

30 août 2024

Ivan Tchomba Ngueko, Alain Kuissu Kamga, El Vilaly Oumouhani





TABLE DES MATIÈRES

1	Dataset links	2
	Prédiction de la volatilité : GARCH vs Rough volatility 2.1 Rough volatility	
3	Processus de Hawkes	8
	Copule pour du pair trading 4.1 Qu'est-ce qu'une copule?	

1 DATASET LINKS

Dans ce mémoire, nous nous sommes focalisés sur les données du S&P 500.

import yfinance as yf

ticker = 'GSPC' data = yf.download(ticker, start='2005-01-01', end='2024-03-21')

Suivant nos besoins bous avons ajuster le ticker (bitcoin, Russel, ...). Yahoo finance link



2

PRÉDICTION DE LA VOLATILITÉ : GARCH VS ROUGH VOLATILITY

Nous étudierons dans cette partie les propriétés des séries temporelles linéaires ainsi que de leurs volatilités et nous construirons deux prédicateurs de la volatilité réalisée (et non implicite) du cours du S&P 500 chacun basés sur des propriétés particulières de celle ci.

2.1 Rough volatility

La volatilité réalisée des actifs financiers à très longtemps été modélisée à l'aide d'une dynamique brownienne standard assumant que ces 2 objets auraient la même régularités. C'est ce qu'on retrouve par exemple dans le modèle de Heston pour le pricing des options en volatilité stochastique. Ici on suppose que la volatilité a la même régularité qu'un mouvement brownien standard. Cependant, l'observation de données historiques des series temporelles financières montre que la volatilité pour ces actifs est moins régulière (au sens régularité des fonctions) qu'un mouvement brownien (voir Figure 1). Mathématiquement, la régularité d'un mouvement brownien est d'ordre 0.5 au sens de la norme de Holder mais nous montrerons que celle de la volatilité est beaucoup plus irrégulière et a un ordre de grandeur 0.1 au sens de la norme de Holder.

Le concept de **rough volatility** est motivé par la nécessité de modèles plus sophistiqués pour capturer les caractéristiques empiriques observées dans les données de volatilité des marchés financiers. Contrairement aux modèles classiques qui supposent une volatilité constante ou suivant un processus stochastique lisse ou une dynamique portée par un mouvement brownien standard, les modèles de rough volatility reconnaissent la rugosité intrinsèque de la volatilité, c'est-à-dire sa nature irrégulière et fractale.

Ici, on suppose

Nous calculons la volatilité historique ou réalisée du S&P 500 dont les données de prix de clôture ont été importé depuis la bibliothèque python yfinance pour illustrer cela. : S&P data

1. Calculer les rendements logarithmiques : Vous devez d'abord calculer les rendements logarithmiques de l'actif. Cela se fait en utilisant la formule suivante :

$$r_t = \log(P_t) - \log(P_{t-1})$$

où r_t est le rendement à l'instant t, P_t est le prix de l'actif à l'instant t, et P_{t-1} est le prix de l'actif à l'instant t-1.

- 2. Calculer la variance réalisée : La variance réalisée est calculée en agrégeant les rendements sur la période définie (21 jours par exemple). C'est-à-dire que vous sommez les carrés des rendements pour chaque période (par exemple, chaque jour si vous calculez la volatilité quotidienne).
- 3. Calculer la volatilité réalisée : La volatilité réalisée est simplement la racine carrée de la variance réalisée.

Notons que cette approche suppose que la moyenne est fixée à zéro, compte tenu de la tendance à la hausse et à la baisse dans le mouvement de l'indice.

Ce comportement de faible régularité (ordre de grandeur de 0.1) est observé pour la volatilité de quasiment toutes les séries temporelles financières. Nous allons prouver cette universalité.

Soit σ_{t0} , σ_{t1} , ..., σ_{tN} la série temporelle calculée pour cette période, où $t_{i+1}-t_i$ correspond à un jour ouvrable. Notre mesure d'échelle $m(q, \Delta)$ est le q-ème moment absolu des incréments de la log-volatilité, défini par :

$$m(q, \Delta) := \frac{1}{N} \sum_{k=1}^{\lfloor N/\Delta \rfloor} \left| \log(\sigma_{k\Delta}) - \log(\sigma_{(k-1)\Delta}) \right|^q,$$



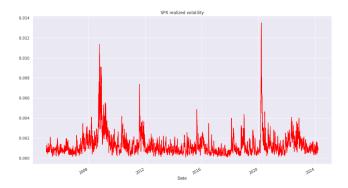


FIGURE 1 – SPX realized volatility

pour différents q>0 et retards Δ . Notre objectif est de revisiter la découverte dans J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough selon laquelle la log-volatilité spot présente les mêmes propriétés d'échelle qu'un mouvement brownien fractionnaire avec un exposant de Hurst H<1/2, et donc, on peut la modéliser avec un tel processus. Il est utile de mentionner les remarques suivantes :

Remarques:

— La quantité $m(q, \Delta)$ est l'équivalent discret de $E[|\log(\sigma_{\Delta}) - \log(\sigma_{0})|^{q}]$. Rappelons que le mouvement brownien fractionnaire W_{H} vérifie :

$$E[|W_{H,t+\Delta} - W_{H,t}|^q] = \widetilde{K}_q \Delta^{qH},$$

— Nous nous attendons à ce que la volatilité se comporte étroitement comme un mouvement brownien fractionnaire. En conséquence, nous observerions la relation suivante :

$$m(q, \Delta) \sim K_q \Delta^{qH}$$
. (1)

Pour vérifier la validité de l'équation (1), nous traçons $log(m(q, \Delta))$ en fonction de $log(\Delta)$ pour différentes valeurs de q. En fonction des résultats de cette première régression, nous pouvons écrire, pour un q donné,

$$m(q, \Delta) \sim K_q \Delta^{\zeta_q}$$
.

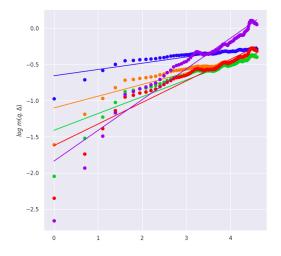


FIGURE 2 – $log(m(q, \Delta))$ en fonction de $log(\Delta)$



D'après le graphique log-log ci-dessus, nous observons que pour chaque $q, m(q, \Delta) \propto \Delta^{\zeta_q}$. Le graphe 3 ci dessous montre l'évolution de ζ_q avec q.

Nous trouvons la relation d'échelle monofractale

 $\zeta_q = qH$, avec $H \approx 0.14$.

Notons cependant que H varie au fil du temps, dans une plage étroite. Notons également que notre estimation de H est biaisée à la hausse car nous avons approximé la variance instantanée v_t par sa moyenne sur chaque journée $\frac{1}{T} \int_0^T v_t dt$.

Ce qui justifie le résultat pour le S&P500. Nous retrouvons notre **ordre de grandeur de 0.1** pour d'autres actifs trouver sur yahoo finance. (Voir 1)

num	names	ticker	h	ν
0	S&P500	ĜSPC	0.148837	0.444150
1	Google	GOOGL	0.097106	0.520188
2	Apple	AAPL	0.104616	0.485646
3	Coca-Cola	KO	0.095554	0.522895
4	Microsoft	MSFT	0.108797	0.508458
5	Gold (Futures)	GC=F	0.089776	0.474940
6	Brent Crude Oil (Futures)	BZ=F	0.125933	0.456607
7	NVIDIA	NVDA	0.097446	0.505514
8	Intel	INTC	0.099904	0.510889

Table 1 – Ordre de grandeur pour d'autres produits financiers

L'universalité de ce résultat est montrée dans l'article : Volatility is rough [1] dans lequel on mentionne des tests sur plus de 6000 actifs.

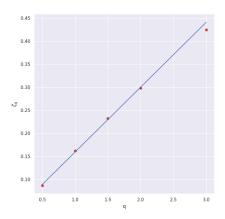


FIGURE 3 – ζ_q en fonction de q

De ce résultat, on peut écrire : $log(\sigma_{t+\Delta}) - log(\sigma_t) = \nu(W_{t+\Delta}^H - W_t^H)$ avec W^H un mouvement brownien fractionnaire d'indice de Hurst H proche de 0.1. (the RFSV (for Rough Fractional Stochastic Volatility))

La motion brownienne fractionnaire (fBm) W_t^H ; $t \in \mathbb{R}$ est le processus gaussien unique avec une moyenne nulle et une fonction de covariance

$$E[W_t^H W_s^H] = \frac{1}{2} \left\{ |t|^{2H} + |s|^{2H} - |t-s|^{2H} \right\}$$

où $H \in (0,1)$ est appelé l'indice de Hurst ou paramètre.

En particulier, lorsque $H=\frac{1}{2}$, fBm est simplement le mouvement brownien.

Si $H > \frac{1}{2}$, les incréments sont positivement corrélés, donc le processus est en tendance.



Si $H < \frac{1}{2}$, les incréments sont négativement corrélés, donc le processus est en réversion.

Il existe une infinité de représentations possibles de fBm en termes de mouvement brownien. Par exemple, avec $\gamma = \frac{1}{2} - H$,

Mandelbrot-Van Ness

$$W_t^H = C_H \left\{ \int_t^{-\infty} dW_s (t-s)^{\gamma} - \int_0^{-\infty} dW_s (-s)^{\gamma} \right\}$$

Le choix

$$C_{H} = \sqrt{2H \frac{\Gamma\left(\frac{3}{2} - H\right)}{\Gamma(2 - 2H)\Gamma\left(H + \frac{1}{2}\right)}}$$

assure que

$$E[W_t^H W_s^H] = \frac{1}{2} \{ t^{2H} + s^{2H} - |t - s|^{2H} \}$$

• La formule de prévision

Nous obtenons la formule de prévision de la variance.

$$E^{P}[v_{t+\Delta}|F_{t}] = \exp\{E^{P}[\log(v_{t+\Delta})|F_{t}] + 2c\nu^{2}\Delta^{2H}\}$$

οù

$$E^{P}[\log v_{t+\Delta}|F_{t}] = \frac{\cos(H\pi)}{\pi} \Delta^{H+\frac{1}{2}} + \frac{1}{2} \int_{t}^{-\infty} \frac{\log v_{s}(t-s+\Delta)}{(t-s)^{H+\frac{1}{2}}} ds.$$

Nous la testons pour le cas de l'indice S&P sur la période '2005-01-01' à '2024-03-21'. La dynamique de la volatilité est bien prédite par le modèle.

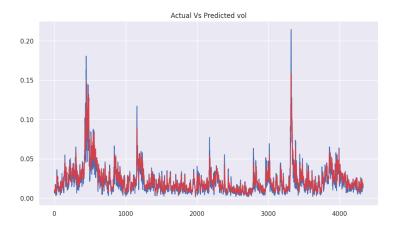


Figure 4 – Actual Vs Predicted vol

2.2 GARCH

Les modèles GARCH (Generalized Autoregressive Conditional Heteroskedasticity) sont des modèles économétriques qui modélisent la volatilité conditionnelle dans une série temporelle financière. Ils sont utilisés pour



capturer les propriétés de persistance et de clustering de la volatilité observées dans de nombreux marchés financiers. La volatilité des séries temporelles financiers n'est pas constantes comme on peut le voir en 1 de plus, on voit que "les grands changements ont tendance à être suivis de grands changements, de signe quelconque, et les petits changements ont tendance à être suivis de petits changements.". Cela s'observe très bien sur la figure 5.

Un modèle GARCH est défini comme suit :

Soit r_t une série temporelle de rendements financiers. Un modèle GARCH(p, q) est défini comme :

$$r_t = \sigma_t \varepsilon_t$$

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

Où:

- ε_t est un processus stochastique i.i.d. avec une distribution normale standardisée.
- σ_t^2 est la variance conditionnelle des rendements à l'instant t.
- ω est une constante.
- α_i et β_i sont les coefficients des termes autorégressifs pour les rendements et les volatilités, respectivement.
- p et q sont les ordres des termes autorégressifs pour les rendements et les volatilités, respectivement.

Les modèles GARCH permettent de capturer les propriétés de persistance de la volatilité, ce qui signifie que les chocs de volatilité ont tendance à persister dans le temps (Figure 5). Cela les rend particulièrement utiles pour modéliser les séries temporelles financières où la volatilité présente des clusters.

Les modèles GARCH offrent plusieurs avantages :

- Flexibilité : Les modèles GARCH permettent de capturer différents schémas de volatilité, allant de la volatilité constante aux modèles plus complexes avec des effets auto régressifs.
- Précision : En capturant les propriétés de persistance et de clustering de la volatilité, les modèles GARCH peuvent fournir des estimations précises de la volatilité future.
- Interprétation : Les coefficients des modèles GARCH ont une interprétation économique directe, ce qui permet de comprendre comment les chocs passés affectent la volatilité future.

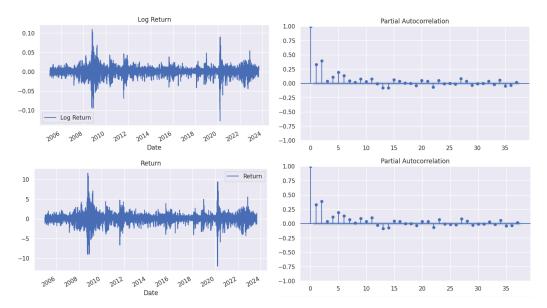


Figure 5 – S&P 500 log-return statistics

Nous utilisons un GARCH(1,1) que nous calibrons sur nos données du S&P500 et nous retrouvons le résultat de la figure 6, montrant l'efficacité de la méthode.



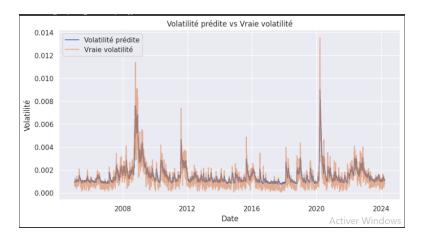


FIGURE 6 – Volatility S&P500 Prediction with GARCH(1,1)

PROCESSUS DE HAWKES

Le processus de Hawkes est un type de processus d'auto-excitation et peut être utilisé pour gérer le comportement d'auto-excitation que l'on trouve dans le volume/la volatilité des actifs financiers. Dans ce cas ci, nous l'appliquons à la mesure de la volatilité du bitcoin. Les grandes tendances présentent souvent une volatilité plus élevée. Les tendances se terminent souvent au moment où la volatilité se disperse. On montre une stratégie de trading simple utilisant le processus de Hawkes qui a une performance décente et des taux de gains assez élevés.

Un processus de Hawkes est un type de processus ponctuel auto-excitation. Un processus de Hawkes est défini comme suit :

Un processus de Hawkes est un processus ponctuel N dont la fonction d'intensité conditionnelle λ est définie par :

$$\lambda(t) = \mu(t) + \sum_{i:\tau_i < t} \nu(t - \tau_i)$$

où $\mu(t)$ est le taux de fond et $\nu(t-\tau_i)$ est la fonction d'excitation qui quantifie l'influence des événements passés sur l'intensité actuelle.

Les processus de Hawkes sont intéressants pour modéliser la volatilité pour plusieurs raisons :

- 1. **Auto-excitation**: Les processus de Hawkes sont des processus auto-excitants, ce qui signifie que l'arrivée d'un événement rend les événements futurs plus probables. Cela peut être utile pour modéliser la volatilité, car les grands mouvements de prix ont tendance à être suivis par d'autres grands mouvements de prix.
- 2. Modélisation des sauts en haute fréquence : Les processus de Hawkes peuvent être utilisés pour modéliser les sauts dans la volatilité. Cela peut être important pour capturer les changements soudains et importants dans la volatilité qui peuvent se produire lors d'événements de marché importants.

Dans notre cas, nous utilisons une version discrète du processus de Hawkes, où la fonction d'excitation est une décroissance exponentielle. C'est-à-dire que l'influence d'un événement diminue exponentielle ment avec le temps. Il est relativement simple et s'accorde à la dynamique La formule pour le processus de Hawkes discret est la suivante :





$$\alpha = e^{-\kappa}$$

$$O = \{o_1, o_2, ..., o_n\}$$

$$o_i = \begin{cases} d_i & \text{si } o_{i-1} \text{ est NaN} \\ o_{i-1} \cdot \alpha + d_i & \text{sinon} \end{cases}$$

$$O = O \cdot \kappa$$

L'idée est que chaque événement augmente la probabilité d'événements futurs. Dans cette fonction, κ est un paramètre de décroissance exponentielle, et est le facteur d'atténuation. Pour chaque point de données, si la valeur précédente n'est pas un nombre (NaN), la valeur actuelle est simplement la valeur de la série de données. Sinon, la valeur actuelle est une combinaison linéaire de la valeur précédente et de la valeur actuelle de la série de données. Enfin, chaque élément de la série de sortie est multiplié par κ .

L'ATR (Average True Range) est un indicateur technique utilisé pour mesurer la volatilité du marché. Il est calculé en prenant la moyenne des vraies amplitudes sur une période de temps spécifiée. Dans ce cas, l'ATR est calculé sur les valeurs logarithmiques des prix hauts, bas et de clôture avec une période de retour en arrière de

Nous calculons ensuite la plage normalisée en prenant la différence entre le logarithme du prix haut et le logarithme du prix bas, et en divisant par l'ATR. Cela donne une mesure de la volatilité relative du marché.

Le processus de Hawkes est appliqué à la plage normalisée avec un paramètre kappa de 0.1. Cette valeur est issue de plusieurs test sur la série temporelle. On veut ainsi modéliser cette volatilité relative. (Voir figure 7)

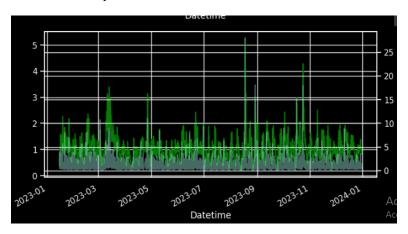


FIGURE 7 – Processus de Hawkes en vert avec $\kappa = 0.1$ vs Plage normalisée en gris pour le bitcoin

un signal de trading basé sur la volatilité. Elle utilise une fenêtre glissante (lookback en heure) pour calculer les quantiles 5% et 95% de la volatilité. Si la volatilité actuelle est inférieure au quantile 5%, elle réinitialise le signal à 0. Si la volatilité actuelle dépasse le quantile 95% et que la volatilité précédente était inférieure ou égale au quantile 95%, elle calcule le changement de prix depuis la dernière fois que la volatilité était inférieure au quantile 5%. Si le changement de prix est positif, le signal est fixé à 1 (long), sinon il est fixé à -1 (short).

Pour différents lookbacks (taille de la fenêtre glissante) 24, 48, 96, 168, 336 et différents κ 0.5, 0.25, 0.1, 0.05, 0.01 on obtient le tableau

Metric	Value
Profit Factor	1.0461661034491434
Long Win Rate	0.5789470637120717
Long Average	0.025258200540659848
Short Win Rate	0.27272714876038695
Short Average	-0.008993903936138019
Time In Market	0.4878640776699029



On peut constater grâce à la Figure 8 que les 'profit factor' sont pour la plupart au dessus de 1 ce qui est encourageant et montre la pertinence de la stratégie et de la modélisation par le processus de Hawkes simplifié.

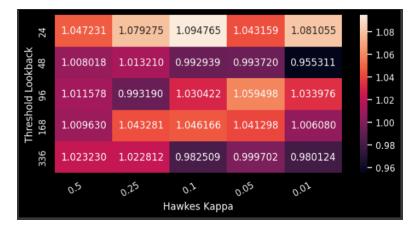


FIGURE 8 – Profit factor = win return / loose return for the grid kappa/lookback

4 COPULE POUR DU PAIR TRADING

La théorie des copules est intimement liée à l'étude des distributions de probabilité multivariées dont les marginales sont données. C'est-à-dire à l'étude de la manière dont des phénomènes aléatoires peuvent se combiner, et en particulier produire des phénomènes encore plus intenses ou au contraire amortir les fluctuations de chacun. Ce problème est d'un intérêt tout particulier en pair trading, où l'étude des rendements des actifs pour chaque paire s'est imposée au travers des activités de stratégie de trading ou d'étude de risque.

Il apparaît que chaque paire d'actifs possède une distribution de rendements qui lui est propre. Cette dernière est souvent asymétrique et peut présenter une queue de distribution lourde (probabilité accrue de réalisation d'événements graves à extrêmes), de sorte qu'aucune fonction de répartition multivariée classique ne peut modéliser une telle variété de comportements individuels. Nous allons montrer que la théorie des copules permet de lever cette contrainte.

4.1 Qu'est-ce qu'une copule?

Une copule (ou n-copule) C est la fonction de répartition jointe d'un vecteur aléatoire (U_1, U_2, \dots, U_n) de \mathbb{R}^n dont les marginales U_1, U_2, \dots, U_n sont uniformes U[0;1]:

$$C(u_1, \dots, u_n) = \Pr(U_1 \le u_1, \dots, U_n \le u_n), \quad u_i \in [0, 1], \quad i = 1, \dots, n$$
 (1)

L'un des principaux attraits de la théorie des copules est donné par le théorème de Sklar :

[Sklar 1959] Soit F une fonction de répartition n-variée (F est le fonction de répartition de (X_1, X_2, \ldots, X_n)) dont les marginales F_1, \ldots, F_n sont continues (F_1 est la fonction de répartition de X_1, \ldots, F_n celle de X_n). Alors il existe une unique copule C telle que :

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$
(2)



où $(x_1, \ldots, x_n) \in \text{Dom} F_1 \times \ldots \times \text{Dom} F_n$

On peut conclure trois choses de ce théorème :

- Tout d'abord, puisque le fonction de répartition multivariée contient toute l'information sur le vecteur aléatoire qu'elle caractérise, cela signifie que la copule contient toute l'information sur la dépendance qui existe entre les composantes du vecteur aléatoire.
- De plus, puisque le théorème ne requiert pas des marginales qu'elles soient identiques, ou même qu'elles appartiennent à la même famille, il est possible de construire des fonctions de répartition multivariées simplement en sélectionnant des fonctions de répartitions marginales continues (par exemple Lognormale, Pareto, Weibull et Gamma) ainsi qu'une copule pour les lier.
- Enfin l'étude des phénomènes aléatoires multivariés peut se scinder en deux parties : l'étude, dans un premier temps, des comportements individuels au travers de l'étude des marginales, puis l'étude de la forme de dépendance existante entre ces comportements individuels qui est l'étude de la copule.

Dans un contexte d'évaluation du capital économique d'une compagnie, cela signifie que l'étude de la distribution des charges sinistres au niveau de la compagnie se réduit à deux étapes clefs : rendre compte du comportement des charges sinistres au niveau de chaque branche, puis refléter la dépendance entre ces branches en choisissant une copule.

Ce théorème fournit aussi une manière de généraliser les modèles préexistants d'évaluation du capital économique ayant recours à des fonctions de répartitions multivariées connues. En effet, puisque toutes ces distributions présentent des marginales continues, il existe une unique copule, implicite, les reliant. Une fois cette dernière isolée, il devient possible de la réutiliser avec d'autres marginales susceptibles de mieux refléter les caractéristiques de la société étudiée.

4.2 Utilisation de copule pour du pair trading

Nous allons étudier à l'aide de divers copule, la corrélation entre les cours du S&P 500 et du Russel 2000 et essayer de construire une stratégie de trading basée sur les 2 actifs.

Nous pouvons remarquer que les 2 actifs ont l'air de suivre une tendance commune, donc semblent corrélés. Nous avons utiliser divers copules pour observer une éventuelle corrélation entre ces actifs et appliquer une strategie de pair trading avec les informations issues des distributions inférées.

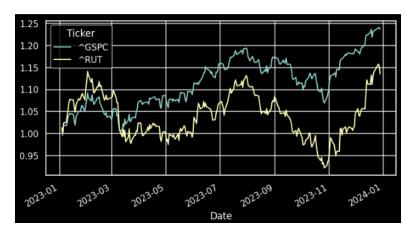


Figure 9 – Évolution des indices S&P500 et Russel 2000

La stratégie de Trading basée sur la Probabilité Conditionnelle et utilise la distribution conjointe de la Copule qui ici est gaussienne apres de nombreux tests pour le choix. La stratégie de trading présentée ici est basée sur la probabilité conditionnelle et utilise la distribution conjointe de la Copule. Elle comprend deux étapes principales : le calcul de la probabilité conditionnelle et l'exécution de la stratégie de trading par paires.



Calcul de la Probabilité Conditionnelle : Pour chaque période de trading, nous ajustons les marginales à l'aide de la distribution de Student sur la période d'ajustement. Ensuite, nous transformons les marginales en utilisant la fonction de répartition cumulative (CDF) de la distribution ajustée. Nous ajustons une copule gaussienne sur les marginales transformées. Enfin, nous calculons la probabilité conditionnelle pour chaque retour de prédiction en utilisant la CDF conditionnelle de la copule.

Stratégie de Trading par Paires : Nous ouvrons une position longue lorsque la probabilité conditionnelle du premier actif est inférieure à 1-cl et celle du deuxième actif est supérieure à cl. Nous ouvrons une position courte lorsque la probabilité conditionnelle du premier actif est supérieure à cl et celle du deuxième actif est inférieure à 1-cl. Nous fermons la position lorsque la probabilité conditionnelle de l'actif long dépasse $1-cl_{\rm close}$ ou celle de l'actif court tombe en dessous de $cl_{\rm close}$.

Calcul des Rendements: Les rendements sont calculés en multipliant les positions décalées par les rendements de trading. Les rendements sont ensuite convertis en rendements simples et la moyenne des rendements est calculée. Enfin, nous calculons le rendement cumulatif de la stratégie de trading.

Les résultats de notre algorithme sont sur la figure 10.

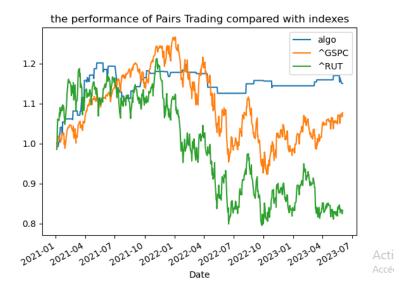


FIGURE 10 – Performance de la stratégie



RÉFÉRENCES

- [1] Gatheral, Jaisson and Rosenbaum. 2014. Volatility is rough. $Jim\ Gatheral\ Baruch\ College,\ City\ University$ of $New\ York$
- [3] Àlex Martorell i Locascio. Modeling volatility using arch and garch processes. Treballs Finals de Grau de Matem'atiques, Facultat de Matem'atiques, Universitat de Barcelona.