



MAP 565 : MÉMOIRE SUR LES SÉRIES TEMPORELLES

Mémoire MAP 565

10 avril 2024

Mathias GRAU, Antoine MARTINEZ



TABLE DES MATIÈRES

1	Introduction	1
2	Evolution des valeurs boursières	2
2.1	Les séries temporelles	2
2.2	Stationnarité d'une série temporelle	5
3	Analyse par des processus de GARCH	6
3.1	Modèles ARCH et GARCH	6
3.2	Prévision de la volatilité	7
4	Théorie des valeurs extrêmes	9
4.1	La distribution généralisée des valeurs extrêmes	9
4.2	Application à notre série temporelle de price closing	9
5	Étude des copules	11
5.1	Présentation du concept	11
5.2	Application à notre jeu de données	13
6	Processus Gaussiens et interpolation par krigage	15
6.1	Processus Gaussien	15
6.2	Krigeage par processus gaussien	15
7	Annexe	19

1 INTRODUCTION

Au cours de ce rapport, nous allons traiter de l'évolution des prix des entreprises ayant un lien avec le domaine quantique afin de capturer leurs liens et leurs évolutions.

Pour cela, nous avons sélectionné les principales entreprises impliquées dans le domaine de l'informatique quantique d'après le classement *Forbes* [14]. Nous avons donc pu sélectionner un dizaine de firmes qui sont :

1. International Business Machines (IBM)
2. Microsoft Corporation (MSFT)
3. Alphabet Inc. (GOOGL)
4. Amazon.com Inc. (AMZN)
5. Intel Corporation (INTC)
6. ...

Nous avons par la suite utilisé l'API de `yahoo_finance` afin de récupérer les données quotidiennes et horaires de prix des cours des actions afin des les analyser par la suite.

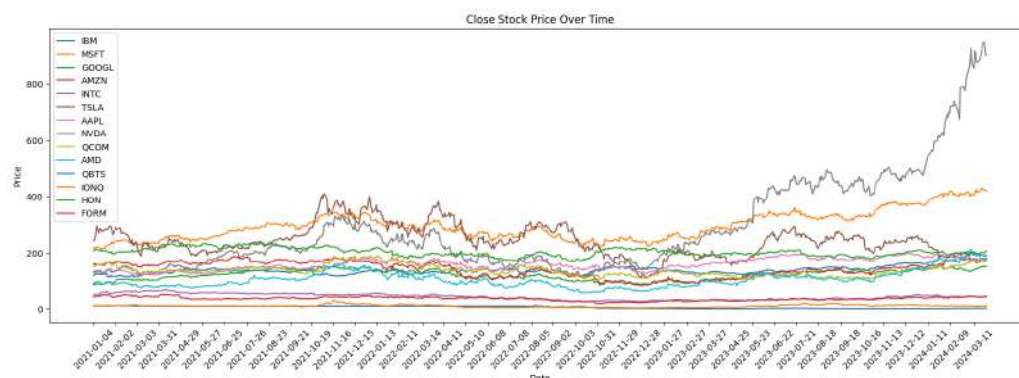


FIGURE 1 – Cours des actions des principaux acteurs dans le domain quantique

Nous allons alors utiliser ce jeu de données afin d'illustrer quelques notions vues en cours :

- La théorie des extrêmes
- Les copules
- Les processus GARCH

ainsi que certaines notions parallèles sur les time series et les processus gaussien avec l'introduction du krigeage.

2

EVOLUTION DES VALEURS BOURSIÈRES

Dans cette première partie, notre objectif principal est d'explorer en détail les schémas et les tendances cachés au sein de la série temporelle qui représente le prix des actions de différentes sociétés présentes dans les ordinateurs quantique. Nous allons nous concentrer sur l'étude des modèles capables d'extraire des informations des données historiques pour prédire le comportement futur.

Ces modèles de séries temporelles sont largement utilisés dans le domaine de la finance quantitative en raison même des données financières. Ces modèles jouent un rôle prépondérant dans la prédiction des actifs financiers.

Notre étude commence par une présentation détaillée des outils analytiques nécessaires pour décortiquer les caractéristiques et les tendances sous-jacentes des séries temporelles. Nous aborderons ensuite les méthodes permettant d'évaluer la stationnarité des séries temporelles, et d'autre qui permettent de se ramener à des séries stationnaires.

2.1 LES SÉRIES TEMPORELLES

Une série temporelle est une suite de valeur indexée par des données temporelles. Ces valeurs sont donc séparées par des intervalles de temps qui sont en temps normal égaux. Dans notre étude les données ont été collectées chaque jour, ce qui nous fait un décalage ou Lag en anglais de 1 journée : $\Delta t = t_{i+1} - t_i = 1\text{jour}$.

Il y a un total de T décalages et donc les valeurs temporelles sont notées : $(t_i)_{1 \leq i \leq T}$. Comme dans de nombreux modèles, connaître l'écart de temps entre deux valeurs peut être primordial pour comprendre l'évolution de la variable d'intérêt.

Une série temporelle est dite linéaire, si elle peut être écrite comme une somme pondérée de perturbations passées, ϵ_t -qui sont également appelées innovations et sont ici supposées représenter un bruit blanc et de la moyenne de la série μ :

$$y_t = \mu + \sum_{i=0}^{+\infty} a_i \epsilon_{t_i}, \text{ avec } a_0 = 1 \text{ et } \epsilon \text{ i.i.d.}$$

Un de nos principaux objectifs est de comprendre le comportement de notre série temporelle en déterminant les différents coefficients $(a_i)_{i \geq 0}$. Il existe ensuite des méthodes qui permettent de prédire le comportement de la série en extrayant des motifs clés à partir de ces données. Parmi les outils disponibles, on peut noter l'analyse de l'autocorrélation ou les statistiques à fenêtre glissante comme les moyennes mobiles.

2.1.1 • DÉCOMPOSITION DES MOTIFS DES SÉRIES TEMPORELLES

Les motifs des séries temporelles que l'on cherche à déterminer peuvent en général être décomposées en plusieurs composantes distinctes.

Le principe de la décomposition des séries temporelles est de découper les séries originales en plusieurs composants indépendants : tendance, saisonnalité et bruit.

Nous avons alors souhaité déterminer la tendance de notre série temporelle en utilisant les moyennes glissantes. Le principe est d'effectuer une moyenne sur un intervalle de temps que l'on peut modifier pour se rendre compte de la tendance plus globale de notre variable.

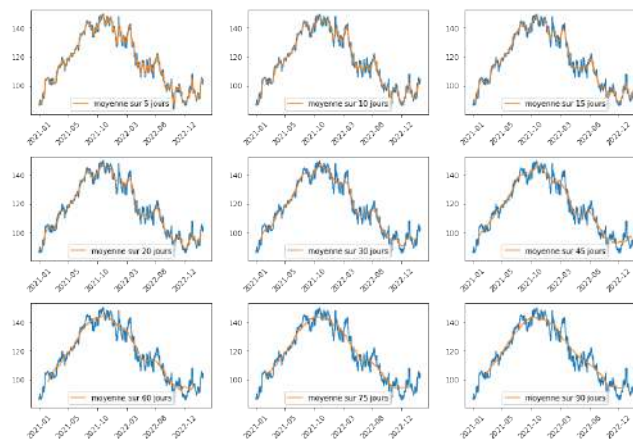


FIGURE 2 – Détermination de la tendance

Avec une moyenne glissante effectuée sur 90 jours nous pouvons alors avoir une idée plus claire de ce qu'est la tendance de notre série temporelle.

Pour la suite de notre décomposition, nous avons décidé d'utiliser une bibliothèque python. En effet le module Statsmodels comprend une méthode simple pour diviser la série temporelle en composantes distinctes de tendance, de saisonnalité et de résidu à l'aide de moyennes mobiles. Nous pouvons alors l'appliquer aux données journalières de price closing :

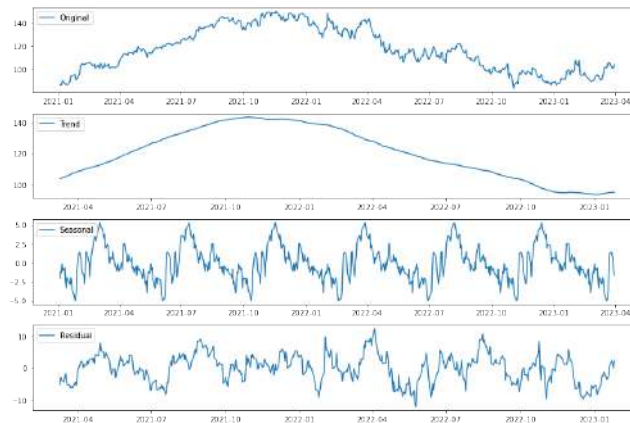


FIGURE 3 – Décomposition des motifs

La partie restante des données, après avoir extrait les composantes de tendance et de saisonnalité, est la composante résiduelle. En supposant que les tendances et les variations saisonnières sont relativement prévisibles, la partie la plus pointue de l'étude réside en l'analyse des données résiduelles qui sont considérées comme plus aléatoires. Nous avons essayer de trouver la période optimale pour la saisonnalité de notre série temporelle et nous

2.1.2 • L'AUTOCORRÉLATION

L'autocorrélation est un indicateur montrant comment les observations d'une série temporelle sont liées entre elles. Il s'agit de la corrélation entre la série et cette même série décalée dans le temps. Nous pouvons donc définir l'autocorrélation de décalage k comme ceci :

$$\gamma(t, t+k) = \frac{1}{\sigma^2} \text{Cov}(y_t, y_{t+k})$$

N'ayant pas accès à la covariance et variance théorique de notre série temporelle, nous allons donc approcher cette valeur par l'autocorrélation empirique de décalage k :

$$\rho(k) = \frac{\sum_{t=k+1}^T (y_{t-k} - \bar{y})(y_t - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

L'autocorrélation de décalage k prend donc des valeurs entre -1 et 1, ce qui va nous indiquer à quel point les valeurs avec un décalage de k sont corrélées.

Nous avons donc tracé l'autocorrélation du price closing de notre série temporelle en modifiant le décalage de 1 à 500 jours :

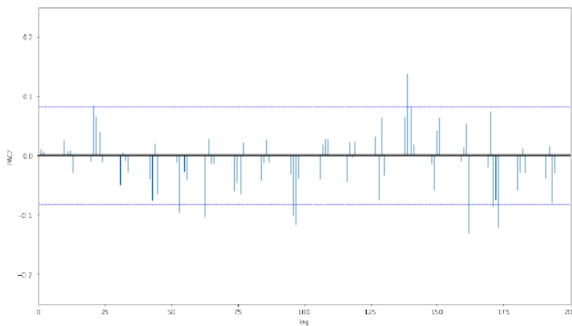


FIGURE 4 – Tracé de l'autocorrélation partielle en fonction du décalage

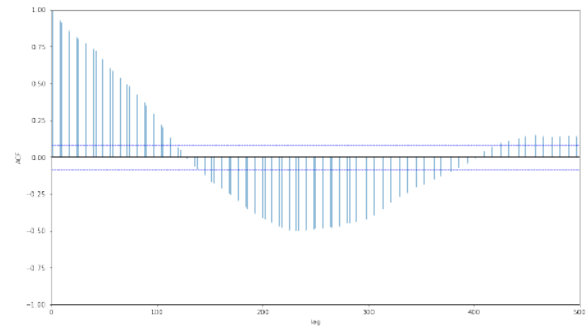


FIGURE 5 – Tracé de l'autocorrélation en fonction du décalage

Dans notre cas, nous pouvons remarquer une oscillations de l'autocorrélation, bien que l'amplitude décroît lorsque le décalage augmente.

Lorsque nous examinons l'autocorrélation pour un décalage k supérieur à 1, nous évaluons la relation entre les observations qui sont séparées par plus d'un pas de temps. Cette notion mesure à la fois la corrélation directe entre ces observations et l'effet indirect des observations intermédiaires. A contrario, l'autocorrélation partielle élimine cet effet indirect et se concentre uniquement sur la dépendance linéaire entre les observations à un décalage donné, en se débarrassant les effets des décalages plus courts. La fonction d'autocorrélation partielle (PACF) révèle ainsi les corrélations qui persistent une fois que les effets des corrélations à court terme ont été éliminés.

2.2 STATIONNARITÉ D'UNE SÉRIE TEMPROELLE

Nous allons maintenant nous intéresser à la notion de stationnarité des séries temporelles. Le principe de stationnarité est le fait que plusieurs valeurs caractéristiques doivent rester constante au cours du temps. En effet, la moyenne, la variance ou la covariance ne doivent pas dépendre du temps lorsqu'elles sont calculées de manière glissante comme nous avons vu avec la moyenne précédemment. Ces caractéristiques traduisent le fait qu'une série stationnaire ne doit pas être sujet à des phénomènes saisonniers ou de tendances globales.

Le fait d'avoir une série stationnaire est un réel avantage pour connaître le comportement de celle-ci. Pour cela nous allons nous intéresser à des transformations permettant de s'y ramener.

Pour rendre une série temporelle stationnaire, nous devons souvent la transformer en plusieurs étapes. Dans un premier temps nous pouvons essayer de supprimer la tendance avec une transformation logarithmique par exemple. Ensuite nous pouvons transformer les données originales en une série de différences d'une période à l'autre et/ou d'une saison à l'autre. Ainsi on peut se ramener après un certain nombre de transformation à une série stationnaire

3

ANALYSE PAR DES PROCESSUS DE GARCH

On suppose dans cette partie que le marché est efficient, c'est à dire que le prix d'une action reflète toutes les informations disponibles.

$$\mathbb{E}[X_t | X_{t-1}, X_{t-2}, \dots] = 0$$

où X_t est le prix/rendement actuel de l'actif, conditionné sur les prix/rendements précédents. Cela justifie régulièrement une modélisation par des lois centrées

La modélisation ARCH va permettre d'essayer de distinguer les périodes de haute et basse volatilité. Cette approche permet de gérer les risques. Elle est particulièrement utile à moyen terme.

3.1 MODÈLES ARCH ET GARCH

Modèle ARCH : X_t peut être le prix, les rendements, ou les rendements logarithmiques d'un actif, modélisé par la somme de ses valeurs historiques au carré X_{t-j}^2 . Ici, nous définissons X_t comme étant les rendements logarithmiques d'un actif.

$$X_t = \sigma_t \epsilon_t \text{ avec } \sigma_t^2 = b_0 + \sum_{j=1}^p b_j X_{t-j}^2$$

où ϵ_t est IID(0,1) et $b_0, b_1, \dots, b_p > 0$.

La variance marginale permet d'incorporer l'hétéroscédasticité en modélisant la volatilité au temps t (σ_t^2), variance conditionnelle de la série chronologique, et est directement influencée par le rendement au carré X_t^2 .

Nous pouvons utiliser tout cela pour voir intuitivement que si X_t oscille rapidement. C'est ainsi que les modèles ARCH capturent les oscillations de prix et leur impact sur la volatilité.

Modèles GARCH :

$$X_t = \sigma_t \epsilon_t \text{ avec } \sigma_t^2 = b_0 + \sum_{j=1}^p b_j X_{t-j}^2 + \sum_{j=1}^n a_j \sigma_{t-j}^2$$

où ϵ_t est IID(0,1) et $b_0, b_1, \dots, b_p, a_0, a_1, \dots, a_n \neq 0$.

Interprétation : Nous pouvons ainsi réduire l'énoncé du modèle en décomposant les parties du modèle en : un effet ARCH qui implique que la variance d'une série temporelle dépend des valeurs passées de la série et effet GARCH étend le concept d'ARCH en introduisant des termes autorégressifs pour modéliser non seulement les variations passées de la volatilité mais aussi les résidus des erreurs de volatilité.

3.1.1 • TEST DE MCLEOD-LI

Le test de McLeod-Li propose les hypothèses nulle et alternative suivantes pour tester le comportement de bruit blanc dans les résidus de la série :

- H_0 : il n'y a pas d'effets ARCH dans la série temporelle
- H_a : il y a des effets ARCH dans la série temporelle

Nous ajustons donc d'abord un modèle GARCH(1,1) puis appliquons le test de McLeod-Li aux résidus. Le résultat du test est une valeur de $p = 10^{-14} < 0.001$ et nous rejetons l'hypothèse H_0 d'absence d'effet ARCH et poursuivons notre étude.

3.1.2 • AJUSTEMENT DU MODÈLE GARCH(P,Q)

Ci-dessous, nous ajustons le modèle GARCH(1,1) aux données. Nous mettons en œuvre le processus en utilisant une division des données en ensembles d'entraînement et de test.

Nous utilisons la fonction `arch_model()` du package `arch` et spécifions que les données ont une moyenne nulle et sont modélisées avec un processus GARCH.

Les paramètres p et q dans un modèle GARCH jouent des rôles distincts dans la modélisation de la volatilité conditionnelle d'une série temporelle financière.

- Le paramètre p contrôle la partie autoregressive (AR) du modèle. Il indique le nombre de termes de la volatilité passée à prendre en compte pour prédire la volatilité actuelle.
- Le paramètre q contrôle la partie moving average (MA) du modèle. Il indique le nombre de termes d'erreur de volatilité passée à inclure dans le modèle.

Après optimisation du modèle de GARCH, on obtient les paramètres suivants pour les prédictions futures : $\sigma_t^2 = \mathbf{0.17} + \mathbf{0.03}X_{t-1}^2 + \mathbf{0.88}\sigma_{t-1}^2$

3.2 PRÉVISION DE LA VOLATILITÉ

Prévision à N Pas en Avance : En utilisant le modèle GARCH mentionné précédemment, nous sommes capables de prévoir les volatilités σ_t^2 à l'instant t . Nous utilisons ensuite ces prévisions pour étendre la prévision de la volatilité aux instants $t + 1, t + 2, \dots, t + n$. Nous ne mettons pas à jour continuellement X_t à mesure que le temps t augmente, mais limitons plutôt nos prévisions pour toutes les dates futures à être faites uniquement à partir de la date actuelle.

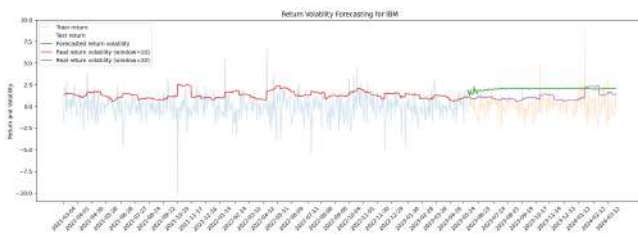


FIGURE 6 – Prédiction de la volatilité

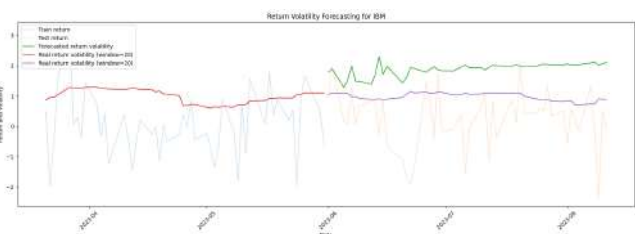


FIGURE 7 – Zoom sur la prédiction

On peut voir une petite tendance croissante à la suite de la période d'entraînement qui se confirme effectivement par les données de test donc la volatilité tend à augmenter en moyenne glissante. On comprend alors que le modèle n'est valable que sur de courtes durées de prédictions et qu'il n'y a plus beaucoup de sens à le prendre en compte pour des périodes de très long terme.

Prévision en Temps Réel d'un Pas en Avant : Nous avons donc modifié la manière d'entraîner et de tester nos prédictions, en partant de données d'entraînement glissante et en testant sur quelques journées suivantes. Nous allons donc ne faire que des prédictions à un pas en avant en se basant sur les informations jusqu'au temps t , et de mettre à jour les informations dans le modèle en temps réel à mesure que le temps passe de t à $t+1$, $t+2$, etc. Plus simplement, chaque jour qui passe, nous mettons à jour X_t , X_{t+1} , X_{t+2} avec le rendement observé réel de ce jour-là pour modéliser un pas en avant.

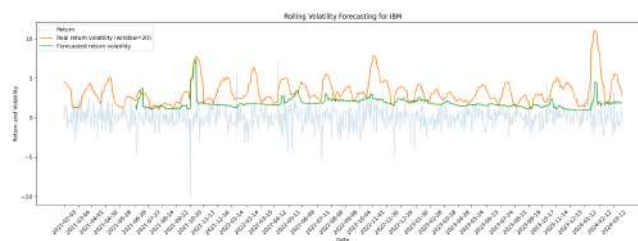


FIGURE 8 – Prévision de la volatilité *1 step forward*

Les résultats sont alors bien plus satisfaisants et aussi plus cohérent avec ce qui peut être fait en réalité si l'on souhaite accéder à la volatilité à court terme.

3.2.1 • EVALUATION DU MODÈLE

Nous faisons une hypothèse dans le modèle $GARCH(P, Q)$ sur la distribution de nos erreurs : si le modèle est correctement spécifié, alors les erreurs sont distribuées normalement avec une moyenne de 0 et une variance constante. Le test de Jarque-Bera vérifie si les erreurs du modèle sont distribuées normalement :

- H_0 : Les données suivent une distribution normale.
- H_a : Les données ne suivent pas une distribution normale.

Nous obtenons à l'issue du test un p-valeur de 10^{-17} qui nous oblige à rejeter le modèle de distribution normale, il convient d'en essayer d'analyser des distributions plus spécifiques. Le modèle n'est donc pas valide dans notre cas, il faudrait envisager des distributions plus spécifiques et adaptées, comme par exemple les distributions de student qui ont des queues plus "épaisses" tout en étant similaires à la loi normale et permettent de mieux cerner les évolutions de retours tels que les nôtres pour lesquels les valeurs extrêmes ont une plus grande probabilité.

4

THÉORIE DES VALEURS EXTRÊMES

La théorie des valeurs extrêmes étudie les distributions des valeurs extrêmes dans un ensemble de données. Ces événements sont en général des phénomènes rares mais qui peuvent être très significatifs, comme les catastrophes naturelles par exemple. Dans notre cas nous allons nous intéresser aux valeurs extrêmes du rendement des actions.

4.1 LA DISTRIBUTION GÉNÉRALISÉE DES VALEURS EXTRÊMES

La distribution généralisée des valeurs extrêmes (GEV) est une des distributions les plus couramment utilisées en théorie des valeurs extrêmes. Elle est définie par trois paramètres : la localisation μ , l'échelle σ et la forme γ .

Ce dernier paramètre γ est central car il va déterminer la nature de la GEV. En effet celle-ci peut représenter trois types de distributions extrêmes : la distribution de Gumbel $\gamma = 0$, la distribution de Fréchet $\gamma > 0$, et la distribution de Weibull $\gamma < 0$.

4.2 APPLICATION À NOTRE SÉRIE TEMPORELLE DE PRICE CLOSING

Le premier objectif est de déterminer un seuil pour notre étude. Nous avons alors choisi un seuil en traçant l'excès moyen de seuil dont la formule est la suivante :

$$e(s) = \frac{\sum_{X_i > s} X_i - s}{\text{Card}(X_i > s)}$$

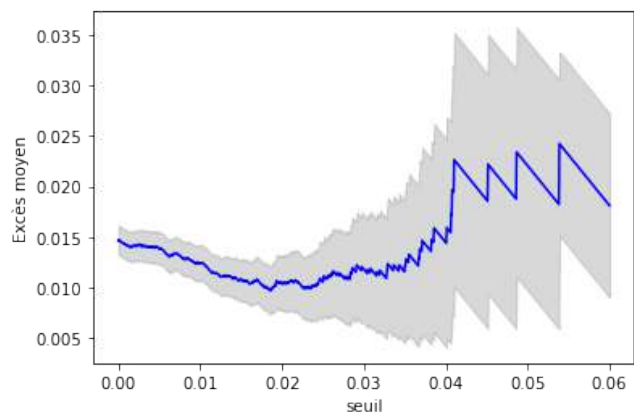


FIGURE 9 – Excès moyen de seuil

Nous avons donc choisi un seuil à 0.02 car le but est que la courbe soit approximativement linéaire au dessus du seuil choisi et on peut voir qu'entre 0.02 et 0.04 la courbe d'excès moyen de seuil semble linéaire.

Nous pouvons alors afficher les données de rendement de nos actions qui sont au dessus du seuil pour se rendre compte de la concentration de celles-ci :

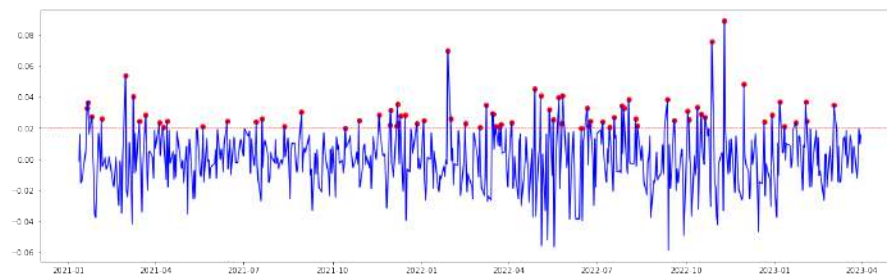


FIGURE 10 – Plot des valeurs au dessus du seuil

En effectuant notre analyse nous trouvons un paramètre $\gamma > 0$ donc nous sommes dans une distribution de Fréchet. Enfin nous avons tracé les courbes Q-Q plots, P-P plot et la densité de probabilité calculé à partir de ce modèle pour vérifier les résultats obtenus :

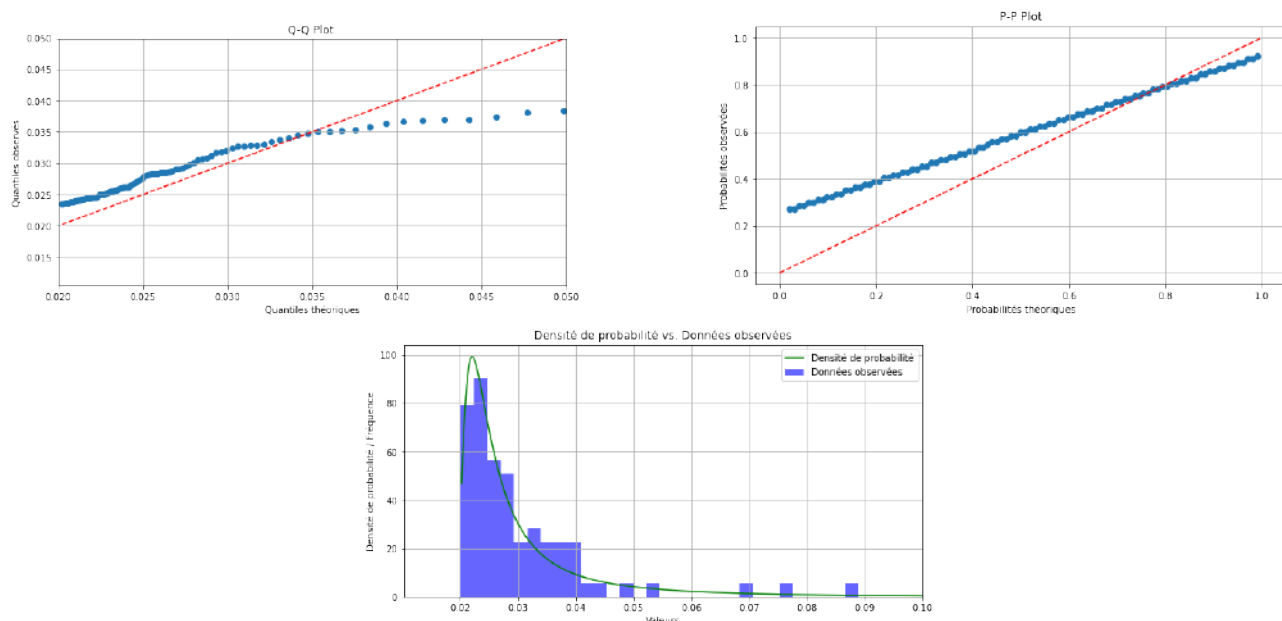


FIGURE 11 – Q-Q plot, P-P plot et densité de probabilité

5

ÉTUDE DES COPULES

5.1 PRÉSENTATION DU CONCEPT

Une copule : est une fonction de répartition, notée C , définie sur $[0, 1]^d$, dont les marges sont uniformes sur $[0, 1]$. Une caractérisation est alors que :

- $C(u_1, \dots, u_d) = 0$ si une des composantes u_i est nulle,
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$,
- C est d-croissante.

En dimension 2, $C(0, v) = C(u, 0) = 0$ pour tout u et v , $C(u, 1) = u$ et $C(1, v) = v$ pour tout u et v , et enfin, la propriété de 2-croissance se traduit par :

$$C(u_1, v_1) - C(u_1, v_2) - C(u_2, v_1) + C(u_2, v_2) \geq 0$$

pour tout $0 \leq u_1 \leq u_2 \leq 1$ et $0 \leq v_1 \leq v_2$.

5.1.1 • THÉORÈME DE SKLAR

Le théorème de Sklar dit que si C est une copule, et si F_1, \dots, F_d sont des fonctions de répartition (univariées), alors : $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$ est une fonction de répartition de dimension d , dont les marges sont précisément F_1, \dots, F_d .

Et réciproquement, si F est une fonction de répartition en dimension d , il existe une copule C telle que : $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$ où les F_i sont les lois marginales de F .

5.1.2 • LIEN AVEC LES LOIS MARGINALES

Si ces lois marginales sont toutes continues, la copule C est alors unique, on a :

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$$

Dans ce cas, on pourra alors parler de la copule associée à un vecteur aléatoire (X_1, \dots, X_d) .

La copule d'un vecteur aléatoire (X_1, \dots, X_d) est alors la fonction de répartition du vecteur aléatoire $(F_1(X_1), \dots, F_d(X_d))$, que l'on notera parfois (U_1, \dots, U_d) .

Un intérêt de la copule est de simuler une variable aléatoire multivariée à partir de sa copule et de ses lois marginales. Il suffit de générer un échantillon (U_1, \dots, U_d) à partir de la copule et de construire l'échantillon voulu grâce à la relation :

$$(X_1, \dots, X_d) = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d))$$

où F_i^{-1} désigne la fonction quantile associée à F_i :

$$F_i^{-1}(q) = \inf\{x : F_i(x) \geq q\}$$

5.1.3 • DENSITÉ DE PROBABILITÉ

Enfin, désignons par $c(\cdot, \cdot)$ la fonction de densité correspondant à la copule $C(\cdot, \cdot)$, c'est-à-dire,

$$c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2).$$

La relation suivante entre la densité jointe $f(\cdot, \cdot)$ et la densité de copule $c(\cdot, \cdot)$ est donc :

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)c(F_1(x_1), F_2(x_2))$$

où $f_1(x_1)$ et $f_2(x_2)$ désignent respectivement les fonctions de densité marginales de X_1 et X_2 . Ainsi, la fonction de densité de la copule peut être exprimée comme suit :

$$c(u_1, u_2) = \frac{f(F_1^{-1}(u_1), F_2^{-1}(u_2))}{f_1(F_1^{-1}(u_1))f_2(F_2^{-1}(u_2))}.$$

5.1.4 • DÉPENDANCE DE LA QUEUE INFÉRIEURE ET SUPÉRIEURE

Soient X_1 et X_2 des variables aléatoires avec une copule $C \in \mathcal{C}^2$ et des fonctions de distribution continues et strictement croissantes F_1 et F_2 , respectivement. Alors les coefficients de dépendance des queues sont donnés par ($\alpha \in (0, 1)$) :

$$\lambda_l = \lim_{\alpha \rightarrow 0} \frac{C(\alpha, \alpha)}{\alpha} \quad (1)$$

$$\lambda_u = 2 - \lim_{\alpha \rightarrow 1} \frac{1 - C(\alpha, \alpha)}{1 - \alpha} \quad (2)$$

5.1.5 • PRINCIPALES COPULES ÉTUDIÉES DANS LE COURS

Les copules gaussiennes : sont basées sur la distribution normale multivariée et sont souvent utilisées lorsque les données présentent une corrélation linéaire. Elles sont caractérisées par une matrice de corrélation, qui mesure les relations linéaires entre les variables aléatoires. $C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$ où Φ_ρ est la fonction de répartition bivariée de la distribution normale avec une corrélation ρ , et Φ^{-1} est la fonction quantile de la distribution normale.

La copule de Clayton : est souvent utilisée pour modéliser la dépendance positive entre les variables aléatoires. Elle est caractérisée par un paramètre de dépendance, qui contrôle la force de la dépendance entre les variables (où θ est le paramètre de dépendance) $C_\theta(u, v) = \max\{0, (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}\}$

La copule de Gumbel : quant à elle, est utilisée pour modéliser la dépendance négative entre les variables aléatoires. Elle est caractérisée par un paramètre de dépendance qui contrôle la queue de la distribution de probabilité (où θ est le paramètre de dépendance) $C_\theta(u, v) = \exp\left(-\left(u^{-\theta} + v^{-\theta}\right)^{-1/\theta}\right)$

5.1.6 • ESTIMATION DES PARAMÈTRES DES COPULES BIVARIÉES

Les méthodes les plus couramment utilisées pour estimer le paramètre θ sont la méthode du maximum de vraisemblance et la méthode des moments. En revanche, et comme nous le ferons dans ce mémoire, nous pouvons aussi estimer le coefficient θ à partir du coefficient de corrélation de Kendall τ calculé empiriquement sur les jeux de données. Ce coefficient de Kendall mesure à la base de manière empirique le degré de dépendance entre les variables aléatoires : il mesure la concordance des paires d'observations et varie entre -1 et 1, où -1 indique une corrélation négative parfaite, 0 indique l'indépendance et 1 indique une corrélation positive parfaite.

Pour les copules de Clayton et de Gumbel, il existe une relation entre le coefficient de corrélation de Kendall (τ) et les paramètres θ : $\tau_{Clayton} = \frac{\theta}{2+\theta}$ et $\tau_{Gumbel} = 1 - \frac{1}{\theta}$

5.2 APPLICATION À NOTRE JEU DE DONNÉES

On dispose ainsi de copules de Clayton et de Gumbel associées aux jeux de données. Dans le cas de de Clayton et de Gumbel, nous obtenons la fonction de répartition de la copule aux paramètres : $(\tau_{Clayton}, \theta_{Clayton}) = (0.5705, 2.6569)$ et $(\tau_{Gumbel}, \theta_{Gumbel}) = (0.5705, 2.3284)$

Pour la copule de Clayton, si $\theta = 0$, nous obtenons la copule d'indépendance. La copule de Clayton est une copule complète : elle interpole entre une limite inférieure de la copule de contre-monotonie ($\theta \rightarrow -1$) et une limite supérieure de la copule de comonotonie ($\theta \rightarrow +\infty$). Ici nous avons une légère comonotonie.

Pour la copule de Gumbel, si ($\theta = 1$), nous obtenons la copule d'indépendance et lorsque ($\theta \rightarrow +\infty$), nous nous approchons de la copule comonotone. Ici, nous avons un theta relativement faible, ce qui signifie que nous sommes proches de la copule indépendante.

On suppose des lois marginales normales pour les retours, ce qui est usuellement fait en pour les données financières dans le cadre des marchés efficients (on aurait pu cependant prendre des distributions de Student pour avoir des queues plus épaisses comme nous l'avons évoqué précédemment) : $\mathcal{N}_{GOOGL}(0.00088, 0.01981^2)$ et $\mathcal{N}_{MSFT}(0.00095, 0.01719^2)$.

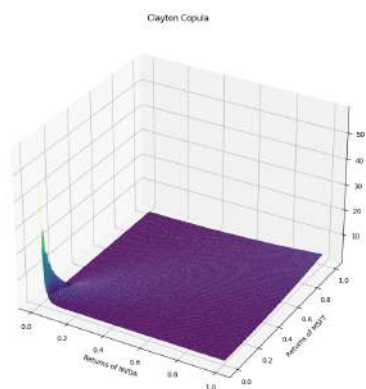


FIGURE 12 – Densité de probabilité de la copule de Clayton ajustée aux données

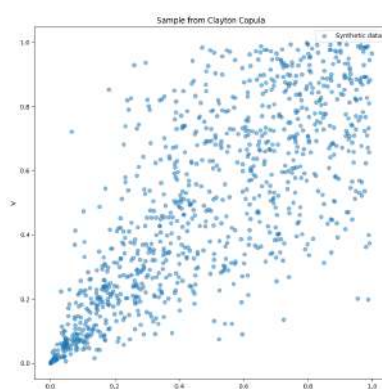


FIGURE 13 – Echantillon uniforme à partir de la copule ajustée

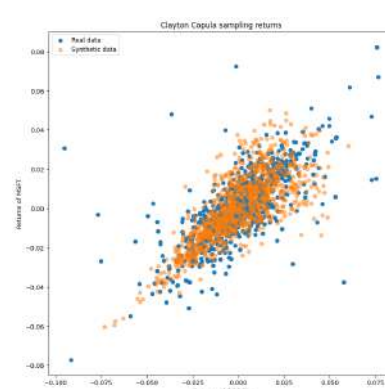


FIGURE 14 – Echantillon de return simulé à partir de la copule ajustée

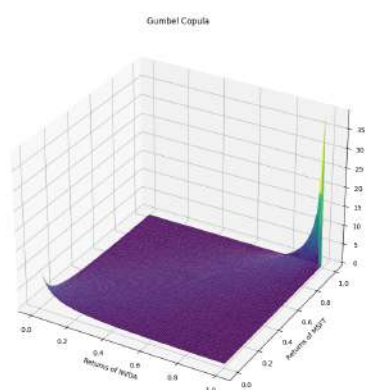


FIGURE 15 – Densité de probabilité de la copule de Gumbel ajustée aux données

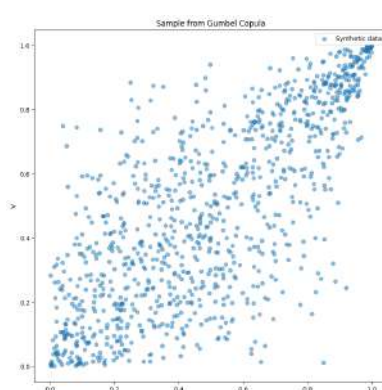


FIGURE 16 – Echantillon uniforme à partir de la copule ajustée

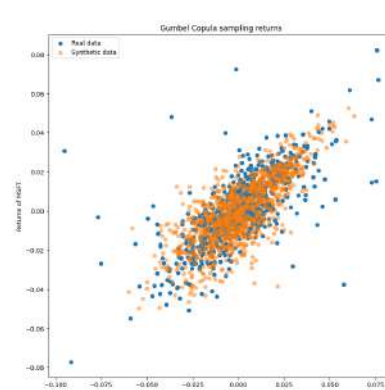


FIGURE 17 – Echantillon de return simulé à partir de la copule ajustée

On peut alors évaluer les dépendances hautes et basses de ces deux copules ajustées :

	Clayton	Gumbel
λ_l	$2^{-\frac{1}{\theta}} = 0.77$	0
λ_u	0	$2 - 2^{\frac{1}{\theta}} = 0.65$

6

PROCESSUS GAUSSIENS ET INTERPOLATION PAR KRIGEAGE

6.1 PROCESSUS GAUSSIEN

Vecteur aléatoire gaussien : On considère $n > 1$ ainsi qu'un vecteur aléatoire $\mathbf{y} = (y_1, \dots, y_n)$. On dit que \mathbf{y} est un vecteur aléatoire gaussien si l'une des deux conditions suivantes (équivalentes) est vérifiée :

1. Pour tous $a_1, \dots, a_n \in \mathbb{R}$, la variable aléatoire $\sum_{i=1}^n a_i y_i$ suit une distribution gaussienne
2. Il existe un vecteur \mathbf{m} de taille n et une matrice positive K de taille $n \times n$ tels que le vecteur \mathbf{y} possède une fonction caractéristique qui s'écrit $\mathbf{u} \rightarrow \exp(i\mathbf{u}^t \mathbf{m} - \frac{1}{2} \mathbf{u}^t K \mathbf{u})$

Noyau : Le choix du *kernel* est crucial car il détermine la fonction de covariance du processus et la nature de la corrélation entre les données. Différents noyaux peuvent être utilisés (noyau linéaire, noyau gaussien, ...) mais nous avons choisi sur le noyau matern, où ν et ρ sont des paramètres à estimer ou définir :

$$K(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} \|x - x'\|}{\rho} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} \|x - x'\|}{\rho} \right)$$

Fonction de covariance : Chaque élément K_{ij} de la matrice \mathbf{K} est calculé en appliquant le noyau sur la paire de points (x_i, x_j) , c'est-à-dire $K_{ij} = k(x_i, x_j)$.

Famille paramétrique Une famille paramétrique de fonctions de covariance stationnaires est de la forme $\{K_\psi, \psi \in \Psi\}$ où K_ψ est une fonction de covariance stationnaire, et Ψ est un sous-ensemble de \mathbb{R}^p . Puisque la variance du processus gaussien stationnaire est constante, on considère $K_\psi : \{\sigma^2 R_\theta, \sigma^2 > 0, \theta \in \Theta\}$. La séparation explicite du paramètre de variance σ^2 et du paramètre de corrélation θ s'avèrera utile lorsque nous aborderons leur estimation par maximum de vraisemblance.

6.2 KRIGEAGE PAR PROCESSUS GAUSSIEN

Pour un vecteur de taille $n = n_1 + n_2$ (n_1 : nombre de points connus, et n_2 ceux recherchés)

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m^{(1)} \\ m^{(2)} \end{bmatrix}, \begin{bmatrix} K_{1,1} & K_{1,2} \\ K_{2,1} & K_{2,2} \end{bmatrix} \right)$$

Alors, si $K_{1,1}$ est inversible, nous avons, conditionnellement à $y^{(1)} = v^1$, y_2 suit une loi de la forme (y_2 correspondant aux données recherchées) :

$$\mathcal{N}(m^{(2)}K_{2,1}K_{1,1}^{-1}(v^{(1)} - m^{(1)}), K_{2,2} - K_{2,1}K_{1,1}^{-1}K_{1,2})$$

6.2.1 • ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE

Il s'agit donc de maximiser la fonction de vraisemblance, afin de calibrer le modèle de manière à capturer au mieux la structure de dépendance spatiale des données. Pour un processus gaussien qui est donnée par :

$$\mathbf{L}(\psi) = \frac{1}{(2\pi)^{n/2}|\mathbf{K}_\psi|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{K}_\psi^{-1} \mathbf{y}\right) \quad (3)$$

où \mathbf{K}_ψ est la matrice de covariance construite à partir des paramètres ψ et des données. Pour trouver les estimations des paramètres, nous maximisons le logarithme de la vraisemblance.

6.2.2 • APPLICATION À NOS DONNÉES

Nous avons donc dans un premier temps pensé à appliquer le krigeage à la prédiction de données futures compte tenu des informations accessibles jusqu'au temps t .



FIGURE 18 – Prédiction du prix futur



FIGURE 19 – Zoom sur la prédiction

Cependant, nous comprenons rapidement que la méthode n'est pas adaptée à la prédiction future de prix, étant donné que le krigeage est spécifique à l'utilisation de prédiction de données intermédiaires lorsque l'on connaît les valeurs en un certain nombre de points. Nous avons donc tenté d'estimer le prix du cours des actions au cours d'une journée sachant la valeur de l'action à chaque heure.

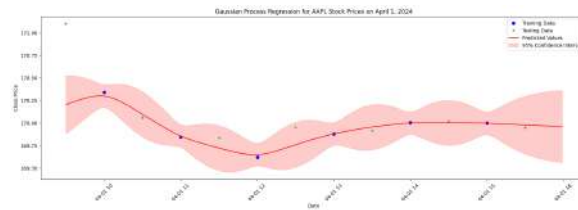


FIGURE 20 – Prédiction du prix toutes les minutes sachant le prix toutes les heures

Nous voyons donc ici que l'intervalle de confiance à 95% estimé par le processus gaussien et le kernel Matern permet de prendre en compte toutes les valeurs du cours du prix chaque minutes alors que l'on connaissait à priori uniquement les valeurs horaires de ces prix.

RÉFÉRENCES

- [1] Boaz Nadler AHARON BIRNBAUM Iain M. Johnstone et Debashis PAUL. “Minmax bounds for sparse PCA with noisy high-dimensional Data”. In : *NIH* (2013).
- [2] Antoine BERGERON. “Copules dynamiques pour séries chronologiques”. In : *Université de Sherbrooke* (2020).
- [3] Eric BOUYÉ. “Copulas for FinanceA Reading Guide and Some Applications”. In : *University Business School London* (2000).
- [4] Martin COUX. “Matériaux texturés activables”. Thèse de doct. Paris Sciences et Lettres (ComUE), 2018.
- [5] Thomas DESCHATRE. “On the control of the difference between two Brownian motions : a dynamic copula approach”. In : *De Gruyter* (2016).
- [6] Vladimir KOLTCHINSKII et Karim LOUNICI. “Concentration inequalities and moment bounds for sample covariance operator”. In : *PSchool of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA.* (2017).
- [7] Karim LOUNICI. “High-dimensional Principal Component Analysis”. In : *M2 Data Science Ecole Polytechnique* ().
- [8] Fadhilah Yusof NORIZZATI SALLEH et Zulkifli YUSOP. “Bivariate copulas functions for flood frequency analysis”. In : *AIP Conference Proceedings* (2016).
- [9] RiskHelder Parra PALARO et Luiz Koodi HOTTA. “Using Conditional Copula to Estimate Value at Risk”. In : *State University of Campinas* (2006).
- [10] Andrew PATTON. “Copula-Based Models for Financial Time Series”. In : *The Oxford-Man Institute* (2006).
- [11] Filippo PIZZOCCHERO et al. “The hot pick-up technique for batch assembly of van der Waals heterostructures”. In : *Nature communications* 7.1 (2016), p. 1-10.
- [12] Frédéric PLANCHET. “Modèles financiers en assurance et analyse dynamique”. In : *ISFA* (2023).
- [13] Volker SCHMITZ. “Copulas and Stochastic Processes”. In : *The Institute of Statistics of Aachen University* (2003).
- [14] Craig S. SMITH. “Top 10 Quantum Computing Companies Making Change”. In : *Forbes* (2023). URL : <https://www.forbes.com/sites/technology/article/top-quantum-computing-companies/?sh=1e1cc90b3a94>.
- [15] Vladimir V TSUKRUK et Srikanth SINGAMANENI. *Scanning probe microscopy of soft matter : fundamentals and practices*. John Wiley & Sons, 2012.
- [16] Roman VERSHYNIN. “High-Dimensional Probability”. In : *University of California, Irvine* (2020).



- [17] Roman VERSHYNIN. “Introduction to the non-asymptotic analysis of random matrices”.
In : *University of Michigan* (2011).

7 ANNEXE

Log transformation du Price closing Une transformation courante pour diminuer la tendance d'une série temporelle consiste à prendre le logarithme pour convertir une croissance exponentielle en une tendance linéaire et stabiliser la variance.

Nous avons donc tracé la transformation logarithmique de notre série temporelle pour voir si la tendance s'atténue.

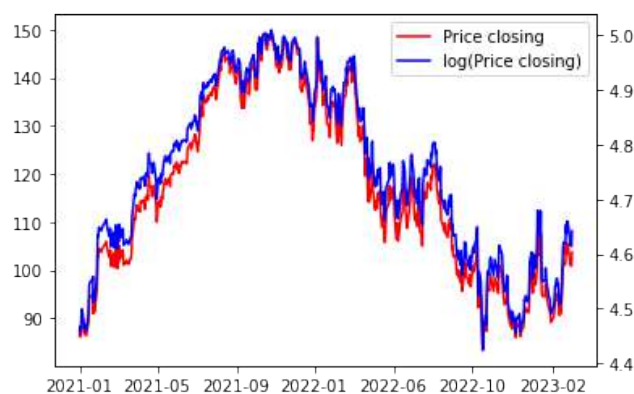


FIGURE 21 – transformation logarithmique

Nous pouvons constater qu'ici l'évolution de price closing ne semble pas clairement être modifiée en appliquant une transformation logarithmique, ceci pouvant être expliqué par le fait que l'amplitude des perturbations n'augmente pas de manière significative au cours du temps.

Tests d'autocorrélation et présence d'effet ARCH et GARCH Afin de comprendre s'il y a effectivement de tels effet au sein de nos données, il convient d'effectuer une analyse d'autocorrélation :

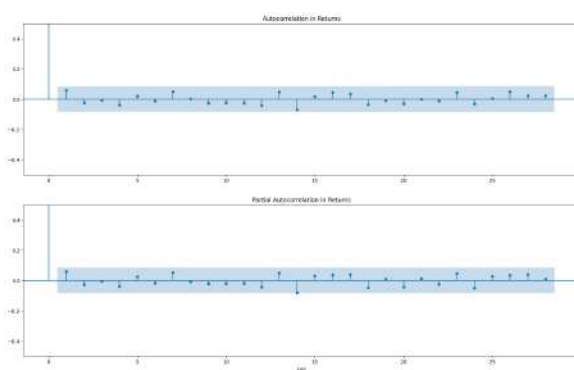


FIGURE 22 – Corrélation des retours

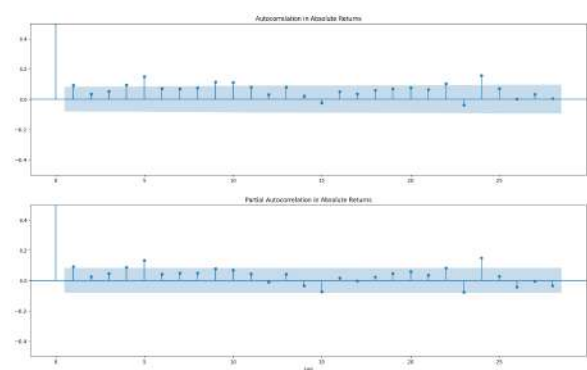


FIGURE 23 – Corrélation absolue

Les graphiques ACF (fonction d'autocorrélation) et PACF (fonction d'autocorrélation partielle) ci-dessus ne montrent pas la présence d'effets ARCH et GARCH. En particulier, nous voudrions des indications pour croire que la série n'est pas indépendante, ce qui serait visuellement perceptible par des retards pour lesquels les valeurs de l'ACF et de la PACF sont en dehors de la zone bleue.

Pour nous assurer qu'il s'agit d'une série indépendante, ou plutôt pour tester que ce n'est pas le cas, nous pouvons répéter le processus ci-dessus en utilisant les rendements au carré ou les rendements absolus. Ainsi, après la transformation précédente en rendements absolus, il y a maintenant un motif indicatif de la présence d'effets ARCH et GARCH. Dans la PACF, il n'y a pas une grande chute avant environ le retard 8, ce qui pourrait indiquer que nous devrions avoir un ordre de P,Q allant jusqu'à 8. Pour des raisons de simplicité et d'illustration, nous utiliserons un modèle GARCH(1,1) avant d'évaluer et de tester différentes spécifications. Nous testerons plus formellement ces effets en utilisant le test de McLeod-Li pour les effets ARCH et GARCH.

Distribution gaussienne L'expression de la densité de probabilité de la distribution gaussienne est donnée par :

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

où μ est la moyenne de la distribution et σ^2 est sa variance.

Distribution de Student L'expression de la densité de probabilité de la distribution de Student est donnée par :

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

où ν est le nombre de degrés de liberté de la distribution de Student et Γ représente la fonction gamma.

Krigeage Universel Considérons maintenant le cas du Krigeage ordinaire ou universel, où la fonction moyenne n'est plus considérée nulle : nous notons par H la matrice $n \times m$ telle que

$$H_{i,j} = g_j(x^{(i)})$$

où la fonction moyenne est supposée être de la forme $\sum_{i=1}^m \beta_i g_i$, avec des fonctions connues g_i et des coefficients inconnus β_i . Nous notons également par h le vecteur $m \times 1$ tel que $h_j = g_j(x_{\text{new}})$. Cependant, nous utilisons toujours explicitement la décomposition σ^2, θ . Par conséquent, nous présenterons les équations du maximum de vraisemblance uniquement dans ce cas.

Le critère de vraisemblance dépend maintenant de β, σ^2 et θ et est donné par :

$$L(\beta, \sigma^2, \theta) := \frac{1}{n} \ln |\sigma^2 R_\theta| + \frac{1}{n\sigma^2} (\mathbf{y} - H\beta)^t R_\theta^{-1} (\mathbf{y} - H\beta) \quad (4)$$

De manière similaire au cas du Krigage simple, le critère de vraisemblance peut être minimisé explicitement par rapport à β et σ^2 , en éliminant $m + 1 = \dim(\beta) + 1$ dimensions dans le problème d'optimisation numérique. Ceci est résumé dans la proposition suivante.

L'estimateur du maximum de vraisemblance de $(\hat{\beta}, \hat{\sigma}^2, \theta)$ est $(\hat{\beta}_{ML}, \hat{\sigma}_{ML}^2, \hat{\theta}_{ML})$, avec :

$$\hat{\theta}_{ML} \in \arg \min_{\theta \in \Theta} L(\theta), \quad (5)$$

avec

$$L(\theta) = \ln(\hat{\sigma}_{ML}^2(\theta)) + \frac{1}{n} \ln |R_\theta|,$$

$$\hat{\sigma}_{ML}^2(\theta) = \frac{1}{n} (\mathbf{y} - H\hat{\beta}_{ML}(\theta))^t R_\theta^{-1} (\mathbf{y} - H\hat{\beta}_{ML}(\theta))^t$$

$$\hat{\beta}_{ML}(\theta) = (H^t R_\theta^{-1} H)^{-1} H^t R_\theta^{-1} \mathbf{y}$$

et

$$\hat{\beta}_{ML} = \hat{\beta}_{ML}(\hat{\theta}_{ML}), \hat{\sigma}_{ML}^2 = \hat{\sigma}_{ML}^2(\hat{\theta}_{ML}).$$

De plus, on peut aussi écrire :

$$\hat{\sigma}_{ML}^2(\theta) = \frac{1}{n} \mathbf{y}^t \Pi_\theta \mathbf{y},$$

avec :

$$\Pi_\theta = R_\theta^{-1} - R_\theta^{-1} H (H^t R_\theta^{-1} H)^{-1} H^t R_\theta^{-1}$$

Test de Jarque-Bera Nous faisons une hypothèse dans le modèle GARCH(P, Q) sur la distribution de nos erreurs : si le modèle est correctement spécifié, alors les erreurs sont distribuées normalement avec une moyenne de 0 et une variance constante. Le test de Jarque-Bera vérifie si les erreurs du modèle sont distribuées normalement.

La formule pour la statistique de test est la suivante :

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4} (K - 3)^2 \right)$$

où n est la taille des données, S est l'asymétrie de l'échantillon (qui mesure la symétrie des données par rapport à la moyenne) et K est le kurtosis de l'échantillon (qui mesure la forme de la distribution, en particulier les queues).

Nous savons que le kurtosis de la distribution normale est de 3, ce qui est utile comme référence. S'il est supérieur à 3, nous avons un échantillon avec des queues plus épaisses que la distribution normale. S'il est inférieur à 3, nous avons un échantillon avec des queues plus minces que la distribution normale.

Nous pouvons comprendre que l'utilisation de la symétrie et de la forme de l'échantillon par rapport aux valeurs connues de la distribution normale nous aiderait à arriver à l'hypothèse suivante :

- H_0 : Les données suivent une distribution normale.
- H_a : Les données ne suivent pas une distribution normale.

Nous obtenons à l'issue du test un p-valeur de 10^{-17} qui nous oblige à rejeter le modèle de distribution normale, il convient d'en essayer d'analyser des distributions plus spécifiques. Le modèle n'est donc pas valide dans notre cas, il faudrait envisager des distributions plus spécifiques et adaptées, comme par exemple les distributions de student qui ont des queues plus "épaisses" tout en étant similaires à la loi normale et permettent de mieux cerner les évolutions de retours tels que les nôtres pour lesquels les valeurs extrêmes ont une plus grande probabilité.

Noyau et fonction de covariance Dans le cadre des processus gaussiens, le choix du noyau (ou *kernel*) est crucial car il détermine la fonction de covariance du processus, influençant ainsi la nature de la corrélation entre les points de données. Différents noyaux peuvent être utilisés, chacun ayant des propriétés et des applications spécifiques :

1. **Noyau Linéaire** :, défini par $K(x, x') = x^\top x'$,
2. **Noyau Polynomial** : défini par $K(x, x') = (x^\top x' + c)^d$, où c et d
3. **Noyau Gaussien (RBF - Radial Basis Function)** : défini par $K(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$, où σ
4. **Noyau Exponentiel** : défini par $K(x, x') = \exp\left(-\frac{\|x-x'\|}{\sigma}\right)$
5. **Noyau Matérn** : défini par $K(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\|x-x'\|}{\rho}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}\|x-x'\|}{\rho}\right)$, où ν et ρ

Fonction de covariance : Chaque élément K_{ij} de la matrice \mathbf{K} est calculé en appliquant le noyau sur la paire de points (x_i, x_j) , c'est-à-dire $K_{ij} = k(x_i, x_j)$. Ainsi, la matrice de covariance \mathbf{K} est définie comme suit :

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

7.0.1 • ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE

Il s'agit donc de maximiser la fonction de vraisemblance, afin de calibrer le modèle de manière à capturer au mieux la structure de dépendance spatiale des données. Pour un processus gaussien qui est donnée par :

$$\mathbf{L}(\psi) = \frac{1}{(2\pi)^{n/2} |\mathbf{K}_\psi|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^\top \mathbf{K}_\psi^{-1} \mathbf{y}\right) \quad (6)$$

où \mathbf{K}_ψ est la matrice de covariance construite à partir des paramètres ψ et des données. Pour trouver les estimations des paramètres, nous maximisons le logarithme de la vraisemblance :

$$\ln \mathbf{L}(\psi) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \ln |\mathbf{K}_\psi| - \frac{1}{2} \mathbf{y}^\top \mathbf{K}_\psi^{-1} \mathbf{y} \quad (7)$$