

Modélisation de volatilité financière : les processus ARCH et GARCH

Thomas Gerault
Nathan Raitière

Février 2023

Table des matières

1	Introduction	2
1.1	Séries financières	2
1.1.1	Qu'est ce que la volatilité?	2
2	Les biens-fondés des modèles ARCH et GARCH	3
2.1	Définitions	3
2.1.1	Bruit blanc	3
2.1.2	Processus stationnaire	3
2.2	Le modèle ARCH	4
2.2.1	Stationnarité	4
2.2.2	Variance conditionnelle constante ou non	5
2.2.3	Vers le processus <i>GARCH</i>	7
2.3	Le processus GARCH	8
2.3.1	Stationnarité	9
2.3.2	Variance conditionnelle non constante	9
2.4	Conclusion, comparaison et validation	10
2.4.1	Validation <i>ARCH</i> et <i>GARCH</i>	11
2.4.2	Prévisions	11
3	Applications aux jeux de données réels	13
3.1	Construction et modélisation pour la série du <i>CAC40</i>	13
3.1.1	Détermination des ordres p et q	14
3.1.2	Estimations et diagnostics de différents modèles <i>GARCH</i> (p, q)	17
3.1.3	Prévisions	18
3.2	Construction et modélisation pour la série <i>SP500</i>	21
3.2.1	Détermination des ordres p et q	22
3.2.2	Estimation du modèle <i>GARCH</i> (p, q)	23
3.2.3	Prévisions	23
4	Les processus : ARMA-GARCH	27
4.1	Définition	27
4.2	Simulation	28
5	Conclusion	29

Chapitre 1

Introduction

Parmi la grande famille des séries chronologiques, il existe une branche particulière qui décrit les marchés financiers, taux de change, actions et autres. Celles-ci possèdent des caractéristiques qui les rendent difficiles à modéliser par des processus classiques tels que les *ARMA*. Cette difficulté est due aux nombreuses propriétés qui caractérisent les séries financières. C'est en 1982 que Robert Engle développe le premier modèle Auto-Regressif Conditionnellement Hétéroscédastique (*ARCH*) qui permet de reproduire artificiellement le comportement de ces séries. Ce modèle représentant une avancée majeure dans la modélisation des séries financières contient rapidement des faiblesses et est vite généralisé par le processus *GARCH* (Auto-Regressif Conditionnellement Hétéroscédastique Généralisé). Depuis, cette classe de modèles qui permet de modéliser le cours d'un actif financier est devenue incontournable.

1.1 Séries financières

Le cours d'un actif financier à la date t , noté X_t , peut représenter un prix, un point d'indice ou encore une valeur de taux. Le modèle ARCH ne permet pas de modéliser le cours de l'actif mais son rendement ou sa volatilité. Le rendement d'un actif est défini comme étant le rapport des valeurs d'un actif à deux dates consécutives. Afin d'atténuer la variabilité, nous affectons au rendement le logarithme, ce qui donne le rendement $\epsilon_t = \log(\frac{X_t}{X_{t-1}})$. Définissons maintenant ce qu'est la volatilité.

1.1.1 Qu'est ce que la volatilité ?

Nous appelons volatilité d'un actif, l'amplitude générée par les variations du rendement. En effet cette volatilité est un indicateur de la dispersion de ses rendements par rapport à sa moyenne. On voit donc ici naître la notion de variance. La volatilité de l'actif sera d'autant plus forte que les cours des marchés sont instables. Elle permet donc de mesurer le risque : plus la volatilité d'un produit est grande, plus fort est le risque associé à ce produit. En effet, si le prix d'un produit varie beaucoup, il n'est pas forcément sûr de pouvoir le revendre avec profit ou même sans perte.

Par la suite, nous étudierons des séries financières dont la variance conditionnelle est non constante. La problématique de ce projet va donc être d'étudier et de montrer en quoi ces processus permettent de meilleures modélisations des séries financières. Nous montrerons dans un premier temps les biens fondés de ces modèles en introduisant leur construction et étudierons leur stationnarité. Nous montrerons ensuite en quoi les modèles simples comme les ARMA ne sont pas appropriés puis, nous illustrerons tout cela sur plusieurs jeux de données comportant de fortes volatilités afin de comprendre dans quelle mesure ces modèles à hétéroscédasticité conditionnelle sont capables de reproduire ces phénomènes.

Chapitre 2

Les biens-fondés des modèles ARCH et GARCH

2.1 Définitions

Pour commencer, définissons les termes de bruit blanc et de processus stationnaire.

2.1.1 Bruit blanc

Soit T un espace des temps. On appelle bruit blanc fort, tout processus $(X_t)_{t \in T}$ de variables indépendantes et identiquement distribuées.

Un bruit blanc (faible) correspond à tout processus $(X_t)_{t \in T}$ de variable non corrélées, de même espérance m et de même variance σ^2 finie. On a alors :

$$\forall t \in T, \quad E[X_t] = m \quad \text{et} \quad \forall (t, s) \in T^2, \quad Cov(X_t, X_s) = \begin{cases} \sigma^2 & \text{si } t = s \\ 0 & \text{sinon} \end{cases}$$

Par la suite, il sera précisé quand nous parlerons de bruit blanc fort. Autrement, la notion de bruit blanc sera considérée au sens faible. Maintenant que nous avons défini ce qu'est un bruit blanc, regardons maintenant ce qu'est un processus stationnaire.

2.1.2 Processus stationnaire

Une série chronologique est dite stationnaire (au second ordre) si sa moyenne et sa covariance sont invariantes pour tout décalage de temps. Autrement dit :

- $\forall t \in T, \quad E[X_t] = m$
- $\forall (t, s) \in T^2$ et pour tout décalage temporel $k \in T, \quad Cov(X_t, X_s) = \Gamma(t, s) = \Gamma(t + k, s + k)$

Afin d'étudier la stationnarité de modèles à hétéroscédasticité conditionnelle, définissons ce qu'est un modèle ARCH.

2.2 Le modèle ARCH

Un modèle X_t est dit autorégressif conditionnellement hétéroscédastique d'ordre p que l'on note $ARCH(p)$, s'il admet l'écriture suivante :

$$\begin{cases} \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 \end{cases}$$

avec ω une constante strictement positive, α_i constantes positives et $\eta_t \sim \mathcal{N}(0, 1)$ un bruit blanc fort. De cette façon, σ_t^2 est strictement positif et définit la variance conditionnelle du processus (Cela sera démontré par la suite).

Étudions maintenant la stationnarité de ces processus. Nous verrons par la suite que la stationnarité est atteinte sous certaines conditions.

2.2.1 Stationnarité

Nous considérerons que nos modèles sont centrés en 0. Montrons qu'un processus $ARCH(1)$ est stationnaire :

$$\begin{aligned} E[\epsilon_t] &= E[\sigma_t]E[\eta_t] \text{ par indépendance} \\ &= 0 \end{aligned}$$

Regardons également la variance :

$$\begin{aligned} Var(\epsilon_t) &= E[\epsilon_t^2] \\ &= E[\sigma_t^2] E[\eta_t^2] \\ &= \omega + \alpha_1 E[\epsilon_{t-1}^2] \\ &= \omega + \alpha_1 (\omega + \alpha_1 E[\epsilon_{t-2}^2]) \\ &= \dots \\ &= \omega + \omega \sum_{k \geq 1} (\alpha_1^k) + \prod_{k \geq 1} (\alpha_1^{k+1} E[\epsilon_{t-k+1}^2]) \\ &= \frac{\omega}{1 - \alpha_1} \quad \text{par somme d'une suite géométrique si } 0 < \alpha_1 < 1 \end{aligned}$$

et on en déduit alors la covariance :

$$\begin{aligned} E[\epsilon_t \epsilon_{t+k}] &= E[\sigma_t \eta_t \sigma_{t+k} \eta_{t+k}] \\ &= E[\sigma_t \sigma_{t+k}] E[\eta_t \eta_{t+k}] \text{ par indépendance} \\ &= E[\eta_t \eta_{t+k}] E[\sqrt{\omega + \alpha_1 X_{t-1}^2} \sqrt{\omega + \alpha_1 X_{t-1+k}^2}] \\ &= \begin{cases} \frac{\omega}{1 - \alpha_1} & \text{si } k = 0 \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

Avec $k = 0$, on obtient la variance du processus. Ainsi, la covariance ne dépend pas du temps. On a donc bien un processus stationnaire sous la condition que le coefficient $\alpha_1 \in [0, 1[$. On peut donc généraliser l'expression de la variance :

$$Var(\epsilon_t) = \frac{\omega}{1 - \sum_{k=1}^p \alpha_k}$$

Un critère de stationnarité est donc le fait d'avoir $\sum_{k=1}^p \alpha_k < 1$.

Maintenant que la stationnarité a été établie, nous allons regarder la variance conditionnelle des processus *AR* et *ARCH*.

2.2.2 Variance conditionnelle constante ou non

Variance conditionnelle : *AR*(1)

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus suivant un *AR*(1). On a alors :

$$\forall t \in \mathbb{Z} \quad X_t = \theta X_{t-1} + \epsilon_t \text{ avec } \epsilon_t \sim \mathcal{N}(0, 1) \text{ un bruit blanc}$$

Afin de calculer la variance conditionnelle, nous avons tout d'abord besoin de calculer l'espérance conditionnelle :

$$\begin{aligned} E[X_t | X_{t-1}] &= E[\theta X_{t-1} + \epsilon_t | X_{t-1}] \\ &= E[\theta X_{t-1} | X_{t-1}] + E[\epsilon_t | X_{t-1}] \quad \text{par linéarité de l'espérance} \\ &= \theta X_{t-1} + E[\epsilon_t] \quad \text{par indépendance entre } \epsilon_t \text{ et } X_{t-1} \\ &= \theta X_{t-1} \end{aligned}$$

Regardons maintenant la variance conditionnelle définie par l'expression suivante :

$$\begin{aligned} Var(X_t | X_{t-1}) &= E[(X_t - E[X_t | X_{t-1}])^2 | X_{t-1}] \\ &= E[(\theta X_{t-1} + \epsilon_t - \theta X_{t-1})^2 | X_{t-1}] \\ &= E[\epsilon_t^2 | X_{t-1}] \\ &= E[\epsilon_t^2] \\ &= \sigma^2 \end{aligned}$$

Pour un processus *AR*(1), on obtient donc une variance conditionnelle constante. Regardons ce qu'il en est pour un processus *ARCH* dont on suppose que cette même variance conditionnelle n'est pas constante.

Variance conditionnelle : *ARCH*(1)

Prenons un modèle *ARCH*(1) dont on rappelle l'écriture :

$$\epsilon_t = \sigma_t \eta_t$$

avec $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$ et $\eta_t \sim \mathcal{N}(0, 1)$ un bruit blanc fort.

On veut montrer que $Var(X_t | \epsilon_{t-1}, \dots, \epsilon_{t-k})$ n'est pas constante. On notera par la suite $I_{t-k} = \epsilon_{t-1}, \dots, \epsilon_{t-k}$ comme étant l'ensemble de l'information au passé pour tout $k \in \mathbb{N}^*$. On sait que :

$$\begin{aligned}
Var(\epsilon_t \mid I_{t-k}) &= E[\epsilon_t^2 \mid I_{t-k}] - (E[\epsilon_t \mid I_{t-k}])^2 \\
&= E[\eta_t^2 (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \mid I_{t-k}] - E[\eta_t \sigma_t \mid I_{t-k}]^2 \\
&= E[\eta_t^2] E[(\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \mid I_{t-k}] - E[\eta_t]^2 E[\sigma_t \mid I_{t-k}]^2 \text{ par indépendance de } \eta_t^2 \\
&= E[(\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \mid I_{t-k}] - 0 \\
&= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \\
&= \sigma_t^2
\end{aligned}$$

On observe que cette expression dépend du temps. La variance conditionnelle n'est donc pas constante dans le cas d'un $ARCH(1)$.

Regardons graphiquement la différence de variance entre un processus classique $ARMA(1,1)$ et un processus $ARCH(1)$ obtenue à l'aide de Python :

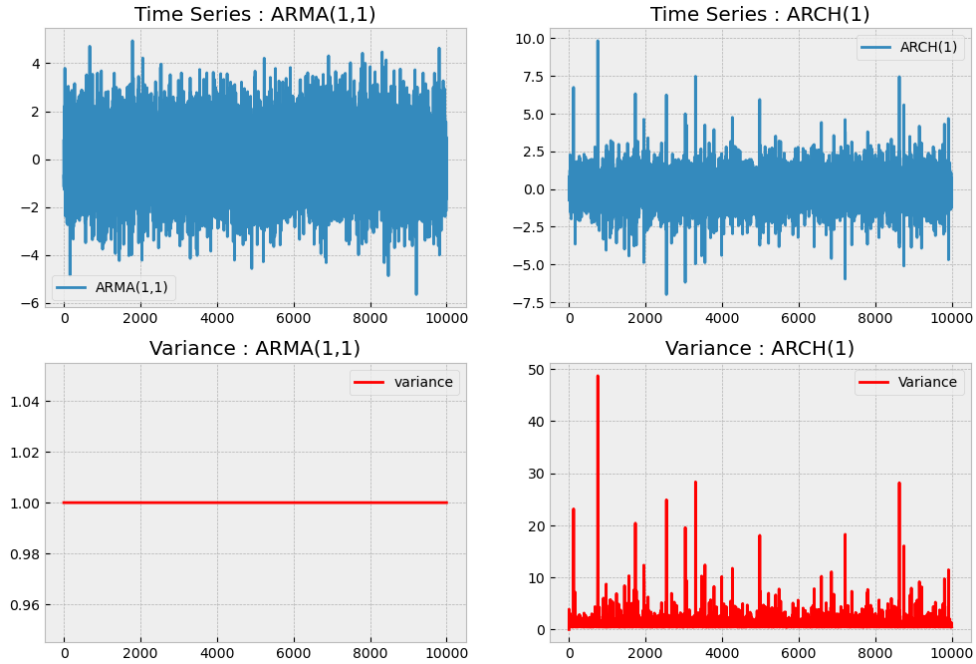


FIGURE 2.1 – Série et volatilité simulées d'un $ARMA$ et d'un $ARCH$

Comme nous pouvons le voir, le processus hétéroscédastique possède des pics bien plus variables que le processus $ARMA(1,1)$. La taille d'échelle sur l'axe y met bien en valeur cette différence. Ainsi, nous pouvons déjà déduire visuellement, en ne regardant que la série, qu'un processus classique telle que les $ARMA$, ne sont plus intéressants pour modéliser des séries à forte volatilité.

Maintenant, pour mettre en évidence le souci de variance, nous avons représenté cette variance comme fonction du temps (graphe en rouge). En comparant les graphes des deux processus, nous pouvons remarquer que la variance conditionnelle dans le cas d'un processus $ARMA$ est constante contrairement à un processus $ARCH$.

2.2.3 Vers le processus *GARCH*

Afin de mieux modéliser la variance conditionnelle, il existe une autre forme de processus appelé *GARCH*. Regardons donc à l'aide d'une simulation les différences entre ces deux processus :

On se propose de simuler 10000 données suivant un *ARCH*(1) de paramètres $\omega = 1$ et $\alpha_1 = 0.9$. On obtient alors les résultats suivants :

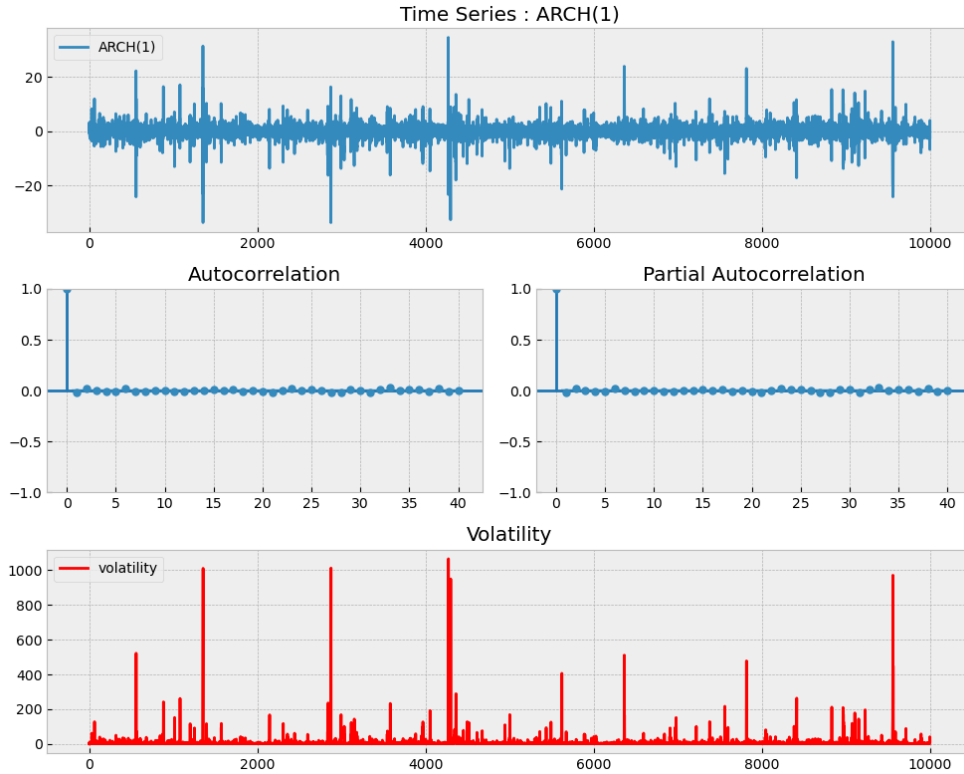


FIGURE 2.2 – Simulation d'un processus *ARCH*(1)

On peut remarquer au vu de l'autocorrélation et de l'expression de sa covariance que ce processus se comporte comme un bruit blanc. Il ne l'est pas pour autant. En effet, la distribution des rendements d'un actif n'est pas gaussienne. De plus, les tests statistiques rejettent l'hypothèse de normalité car les distributions sont à queues lourdes. On peut faire la représentation d'un QQplot pour s'en convaincre :

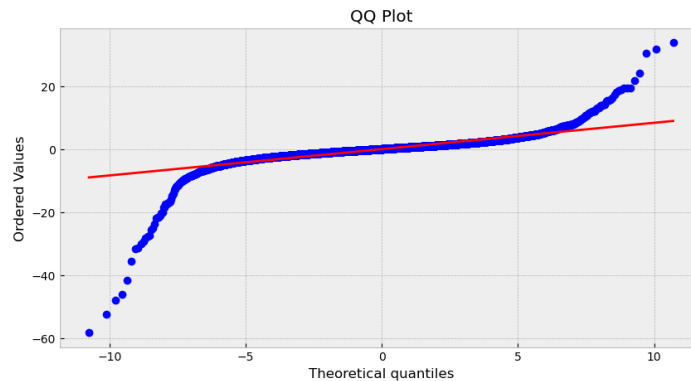


FIGURE 2.3 – QQplot des données simulées d'un *ARCH*(1)

Malheureusement, le processus *ARCH* possède des lacunes. En effet, celui ci ne prend pas en compte la variance conditionnelle passée dans la modélisation. Si l'on regarde l'expression d'un processus *ARCH* on peut voir que la volatilité n'est modélisée uniquement que par les rendements passés $(\alpha_i \epsilon_{t-i})$. C'est pourquoi, en 1986, Bollerslev généralise le modèle *ARCH* (Generalized ARCH) en ajoutant le terme de la variance conditionnelle passée dans la modélisation de la volatilité. La première équation du modèle *ARCH* reste inchangée, en revanche la deuxième équation devient :

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Comme le modèle ARCH, le paramètre ω est strictement positif et les paramètres α_i , β_j sont positifs. Le modèle *GARCH* reproduit les mêmes caractéristiques que le modèle *ARCH* et aura tendance à accentuer le regroupement des extrêmes, comme nous pouvons le voir sur le graphe ci dessous. Les paramètres utilisés pour la simulation sont $\omega = 1$, $\alpha_1 = 0.3$ et $\beta_1 = 0.6$.

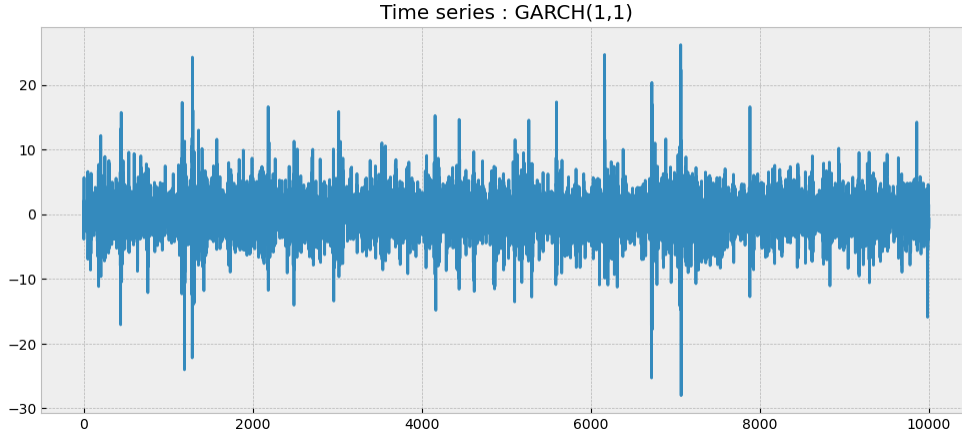


FIGURE 2.4 – Simulation *GARCH*(1,1)

On peut voir que les extrêmes se regroupent contrairement au processus *ARCH* où nous avons simplement des pics de grandes amplitudes un peu éparpillés. Définissons maintenant un processus *GARCH*.

2.3 Le processus GARCH

Comme introduit précédemment, on définit un processus *GARCH*(p, q) comme suit :
On dit qu'un processus ϵ_t est un processus *GARCH*(p, q) s'il vérifie l'écriture :

$$\begin{cases} \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \end{cases}$$

avec ω une constante strictement positive, α_i et β_j des constantes positives et $\eta_t \sim \mathcal{N}(0, 1)$ un bruit blanc fort.

2.3.1 Stationnarité

Comme pour le modèle *ARCH*, nous allons étudier sous quelles conditions nous avons la stationnarité. Nous considérerons par la suite que nos modèles sont centrés en 0. Nous ferons la démonstration pour un processus *GARCH*(1,1). Pour l'espérance, nous avons le même résultat que pour le modèle *ARCH* :

$$E[\epsilon_t] = 0$$

Pour la variance, nous avons cette fois ci :

$$\begin{aligned} \text{Var}(\epsilon_t) &= E[\epsilon_t^2] \\ &= E[\sigma_t^2] E[\eta_t^2] \\ &= \omega + \alpha_1 E[\epsilon_{t-1}^2] + \beta_1 E[\sigma_{t-1}^2] \\ &= \omega + (\alpha_1 + \beta_1) E[\epsilon_{t-1}^2] \\ &= \omega + (\alpha_1 + \beta_1) (\omega + (\alpha_1 + \beta_1) E[\epsilon_{t-2}^2]) \\ &= \dots \\ &= \omega + \omega \sum_{k \geq 1} ((\alpha_1 + \beta_1)^k) + \prod_{k \geq 1} (\alpha_1 + \beta_1)^{k+1} E[\epsilon_{t-k+1}^2] \\ &= \frac{\omega}{1 - (\alpha_1 + \beta_1)} \quad \text{par somme d'une suite géométrique si } 0 < \alpha_1 + \beta_1 < 1 \end{aligned}$$

On en déduit la covariance de la même manière que pour le modèle *ARCH* :

$$\text{Cov}(\epsilon_t, \epsilon_{t+k}) = \begin{cases} \frac{\omega}{1 - (\alpha_1 + \beta_1)} & \text{si } k = 0 \\ 0 & \text{sinon} \end{cases}$$

Avec $k = 0$, on obtient la variance du processus. Ainsi, on peut voir que la covariance ne dépend pas du temps. Cela implique qu'on a donc bien un processus stationnaire sous la condition tout de même que le coefficient $(\alpha_1 + \beta_1) \in [0, 1[$. Ainsi, on peut généraliser l'expression de la variance :

$$\text{Var}(\epsilon_t) = \frac{\omega}{1 - \sum_{k=1}^{\max(p,q)} (\alpha_k + \beta_k)}$$

Nous pouvons donc avoir comme critère de stationnarité $\sum_{k=1}^p (\alpha_k + \beta_k) < 1$. Maintenant que la stationnarité a été établie, nous allons montrer que la variance conditionnelle dans un processus *GARCH* n'est pas constante.

2.3.2 Variance conditionnelle non constante

La variance conditionnelle d'un processus classique a été établie dans la section I.1.2

Variance conditionnelle : *GARCH*(1,1)

Prenons un modèle *GARCH*(1,1) dont on rappelle l'écriture :

$$\epsilon_t = \sigma_t \eta_t$$

avec $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$ et $\eta_t \sim \mathcal{N}(0, 1)$ un bruit blanc fort.

On veut montrer que $Var(\epsilon_t | \epsilon_{t-1}, \dots, \epsilon_{t-k})$ n'est pas constante. On notera par la suite $I_{t-k} = \epsilon_{t-1}, \dots, \epsilon_{t-k}$ comme étant l'ensemble de l'information au passé pour tout $k \in \mathbb{N}^*$. On sait que :

$$\begin{aligned}
Var(\epsilon_t | I_{t-k}) &= E[\epsilon_t^2 | I_{t-k}] - (E[\epsilon_t | I_{t-k}])^2 \\
&= E[\eta_t^2 (\omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2) | I_{t-k}] - E[\eta_t \sigma_t | I_{t-k}]^2 \\
&= E[\eta_t^2] E[(\omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2) | I_{t-k}] - E[\eta_t]^2 E[\sigma_t | I_{t-k}]^2 \text{ par indépendance de } \eta_t^2 \\
&= E[(\omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2) | I_{t-k}] - 0 \\
&= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \sigma_t^2
\end{aligned}$$

On peut voir que cette expression dépend du temps ce qui implique que la variance conditionnelle n'est donc pas constante dans ce cas ci.

2.4 Conclusion, comparaison et validation

On se propose pour terminer cette section consacrée à montrer les différences de variance d'effectuer deux superpositions de processus : une superposition d'un $AR(1)$ avec un $ARCH(1)$ et une superposition d'un $ARMA(1,1)$ avec un $GARCH(1,1)$.

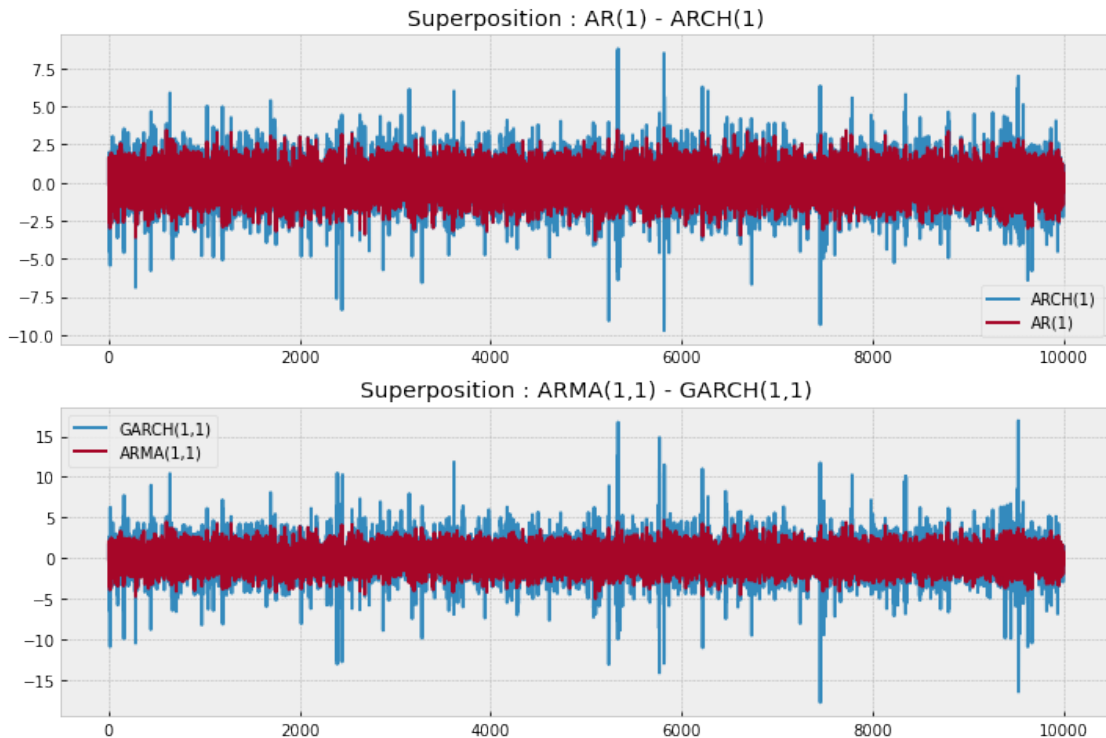


FIGURE 2.5 – Comparaison $ARMA$ - $GARCH$

On peut constater visuellement que les modélisations $ARMA(1,1)$ et $AR(1)$ possèdent des pics bien plus petit que les processus $GARCH(1,1)$ et $ARCH(1)$. Comparons maintenant la qualité de nos simulations $ARCH$ et $GARCH$.

2.4.1 Validation *ARCH* et *GARCH*

Pour montrer la validité de nos simulations, nous avons appliqué la fonction *arch model* du package *arch* de python sur les données que nous avons simulé. Nous avons compilé un *GARCH*(1,1) avec pour paramètre $\omega = 1$, $\alpha_1 = 0.5$ et $\beta_1 = 0.3$ et un *ARCH*(1) avec pour paramètre $\omega = 1$ et $\alpha_1 = 0.5$. En voici les estimations :

Mean Model					
	coef	std err	t	P> t	95.0% Conf. Int.
mu	5.0601e-03	1.083e-02	0.467	0.640	[-1.617e-02,2.629e-02]
Volatility Model					
	coef	std err	t	P> t	95.0% Conf. Int.
omega	0.9845	2.134e-02	46.142	0.000	[0.943, 1.026]
alpha[1]	0.4806	1.885e-02	25.497	2.140e-143	[0.444, 0.518]

FIGURE 2.6 – Paramètres estimés d'un ARCH(1)

Mean Model					
	coef	std err	t	P> t	95.0% Conf. Int.
mu	-0.0154	1.504e-02	-1.024	0.306	[-4.488e-02,1.407e-02]
Volatility Model					
	coef	std err	t	P> t	95.0% Conf. Int.
omega	0.9745	5.165e-02	18.868	2.077e-79	[0.873, 1.076]
alpha[1]	0.5104	2.042e-02	24.989	7.970e-138	[0.470, 0.550]
beta[1]	0.2993	2.020e-02	14.811	1.246e-49	[0.260, 0.339]

FIGURE 2.7 – Paramètres estimés d'un GARCH(1,1)

Les paramètres comme la moyenne μ , ω , α_1 et β_1 ont été estimés par maximum de vraisemblance. Nous obtenons pour nos deux modèles les estimations ci-dessus ainsi que leurs intervalles de confiance respectifs. Nous pouvons constater que nos paramètres sont relativement proches de leur vraie valeur et sont de plus significatif au seuil de 0.1. Nous pouvons alors dire que notre simulation représente bien les processus *ARCH* et *GARCH*.

2.4.2 Prévisions

Un des objectifs très importants de ces modèles est de faire des prévisions de volatilité. Ainsi, nous obtiendrions une estimation de la variabilité des rendements d'un actif. On s'intéresse dans un premier temps aux prévisions du modèle *ARCH*(p).

Prévisions du modèle *ARCH*(p)

Rappelons l'écriture d'un *ARCH*(p) :

$$\begin{cases} \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 \end{cases}$$

Avec cette dernière expression de σ_t^2 , on peut prévoir les valeurs futures de la variances conditionnelles. Ainsi, si on prend un exemple avec un $ARCH(2)$, on aura :

Soit la série ϵ_t ayant comme valeur à l'instant t 0.83 et -0.52 à l'instant $t - 1$. On a également les coefficients $\omega = 0.01$ et $\alpha_1 = 0.10$ ainsi que $\alpha_2 = 0.4$ comme paramètres du modèle $ARCH(2)$. Le calcul pour déterminer la prévision à l'instant $t + 1$ va être :

$$\begin{aligned}\sigma_{t+1}^2 &= \omega + \alpha_1 \epsilon_t^2 + \alpha_2 \epsilon_{t-1}^2 \\ &= 0.01 + 0.1 * 0.83^2 + 0.4 * (-0.52)^2 \\ &= 0.187\end{aligned}$$

On aura donc une volatilité prédite à 0.187 à l'instant $t + 1$. Si on généralise pour un $ARCH(p)$, on obtient l'expression :

$$\sigma_{t+1}^2 = \omega + \alpha_1 \epsilon_t^2 + \dots + \alpha_p \epsilon_{t-p}^2$$

Regardons maintenant comment s'effectuent les prévisions dans la cas d'un modèle $GARCH(p, q)$

Prévisions du modèle $GARCH(p, q)$

Rappelons tout d'abord l'expression du modèle :

$$\begin{cases} \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \end{cases}$$

avec $\sum_{i=1}^{max(p,q)} \alpha_i + \beta_i < 1$. Comme avec le modèle $ARCH(p)$, on utilise la dernière expression de σ_t^2 pour obtenir des prévisions de volatilité. Si on prend un exemple d'un $GARCH(2, 1)$ avec les mêmes paramètres que pour notre $ARCH(2)$ avec en plus le coefficient $\beta_1 = 0.3$ et la volatilité $\sigma_t^2 = 0.6$ à l'instant t . Le calcul pour déterminer la prévision à l'instant $t + 1$ va être :

$$\begin{aligned}\sigma_{t+1}^2 &= \omega + \alpha_1 \epsilon_t^2 + \alpha_2 \epsilon_{t-1}^2 + \beta_1 \sigma_t^2 \\ &= 0.01 + 0.1 * 0.83^2 + 0.4 * (-0.52)^2 + 0.3 * 0.6 \\ &= 0.295\end{aligned}$$

On aura donc une volatilité prédite à 0.295 à l'instant $t + 1$ en modélisant avec un $GARCH(2, 1)$. Si on généralise, on obtient l'expression :

$$\sigma_{t+1}^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Cette section étant terminée, nous allons maintenant regarder quelques exemples d'applications à des séries financières en particulier sur la série du *CAC40* et du *SP500*. Nous chercherons quels modèles modélisera au mieux ces séries et nous ferons des prévisions que nous calculerons grâce à la commande *forecast* de python dans un premier temps, puis nous les calculerons à la main pour comparer.

Chapitre 3

Applications aux jeux de données réels

Cette section va avoir pour but d'illustrer plusieurs tentatives de modélisation de série financières existantes tel que le *CAC40* ou le *SP500* par des modèles à hétéroscédasticité conditionnelle. On tentera plusieurs modélisations de ces séries financières par des *ARCH* et des *GARCH*.

3.1 Construction et modélisation pour la série du *CAC40*

Regardons tout d'abord la série du *CAC40* en données journalières. En voici une représentation graphique de la série :



FIGURE 3.1 – Prix de clôture de l'indice boursier journalier du CAC 40

Comme les modèles à hétéroscédasticité conditionnelle travaillent sur les rendements des séries financières, nous allons appliquer un changement à cette série. Nous allons appliquer le log-retour journalier qui se traduit par la série :

$$X_t = \log\left(\frac{S_t}{S_{t-1}}\right)$$

où S_t est la série financière originale au temps t . Voici à quoi ressemble la nouvelle série X_t :

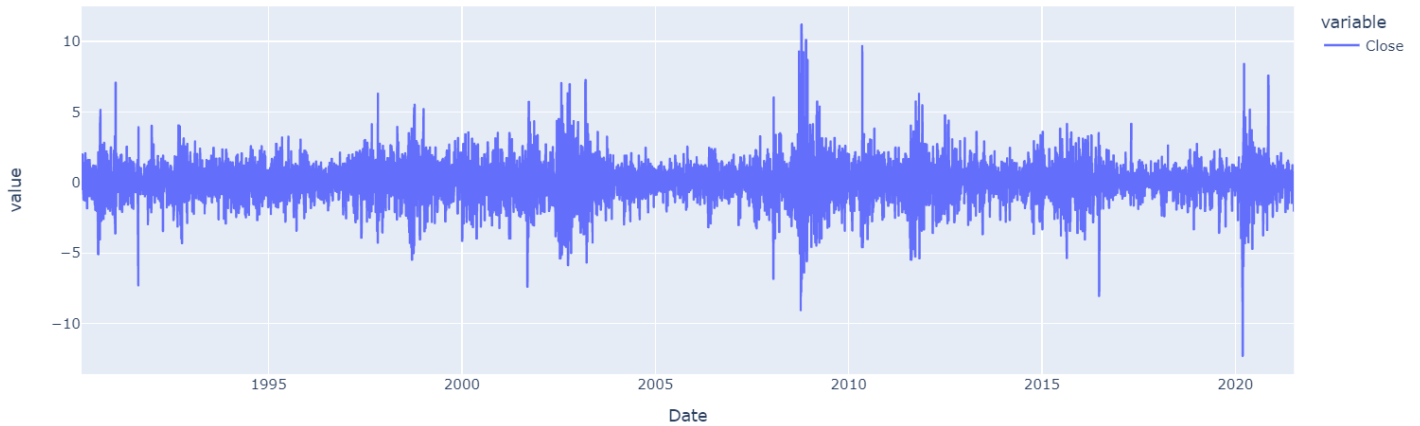


FIGURE 3.2 – Prix de clôture de l’indice boursier journalier du CAC 40 appliqué au log-retour journalier

Maintenant que la série est prête à être travaillée, nous allons suivre la méthode de modélisation suivante :

- Détermination des ordres p et q
- Estimation et diagnostic du modèle
- Prévisions

3.1.1 Détermination des ordres p et q

Avant de regarder les ACF et PACF, effectuons tout d’abord les test ADF et $KPSS$ pour tester la stationnarité de la série.

```
ADF p-val : 0.0
KPSS p-val : 0.1
```

FIGURE 3.3 – Résultats des tests ADF et KPSS de la série différenciée

L’hypothèse nulle du test ADF étant la non stationnarité de la série, on obtient une p-value égale à 0. On rejette donc la non stationnarité. Quant au test KPSS, celui-ci a comme hypothèse nulle la stationnarité de la série. Par conséquent, avec une p-value à hauteur de 0.1, on ne rejette pas l’hypothèse de stationnarité au seuil de 5 de la série. Continuons alors notre modélisation et déterminons les ordres p et q des modèles ARCH et GARCH.

Comme pour les modèles $ARMA$, les ordres p et q peuvent être déterminé en regardant la fonction d’auto-corrélation partielle de ϵ_t^2 . Nous regardons la série au carré. En effet, la fonction d’auto-corrélation partielle ne peut plus nous renseigner. Regardons l’ACF et la PACF d’un $GARCH(3, 1)$:

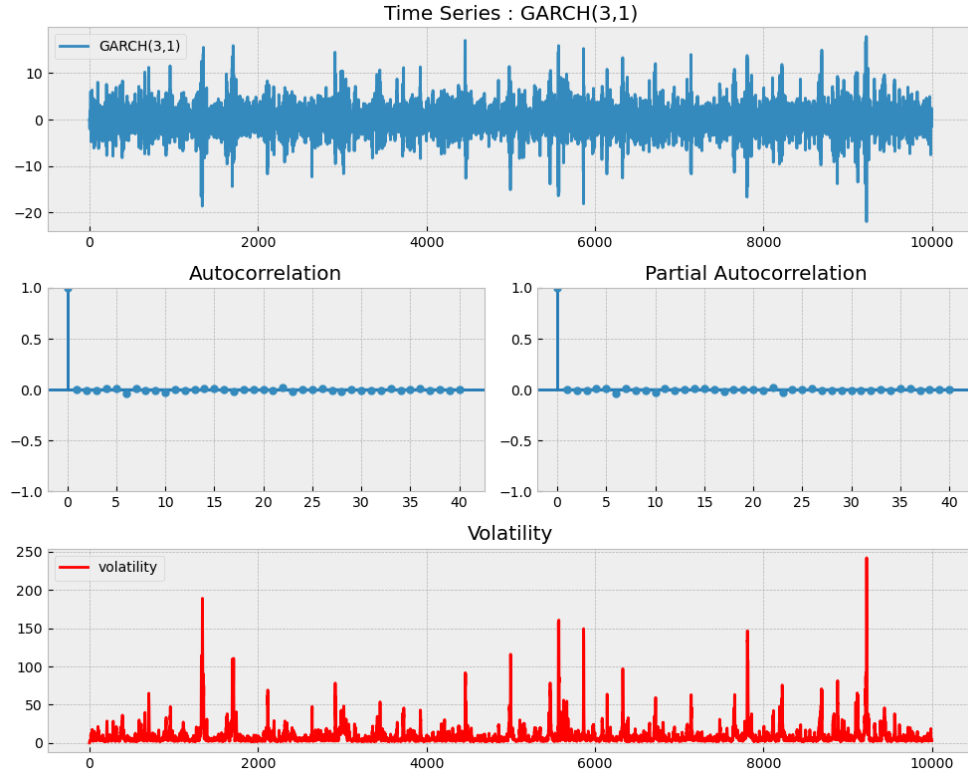


FIGURE 3.4 – Simulation d'un processus $GARCH(3, 1)$

On remarque qu'il n'y a pas de pics remarquable ce qui implique qu'il n'y pas d'auto-corrélation partiel. Ceci est normal car en regardant comment les auto-corrélations partielles sont calculées, on observe une particularité. Soit $\alpha(h)$ la fonctions d'autocorrélation partielle :

$$\alpha(h) = \frac{\begin{vmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(h-2) & \gamma(1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(h-3) & \gamma(2) \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(h-1) & \gamma(h-2) & \dots & \gamma(1) & \gamma(h) \end{vmatrix}}{\begin{vmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(h-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h-1) & \gamma(h-2) & \dots & \gamma(0) \end{vmatrix}}$$

En rappelant que la covariance pour les processus $ARCH$ et $GARCH$ est égale à 0 pour $h \neq 0$, on remarque que le terme en bas de la diagonale sur le déterminant du numérateur est nulle. Par conséquent, pour tout $h \geq 1$, la PACF sera nulle. C'est donc pour cela que l'on analyse l'ACF et la PACF sur les rendements au carré. En effet, on calcule la série au carré pour déterminer les ordres p et q car la variance conditionnelle devient proche d'un $ARMA$. Regardons par exemple un modèle $ARCH(1)$ mis au carré. On obtient la relation suivante :

$$\begin{aligned} \epsilon_t^2 &= \sigma_t^2 \eta_t^2 \\ &= \eta_t^2 (\omega + \alpha_1 \epsilon_{t-1}^2) \end{aligned}$$

Ainsi, en regardant l'ACF et la PACF sur les rendements au carré, on obtient :

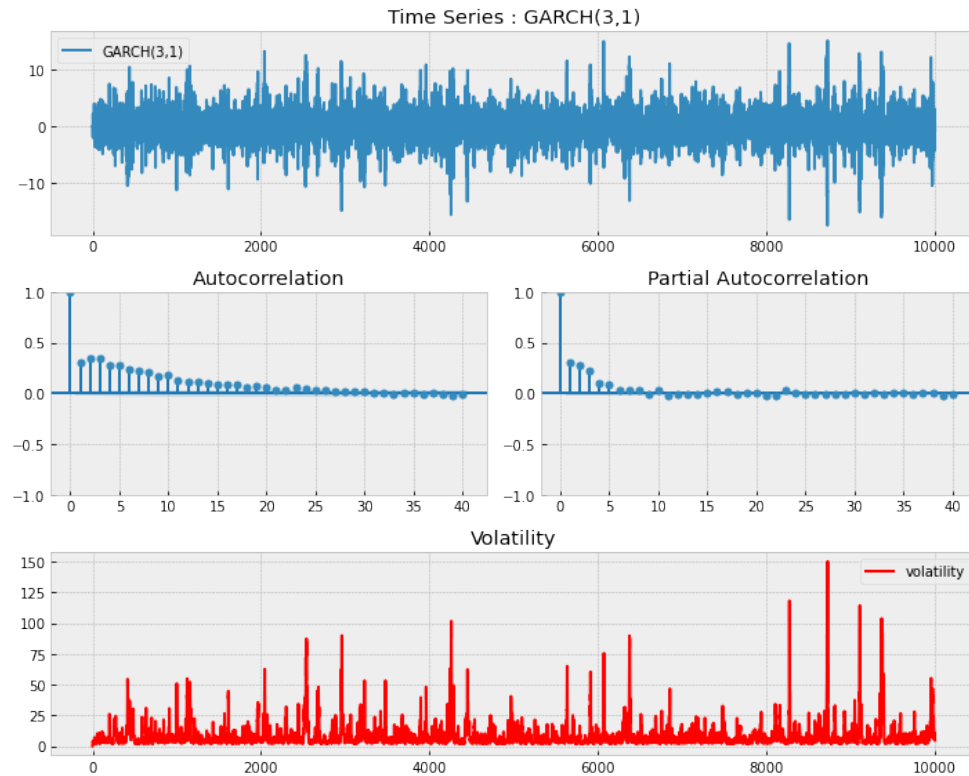


FIGURE 3.5 – Simulation avec ACF et $PACF$ des rendements aux carrés

L'ACF n'est pas pour autant interprétable mais on reconnaît la décroissance rapide des $ARMA$. Tandis que pour la PACF, on peut constater trois pics remarquables (les deux suivants sont considérés comme des artefacts) et estimer l'ordre $p = 3$. Regardons maintenant ce que donne l'ACF et la PACF dans le cas de notre première application, le CAC40 :

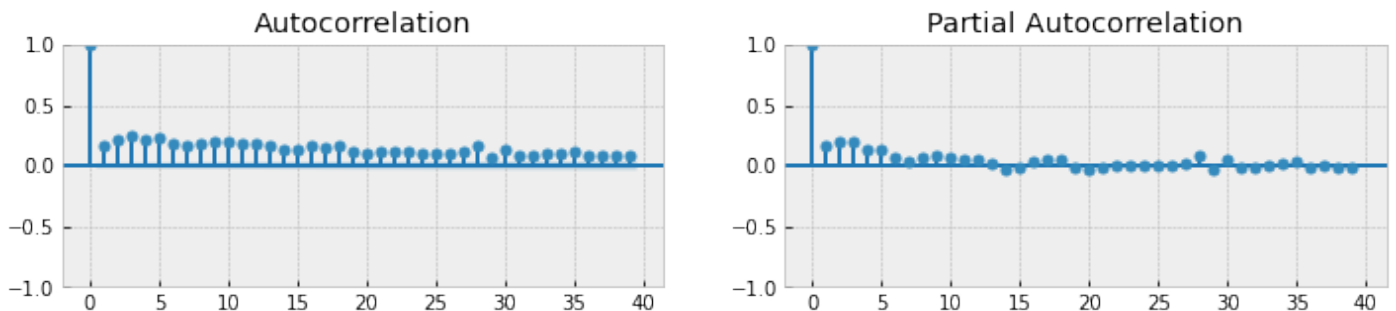


FIGURE 3.6 – ACF et PACF des rendements au carré

On peut observer que l'*ACF* n'est pas interprétable car presque tous les pics sont significatifs. Pour la *PACF*, nous pouvons identifier le pic 5 comme étant le dernier significatif. On retient alors l'ordre $p = 5$. Mais comment déterminer l'ordre q ? Pour cela, on a recours au critère des *AIC*, c'est à dire que nous allons tester toute une série de combinaison (p, q) et comparer chaque *AIC* pour identifier quelle modèle possède celui qui est minimale. Ainsi, pour $p \in \{1, 2, 3, 4, 5, 6, 7\}$ et $q \in \{0, 1, 2\}$. Voici les *AIC* obtenus :

	0	1	2
1	26951.443818	25187.615183	25189.615183
2	26349.354598	25180.167415	25181.835512
3	25864.743596	25182.167415	25180.308904
4	25638.751741	25184.167415	25182.308899
5	25503.224375	25186.167414	25184.308898
6	25402.039311	25188.167415	25184.303881
7	25359.312695	25190.167415	25186.303878

FIGURE 3.7 – *AIC* issus de $GARCH(p, q)$

L'*AIC* minimum est obtenu aux ordres $p = 2$ et $q = 1$ qui est néanmoins très proche de l'*AIC* du modèle $GARCH(3, 2)$. Les modèles avec $q = 0$ paraissent peu adaptés, qui sont entre autre des modèles *ARCH*. En effet, un modèle $ARCH(p)$ est en réalité un modèle $GARCH(p, q)$ avec l'ordre $q = 0$. Nous allons maintenant nous concentrer sur les modèles $GARCH$. Par la suite nous testerons différents modèles $GARCH(p, q)$ afin de pouvoir les comparer et trouver le meilleur modèle. Mais pour l'instant, retenons le modèle $GARCH(2, 1)$ pour la suite de notre modélisation.

3.1.2 Estimations et diagnostics de différents modèles $GARCH(p, q)$

Commençons par entraîner un modèle $GARCH(2, 1)$ sur l'ensemble des données. Voici les résultats obtenus :

Mean Model					
	coef	std err	t	P> t	95.0% Conf. Int.
mu	0.0546	1.218e-02	4.479	7.507e-06	[3.068e-02, 7.842e-02]
Volatility Model					
	coef	std err	t	P> t	95.0% Conf. Int.
omega	0.0375	1.041e-02	3.605	3.121e-04	[1.713e-02, 5.795e-02]
alpha[1]	0.0641	1.746e-02	3.671	2.421e-04	[2.986e-02, 9.828e-02]
alpha[2]	0.0498	2.528e-02	1.971	4.868e-02	[2.887e-04, 9.938e-02]
beta[1]	0.8672	2.116e-02	40.985	0.000	[0.826, 0.909]

FIGURE 3.8 – $GARCH(2, 1)$ sur l'ensemble des données

Le modèle obtenu s'écrit donc :

$$\begin{cases} \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= 0.0375 + 0.0641\epsilon_{t-1}^2 + 0.0498\epsilon_{t-2}^2 + 0.8672\sigma_{t-1}^2 \end{cases}$$

avec $\eta_t \sim \mathcal{N}(0, 1)$ un bruit blanc fort. Les paramètres ω , α_1 , α_2 et β_1 sont estimés par maximum de vraisemblance conditionnelle. Leur p-value sont toutes inférieures à 0.01 donc tous les coefficients sont significatifs. Maintenant que nous avons le modèle estimé, regardons les prévisions de ce modèle.

3.1.3 Prévisions

Pour générer des prévisions, nous procéderons d'abord par renseigner une date qui sera la date t où l'on souhaite couper nos données, et prédire celles qui se trouvent après cette date butoire. Ensuite, nous entraînerons un nouveau modèle $GARCH(2, 1)$ sur nos données tronquées, les coefficients seront exactement les mêmes (à 10^{-3} près). Puis, nous appliquerons la commande *forecasts* de python en renseignant la date suivante à la date butoire. Ainsi, nous obtiendrons des prévisions sur le reste des données. Nos données du *CAC40* allant de Avril 1990 à juillet 2021, nous allons tronqué nos données jusqu'au 31 décembre 2020 et essayer de prédire la volatilité de l'année 2021. Une fois les valeurs préditent, nous les représentons en même temps que la volatilité réel pour pouvoir comparer visuellement :

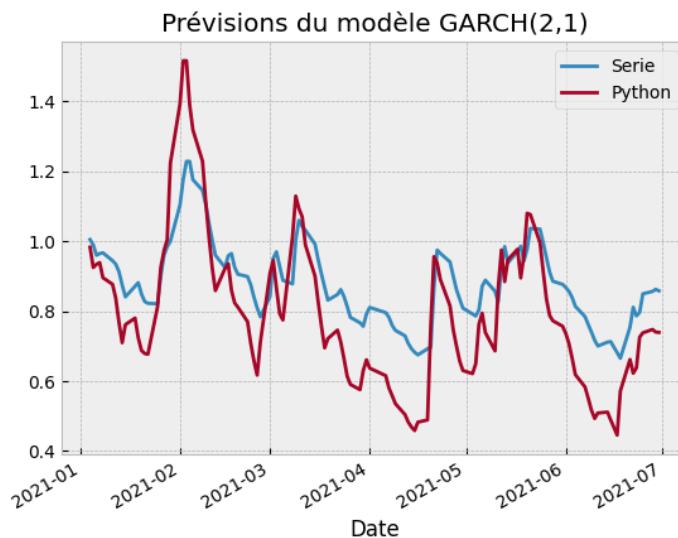


FIGURE 3.9 – Représentation des prédictions de la volatilité

On observe que la courbe des prévisions (courbe en rouge) est généralement inférieure à la réalité, mais on peut constater un pic qui est surestimé au moment d'une forte variation de la volatilité. On peut donc déduire que nos prédictions sont sensibles aux fortes variations. Pour mesurer, l'erreur de prédiction, nous avons recours à la mesure du *Root Mean Square Error* (RMSE) qui mesure les différence entre les prévisions et les vraies valeurs. Plus ce critère est proche de 0, plus cela signifie que les prédictions sont proches des valeurs réels. Pour notre modélisation $GARCH(2, 1)$, nous obtenons une *RMSE* égale à 0.1436 ce qui est assez proche de 0, on peut donc dire que nos prévisions sont plutôt correct. Nous cherchons tout de même s'il n'est pas possible de trouver un meilleur modèle permettant de minimiser encore plus le *RMSE*. Pour cela, nous procédons comme pour la recherche de l'*AIC* optimale, nous testons une série de combinaison (p, q) où nous générons des prévisions pour chaque modèle construit. Pour chacune de ces prévisions, nous calculons leur *RMSE* que nous stockons dans un tableau. Pour toutes combinaisons $(p, q) \rightarrow \{1, 2, 3, 4, 5, 6, 7\} \times \{1, 2, 3\}$, voici le tableau résumant les valeurs des *RMSE*.

	0	1	2
1	0.421608	0.14816	0.148157
2	0.332582	0.14363	0.144313
3	0.272027	0.143633	0.143821
4	0.23041	0.14363	0.149552
5	0.20785	0.143629	0.14382
6	0.192607	0.143627	0.151476
7	0.184817	0.143635	0.151476

FIGURE 3.10 – Tableau représentant toutes les $RMSE$ des modèles testés

On peut constater que les meilleurs score de $RMSE$ sont obtenu avec le modèle $GARCH(5, 1)$ et notre modèle $GARCH(2, 1)$. A remarquer que l'on retrouve l'ordre $p = 5$ de la $PACF$. Ces modèles ne sont pas à retenir absolument car d'autres modèles possèdent une $RMSE$ pratiquement égale, à 10^{-4} près. Les graphiques étant très similaire, nous ne le représenterons pas.

Maintenant, nous avons obtenu ces prévisions à l'aide de la commande `forecasts` de python, nous allons donc essayer de les retrouver par le calcul à la main en suivant les indications émisent dans la section 2.4.2. Voici un tableau référencant les valeurs obtenu avec python dans la colonne "Python" et les valeurs calculées dans la colonne "Calculer".

	Python	Calculer
Date		
2021-01-04	0.982950	0.979258
2021-01-05	0.924860	0.921985
2021-01-06	0.934107	0.936453
2021-01-07	0.938932	0.946053
2021-01-08	0.895808	0.900001
...
2021-06-24	0.726006	0.725291
2021-06-25	0.738188	0.739188
2021-06-28	0.747865	0.736077
2021-06-29	0.740414	0.731803
2021-06-30	0.739433	0.730012
126 rows × 2 columns		

FIGURE 3.11 – Tableau des prévisions calculées

Les valeurs sont très proches les unes des autres, nous avons même un coefficient de corrélation de Pearson à hauteur de 0.9996 entre nos deux distributions.

Pour terminer cette première application, on se propose de représenter quatres prévisions issu de modèle différents. On prendra nos deux meilleur modèles qui sont des modèles *GARCH*, puis on prendra deux modèles *ARCH* pour comparer le prévisions de ces deux processus que l'on étudie. Voici les représentations de nos quatre modèles sélectionné :

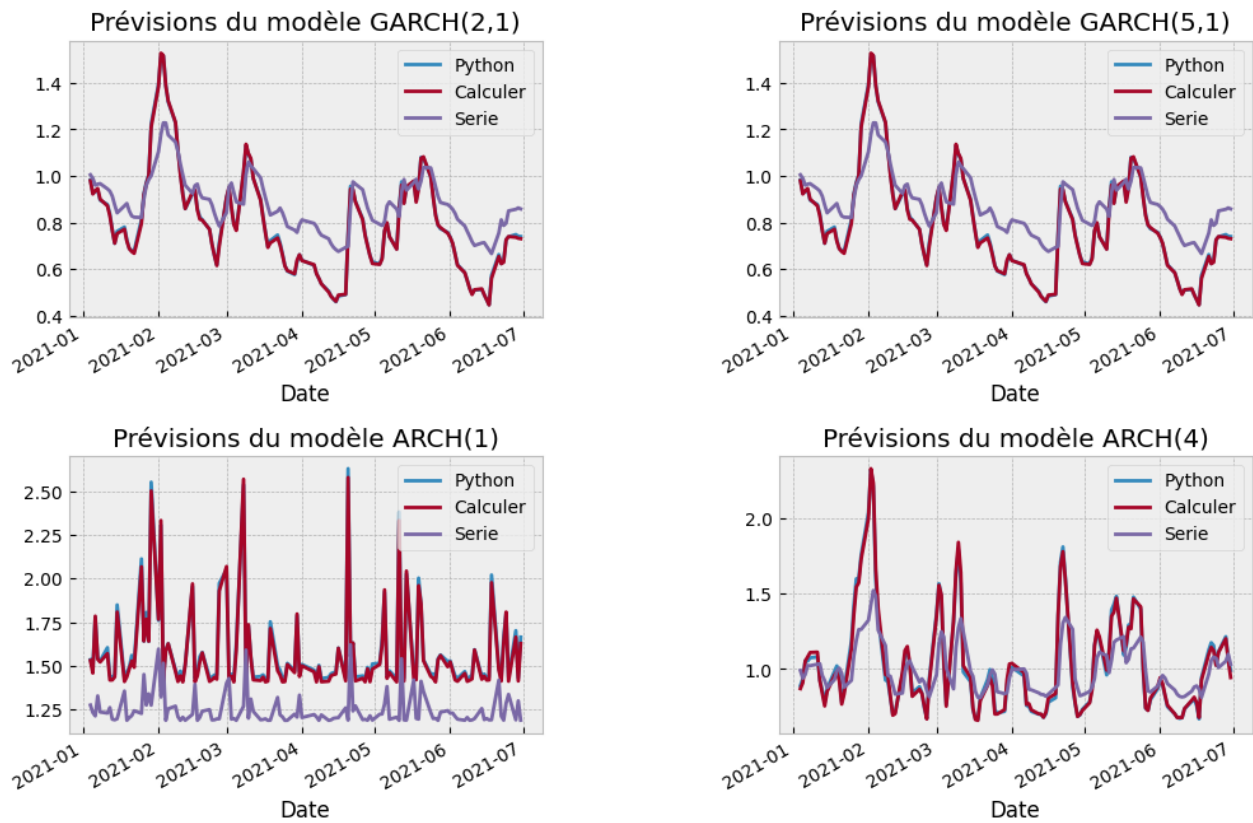


FIGURE 3.12 – Graphes de prévisions calculées par les modèles *ARCH* et *GARCH*

On constate que les prévisions issus des modèles *GARCH* sont plus précise vis à vis des valeurs réel (courbe en violet), les prévisions des modèles *GARCH* sont donc bien meilleures que les modèles *ARCH*. De plus, on peut voir que la courbe des valeurs calculées (courbe en bleu) par python sont partiquement confondu avec la courbe des prévisions calculées (courbe en rouge). Cette application aux données du *CAC40* étant terminé, on se propose une seconde application avec les données du *SP500* que nous introduisons dans la section qui suit.

3.2 Construction et modélisation pour la série $SP500$

L'indice $SP500$ est un indice boursier qui mesure la performance d'environ 500 sociétés aux États-Unis (telles que Apple, Microsoft ou encore Amazon). Il inclut des sociétés de 11 secteurs différents afin de donner une image de la santé du marché boursier américain et de son économie en général. Ces données sont également très connues, c'est pourquoi nous les traitons dans ce rapport. Voici à quoi ressemble la série :

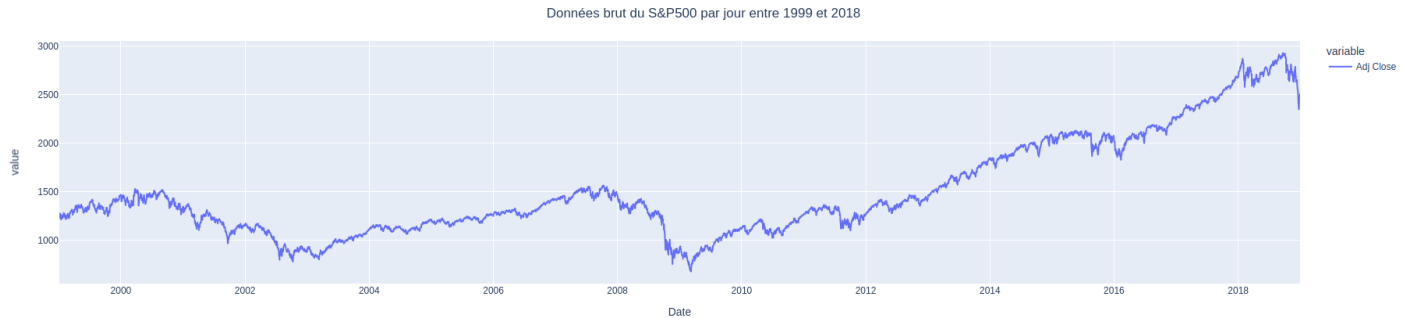


FIGURE 3.13 – Prix de clôture de l'indice boursier journalier du SP 500

Comme pour l'application au $CAC40$, on applique le log-retour journalier à cette série pour obtenir les rendements. Voici à quoi ressemble nos rendements :

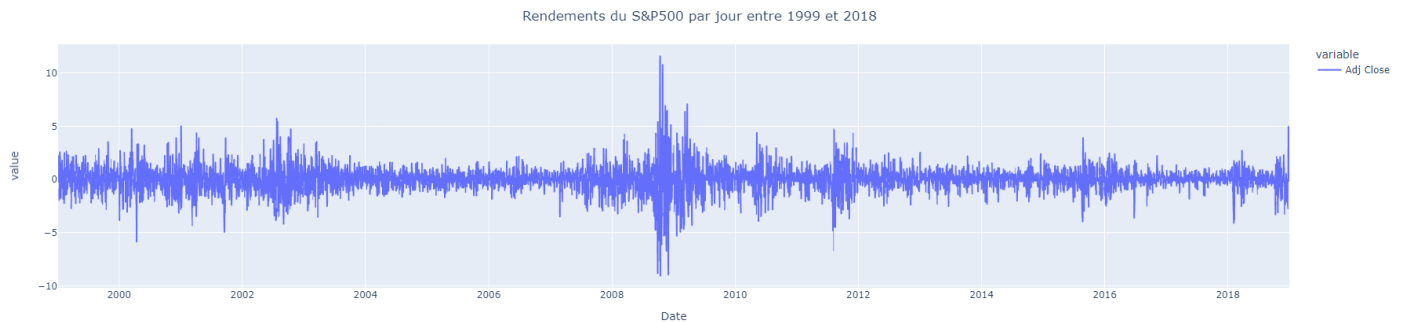


FIGURE 3.14 – Rendements de l'indice boursier journalier du SP 500

Contrairement à la série d'origine, les rendements semblent être stationnaire. Nous le testons grâce aux tests ADF et $KPSS$:

```
ADF p-val : 1.8847790799642558e-29
KPSS p-val : 0.1
```

FIGURE 3.15 – Résultats des tests ADF et $KPSS$ de la série différenciée

L'hypothèse nulle du test ADF étant la non stationnarité de la série, on obtient une p-value très proche de 0. On rejette donc la non stationnarité au seuil de 1%. Quant au test $KPSS$, celui-ci a comme hypothèse nulle la stationnarité de la série. Par conséquent, avec une p-value à hauteur de 0.1, on ne rejette pas l'hypothèse de stationnarité au seuil de 5%. Les tests sont bien en accord avec la stationnarité de la série. Nous pouvons donc continuer notre modélisation. L'étape suivante va consister à déterminer les ordres p et q des modèles $ARCH$ et $GARCH$.

3.2.1 Détermination des ordres p et q

Pour déterminer les ordres p et q , on a recours à la *PACF* pour identifier le paramètre p et à l'*ACF* pour identifier le paramètre q . Comme vu précédemment, on applique cette étape aux rendements mis au carré, voici le résultat :

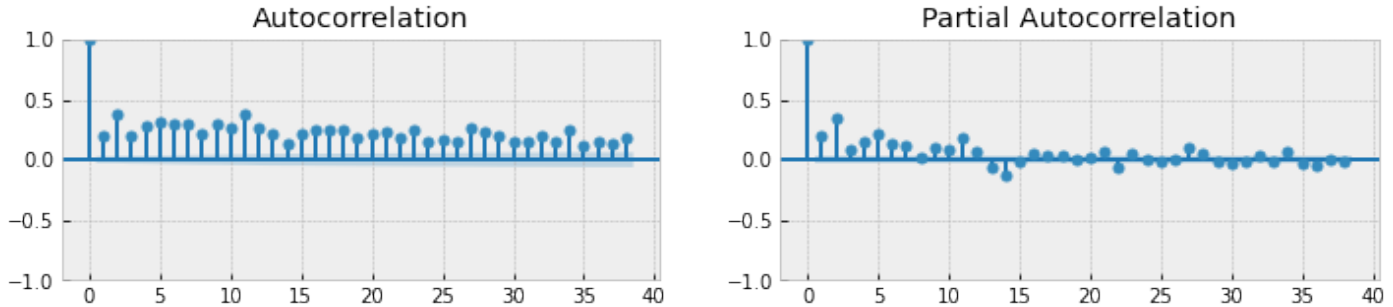


FIGURE 3.16 – ACF et PACF des rendements aux carrés

Malheureusement, l'*ACF* n'est pas vraiment interprétable. Quant à la *PACF* on distingue quelques pics significatifs. Le dernier considéré est le pic 11. Les pics suivants pourraient être discutable. Nous les considérons ici comme des artefacts. En revanche, prendre $p = 11$ est un ordre beaucoup trop grand. En général, les ordres p et q sont relativement petit (à hauteur de 1,2,3 voir 4 maximum). Il n'est donc pas raisonnable de considérer un modèle avec $p = 11$. Nous allons donc avoir recours à l'AIC afin de trouver les ordres p et q . Nous allons ainsi tester toute une série de combinaison (p, q) afin d'entraîner un $GARCH(p, q)$. Pour chacune de ces combinaisons, nous allons prendre la valeur de l'AIC et ainsi prendre celle qui minimise parmi tous les modèles. Pour toutes combinaisons $(p, q) \rightarrow \{1, 2, 3, 4, 5, 6, 7\} \times \{1, 2, 3\}$, voici un tableau résumant les valeurs de l'AIC obtenu.

	0	1	2
1	15613.28536	13881.436954	13883.436954
2	14853.598648	13875.39208	13873.483431
3	14515.941724	13877.392079	13875.475552
4	14261.160188	13879.392079	13877.475553
5	14132.890046	13881.39208	13879.475552
6	14073.280024	13883.392079	13881.475552
7	14015.108815	13884.567381	13883.475552

FIGURE 3.17 – Tableau des AIC

En comparant les valeurs de ces AIC, nous pouvons nous apercevoir que l'AIC minimale serait pour les ordres $p = 2$ et $q = 2$. Par la suite, nous testerons alors un $GARCH(2, 2)$ mais aussi les autres combinaisons qui sont relativement proches la valeur de l'AIC minimale. Maintenant que nous avons déterminé les ordres p et q , nous pouvons passer à l'estimation du modèles.

3.2.2 Estimation du modèle $GARCH(p, q)$

Grâce à la bibliothèque *arch model* de python, nous pouvons modéliser un modèle $GARCH(2, 2)$ et ainsi estimer les paramètres. Voici la sortie obtenue du modèle entraîné :

Mean Model					
	coef	std err	t	P> t	95.0% Conf. Int.
mu	0.0572	1.151e-02	4.972	6.642e-07	[3.467e-02, 7.980e-02]
Volatility Model					
	coef	std err	t	P> t	95.0% Conf. Int.
omega	0.0321	8.849e-03	3.628	2.858e-04	[1.476e-02, 4.944e-02]
alpha[1]	0.0675	2.173e-02	3.107	1.892e-03	[2.492e-02, 0.110]
alpha[2]	0.1115	2.356e-02	4.730	2.248e-06	[6.527e-02, 0.158]
beta[1]	0.2636	0.266	0.992	0.321	[-0.257, 0.784]
beta[2]	0.5341	0.244	2.193	2.831e-02	[5.674e-02, 1.011]

FIGURE 3.18 – Estimation des paramètres du modèle $GARCH(2, 2)$

On obtient le modèle suivant :

$$\begin{cases} \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= 0.0321 + 0.0675\epsilon_{t-1}^2 + 0.1115\epsilon_{t-2}^2 + 0.2636\sigma_{t-1} + 0.5341\sigma_{t-2} \end{cases}$$

Avec $\eta_t \sim \mathcal{N}(0, 1)$ un bruit blanc. Les paramètres ont été estimés par maximum de vraisemblance conditionnelle. Maintenant que le modèle les paramètres ont été estimés, intéressons-nous aux prévisions de ce modèle.

3.2.3 Prévisions

L'étape des prévisions se déroulera comme pour l'exemple précédent. Nous allons tout d'abord renseigner une date qui indiquera l'instant t où l'on désire couper la série. Les données suivantes seront ainsi prédites par la commande *Forecast* de Python puis comparées avec les prévisions que nous obtenons à la main. Pour cela on entraîne un même modèle $GARCH(2, 2)$ sur la série tronquée. On applique ensuite la commande *forecast* en indiquant le jour suivant à la date butoir. Les paramètres estimés sont les mêmes à 10^{-3} près. Voici la représentation graphique de nos prévisions sur l'année 2018 :

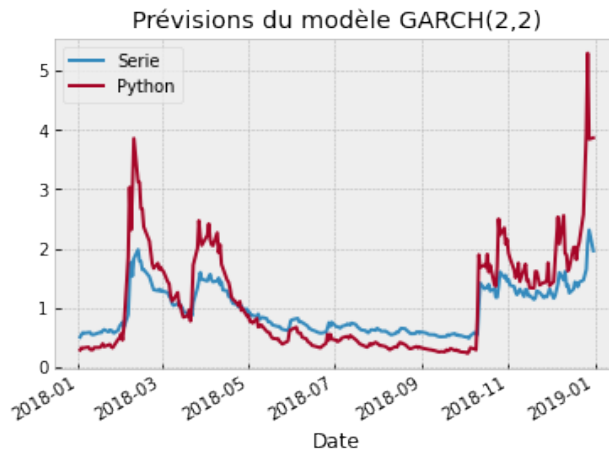


FIGURE 3.19 – Prévisions du modèle $GARCH(2, 2)$

Comme nous pouvons le voir, les prévisions sont parfois surestimées mais celles-ci restent néanmoins corrects. Par la suite, nous avons calculé une *Root Mean Square Error* (*RMSE*) pour mesurer les différences entre les valeurs de nos prévisions et celles de notre série. Plus cette valeur est proche de 0, plus les prédictions sont proches des vraies valeurs. Nous obtenons une *RMSE* à hauteur de 0.516 ce qui est assez élevé et normal au vu du graphique obtenu. Comme le *RMSE* est relativement élevé, nous allons chercher quel modèle minimise le critère du *RMSE*. Pour cela, on procède comme vu précédemment avec les AIC. Nous testons plusieurs combinaisons d'ordre p et q puis effectuons des prévisions pour chacun des modèles entraînés. Ainsi nous pouvons calculer le *RMSE* que l'on inscrit dans un tableau. Pour toutes combinaisons $(p, q) \rightarrow \{1, 2, 3, 4, 5, 6, 7\} \times \{1, 2, 3\}$, voici un tableau résumant les *RMSE* calculés :

	0	1	2
1	0.871049	0.488549	0.488558
2	1.062168	0.514079	0.515882
3	1.103283	0.514063	0.515697
4	1.020929	0.51406	0.515704
5	0.906038	0.514071	0.515724
6	0.833973	0.514065	0.515704
7	0.789412	0.546893	0.528092

FIGURE 3.20 – Tableau regroupant les RMSE de différents modèles $GARCH(p, q)$

Nous pouvons voir que le *RMSE* minimal est obtenu avec un $GARCH(1, 1)$ à hauteur de 0.4885 cependant la valeur du *RMSE* du $GARCH(1, 2)$ est pratiquement similaire. Représentons alors les prévisions du modèle $GARCH(1, 1)$:

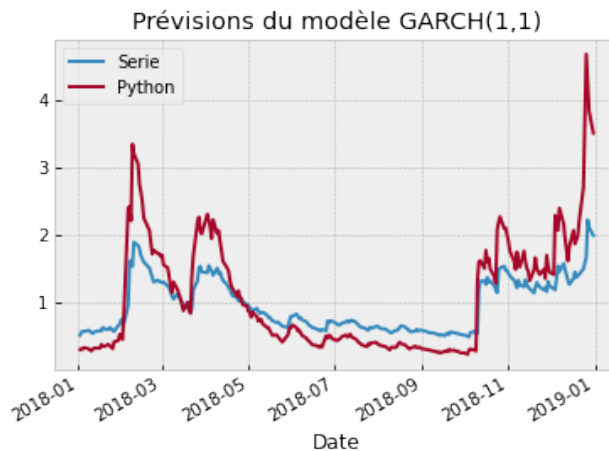


FIGURE 3.21 – Prévisions du modèle $GARCH(1, 1)$

Visuellement, comparé aux prévisions obtenu par le modèle $GARCH(2, 2)$, on ne distingue pas d'amélioration net. Ceci n'est pas surprenant puisque leurs *RMSE* respectifs est assez proches. Maintenant que nous savons générer des prévisions, nous allons essayer de les retrouver par le calcul à la main grâce aux expressions décrites en section 2.4.2.

Voici un tableau indiquant dans la colonne nommé "Python" les prévisions générées par la commande *forecasts* et dans la colonne "Calculer" les prévisions que nous avons calculé à la main :

	Python	Calculer
Date		
2018-01-02	0.300110	0.323764
2018-01-03	0.315421	0.337755
2018-01-04	0.308480	0.326834
2018-01-05	0.330229	0.353541
2018-01-08	0.311570	0.326405
...
2018-12-24	2.705965	2.763023
2018-12-26	4.681160	5.003688
2018-12-27	4.267176	4.471440
2018-12-28	3.839358	3.968785
2018-12-31	3.512197	3.605923
251 rows × 2 columns		

FIGURE 3.22 – Tableau de comparaison des prévisions

Nous pouvons voir que les valeurs sont proches les unes des autres. Pour s'en convaincre, on applique un coefficient de corrélation de Pearson entre ces deux distributions. Nous obtenons une corrélation de 0.99905. On se propose alors de représenter ces valeurs calculées en les superposant aux prévisions de *forecasts* pour quatre modèles différents. Selon le tableau des *RMSE*, nous choisissons les deux meilleurs modèles, c'est à dire les modèles *GARCH*(1,1) et *GARCH*(1,2), puis deux autres modèles avec des *RMSE* bien plus grandes tel que les modèles *ARCH*(1) et *ARCH*(4).

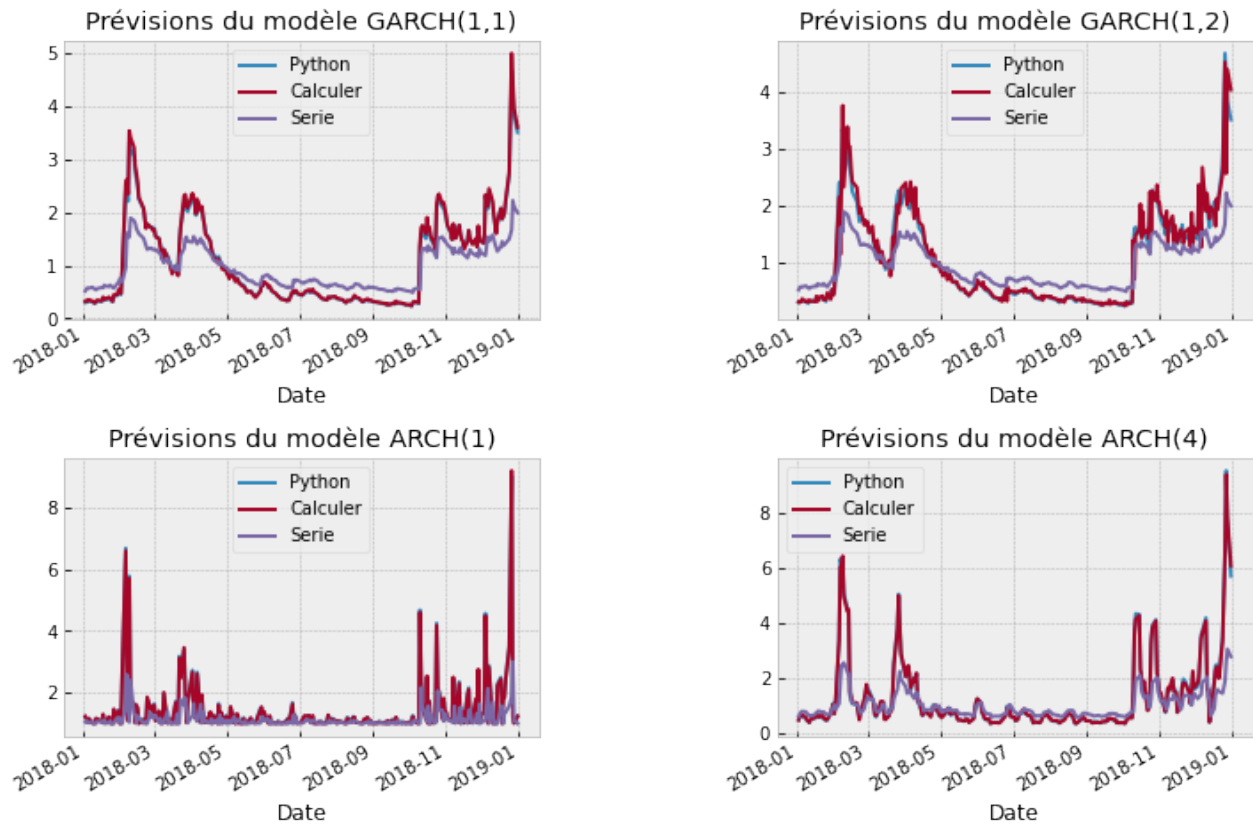


FIGURE 3.23 – Graphes des prévisions calculées par plusieurs modèles

Nous observons que la courbe des valeurs calculées à la main (courbe en rouge) recouvre pratiquement les valeurs générées par *forecasts* (courbe en bleu). Nous pouvons également conclure par le fait que les modèles $GARCH(p, q)$ sont plus adaptés à cette série que les modèles $ARCH(p)$.

Cette section étant terminée, avant de conclure ce rapport on se propose une petite ouverture en introduisant simplement les processus dit : *ARMA – GARCH*

Chapitre 4

Les processus : ARMA-GARCH

Les modèles $ARMA(P, Q) - GARCH(p, q)$ (Autoregressive Moving Average - Generalized Autoregressive Conditional Heteroskedasticity), sont des modèles similaires aux processus $ARCH$ et $GARCH$. Ils sont utilisés pour modéliser la volatilité des séries chronologiques. Cependant, ces modèles ont une particularité. Ils combinent deux modèles, les $ARMA$ et les $GARCH$. Le modèle $ARMA$ est utilisé pour modéliser la tendance de la série à l'aide des valeurs passées de la série. Les ordres P et Q sont respectivement les ordres classiques des $ARMA$ identifiant l'ordre auto-regressif et moyenne mobile. Quand au processus $GARCH$, lui est utilisé pour modéliser la volatilité de la série en utilisant en particulier les erreurs ϵ_t du modèle $ARMA$.

4.1 Définition

Ainsi, en combinant ces deux modèles, les $ARMA - GARCH$ vont modéliser à la fois la tendance et la volatilité de la série, ce qui est un atout majeur pour travailler sur les séries financières où la volatilité est la plupart du temps très élevée et variable. Un tel modèle, supposé centré en 0, se définit comme suit :

$$\begin{aligned}X_t &= \sum_{i=1}^P \phi_i X_{t-i} + \epsilon_t - \sum_{j=1}^Q \theta_j \epsilon_{t-j} \\ \epsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2\end{aligned}$$

Avec $\eta \sim \mathcal{N}(0, 1)$, un bruit blanc, les coefficients ω , α_i et β_j sont définis comme dans un processus $GARCH$ et les paramètres ϕ_i , θ_j sont les coefficients des processus $ARMA$. Nous allons proposer une simulation d'un tel processus et les comparer avec les modèles $ARMA$ et $GARCH$ classiques.

4.2 Simulation

Pour l'exemple, nous simulerons un processus $ARMA(1,1)$ avec les paramètres $\phi_1 = 0.5$ et $\theta_1 = 0.3$, un $GARCH(3,1)$ avec comme paramètres : $\omega = 1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.1$, $\alpha_3 = 0.2$ et $\beta = 0.4$. En voici une représentation :

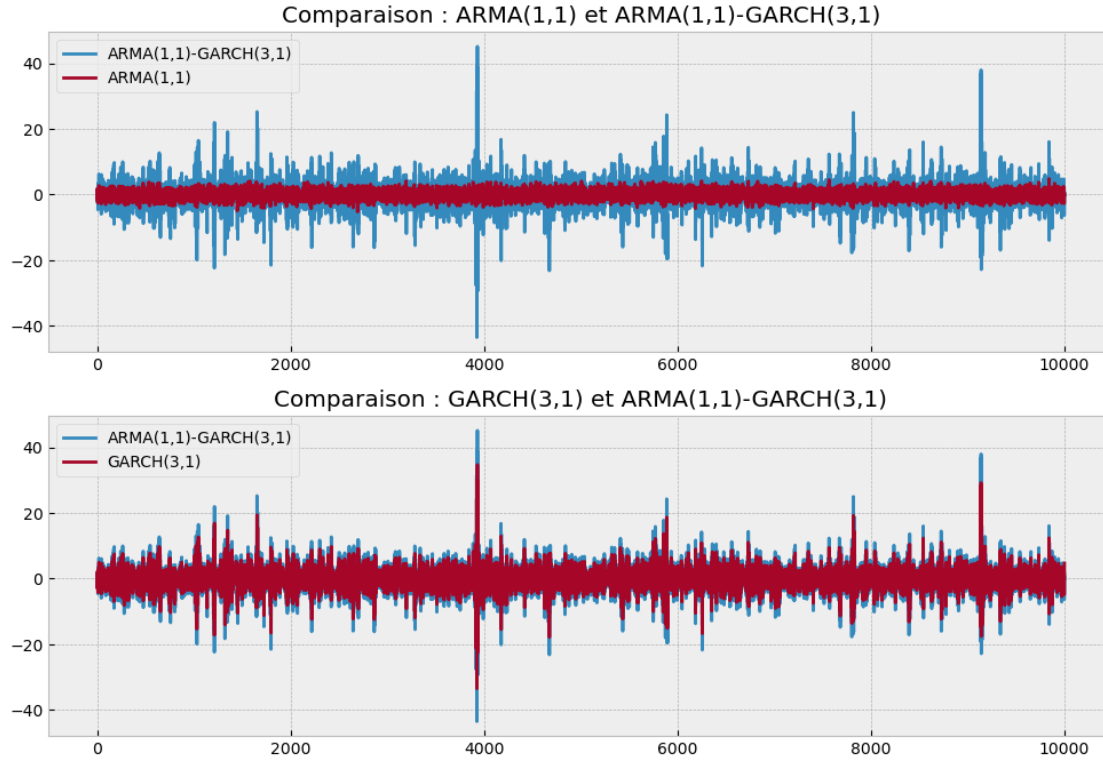


FIGURE 4.1 – Comparaison des différents processus

Chapitre 5

Conclusion

En conclusion de ce projet, nous avons étudié les modélisations des processus à hétéroscédasticités conditionnelle, et ils ont été étudiés sur le cas des séries financières. Le but était donc d'étudier ces processus et de montrer dans quelles mesures ces modèles sont capables de reproduire des phénomènes financiers tel que la volatilité. Il a été démontré dans un premier temps que les modèles simples de modélisation de série temporelle comme les *ARMA* ne sont pas adaptés à ce type de série chronologique. Pour cause, la variance conditionnelle est constante dans un processus *ARMA* ce qui n'est pas le cas pour nos modèles *ARCH* et *GARCH*. Nos simulations nous ont permis de mettre en application ces affirmations. Nous avons ensuite mis en application ces méthodes sur deux jeux de données, le *CAC40* et les données du *SP500* et cherché la meilleure modélisation en comparant la qualité des prévisions.

Nous n'avons étudié seulement deux modèles permettant de modéliser la volatilité financière, mais il en existe plusieurs notamment les processus *ARMA* – *GARCH* introduit dans la dernière section qui combinent à la fois une modélisation *ARMA* avec des erreurs modélisées par un modèle *GARCH*. Il existe également par exemple les processus comme les *EGARCH* (Exponential Generalized Autoregressive Conditional Heteroskedasticity) où l'idée est que la variance peut être exprimée en termes exponentiels mais ceux-ci sont sujets à d'autres études plus complexes.