

Mémoire
Prévision de l'indice CAC 40

DABERE Abasse, EDDAOUDI Oussama

Avril 2024

Table des matières

1	Introduction	2
2	Visualisation des données	3
3	Analyse exploratoire des données	3
3.1	Analyse statistique	3
3.2	Analyse en Composantes Principales	4
3.3	Test de stationnarité	4
3.4	Fonction d'autocorrélation et autocorrélation partielle	5
4	Répartition des données et prévisions	6
4.1	Résultats ARIMAX(0,1,1)	7
4.2	Résultats XGBoost	7
4.3	Comparaison	8
5	Utilisation des copules	8
5.1	Analyse Statistique des Rendements	8
5.2	Tests de Stationnarité	9
5.3	Analyse de Corrélation	10
5.4	Transformation des Marginales	11
5.5	Sélection et Évaluation des Copules	11
5.6	Analyse Quantitative et Interprétation des Résultats	13
5.7	Conclusion	15
6	Processus de Hawkes	15
6.1	Définition	15
6.2	Évaluation de la qualité du modèle	15
6.3	Ajustement et résultats	16
7	Conclusion	16

1 Introduction

Les marchés financiers, en raison de leur complexité et de leur dynamisme, présentent un défi significatif pour la modélisation et la prévision des prix des actions. Cette complexité a suscité un intérêt croissant parmi les chercheurs et les praticiens, conduisant à une exploration approfondie des méthodes et des modèles. Une approche classique, l'ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables), a longtemps été un pilier dans l'analyse des séries temporelles financières. Cette méthode offre une base solide en capturant les tendances et les motifs saisonniers.

Dans cette mémoire, nous allons alors explorer l'utilisation de cette méthode sophistiquée pour prédire le cours de l'indice CAC 40 de la bourse de Paris. Pour ce faire, nous allons travailler sur le cours de celui-ci sur la période allant de janvier 2014 jusqu'à avril 2024. Nous comparerons ensuite les résultats de prévision avec ceux d'un autre algorithme réputé en machine learning, à savoir XGBoost. Nous utiliserons ensuite les copules pour explorer la corrélation deux à deux entre les actions de LVMH, TotalEnergies et BNP Paribas, représentant divers secteurs clés.

En conclusion, nous examinerons le rôle croissant de la cryptomonnaie dans l'économie mondiale au cours des dernières années, en mettant particulièrement en lumière le Bitcoin. Depuis son lancement en 2009 par son créateur anonyme Satoshi Nakamoto, le Bitcoin a connu une appréciation inattendue par rapport au dollar américain. Nous aborderons alors l'application des processus de Hawkes pour modéliser les flux d'ordres du Bitcoin dans cette perspective.

Données

Dans le cadre de notre étude, nous avons choisi de nous appuyer sur l'évolution de l'indice CAC 40, tel que décrit précédemment dans l'introduction. Contrairement au modèle ARIMA traditionnel, ARIMAX offre la possibilité d'inclure davantage de variables exogènes, enrichissant ainsi la modélisation.

Ces variables exogènes, ou facteurs extérieurs, sont supposées avoir un impact sur les fluctuations de l'indice boursier. Pour notre étude, nous avons sélectionné les cours des actifs de plusieurs entreprises, toutes faisant partie intégrante de la composition du CAC 40. Ces entreprises représentent une diversité de secteurs économiques clés et leur performance peut influencer l'évolution de l'indice.

Les entreprises sélectionnées et les liens vers leurs données sont résumés ci-dessous :

- CAC 40, Crédit Agricole S.A., Carrefour S.A.
- Airbus SE, Atos SE, Danone S.A
- BNP Paribas SA, Capgemini SE, Vinci SA
- Engie SA, Thales S.A., Kering SA
- Orange S.A., Pernod Ricard SA, Saint-Gobain
- Vivendi SE, LVMH, TotalEnergies
- Bitcoin

2 Visualisation des données

Au cœur de cette analyse réside notre ambition d’explorer la dynamique changeante de l’indice CAC 40, tout en tenant compte des indices sectoriels contenus dans notre base de données. La figure ci-dessous illustre de manière éloquente cette évolution sur la période examinée.

La série chronologique couvre la période allant de 2014 à 2024, reflétant une amplitude variant entre 3000 et 9000. Une tendance haussière se dessine, ponctuée par une variabilité prononcée, particulièrement visible au début de l’année 2020 en raison de la crise sanitaire liée à la COVID-19. Cette série témoigne également d’une phase de reprise, marquée par un pic atteint au début de l’année 2022, suivi d’une décroissance post-2022, et enfin d’une reprise à partir de 2023.



FIGURE 1 – Évolution de l’indice CAC 40

3 Analyse exploratoire des données

3.1 Analyse statistique

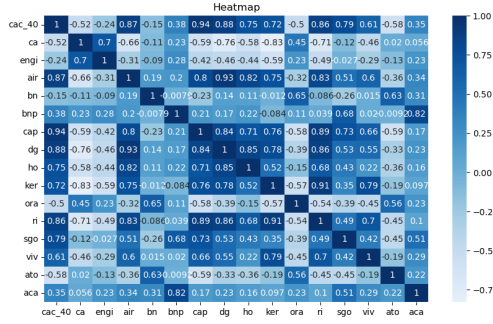
Le tableau ci-dessous résume les principales statistiques de 6 variables de nos données, incluant le nombre d’observations, la moyenne, l’écart-type, les valeurs minimale et maximale, les quartiles.

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
cac_40	2625.0	5525.71	1002.59	3754.84	4719.17	5319.17	6285.24	8205.81
ca	2625.0	19.57	4.97	12.33	15.83	17.35	23.20	32.79
engi	2625.0	14.22	2.40	8.70	12.55	13.82	15.03	21.09
air	2625.0	88.89	29.78	40.54	60.41	91.98	113.70	171.60
bn	2625.0	60.59	7.00	46.83	54.78	60.13	65.04	82.00
bnp	2625.0	51.92	8.66	24.67	46.00	52.66	58.05	68.89

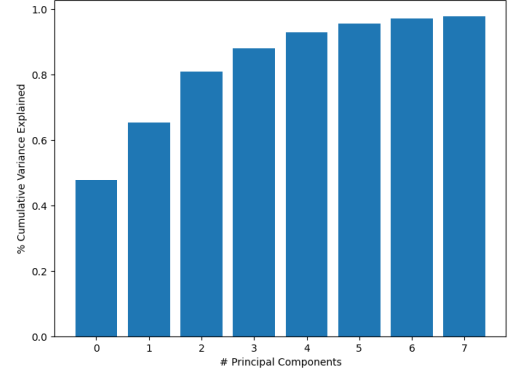
TABLE 1 – Statistiques résumées des variables

3.2 Analyse en Composantes Principales

Au cours de notre traitement des données, nous avons observé une forte corrélation entre plusieurs composantes du CAC 40, ce que nous avons pu visualiser facilement à l'aide du heatmap tracé ci-dessous. Parmi ces composantes, nous avons relevé notamment une corrélation significative entre des entreprises telles que Crédit Agricole et BNP Paribas. Dans le but de réduire cette corrélation et d'optimiser l'exécution du modèle, nous avons envisagé d'appliquer une Analyse en Composantes Principales (PCA) à nos données.



(a) Heatmap



(b) Cumulative Explained Variance

FIGURE 2 – Principal Component Analysis (PCA)

Nous avons sélectionné les 8 premières composantes principales, car elles peuvent expliquer environ 98% de la variance des données.

3.3 Test de stationnarité

Avant de mettre en œuvre le modèle SARIMAX, il est crucial de déterminer les valeurs de (p, d, q). Ces paramètres représentent respectivement l'ordre d'autorégression, définissant le nombre de valeurs passées qui influencent significativement la valeur actuelle de la série temporelle. Le paramètre d, quant à lui, reflète l'ordre de différenciation, permettant d'éliminer toute tendance prédominante de la série et de la rendre stationnaire. Enfin, le paramètre q correspond à l'ordre de la moyenne mobile, indiquant le nombre d'erreurs passées qui influent de manière significative sur la valeur actuelle de la série.

Dans cette étape, nous utiliserons deux tests statistiques pour évaluer la stationnarité de la série CAC 40 : le test Augmented Dickey Fuller (ADF) et le test Kwiatkowski–Phillips–Schmidt–Shin (KPSS).

Les hypothèses du test ADF sont les suivantes : H_0 : la série n'est pas stationnaire contre H_a : la série est stationnaire. En revanche, Les hypothèses du test KPSS sont les suivantes : H_0 : la série est stationnaire contre H_a : la série est non-stationnaire.

Test Statistique	Statistique	p-value	Stationnarité
ADF	-0.794	0.82	Non stationnaire
KPSS	6.751	0.01	Non stationnaire

TABLE 2 – Résultats des tests de stationnarité

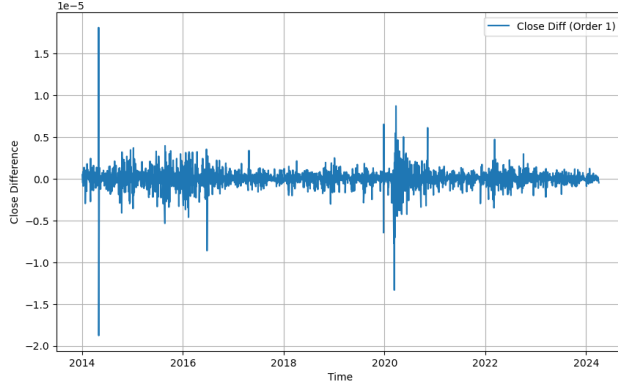


FIGURE 3 – Différence d'ordre 1 du CAC 40

Test	Statistique	p-value	Stationnarité
ADF	-57.324	0.0	Stationnaire
KPSS	0.055	0.1	Stationnaire

FIGURE 4 – Résultats des tests

Les tests nous confirment que la série est bel et bien stationnaire. Nous pouvons donc passer à l'étape suivante de la sélection de notre modèle. Celle-ci consiste à déterminer les ordres p et q , respectivement de l'autorégression et de la moyenne mobile, que nous utiliserons pour prédire les tendances du CAC 40.

3.4 Fonction d'autocorrélation et autocorrélation partielle

La fonction d'autocorrélation (ACF) est une mesure statistique qui évalue la corrélation linéaire entre les observations d'une série temporelle décalées de différentes périodes, appelées "lags". Elle est généralement représentée mathématiquement par la formule suivante :

$$R(k) = \frac{\sum_{t=k+1}^n (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

Cette formule calcule la corrélation entre les observations à un décalage k de la série temporelle. Plus précisément, elle mesure la corrélation entre les valeurs actuelles et les valeurs décalées de k périodes en arrière.

La fonction d'autocorrélation partielle (PACF), quant à elle, évalue la corrélation linéaire entre les observations d'une série temporelle, en tenant compte de l'influence des observations situées entre elles. Elle est définie mathématiquement par la formule suivante :

$$\phi_{k,k} = \frac{\gamma_{k,k}}{\sqrt{\gamma_{0,0}\gamma_{k,k}}}$$

Dans cette formule, $\gamma_{k,k}$ représente l'ACF de lag k , c'est-à-dire la corrélation entre les observations à un décalage de k périodes, $\gamma_{0,0}$ est la variance de la série, et $\phi_{k,k}$ est la PACF de lag k . La PACF permet d'isoler la corrélation directe entre les observations à un lag particulier, en éliminant l'influence des observations intermédiaires.

Nous avons alors tenté de tracer les courbes ACF et PACF pour la série temporelle de l'indice CAC 40, ainsi que pour sa différence d'ordre 1. Voici les résultats que nous avons obtenus :

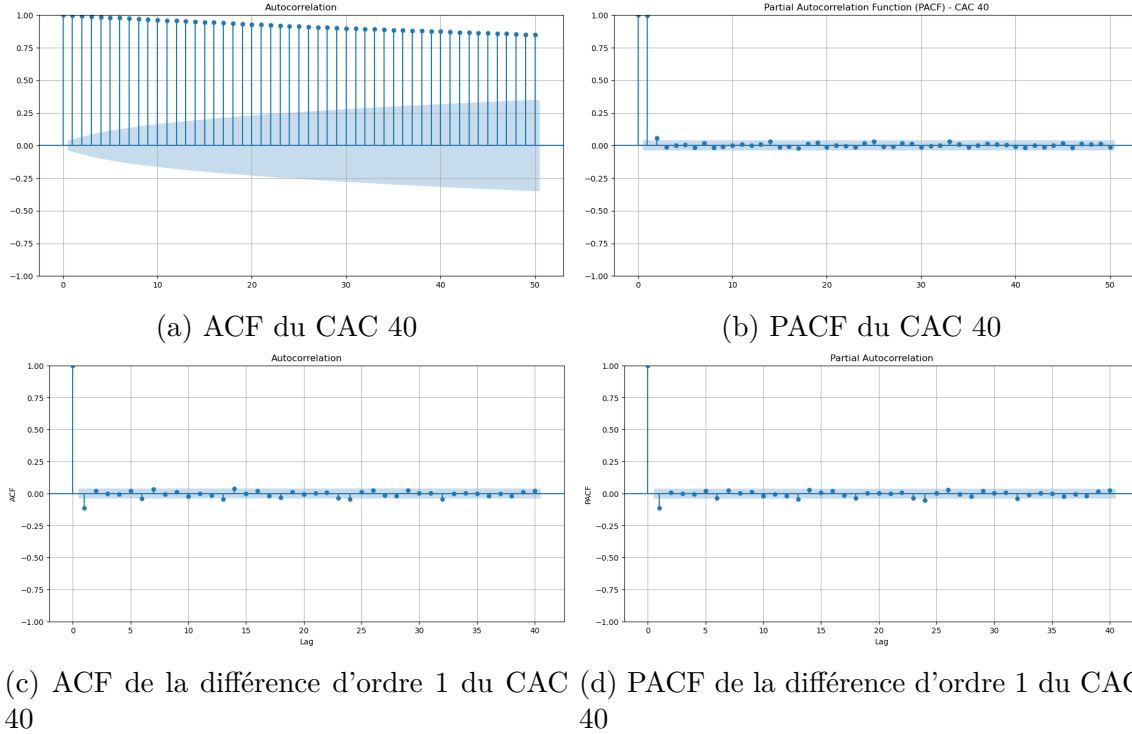


FIGURE 5 – ACF et PACF

ACF (Auto-Correlation Function) : La décroissance rapide des autocorrélations dans la fonction ACF suggère l'absence d'une composante saisonnière significative. Cependant, quelques pics résiduels peuvent indiquer des dépendances temporelles à court terme.

PACF (Partial Auto-Correlation Function) : Des pics significatifs dans la PACF aux premiers retards peuvent indiquer la présence d'une composante $AR(p)$ potentielle dans le modèle ARIMA.

4 Répartition des données et prévisions

Dans cette section, nous allons tout d'abord répartir les données dont nous disposons en données d'entraînement et en données de test. Nous prendrons 80% des données disponibles comme données d'entraînement et le reste comme données de test. Ensuite, nous examinerons la performance de deux modèles : ARIMAX(1,1,0) et XGBoost, en termes de prévisions. Nous avons également testé d'autres variantes de ARIMAX, notamment ARIMAX(0,1,1) et ARIMAX(1,1,1), mais elles produisent précisément les mêmes résultats que ARIMAX(0,1,1). Par conséquent, nous ne présenterons que cette dernière et la comparerons avec XGBoost.

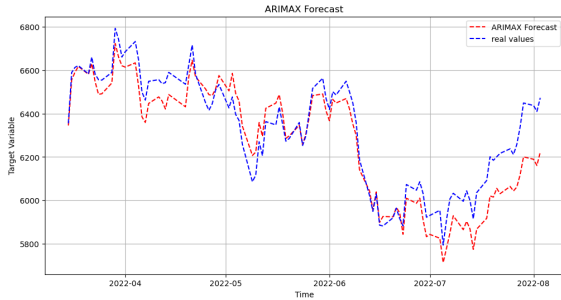


FIGURE 6 – Répartition des données en ensembles d'entraînement et de test

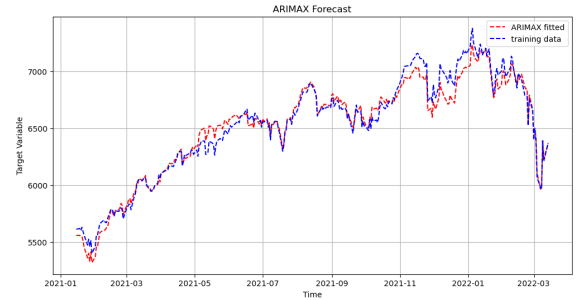
4.1 Résultats ARIMAX(0,1,1)

Nous allons désormais présenter les résultats prévisionnels obtenus par le modèle SARIMAX. Dans un premier temps, nous illustrons le comportement du modèle sur une partie des données d'entraînement, couvrant la période de janvier 2021 à mars 2023. On peut observer que le modèle se comporte généralement bien sur ces données d'entraînement, parvenant ainsi à capturer la variance et la tendance du CAC 40.

Nous allons ensuite examiner le comportement du modèle sur une partie des données de test, allant d'avril 2022 à août 2022. Nous constatons que le modèle parvient également à capturer la tendance et la variance de la série, démontrant ainsi une capacité de généralisation satisfaisante.



(a) Capacité à généraliser



(b) Données d'entraînement ajustées

FIGURE 7 – Prévisions sur les données test et ajustement sur les données d'entraînement

4.2 Résultats XGBoost

Finalement, nous allons comparer les prévisions effectuées par le modèle XGBoost. Nous nous limiterons à tracer uniquement le comportement sur les données de test sur la même période que ARIMAX.

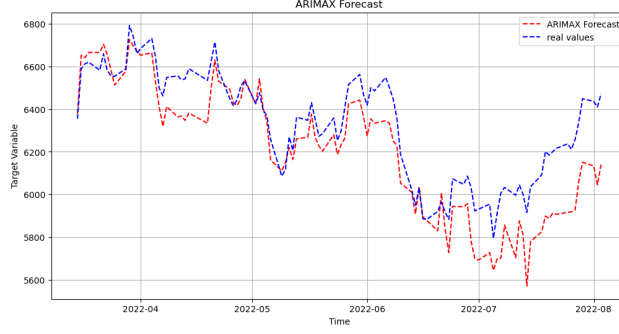


FIGURE 8 – Comportement de XGBoost

4.3 Comparaison

Enfin, voici deux tableaux résumant différentes métriques d'évaluation des deux modèles utilisés :

	Entraînement			Test		
	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE
XGBoost	28.13	0.34	17.48	569.38	6.75	486.40
SARIMAX	125.87	0.37	18.61	541.62	6.44	464.54

TABLE 3 – Performances des modèles

Les deux modèles présentent généralement de bonnes performances en termes de prévisions et parviennent à généraliser les données, bien que SARIMAX semble avoir une légère avance sur XGBoost en termes de généralisation, avec une capacité à prédire des données sur lesquelles il n'a pas été entraîné. De plus, SARIMAX se distingue par un temps d'exécution considérablement plus court que XGBoost, prenant seulement quelques secondes comparativement à environ une minute pour ce dernier. En conclusion, tant SARIMAX que XGBoost se révèlent être des modèles de prévision efficaces pour les séries temporelles, avec une légère préférence pour SARIMAX qui semble mieux adapté à ce type de données.

5 Utilisation des copules

Les copules analysent la dépendance entre les rendements des actifs, essentielles pour comprendre les marchés financiers. Notre étude sur le CAC40 emploie ces outils pour explorer la corrélation deux à deux entre LVMH, TotalEnergies et BNP Paribas, représentant divers secteurs clés. Ceci offre une vue globale des interactions du marché et des secteurs, aidant à modéliser leur interdépendance et leur impact sur la performance du CAC40.

5.1 Analyse Statistique des Rendements

Le rendement journalier d'une action est calculé en utilisant la formule suivante :

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

où R_t est le rendement au temps t , P_t est le prix de l'action au temps t , et P_{t-1} est le prix de l'action au temps $t - 1$.

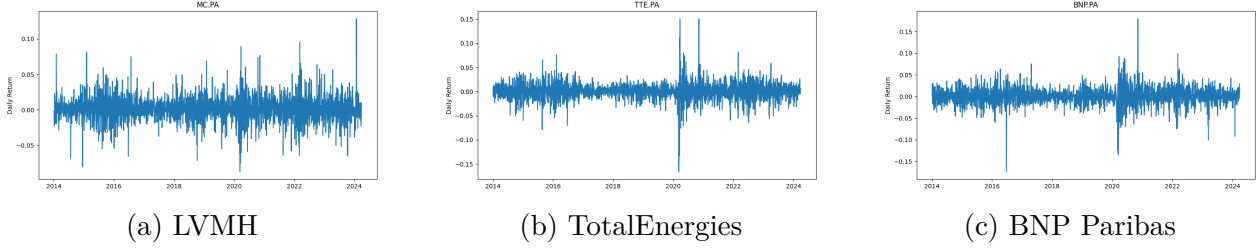


FIGURE 9 – rendements journaliers

Les statistiques descriptives, incluant la moyenne, l'écart type, l'asymétrie (skewness), et le kurtosis des rendements journaliers sont présentées dans le tableau suivant :

Statistique	LVMH	TotalEnergies	BNP Paribas
Nombre d'observations	2623	2623	2623
Moyenne	0.000918	0.000523	0.000436
Écart type	0.017074	0.017369	0.019796
Asymétrie	0.22899	-0.21249	-0.30631
Kurtosis	4.19995	12.57173	8.79411

TABLE 4 – Statistiques descriptives des rendements journaliers

Les variations observées en termes de volatilité, d'asymétrie et de kurtosis suggèrent des caractéristiques uniques pour chaque action. Ces différences soulignent l'importance d'explorer diverses copules pour une modélisation précise de la dépendance. L'asymétrie, tant positive que négative, indique des dépendances potentiellement non linéaires, rendant les copules non gaussiennes appropriées. De plus, un kurtosis élevé révèle la présence de comportements extrêmes, ce qui suggère l'efficacité des copules comme Clayton ou Gumbel pour modéliser ces caractéristiques.

5.2 Tests de Stationnarité

Nous avons appliqué le test ADF aux séries temporelles des rendements journaliers pour les actions 'MC.PA', 'TTE.PA', et 'BNP.PA'. Le test fournit plusieurs résultats clés, dont la statistique de test et la valeur P.

Action	Statistique ADF	Valeur P
LVMH	-53.4434	0.0
TotalEnergies	-13.9758	4.26e-26
BNP Paribas	-12.9113	4.05e-24

TABLE 5 – Résultats du Test Augmenté de Dickey-Fuller

Pour chacune des séries de rendements, la statistique ADF est significativement inférieure aux valeurs critiques pour les seuils de 1%, 5% et 10%, et les valeurs P sont extrêmement faibles. Cette forte évidence de stationnarité permet de rejeter l'hypothèse nulle de présence d'une racine unitaire. Ainsi, ces résultats valident l'application de méthodes statistiques et économétriques ultérieures qui supposent ou requièrent la stationnarité des séries temporelles, telle que la modélisation des dépendances à l'aide de copules.

5.3 Analyse de Corrélation

Corrélation Linéaire - Coefficient de Pearson : Le coefficient de Pearson, noté r , mesure la relation linéaire entre deux variables. Il est calculé comme suit :

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

où X_i et Y_i représentent les valeurs des deux variables et \bar{X} , \bar{Y} leurs moyennes respectives.

Corrélation de Rang - Tau de Kendall : Le tau de Kendall, noté τ , est calculé comme :

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sgn}(X_i - X_j) \text{sgn}(Y_i - Y_j)$$

où sgn est la fonction signe, X_i et Y_i sont les valeurs des observations.

Corrélation de Rang - Rho de Spearman : Le rho de Spearman, noté ρ , est défini comme :

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

où d_i est la différence entre les rangs de chaque observation des deux variables.

Les tableaux suivants résument les résultats des corrélations de Pearson, de Kendall et de Spearman pour les rendements journaliers de nos actions.

Coefficient	LVMH	TotalEnergies	BNP Paribas
Pearson LVMH	1.000	0.4578	0.5294
Pearson TotalEnergies	0.4578	1.000	0.6058
Pearson BNP Paribas	0.5294	0.6058	1.000
Kendall LVMH	1.000	0.2786	0.3306
Kendall TotalEnergies	0.2786	1.000	0.3723
Kendall BNP Paribas	0.3306	0.3723	1.000
Spearman LVMH	1.000	0.3975	0.4719
Spearman TotalEnergies	0.3975	1.000	0.5213
Spearman BNP Paribas	0.4719	0.5213	1.000

TABLE 6 – Résultats de Corrélation

Les résultats de corrélation entre les actions LVMH, TotalEnergies et BNP Paribas révèlent plusieurs aspects importants :

- **Corrélation de Pearson** : Nous observons des corrélations allant de modérées à élevées. En particulier, la corrélation entre LVMH et TotalEnergies (0.46) ainsi qu'entre BNP Paribas et TotalEnergies (0.61) suggère une synchronisation notable dans leurs mouvements.
- **Corrélation de Kendall** : Les valeurs sont généralement plus basses que celles de Pearson, indiquant des corrélations plus conservatrices. Cela est particulièrement notable entre LVMH et TotalEnergies, ainsi qu'entre BNP Paribas et TotalEnergies.
- **Corrélation de Spearman** : Les tendances sont similaires à celles observées avec Pearson, quoique légèrement inférieures. Cela confirme les corrélations observées, particulièrement entre LVMH et TotalEnergies ainsi qu'entre BNP Paribas et TotalEnergies.

Ces divers degrés de corrélation impliquent qu'une sélection appropriée de copules est nécessaire pour modéliser efficacement les dépendances entre ces actions. Les copules telles que Clayton ou Gumbel, qui modélisent bien les dépendances dans les événements extrêmes, peuvent être particulièrement pertinentes pour ces séries temporelles.

5.4 Transformation des Marginales

La transformation des marginales, essentielle en analyse des copules, convertit les données en marginales uniformes $[0, 1]$ pour modéliser la dépendance entre variables. Elle transforme les rendements en quantiles via la fonction de distribution empirique (CDF).

La transformation a été appliquée aux rendements des actions telles que LVMH, TotalEnergies, et BNP Paribas. Les rendements sont transformés en leurs quantiles correspondants dans leur propre distribution, ce qui facilite une analyse cohérente et normalisée dans le cadre des modèles de copules.

5.5 Sélection et Évaluation des Copules

5.5.1 Présentation des Copules Choisies

La sélection des copules pour notre étude est fondée sur plusieurs critères. Premièrement, les caractéristiques statistiques des rendements, telles que la diversité, l'asymétrie, et les kurtosis élevés, guident le choix des copules appropriées. Deuxièmement, une analyse des corrélations de Pearson, Kendall et Spearman révèle différents niveaux de dépendance entre les paires de stocks, orientant vers des copules spécifiques comme Clayton, Gumbel pour les dépendances asymétriques et extrêmes, et la Gaussienne pour les corrélations linéaires plus fortes. Enfin, la stationnarité des séries de rendements, confirmée par le Test Augmenté de Dickey-Fuller, valide l'utilisation des copules choisies.

- **Copule de Clayton** : La copule de Clayton est particulièrement efficace pour modéliser la dépendance dans les queues inférieures des distributions. Sa fonction de copule

est donnée par :

$$C(u, v) = \max \left(u^{-\theta} + v^{-\theta} - 1, 0 \right)^{-1/\theta}$$

où $\theta > 0$ contrôle l'intensité de la dépendance.

- **Copule de Gumbel** : La copule de Gumbel est adaptée pour capturer la dépendance dans les queues supérieures. Sa fonction de copule est :

$$C(u, v) = \exp \left[- \left((-\log u)^\theta + (-\log v)^\theta \right)^{1/\theta} \right]$$

avec $\theta \geq 1$.

- **Copule Gaussienne** : La copule Gaussienne, quant à elle, modélise la dépendance dans la partie centrale de la distribution conjointe et est définie par :

$$C(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

où Φ_ρ est la fonction de distribution conjointe de deux variables normales standard avec corrélation ρ .

Les paramètres spécifiques des copules relatifs à nos données sont récapitulés dans le tableau ci-dessous, dérivés en utilisant la bibliothèque `copulas.bivariate`.

Paire d'Actions	Clayton	Gumbel	Gaussienne
LVMH - TotalEnergies	0.7724	1.3862	0.4324
LVMH - BNP Paribas	0.9878	1.4939	0.5095
TotalEnergies - BNP Paribas	1.1862	1.5931	0.5646

TABLE 7 – Paramètres des copules pour les paires d'actions sélectionnées

5.6 Analyse Quantitative et Interprétation des Résultats

5.6.1 Analyse Quantitative

Les performances des différentes copules ont été évaluées sur la base de la log-vraisemblance, de l'AIC et du BIC. Les résultats indiquent que les copules de Clayton et Gumbel ont des performances supérieures à celles de la copule Gaussienne. Les détails pour chaque paire d'actions sont les suivants :

Copule	Paire d'Actions	Log-Likelihood	AIC / BIC
Clayton	MC.PA - TTE.PA	262.73	-523.45 / -517.58
Clayton	MC.PA - BNP.PA	349.74	-697.47 / -691.60
Clayton	TTE.PA - BNP.PA	455.14	-908.28 / -902.41
Gumbel	MC.PA - TTE.PA	255.63	-509.25 / -503.38
Gumbel	MC.PA - BNP.PA	352.39	-702.79 / -696.92
Gumbel	TTE.PA - BNP.PA	477.87	-953.74 / -947.86
Gaussienne	MC.PA - TTE.PA	-7209.93	14421.85 / 14427.73
Gaussienne	MC.PA - BNP.PA	-7087.22	14176.44 / 14182.31
Gaussienne	TTE.PA - BNP.PA	-6977.87	13957.74 / 13963.61

La log-vraisemblance positive pour les copules de Clayton et Gumbel indique un meilleur ajustement par rapport à un modèle de base sans relation. Les valeurs AIC et BIC plus basses (plus négatives) pour ces copules suggèrent également un meilleur ajustement. Par exemple, pour la paire 'TTE.PA - BNP.PA', la copule de Clayton présente les meilleures valeurs d'AIC et de BIC, indiquant le meilleur ajustement parmi les modèles de Clayton. En revanche, les valeurs de log-vraisemblance négatives et significativement inférieures pour la copule Gaussienne indiquent un mauvais ajustement pour ces données.

5.6.2 Analyse Graphique

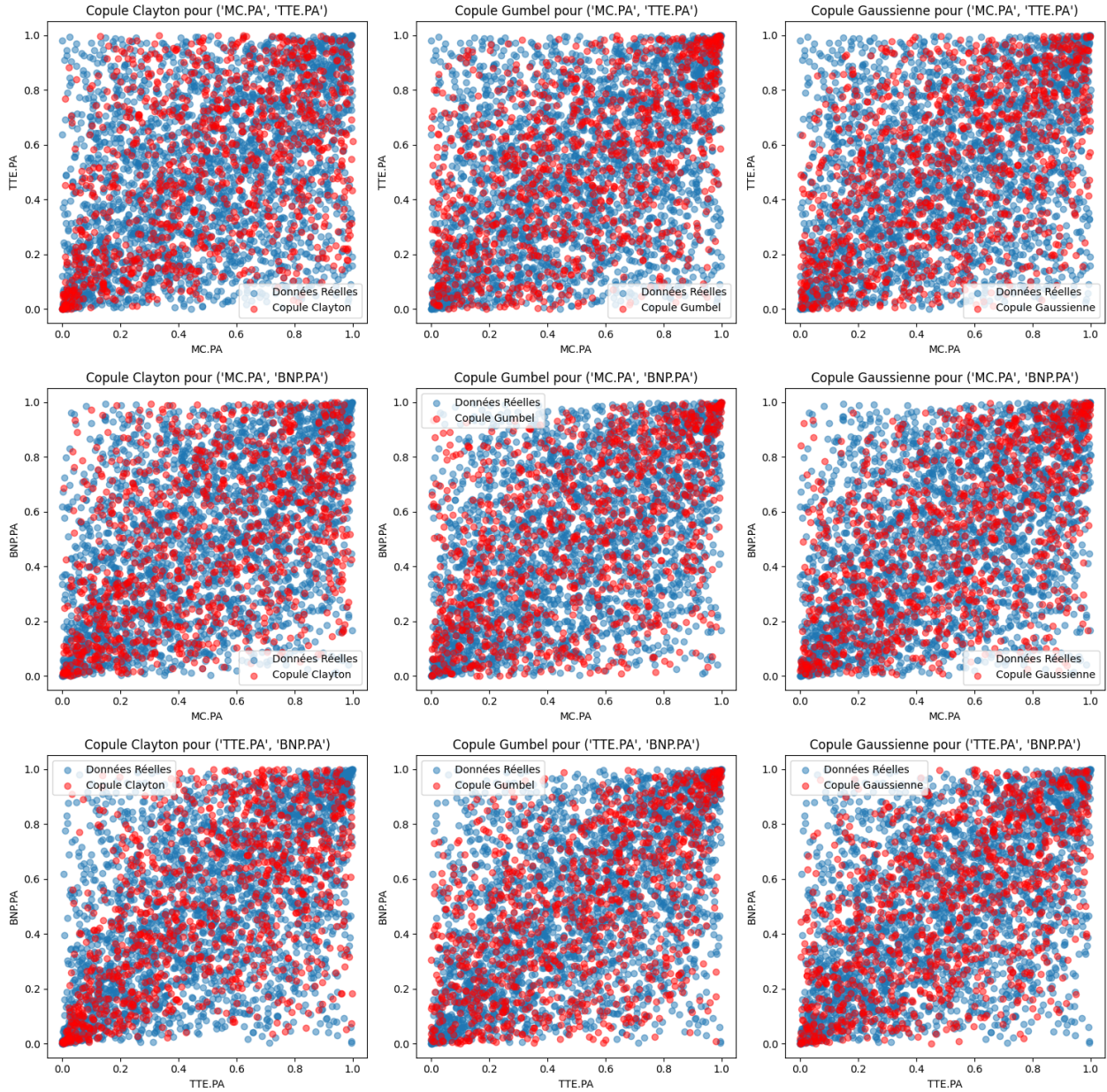


FIGURE 10 – Distributions conjointes des données simulées et réelles.

- **Pour la paire MC.PA et TTE.PA :** La copule Gumbel montre un meilleur chevauchement entre les données simulées et réelles, ce qui indique une dépendance plus forte lorsque les rendements sont élevés. Ceci est cohérent avec nos statistiques AIC et BIC.
- **Pour la paire TTE.PA et BNP.PA :** De même, la copule Gumbel semble offrir un ajustement supérieur, avec des données simulées qui reflètent plus fidèlement la distribution des données réelles, en particulier dans les valeurs de rendements élevés.

- **Pour la paire MC.PA et BNP.PA :** La copule Clayton semble fournir une meilleure représentation de la dépendance dans les rendements faibles.

5.7 Conclusion

Les copules sélectionnées pour notre modèle, basées sur l'analyse graphique et quantitative, sont :

- **Copule de Gumbel** pour les paires 'MC.PA - TTE.PA' et 'TTE.PA - BNP.PA', choisie pour sa performance supérieure démontrée par des valeurs plus basses d'AIC et BIC et sa capacité à mieux capturer les dépendances dans les queues supérieures.
- **Copule de Clayton** pour la paire 'MC.PA - BNP.PA', sélectionnée en raison de sa meilleure adéquation pour modéliser la dépendance dans les queues inférieures, confirmée par l'analyse graphique et des valeurs d'AIC et BIC favorables.

Ce choix reflète la capacité des copules à modéliser avec précision et robustesse les spécificités des dépendances observées dans les données financières.

6 Processus de Hawkes

Les processus de Hawkes sont particulièrement pertinents dans la modélisation des événements financiers, où ils permettent d'étudier les séquences d'arrivées d'ordres avec une autocorrélation temporelle. Ces processus décrivent la manière dont les événements antérieurs influencent la probabilité d'occurrences futures, par le biais d'une fonction d'excitation qui ajuste l'intensité basée sur l'historique des événements.

6.1 Définition

Dans le cadre des données du Bitcoin, nous considérons $N(t)$ comme le nombre d'ordres exécutés jusqu'au temps t . La fonction d'intensité conditionnelle $\lambda(t | H_t)$ est définie par :

$$\lambda(t | H_t) = \mu + \sum_{i=1}^{N(t)} \phi(t - t_i)$$

Ici, μ représente l'intensité de base des arrivées d'ordres, et $\phi(t - t_i)$ est la fonction d'excitation déterminant l'impact de chaque ordre antérieur sur la dynamique des arrivées futures.

6.2 Évaluation de la qualité du modèle

Les temps transformés d'occurrence d'ordres, t'_i , obtenus par l'intégration de la fonction d'intensité conditionnelle, devraient suivre une distribution uniforme si le modèle est correct. La qualité de l'ajustement peut être évaluée à l'aide du test de Kolmogorov-Smirnov, en vérifiant l'uniformité des t'_i sur l'intervalle pertinent.

6.3 Ajustement et résultats

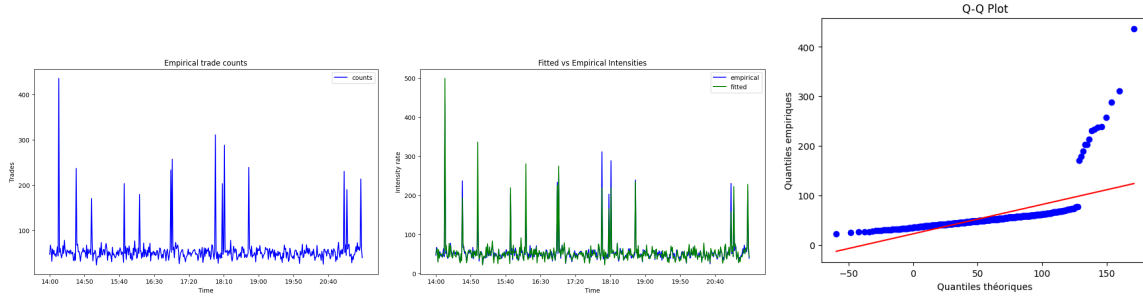
Notre jeu de données comprend les horodatages des ordres du Bitcoin. En nous concentrant sur des fenêtres temporelles où l'intensité de base est supposée constante, nous ajustons un modèle de Hawkes en utilisant la bibliothèque Python dédiée.

La fonction d'excitation adoptée est de forme exponentielle avec les paramètres α et β :

$$\phi(t) = \alpha e^{-\beta t}$$

Les paramètres μ , α et β sont estimés par maximum de vraisemblance, donnant les valeurs (0.1, 1, 0.5) pour notre période d'étude.

Les figures suivantes illustrent les résultats de l'ajustement :



Suite à l'ajustement des paramètres, nous observons un profil d'intensité qui montre un effet de regroupement caractéristique des processus de Hawkes, ce qui est confirmé par le tracé Q-Q des résidus. Cette observation met en évidence la présence significative de l'auto-excitation dans la dynamique des ordres du Bitcoin, soulignant l'importance de considérer des phénomènes tels que l'influence des événements passés sur les événements futurs dans la modélisation des marchés financiers, en particulier dans le contexte des cryptomonnaies.

7 Conclusion

Ce mémoire a exploré en profondeur diverses méthodes quantitatives et modèles statistiques pour analyser et prédire les dynamiques des marchés financiers. L'application de modèles ARIMAX et XGBoost sur l'indice CAC 40 a révélé leurs capacités prédictives robustes, avec une légère préférence pour le modèle SARIMAX en termes de performance et de temps d'exécution. L'usage innovant des copules pour examiner les corrélations entre les actions de LVMH, TotalEnergies et BNP Paribas a fourni des insights précieux sur les interdépendances sectorielles, contribuant à une compréhension plus nuancée des marchés. Enfin, l'adaptation des processus de Hawkes pour modéliser les ordres d'arrivée du Bitcoin a démontré l'efficacité de cette approche pour capturer les dynamiques temporelles et les autocorrélations dans les données des marchés de cryptomonnaies. Ces approches diversifiées illustrent la richesse et la complexité de l'analyse financière moderne, soulignant l'importance d'adopter des outils et des méthodologies variés pour une compréhension complète du comportement des marchés.