

Correction Du TP 1 : Méthode Box-Jenkins

Réponses aux questions de cours :

- 1- La différence entre un processus DS et TS :
 - Le Processus déterministe TS est caractérisé par une Moyenne instable et une variance stable durant toute la période d'étude d'une série.
 - Un processus stochastique (aléatoire) DS est caractérisé par une moyenne stable et une variance instable.
- 2- Comparativement au test Dickey Fuller simple, le test de Dickey Fuller Augmenté prend en considération le fait que l'erreur ne peut pas toujours vérifier l'hypothèse d'absence d'autocorrélation, donc permet à l'erreur d'être autocorréllée.
- 3- Un corrélogramme est caractérisé par la représentation graphique de la fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF).
- 4- La méthode utilisée pour l'estimation d'un processus ARMA est la méthode des moindres carrés ordinaires (MCO).

Exercice Appliqué sur ordinateur :

- 1- Représentation graphique et commentaire :

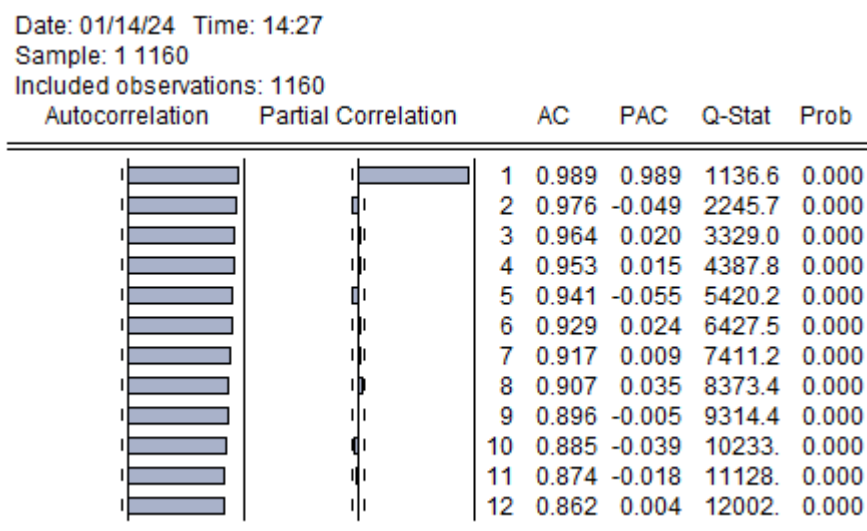


La représentation graphique de la série du CAC40 nous renseigne sur l'évolution à travers le temps de cette série, nous remarquons une forte volatilité de la série avec des perturbations ou des chocs à des périodes bien précises. Particulièrement avec la période de l'intervalle à

environ [200 ; 300] des abscisses, la période entre [700 ; 800], et dernièrement la période [1050 ; 1160], qui sont marqué par des chocs particuliers.

Le graphique également nous renseigne sur une non stationnarité de la série, où nous présumons une non-stationnarité de type DS, avec une variance particulièrement instable durant toute l'évolution de la série temporelle.

- Corrélogramme de la série CAC40 :



Ici avec le corrélogramme de la série CAC40 nous pouvons également confirmer que la série n'est pas stationnaire, avec des termes de la fonction d'autocorrélation (ACF) significatifs et qui diminue en valeur progressivement du 1^{er} retard au dernier retard.

2- Test de stationnarité de Dickey-Fuller augmenté :

Model 3 : (Avec tendance et constante)

Null Hypothesis: CAC40 has a unit root		
Exogenous: Constant, Linear Trend		
Lag Length: 0 (Automatic - based on SIC, maxlag=22)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.053788	0.5705
Test critical values:	1% level	-3.965984
	5% level	-3.413694
	10% level	-3.128911
*MacKinnon (1996) one-sided p-values.		
Augmented Dickey-Fuller Test Equation		
Dependent Variable: D(CAC40)		
Method: Least Squares		
Date: 01/14/24 Time: 13:58		
Sample (adjusted): 2 1160		

Included observations: 1159 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
CAC40(-1)	-0.008447	0.004113	-2.053788	0.0402
C	14.93911	7.437617	2.008589	0.0448
@TREND("1")	0.002103	0.001896	1.108866	0.2677
R-squared	0.003870	Mean dependent var		0.417204
Adjusted R-squared	0.002147	S.D. dependent var		20.64858
S.E. of regression	20.62640	Akaike info criterion		8.893606
Sum squared resid	491818.4	Schwarz criterion		8.906691
Log likelihood	-5150.845	Hannan-Quinn criter.		8.898543
F-statistic	2.245643	Durbin-Watson stat		1.867318
Prob(F-statistic)	0.106321			

A travers la lecture du tableau représentant le test de stationnarité de ADF, au modèle n°3, avec tendance et constante. Nous remarquons que la tendance n'est pas significative, avec un t-statistic à (1,10) est inférieur à la valeur tabulée de Dickey-Fuller qui est de (2.78) à 5% de risque. Nous éliminons donc la tendance de notre modèle, et nous passons au modèle n°2 (avec constante).

Modèle 2 : (Avec constante)

Null Hypothesis: CAC40 has a unit root Exogenous: Constant Lag Length: 0 (Automatic - based on SIC, maxlag=22)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.805839	0.3779
Test critical values:	1% level		-3.435777	
	5% level		-2.863824	
	10% level		-2.568037	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(CAC40) Method: Least Squares Date: 01/14/24 Time: 13:59 Sample (adjusted): 2 1160 Included observations: 1159 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
CAC40(-1)	-0.007093	0.003928	-1.805839	0.0712
C	13.63563	7.344864	1.856485	0.0636
R-squared	0.002811	Mean dependent var		0.417204
Adjusted R-squared	0.001949	S.D. dependent var		20.64858
S.E. of regression	20.62845	Akaike info criterion		8.892943
Sum squared resid	492341.5	Schwarz criterion		8.901667
Log likelihood	-5151.461	Hannan-Quinn criter.		8.896235
F-statistic	3.261054	Durbin-Watson stat		1.867860

Prob(F-statistic)	0.071203
-------------------	----------

A travers la lecture du tableau du modèle n°2, nous concluons ici que la constante avec un t-Statistic à (1,85) est inférieure à la valeur tabulée de Dickey-Fuller avec le nombre d'observations à l'infini ($T=\infty$), qui est de (2,52) à 5% de risque. La constante n'est pas significative, nous l'éliminons de notre modèle et nous Passons au modèle n°1 (sans tendance et sans constante).

Modèle n°1 : (sans tendance et sans constante)

Null Hypothesis: CAC40 has a unit root Exogenous: None Lag Length: 0 (Automatic - based on SIC, maxlag=22)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			0.536641	0.8318
Test critical values:	1% level		-2.566959	
	5% level		-1.941097	
	10% level		-1.616515	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(CAC40) Method: Least Squares Date: 01/14/24 Time: 14:00 Sample (adjusted): 2 1160 Included observations: 1159 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
CAC40(-1)	0.000174	0.000324	0.536641	0.5916
R-squared	-0.000160	Mean dependent var		0.417204
Adjusted R-squared	-0.000160	S.D. dependent var		20.64858
S.E. of regression	20.65023	Akaike info criterion		8.894192
Sum squared resid	493808.1	Schwarz criterion		8.898554
Log likelihood	-5153.184	Hannan-Quinn criter.		8.895838
Durbin-Watson stat	1.875893			

Ici, à travers la lecture des résultats du tableau, le t-statistic de CAC40(-1) qui est de (0.53) est supérieur à sa valeur ADF tabulée de (-1,94), ce qui signifie que la variable CAC40 est non stationnaire en niveau avec un processus DS, nous devons passer à la première différence et vérifier la stationnarité de la série.

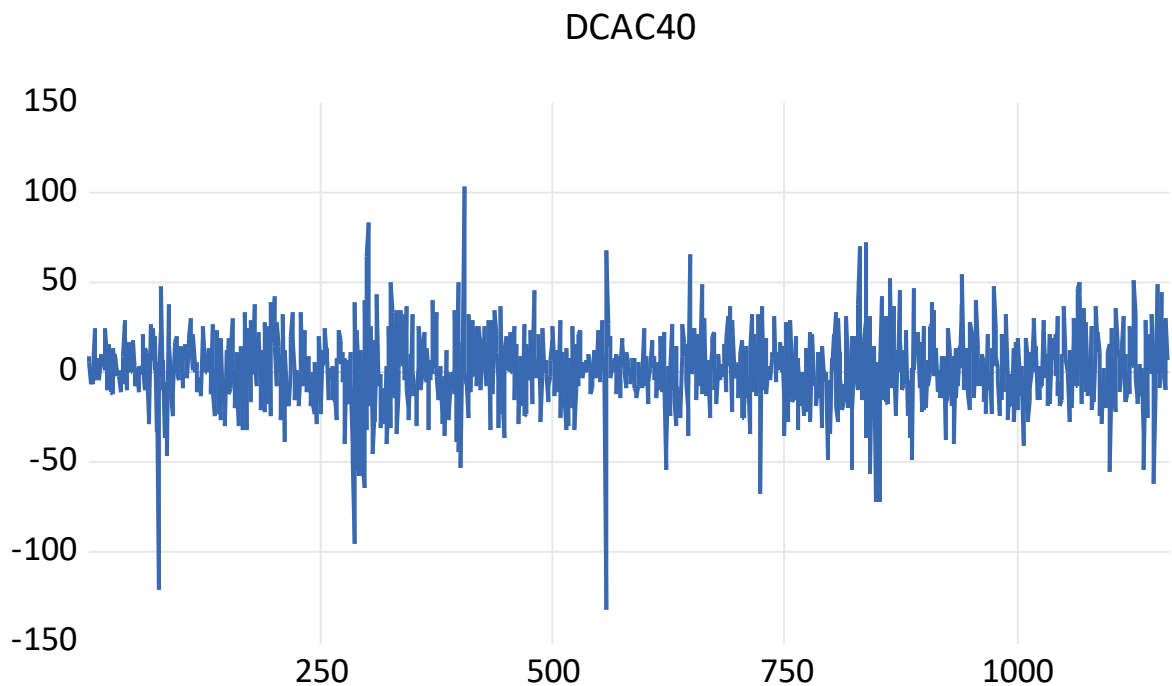
Model n°1 à la première différence : (sans tendance et sans constante)

Null Hypothesis: D(CAC40) has a unit root				
Exogenous: None				
Lag Length: 0 (Automatic - based on SIC, maxlag=22)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-31.95720	0.0000
Test critical values:	1% level		-2.566960	
	5% level		-1.941097	
	10% level		-1.616515	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(CAC40,2)				
Method: Least Squares				
Date: 01/14/24 Time: 14:00				
Sample (adjusted): 3 1160				
Included observations: 1158 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(CAC40(-1))	-0.937653	0.029341	-31.95720	0.0000
R-squared	0.468843	Mean dependent var		-0.001926
Adjusted R-squared	0.468843	S.D. dependent var		28.29298
S.E. of regression	20.62007	Akaike info criterion		8.891270
Sum squared resid	491941.9	Schwarz criterion		8.895635
Log likelihood	-5147.046	Hannan-Quinn criter.		8.892917
Durbin-Watson stat	1.998392			

A travers la première différence nous remarquons ici que la variable différenciée $d(CAC40(-1))$ est stationnaire, avec son t-statistic à (-31,95) inférieur à la valeur tabulée de ADF, qui est de (-1,94) à 5% de risque. Nous devons alors créer et utiliser la variable différenciée stationnaire durant tout le reste des étapes de la méthodologie Box Jenkins.

3- Estimation de la série différenciée et représentation graphique :

Ici, nous avons créé une variable nommée DCAC40 qui représente la variable stationnaire différenciée de la variable initiale CAC40.



Le graphique de la série DCAC40 nous montre une série parfaitement stationnaire, avec une moyenne et une variance pratiquement stable durant toute la période de notre étude. A certains détails près, nous remarquons à certaines période une variation plus prononcée que les autres.

4- Corrélogramme de la série différenciée stationnaire :

Date: 01/14/24 Time: 14:14
Sample (adjusted): 2 1160
Included observations: 1159 after adjustments

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.062	0.062	4.4625	0.035
		2 -0.010	-0.014	4.5772	0.101
		3 -0.039	-0.038	6.3597	0.095
		4 0.059	0.064	10.409	0.034
		5 -0.015	-0.024	10.665	0.058
		6 -0.020	-0.018	11.145	0.084
		7 -0.017	-0.010	11.493	0.119
		8 -0.008	-0.012	11.572	0.171
		9 0.044	0.046	13.813	0.129
		10 0.021	0.016	14.347	0.158
		11 0.018	0.016	14.712	0.196
		12 0.040	0.043	16.558	0.167

Le corrélogramme a cette étape nous renseigne sur l'ordre de notre ARMA(p,q) à prendre en compte, à la partie « Autocorrélation » nous remarquons que le premier retard est significatif et le reste des retards ne le sont pas, ce qui fait que nous avons un MA(1). Sur la partie « Partial Corrélation », même constat également, le premier terme qui représente le premier retard est significatif, ce qui fait que nous avons un AR(1). Ce qui fait au total un ARMA(1,1).

Remarque : Ne prenez en considération que les trois premiers retards sur votre corrélogramme (max = 3 retards).

5- Modèles ARMA retenus :

A partir d'un modèle ARMA(1,1), nous pourrions estimer :

- un ARMA(1,1) avec constante.

Dependent Variable: DCAC40 Method: ARMA Maximum Likelihood (OPG - BHHH) Date: 01/14/24 Time: 14:31 Sample: 2 1160 Included observations: 1159 Convergence achieved after 21 iterations Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.417758	0.655863	0.636959	0.5243
AR(1)	-0.025527	0.375966	-0.067898	0.9459
MA(1)	0.088188	0.372922	0.236477	0.8131
SIGMASQ	424.3261	10.59145	40.06308	0.0000
R-squared	0.003920	Mean dependent var		0.417204
Adjusted R-squared	0.001333	S.D. dependent var		20.64858
S.E. of regression	20.63481	Akaike info criterion		8.895285
Sum squared resid	491793.9	Schwarz criterion		8.912732
Log likelihood	-5150.818	Hannan-Quinn criter.		8.901869
F-statistic	1.515033	Durbin-Watson stat		1.999976
Prob(F-statistic)	0.208913			

- Un ARMA(1,1) sans constante.

Dependent Variable: DCAC40 Method: ARMA Maximum Likelihood (OPG - BHHH) Date: 01/14/24 Time: 14:34 Sample: 2 1160 Included observations: 1159 Convergence achieved after 24 iterations Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.023450	0.371323	-0.063153	0.9497
MA(1)	0.086461	0.368298	0.234758	0.8144
SIGMASQ	424.4810	10.48973	40.46634	0.0000
R-squared	0.003556	Mean dependent var		0.417204
Adjusted R-squared	0.001832	S.D. dependent var		20.64858
S.E. of regression	20.62965	Akaike info criterion		8.893925
Sum squared resid	491973.5	Schwarz criterion		8.907010
Log likelihood	-5151.029	Hannan-Quinn criter.		8.898862
Durbin-Watson stat	1.999933			

- Un MA(1) avec constante.

Dependent Variable: DCAC40				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 01/14/24 Time: 14:32				
Sample: 2 1160				
Included observations: 1159				
Convergence achieved after 12 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.417824	0.651939	0.640894	0.5217
MA(1)	0.062896	0.023178	2.713603	0.0068
SIGMASQ	424.3290	10.58295	40.09552	0.0000
R-squared	0.003913	Mean dependent var		0.417204
Adjusted R-squared	0.002189	S.D. dependent var		20.64858
S.E. of regression	20.62596	Akaike info criterion		8.893566
Sum squared resid	491797.3	Schwarz criterion		8.906652
Log likelihood	-5150.822	Hannan-Quinn criter.		8.898504
F-statistic	2.270493	Durbin-Watson stat		2.000290
Prob(F-statistic)	0.103722			

- Un AR(1) avec constante.

Dependent Variable: DCAC40				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 01/14/24 Time: 14:33				
Sample: 2 1160				
Included observations: 1159				
Convergence achieved after 8 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.417978	0.653948	0.639161	0.5228
AR(1)	0.061929	0.023384	2.648354	0.0082
SIGMASQ	424.3595	10.58949	40.07366	0.0000
R-squared	0.003841	Mean dependent var		0.417204
Adjusted R-squared	0.002118	S.D. dependent var		20.64858
S.E. of regression	20.62670	Akaike info criterion		8.893638
Sum squared resid	491832.7	Schwarz criterion		8.906724
Log likelihood	-5150.863	Hannan-Quinn criter.		8.898576
F-statistic	2.228733	Durbin-Watson stat		1.998042
Prob(F-statistic)	0.108127			

- Un AR(1) sans constante.

Dependent Variable: DCAC40				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 01/14/24 Time: 14:35				
Sample: 2 1160				
Included observations: 1159				
Convergence achieved after 8 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.062302	0.023354	2.667684	0.0077
SIGMASQ	424.5132	10.48391	40.49188	0.0000
R-squared	0.003480	Mean dependent var		0.417204
Adjusted R-squared	0.002619	S.D. dependent var		20.64858
S.E. of regression	20.62152	Akaike info criterion		8.892275
Sum squared resid	492010.8	Schwarz criterion		8.900998
Log likelihood	-5151.073	Hannan-Quinn criter.		8.895567
Durbin-Watson stat	1.998054			

A travers tous les résultats de nos estimations, nous constatons que les modèles : ARMA(1,1) avec constante, ARMA(1,1) sans constante, MA(1) avec constante, Un AR(1) avec constante. Sont tous rejetés car :

- On a des termes non significatifs avec une probabilité supérieur a (0,05) qui est 5% de risque.
- Ou bien un t-statistic inférieur a la valeur tabulée de Student a ($n > 30$) qui est de (1,96).

Remarque : Quand le nombre d'observation ($n > 30$) alors la valeur tabulée de student est de (1,96).

- Le modèle retenu est donc un AR(1) sans constante, car tous ses termes sont significatifs:

Dependent Variable: DCAC40				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 01/14/24 Time: 14:35				
Sample: 2 1160				
Included observations: 1159				
Convergence achieved after 8 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.062302	0.023354	2.667684	0.0077
SIGMASQ	424.5132	10.48391	40.49188	0.0000
R-squared	0.003480	Mean dependent var		0.417204
Adjusted R-squared	0.002619	S.D. dependent var		20.64858
S.E. of regression	20.62152	Akaike info criterion		8.892275
Sum squared resid	492010.8	Schwarz criterion		8.900998
Log likelihood	-5151.073	Hannan-Quinn criter.		8.895567
Durbin-Watson stat	1.998054			

Avec un AR(1) significatif ($2,66 > 1,96$) et une probabilité ($0,007 < 0,05$) de risque ; un coefficient de détermination (R-squared) a 0,3% très faible, de même pour le coefficient de détermination ajusté (Adjusted R-Squared) a 0,2%. On remarque que ce modèle n'a pas de F-statistic qui est un test de significativité global (Test de Fisher), et ce car il y'a qu'un seul terme dans notre modèle actuel et c'est le terme AR(1).

Remarque : Ici nous pouvons également estimer un modèle MA(1) sans constante, qui est également significatif statistiquement, avec également des critères d'informations plus minimisé que le AR(1) sans constante, mais pour des raisons de bonne spécification du modèle, et afin d'éviter que notre série soit représentée qu'avec une erreur retardée, j'ai préféré prendre que le AR(1) sans constante.

- Notre modèle final AR(1) sans constante sera ainsi :

$$DCAC40 = 0.06 * DCAC40(-1) + \varepsilon_t$$

6- Autocorrélation des erreurs :

Date: 01/14/24 Time: 14:43

Sample (adjusted): 2 1160

Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.001	0.001	0.0003	
		2 -0.011	-0.011	0.1519	0.697
		3 -0.043	-0.043	2.2552	0.324
		4 0.063	0.063	6.8475	0.077
		5 -0.017	-0.019	7.1979	0.126
		6 -0.018	-0.019	7.5946	0.180
		7 -0.016	-0.011	7.8781	0.247
		8 -0.010	-0.016	7.9928	0.333
		9 0.043	0.044	10.185	0.252
		10 0.018	0.018	10.551	0.308
		11 0.014	0.015	10.781	0.375
		12 0.039	0.045	12.599	0.320

A travers l'analyse du corrélogramme des résidus d'estimation de notre modèle final AR(1) sans constante : Nous déduisons une absence d'autocorrélation des erreurs, avec des termes non significatifs et des probabilités supérieurs au risque de 5%.

7- Test d'hétéroscédasticité :

- Corrélogramme des résidus au carré :

Date: 01/14/24 Time: 14:45

Sample (adjusted): 2 1160

Included observations: 1159 after adjustments

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.106	0.106	13.078	0.000
		2 0.101	0.091	25.024	0.000
		3 0.096	0.078	35.719	0.000
		4 0.043	0.018	37.858	0.000
		5 0.100	0.081	49.571	0.000
		6 0.055	0.028	53.163	0.000
		7 0.007	-0.021	53.220	0.000
		8 0.031	0.011	54.360	0.000
		9 0.069	0.058	59.881	0.000
		10 0.111	0.092	74.264	0.000
		11 0.045	0.009	76.596	0.000
		12 0.031	0.001	77.727	0.000

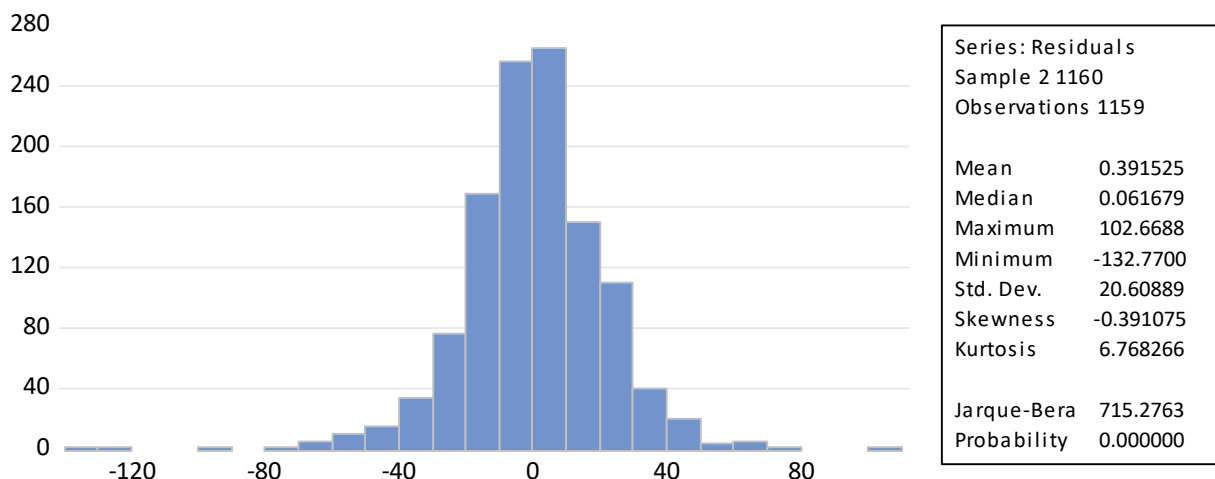
A travers le corrélogramme des résidus au carré, nous pouvons en déduire que les probabilités sont inférieures au seuil du risque de 5%, mais également nous remarquons des termes sur la fonction d'autocorrélation (ACF) significatifs. Nous allons donc tester l'hétéroscédasticité avec le test de ARCH à un degré de 3 retards pour les résidus au carré.

Heteroskedasticity Test: ARCH				
F-statistic	9.996329	Prob. F(3,1152)	0.0000	
Obs*R-squared	29.32960	Prob. Chi-Square(3)	0.0000	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 01/14/24 Time: 14:46				
Sample (adjusted): 5 1160				
Included observations: 1156 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	318.9724	35.43470	9.001696	0.0000
RESID^2(-1)	0.089118	0.029374	3.033891	0.0025
RESID^2(-2)	0.083522	0.029391	2.841777	0.0046
RESID^2(-3)	0.077978	0.029375	2.654541	0.0081
R-squared	0.025372	Mean dependent var	425.4914	
Adjusted R-squared	0.022834	S.D. dependent var	1018.269	
S.E. of regression	1006.577	Akaike info criterion	16.66995	
Sum squared resid	1.17E+09	Schwarz criterion	16.68744	
Log likelihood	-9631.232	Hannan-Quinn criter.	16.67655	
F-statistic	9.996329	Durbin-Watson stat	2.002761	
Prob(F-statistic)	0.000002			

Le test de ARCH révèle la présence d'une hétéroscédasticité significative pour les trois termes de retards des résidus au carré avec des probabilités inférieures au risque de 5%.

Remarque : Au-delà de trois termes significatifs pour les résidus au carré sur le test de ARCH, on peut envisager un GARCH(1,1) comme modèle, afin de prendre en considération et corriger l'hétéroscédasticité de notre modèle.

8- Test de normalité des erreurs :



A travers le test de normalité des erreurs, la statistique de Jarque-Bera avec une probabilité inférieure à 5% de risque, indique que les résidus d'estimation ne suivent pas une loi normale.

En conclusion :

Nous ne pouvons utiliser ce modèle à des fins de prévisions, car un modèle efficace doit vérifier ces trois règles à la fois, et ce après vérification que tous les termes du modèle sont significatifs, et ainsi être globalement significatif à travers le test de Fisher pour les modèles à plus de deux termes, les résidus du modèle doivent être caractérisés par un Bruit Blanc et donc vérifier :

- Absence d'autocorrélation.
- Absence d'hétéroscédasticité.
- Doivent suivre une loi normale.