .0680239    2.652421
-------------------------------------------------------------------------------
```

- Test Philips-Perron (PP) : `pperron LR, trend regress lags(0)`

```
Phillips-Perron test for unit root                 Number of obs   =        76
                                                   Newey-West lags =         0

                                 ---------- Interpolated Dickey-Fuller ---------
                 Test            1% Critical       5% Critical      10% Critical
               Statistic            Value             Value            Value
------------------------------------------------------------------------------
 Z(rho)          -7.955            -26.584           -20.268          -17.164
 Z(t)            -2.086             -4.093            -3.474           -3.164
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5540

------------------------------------------------------------------------------
LR           |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          LR |
         L1. |   .8953279     .05018    17.84   0.000     .7953193    .9953366
      _trend |   .0011553   .0005327     2.17   0.033     .0000936     .002217
       _cons |   1.360223   .6483696     2.10   0.039     .0680239    2.652421
------------------------------------------------------------------------------
```

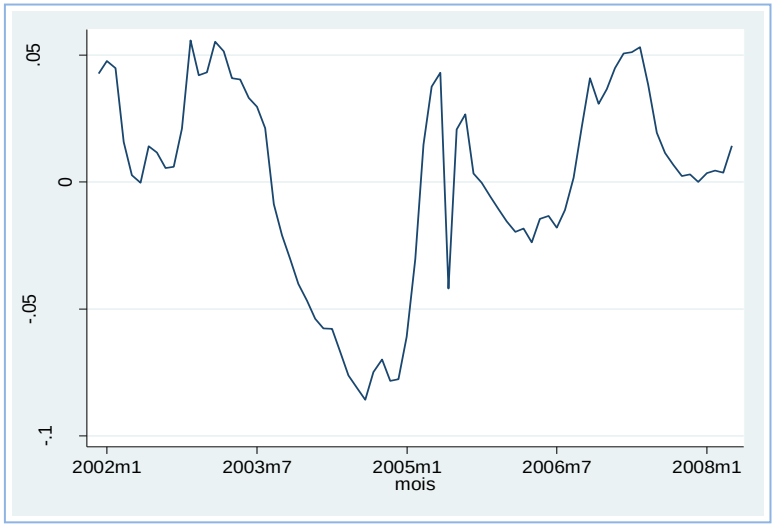
I.2. Stationnarisation de la série

- *Transformations de la série* :

```
reg LR mois
predict W, resid
```

```
      Source |       SS       df       MS              Number of obs =      77
-------------+------------------------------           F(  1,    75) = 2668.13
       Model |   4.1770681     1   4.1770681           Prob > F      =  0.0000
    Residual |  .117415615    75  .001565542           R-squared     =  0.9727
-------------+------------------------------           Adj R-squared =  0.9723
       Total |  4.29448372    76  .056506365           Root MSE      =  .03957
------------------------------------------------------------------------------
          LR |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        mois |   .0104792   .0002029    51.65   0.000      .010075    .0108833
       _cons |   7.659687   .1098467    69.73   0.000     7.440862    7.878513
------------------------------------------------------------------------------
```

- *Plot de la série transformée* : `twoway (tsline W)`



- *Test ADF sur la série transformée « W »* : commandes→

```
dfuller W, trend regress lags(0)
```

```
Dickey-Fuller test for unit root                    Number of obs   =        76

                                 ---------- Interpolated Dickey-Fuller ---------
                    Test         1% Critical       5% Critical      10% Critical
                Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.086            -4.093            -3.474            -3.164
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.5540

------------------------------------------------------------------------------
D.W          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           W |
         L1. |  -.1046721     .05018    -2.09   0.040    -.2046807   -.0046634
      _trend |   .0000584   .0000898     0.65   0.517    -.0001206    .0002375
       _cons |  -.0026437   .0039804    -0.66   0.509    -.0105766    .0052891
------------------------------------------------------------------------------
```

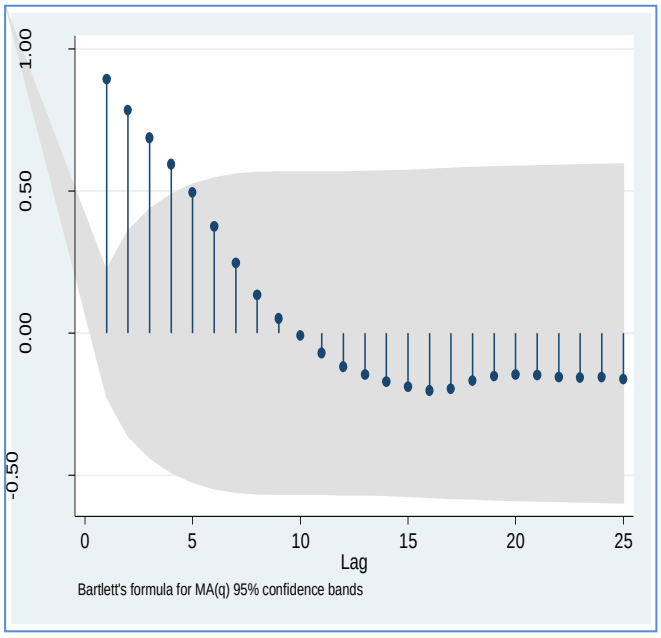
```
dfuller W, nocons regress lags(0)
```

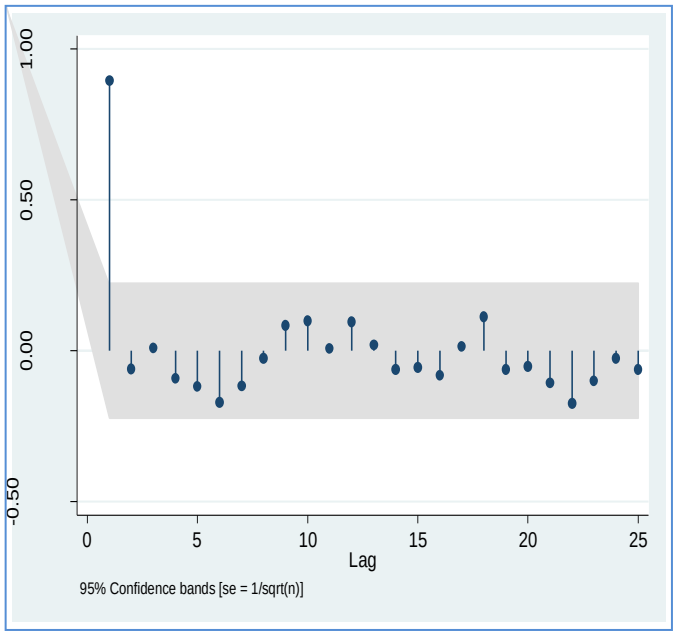
```
Dickey-Fuller test for unit root                    Number of obs   =        76

                                 ---------- Interpolated Dickey-Fuller ---------
                    Test         1% Critical       5% Critical      10% Critical
                Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.112            -2.610            -1.950            -1.610

------------------------------------------------------------------------------
         D.W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           W |
         L1. |  -.1048985   .0496609    -2.11   0.038    -.2038281   -.0059688
------------------------------------------------------------------------------
```

II. Identification du modèle optimal (provisoire)

_____Graphes des fonctions AC et PAC :

```
ac W, lags(25)
pac W, lags(25)
```





Fonctions AC et PAC : Graphes et coefficients calculés

```
corrgram W, lags(25)
```

LAG	AC	PAC	Q	Prob>Q	[Autocorrelation]	[Partial Autocor]
1	0.8935	0.8951	63.905	0.0000	|-------	|-------
2	0.7849	-0.0600	113.87	0.0000	|------	|
3	0.6891	0.0084	152.9	0.0000	|-----	|
4	0.5956	-0.0922	182.47	0.0000	|----	|
5	0.4946	-0.1184	203.13	0.0000	|---	|
6	0.3759	-0.1702	215.24	0.0000	|---	-|
7	0.2478	-0.1174	220.58	0.0000	|-	|
8	0.1344	-0.0240	222.17	0.0000	|-	|
9	0.0518	0.0848	222.41	0.0000	|	|
10	-0.0074	0.0998	222.42	0.0000	|	|
11	-0.0690	0.0066	222.85	0.0000	|	|
12	-0.1179	0.0966	224.15	0.0000	|	|
13	-0.1456	0.0192	226.17	0.0000	-|	|
14	-0.1703	-0.0615	228.97	0.0000	-|	|
15	-0.1886	-0.0554	232.46	0.0000	-|	|
16	-0.2011	-0.0815	236.49	0.0000	-|	|
17	-0.1949	0.0136	240.35	0.0000	-|	|
18	-0.1675	0.1122	243.24	0.0000	-|	|
19	-0.1518	-0.0622	245.66	0.0000	-|	|
20	-0.1450	-0.0523	247.9	0.0000	-|	|
21	-0.1475	-0.1063	250.27	0.0000	-|	|
22	-0.1540	-0.1743	252.89	0.0000	-|	-|
23	-0.1567	-0.1006	255.66	0.0000	-|	|
24	-0.1540	-0.0249	258.38	0.0000	-|	|
25	-0.1613	-0.0610	261.42	0.0000	-|	|

III. Estimation du modèle provisoire retenu « ARIMA(1,0,0)

Commande : `arima W, noconstant arima(1,0,0)`

```
(setting optimization to BHHH)
Iteration 0:   log likelihood =  204.03873
Iteration 1:   log likelihood =  204.04041
Iteration 2:   log likelihood =  204.04057
Iteration 3:   log likelihood =  204.04059
Iteration 4:   log likelihood =  204.04059

ARIMA regression

Sample:  2001m12 to 2008m4                      Number of obs      =        77
                                                Wald chi2(1)       =    156.95
Log likelihood =  204.0406                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |                 OPG
           W |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ar |
        L1. |   .8977319   .0716587    12.53   0.000     .7572833     1.03818
-------------+----------------------------------------------------------------
       /sigma |   .0169175   .0009564    17.69   0.000     .0150431    .0187919
------------------------------------------------------------------------------
```

Autre commande (préférable) : `reg W L.W, nocons`

```
      Source |       SS       df       MS              Number of obs =      76
-------------+------------------------------           F(  1,    75) =  324.87
       Model |  .093910459     1  .093910459           Prob > F      =  0.0000
    Residual |  .021680064    75  .000289068           R-squared     =  0.8124
-------------+------------------------------           Adj R-squared =  0.8099
       Total |  .115590524    76  .001520928           Root MSE      =    .017

------------------------------------------------------------------------------
           W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           W |
        L1. |   .8951015   .0496609    18.02   0.000     .7961719    .9940312
------------------------------------------------------------------------------
```

IV. Validation du modèle estimé

- *Test d'hétéroscédasticité*[1] *:* `estat archlm`

```
LM test for autoregressive conditional heteroskedasticity (ARCH)
------------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+----------------------------------------------------------------
       1     |          9.972               1                      0.0016
------------------------------------------------------------------------------
         H0: no ARCH effects      vs.  H1: ARCH(p) disturbance
```

[1] Notre modèle ARIMA(1,0,0) estimé souffre d'hétéroscédasticité : prob-chi2 < 5% (c'est un problème moins tributaire/moins grave pour les séries temporelles).

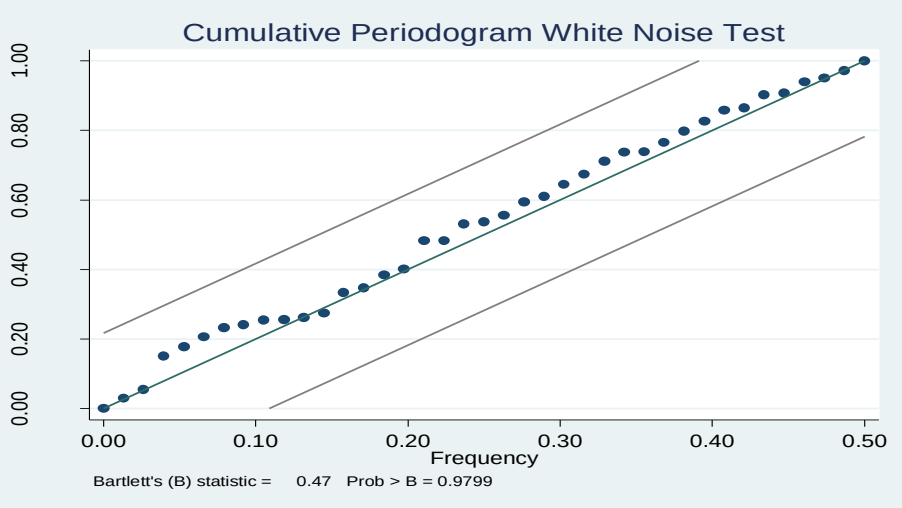
- *Test de bruit blanc (Test portemanteau) :*

*Commande (Test de **Box-Ljung**)*

```
reg W L.W, nocons
predict e, res
wntestq
```

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    28.4631
 Prob > chi2(36)           =     0.8103
```

*Autre commande (test de **Bartlett**[1]) :* `wntestq`



- *Test de linéarité*

```
reg W L.W, nocons
predict WP, xb
gen WP2=WP^2
gen WP3=WP^3
reg W  WP2 WP3 L.W, nocons
```

```
      Source |       SS       df       MS              Number of obs =      76
-------------+------------------------------           F(  3,    73) =  107.87
       Model |  .094314351     3  .031438117           Prob > F      =  0.0000
    Residual |  .021276172    73  .000291454           R-squared     =  0.8159
-------------+------------------------------           Adj R-squared =  0.8084
       Total |  .115590524    76  .001520928           Root MSE      =  .01707

------------------------------------------------------------------------------
           W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         WP2 |  -.7995694   1.748166    -0.46   0.649    -4.283658     2.68452
         WP3 |   17.57805   52.44961     0.34   0.738    -86.95387      122.11
           W |
          L1. |   .8339182    .125565     6.64   0.000     .5836675    1.084169
------------------------------------------------------------------------------
```

[1] Les résidus sont des bruits blancs (accepter H0 : Cfr Test de Box-Ljung), évoluant dans les limites/intervalle de confiance (Cfr Graphique : Test de Bartlett).

- *Performance du modèle estimé*

```
label var e "Error ARIMA"
label var W "LR Stationnaire"
line W WP e mois, sort
```

- *Test de Normalité des résidus :*

Commande (Test de Jarque-Berra) : `sktest e`

```
                    Skewness/Kurtosis tests for Normality
                                                   ------- joint ------
    Variable |  Pr(Skewness)   Pr(Kurtosis)  adj chi2(2)     Prob>chi2
-------------+-------------------------------------------------------
           e |      0.062          0.000          18.50        0.0001
```

Commande (Test de Shapiro-Wilk) : `swilk e`

```
                    Shapiro-Wilk W test for normal data
    Variable |     Obs        W          V          z      Prob>z
-------------+-------------------------------------------------
           e |      76    0.87218      8.412      4.653   0.00000
```

- *Test d'absence d'autocorrélation des erreurs*

Commande (autocorrélation d'ordre 1) : `estat durbinalt`

```
Durbin's alternative test for autocorrelation
------------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+----------------------------------------------------------------
       1     |          0.242               1                   0.6225
------------------------------------------------------------------------------
                        H0: no serial correlation
```

Commande (autocorrélation d'ordre 2): `estat bgodfrey`

```
Breusch-Godfrey LM test for autocorrelation
------------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+----------------------------------------------------------------
       1     |          0.248               1                   0.6184
------------------------------------------------------------------------------
                        H0: no serial correlation
```

VI. Prévision (Cfr Part 1 : Pratique sur EViews)

***