



ÉCOLE CENTRALE LYON

MOS 6.5 - ECONOMETRIE DES SÉRIES TEMPORELLES
PROJET
RAPPORT

L'étude d'une série temporelle du prix de l'or

Élèves :
MARIEM NAIMI

Enseignant :
Christian DE PERETTI

16 juin 2024

Table des matières

1	Introduction	3
1.1	Contexte	3
1.2	Source de données	3
2	Statistiques descriptives et régression linéaire	3
2.1	Visualisation de l'Évolution Temporelle du Prix High de l'Or	3
2.2	Étude de la distribution de la série	4
2.3	Analyse de la Corrélation Temporelle	5
2.3.1	Calcul et Visualisation de l'Autocorrélation (ACF) et de l'Autocorrélation Partielle (PACF)	5
2.3.2	Focus sur les Derniers 365 Jours du Prix High de l'Or	6
2.3.3	Analyse de la Corrélation avec un Lag (Lag Plot) - Premiers 60 Jours	6
2.3.4	Analyse de l'Autocorrélation sur l'Année 2023	8
2.3.5	l'autocorrélation et l'autocorrélation partielle sur une plage de 70 lags	8
2.4	Pré-traitement : Stationnarité	9
2.4.1	Test de Dickey-Fuller augmenté	9
2.4.2	Test de Kwiatkowski Phillips Schmidt Shin (test KPSS)	10
2.5	Stationnarisation	11
2.5.1	Première différence :	11
2.5.2	Rendements logarithmiques :	13
3	Modèles univariés	16
3.1	Modélisation AR	16
3.1.1	Modèle AR(3) complet	16
3.1.2	Modèle AR(2) complet	17
3.1.3	Modèle AR(1) avec constante	18
3.1.4	Modèle AR(1) sans constante	18
3.1.5	Comparaison des modèles	19
3.2	Modélisation MA	20
3.2.1	Modèle MA(2) complet	21
3.2.2	Modèle MA(1) avec constante	22
3.2.3	Modèle MA(1) sans constante	22
3.3	Conclusions sur les modèles AR et MA	23
3.4	Modélisation ARMA	23
3.5	Modélisation GARCH	25
3.5.1	Modèle GARCH(2,2)	27
3.5.2	Modèle GARCH(1,2)	28
3.5.3	GARCH(1,1)	29
3.5.4	GARCH(1,1)-GED	29
3.5.5	Restriction IGARCH	30
3.5.6	IGARCH(1,1)- GED	31
3.5.7	ARMA(2,2)-GARCH(1,1) GED	32
3.5.8	AR(1)-GARCH(1,1)-GED	34

4	Modèles multivariés	34
4.1	Test de cointégration	34
4.1.1	Test de cointégration entre le prix high et le volume	35
4.2	Modèle VAR	38
5	Éventuelles Améliorations - Prédictions	42
5.1	EWMA Modèle à Moyenne Mobile Exponentielle Pondérée	42
6	Conclusion	43

1 Introduction

1.1 Contexte

L'or, depuis des millénaires, a joué un rôle crucial dans l'économie mondiale en tant que réserve de valeur et actif refuge. Historiquement, l'or a été utilisé comme monnaie, moyen d'échange et mesure de valeur, et même aujourd'hui, il demeure un actif crucial dans les portefeuilles d'investissement, offrant une diversification et une protection contre l'incertitude économique et les risques inflationnistes.

La volatilité des prix de l'or est une caractéristique remarquable de ce marché. Les prix de l'or sont influencés par une multitude de facteurs économiques, géopolitiques et financiers, notamment l'inflation, les taux d'intérêt, les mouvements des devises, les tensions géopolitiques et la demande industrielle et de bijouterie. Cette complexité rend la modélisation des prix de l'or difficile mais essentielle pour les investisseurs, les gouvernements et les institutions financières.

La capacité à modéliser et à prédire les prix de l'or revêt une importance significative dans de nombreux contextes. Pour les investisseurs, une meilleure compréhension des tendances et des cycles des prix de l'or peut fournir des indications précieuses pour la prise de décision en matière d'allocation d'actifs et de gestion des risques. Les gouvernements et les banques centrales surveillent également de près les prix de l'or en tant qu'indicateur de la santé économique globale et comme référence pour la politique monétaire. De plus, dans le contexte actuel de mondialisation financière et de marchés interconnectés, la capacité à prédire les prix de l'or peut contribuer à atténuer les risques systémiques et à stabiliser les marchés.

1.2 Source de données

Les données utilisées dans cette étude proviennent de Yahoo Finance, une plateforme largement reconnue pour sa fiabilité et sa crédibilité en matière de données financières. Les données spécifiques concernent les contrats à terme sur l'or, symbolisés par "Gold Jun 24 (GC=F)". Chaque entrée de données comprend les informations suivantes :

- **Date** : La date de la transaction.
- **Ouverture** : Le prix d'ouverture du contrat pour la journée de négociation.
- **Élevé** : Le prix le plus élevé atteint pendant la journée.
- **Faible** : Le prix le plus bas atteint pendant la journée.
- **Clôture** : Le prix auquel le contrat a été clôturé à la fin de la journée.
- **clôture ajusté** : Le prix de clôture ajusté pour tenir compte de tout événement tel que les dividendes ou les émissions d'actions survenues après la fermeture du marché.
- **Volume** : Le nombre d'unités de contrat négociées au cours de la journée.

2 Statistiques descriptives et régression linéaire

2.1 Visualisation de l'Évolution Temporelle du Prix High de l'Or

La figure ci-dessous montre la série temporelle des prix High de l'or depuis le début de l'année 2008 jusqu'à la fin de la période de données disponibles.

L'analyse de la série temporelle des prix élevés de l'or révèle des tendances dynamiques au fil du temps.

Observées sur la période de 2008 à la fin des données disponibles, des périodes de hausses et de baisses significatives sont apparentes. Notamment, une hausse notable se manifeste vers 2011, coïncidant probablement avec des facteurs tels que l'instabilité financière mondiale, stimulant la demande d'or en tant que valeur refuge. Cependant, entre 2012 et 2017, une tendance à la baisse se dessine, suggérant des changements économiques ou politiques ayant un impact négatif sur la demande d'or. Une reprise vers des valeurs hausses post-2017 indique possiblement un retour de l'intérêt pour l'or en période d'incertitude économique.

La non-stationnarité observée nécessite une approche de modélisation plus complexe, telle que l'utilisation de modèles ARIMA, et une décomposition de la série temporelle pour mieux comprendre les composantes sous-jacentes.



FIGURE 1 – la série temporelle des prix High de l'or

2.2 Étude de la distribution de la série

L'histogramme de la série est donné dans la figure suivante :

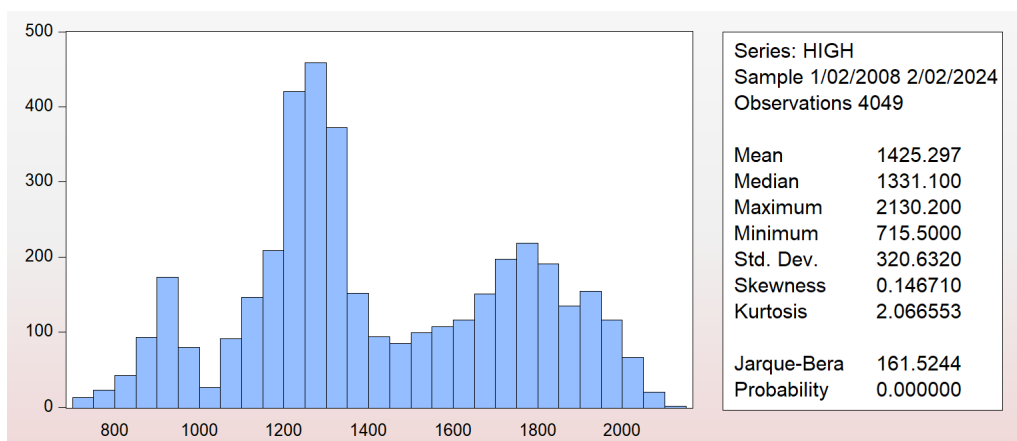


FIGURE 2 – Histogramme et statistiques descriptives de la série

On remarque que :

- Le **Skewness** est positif. La série est donc asymétrique et étalée vers la droite. Dans le contexte spécifique du prix de l'or, un skewness positif implique que les chocs positifs (augmentations de prix) ont un impact plus important que les chocs négatifs (baisses de prix). En d'autres termes, les variations positives des prix de l'or sont plus fréquentes et/ou plus importantes que les variations négatives. L'idée que cette asymétrie puisse être le signe d'une non-linéarité dans le processus d'évolution du prix de l'or est une interprétation plausible. La non-linéarité pourrait résulter de divers facteurs, tels que des réactions différentes du marché aux hausses et aux baisses des prix, des changements de comportement des investisseurs, ou d'autres forces économiques spécifiques au marché de l'or.
- une **kurtosis** de 2.09 suggère une distribution où les événements extrêmes sont moins fréquents et les queues sont relativement proches de celles d'une distribution normale. Cela peut indiquer que les variations de prix de l'or ne sont pas excessivement extrêmes.

2.3 Analyse de la Corrélation Temporelle

2.3.1 Calcul et Visualisation de l'Autocorrélation (ACF) et de l'Autocorrélation Partielle (PACF)

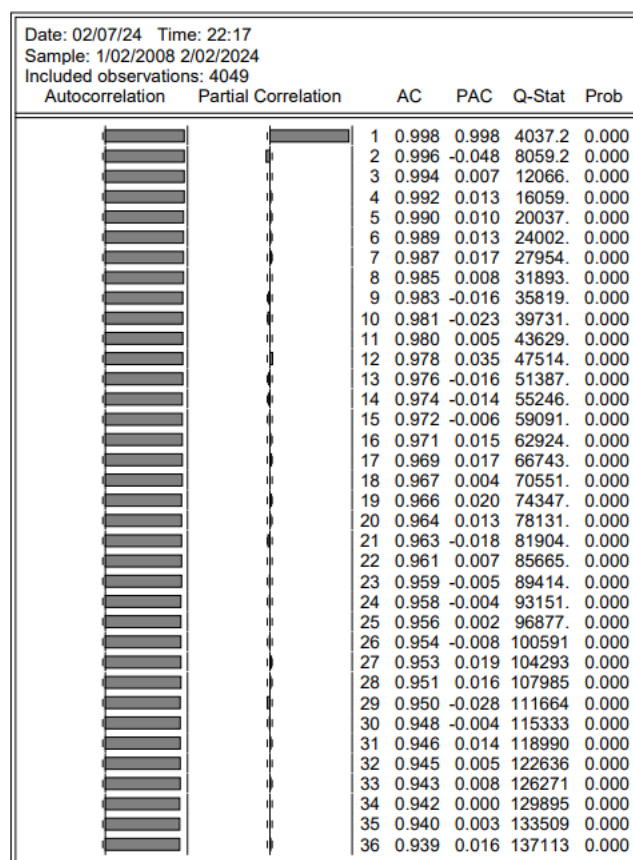


FIGURE 3 – Correlogram of HIGH

Dans cette analyse, nous évaluons les hypothèses concernant l'auto-corrélation des erreurs dans la série temporelle des prix élevés de l'or. Les hypothèses nulle (H_0) et de recherche (H_1) sont formulées comme suit :

- **Hypothèse Nulle (H_0)** : Il n'y a pas d'auto-corrélation des erreurs d'ordre 1 à r .
- **Hypothèse de Recherche (H_1)** : Il y a une auto-corrélation des erreurs d'ordre 1 à r .

Les observations sont les suivantes :

- Toutes les auto-corrélations sont significativement différentes de 0 et ne diminuent presque pas. Cela indique une non-stationnarité de la série temporelle.
- L'auto-corrélation partielle est quasiment nulle sauf pour la première valeur, qui est quasiment égale à 1. Cela suggère que la série pourrait suivre un processus autorégressif AR(1) avec un coefficient de 1, indiquant une racine unitaire et confirmant ainsi la non-stationnarité.

Ces résultats soulignent la nécessité d'appliquer des transformations, telles que la différenciation, pour rendre les données plus stationnaires avant de modéliser les prix de l'or. Cette approche est cruciale pour garantir la validité des hypothèses dans le cadre de la modélisation des séries temporelles financières.

2.3.2 Focus sur les Derniers 365 Jours du Prix High de l'Or

On remarque que la série temporelle des prix élevés de l'or ne présente pas de motifs clairs ou réguliers. Cela peut être dû à des fluctuations aléatoires ou à des facteurs externes imprévisibles qui influent sur les prix. Aussi, les variations dans les prix de l'or peuvent être influencées par une multitude de facteurs économiques, politiques et géopolitiques. Cette complexité peut rendre difficile l'identification de motifs clairs à partir d'une simple visualisation.

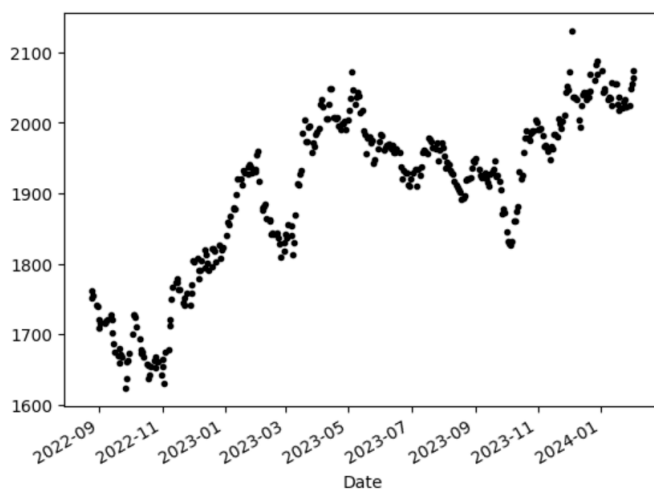


FIGURE 4 – une version simplifiée de la série temporelle se concentrant uniquement sur les derniers 365 jours

2.3.3 Analyse de la Corrélation avec un Lag (Lag Plot) - Premiers 60 Jours

Les figures ci dessous présentent Lag Plot sur l'Ensemble des Données et sur les 60 Derniers Jours. Ces graphiques sont générés pour examiner la relation linéaire entre les

valeurs consécutives.

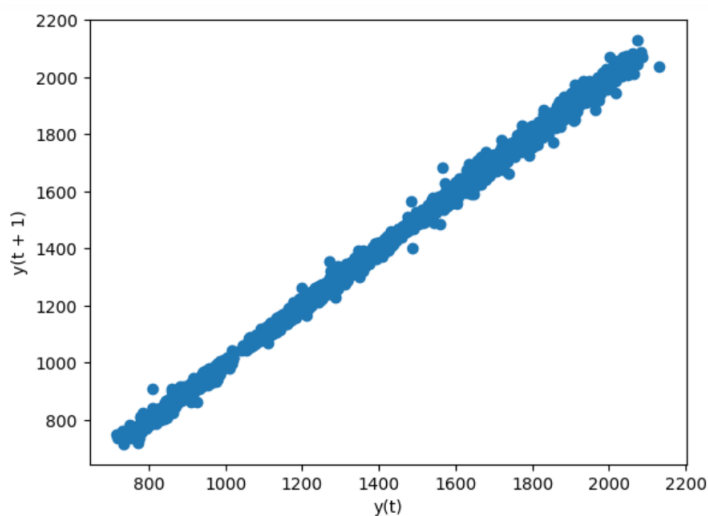


FIGURE 5 – Lag Plot (Diagramme de Retard) sur l'Ensemble des Données

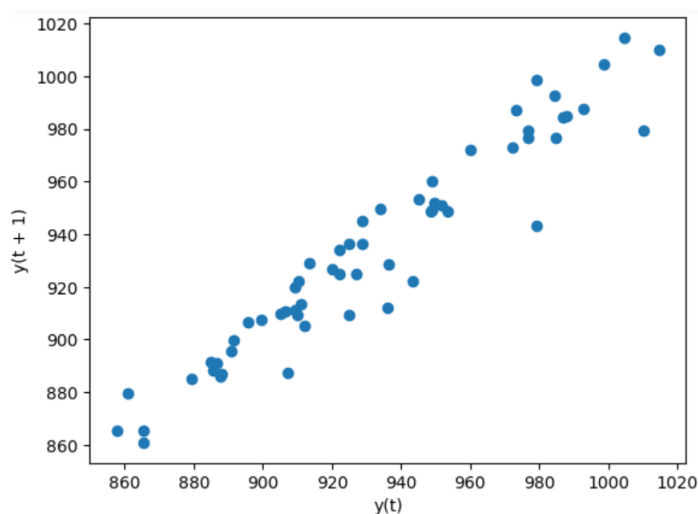


FIGURE 6 – Lag Plot (Diagramme de Retard) sur les 60 Derniers Jours

Dans le lag plot sur l'ensemble des données, une relation parfaitement linéaire entre chaque jour et son jour suivant est observée. Cela suggère une forte corrélation ou une dépendance linéaire entre les valeurs consécutives sur l'ensemble de la période.

Cependant, lorsque l'on se concentre sur les 60 derniers jours dans le lag plot zoomé, la linéarité apparente disparaît. Cela peut indiquer que sur une période plus courte, la corrélation entre les valeurs consécutives est moins prononcée, et d'autres tendances ou variations deviennent apparentes.

2.3.4 Analyse de l'Autocorrélation sur l'Année 2023

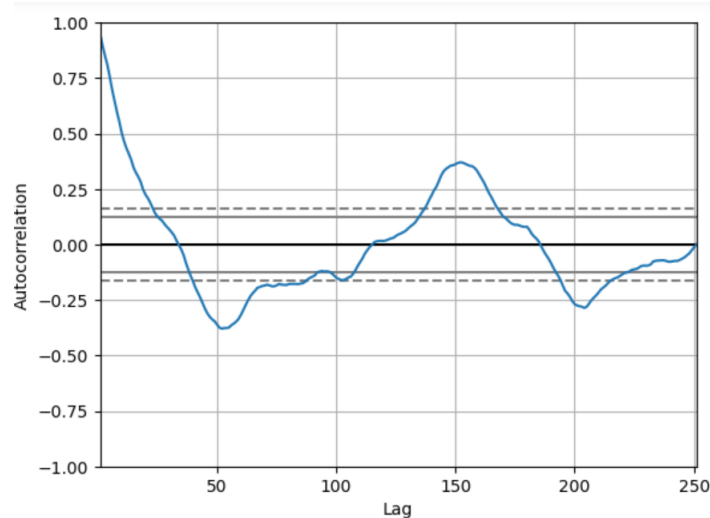


FIGURE 7 – l'Autocorrélation sur l'Année 2023

On Observe :

- Une autocorrélation d'environ -0.35 à un lag de 50 indique une corrélation négative entre les valeurs actuelles et celles qui remontent à environ 50 jours auparavant.
- Une autocorrélation d'environ 0.4 à un lag de 150 indique une corrélation positive entre les valeurs actuelles et celles qui remontent à environ 150 jours auparavant.

La variation d'autocorrélation entre l'ensemble de la série et l'année 2023 peut indiquer des changements dans les motifs temporels au fil du temps.

En effet cette différence dans les signes d'autocorrélation et les lags associés peut être utilisée pour ajuster des modèles de séries temporelles spécifiques pour différentes périodes.

2.3.5 l'autocorrélation et l'autocorrélation partielle sur une plage de 70 lags

Une plage de lags étendue de 70 jours permet de mieux comprendre les corrélations directes entre les valeurs actuelles et celles des jours précédents, ce qui aide ainsi à affiner la modélisation des séries temporelles.

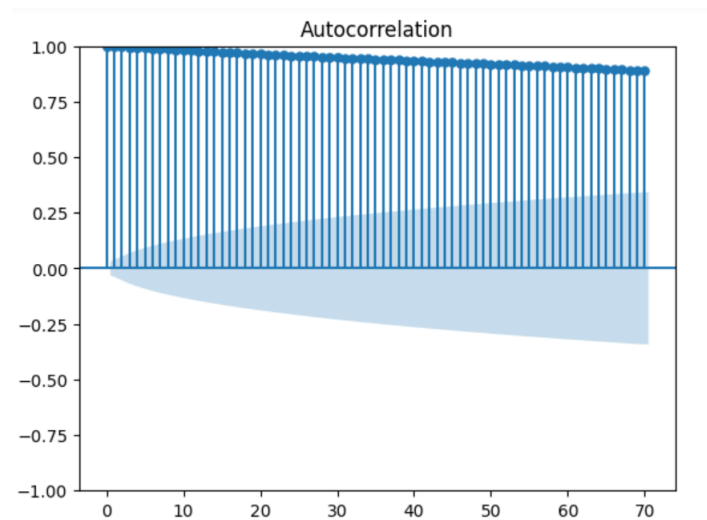


FIGURE 8 – l'autocorrélation sur une plage étendue de lags

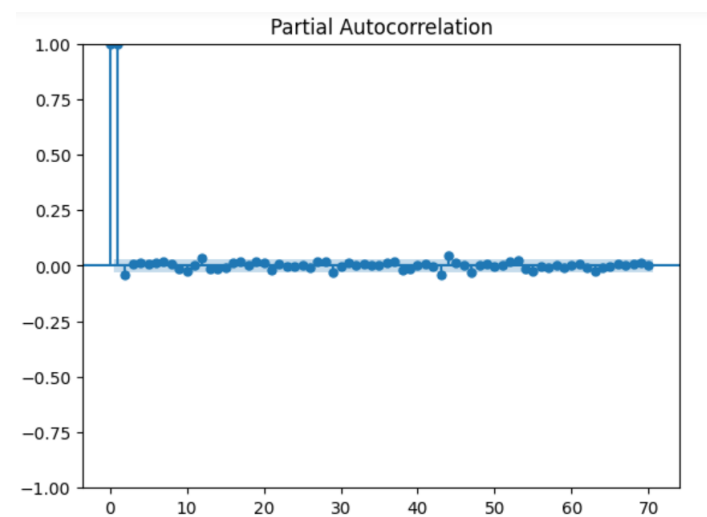


FIGURE 9 – l'autocorrélation partielle sur une plage étendue de lags

L'analyse de l'acf et la pcacf sur une plage étendue de lags confirme la persistance des corrélations directes entre les valeurs actuelles et celles des jours précédents, corroborant ainsi la stabilité des motifs temporels identifiés dans l'ensemble de la série temporelle des prix élevés de l'or.

2.4 Pré-traitement : Stationnarité

2.4.1 Test de Dickey-Fuller augmenté

Le test Augmented Dickey-Fuller (ADF) évalue la stationnarité d'une série temporelle à travers trois modèles. Le premier modèle, sans tendance ni drift, s'exprime comme suit :

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t$$

Le deuxième modèle, intégrant uniquement le drift, est formulé par :

$$\Delta y_t = a_0 + \gamma y_{t-1} + \epsilon_t$$

Enfin, le troisième modèle, considérant à la fois la tendance et le drift, est défini par :

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \epsilon_t$$

Dans ces équations, ϵ_t suit une distribution normale $\mathcal{N}(0, \sigma^2)$.

Les hypothèses testées sont les suivantes :

- Hypothèse Nulle (H_0) : $\gamma = 0$ (non-stationnarité)
- Hypothèse Alternative (H_1) : $\gamma < 1$ (stationnarité)

Les étapes du test ADF impliquent :

- l'estimation de γ par la méthode des moindres carrés ordinaires (MCO)
- le calcul de la statistique t_γ
- la détermination des valeurs critiques dans la table de Dickey-Fuller
- si $t_\gamma < \text{valeur critique}$, on rejette H_0 , indiquant une série stationnaire.

Augmented Dickey-Fuller Unit Root Test on HIGH

Null Hypothesis: HIGH has a unit root Exogenous: Constant Lag Length: 1 (Automatic - based on SIC, maxlag=30)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.391943	0.5878
Test critical values:				
1% level			-3.431781	
5% level			-2.862058	
10% level			-2.567089	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(HIGH) Method: Least Squares Date: 02/08/24 Time: 10:13 Sample (adjusted): 1/04/2008 2/02/2024 Included observations: 4047 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
HIGH(-1)	-0.000958	0.000688	-1.391943	0.1640
D(HIGH(-1))	0.087974	0.015666	5.615578	0.0000
C	1.633224	1.005230	1.624726	0.1043
R-squared	0.008121	Mean dependent var	0.294144	
Adjusted R-squared	0.007631	S.D. dependent var	14.07653	
S.E. of regression	14.02272	Akaike info criterion	8.119975	
Sum squared resid	795198.4	Schwarz criterion	8.124650	
Log likelihood	-16427.77	Hannan-Quinn criter.	8.121631	
F-statistic	16.55562	Durbin-Watson stat	1.997794	
Prob(F-statistic)	0.000000			

Augmented Dickey-Fuller Unit Root Test on HIGH

Null Hypothesis: HIGH has a unit root Exogenous: Constant, Linear Trend Lag Length: 1 (Automatic - based on SIC, maxlag=30)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-2.062756	0.5659
Test critical values:				
1% level			-3.960307	
5% level			-3.410916	
10% level			-3.127263	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(HIGH) Method: Least Squares Date: 02/08/24 Time: 10:13 Sample (adjusted): 1/04/2008 2/02/2024 Included observations: 4047 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
HIGH(-1)	-0.001979	0.000960	-2.062756	0.0392
D(HIGH(-1))	0.088467	0.015667	5.646781	0.0000
C	2.275314	1.089458	2.088483	0.0368
@TREND("1/02/2008")	0.000402	0.000263	1.527214	0.1268
R-squared	0.008693	Mean dependent var	0.294144	
Adjusted R-squared	0.007958	S.D. dependent var	14.07653	
S.E. of regression	14.02041	Akaike info criterion	8.119893	
Sum squared resid	794739.9	Schwarz criterion	8.126125	
Log likelihood	-16426.60	Hannan-Quinn criter.	8.122101	
F-statistic	11.81818	Durbin-Watson stat	1.997878	
Prob(F-statistic)	0.000000			

(a) Résultats du test ADF type constante

(b) Résultats du test ADF type constante
+ tendance déterministe

FIGURE 10 – Résultat du test ADF

Les résultats indiquent que, dans les deux cas, la p-value est supérieure à 0.05. Cela signifie qu'on ne peut pas rejeter l'hypothèse nulle (H_0) selon laquelle la série est non stationnaire. De plus, le coefficient de tendance (TREND) est de l'ordre de 10^{-4} . Cela pourrait indiquer une petite pente pour la tendance, mais son impact pourrait être insignifiant.

2.4.2 Test de Kwiatkowski Phillips Schmidt Shin (test KPSS)

Le test de Kwiatkowski, Phillips, Schmidt et Shin (KPSS) constitue un outil supplémentaire dans l'évaluation de la stationnarité d'une série temporelle. Contrairement aux tests précédents, le KPSS pose une hypothèse nulle différente : H_0 stipule que la série (y_t) est stationnaire. L'inclusion de ce test offre une perspective complémentaire pour vérifier la cohérence des résultats obtenus précédemment.

KPSS Unit Root Test on HIGH

Null Hypothesis: HIGH is stationary Exogenous: Constant, Linear Trend Bandwidth: 51 (Newey-West automatic) using Bartlett kernel				
				LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic				0.786744
Asymptotic critical values*:				0.216000
1% level				0.146000
5% level				0.119000
10% level				0.119000
*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)				
Residual variance (no correction)				52800.80
HAC corrected variance (Bartlett kernel)				2648745.
KPSS Test Equation Dependent Variable: HIGH Method: Least Squares Date: 02/08/24 Time: 10:29 Sample: 1/02/2008 2/02/2024 Included observations: 4049				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1038.177	7.222760	143.7368	0.0000
@TREND("1/02/2008")	0.191265	0.003090	61.89262	0.0000
R-squared	0.486271	Mean dependent var	1425.297	
Adjusted R-squared	0.486144	S.D. dependent var	320.6320	
S.E. of regression	229.8410	Akaike info criterion	13.71315	
Sum squared resid	2.14E+08	Schwarz criterion	13.71626	
Log likelihood	-27760.27	Hannan-Quinn criter.	13.71425	
F-statistic	3830.696	Durbin-Watson stat	0.003750	
Prob(F-statistic)	0.000000			

FIGURE 11 – Résultat du test KPSS

Une nouvelle fois, le test de Kwiatkowski, Phillips, Schmidt et Shin (KPSS) confirme que la série n'est pas stationnaire. En effet, la valeur de la statistique dépasse les valeurs critiques aux seuils de risque de 1

2.5 Stationnarisation

Dans cette partie, on cherche à trouver le traitement convenable pour obtenir une série stationnaire.

Une série temporelle est considérée stationnaire au sens strict (ou fort) si elle satisfait à trois conditions :

1. $E[X_t] = \mu$, pour tout t
2. $\text{Var}(X_t) = \sigma^2$, pour tout t
3. $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$, pour tout t

Une série temporelle est considérée stationnaire au sens large (ou faible) si elle satisfait seulement les deux premières conditions (moyenne et variance constantes), mais la covariance peut dépendre du temps.

2.5.1 Première différence :

Nous allons maintenant tenter de rendre la série temporelle stationnaire en travaillant avec la série de première différence. En posant $D_t = y_{t+1} - y_t$, examinons son évolution ainsi que sa distribution. La figure ci-dessous illustre l'évolution de la série de différences des prix élevés de l'or.

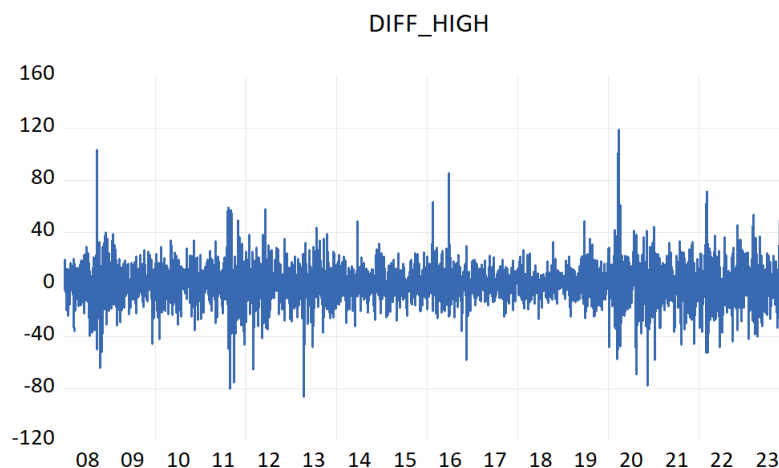


FIGURE 12 – Évolution de la série de première différence

Elle semble stationnaire d'après ce graphique, malgré la présence de deux pics, l'un en 2008 et l'autre en 2020, correspondant à la période de la crise financière mondiale et à la période de la pandémie de COVID-19, respectivement. Ces anomalies pourraient être attribuées à une perturbation ou à une information externe spécifique à ces deux périodes.

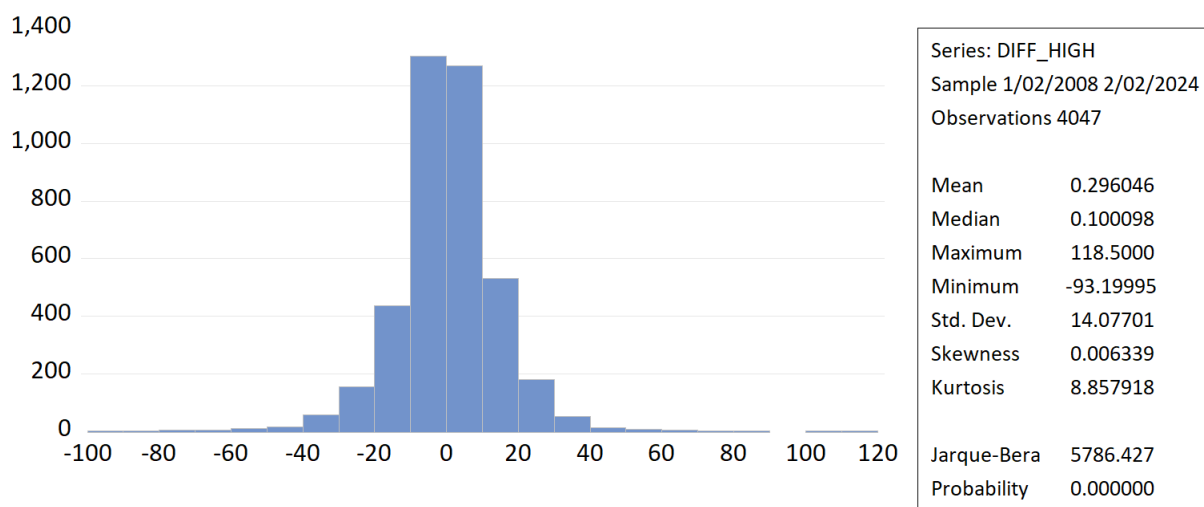
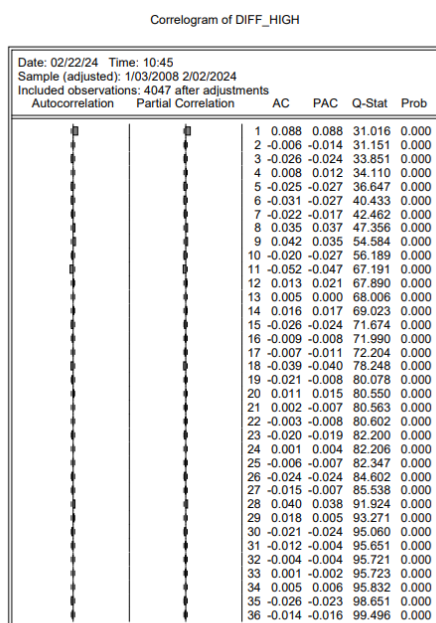
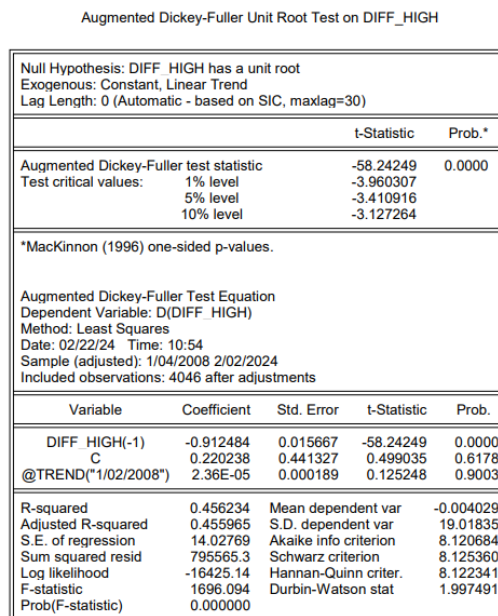


FIGURE 13 – Histogramme de la série de première différence

l'histogramme des premières différences des prix élevés de l'or semble montrer une légère asymétrie positive ($\text{skewness} = 0,0063$) et une distribution plus concentrée autour de la moyenne avec des queues plus épaisses ($\text{kurtosis} = 8,857 > 3$), ce qui pourrait être lié aux pics observés en 2008 et 2020.



(a) corrélogramme



(b) ADF type constante + tendance déterministe

FIGURE 14 – Résultats du test ADF et corrélogramme pour la première différence

Finalement, on voit que les auto-corrélations et les auto-corrélations partielles sont quasiment nulles. et le Dickey-Fuller donne une p-value (quasiment) nulle et donc inférieure à 0.05. On rejette alors H_0 et donc on conclut grâce à ce test que la série est stationnaire.

2.5.2 Rendements logarithmiques :

L'utilisation des rendements logarithmiques $r_t = \log\left(\frac{y_{t+1}}{y_t}\right)$ a démontré son efficacité dans le processus de stationnarisation de la série temporelle représentant les prix de l'or.

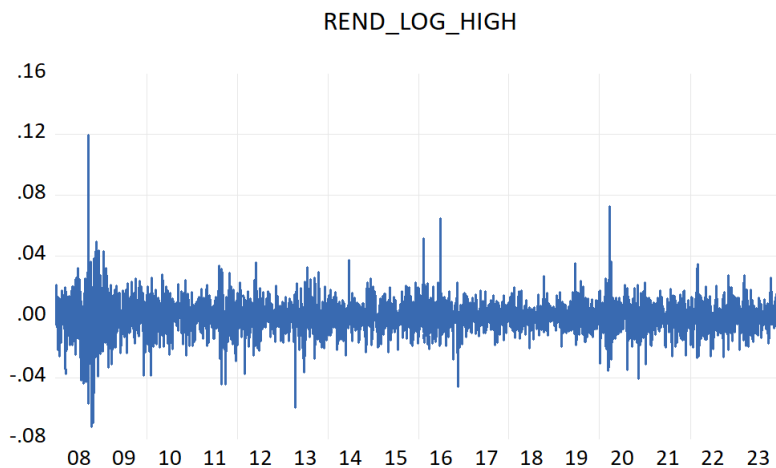


FIGURE 15 – Évolution de la série des rendements

L'observation de cette série suggère une apparente stationnarité, caractérisée par une moyenne et une variance stables. Toutefois, il est crucial de noter la présence inhabituelle d'un pic au milieu de la série, survenu en 2008, année marquée par la crise financière mondiale. Cette anomalie pourrait être attribuée à une perturbation ou à une information externe spécifique à cette période de crise. Étant donné que la série concerne les prix de l'or, il est plausible que des événements économiques, géopolitiques ou financiers liés à la crise de 2008 aient influencé cette variation.

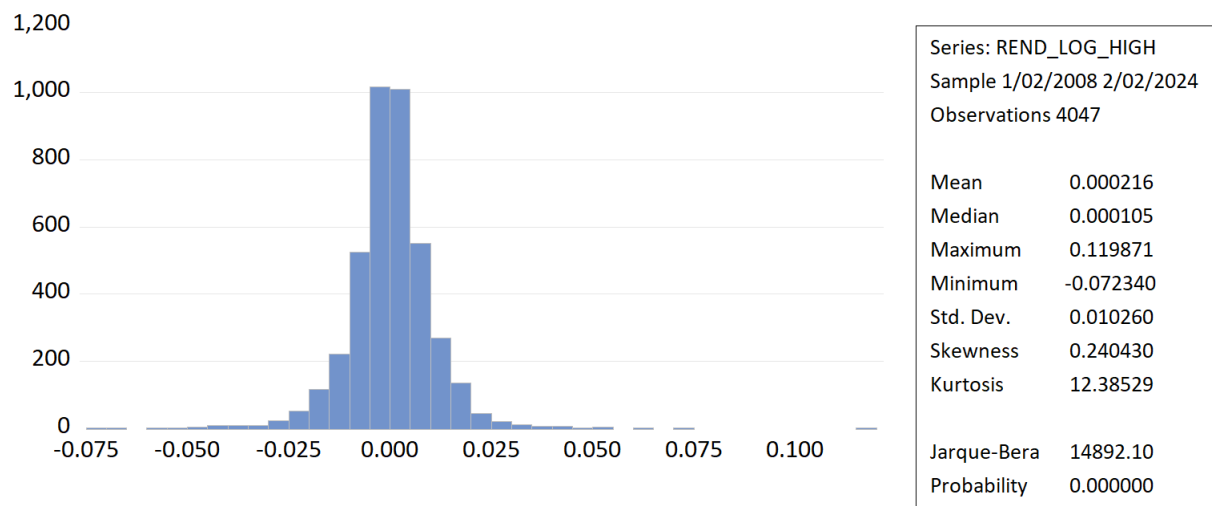
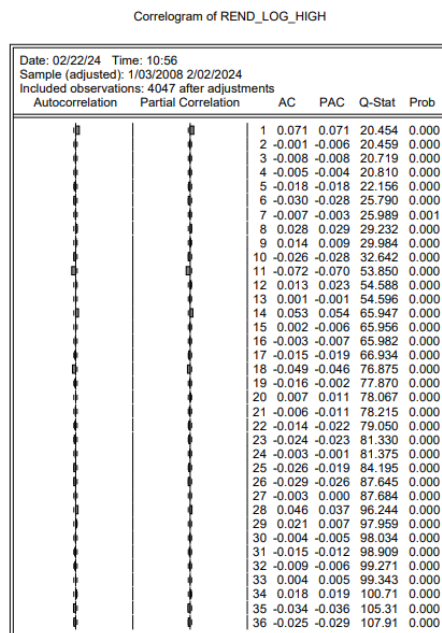
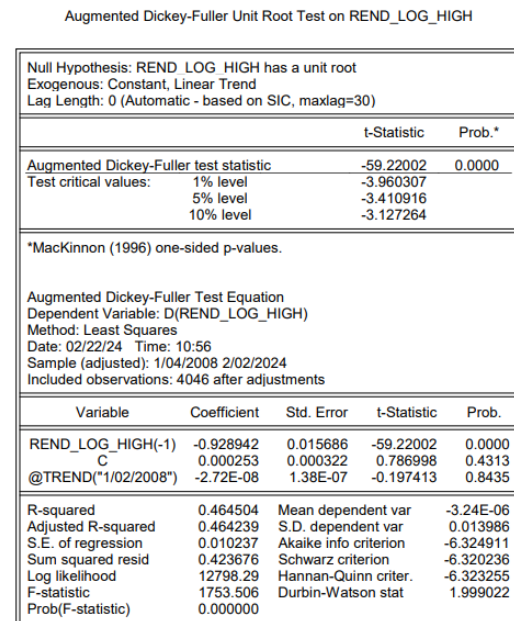


FIGURE 16 – Histogramme de la série de rendements logarithmique

L'histogramme des logarithmes des rendements des prix élevés de l'or révèle des caractéristiques intéressantes de la distribution des données. La légère asymétrie positive, illustrée par une skewness de 0,24, suggère une queue légèrement plus longue du côté droit de la distribution. Parallèlement, la kurtosis élevée de 12,38, bien au-dessus de la valeur de référence de 3 pour une distribution normale, souligne une concentration plus importante des observations autour de la moyenne, accompagnée de queues plus épaisses. Cette distribution plus « lourde » pourrait être associée aux pics observés en 2008, indiquant des variations plus extrêmes par rapport à une distribution normale.



(a) corrélogramme



(b) ADF type constante + tendance déterministe

FIGURE 17 – Résultats du test ADF et corrélograme pour log rendement

Finalement, on voit que les auto-corrélations et les auto-corrélations partielles sont quasiment nulles. et le Dickey-Fuller donne une p-value (quasiment) nulle et donc inférieure à 0.05. On rejette alors H_0 et donc on conclut grâce à ce test que la série est stationnaire.

Les rendements logarithmiques stationnarisent la série de manière convenable, nous travaillons dans la suite avec cette série.

Pourquoi favorisation de logarithmes des rendements et par rapport à la simple différence ? La stationnarisation par les logarithmes des rendements présente des avantages significatifs dans l'étude des séries de prix de l'or en finance.

En utilisant les log rendements, on capture les variations relatives plutôt que les niveaux absolus, permettant une comparaison plus robuste des mouvements de prix sur différentes périodes.

De plus, cette approche facilite l'interprétation des résultats en termes de pourcentage de variation. La stationnarité des log rendements contribue à stabiliser la moyenne et la variance, rendant l'analyse statistique et la modélisation plus fiables. Ceci est particulièrement crucial dans la finance, où la compréhension précise des variations des prix est essentielle pour la prise de décision et la gestion des risques.

3 Modèles univariés

3.1 Modélisation AR

Dans un modèle AR, la valeur actuelle de la série temporelle est modélisée comme une combinaison linéaire de ses valeurs précédentes. Mathématiquement, un modèle AR d'ordre p peut être exprimé comme suit :

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

Où X_t est la valeur de la série temporelle au temps t , c est une constante, $\phi_1, \phi_2, \dots, \phi_p$ sont les coefficients AR, $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ sont les valeurs précédentes de la série temporelle et ε_t est le terme d'erreur aléatoire à chaque période.

Dans le contexte du modèle que nous traitons, le modèle AR peut être pertinent pour **capturer les dépendances temporelles dans la série temporelle des prix de l'or**, ce qui peut être utile pour la prévision future des prix.

Pour déterminer le nombre optimal de retards p dans un modèle autorégressif (AR), plusieurs stratégies peuvent être utilisées. Une approche courante est d'utiliser l'analyse de l'autocorrélation partielle (PACF) pour examiner les corrélations entre la série temporelle et ses retards, en excluant les corrélations dues aux retards intermédiaires. Les pics significatifs dans la PACF indiquent le nombre de retards à inclure dans le modèle.

En outre, pour s'assurer que le modèle contient les retards les plus significatifs, une méthode consiste à choisir initialement un grand nombre de retards p pour estimer un modèle AR(p), puis à éliminer progressivement les variables non significatives. À chaque étape, la variable la moins significative est retirée du modèle et le processus est répété jusqu'à ce qu'il ne reste que des variables significatives dans le modèle final.

En se basant sur l'analyse statistique faite auparavant, un modèle AR(3) complet présente un bon compromis pour le modèle de base.

3.1.1 Modèle AR(3) complet

Dependent Variable: RENDEMENT				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 02/29/24 Time: 16:32				
Sample: 1/03/2008 2/02/2024				
Included observations: 4047				
Convergence achieved after 13 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000172	1.256384	0.2090
AR(1)	0.071462	0.007907	9.038051	0.0000
AR(2)	-0.005705	0.009887	-0.576991	0.5640
AR(3)	-0.007534	0.011218	-0.671547	0.5019
SIGMASQ	0.000105	1.02E-06	102.8832	0.0000
R-squared	0.005147	Mean dependent var		0.000216
Adjusted R-squared	0.004162	S.D. dependent var		0.010260
S.E. of regression	0.010239	Akaike info criterion		-6.324076
Sum squared resid	0.423715	Schwarz criterion		-6.316286
Log likelihood	12801.77	Hannan-Quinn criter.		-6.321316
F-statistic	5.227739	Durbin-Watson stat		1.999818
Prob(F-statistic)	0.000337			
Inverted AR Roots	.12-.18i	.12+.18i	-.17	

FIGURE 18 – Modèle AR(3) complet

L'analyse du modèle AR(3) complet révèle que seul le coefficient AR(1) est statistiquement significatif, indiquant une forte autocorrélation avec la valeur précédente. Les autres coefficients AR ne sont pas significatifs. Le modèle présente une faible capacité à expliquer la variance des données, comme en témoigne le faible R-carré ajusté de 0.004162. Bien que l'AIC soit relativement bas à -6.324, suggérant une certaine adéquation du modèle, des tests et des diagnostics supplémentaires pourraient être nécessaires pour confirmer sa pertinence. Les racines inverses AR se situent toutes dans la région de convergence, ce qui est positif pour la stabilité du modèle.

3.1.2 Modèle AR(2) complet

Dependent Variable: RENDEMENT				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 02/29/24 Time: 16:52				
Sample: 1/03/2008 2/02/2024				
Included observations: 4047				
Convergence achieved after 21 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000173	1.251286	0.2109
AR(1)	0.071508	0.007908	9.042073	0.0000
AR(2)	-0.006244	0.009786	-0.638077	0.5235
SIGMASQ	0.000105	1.00E-06	104.3669	0.0000
R-squared	0.005090	Mean dependent var		0.000216
Adjusted R-squared	0.004352	S.D. dependent var		0.010260
S.E. of regression	0.010238	Akaike info criterion		-6.324514
Sum squared resid	0.423739	Schwarz criterion		-6.318281
Log likelihood	12801.65	Hannan-Quinn criter.		-6.322306
F-statistic	6.895072	Durbin-Watson stat		1.999854
Prob(F-statistic)	0.000125			
Inverted AR Roots	.04+.07i	.04-.07i		

FIGURE 19 – Modèle AR(2) complet

Suite à l'analyse du modèle AR(2) complet, les résultats révèlent que le coefficient AR(1) est statistiquement significatif avec un t-statistic élevé de 9.04 et une p-valeur très faible de 0.000, indiquant une forte autocorrélation avec la valeur précédente. Cependant, le coefficient AR(2) ne parvient pas à atteindre un niveau de significativité statistique, avec un t-statistic de -0.64 et une p-valeur de 0.524.

3.1.3 Modèle AR(1) avec constante

Dependent Variable: RENDEMENT
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/29/24 Time: 16:59
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000174	1.244011	0.2136
AR(1)	0.071064	0.007904	8.991086	0.0000
SIGMASQ	0.000105	1.00E-06	104.3637	0.0000

R-squared	0.005051	Mean dependent var	0.000216
Adjusted R-squared	0.004559	S.D. dependent var	0.010260
S.E. of regression	0.010237	Akaike info criterion	-6.324969
Sum squared resid	0.423755	Schwarz criterion	-6.320295
Log likelihood	12801.57	Hannan-Quinn criter.	-6.323313
F-statistic	10.26587	Durbin-Watson stat	1.998872
Prob(F-statistic)	0.000036		

Inverted AR Roots	.07
-------------------	-----

FIGURE 20 – Modèle AR(1) avec constante

Dans l'analyse du modèle AR(1) complet, les résultats révèlent que le coefficient AR(1) est statistiquement significatif, avec un t-statistic élevé de 8.99 et une p-valeur très faible de 0.000, ce qui indique une forte autocorrélation avec la valeur précédente. Cependant, la constante c n'est pas significativement différente de zéro, avec une p-valeur de 0.2136, ce qui suggère qu'elle n'ajoute pas de valeur significative au modèle. Malgré cela, la série temporelle semble bien être modélisée par le modèle AR(1), étant donné la significativité du coefficient AR(1) et sa capacité à capturer la dynamique temporelle des données. La racine inverse associée à la constante indique également que le modèle est stable.

3.1.4 Modèle AR(1) sans constante

Dependent Variable: RENDEMENT
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/29/24 Time: 16:59
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000174	1.244011	0.2136
AR(1)	0.071064	0.007904	8.991086	0.0000
SIGMASQ	0.000105	1.00E-06	104.3637	0.0000

R-squared	0.005051	Mean dependent var	0.000216
Adjusted R-squared	0.004559	S.D. dependent var	0.010260
S.E. of regression	0.010237	Akaike info criterion	-6.324969
Sum squared resid	0.423755	Schwarz criterion	-6.320295
Log likelihood	12801.57	Hannan-Quinn criter.	-6.323313
F-statistic	10.26587	Durbin-Watson stat	1.998872
Prob(F-statistic)	0.000036		

Inverted AR Roots	.07
-------------------	-----

FIGURE 21 – Modèle AR(1) sans constante

Dans l'analyse du modèle AR(1) sans constante, nous constatons que le coefficient AR(1) est toujours statistiquement significatif, avec un t-statistic de 9.05 et une p-valeur

de 0.000. Cela suggère une forte autocorrélation avec la valeur précédente, ce qui est cohérent avec les résultats précédents du modèle AR(1) avec constante.

Comparativement au modèle AR(1) avec constante, le modèle sans constante présente des statistiques de régression similaires en termes de significativité du coefficient AR(1) et de performance globale du modèle, telles que l'AIC et le R-carré ajusté. Cependant, la suppression de la constante n'a pas d'impact significatif sur la significativité du coefficient AR(1) ni sur les autres métriques de performance du modèle.

3.1.5 Comparaison des modèles

Pour comparer les modèles AR estimés, plusieurs métriques sont couramment utilisées. Le R-carré ajusté (R_{adj}^2) est une mesure de la qualité de l'ajustement du modèle qui tient compte du nombre de variables dans le modèle. Il est défini comme :

$$R_{\text{adj}}^2 = 1 - \frac{(n-1)}{(n-p)} \times (1 - R^2)$$

où n est le nombre d'observations et p est le nombre de prédicteurs dans le modèle.

Le critère d'information d'Akaike (AIC) est une mesure de la qualité du modèle qui prend en compte la qualité de l'ajustement et la complexité du modèle. Il est calculé comme suit :

$$\text{AIC} = -2 \times \ln(L) + 2k$$

où L est la fonction de vraisemblance maximisée du modèle et k est le nombre de paramètres estimés dans le modèle.

Le critère de Hannan-Quinn (HQ) est similaire à l'AIC mais pénalise davantage la complexité du modèle, ce qui le rend plus adapté à des échantillons de taille plus petite. Il est calculé comme suit :

$$\text{HQ} = -2 \times \ln(L) + 2k \times \ln(\ln(n))$$

Le critère de Schwarz (SC), également connu sous le nom de critère du nombre minimum de Schwartz (BIC), est une autre mesure de la qualité du modèle qui pénalise la complexité du modèle. Il est défini comme :

$$\text{SC} = -2 \times \ln(L) + k \times \ln(n)$$

Ces métriques peuvent être utilisées pour comparer différents modèles AR estimés. Un modèle avec un R_{adj}^2 plus élevé, un AIC, un HQ ou un SC plus faible est généralement considéré comme préférable car il offre un meilleur ajustement aux données tout en étant plus parcimonieux en termes de complexité du modèle.

Modèle	R^2 ajusté	AIC	Hannan-Quinn	Schwarz
c AR(1) AR(2) AR(3)	0.004162	-6.324076	-6.321316	-6.316286
c AR(1) AR(2)	0.005090	-6.324514	-6.322306	-6.318281
c AR(1)	0.004559	-6.324969	-6.323313	-6.320295
AR(1)	0.004423	-6.325079	-6.323975	-6.321962

TABLE 1 – Tableau comparatif des modèles

Sur la base des métriques présentées, le modèle AR(1) sans constante présente les meilleures performances en termes d'AIC, de Hannan-Quinn et de Schwarz, indiquant qu'il offre le meilleur ajustement aux données tout en étant le plus parcimonieux. Cependant, il est important de noter que le modèle AR(2) complet présente le R^2 ajusté le plus élevé, ce qui suggère qu'il explique une plus grande proportion de la variance des données par rapport aux autres modèles.

En conclusion, nous retenons le modèle AR(1) sans constante comme le choix optimal pour modéliser la série temporelle étudiée, en raison de ses performances globalement supérieures en termes d'AIC, de Hannan-Quinn et de Schwarz, tout en tenant compte de sa capacité à expliquer la variance des données.

3.2 Modélisation MA

Le modèle MA (Moyenne Mobile) est un type de modèle utilisé pour modéliser les séries temporelles qui suppose que la valeur actuelle de la série dépend linéairement des erreurs d'observation passées. Mathématiquement, un modèle MA d'ordre q (MA(q)) est défini comme suit :

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

où Y_t est la valeur observée à un instant t , μ est la moyenne de la série temporelle, ε_t est le terme d'erreur blanc à l'instant t , et $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients de la moyenne mobile correspondant aux retards q précédents.

La validité du modèle MA repose sur plusieurs hypothèses, notamment :

1. **Stationnarité** : La série temporelle doit être stationnaire.
2. **Indépendance des erreurs** : Les erreurs d'observation doivent être indépendantes et identiquement distribuées.
3. **Finiesse de l'ordre** : Le modèle MA suppose que le nombre de termes de moyenne mobile (q) est fini. Cet ordre est souvent déterminé par des méthodes telles que l'analyse de l'autocorrélation partielle (PACF) pour sélectionner le modèle qui minimise l'erreur de prédiction.

La sélection de l'ordre q dans un modèle MA est cruciale pour garantir un ajustement approprié aux données.

3.2.1 Modèle MA(2) complet

Dependent Variable: RENDEMENT
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 03/01/24 Time: 14:48
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 38 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000173	1.249764	0.2115
MA(1)	0.071470	0.007908	9.038121	0.0000
MA(2)	-6.55E-05	0.009937	-0.006595	0.9947
SIGMASQ	0.000105	1.00E-06	104.3509	0.0000

R-squared	0.005083	Mean dependent var	0.000216
Adjusted R-squared	0.004345	S.D. dependent var	0.010260
S.E. of regression	0.010238	Akaike info criterion	-6.324507
Sum squared resid	0.423742	Schwarz criterion	-6.318274
Log likelihood	12801.64	Hannan-Quinn criter.	-6.322299
F-statistic	6.885819	Durbin-Watson stat	1.999759
Prob(F-statistic)	0.000127		

Inverted MA Roots	.00	-.07
-------------------	-----	------

FIGURE 22 – Modèle MA(2) complet

Le modèle MA(2) appliqué à la série des rendements logarithmiques sur le prix de l'or montre que :

Le terme d'autorégression d'ordre 1 ($MA(1)$) est significatif avec un coefficient de 0.071470, indiquant une corrélation positive entre les rendements actuels et les résidus retardés de périodes précédentes.

Le terme d'autorégression d'ordre 2 ($MA(2)$) n'est pas significatif, avec un coefficient proche de zéro -6.55×10^{-5} , et une p-value élevée 0.9947, suggérant qu'il n'ajoute pas de valeur significative au modèle.

Le terme constant (C) n'est pas significatif non plus, avec une p-value de 0.2115.

La racine inversée pour le terme MA(1) est non significative, indiquant que le processus MA est stable.

3.2.2 Modèle MA(1) avec constante

Dependent Variable: RENDEMENT
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 03/01/24 Time: 15:02
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 40 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000173	1.249950	0.2114
MA(1)	0.071475	0.007873	9.078147	0.0000
SIGMASQ	0.000105	1.00E-06	104.3513	0.0000

R-squared	0.005083	Mean dependent var	0.000216
Adjusted R-squared	0.004591	S.D. dependent var	0.010260
S.E. of regression	0.010236	Akaike info criterion	-6.325001
Sum squared resid	0.423742	Schwarz criterion	-6.320327
Log likelihood	12801.64	Hannan-Quinn criter.	-6.323345
F-statistic	10.33127	Durbin-Watson stat	1.999769
Prob(F-statistic)	0.000033		

Inverted MA Roots	-.07
-------------------	------

FIGURE 23 – Modèle MA(1) avec constante

Les résultats indiquent une corrélation significative entre les rendements actuels et les résidus retardés d'une période précédente (MA(1)). Cependant, le terme constant n'est pas significatif. La racine inversée pour le terme MA(1) est non significative, suggérant que le processus MA est stable.

3.2.3 Modèle MA(1) sans constante

Dependent Variable: RENDEMENT
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 03/01/24 Time: 15:07
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 35 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.071834	0.007867	9.130489	0.0000
SIGMASQ	0.000105	1.00E-06	104.5676	0.0000

R-squared	0.004697	Mean dependent var	0.000216
Adjusted R-squared	0.004451	S.D. dependent var	0.010260
S.E. of regression	0.010237	Akaike info criterion	-6.325107
Sum squared resid	0.423906	Schwarz criterion	-6.321991
Log likelihood	12800.85	Hannan-Quinn criter.	-6.324003
Durbin-Watson stat	1.999710		

Inverted MA Roots	-.07
-------------------	------

FIGURE 24 – Modèle MA(1) sans constante

Les mesures de qualité de l'ajustement du modèle, telles que l'AIC, le critère de Schwarz et le critère de Hannan-Quinn, sont également très similaires entre les deux modèles, avec des valeurs légèrement meilleures pour le modèle MA(1) sans constante. Le coefficient de détermination (R^2) et le coefficient de détermination ajusté (R^2 ajusté) sont également très proches entre les deux modèles, indiquant une faible capacité à expliquer la variation des rendements. D'où la nécessité de considérer des modèles plus complets.

Modèle	R^2 ajusté	AIC	Hannan-Quinn	Schwarz
c MA(1) MA(2)	0.004345	-6.324507	-6.322299	-6.318274
c MA(1)	0.004591	-6.325001	-6.323345	-6.320327
MA(1)	0.004451	-6.325107	-6.324003	-6.321991

TABLE 2 – Tableau comparatif des modèles

3.3 Conclusions sur les modèles AR et MA

En tenant compte des caractéristiques de la série de rendement logarithmique sur le prix de l'or, qui peut être volatile et soumise à des influences complexes, il est apparent que les modèles AR et MA simples ne capturent pas efficacement la dynamique sous-jacente des rendements.

Des modèles plus sophistiqués, tels que les modèles **ARMA** ou les modèles **GARCH**, pourraient être explorés pour mieux prendre en compte la volatilité et les tendances temporelles des rendements de l'or.

3.4 Modélisation ARMA

Dans cette section, nous procéderons à l'ajustement d'un modèle ARMA(p,q) en utilisant la méthode "Backward" (afin d'éviter le problème des variables omises). La structure du modèle ARMA(p,q) est définie comme suit :

$$r_t = c + a_1 r_{t-1} + \dots + a_p r_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}$$

où (ε_t) suit une distribution de bruit blanc $\mathcal{N}(0, \sigma^2)$.

Nous entreprendrons maintenant la modélisation de la série en ajustant les paramètres du modèle ARMA afin de capturer les motifs temporels et les structures de dépendance présents dans les rendements de la série.

Comment on choisit les ordres au départ ? L'ACF et le PACF fournissent des indices sur la dépendance temporelle dans la série. Les pics significatifs dans l'ACF à des retards multiples suggèrent des ordres potentiels pour q , tandis que les pics dans le PACF peuvent indiquer des ordres potentiels pour p .

Commençons par considérer des ordres élevés, par exemple ARMA(3,3), tout en notant que des ordres plus élevés peuvent devenir compliqués numériquement.

On donne ici les premiers résultats obtenus pour un ARMA(3, 3) :

Dependent Variable: LOG_REND_HIGH
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/22/24 Time: 21:03
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 50 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000171	1.263327	0.2065
AR(1)	0.892359	0.247963	3.598762	0.0003
AR(2)	-0.631599	0.287522	-2.196700	0.0281
AR(3)	-0.256805	0.227728	-1.127685	0.2595
MA(1)	-0.823071	0.246210	-3.342968	0.0008
MA(2)	0.577474	0.275695	2.094610	0.0363
MA(3)	0.295626	0.221382	1.335368	0.1818
SIGMASQ	0.000104	1.08E-06	96.63918	0.0000
R-squared	0.010436	Mean dependent var	0.000216	
Adjusted R-squared	0.008721	S.D. dependent var	0.010260	
S.E. of regression	0.010215	Akaike info criterion	-6.327902	
Sum squared resid	0.421462	Schwarz criterion	-6.315437	
Log likelihood	12812.51	Hannan-Quinn criter.	-6.323486	
F-statistic	6.085016	Durbin-Watson stat	2.002104	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.58+.78i	.58-.78i	-.27	
Inverted MA Roots	.57+.78i	.57-.78i	-.32	

FIGURE 25 – Modèle ARMA(3,3)

On peut commencer par observer que l'algorithme a bien convergé, et aucune incohérence n'est constatée dans les valeurs des coefficients obtenus. Cependant, il est observé que certains coefficients ont des valeurs négatives, suggérant des alternances de hausses et de baisses, peut-être dues à un cycle sous-jacent. L'examen des inverses des racines indique qu'ils sont assez éloignés de 1, écartant ainsi le problème de non stationnarité. Néanmoins, on remarque que le coefficient AR(3) présente une p-value supérieure à 0.05, ce qui le rend non significatif sur le plan statistique, et c'est également le coefficient ayant la p-value la plus élevée. Par conséquent, ce régresseur est retiré du modèle, conduisant à la mise à jour des résultats comme suit :

Dependent Variable: LOG_REND_HIGH
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/22/24 Time: 21:15
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 53 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000216	0.000177	1.217925	0.2233
AR(1)	0.710948	0.054478	13.05025	0.0000
AR(2)	-0.876707	0.050761	-17.27138	0.0000
MA(1)	-0.640252	0.054753	-11.69347	0.0000
MA(2)	0.835398	0.047892	17.44344	0.0000
MA(3)	0.082911	0.008688	9.543152	0.0000
SIGMASQ	0.000104	1.10E-06	95.04752	0.0000
R-squared	0.008200	Mean dependent var	0.000216	
Adjusted R-squared	0.006727	S.D. dependent var	0.010260	
S.E. of regression	0.010225	Akaike info criterion	-6.326153	
Sum squared resid	0.422415	Schwarz criterion	-6.315246	
Log likelihood	12807.97	Hannan-Quinn criter.	-6.322289	
F-statistic	5.566752	Durbin-Watson stat	2.000496	
Prob(F-statistic)	0.000009			
Inverted AR Roots	.36+.87i	.36-.87i		
Inverted MA Roots	.37+.88i	.37-.88i	-.09	

FIGURE 26 – Modèle ARMA(2,3) avec constante

une p-value de 0.2233 pour le coefficient de c indique que ce coefficient n'est pas statistiquement significatif au seuil de 0.05. La faible magnitude du coefficient c peut également contribuer à l'absence de significativité statistique. Par conséquent, la constante est retirée du modèle, conduisant à la mise à jour des résultats comme suit :

Dependent Variable: LOG_REND_HIGH
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/22/24 Time: 21:16
Sample: 1/03/2008 2/02/2024
Included observations: 4047
Convergence achieved after 91 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.537076	0.021315	25.19765	0.0000
AR(2)	-0.952030	0.021657	-43.96048	0.0000
MA(1)	-0.467441	0.021923	-21.32204	0.0000
MA(2)	0.905846	0.022793	39.74256	0.0000
MA(3)	0.084035	0.008324	10.09611	0.0000
SIGMASQ	0.000104	1.04E-06	100.0283	0.0000
R-squared	0.008100	Mean dependent var	0.000216	
Adjusted R-squared	0.006873	S.D. dependent var	0.010260	
S.E. of regression	0.010225	Akaike info criterion	-6.326536	
Sum squared resid	0.422457	Schwarz criterion	-6.317187	
Log likelihood	12807.75	Hannan-Quinn criter.	-6.323224	
Durbin-Watson stat	1.997452			
Inverted AR Roots	.27+ .94i	.27- .94i		
Inverted MA Roots	.28-.94i	.28+.94i	-.09	

FIGURE 27 – Modèle ARMA(2,3) sans constante

On observe que toutes les variables estimées dans ce modèle sont statistiquement significatives au seuil de 5%, indiquant une robustesse dans la capture des relations entre les variables. Les coefficients, encore une fois négatifs, suggèrent une tendance globale à la baisse dans la série temporelle des rendements logarithmiques des prix élevés de l'or. Le faible R^2 de 0.81% souligne une capacité relativement limitée du modèle à expliquer la variance des observations. Malgré cela, le modèle peut fournir des approximations intéressantes dans certaines perspectives, même si une part importante de la variabilité reste inexpliquée.

Passant à la partie inférieure de l'analyse, la proximité de la statistique de Durbin-Watson à 2 suggère l'absence d'autocorrélation d'ordre 1 dans les résidus du modèle, renforçant ainsi la validité de la structure temporelle du modèle.

Les critères d'information (AIC, BIC, Hannan-Quinn) permettent de comparer les modèles ajustés, tenant compte du compromis entre ajustement et complexité. Une évaluation graphique des racines du polynôme caractéristique peut fournir des informations supplémentaires sur la stabilité du modèle.

En considérant que la série concerne le log de rendement des prix élevés de l'or, ces résultats suggèrent la nécessité de considérer d'autres facteurs ou d'explorer des modèles plus complexes pour améliorer la capacité prédictive du modèle dans un contexte financier particulièrement volatil.

3.5 Modélisation GARCH

Les modèles ARCH (Autoregressive Conditional Heteroskedasticity) et GARCH (Generalized Autoregressive Conditional Heteroskedasticity) sont des modèles couramment utilisés pour modéliser la volatilité conditionnelle des séries temporelles financières. Ils ont été développés pour capturer les propriétés de variance hétéroscédastique conditionnelle observées dans de nombreux marchés financiers, où la volatilité des rendements peut varier dans le temps.

L'intérêt de ces modèles dans le cas de la série des prix de l'or réside dans la nature particulière de cette série. Les prix de l'or présente une faible autocorrélation comme on a pu le voir dans l'étude précédente, ce qui signifie que les variations dans les prix ne sont pas directement prévisibles à partir des prix passés. Cependant, la volatilité des prix de l'or

peut varier considérablement dans le temps, en particulier en réponse à des événements économiques, politiques ou géopolitiques. Les modèles ARCH et GARCH permettent de capturer cette variation temporelle de la volatilité et d'améliorer la modélisation des risques associés aux prix de l'or.

La formulation mathématique générale de ce type de modèles est la suivante. Soit r_t une série temporelle de rendements et ε_t les résidus associés. Le modèle ARCH est défini comme suit :

$$\varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

où z_t est une série de variables aléatoires iid centrées, σ_t^2 est la variance conditionnelle des résidus à l'instant t , et $\alpha_0, \alpha_1, \dots, \alpha_p$ sont les paramètres du modèle. Le terme $\sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$ représente la contribution de p périodes précédentes à la variance conditionnelle actuelle.

Le modèle GARCH est une extension du modèle ARCH qui permet de capturer à la fois les effets d'autorégression et d'hétéroscédasticité conditionnelle. Il est défini par :

$$\varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

où β_1, \dots, β_q sont les paramètres du modèle GARCH.

Les hypothèses d'application des modèles ARCH et GARCH incluent l'iid des résidus, l'absence d'autocorrélation série et la stationnarité. De plus, les paramètres α_i et β_j doivent être positifs pour garantir la positivité de la variance conditionnelle.

3.5.1 Modèle GARCH(2,2)

Dependent Variable: RENDEMENT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 03/17/24 Time: 12:46
Sample (adjusted): 1/03/2008 2/02/2024
Included observations: 4047 after adjustments
Convergence achieved after 35 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*RESID(-2)^2 + C(5)*GARCH(-1)
+ C(6)*GARCH(-2)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000195	0.000137	1.426731	0.1537

Variance Equation

C	1.94E-06	4.70E-07	4.122092	0.0000
RESID(-1)^2	0.102078	0.009337	10.93209	0.0000
RESID(-2)^2	-0.040758	0.016985	-2.399611	0.0164
GARCH(-1)	0.735230	0.172683	4.257692	0.0000
GARCH(-2)	0.183726	0.158352	1.160236	0.2460

R-squared	-0.000004	Mean dependent var	0.000216
Adjusted R-squared	-0.000004	S.D. dependent var	0.010260
S.E. of regression	0.010260	Akaike info criterion	-6.521106
Sum squared resid	0.425909	Schwarz criterion	-6.511758
Log likelihood	13201.46	Hannan-Quinn criter.	-6.517794
Durbin-Watson stat	1.857637		

FIGURE 28 – Résultats du modèle GARCH(2,2)

Le modèle GARCH(2,2) a été utilisé pour modéliser la volatilité conditionnelle des rendements logarithmiques de la série temporelle des prix de l'or. Les résultats du modèle sont résumés ci-dessous.

- Les coefficients associés aux termes $RESID(-1)^2$, $RESID(-2)^2$ et $GARCH(-1)$ sont significatifs et positifs, suggérant une réaction positive de la volatilité conditionnelle aux chocs précédents et une persistance temporelle dans la volatilité.
- Le coefficient associé au terme $RESID(-2)^2$ est significatif et négatif, ce qui est moins courant. Cela peut indiquer une certaine forme de réversion à la moyenne dans la volatilité conditionnelle.
- Les métriques d'évaluation du modèle, telles que l'AIC et le critère de Schwarz, indiquent une bonne adéquation du modèle aux données, avec des valeurs négatives et relativement basses.
- Le coefficient de détermination R^2 est négatif, ce qui est typique pour les modèles de volatilité. Il mesure l'ajustement global du modèle aux données, en d'autres termes, la qualité de l'équation de la moyenne conditionnelle.
- Le coefficient GARCH(-2) est non significatif avec une p-value élevée. Donc, dans l'optique d'une démarche backward, on peut commencer par enlever ce coefficient.

3.5.2 Modèle GARCH(1,2)

Dependent Variable: RENDEMENT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 03/17/24 Time: 12:49
Sample (adjusted): 1/03/2008 2/02/2024
Included observations: 4047 after adjustments
Failure to improve likelihood (non-zero gradients) after 0 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*RESID(-2)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000216	0.000308	0.700200	0.4838

Variance Equation				
C	7.25E-05	1.43E-05	5.078375	0.0000
RESID(-1)^2	0.133333	0.017367	7.677475	0.0000
RESID(-2)^2	0.044444	0.043039	1.032665	0.3018
GARCH(-1)	0.533333	0.090554	5.889656	0.0000

R-squared	0.000000	Mean dependent var	0.000216
Adjusted R-squared	0.000000	S.D. dependent var	0.010260
S.E. of regression	0.010260	Akaike info criterion	-6.238982
Sum squared resid	0.425907	Schwarz criterion	-6.231192
Log likelihood	12629.58	Hannan-Quinn criter.	-6.236222
Durbin-Watson stat	1.857645		

FIGURE 29 – Résultats du modèle GARCH(1,2)

- Les coefficients associés aux termes $RESID(-1)^2$ et $GARCH(-1)$ sont significatifs et positifs, suggérant une réaction positive de la volatilité conditionnelle aux chocs précédents et une persistance temporelle dans la volatilité.
- Les métriques d'évaluation du modèle, telles que l'AIC et le critère de Schwarz, indiquent une bonne adéquation du modèle aux données, avec des valeurs négatives et relativement basses.

3.5.3 GARCH(1,1)

Dependent Variable: RENDEMENT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 03/17/24 Time: 12:52
Sample (adjusted): 1/03/2008 2/02/2024
Included observations: 4047 after adjustments
Convergence achieved after 25 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000176	0.000134	1.317018	0.1878

Variance Equation				
C	2.46E-06	3.46E-07	7.091737	0.0000
RESID(-1)^2	0.071537	0.004672	15.31191	0.0000
GARCH(-1)	0.904190	0.007366	122.7534	0.0000

R-squared	-0.000015	Mean dependent var	0.000216
Adjusted R-squared	-0.000015	S.D. dependent var	0.010260
S.E. of regression	0.010260	Akaike info criterion	-6.519980
Sum squared resid	0.425913	Schwarz criterion	-6.513748
Log likelihood	13197.18	Hannan-Quinn criter.	-6.517772
Durbin-Watson stat	1.857617		

FIGURE 30 – Résultats du modèle GARCH(1,1)

L'analyse du modèle GARCH(1,1) révèle des résultats significatifs et cohérents avec les attentes. Les coefficients associés à la variation conditionnelle des erreurs au carré au décalage précédent ainsi qu'à la variation conditionnelle de GARCH au décalage précédent sont tous deux positifs et statistiquement significatifs. Cela indique que les erreurs passées ainsi que la volatilité conditionnelle précédente contribuent de manière importante à la volatilité actuelle des rendements logarithmiques sur le prix de l'or.

En outre, les métriques de performance du modèle montrent un bon ajustement aux données observées. L'AIC et le critère de Schwarz sont tous deux négatifs, indiquant une bonne adéquation du modèle aux données. Ces résultats suggèrent que le modèle GARCH(1,1) est capable de capturer efficacement la volatilité conditionnelle de la série temporelle.

Dans l'ensemble, ces résultats confirment la pertinence du modèle GARCH(1,1) pour modéliser la volatilité des rendements logarithmiques sur le prix de l'or.

3.5.4 GARCH(1,1)-GED

La modélisation de la volatilité des séries financières à l'aide de modèles GARCH est souvent réalisée en supposant une loi conditionnelle gaussienne pour les résidus. Cependant, cette hypothèse peut être insuffisante pour capturer la complexité des données financières, notamment en ce qui concerne les queues épaisses et l'asymétrie des distributions des rendements. Afin de remédier à cette limitation, il est possible d'introduire une loi conditionnelle non gaussienne, telle que la loi de Student ou la loi généralisée de Laplace, dans le modèle GARCH.

Dependent Variable: RENDEMENT
Method: ML ARCH - Generalized error distribution (GED) (BFGS / Marquardt steps)
Date: 03/17/24 Time: 14:05
Sample (adjusted): 1/03/2008 2/02/2024
Included observations: 4047 after adjustments
Convergence achieved after 37 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000114	0.000116	0.990199	0.3221

Variance Equation

C	1.28E-06	3.60E-07	3.542858	0.0004
RESID(-1)^2	0.045979	0.006226	7.385616	0.0000
GARCH(-1)	0.940459	0.008529	110.2722	0.0000

GED PARAMETER	1.198723	0.032285	37.12938	0.0000
---------------	----------	----------	----------	--------

R-squared	-0.000098	Mean dependent var	0.000216
Adjusted R-squared	-0.000098	S.D. dependent var	0.010260
S.E. of regression	0.010260	Akaike info criterion	-6.598766
Sum squared resid	0.425949	Schwarz criterion	-6.590976
Log likelihood	13357.60	Hannan-Quinn criter.	-6.596006
Durbin-Watson stat	1.857463		

FIGURE 31 – Résultats du modèle GARCH(1,1)-GED

Le modèle GARCH(1,1) avec distribution GED présente des résultats significatifs pour les coefficients associés à la variance conditionnelle et à la distribution des erreurs. Le coefficient estimé pour le paramètre de la distribution GED est statistiquement significatif avec une p -value de 0, indiquant son effet significatif sur la distribution des erreurs. Cela suggère que la distribution GED capture mieux les queues épaisses et les queues minces de la distribution des rendements, ce qui peut être crucial pour modéliser les événements extrêmes, souvent observés dans les séries financières telles que les prix de l'or.

En ce qui concerne les métriques de performance, bien que les métriques telles que l'AIC, le critère de Schwarz et la log-vraisemblance montrent une légère amélioration par rapport au modèle GARCH(1,1) standard, les valeurs restent assez proches et ne suggèrent pas de différence significative dans l'ajustement du modèle. Cela indique que l'ajout de la distribution GED n'a pas affecté de manière significative les paramètres du modèle GARCH(1,1) mais a permis une meilleure capture des caractéristiques de la distribution des erreurs.

3.5.5 Restriction IGARCH

L'introduction de la restriction IGARCH (Integrated GARCH) dans un modèle GARCH peut être justifiée dans le but de capturer les variations persistantes de la volatilité observées dans les séries financières. Contrairement au modèle GARCH standard, où la volatilité conditionnelle est strictement stationnaire, le modèle IGARCH permet à la volatilité conditionnelle d'être intégrée d'ordre un, ce qui signifie qu'elle peut présenter une tendance à long terme. Cette flexibilité permet de mieux capturer les comportements de persistance observés dans la volatilité des séries temporelles financières.

Dependent Variable: RENDEMENT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 03/17/24 Time: 14:11
Sample (adjusted): 1/03/2008 2/02/2024
Included observations: 4047 after adjustments
Convergence achieved after 25 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2)*RESID(-1)^2 + (1 - C(2))*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000142	0.000114	1.241931	0.2143
Variance Equation				
RESID(-1)^2	0.042032	0.001939	21.68191	0.0000
GARCH(-1)	0.957968	0.001939	494.1566	0.0000
R-squared	-0.000052	Mean dependent var		0.000216
Adjusted R-squared	-0.000052	S.D. dependent var		0.010260
S.E. of regression	0.010260	Akaike info criterion		-6.501273
Sum squared resid	0.425929	Schwarz criterion		-6.498157
Log likelihood	13157.33	Hannan-Quinn criter.		-6.500169
Durbin-Watson stat	1.857548			

FIGURE 32 – Résultats du modèle IGARCH(1,1)

En analysant les résultats pour le modèle IGARCH(1,1), nous observons que le coefficient estimé pour la variation conditionnelle des erreurs au carré au décalage précédent est positif avec une p -value de 0, tandis que le coefficient associé à la variation conditionnelle de GARCH au décalage précédent est proche de 1 avec une p -value de 0.

Les métriques de performance telles que l'AIC et le critère de Schwarz montrent une légère détérioration par rapport au modèle GARCH(1,1) standard, indiquant une moins bonne adéquation du modèle IGARCH aux données observées.

3.5.6 IGARCH(1,1)- GED

Date: 03/17/24 Time: 14:16
Sample (adjusted): 1/03/2008 2/02/2024
Included observations: 4047 after adjustments
Failure to improve likelihood (singular hessian) after 4 iterations
Coefficient covariance computed using outer product of gradients
WARNING: Singular covariance - coefficients are not unique
Presample variance: backcast (parameter = 0.7)
GARCH = C(2)*RESID(-1)^2 + (1 - C(2))*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000216	NA	NA	NA
Variance Equation				
RESID(-1)^2	0.000105	NA	NA	NA
GARCH(-1)	0.999895	NA	NA	NA
GED PARAMETER	1.999832	NA	NA	NA
Mean dependent var	0.000216	S.D. dependent var		0.010260

FIGURE 33 – Résultats du modèle IGARCH(1,1)-GED

Les résultats de ce modèle GARCH avec distribution des erreurs généralisée (GED) indiquent un avertissement important concernant une hessienne singulière, ce qui signifie que la matrice de covariance des coefficients n'est pas unique. Cela peut être dû à plusieurs raisons, telles que la spécification inadéquate du modèle ou des problèmes numériques lors de l'estimation des paramètres.

3.5.7 ARMA(2,2)-GARCH(1,1) GED

Dependent Variable: RENDEMENT
Method: ML ARCH - Generalized error distribution (GED) (BFGS / Marquardt steps)
Date: 03/17/24 Time: 14:22
Sample (adjusted): 1/07/2008 2/02/2024
Included observations: 4045 after adjustments
Convergence achieved after 72 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/03/2008 1/04/2008
Presample variance: backcast (parameter = 0.7)
GARCH = C(6) + C(7)*RESID(-1)^2 + C(8)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000109	0.000120	0.907516	0.3641
AR(1)	-0.167838	0.395046	-0.424857	0.6709
AR(2)	-0.434717	0.159988	-2.717186	0.0066
MA(1)	0.225274	0.395225	0.569990	0.5687
MA(2)	0.437323	0.163300	2.678025	0.0074

Variance Equation				
C	1.37E-06	3.79E-07	3.627228	0.0003
RESID(-1)^2	0.047821	0.006470	7.391247	0.0000
GARCH(-1)	0.937420	0.008932	104.9520	0.0000

GED PARAMETER	1.208136	0.032794	36.84001	0.0000
---------------	----------	----------	----------	--------

R-squared	0.005061	Mean dependent var	0.000214
Adjusted R-squared	0.004076	S.D. dependent var	0.010262
S.E. of regression	0.010241	Akaike info criterion	-6.601431
Sum squared resid	0.423676	Schwarz criterion	-6.587402
Log likelihood	13360.39	Hannan-Quinn criter.	-6.596461
Durbin-Watson stat	1.970887		

Inverted AR Roots	-.08+.65i	-.08-.65i
Inverted MA Roots	-.11-.65i	-.11+.65i

FIGURE 34 – Résultats du modèle ARMA(2,2)-GARCH(1,1)-GED

Dans l'étude précédente, on a négligé le modèle ARMA(p,q) compte tenu des observations précédentes : la série de rendement logarithmique du prix de l'or semble avoir le comportement d'un bruit. Pour s'assurer de cette hypothèse, on introduit le modèle ARMA(2,2)-GARCH(1,1) GED plutôt que de se limiter uniquement au modèle GARCH(1,1) GED. Ceci repose sur l'idée que le comportement de la série temporelle pourrait être mieux capturé en tenant compte à la fois de la structure autorégressive (ARMA) et de la volatilité conditionnelle (GARCH). Cette décision est motivée par la nécessité d'explorer un modèle plus complet qui prend en compte à la fois les dépendances autorégressives et les effets de volatilité.

Analysons les résultats du modèle ARMA(2,2)-GARCH(1,1) GED :

- Les coefficients AR(2) et MA(2) indiquent les relations linéaires entre les valeurs passées de la série temporelle et la valeur actuelle, ainsi que les résidus passés. Les

coefficients significatifs suggèrent des relations statistiquement significatives qui peuvent être exploitées pour modéliser le comportement de la série.

- Les coefficients du modèle GARCH déterminent la dynamique de la volatilité conditionnelle de la série temporelle. Les coefficients non significatifs pourraient remettre en question la pertinence du modèle GARCH.
- Le paramètre de forme de la distribution GED est estimé à 1.208136, ce qui caractérise le degré d'asymétrie et de leptokurtosis de la distribution des erreurs. Une valeur différente de 2 (cas de la distribution normale) suggère des queues de distribution plus épaisses, ce qui est souvent observé dans les données financières.
- Les critères d'information tels que l'AIC et le BIC sont utilisés pour évaluer la qualité du modèle. Un modèle avec un AIC ou un BIC plus bas est préféré car il offre un meilleur compromis entre l'ajustement du modèle et sa complexité.

En conclusion, l'introduction du modèle ARMA(2,2)-GARCH(1,1) GED permet d'explorer plus en profondeur la structure de la série temporelle en capturant à la fois les dépendances autorégressives et les effets de volatilité. Les résultats obtenus fournissent des informations précieuses sur les relations statistiques et la dynamique de la volatilité dans la série temporelle des rendements, ce qui peut aider à une meilleure modélisation et à des prévisions plus précises.

Même si ARMA seul n'a pas donné des résultats satisfaisants, l'ajout de la composante GARCH peut améliorer la modélisation en prenant en compte l'hétéroscédasticité conditionnelle.

3.5.8 AR(1)-GARCH(1,1)-GED

Dependent Variable: RENDEMENT
Method: ML ARCH - Generalized error distribution (GED) (BFGS / Marquardt steps)
Date: 03/17/24 Time: 14:58
Sample (adjusted): 1/04/2008 2/02/2024
Included observations: 4046 after adjustments
Convergence achieved after 37 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000118	0.000123	0.962964	0.3356
AR(1)	0.057868	0.014230	4.066598	0.0000

Variance Equation

C	1.36E-06	3.75E-07	3.619709	0.0003
RESID(-1)^2	0.047599	0.006432	7.400522	0.0000
GARCH(-1)	0.937824	0.008863	105.8157	0.0000

GED PARAMETER	1.209068	0.032685	36.99107	0.0000
---------------	----------	----------	----------	--------

R-squared	0.004801	Mean dependent var	0.000214
Adjusted R-squared	0.004555	S.D. dependent var	0.010260
S.E. of regression	0.010237	Akaike info criterion	-6.602235
Sum squared resid	0.423787	Schwarz criterion	-6.592884
Log likelihood	13362.32	Hannan-Quinn criter.	-6.598922
Durbin-Watson stat	1.972620		

Inverted AR Roots	.06
-------------------	-----

FIGURE 35 – Résultats du modèle AR(1)-GARCH(1,1)-GED

L'analyse des résultats du modèle AR(1)-GARCH(1,1) GED montre des coefficients significatifs pour les termes AR(1) et GARCH(1), ainsi qu'un paramètre de forme de distribution GED significatif.

En termes de critères d'information AIC et BIC sont relativement bas, ce qui suggère que le modèle AR(1)-GARCH(1,1) GED offre un bon ajustement aux données tout en étant relativement simple.

En conclusion, le modèle AR(1)-GARCH(1,1) GED semble être le plus adapté pour modéliser la série temporelle des rendements. Il capture à la fois les dépendances autorégressives et les effets de volatilité conditionnelle, ce qui en fait un choix robuste pour l'analyse et la prévision des rendements financiers.

4 Modèles multivariés

4.1 Test de cointégration

La cointégration se réfère à la relation à long terme entre des séries temporelles qui sont stationnaires lorsque combinées linéairement. Plus formellement, si deux séries temporelles Y_t et X_t sont cointégrées, alors il existe une combinaison linéaire de ces séries, souvent appelée vecteur de cointégration, qui est stationnaire. La cointégration est souvent représentée par le modèle de régression suivant :

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

où ε_t est un terme d'erreur stationnaire. Les hypothèses sous-jacentes à la cointégration comprennent la stationnarité des séries individuelles, la non-corrélation des erreurs, et l'absence de racines unitaires dans le vecteur de cointégration.

Pour analyser la série de prix high de l'or, il est important de considérer plusieurs aspects. Tout d'abord, l'évolution des prix high de l'or au fil du temps peut être étudiée en relation avec d'autres variables telles que le volume des échanges, les taux de change, les indices boursiers ou les indicateurs macroéconomiques. L'analyse des relations à long terme entre le prix high de l'or et ces variables peut être réalisée en utilisant des techniques de cointégration. La cointégration entre le prix high de l'or et le volume des échanges pourrait indiquer l'existence d'une relation à long terme entre ces deux séries, ce qui pourrait avoir des implications sur la liquidité du marché de l'or et son impact sur les prix. En outre, il est important d'examiner les résidus du modèle de cointégration pour évaluer l'efficacité du modèle et l'absence d'effets de mémoire à court terme.

4.1.1 Test de cointégration entre le prix high et le volume

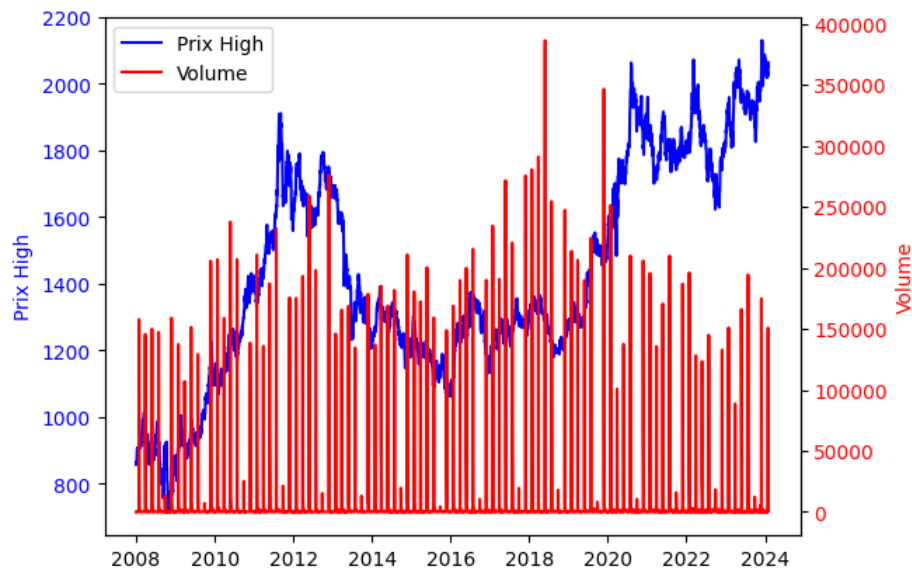


FIGURE 36 – Prix high vs Volume

Les observations réalisées sur les variations du volume des échanges et des prix high de l'or suggèrent des dynamiques complexes sur le marché financier. En particulier, des périodes de synchronisation entre les variations du volume des échanges et des prix high, telles que celles observées pendant le printemps arabe en 2011, peuvent indiquer une possible cointégration entre les deux séries. Cette cointégration impliquerait une relation à long terme entre le volume des échanges et les prix high, malgré les fluctuations à court terme causées par des événements spécifiques ou d'autres facteurs temporaires. Il est important de noter que la cointégration ne suggère pas nécessairement une relation de cause à effet directe entre les séries, car d'autres variables ou facteurs non observés peuvent

également influencer les deux. Pour confirmer la présence de cointégration, des tests statistiques spécifiques, tels que le **test de Johansen**, peuvent être utilisés pour évaluer la force de la relation à long terme entre les deux séries.

Test de Cointégration de Johansen Le test de Johansen repose sur deux hypothèses principales :

- **Hypothèse nulle (H_0)** : L'hypothèse nulle du test de Johansen est que le nombre de vecteurs de cointégration est égal ou inférieur à un nombre spécifié. Formellement, cela s'exprime comme :

$$H_0 : r \leq k$$

où r représente le nombre de vecteurs de cointégration et k est un nombre spécifié.

- **Hypothèse alternative (H_1)** : L'hypothèse alternative est que le nombre de vecteurs de cointégration est supérieur à un. Formellement, cela s'exprime comme :

$$H_1 : r > k$$

où r représente le nombre de vecteurs de cointégration et k est un nombre spécifié.

Le test de Johansen utilise deux métriques principales pour évaluer ces hypothèses :

- **Statistique de trace** : La statistique de trace est la somme des valeurs propres des matrices de cointégration pour différents nombres de vecteurs de cointégration. Elle est utilisée pour tester l'hypothèse nulle de cointégration.
- **Max-eigenvalue** : Cette métrique examine la plus grande valeur propre des matrices de cointégration pour différents nombres de vecteurs de cointégration. Elle est également utilisée pour tester l'hypothèse nulle de cointégration.

Le rejet de l'hypothèse nulle indique la présence de cointégration dans le système de séries temporelles, ce qui suggère une relation à long terme entre les séries malgré les variations à court terme.

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.** Critical Value
None *	0.170165	758.1672	25.87211	0.0000
At most 1	0.000996	4.030832	12.51798	0.7376
Trace test indicates 1 cointegrating equation(s) at the 0.05 level				
* denotes rejection of the hypothesis at the 0.05 level				
**Mackinnon-Haug-Michelis (1999) p-values				
Unrestricted Cointegration Rank Test (Max-eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.** Critical Value
None *	0.170165	754.1364	19.38704	0.0000
At most 1	0.000996	4.030832	12.51798	0.7376
Max-eigenvalue test indicates 1 cointegrating equation(s) at the 0.05 level				
* denotes rejection of the hypothesis at the 0.05 level				
**Mackinnon-Haug-Michelis (1999) p-values				

FIGURE 37 – Résultats du test de trace et de valeurs propres de Johansen

Les résultats des tests de rang de cointégration suggèrent qu'il existe une relation à long terme entre les séries temporelles du volume des échanges et des prix high de l'or, bien que les conclusions puissent être nuancées. La statistique de trace et la max-eigenvalue indiquent toutes deux la présence potentielle d'une cointégration, avec des valeurs de 758.1672 et 754.1364 respectivement, dépassant largement les seuils critiques aux niveaux de signification de 0.05.

Cependant, la p-valeur associée à l'hypothèse "At most 1" est de 0.7376, ce qui dépasse largement le seuil de 0.05. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle il n'y a au plus qu'un vecteur de cointégration dans le système de séries temporelles.

Normalized Coefficients

1 Cointegrating Equation		
Log-Likelihood: -63471.97		
Normalized cointegrating coefficients (standard error in parentheses)		
HIGH	VOLUME	@TREND
1.000000	0.910073 (0.03164)	0.007493 (0.39805)
Adjustment coefficients (standard error in parentheses)		
D(HIGH)	4.56E-05 (1.6E-05)	
D(VOLUME)	-0.932788 (0.03250)	

FIGURE 38 – Résultats du test de Johansen - Coefficients normalisés

Les résultats des tests indiquent une relation à long terme entre les séries temporelles du prix high (*HIGH*) et du volume des échanges (*VOLUME*) de l'or, ainsi que la tendance (*@TREND*), comme représenté par la matrice de cointégration β suivante :

$$\beta = \begin{bmatrix} 1.0 & 0.910073 & 0.007493 \end{bmatrix}$$

Cela signifie que chaque unité de variation dans le prix high est associée à une variation de 1.0 dans le prix high lui-même, de 0.910073 dans le volume des échanges, et de 0.007493 dans la tendance.

Les coefficients d'ajustement associés aux variations des variables sont représentés par la matrice α suivante :

$$\alpha = \begin{bmatrix} 4.56 \times 10^{-5} \\ -0.932788 \end{bmatrix}$$

Ces coefficients indiquent la vitesse à laquelle les variables s'ajustent en réponse à des chocs ou des variations dans une période donnée. Par exemple, une variation positive de 1 unité dans le prix high se traduira par une augmentation de 4.56×10^{-5} unités dans le prix high lui-même, tandis qu'une variation négative de 1 unité dans le volume des échanges entraînera une diminution de 0.932788 unités dans le volume des échanges. On note que ces coefficients sont bien opposés.

4.2 Modèle VAR

Un modèle VAR (Vector Autoregression) est un modèle économétrique utilisé pour analyser les relations dynamiques entre plusieurs séries temporelles. Dans un modèle VAR, chaque variable endogène est modélisée en fonction de ses retards (lags) ainsi que des retards des autres variables endogènes.

Les variations des taux d'intérêt et le cours de l'or suivent une corrélation inverse, ou négative. Lorsque les banques centrales décident d'augmenter les taux d'intérêt, le prix de l'or diminue et inversement. Pourquoi ? Parce que la hausse et la baisse des taux d'intérêt rendent les produits d'investissement plus ou moins attractifs.

Pour ce faire, on se propose d'étudier le modèle VAR les deux variables, le prix élevé de l'or (Gold High Price) et le taux d'intérêt à 10 ans des bons du Trésor américain (CBOE Interest Rate 10-Year T-Note), avec des lags de 1 et 2, la spécification peut être exprimée comme suit :

Pour la variable Gold High Price (GHP) :

$$\begin{aligned} GHP_t = & \alpha_{10} + \beta_{11}GHP_{t-1} + \beta_{12}GHP_{t-2} \\ & + \gamma_{11}IR_{t-1} + \gamma_{12}IR_{t-2} + \epsilon_{1t} \end{aligned}$$

Pour la variable CBOE Interest Rate 10-Year T-Note (IR) :

$$\begin{aligned} IR_t = & \alpha_{20} + \beta_{21}GHP_{t-1} + \beta_{22}GHP_{t-2} \\ & + \gamma_{21}IR_{t-1} + \gamma_{22}IR_{t-2} + \epsilon_{2t} \end{aligned}$$

Où :

- GHP_t est le prix élevé de l'or à l'instant t .
- IR_t est le taux d'intérêt à 10 ans des bons du Trésor américain à l'instant t .
- α_{10}, α_{20} sont les termes constants.
- $\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}$ sont les coefficients de régression des lags des variables.
- $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$ sont les coefficients de régression croisée entre les variables.
- $\epsilon_{1t}, \epsilon_{2t}$ sont les termes d'erreur.

l'analyse initiale indique que les séries ne sont pas stationnaires même après normalisation.



FIGURE 39 – Les deux séries avant stationnarisation

Dans notre cas, nous avons choisi d'utiliser la première différence pour stationnariser deux séries.

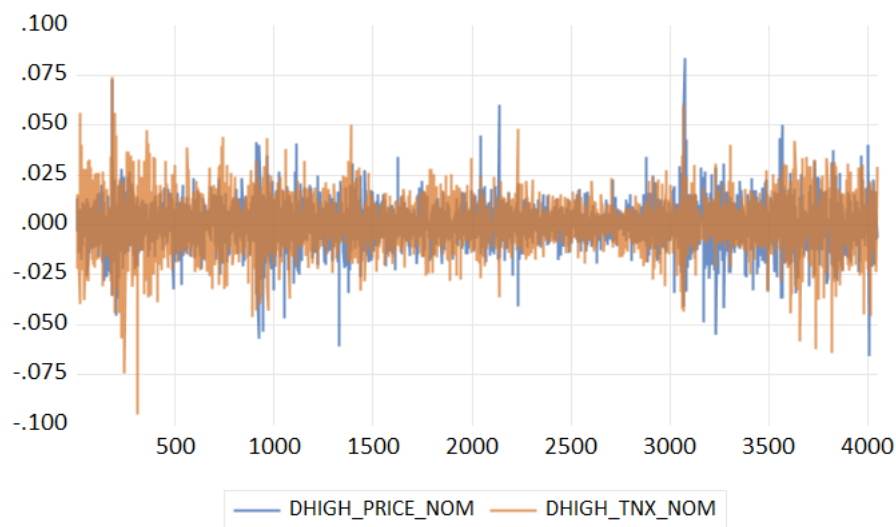


FIGURE 40 – Les deux séries après stationnarisation utilisées pour le VAR

Vector Autoregression Estimates

Vector Autoregression Estimates		
Date: 03/13/24 Time: 15:32		
Sample (adjusted): 4 4046		
Included observations: 4043 after adjustments		
Standard errors in () & t-statistics in []		
	DHIGH_PRI	DHIGH_TNX_NOM
DHIGH_PRICE_NOM(-1)	0.074789 (0.01592) [4.69657]	-0.052746 (0.01929) [-2.73444]
DHIGH_PRICE_NOM(-2)	-0.018084 (0.01587) [-1.13915]	0.042117 (0.01923) [2.19024]
DHIGH_TNX_NOM(-1)	-0.070427 (0.01314) [-5.35960]	0.062918 (0.01592) [3.95278]
DHIGH_TNX_NOM(-2)	-0.006498 (0.01318) [-0.49301]	-0.030655 (0.01597) [-1.92007]
C	0.000196 (0.00016) [1.26102]	5.86E-06 (0.00019) [0.03112]
R-squared	0.015020	0.008251
Adj. R-squared	0.014045	0.007269
Sum sq. resids	0.394478	0.578835
S.E. equation	0.009884	0.011973
F-statistic	15.39438	8.398629
Log likelihood	12931.65	12156.50
Akaike AIC	-6.394583	-6.011130
Schwarz SC	-6.386786	-6.003333
Mean dependent	0.000208	3.55E-06
S.D. dependent	0.009954	0.012017
Determinant resid covariance (dof adj.)	1.37E-08	
Determinant resid covariance	1.36E-08	
Log likelihood	25136.45	
Akaike information criterion	-12.42960	
Schwarz criterion	-12.41401	
Number of coefficients	10	

FIGURE 41 – Résultats du test de Johansen - Coefficients normalisés

Analyses des Résultats

- La signification des coefficients entre la Gold High Price et son propre retard avec lag de 1, indique une forte autocorrélation. Cela signifie que le prix élevé de l'or d'un jour est un bon prédicteur de son prix le jour suivant. Cette tendance peut être attribuée à la nature du marché de l'or, qui est souvent considéré comme un investissement refuge en période d'incertitude économique ou de volatilité sur les marchés financiers. Les investisseurs se tournent vers l'or comme une valeur sûre, ce qui entraîne une continuité ou une augmentation de sa valeur au fil du temps.
- La signification des coefficients entre la Gold High Price et les taux d'intérêt, avec un lag de 1, suggère une interaction importante entre les prix de l'or et les politiques monétaires reflétées par les taux d'intérêt. Un taux d'intérêt élevé peut augmenter le coût d'opportunité de détenir de l'or, ce qui pourrait réduire la demande d'or et,

par conséquent, son prix. Inversement, des taux d'intérêt bas peuvent encourager l'investissement dans l'or.

- Le fait que les taux d'intérêt soient significativement liés à la Gold High Price avec des retards de 1 et 2 indique que les variations des taux d'intérêt influencent le prix de l'or, mais avec un certain délai. Cela peut refléter le temps nécessaire pour que les changements de politique monétaire aient un impact perceptible sur l'économie réelle et, par extension, sur les marchés des métaux précieux. Cela peut aussi refléter les anticipations du marché vis-à-vis des futures politiques monétaires et leur effet sur le prix de l'or.

Ces dynamiques offrent des perspectives importantes pour les investisseurs et les décideurs politiques, en soulignant l'importance de surveiller les politiques monétaires et leur impact sur les investissements en or. Pour les investisseurs, ces informations peuvent aider à informer les stratégies d'allocation d'actifs, surtout dans le contexte de diversification de portefeuille et de gestion des risques.

Par contre, le R-squared très faible impliquant la Gold High Price et les taux d'intérêt à 10 ans des bons du Trésor américain (CBOE Interest Rate 10 Year T No) indique que les modèles n'expliquent qu'une très petite partie de la variabilité des séries temporelles analysées.

La présence de coefficients significatifs dans un contexte de R^2 faible signifie que bien que certaines relations dynamiques soient présentes et détectables, elles ne suffisent pas à elles seules pour expliquer de manière exhaustive la variabilité observée dans les données. Cela souligne l'importance d'explorer d'autres variables, d'améliorer la spécification du modèle, ou de considérer d'autres approches modélisatrices pour obtenir une compréhension plus complète et précise des dynamiques à l'œuvre.

On complète cette étude avec le test de causalité de Granger qui donne le résultat suivant.

VAR Granger Causality/Block Exogeneity Wald Tests			
Date: 03/13/24 Time: 15:33			
Sample: 1 4046			
Included observations: 4043			
Dependent variable: DHIGH_PRICE_NOM			
Excluded	Chi-sq	df	Prob.
DHIGH_TNX_NOM	29.28683	2	0.0000
All	29.28683	2	0.0000
Dependent variable: DHIGH_TNX_NOM			
Excluded	Chi-sq	df	Prob.
DHIGH_PRICE_NOM	11.50868	2	0.0032
All	11.50868	2	0.0032

FIGURE 42 – Résultats du test de causalité de Granger

Toutes les valeurs de p-valeur obtenues à partir du test de causalité de Granger sont

inférieures à 5%, cela suggère que les variables testées ont une relation de causalité significative entre elles. Dans notre cas, cela signifie que le prix élevé de l'or (Gold High Price) et le taux d'intérêt à 10 ans des bons du Trésor américain (CBOE Interest Rate 10-Year T-Note) sont susceptibles d'avoir une influence causale l'un sur l'autre.

Cependant, il est important de noter que même si les p-valeurs sont significatives, cela ne garantit pas que la relation est forte ou substantielle. La faible valeur du coefficient de détermination (R-squared), de l'ordre de 0.01, suggère que le modèle explique très peu de la variance observée dans les données. Cela indique que le modèle ne capture peut-être pas complètement les relations complexes entre les variables ou qu'il manque des facteurs importants.

5 Éventuelles Améliorations - Prédictions

5.1 EWMA Modèle à Moyenne Mobile Exponentielle Pondérée

Le modèle EWMA (*Exponentially Weighted Moving Average*), est une méthode de prévision et de suivi des données temporelles qui **accorde plus de poids aux observations récentes tout en atténuant l'impact des observations plus anciennes**. Ce modèle est pertinent pour le prix de l'or en raison de la nature dynamique et volatile de ce marché, où les événements récents peuvent avoir un impact plus significatif sur les prix que les données historiques plus anciennes.

Mathématiquement, le modèle EWMA est défini par la formule suivante :

$$S_t = \alpha \times Y_t + (1 - \alpha) \times S_{t-1}$$

Où : - S_t est la valeur EWMA à l'instant t , - Y_t est la valeur observée à l'instant t , - α est le facteur de lissage (compris entre 0 et 1), qui détermine le poids relatif des observations récentes par rapport aux observations historiques.

Les hypothèses derrière le modèle EWMA comprennent l'idée que les données sont stationnaires et que les erreurs de mesure sont aléatoires et indépendantes. De plus, le choix du facteur de lissage α influence la sensibilité du modèle aux variations des données, avec des valeurs plus élevées de α donnant plus de poids aux observations récentes et vice versa.

En incorporant le modèle EWMA dans l'analyse des prix de l'or, nous pouvons mieux capturer les tendances à court terme et les variations rapides du marché, ce qui peut être précieux pour la prise de décision en matière d'investissement et de gestion des risques. Cependant, il est important de reconnaître que le modèle EWMA repose sur des hypothèses spécifiques sur la nature des données et peut ne pas être approprié dans tous les contextes, notamment lorsque les données présentent des structures de dépendance temporelle plus complexes.



FIGURE 43 – Modèle EWMA avec différentes valeurs de span

Le paramètre "span" dans les modèles EWMA définit la période sur laquelle la pondération décroît exponentiellement.

Un span plus élevé est souvent privilégié pour lisser les données et capturer les tendances à long terme, ce qui est utile lorsque la série temporelle est sujette à de nombreuses fluctuations aléatoires ou bruits. D'autre part, un span plus faible est préféré pour détecter les variations rapides et les changements soudains dans les données, ce qui peut être important pour les marchés volatils ou pour suivre les tendances à court terme.

Voici les résultats des métriques de performance pour les modèles EWMA avec différentes valeurs de span :

Span	RMSE	MAE	R^2
3	8.44	6.07	0.9993
6	15.00	10.98	0.9978
12	23.13	17.23	0.9948

Les résultats montrent que le modèle EWMA avec un span de 3 donne les meilleures performances en termes de RMSE, MAE et R^2 , indiquant une meilleure adéquation entre les prévisions du modèle et les données réelles. Cela suggère que pour cette série temporelle, un span plus faible est plus approprié pour capturer les variations rapides et les tendances à court terme.

6 Conclusion

L'étude de la série temporelle du prix de l'or a fourni des perspectives importantes sur les tendances et les motifs sous-jacents dans les mouvements des prix. Nos analyses ont utilisé des techniques variées, notamment la régression linéaire, l'analyse de corrélation,

ainsi que les modèles ARMA et GARCH, pour comprendre la complexité du marché de l'or.

Cependant, il est crucial de reconnaître que les modèles traditionnels peuvent parfois être insuffisants pour capturer pleinement la dynamique des prix de l'or, en raison de la nature complexe et souvent non linéaire des données financières. Dans ce contexte, l'utilisation de modèles tels que les réseaux de neurones LSTM (Long Short-Term Memory) pourrait être explorée comme une approche alternative.

Les réseaux de neurones LSTM sont connus pour leur capacité à capturer les dépendances à long terme dans les séries temporelles, ce qui en fait des candidats prometteurs pour la modélisation et la prédiction des prix de l'or. Leur capacité à apprendre des motifs complexes et non linéaires à partir des données pourrait potentiellement améliorer les performances de prédiction par rapport aux modèles traditionnels.

En conclusion, cette étude a jeté les bases pour une compréhension approfondie des prix de l'or, mais elle ne représente qu'une étape dans un processus continu d'analyse et de recherche. L'exploration de nouvelles approches, telles que les réseaux de neurones LSTM, est nécessaire pour relever les défis persistants de la modélisation et de la prédiction des prix de l'or dans un environnement financier en constante évolution.