

M2 – Ingénierie Economique et Financière



PROJET D'ECONOMETRIE

AMATO Virgile – LAKEHAL AYAT Maïssane

## **Introduction**

Nous avons choisi d'étudier et de modéliser, dans ce projet, l'évolution hebdomadaire du VIX, indice phare de la volatilité des marchés, et calculé comme la volatilité implicite du S&P 500, principal indice action des Etats-Unis.

La période d'observation va du 24/10/1997 au 13/10/2017, pour un échantillon total de 1042 observations.

Notre choix d'étude s'est porté sur le VIX afin de modéliser la volatilité globale du marché, le VIX en étant un bon indicateur.

En effet, nous pensions qu'une non-linéarité pourrait être observée sur cet indicateur, dans la mesure où, sur la période considérée, les marchés financiers ont dû faire face à de nombreuses crises (bulle technologique et immobilière des années 2000, crises des sub-primes en 2008 et crise de la dette en 2010), qui pourrait logiquement faire varier la volatilité dans deux régimes distincts.

On peut facilement imaginer un modèle à 2 régimes, le premier, classique, pour les périodes où l'économie mondiale est dans un période de croissance, et un régime secondaire en période de crises économiques et financières.

Afin de modéliser, d'estimer, et de tester les caractéristiques du modèle non-linéaire SETAR associé au VIX hebdomadaire entre 1997 et 2017, nous procéderons de la manière suivante :

- Dans un premier temps, étude de la stationnarité de la série.
- Si nécessaire, stationnarisation de la série.
- Estimation du modèle ARMA correspondant, en essayant toutefois de travailler sur un modèle AR, afin de pouvoir estimer un modèle SETAR par la suite.
- Estimation du modèle SETAR associé à la série, avec un décalage temporel de la variable de changement de régime de 1.
- Test du modèle AR contre le modèle SETAR.
- Estimation du décalage temporel optimal pour la variable de changement de régime, en minimisant la probabilité critique du test AR vs SETAR.
- Enfin, test d'un modèle AR contre un modèle STAR.

L'intégralité du projet et des conclusions sont tirés de routines et de fonctions Matlab, dont les résultats ont par la suite été vérifiés avec EViews.

Aussi, on choisira, pour les tests statistiques de ce projet, un risque de première espèce  $\alpha = 5\%$ .

Enfin, dans un souci de clarté, les tests seront résumés, tout au long du projet, dans des tableaux produit via Microsoft Excel, ces résultats pourront bien évidemment être vérifiés, avec les routines et fonctions Matlab associées, ou encore à l'aide des sorties EViews en Annexe.

**Question 1 : Choisissez une série financière ou macroéconomique. Vérifier la stationnarité de cette série à l'aide de tests de racine unitaire ou de stationnarité. Par la suite, on travaillera avec la série stationnarisée.**

Dans un premier temps, nous commençons par étudier la stationnarité de la série, et ce afin de travailler par la suite sur une série stationnaire.

Nous procéderons à cet effet à deux tests :

- Test de Dickey-Fuller Augmenté (ADF)  
Afin de tester la présence de racine unitaire dans le modèle, et de déterminer le type de série auquel nous avons à faire (DS ou TS).
- Test de Philippe Perron (PP)  
Afin de vérifier la stationnarité du modèle issu du test ADF.

### 1.1) Test ADF

Les détails du test sont disponibles en Annexe 1.1.  
Nous récapitulons le test dans le tableau ci-dessous :

Test ADF						
	Racine Unitaire			Significativité du Coefficient		
	T-stat	Valeur Tabulée	P-critique	T-stat	Valeur Tabulée	P-critique
Modèle 3	-5,3	-3,41	0,01%	1,58	1,96	11,36%
Modèle 2	-5,05	-2,86	0%	4,74	1,96	0%
Modèle 1	-1,75	-1,94	7,55%	-1,75	1,96	7,98%

Attention, la valeur tabulée pour le test de significativité du coefficient est à chercher dans la table de DF si l'hypothèse de racine unitaire n'a pu être rejetée, mais se trouve dans la table de Student si l'hypothèse de racine unitaire est rejetée.

#### Modèle 3 :

→ Racine unitaire : Test de Racine Unitaire.

Rejet de l'hypothèse nulle de racine unitaire :

Par la valeur critique :  $-5,3 > -3,41$

Par la P-critique :  $0,01\% < 5\%$

→ Significativité du coefficient : Test de Significativité Student.

Non rejet de l'hypothèse nulle de non-significativité du coefficient associé à la tendance :

Par la valeur critique :  $|1,58| < 1,96$

Par la P-critique :  $11,36\% > 5\%$

→ On ne choisit donc pas le modèle 3.

#### Modèle 2 :

→ Racine unitaire : Test de Racine Unitaire.

Rejet de l'hypothèse nulle de racine unitaire :

Par la valeur critique :  $-5,05 > -2,86$

Par la P-critique :  $0\% < 5\%$

→ Significativité du coefficient : Test de Significativité Student.

Rejet de l'hypothèse nulle de non-significativité du coefficient associé à la tendance :

Par la valeur critique :  $|4,74| > 1,96$

Par la P-critique :  $0\% < 5\%$

→ On choisit donc le modèle 2.

Nous sommes donc en présence d'un processus  $X_t$  tel que la variable aléatoire  $X_t$  :

$$X_t \sim I(0) + c$$

$$\text{Et : } \Delta X_t = c + \rho X_{t-1} + \varepsilon_t$$

## 1.2) Test PP

Les détails du test sont disponibles en Annexe 1.2.

Nous récapitulons le test dans le tableau ci-dessous :

Test PP						
	Racine Unitaire			Significativité du Coefficient		
	T-stat	Valeur Tabulée	P-critique	T-stat	Valeur Tabulée	P-critique
Modèle 3	-5,94	-3,41	0,00%	1,88	1,96	6,09%
Modèle 2	-5,61	-2,86	0%	5,6	1,96	0%
Modèle 1	-1,74	-1,94	7,77%	-2,19	1,96	2,85%

Les résultats et le choix du modèle issus du test ADF sont confirmés par le test PP.

→ Puisque notre série est  $I(0)$  (i.e. stationnaire), nous n'avons pas besoin de la stationnariser.

Nous pouvons donc passer à l'estimation du modèle ARMA correspondant.

**Question 2 : Estimez le modèle ARMA correspondant, en cherchant de préférence un modèle AR(p)**

Avec la série stationnaire ou stationnarisée obtenue à la question précédente, nous pouvons maintenant estimer le modèle ARMA (Auto Regressive and Moving Average Model) issu de cette série.

Nous procéderons, pour ce faire, de la manière suivante :

- Etudes préliminaires de l'autocorrélogramme et de l'autocorrélogramme partiel de la série afin d'approximer graphiquement les retards optimaux pour la partie AR.
- Estimation des paramètres du modèle ARMA à l'aide de la méthode du maximum de vraisemblance.

## 1 Etudes préliminaires : Autocorrélation et Autocorrélogramme

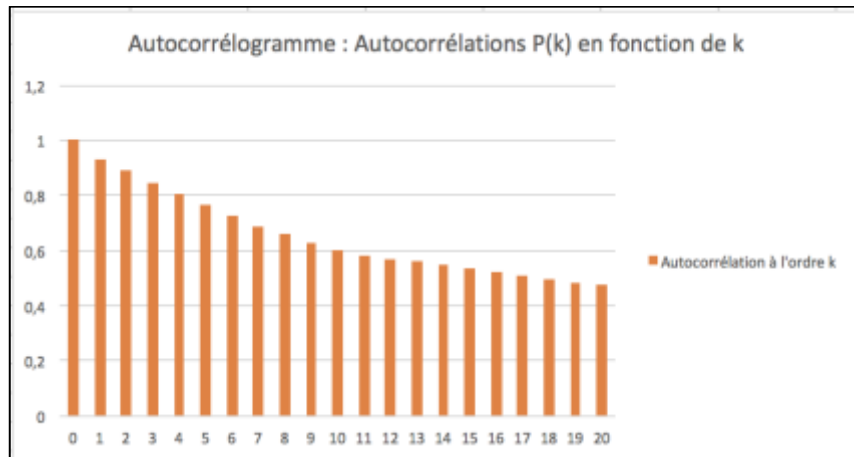
On commence par étudier l'autocorrélogramme (en niveau), pour étudier la corrélation de la variable avec elle-même à différents écarts de périodes (ce qu'on appelle l'autocorrélation). De même l'autocorrélogramme partiel permet d'étudier la corrélation de la variable avec elle-même à différents écarts de périodes, indépendamment des autres périodes.

### 1.1) Autocorrélogramme

L'autocorrélation est calculé par la formule suivante :

$$\rho_k = \frac{E((X_t - \mu)(X_{t+k} - \mu))}{\sigma^2}$$

L'autocorrélogramme est de la forme suivante :



On peut voir qu'il y'a une très forte autocorrélation à l'ordre 1, et que l'autocorrélation n'est pas linéairement décroissante : elle est très décroissante entre les décalage 1 et 6, puis décroît de moins en moins fortement pour finir par être presque constante.

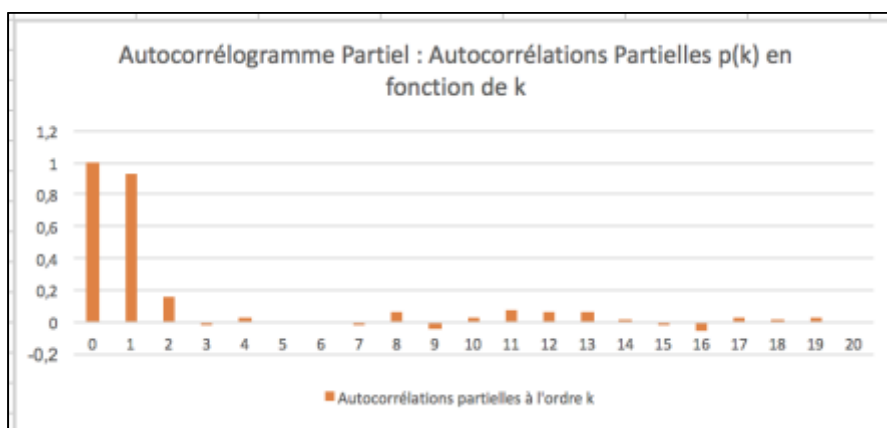
## 1.2) Autocorrélogramme Partiel

L'autocorrélation partielle est calculée par la formule suivante :

$$x_{t+1-m} = c_1(x_t - m) + c_2(x_{t-1} - m) + \dots + c_{k-1}(x_{t-k} - m) + p_k(x_{t-k+1} - m)$$

Où l'autocorrélation partielle  $p_k$  est le dernier coefficient de cette équation.

L'autocorrélogramme partiel est de la forme suivante :



On observe une très forte autocorrélation partielle à l'ordre 1, une autocorrélation partielle significative à l'ordre 2, puis des autocorrélations partielles faibles et alternativement positive et négative.

Nous décidons de tester le modèle ARMA(p, q) de notre série pour des décalages temporels p et q allant de 0 jusqu'à 3.

## 2 Estimation du modèle ARMA correspondant

On commence tout d'abord par sélectionner le nombre de retards p et q optimaux ( $p^*$  et  $q^*$ ) des parties AR et MA de notre modèle.

Pour cela, nous allons calculer le critère d'Aikaike (AIC) des modèles ARMA(p,q) de notre série pour des décalages temporels p et q allant de 0 jusqu'à 3.

Nous sélectionnerons ensuite le nombre de retards minimisant ce critère.

Nous pourrions ensuite estimer les paramètres optimaux du modèle ARMA( $p^*$ ,  $q^*$ )

### 2.1) Estimation des retards $p^*$ et $q^*$ du modèle ARMA(p,q)

On peut voir en Annexe 1.1 que EViews a, automatiquement, dans les tests ADF précédents, sélectionné un décalage temporel de 1 pour la série, sur un décalage maximum de 21.

On peut penser qu'à priori c'est le décalage optimal.

Pour choisir le nombre de retards p et q respectivement du processus AR(p) et MA(q), nous estimons les modèles ARMA(p, q) avec p et q allant de 0 à 3.

Nous choisissons ensuite le nombre de retards qui minimise le critère d'information de Akaike (AIC).

Le modèle est le suivant :

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \varphi_j \varepsilon_{t-j} ; \text{ Avec : } p, q \in \{0, 3\}$$

Le détail des 16 sorties EViews de chaque estimation est disponibles en Annexes 2.1. Voici un tableau récapitulatif de nos résultats :

AIC du modèle ARMA(p,q) en fonction de p et q				
p \ q	0	1	2	3
0	7.1506	6.3080	5.8436	5.5891
1	5.0944	5.0740	5.0750	5.0759
2	5.0718	5.0754	5.0761	5.0779
3	5.0744	5.0773	5.0787	5.0804

le nombre de retards  $p^*$  et  $q^*$  optimaux sont donc tels que :

- $p^* = 2$
- $q^* = 0$

Et le modèle suit un ARMA( $p^*$ ,  $q^*$ ) = AR( $p^*$ ) = AR(2), i.e. :

$$X_t \sim AR(2)$$

### 2.2) Estimation des coefficients du modèle

Voici un tableau récapitulatif de l'estimation du modèle AR(2) :

<b>AR(2)</b>				
	<b>Coefficient estimé</b>	<b>Ecart-Type</b>	<b>T-stat</b>	<b>P-critique</b>
$c$	1,165	0,246	4,737	0%
$X_{t-1}$	0,784	0,031	24,432	0%
$X_{t-2}$	0,159	0,031	5,171	0%
$R^2$	0,875			
$SCR$	9663,67			
$AIC$	5,072			
$SIC$	5,086			

Le modèle est donc de la forme suivante :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

Avec :

$$c = 1,165$$

$$\phi_1 = 0,784$$

$$\phi_2 = 0,159$$

Qui sont tous les 3 significatifs.

La sortie EViews de l'estimation du modèle AR(2) est disponible en Annexe 2.2.

De plus, on peut voir que le coefficient de détermination  $R^2$  est plutôt élevé (87,5%), ce qui montre un modèle de qualité.

### **2.3) Spécification du modèle**

Nous devons vérifier :

- Que tous les coefficients de notre modèle peuvent être considérés comme étant significatif, par un test de significativité de Student.
- Que nos erreurs ne sont pas autocorrélées, qu'elles sont homoscédastiques, et qu'elles peuvent être considérées comme suivant une Loi Normale.

#### **a) Significativité des coefficients (Test de Significativité de Student) :**

On peut, dans le tableau récapitulatif de l'estimation de l'AR(2), voir que nos coefficients sont tous significativement différent de 0 :

- Leurs p-values sont toutes inférieures à 5% (niveau de risque de première espèce choisi).
- Leurs t-stat sont toutes supérieures à 1,96 (valeur tabulée de Student pour un risque de 5%).

#### **b) Autocorrélation des erreurs (Test de Ljung Box ) :**

Voici un tableau récapitulatif du test de Ljung Box d'autocorrélation des erreurs :

Test de Ljung Box d'Autocorrélation des erreurs				
Décalages temporels	Autocorrélation	Autocorr. Partielle	Q-Stat	P-Critique
1	0,002	0,002	0,0049	94,40%
2	0,008	0,008	0,0767	95,20%
3	-0,027	-0,027	0,8273	84,30%
4	0,014	0,015	1,0454	90,30%
5	0,003	0,004	1,0582	95,80%
6	0,011	0,01	1,1915	97,70%
7	-0,054	-0,053	4,1995	75,70%
8	0,053	0,054	7,2048	51,50%
9	-0,044	-0,043	9,1988	41,90%
10	-0,074	-0,078	14,916	13,50%
11	-0,059	-0,054	18,584	6,90%
12	-0,029	-0,032	19,491	7,70%
13	0,027	0,026	20,273	8,90%
14	0,054	0,05	23,317	5,50%
15	0,048	0,056	25,784	4%
16	-0,021	-0,025	26,258	5%
17	-0,02	-0,003	26,261	7%
18	-0,08	-0,007	26,335	9,20%
19	0,035	0,0028	27,655	9%
20	0,055	0,051	30,891	5,70%

On a  $30,891 < 31,41$ , les résidus de notre modèle ne peuvent donc pas être considérés comme étant autocorrélés.

Le détail du test est disponible en Annexe 2.3.

### c) Homoscédastité des erreurs :

Voici un tableau récapitulatif du test de d'Homoscédasticité des erreurs :

Test d'Hétéroscédastité des erreurs				
Décalages temporels	Autocorrélation	Autocorr. Partielle	Q-Stat	P-Critique
1	0,158	0,158	26,094	0,00%
2	0,131	0,109	44,028	0,00%
3	0,358	0,334	177,94	0,00%
4	0,308	0,245	277,43	0,00%
5	0,069	-0,042	282,42	0,00%
6	0,074	-0,103	288,24	0,00%
7	0,247	0,081	352,27	0,00%
8	0,062	-0,036	356,27	0,00%
9	0,054	0,041	359,32	0,00%
10	0,084	-0,015	366,76	0,00%

On voit que :

- Les p-value sont toutes inférieures à 5%
- La q-stat à l'ordre 10 est de  $366,76 > 31,41$  (quantile du Khi2 correspondant)

→ Nous sommes donc ici en présence de résidus hétéroscédastiques, c'est-à-dire dont la variance n'est pas constante.

Le détail du test est disponible en Annexe 2.4.



**d) Test de normalité des erreurs (Test de Jarque-Bera) :**

La statistique de Jarque-Bera du modèle est égale à  $3134,284 > \chi^2_{0,05}(2) = 5,991$ . De même, sa probabilité critique de 0% est  $< 5\%$ .

On rejette donc l'hypothèse nulle de normalité des erreurs : on ne peut pas considérer que les erreurs suivent une loi Normale.

Le détail du test est disponible en Annexe 2.5.

→ Ainsi, les coefficients de notre modèle sont significatifs et les résidus de notre modèle sont :

- Non autocorrélés
- Hétéroscédastiques
- Non normaux

Le principal problème que nous voulions éviter était l'autocorrélation des erreurs, créant des modèles où la variance n'est plus minimum pour les estimateurs du modèle.

L'hétéroscédasticité et la non normalité des erreurs ne modifiant pas ni le biais ni la variance de nos estimateurs, nous pouvons conclure que le modèle est malgré tout bien spécifié.

**Question 3 : Estimez le modèle SETAR avec  $q_t = X_{t-1}$  comme variable de changement de régime. On suppose qu'il y'a deux régimes et que le nombre de retards dans chaque régime est celui de l'AR(p) de la question précédente.**

**1) Le modèle**

Nous avons défini, à la question précédente, que nous étudierions le modèle comme un processus AR(2).

Le modèle non linéaire TAR (Threshold AutoRegressive Model) associé est donc de la forme :

$$X_t = \begin{cases} c_1 + \sum_{i=1}^p \phi_{1,i} X_{t-i} + \varepsilon_t & , si \ q \leq c \\ c_2 + \sum_{i=1}^p \phi_{2,i} X_{t-i} + \varepsilon_t & , si \ q > c \end{cases}$$

Avec :

$c_1, c_2$  : les constantes des modèles 1 et 2 respectivement

$\phi_1, \phi_2$  : les vecteurs de coefficients des modèles 1 et 2 respectivement

$\varepsilon_t$  : le résidu

$q$  : variable de changement de régime

$c$  : seuil de changement de régime

$p$  : nombre de retards du modèle

Et où chaque régime suit un AR(2), on a donc  $p = 2$ .

De plus, nous estimons ici un modèle SETAR (Self Exciting TAR), qui est un modèle TAR où la variable de changement de régime  $q$  est la variable expliquée décalée dans le temps.

Nous procéderons à une estimation par MCO séquentiels, c'est-à-dire :

- Séries de MCO sur le modèle où l'on fait varier  $c$ , la variable de changement de régime, sur un ensemble de valeurs possibles
- On récupère la valeur de  $c$  qui minimise la variance résiduelle,  $c^*$ .
- Estimation des paramètres du modèle à l'aide du seuil de changement de régime optimal,  $c^*$ .
- Nous obtenons aussi les coefficients estimés optimaux du modèle SETAR.

## 2) Estimation du seuil de changement de régime optimale.

Nous commençons par mettre en place la régression de notre modèle, notamment en donnant forme à la matrice des régresseurs.

Les détails de la mise en place de l'estimation est disponible en Annexe 3.1.

Avec une réécriture du modèle à l'aide de variables indicatrices, on définit une matrice des régresseurs classiques nous permettant d'estimer les  $2 \cdot (k+1)$  paramètres à estimer (avec  $k$  le nombre de variables explicatives, soit dans un modèle AR le nombre de décalages temporels, ici, 2).

Afin d'avoir un modèle SETAR optimal, il est nécessaire de trouver le « vrai » seuil à partir duquel la variable expliquée peut être considérée comme étant régie par le régime 2.

Afin de trouver ce seuil, nous allons procéder de la façon suivante :

- On commence par trier par ordre croissant le vecteur de la variable expliquée.
- On supprime une certaine fraction en haut et en bas de cet échantillon, afin de supprimer les valeurs extrêmes.
- On essaye ensuite chacun des seuils dans le vecteur nouvellement créé, en effectuant une estimation des paramètres du modèle SETAR pour chaque seuil. On prend soin de récupérer l'écart type résiduel de chaque régression.
- On sélectionne ensuite le seuil  $c$  qui minimise l'écart type résiduel du modèle.

On aurait également pu définir un autre critère de sélection, comme par exemple la minimisation de l'AIC, la maximisation du  $R^2$ , etc....

On trouve à la suite de cet procédure la valeur de  $c$  qui minimise la variance du modèle.

$$c^* = 25,07$$

Voici un tableau montrant quelques résultats des valeurs prises par  $c$  et de la variance résiduel du modèle associé :

Estimation du seuil de changement de régime optimal	
Valeur de c	Variance résiduelle
12,62	9,3318
14,81	9,3186
16,04	9,3205
18,94	9,3061
21,76	9,3071
25,07	9,3022
27,11	9,3213
27,52	9,3302

La valeur de C qui minimise la variance résiduelle est 25,07, qui donne une variance résiduelle de 9,3022. C'est le seuil que nous choisissons comme seuil de changement de régime optimal.

### 3) Estimation du modèle optimal.

Doté maintenant du seuil optimal de changement de régime de notre modèle, nous pouvons déterminer les coefficients du modèle : ce sont les coefficients optimaux du modèle SETAR. Les coefficients optimaux de notre modèle sont les suivants :

SETAR(2,2) d = 1 c* = 25,07				
	Coefficient estimé	Ecart-Type	T-stat	P-critique
c1	0,546	0,196	1,230	11%
X(-1)	0,824	0,002	16,770	0%
X(-2)	0,160	0,002	3,628	0%
c2	0,743	0,597	0,962	17%
X(-1)	0,795	-0,015	17,098	0%
X(-2)	0,155	-0,002	3,582	0%
SCR	9627,8			

On voit que, les constantes mises à part, nos coefficients sont significativement différents de 0.

La différence entre les deux régimes tient surtout à la valeur de la constante, qui est légèrement plus importante pour le régime 2. Les deux coefficients associés à X(t-1) et X(t-2) sont presque similaires dans les deux régimes.

On constate ainsi que l'impact des valeurs passées (X(t-1) et X(t-2)) sur la valeur présente est quasiment constant entre les deux régimes. Seul le niveau de base, constant et indépendant des valeurs passées (la constante) est plus élevé et donne ainsi des valeurs du VIX plus élevées lorsqu'il se trouve dans le régime 2.

Aussi, on calcule à titre informatif la proportion d'occurrence du régime 1 et du régime 2 dans notre modèle.

On constate que dans 78% des cas, le modèle est dans le régime 1, tandis que dans 22% des cas il est dans le régime 2.

On peut donc supposer que le régime 2 représente les périodes plus agitées sur les marchés, sans pour autant qu'il corresponde exactement aux périodes de crises financières.

(Nous n'avons pas été 1 jour sur 5 en crise financière depuis 1997, et fort heureusement !).

Nous retrouvons ces résultats dans le graphique en Annexe 3.2, qui montre les occurrences du régime 2 en fonction du temps. On constate que celui-ci intervient, sans surprises, principalement entre :

- 1998 et 2004 (Apogée de la bulle technologique/bulle internet des années 2000, + bulle immobilière américaine)
- 2007 et 2012 (Crises des Subprimes en 2008, Crise de la dette en 2010).

Ce qui confirme bien notre hypothèse que le régime 2 correspond aux périodes de « turbulences » voir de « crises » sur les marchés financiers.

**Question 4 : Faites le test d'un modèle AR(p) (hypothèse nulle) contre un modèle SETAR(p) (hypothèse alternative).  
Reportez la probabilité critique de ce test et commentez.**

Afin de vérifier que nous utilisons un modèle AR à seuil (Threshold Auto Regressive Model : TAR) à bon escient, c'est-à-dire que nous n'utilisons pas un modèle SETAR non-linéaire alors qu'un modèle AR linéaire apporterait autant d'informations, nous allons effectuer le test d'un modèle AR contre un modèle SETAR.

Nous allons pour cela appliquer la procédure de test suivante :

- Calcul de la statistique du test à l'aide de l'estimation d'un modèle AR et du modèle SETAR correspondant
- Calcul de la probabilité critique du test par une méthode de simulation numérique
- Conclusion du test à l'aide de la zone de rejet

### 1) Paramétrage du test et calcul de la F-stat

Hypothèse du test :

$$\begin{cases} H0 : \text{Modèle AR, le modèle est linéaire} \\ H1 : \text{Modèle SETAR, le modèle est non linéaire} \end{cases}$$

Statistique du test :

$$F = T * \left( \frac{\widehat{\sigma}_{\varepsilon}^2 - \widehat{\sigma}_{\varepsilon}^2(\widehat{c})}{\widehat{\sigma}_{\varepsilon}^2(\widehat{c})} \right)$$

Avec :

- T = nombres d'observations disponibles
- $\widehat{\sigma}_{\varepsilon}^2(\widehat{c})$  = variance résiduelle du modèle SETAR
- $\widehat{\sigma}_{\varepsilon}^2$  = variance résiduelle du modèle AR

De la question 2), on sait :

$$\rightarrow \widehat{\sigma}_{\varepsilon}^2 = \frac{SCR}{(n-(k+1))} : \frac{9663,67}{(1041-3)} = 9,31$$

De même, de la question 3), on sait :

$$\rightarrow \widehat{\sigma}_{\varepsilon}^2(\widehat{c}) : 9,3022$$

$$d'o\grave{u} : F = 1041 * \left( \frac{9,31 - 9,3022}{9,3022} \right) = 0,8729$$

## 2) Estimation de la distribution de la F-stat par simulation numérique

Afin d'évaluer la distribution de probabilités de la F-stat, et d'ainsi pouvoir donner un intervalle de confiance pour la F-stat observé de la série, nous procédons à une optimisation numérique :

- Nous commençons par générer 20 séries aléatoires de variables suivant une loi normale centrée réduite
- Nous effectuons la régression, pour chaque tirage, du vecteur aléatoire gaussien contre les  $X(t-1)$  et  $X(t-2)$ , pour le modèle AR(2) comme pour le modèle SETAR(2,2).
- Nous récupérons de ces régressions la variance résiduelle, pour chaque tirage, des deux modèle AR et SETAR correspondant, et nous calculons la F-statistique comme nous l'avons fait à l'étape précédente.
- Nous obtenons ainsi une distribution de probabilités estimée pour la F-stat, et on peut de cette façon conclure sur le test.

Voici un tableau récapitulatif de l'estimation :

Estimation de la distribution de la F-stat		
Numéro du tirage aléatoire	F-stat	F-stat > 0,8729 ?
1	13,94578635	VRAI
2	5,478377991	VRAI
3	9,11726066	VRAI
4	2,715193407	VRAI
5	6,731185434	VRAI
6	22,51126606	VRAI
7	1,583665102	VRAI
8	3,66745193	VRAI
9	10,96283662	VRAI
10	1,623469317	VRAI
11	-0,905422908	FAUX
12	10,46407311	VRAI
13	7,190624934	VRAI
14	6,160107289	VRAI
15	10,83804353	VRAI
16	13,78496532	VRAI
17	0,7580457	FAUX
18	-3,406133908	FAUX
19	4,970456916	VRAI
20	4,344061513	VRAI
P(F-stat > 0,8729) = P-critique		85%
Conclusion du test		On ne rejette pas H0

## 3) Conclusion du test

Nous obtenons une probabilité critique, calculée comme la proportion des valeurs des F-statistiques simulées supérieures à la F-statistique observée de notre modèle.

Nous obtenons ici une probabilité critique de 85%, et pouvons ainsi conclure au non rejet de H0.

Nous pouvons considérer que le modèle SETAR non-linéaire n'apporte pas plus d'informations que le modèle AR linéaire.

Ainsi le VIX ne semble pas adopter deux régimes différents, et peut tout aussi bien être modélisé par un seul régime linéaire autorégressif, qui apporterait autant d'informations et aurait une même qualité de prédiction, que le modèle non linéaire SETAR.

Pour le bon déroulé de ce projet, nous continuerons néanmoins par la suite à travailler sur notre modèle non linéaire SETAR.

#### **Question 5 : Estimer le retard d optimal pour la variable de changement de régime**

Après avoir estimé le seuil de changement de régime,  $c$ , optimal, il nous reste encore un paramètre à optimiser.

En effet, nous avons, au début du projet, postuler de manière arbitraire que nous prendrions comme variable de changement de régime  $q_t = X_{t-d}$  (définition d'un modèle SETAR), avec  $d = 1$ .

Néanmoins, nous devons également optimiser ce paramètre  $d$ , afin de déterminer la variable de changement de régime optimale.

Nous aurons ainsi un modèle SETAR avec le seuil de changement de régime optimal ainsi que la variable de changement de régime optimale.

Nous allons donc :

- Estimer la valeur  $d^*$  de la variable de changement de régime,  $X_{t-d}$ .
- Estimer les paramètres de notre modèle final optimisé.

#### **1) Estimation du niveau $d^*$ optimal pour la variable de changement de régime $X_{t-d}$**

Afin d'estimer le retard optimal de la variable  $q_t = X_{t-d}$ , nous allons procéder de la manière suivante :

- Définition d'un intervalle des possibles pour la valeur  $d$ . Nous essayerons les valeurs 1 à 5.
- Régression SETAR en fonction de chaque valeur possible de  $d$ , en prenant soin de calculer, pour chaque nouvelle valeur de  $d$ , le seuil de changement de régime optimal  $c^*$ .
- Test d'un modèle AR contre le modèle SETAR( $c^*$ ) pour chacune des régressions
- On sélectionne le retard  $d$  qui minimise la probabilité critique du test

Voici un tableau récapitulatif des résultats :

<b>Estimation de la variable de changement de régime optimal</b>			
<b>Décalage de <math>q(t) = x(t-d)</math></b>	<b>Seuil C optimal</b>	<b>P-Critique AR vs SETAR</b>	<b>Variance résiduelle</b>
1	25,07	0,85	9,293
2	26,62	0,1	9,197
3	16,19	0	9,159
4	16,41	0,05	9,168
5	16,91	0,45	9,213

On constate ainsi que :

- La valeur du décalage temporel  $d^*$  optimal pour la variable de changement de régime,  $q_t = X_{t-d}$ , est de 3, soit  $q_t = X_{t-3}$ .
- Le seuil de changement de régime  $c^*$  optimal associé à ce niveau de  $d$ , est de 16,19.
- La variance résiduelle est quant à elle de 9,159.
- Enfin, notre modèle non-linéaire SETAR apporte plus d'information qu'un modèle AR linéaire, puisque sa probabilité critique est de  $0\% < 5\%$ .

## 2) Estimation des paramètres du modèle SETAR final

Voici un tableau récapitulatif présentant le modèle SETAR estimé final :

SETAR(2,2) $d^* = 3$ $c^* 16,19$				
	Coefficient estimé	Ecart-Type	T-stat	P-critique
$c1$	3,926	0,878	4,190	1,51E-05
$X(-1)$	0,632	0,007	7,685	1,77E-14
$X(-2)$	0,090	0,009	0,974	17%
$c2$	1,730	0,134	4,725	1,31E-06
$X(-1)$	0,796	0,001	24,060	0%
$X(-2)$	0,131	0,001	3,905	5,02E-05

On peut observer que seul 1 coefficient est non significatif.

Le modèle final est donc de la forme :

$$X_t = \begin{cases} 3,926 + \sum_{i=1}^p \phi_{1,i} X_{t-i} + \varepsilon_t & , si X_{t-3} \leq 16,19 \\ 1,73 + \sum_{i=1}^p \phi_{2,i} X_{t-i} + \varepsilon_t & , si X_{t-3} > 16,19 \end{cases}$$

Ce modèle est le modèle SETAR final optimisé pour la série du VIX hebdomadaire entre 1997 et 2017.

### Question 6 : Faites le test d'un modèle AR contre un modèle STAR

Afin de tenter d'apporter une linéarisation de notre modèle SETAR non linéaire, nous allons tester l'hypothèse nulle d'un modèle AR contre un modèle STAR (Smooth TAR), afin de déterminer si le passage d'un régime à l'autre ne devrait pas être moins brusque, et au contraire se faire de manière progressive, avec une fonction de transition.

Pour ce faire, nous aurons recours à deux tests :

- Le Test du Multiplicateur de Lagrange, qui teste l'hypothèse nulle d'un modèle AR contre un modèle STAR
- Le test de Fisher, qui teste la même chose que le test du Multiplicateur de Lagrange, mais qui est préférable pour de petites tailles d'échantillon.

On considère le modèle STAR suivant :

$$X_t = \phi'_1 z_t * (1 - G(x_{t-d}, \gamma, c)) + \phi'_2 z_t * G(x_{t-d}, \gamma, c) + \varepsilon_t$$

Avec :

- $z_t$  : la matrice des variables explicatives, ici donc (1,  $X(t-1)$ ,  $X(t-2)$ )
- $\phi_n$  : le vecteur des coefficients du régime  $n$
- $G(x_{t-1}, \gamma, c)$  : la fonction de transition du modèle STAR

Voici un tableau récapitulatif de nos résultats :

<b>Test d'un modèle STAR</b>	
<b>Test du Multiplicateur de Lagrange</b>	
<b>Variable de changement de régime = <math>X(t-1)</math></b>	
<b>LM Stat</b>	29,2059
<b>Valeur critique</b>	12,5916
<b>Probabilité critique</b>	0,00005
<b>F-stat</b>	4,9711
<b>Valeur critique</b>	2,1074
<b>Probabilité critique</b>	0,00004

On rejette ainsi, avec les deux tests, l'hypothèse nulle d'un modèle AR : un modèle STAR est plus pertinent et apporte plus d'informations qu'un modèle AR classique.

Notre modèle sur le VIX est donc bien un modèle à seuil, mais il peut être pertinent, au lieu d'utiliser des seuils fixe créant de brusques changements de régimes (modèle SETAR), d'utiliser un modèle à seuil avec fonction de transition (modèle STAR), permettant une linéarisation du changement de régime.

## **Conclusion**

Ainsi, nous sommes parvenus à modéliser, tout au long de ce projet, pour le VIX hebdomadaire entre 1997 et 2017, un modèle SETAR(2,2), à 2 régimes, où chaque régime suit un AR(2).

Lorsque la variable de changement de régime  $q_t = X_{t-1}$ , on trouve  $c^* = 26,07$ . Les coefficients sont, constantes mises à part, significatives, mais l'on ne peut pas dire, après test, que le modèle non linéaire SETAR soit plus pertinent que le modèle AR linéaire associé. En revanche, lorsque l'on optimise  $d$ , et que l'on prend  $q_t^* = X_{t-3}$ , comme variable de changement de régime, on trouve  $c^* = 16,19$ , et le modèle s'avère apporter plus d'informations que le modèle AR correspondant.

Par ailleurs, les coefficients de ce modèle (sauf un) sont significatifs, et l'on dispose d'un bon  $R^2$ .

Enfin, on constate après test qu'il pourrait être pertinent d'étudier, au lieu d'un modèle SETAR, à changement de régime brutaux en fonction du niveau de la variable expliquée, un modèle où le changement de régime serait linéaire et dépendrait d'une fonction de transition : un modèle STAR.



## Annexe 1.1 : Test ADF

### 1) Test de Dickey Fuller Augmenté

Modèles :

$$(3) \Delta X_t = c + bt + \rho X_{t-1} + \varepsilon_t$$

$$(2) \Delta X_t = c + \rho X_{t-1} + \varepsilon_t$$

$$(1) \Delta X_t = \rho X_{t-1} + \varepsilon_t$$

Hypothèses du test :

$$H_0: \rho = 0$$

*vs*

$$H_1: \rho < 0$$

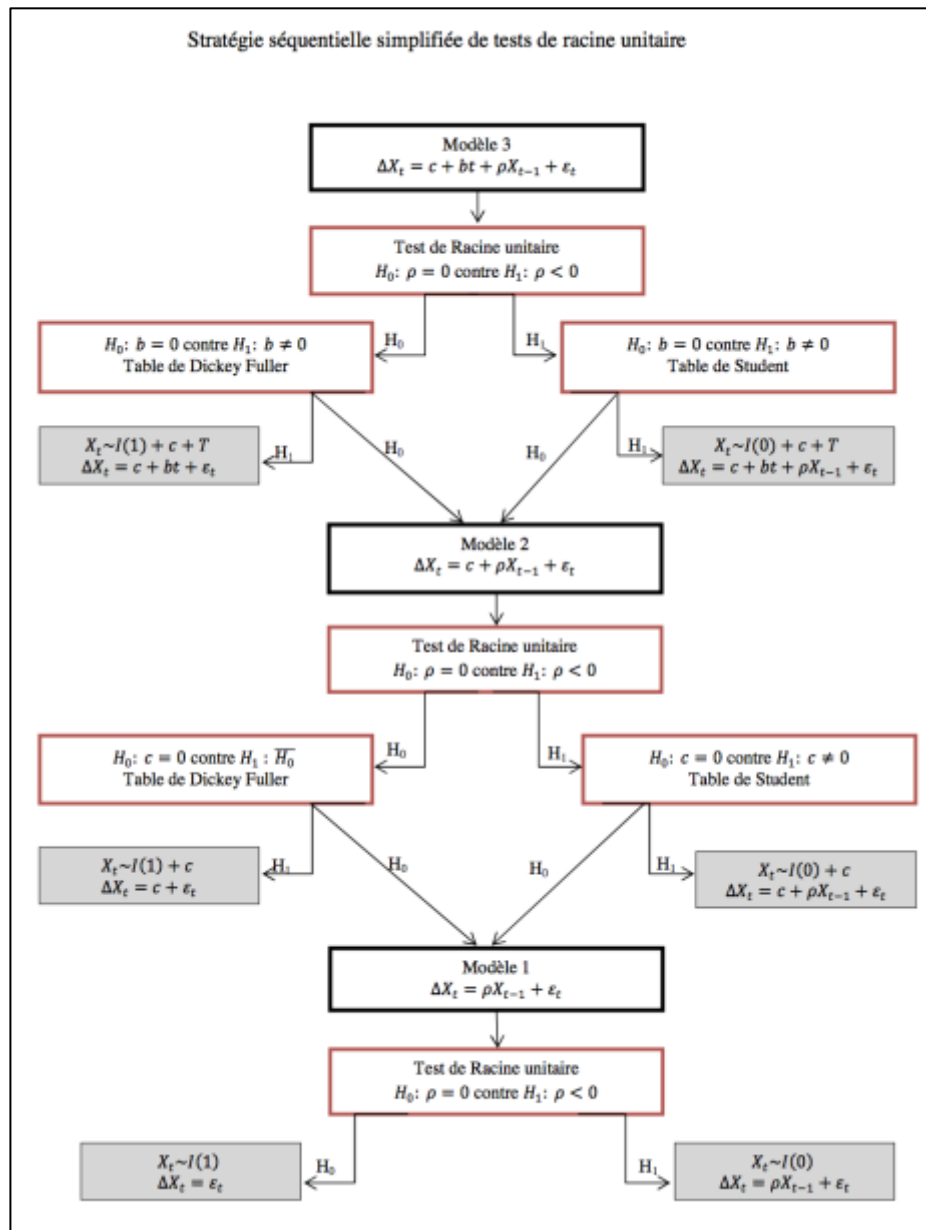
Statistique du test :

$$t_{\text{obs}} = \frac{\hat{\rho}}{\sigma_{\rho}}$$

si  $t_{\text{obs}} < VC = t_{\text{tab}}$  (tabulé par Dickey Fuller), alors on rejette l'hypothèse nulle de présence de racine unitaire.

On rejette  $H_0$  si  $t_{\text{obs}} < VC = t_{\text{tab}}$

On suit la stratégie séquentielle de tests de racine unitaire « classique », qui consiste à tester en premier lieu le Modèle 3, puis le 2, et enfin le 1 :



Voici le résumé de la stratégie séquentielle de test de racine unitaire que nous avons mené sur le cours mensuel du VIX :

**a) Modèle 3 :**

$$\Delta X_t = c + bt + \rho X_{t-1} + \varepsilon_t$$

Avec :

$\Delta X_t$  = Variation mensuelle du VIX

$X_{t-1}$  = Cours du VIX

$\rho$  = coefficient associé au cours du VIX

$\varepsilon_t$  = résidus (bruit blanc)

$t$  = tendance linéaire (grandit linéairement avec t)

$b$  = coefficient associée à la tendance linéaire

Voici la sortie EViews (Quick < Series Statistic < Unit Root Test) d'un test ADF :

- en niveau (On teste le processus de la variable  $X_t$ , quand bien même dans le modèle la variable à expliquer est  $\Delta X_t$ ), on sélectionne donc « Test for unit root in : » Level.
- Du modèle 3, avec tendance et constante. On sélectionne donc « Include in test equation : » Trend and intercept.

Null Hypothesis: SERIES01 has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 1 (Automatic - based on SIC, maxlag=21)				
			t-Statistic	Prob.*
<b>Augmented Dickey-Fuller test statistic</b>			<b>-5.298101</b>	<b>0.0001</b>
Test critical values:	1% level		-3.966888	
	5% level		-3.414136	
	10% level		-3.129172	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(SERIES01)				
Method: Least Squares				
Date: 10/20/17 Time: 13:14				
Sample (adjusted): 3 1043				
Included observations: 1041 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SERIES01(-1)	-0.062296	0.011758	-5.298101	0.0000
D(SERIES01(-1))	-0.156633	0.030886	-5.071334	0.0000
C	1.012555	0.263964	3.835958	0.0001
@TREND("1")	0.000526	0.000332	1.583462	0.1136
R-squared	0.060298	Mean dependent var		0.012988
Adjusted R-squared	0.057579	S.D. dependent var		3.140760
S.E. of regression	3.048999	Akaike info criterion		5.071339
Sum squared resid	9640.360	Schwarz criterion		5.090351
Log likelihood	-2635.632	Hannan-Quinn criter.		5.078551
F-statistic	22.18032	Durbin-Watson stat		1.981210
Prob(F-statistic)	0.000000			

On peut voir que l'hypothèse de racine unitaire est rejetée à 5% :  $-5,29 < -3,41$  et  $0,01\% < 5\%$ .

Nous n'avons donc pas de racine unitaire pour le coefficient associé à  $X_{t-1}$ .

Il nous faut maintenant tester la significativité du coefficient associé au trend,  $b$ .

L'hypothèse de non-significativité du coefficient associé au trend n'est en revanche pas rejetée :  $|1,58| < 1,96$  et  $11,36\% > 5\%$ .

(La valeur tabulée du test de significativité est cherchée dans la table de Student car le test de racine unitaire a été rejeté)

Le coefficient associé au trend est donc statistiquement proche de ou égale à 0, on ne peut donc pas utiliser le modèle 3, et, suivant la stratégie séquentielle de test que nous avons posée, nous passons au modèle 2.

#### b) Modèle 2 :

$$\Delta X_t = c + \rho X_{t-1} + \varepsilon_t$$

Voici la sortie EViews , d'un test ADF en niveau pour le modèle 2 (Intercept) :

Null Hypothesis: SERIES01 has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic - based on SIC, maxlag=21)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-5.053875	0.0000
Test critical values:	1% level		-3.436413	
	5% level		-2.864106	
	10% level		-2.568188	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(SERIES01)				
Method: Least Squares				
Date: 10/20/17 Time: 13:19				
Sample (adjusted): 3 1043				
Included observations: 1041 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SERIES01(-1)	-0.056307	0.011141	-5.053875	0.0000
D(SERIES01(-1))	-0.159550	0.030853	-5.171246	0.0000
C	1.165034	0.245952	4.736844	0.0000
R-squared	0.058026	Mean dependent var		0.012988
Adjusted R-squared	0.056211	S.D. dependent var		3.140760
S.E. of regression	3.051212	Akaike info criterion		5.071832
Sum squared resid	9663.670	Schwarz criterion		5.086092
Log likelihood	-2636.889	Hannan-Quinn criter.		5.077241
F-statistic	31.97038	Durbin-Watson stat		1.982382
Prob(F-statistic)	0.000000			

On peut voir que l'hypothèse de racine unitaire est rejetée à 5% :  $-5,05 < -2,87$  et  $0\% < 5\%$ .

De même, tous les coefficients sont significatifs, notamment celui de la constante :

$|4,73| > 1,96$  et  $0\% < 5\%$ .

Nous pouvons donc utiliser le modèle 2, et nous sommes ici en présence d'un processus  $X_t$  tel que :

$$X_t \sim I(0) + c$$

$$\text{et : } \Delta X_t = c + \rho X_{t-1} + \varepsilon_t$$

Nous testons quand même le modèle 1, il devrait en toute rigueur ne pas être stationnaire, afin d'être sûr que nous choisissons le bon modèle.

### c) Modèle 1 :

$$\Delta X_t = \rho X_{t-1} + \varepsilon_t$$

Voici la sortie EViews, d'un test ADF, en niveau, pour le modèle 1 (None) :

Null Hypothesis: SERIES01 has a unit root				
Exogenous: None				
Lag Length: 1 (Automatic - based on SIC, maxlag=21)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.753596	0.0755
Test critical values:				
	1% level		-2.567185	
	5% level		-1.941128	
	10% level		-1.616495	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(SERIES01)				
Method: Least Squares				
Date: 10/20/17 Time: 13:22				
Sample (adjusted): 3 1043				
Included observations: 1041 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SERIES01(-1)	-0.007590	0.004328	-1.753596	0.0798
D(SERIES01(-1))	-0.183717	0.030741	-5.976311	0.0000
R-squared	0.037664	Mean dependent var		0.012988
Adjusted R-squared	0.036737	S.D. dependent var		3.140760
S.E. of regression	3.082529	Akaike info criterion		5.091297
Sum squared resid	9872.562	Schwarz criterion		5.100803
Log likelihood	-2648.020	Hannan-Quinn criter.		5.094903
Durbin-Watson stat	1.988092			

Le processus apparait comme étant non-stationnaire dans le modèle 1 :

-1,75 > -1,95 et 7,55% > 5%.

Le modèle est donc supposé présenter une racine unitaire pour le coefficient associé à la variable explicative.

De même, la variable explicative n'apparait pas comme étant statistiquement significative ( $|-1,75| < 1,96$  et  $7,98\% > 5\%$ ).

Nous vérifions que le modèle choisi est bien stationnaire, et qu'il est le « bon » modèle à choisir, avec un test de Phillippe Perron.

## Annexe 1.2 : Test Philippe Perron

Modèle :

$$(2) \Delta X_t = c + \rho X_{t-1} + \varepsilon_t$$

Hypothèses du test :

$$H_0: \rho = 0$$

vs

$$H_1: \rho < 0$$

Statistique du test :

$$t_{obs}^{PP} = f(t_{obs}^{ADF}) \rightarrow \text{c'est une fonction de la statistique du test ADF.}$$

Règle de décision :

Si  $t_{obs}^{PP} < t_{tab}^{DF}$ , alors on rejette  $H_0$ . Il est important de noter que l'on se sert des valeurs tabulées de Dickey Fuller.

### a) Modèle 3 :

Voici la sortie Eviews du test de Phillippe Perron, en niveau, sur le modèle 3 (Trend & Intercept) :

Null Hypothesis: SERIES01 has a unit root  
Exogenous: Constant, Linear Trend  
Bandwidth: 8 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-5.949600	0.0000
Test critical values:		
1% level	-3.966879	
5% level	-3.414131	
10% level	-3.129170	

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

Residual variance (no correction)	9.481423
HAC corrected variance (Bartlett kernel)	8.328385

Phillips-Perron Test Equation  
Dependent Variable: D(SERIES01)  
Method: Least Squares  
Date: 10/20/17 Time: 13:43  
Sample (adjusted): 2 1043  
Included observations: 1042 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SERIES01(-1)	-0.073679	0.011668	-6.314347	0.0000
C	1.186876	0.264144	4.493297	0.0000
@TREND("1")	0.000628	0.000335	1.876122	0.0609

R-squared	0.036973	Mean dependent var	0.013013
Adjusted R-squared	0.035120	S.D. dependent var	3.139252
S.E. of regression	3.083634	Akaike info criterion	5.092970
Sum squared resid	9879.643	Schwarz criterion	5.107218
Log likelihood	-2650.437	Hannan-Quinn criter.	5.098374
F-statistic	19.94516	Durbin-Watson stat	2.273652
Prob(F-statistic)	0.000000		

Le test confirme bien le rejet de l'hypothèse nulle de présence de racine unitaire ( $-5,95 < -2,87$  et  $0\% < 5\%$ ).

De même, il confirme également la non-significativité du coefficient associé à la tendance.

## b) Modèle 2 :

Voici la sortie Eviews du test de Phillippe Perron, en niveau, sur notre modèle 2 (Intercept) :

Null Hypothesis: SERIES01 has a unit root  
Exogenous: Constant  
Bandwidth: 8 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-5.614644	0.0000
Test critical values:		
1% level	-3.436407	
5% level	-2.864103	
10% level	-2.568186	

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

Residual variance (no correction)	9.513544
HAC corrected variance (Bartlett kernel)	8.148329

Phillips-Perron Test Equation  
Dependent Variable: D(SERIES01)  
Method: Least Squares  
Date: 10/20/17 Time: 13:26  
Sample (adjusted): 2 1043  
Included observations: 1042 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SERIES01(-1)	-0.066717	0.011076	-6.023502	0.0000
C	1.372728	0.245161	5.599294	0.0000
R-squared	0.033711	Mean dependent var	0.013013	
Adjusted R-squared	0.032782	S.D. dependent var	3.139252	
S.E. of regression	3.087368	Akaike info criterion	5.094432	
Sum squared resid	9913.112	Schwarz criterion	5.103931	
Log likelihood	-2652.199	Hannan-Quinn criter.	5.098035	
F-statistic	36.28258	Durbin-Watson stat	2.281857	
Prob(F-statistic)	0.000000			

Le test confirme bien le rejet de l'hypothèse nulle de présence de racine unitaire ( $-5,61 < -2,87$  et  $0\% < 5\%$ ).

De même, il confirme également la significativité des coefficients de la régression, notamment celui de la constante.

### c) Modèle 1 :

Voici la sortie Eviews du test de Phillippe Perron, en niveau, sur notre modèle 2 (Intercept) :



Null Hypothesis: SERIES01 has a unit root				
Exogenous: None				
Bandwidth: 8 (Newey-West automatic) using Bartlett kernel				
		Adj. t-Stat	Prob.*	
<hr/>				
Phillips-Perron test statistic		-1.739873	0.0777	
Test critical values:				
	1% level	-2.567183		
	5% level	-1.941127		
	10% level	-1.616495		
<hr/>				
*MacKinnon (1996) one-sided p-values.				
<hr/>				
Residual variance (no correction)		9.800341		
HAC corrected variance (Bartlett kernel)		6.441714		
<hr/>				
Phillips-Perron Test Equation				
Dependent Variable: D(SERIES01)				
Method: Least Squares				
Date: 10/20/17 Time: 13:44				
Sample (adjusted): 2 1043				
Included observations: 1042 after adjustments				
<hr/>				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
<hr/>				
SERIES01(-1)	-0.009613	0.004384	-2.192896	0.0285
<hr/>				
R-squared	0.004581	Mean dependent var	0.013013	
Adjusted R-squared	0.004581	S.D. dependent var	3.139252	
S.E. of regression	3.132053	Akaike info criterion	5.122214	
Sum squared resid	10211.96	Schwarz criterion	5.126963	
Log likelihood	-2667.673	Hannan-Quinn criter.	5.124015	
Durbin-Watson stat	2.345175			
<hr/>				

Le test confirme bien le non-rejet de l'hypothèse nulle de présence de racine unitaire (-1,73 < -1,94 et 7,77% > 5%).

De même, il confirme également la non-significativité du coefficient associé à la variable explicative.

## Annexe 2.1 : 16 Sorties EViews des estimations de $p^*$ et $q^*$ du modèle ARMA( $p^*$ , $q^*$ )

### a) ARMA(3,p) avec $p = \{0, 1, 2, 3\}$

Dependent Variable: SERIES01  
Method: ARMA Maximum Likelihood (OPG - BHHH)  
Date: 10/20/17 Time: 15:09  
Sample: 4 1043  
Included observations: 1040  
Convergence achieved after 44 iterations  
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.116189	8.693985	0.128386	0.8979
SERIES01(-1)	0.508392	4.711689	0.107900	0.9141
SERIES01(-2)	0.737718	0.588867	1.252775	0.2106
SERIES01(-3)	-0.299984	3.757070	-0.079845	0.9364
MA(1)	0.279531	4.712650	0.059315	0.9527
MA(2)	-0.352923	3.177867	-0.111057	0.9116
MA(3)	0.033564	0.475003	0.070662	0.9437
SIGMASQ	9.269349	0.203417	45.56823	0.0000
R-squared	0.875376	Mean dependent var	20.41411	
Adjusted R-squared	0.874531	S.D. dependent var	8.628443	
S.E. of regression	3.056338	Akaike info criterion	5.080475	
Sum squared resid	9640.123	Schwarz criterion	5.118529	
Log likelihood	-2633.847	Hannan-Quinn criter.	5.094911	
F-statistic	1035.557	Durbin-Watson stat	1.986961	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.40	.11	-.78	

Dependent Variable: SERIES01  
Method: ARMA Maximum Likelihood (OPG - BHHH)  
Date: 10/20/17 Time: 15:10  
Sample: 4 1043  
Included observations: 1040  
Convergence achieved after 30 iterations  
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.122960	0.488740	4.343742	0.0000
SERIES01(-1)	-0.005330	0.226438	-0.023540	0.9812
SERIES01(-2)	0.803411	0.166422	4.827544	0.0000
SERIES01(-3)	0.099299	0.052190	1.902638	0.0574
MA(1)	0.795377	0.223609	3.557005	0.0004
SIGMASQ	9.271526	0.201028	46.12049	0.0000
R-squared	0.875347	Mean dependent var	20.41411	
Adjusted R-squared	0.874744	S.D. dependent var	8.628443	
S.E. of regression	3.053740	Akaike info criterion	5.077326	
Sum squared resid	9642.387	Schwarz criterion	5.105867	
Log likelihood	-2634.210	Hannan-Quinn criter.	5.088153	
F-statistic	1452.200	Durbin-Watson stat	1.990603	
Prob(F-statistic)	0.000000			
Inverted MA Roots	-.80			

Dependent Variable: SERIES01  
Method: ARMA Maximum Likelihood (OPG - BHHH)  
Date: 10/20/17 Time: 15:10  
Sample: 4 1043  
Included observations: 1040  
Convergence achieved after 40 iterations  
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.578380	0.727206	2.170472	0.0302
SERIES01(-1)	0.249641	0.427439	0.584040	0.5593
SERIES01(-2)	0.758266	0.235339	3.222013	0.0013
SERIES01(-3)	-0.084103	0.188588	-0.445962	0.6557
MA(1)	0.538869	0.427860	1.259452	0.2082
MA(2)	-0.172888	0.148016	-1.168031	0.2431
SIGMASQ	9.270008	0.201000	46.11955	0.0000
R-squared	0.875367	Mean dependent var	20.41411	
Adjusted R-squared	0.874643	S.D. dependent var	8.628443	
S.E. of regression	3.054967	Akaike info criterion	5.078712	
Sum squared resid	9640.808	Schwarz criterion	5.112009	
Log likelihood	-2633.930	Hannan-Quinn criter.	5.091343	
F-statistic	1209.223	Durbin-Watson stat	1.988214	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.23	-.76		

Dependent Variable: SERIES01  
Method: Least Squares  
Date: 10/20/17 Time: 15:11  
Sample (adjusted): 4 1043  
Included observations: 1040 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.183671	0.248868	4.756229	0.0000
SERIES01(-1)	0.786384	0.031273	25.14612	0.0000
SERIES01(-2)	0.170588	0.039412	4.328322	0.0000
SERIES01(-3)	-0.014179	0.031300	-0.453005	0.6506
R-squared	0.875104	Mean dependent var	20.41411	
Adjusted R-squared	0.874742	S.D. dependent var	8.628443	
S.E. of regression	3.053763	Akaike info criterion	5.074465	
Sum squared resid	9661.184	Schwarz criterion	5.093492	
Log likelihood	-2634.722	Hannan-Quinn criter.	5.081683	
F-statistic	2419.625	Durbin-Watson stat	1.986037	
Prob(F-statistic)	0.000000			

### b) ARMA(2,p) avec $p = \{0, 1, 2, 3\}$

Dependent Variable: SERIES01  
Method: ARMA Maximum Likelihood (OPG - BHHH)  
Date: 10/20/17 Time: 15:11  
Sample: 3 1043  
Included observations: 1041  
Convergence achieved after 54 iterations  
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.781674	0.406711	4.380686	0.0000
SERIES01(-1)	0.140213	0.220567	0.635693	0.5251
SERIES01(-2)	0.773663	0.209793	3.687742	0.0002
MA(1)	0.647392	0.223562	2.895803	0.0039
MA(2)	-0.099573	0.045943	-2.167308	0.0304
MA(3)	-0.005665	0.024004	-0.236000	0.8135
SIGMASQ	9.262393	0.202063	45.83905	0.0000
R-squared	0.875541	Mean dependent var	20.40363	
Adjusted R-squared	0.874819	S.D. dependent var	8.630913	
S.E. of regression	3.053702	Akaike info criterion	5.077995	
Sum squared resid	9642.151	Schwarz criterion	5.111266	
Log likelihood	-2636.096	Hannan-Quinn criter.	5.090616	
F-statistic	1212.324	Durbin-Watson stat	1.986432	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.16	-.04	-.77	

Dependent Variable: SERIES01  
Method: ARMA Maximum Likelihood (OPG - BHHH)  
Date: 10/20/17 Time: 15:12  
Sample: 3 1043  
Included observations: 1041  
Convergence achieved after 33 iterations  
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.809461	0.366328	4.939456	0.0000
SERIES01(-1)	0.131544	0.155572	0.845549	0.3980
SERIES01(-2)	0.780968	0.149283	5.231444	0.0000
MA(1)	0.655625	0.160140	4.094079	0.0000
MA(2)	-0.096910	0.043454	-2.230187	0.0259
SIGMASQ	9.262351	0.195380	47.40679	0.0000
R-squared	0.875541	Mean dependent var	20.40363	
Adjusted R-squared	0.874940	S.D. dependent var	8.630913	
S.E. of regression	3.052220	Akaike info criterion	5.076100	
Sum squared resid	9642.108	Schwarz criterion	5.104618	
Log likelihood	-2636.110	Hannan-Quinn criter.	5.086918	
F-statistic	1456.203	Durbin-Watson stat	1.985443	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.12	-.78		

Dependent Variable: SERIES01				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/20/17 Time: 15:13				
Sample: 3 1043				
Included observations: 1041				
Convergence achieved after 48 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.330497	0.259298	5.131158	0.0000
SERIES01(-1)	0.864052	0.128085	5.184450	0.0000
SERIES01(-2)	0.271590	0.123045	2.207235	0.0275
MA(1)	0.123594	0.138853	0.890105	0.3736
SIGMASQ	9.280406	0.184855	50.20381	0.0000
R-squared	0.875299	Mean dependent var	20.40363	
Adjusted R-squared	0.874817	S.D. dependent var	8.630913	
S.E. of regression	3.053718	Akaike info criterion	5.075403	
Sum squared resid	9660.903	Schwarz criterion	5.098169	
Log likelihood	-2636.747	Hannan-Quinn criter.	5.084418	
F-statistic	1817.964	Durbin-Watson stat	1.988050	
Prob(F-statistic)	0.000000			
Inverted MA Roots	-.12			

Dependent Variable: SERIES01				
Method: Least Squares				
Date: 10/20/17 Time: 15:13				
Sample (adjusted): 3 1043				
Included observations: 1041 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.165034	0.245952	4.736844	0.0000
SERIES01(-1)	0.784143	0.030833	25.43164	0.0000
SERIES01(-2)	0.159550	0.030853	5.171246	0.0000
R-squared	0.875263	Mean dependent var	20.40363	
Adjusted R-squared	0.875023	S.D. dependent var	8.630913	
S.E. of regression	3.051212	Akaike info criterion	5.071832	
Sum squared resid	9663.670	Schwarz criterion	5.086092	
Log likelihood	-2636.889	Hannan-Quinn criter.	5.077241	
F-statistic	3641.755	Durbin-Watson stat	1.982382	
Prob(F-statistic)	0.000000			

### c) ARMA(1,p) avec p = {0, 1, 2, 3}

Dependent Variable: SERIES01				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/20/17 Time: 15:14				
Sample: 2 1043				
Included observations: 1042				
Convergence achieved after 34 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.955515	0.208280	4.587640	0.0000
SERIES01(-1)	0.953839	0.006859	139.0581	0.0000
MA(1)	-0.166073	0.020102	-8.261586	0.0000
MA(2)	0.034050	0.019641	1.733651	0.0833
MA(3)	-0.032560	0.021031	-1.548201	0.1219
SIGMASQ	9.268024	0.192987	48.02412	0.0000
R-squared	0.875531	Mean dependent var	20.39331	
Adjusted R-squared	0.874930	S.D. dependent var	8.633197	
S.E. of regression	3.053146	Akaike info criterion	5.075993	
Sum squared resid	9657.281	Schwarz criterion	5.104490	
Log likelihood	-2638.592	Hannan-Quinn criter.	5.086802	
F-statistic	1457.473	Durbin-Watson stat	1.988412	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.34	-.09+.29i	-.09-.29i	

Dependent Variable: SERIES01				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/20/17 Time: 15:14				
Sample: 2 1043				
Included observations: 1042				
Convergence achieved after 37 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.022219	0.188078	5.435082	0.0000
SERIES01(-1)	0.950554	0.005329	178.3797	0.0000
MA(1)	-0.163125	0.019764	-8.253500	0.0000
MA(2)	0.031697	0.019853	1.596583	0.1107
SIGMASQ	9.276652	0.178064	52.09733	0.0000
R-squared	0.875415	Mean dependent var	20.39331	
Adjusted R-squared	0.874935	S.D. dependent var	8.633197	
S.E. of regression	3.053094	Akaike info criterion	5.075001	
Sum squared resid	9666.272	Schwarz criterion	5.098748	
Log likelihood	-2639.076	Hannan-Quinn criter.	5.084008	
F-statistic	1821.662	Durbin-Watson stat	1.988800	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.08+.16i	.08-.16i		

Dependent Variable: SERIES01				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/20/17 Time: 15:15				
Sample: 2 1043				
Included observations: 1042				
Convergence achieved after 29 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.956964	0.181325	5.277612	0.0000
SERIES01(-1)	0.953774	0.005127	186.0463	0.0000
MA(1)	-0.163280	0.019542	-8.355259	0.0000
SIGMASQ	9.285379	0.178005	52.16345	0.0000
R-squared	0.875298	Mean dependent var	20.39331	
Adjusted R-squared	0.874938	S.D. dependent var	8.633197	
S.E. of regression	3.053058	Akaike info criterion	5.074022	
Sum squared resid	9675.364	Schwarz criterion	5.093019	
Log likelihood	-2639.565	Hannan-Quinn criter.	5.081228	
F-statistic	2428.615	Durbin-Watson stat	1.996966	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.16			

Dependent Variable: SERIES01				
Method: Least Squares				
Date: 10/20/17 Time: 15:15				
Sample (adjusted): 2 1043				
Included observations: 1042 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.372728	0.245161	5.599294	0.0000
SERIES01(-1)	0.933283	0.011076	84.26071	0.0000
R-squared	0.872234	Mean dependent var	20.39331	
Adjusted R-squared	0.872111	S.D. dependent var	8.633197	
S.E. of regression	3.087368	Akaike info criterion	5.094432	
Sum squared resid	9913.112	Schwarz criterion	5.103931	
Log likelihood	-2652.199	Hannan-Quinn criter.	5.098035	
F-statistic	7099.867	Durbin-Watson stat	2.281857	
Prob(F-statistic)	0.000000			

d) MA(p) avec p = {0, 1, 2, 3}

Dependent Variable: SERIES01				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/20/17 Time: 15:16				
Sample: 1 1043				
Included observations: 1043				
Convergence achieved after 83 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.35749	0.485944	41.89269	0.0000
MA(1)	1.004654	0.022500	44.65070	0.0000
MA(2)	0.851731	0.030885	27.57792	0.0000
MA(3)	0.417660	0.021608	19.32897	0.0000
SIGMASQ	15.49102	0.338647	45.74388	0.0000
R-squared	0.792068	Mean dependent var	20.38297	
Adjusted R-squared	0.791267	S.D. dependent var	8.635510	
S.E. of regression	3.945331	Akaike info criterion	5.589187	
Sum squared resid	16157.13	Schwarz criterion	5.612916	
Log likelihood	-2909.761	Hannan-Quinn criter.	5.598187	
F-statistic	988.5065	Durbin-Watson stat	1.668325	
Prob(F-statistic)	0.000000			
Inverted MA Roots	-17+.77i	-17-.77i	-67	

Dependent Variable: SERIES01				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/20/17 Time: 15:16				
Sample: 1 1043				
Included observations: 1043				
Convergence achieved after 103 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.37098	0.427570	47.64363	0.0000
MA(1)	0.927254	0.014956	61.99708	0.0000
MA(2)	0.568906	0.016062	35.41889	0.0000
SIGMASQ	20.02506	0.490504	40.82548	0.0000
R-squared	0.731209	Mean dependent var	20.38297	
Adjusted R-squared	0.730433	S.D. dependent var	8.635510	
S.E. of regression	4.483543	Akaike info criterion	5.843694	
Sum squared resid	20886.14	Schwarz criterion	5.862677	
Log likelihood	-3043.486	Hannan-Quinn criter.	5.850894	
F-statistic	942.1532	Durbin-Watson stat	1.346833	
Prob(F-statistic)	0.000000			
Inverted MA Roots	-46-.59i	-46+.59i		

Dependent Variable: SERIES01				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/20/17 Time: 15:17				
Sample: 1 1043				
Included observations: 1043				
Convergence achieved after 5 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.37486	0.388432	52.45408	0.0000
MA(1)	0.732425	0.015687	46.69014	0.0000
SIGMASQ	31.93372	0.828287	38.55393	0.0000
R-squared	0.571363	Mean dependent var	20.38297	
Adjusted R-squared	0.570538	S.D. dependent var	8.635510	
S.E. of regression	5.659137	Akaike info criterion	6.308029	
Sum squared resid	33306.87	Schwarz criterion	6.322267	
Log likelihood	-3286.637	Hannan-Quinn criter.	6.313429	
F-statistic	693.1465	Durbin-Watson stat	0.909838	
Prob(F-statistic)	0.000000			
Inverted MA Roots	-.73			

Dependent Variable: SERIES01				
Method: Least Squares				
Date: 10/20/17 Time: 15:18				
Sample: 1 1043				
Included observations: 1043				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.38297	0.267390	76.22925	0.0000
R-squared	0.000000	Mean dependent var	20.38297	
Adjusted R-squared	0.000000	S.D. dependent var	8.635510	
S.E. of regression	8.635510	Akaike info criterion	7.150601	
Sum squared resid	77704.07	Schwarz criterion	7.155347	
Log likelihood	-3728.038	Hannan-Quinn criter.	7.152401	
Durbin-Watson stat	0.132028			

## Annexe 2.2 : Sortie EViews de l'estimation du modèle $AR(p^*) = AR(2)$

Dependent Variable: SERIES01				
Method: Least Squares				
Date: 10/20/17 Time: 16:15				
Sample (adjusted): 3 1043				
Included observations: 1041 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.165034	0.245952	4.736844	0.0000
SERIES01(-1)	0.784143	0.030833	25.43164	0.0000
SERIES01(-2)	0.159550	0.030853	5.171246	0.0000
R-squared	0.875263	Mean dependent var	20.40363	
Adjusted R-squared	0.875023	S.D. dependent var	8.630913	
S.E. of regression	3.051212	Akaike info criterion	5.071832	
Sum squared resid	9663.670	Schwarz criterion	5.086092	
Log likelihood	-2636.889	Hannan-Quinn criter.	5.077241	
F-statistic	3641.755	Durbin-Watson stat	1.982382	
Prob(F-statistic)	0.000000			

### Annexe 2.3 : Test de Ljung Box – Autocorrélation des erreurs

#### Mise en place du test

Hypothèse du test :

$H_0$ : Les résidus ne sont pas autocorréllés

vs

$H_1$ : les résidus sont autocorréllés

Statistique du test :

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$$

avec  $n$  le nombre d'observations,  $\hat{\rho}_k^2$  le coefficient d'autocorrélation à l'ordre  $k$  élevé au carré, et  $h$  le degré de décalage temporel pour la Q-stat.

On va enfaite calculer cette Q-stat pour les 20 premiers décalages temporels de la série, puis sommer ces 20 Q-stats, pour obtenir la statistique du test.

Sous  $H_0$ , la Q-stat suit une loi de Khi2 à  $h$  degrés de libertés.

$$Q_h \sim \chi^2_{1-\alpha, h}$$

Pour la série considérée,  $\chi^2_{1-\alpha, h} = 31,41$

On rejette  $H_0$  si  $Q_h > \chi^2_{1-\alpha, h}$ .

Commande EViews :

“View > Residuals Diagnostic > Correlogram – Q Statistics “

Date: 10/20/17 Time: 16:36  
Sample: 1 1043  
Included observations: 1041  
Q-statistic probabilities adjusted for 2 dynamic regressors

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.002	0.002	0.0049	0.944
		2	0.008	0.008	0.0767	0.962
		3	-0.027	-0.027	0.8273	0.843
		4	0.014	0.015	1.0454	0.903
		5	0.003	0.004	1.0582	0.958
		6	0.011	0.010	1.1915	0.977
		7	-0.054	-0.053	4.1995	0.757
		8	0.053	0.054	7.2048	0.515
		9	-0.044	-0.043	9.1988	0.419
		10	-0.074	-0.078	14.916	0.135
		11	-0.059	-0.054	18.584	0.069
		12	-0.029	-0.032	19.491	0.077
		13	0.027	0.026	20.273	0.089
		14	0.054	0.050	23.317	0.055
		15	0.048	0.056	25.784	0.040
		16	-0.021	-0.025	26.258	0.050
		17	-0.002	-0.003	26.261	0.070
		18	-0.008	-0.007	26.335	0.092
		19	0.035	0.028	27.655	0.090
		20	0.055	0.051	30.891	0.057
		21	-0.006	-0.016	30.926	0.075
		22	0.013	0.009	31.101	0.094
		23	0.041	0.040	32.871	0.083
		24	-0.013	0.001	33.060	0.103
		25	-0.039	-0.028	34.687	0.094
		26	-0.017	-0.008	34.997	0.112
		27	0.018	0.016	35.338	0.131
		28	0.001	-0.013	35.340	0.160
		29	-0.037	-0.032	36.835	0.150
		30	-0.018	-0.004	37.196	0.172
		31	0.018	0.024	37.541	0.194
		32	-0.007	-0.006	37.596	0.228
		33	0.015	0.018	37.829	0.258
		34	-0.005	-0.007	37.855	0.298
		35	0.008	-0.002	37.931	0.337
		36	0.041	0.036	39.730	0.307

Lorsque la probabilité critique d'un coefficient d'autocorrélation est supérieur à  $\alpha$  (risque de première espèce), on ne rejette pas l'hypothèse nulle de non significativité du coefficient. On peut observer que les coefficients d'autocorrélation entre les erreurs du modèle sont presque toutes non-significatives, seul le coefficient autocorrélation des résidus à l'ordre 15 s'avère significatif.

On peut donc supposer une absence d'autocorrélations entre nos résidus dans le modèle.

La Q-stat est la somme cumulée des Q-stat.

Nous avons décidé de retenir l'ordre 20 pour la statistique du test.

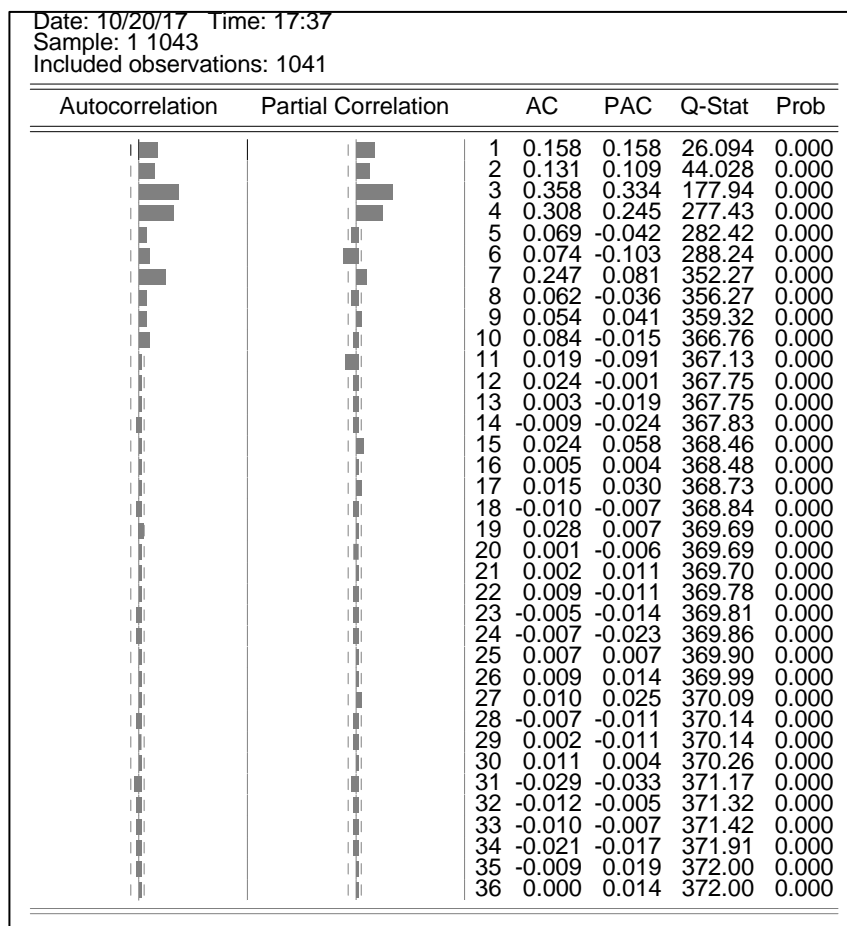
On peut voir que la somme cumulée des Q-stat jusqu'à l'ordre 20 de décalage temporel donne un Q-stat de 30,891.

$30,891 < 31,41$ , on ne peut donc pas rejeter l'hypothèse nulle de non-autocorrélation entre les résidus.

## Annexe 2.4 : Test d'Hétéroscédasticité des erreurs du modèle

Commande EViews :

“View > Residuals Diagnostic > Correlogram Squared Residual”

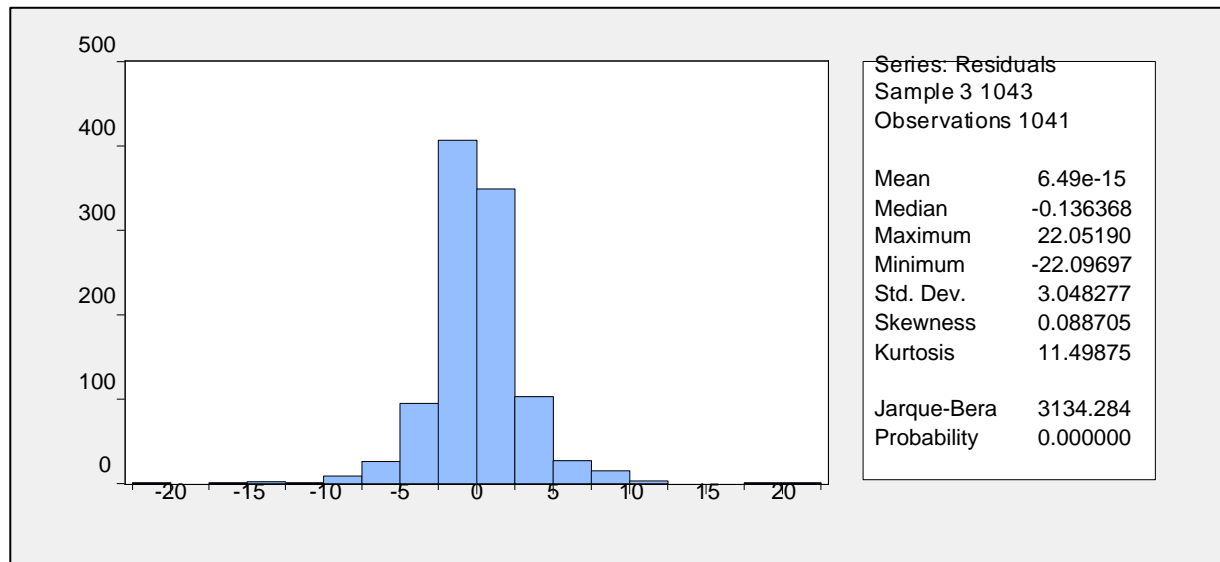




## Annexe 2.5 : Test de Jarque-Bera – Normalité des erreurs

Commande EViews :

“View > Residuals Diagnostic > Histogram – Normality Test”



### Annexe 3.1 : Mise en place de la régression pour l'estimation du modèle SETAR

On commence par définir notre seuil de changement de régime  $c$  comme la moyenne de l'échantillon.

On trouve  $\bar{X} = c = XXX$ .

On a défini  $q$  comme étant égal à  $X_{t-1}$ .

Ainsi, lorsque  $X_{t-1}$  est inférieur ou égale à  $c$ , nous nous trouvons dans le régime 1.

Lorsque  $X_{t-1}$  est en revanche supérieur à  $c$ , nous sommes dans le régime 2.

Afin de pouvoir modéliser cela comme un modèle linéaire multivarié classique, et donc de pouvoir effectuer une régression linéaire simple afin d'estimer les paramètres du modèle, on introduit une variable indicatrice  $1(X_{t-1} \leq c)$  :

$$1(X_{t-1} \leq c) = \begin{cases} 1, & \text{si } q = X_{t-1} \leq c \\ 0, & \text{sinon} \end{cases}$$

Nous remarquons dès lors que le modèle peut être écrit sous la forme similaire suivante :

$$X_t = \left[ c_1 + \sum_{i=1}^p \phi_{1,i} X_{t-i} + \varepsilon_t \right] \times 1(X_{t-1} \leq c) + \left[ c_2 + \sum_{i=1}^p \phi_{2,i} X_{t-i} + \varepsilon_t \right] \times (1 - 1(X_{t-1} \leq c))$$

On a  $Y$ , le vecteur de la variable expliquée.

Si  $Z$  est le vecteur  $(1, X_{t-1}, X_{t-2}, \dots, X_{t-p})$

Alors on peut définir la matrice  $X$  comme :

$$X = [ Z * 1(X_{t-1} \leq c) \quad Z * (1 - 1(X_{t-1} \leq c)) ]$$

En notant :  $1(X_{t-1} \leq c) = I$

La matrice des régresseurs  $X$  est donc de la forme suivante :

$$\begin{pmatrix} 1 \times I & X_{t-1} \times I & X_{t-2} \times I & 1 \times (1-I) & X_{t-1} \times (1-I) & X_{t-2} \times (1-I) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 \times I & X_{n-1} \times I & X_{n-2} \times I & 1 \times (1-I) & X_{n-1} \times (1-I) & X_{n-2} \times (1-I) \end{pmatrix}$$

Régime 1 :

- Une colonne pour la constante du régime 1 : vaut 1 si  $q \leq c$ , 0 sinon.
- Une colonne pour la variable explicative  $X_{t-1}$  du régime 1 : c'est  $Y$  décalé de 1 périodes. Vaut  $X_{t-1}$ , si  $q \leq c$ , 0 sinon.

- Une colonne pour la variable explicative  $X_{t-2}$  du régime 1 : c'est Y décalé de 2 périodes. Vaut  $X_{t-2}$ , si  $q \leq c$ , 0 sinon.
- ...
- Une colonne pour la variable explicative  $X_{t-p}$  du régime 1 : c'est Y décalé de p périodes. Vaut  $X_{t-p}$ , si  $q \leq c$ , 0 sinon.

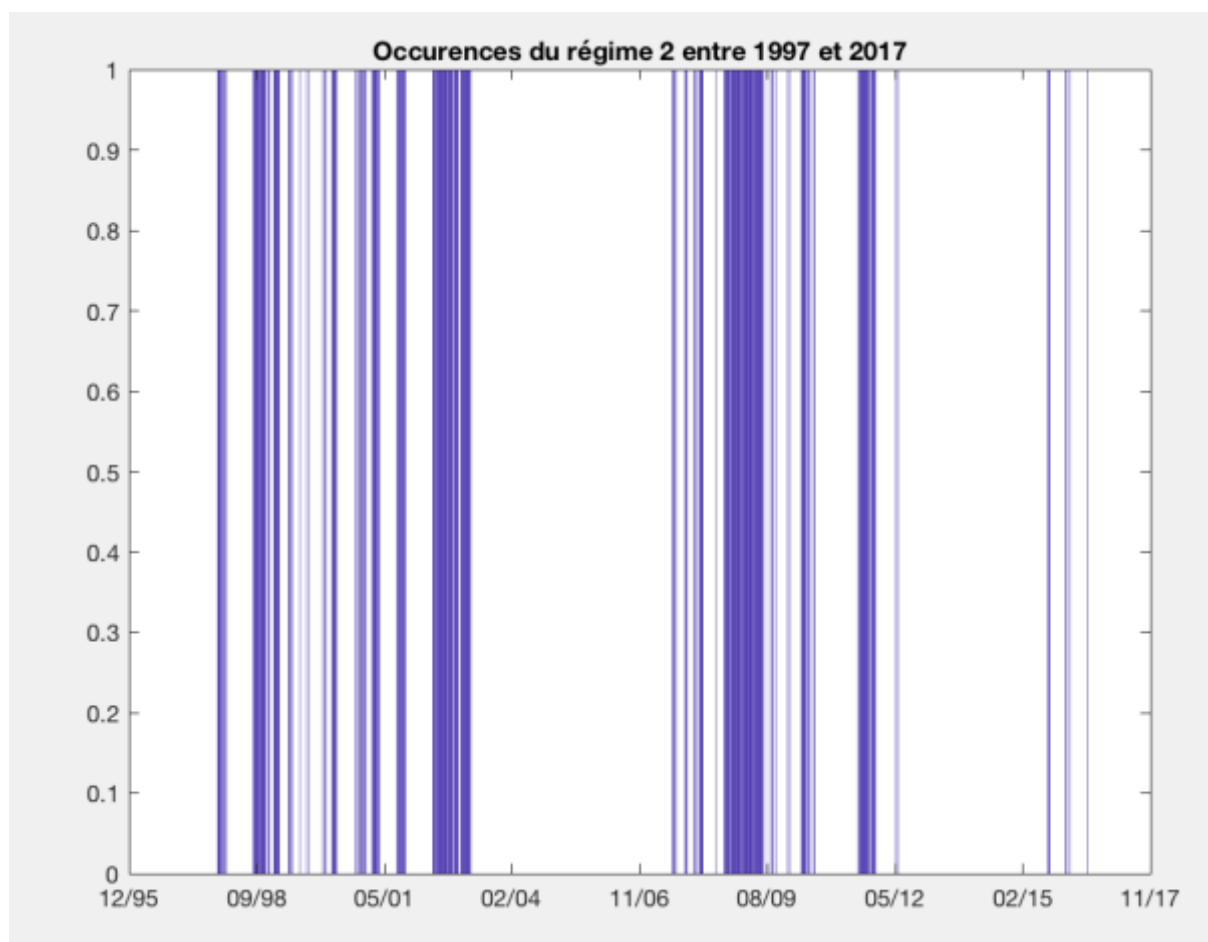
Régime 2 :

- Une colonne pour la constante du modèle 2 : vaut 1 si  $q > c$ , 0 sinon.
- Une colonne pour la variable explicative  $X_{t-1}$  du régime 2 : c'est Y décalé de 1 périodes. Vaut  $X_{t-1}$ , si  $q > c$ , 0 sinon.
- Une colonne pour la variable explicative  $X_{t-2}$  du régime 2 : c'est Y décalé de 2 périodes. Vaut  $X_{t-2}$ , si  $q > c$ , 0 sinon.
- ...
- Une colonne pour la variable explicative  $X_{t-p}$  du régime 2 : c'est Y décalé de p périodes. Vaut  $X_{t-p}$ , si  $q > c$ , 0 sinon.

La matrice X contient donc  $2*(k+1)$  colonnes, avec k le nombre de paramètres à estimer. Avec cette disposition, on peut facilement effectuer une régression linéaire classique de X sur Y.

On obtient ainsi les  $2*(k+1)$  coefficients estimés du modèle par la formule :

$$\hat{\theta} = (X'X)^{-1}(X'Y)$$



Sur l'axe des ordonnées, Y vaut 1 lorsque nous nous trouvons dans le régime 2, et 0 sinon. On peut voir que les périodes où intervient le régime deux sont principalement entre 1998 et 2003, puis entre 2007 et 2010, avant de reprendre entre 2011 et 2012. Plus précisément, les deux pics d'occurrence du régime 2 se situent entre 2002 et 2003, et entre 2008 et 2009.