

# MÉTHODES D'ALLOCATION OPTIMAL DE PORTEFEUILLE

---

SLIMANI MAYSAS, YAKOUB MÉLANIE, ZHOU JINGYI

# Contents

<b>1</b>	<b>Abstract</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
<b>3</b>	<b>Première méthode : Allocation de portefeuille basé sur la théorie des graphes</b>	<b>5</b>
3.0.1	Présentation du projet et objectifs . . . . .	5
3.0.2	Importation des données . . . . .	5
3.0.3	Statistiques descriptives . . . . .	6
3.1	Création du réseau . . . . .	8
3.1.1	Causalité à la Granger . . . . .	8
3.1.2	Matrice d'Adjacence . . . . .	9
3.2	Méthode de Clustering . . . . .	10
3.3	Simulation : Achat / vente . . . . .	11
3.3.1	Construction . . . . .	11
3.3.2	Résultats . . . . .	12
<b>4</b>	<b>Deuxième approche : Méthode</b>	<b>15</b>
<b>5</b>	<b>Conclusion</b>	<b>18</b>

# 1 Abstract

Ce mémoire se focalise sur la sélection d'actifs dans la gestion de portefeuille, une préoccupation majeure pour de nombreux traders évoluant sur les marchés financiers. L'objectif premier d'un portefeuille financier est d'optimiser le rendement tout en minimisant le risque, nécessitant une allocation judicieuse des actifs. Pour cela, les investisseurs s'appuient sur des théories existantes, en particulier la Théorie Moderne du Portefeuille (TMP) introduite par Harry Markowitz en 1950.

La TMP met en avant la sélection d'actifs dans le but de maximiser les rendements tout en minimisant le risque. L'importance de la diversification, basée sur des actifs non corrélés, est soulignée pour atténuer l'impact des chocs de marché. La matrice de corrélation de Markowitz offre une vision statique des relations entre actifs, ne tenant pas compte de l'évolution future des corrélations entre ces derniers. Face à cette contrainte, nous cherchons à introduire une perspective dynamique en tenant compte de l'évolution des relations entre actifs au fil du temps, y compris les effets "spillover" (la propagation de l'événement du secteur à un autre) et feedback (où la situation a des répercussions sur elle-même).

Pour répondre à notre problématique, nous commencerons d'abord par présenter notre base de données contenant 147 actifs du S&P500 que nous avons préalablement transformé en rendement. Ensuite, nous utiliserons le test statistiques de causalité de Granger afin d'établir notre matrice d'adjacence sur la base d'un seuil, qui nous servira à établir les relations causales entre nos actifs. Cette matrice sera ensuite utilisée pour calculer nos clusters à l'aide de la théorie des graphes, en nous appuyant sur l'article de Giorgio Fagiolo intitulé "Clustering in Complex Directed Networks," rédigé en 2007. Ces étapes nous permettront de sélectionner les actifs les moins systémiques qui constitueront notre portefeuille à investir. Enfin, nous conclurons par une simulation d'achat et de revente, évaluant ainsi la capacité de notre modèle.

## 2 Introduction

La sélection des actifs d'un portefeuille est une problématique partagée par un grand nombre de traders opérant sur les marchés financiers. Il s'agit de déterminer, parmi les actifs sur lesquels il est possible d'investir, la répartition du capital constituant le meilleur portefeuille possible.

Le but premier d'un portefeuille financier pour l'investisseur est d'obtenir un rendement le plus élevé possible tout en minimisant sa prise de risque. Pour cela, il faut tout d'abord une bonne allocation de son portefeuille. Cette allocation est une sorte de feuille de route optimale ayant pour but de maximiser la cohérence et le rendement du portefeuille à long terme. Tout cela dépend d'un certain nombre de critères (horizon d'investissement, la volonté à prendre du risque, la fiscalité etc...) qui ne rentreront pas dans le cadre de notre analyse. Chaque investisseur cherche alors son portefeuille efficient en se basant sur les théories existantes.

La théorie moderne du portefeuille repose sur une sélection d'actifs financiers pour laquelle la rentabilité sera maximale tout en ayant un niveau de risque minimum. Depuis 1950 apparaît la notion de gestion moderne de portefeuille qui repose sur la théorie moderne du choix de portefeuille dont le point de départ est dû à Harry Markowitz. En principe, les actifs entrant dans un portefeuille doivent être non-corrélés car lorsqu'ils le sont, ils ont tendance à fluctuer de manière similaire, pouvant entraîner une baisse importante du rendement du portefeuille en cas de choc sur les marchés financiers. Les relations de corrélations entre les différents titres d'un portefeuille sont donc au centre de cette théorie, qu'il modélisera à travers une matrice de corrélation qui permet de représenter graphiquement les liens entre différents actifs d'un portefeuille. A partir des résultats de cette matrice, Markowitz met en avant le principe de diversification. L'investisseur devra s'appliquer à diversifier son portefeuille sur des valeurs qui ont des rendements tels qu'ils n'entraîneront pas de fluctuations importantes sur le rendement total du portefeuille. L'objectif ici est d'obtenir une covariance minimale entre toutes les valeurs présentes dans le portefeuille. Les travaux de Markowitz ont donc abouti à une solution générale au problème de maximisation du rendement et de minimisation du risque.

Même si la matrice de corrélation de Markowitz est un outil utile pour aider les investisseurs à comprendre les relations entre les actifs et déterminer la meilleure répartition de leurs investissements, cette dernière s'avère être statique. En effet, cette matrice est basée sur des données historiques de rendement des actifs, ne tenant pas compte de l'évolution future des corrélations entre ces derniers. Cela signifie qu'elle ne reflète pas les changements potentiels dans les relations entre les actifs qui pourraient survenir dans le futur et ne prend pas en compte l'effet spillover (la propagation de l'événement du secteur à un autre) et feedback (où la situation a des répercussions sur elle-même) possible, pouvant entraîner un effet boule de neige.

Dans ce mémoire, nous commencerons d'abord par présenter notre projet ainsi que les données que nous avons sélectionnées pour le mettre en œuvre. Nous entrerons ensuite au cœur du sujet en exposant la méthodologie

de la création de notre réseau, en expliquant notre choix de causalité et en mettant en œuvre notre matrice d'adjacence à travers une représentation graphique. Dans une deuxième étape, nous expliquerons la méthode de regroupement choisie dans le cadre de ce mémoire, et nous conclurons par une simulation d'achat et de revente qui testera la capacité de notre modèle.

Au cours de ce projet, nous supposerons certaines hypothèses pour encadrer notre travail. Tout d'abord, nous excluons les coûts de transaction de notre thèse. Les coûts de transaction sont généralement faibles en proportion des prix des actifs et varient à la fois d'un actif à l'autre, ce qui rend difficile la comparaison de la performance de différents actifs, mais aussi dans le temps, en fonction des conditions du marché et des changements réglementaires. Deuxièmement, nous supposerons que les séries de rendements utilisées sont stationnaires.

### **3 Première méthode : Allocation de portefeuille basé sur la théorie des graphes**

#### **3.0.1 Présentation du projet et objectifs**

Dans le cadre de notre analyse, nous nous intéresserons aux effets feedback et spillover d'un réseau dynamique que nous créerons à partir des actifs de nos entreprises du S&P500. L'objectif principal de ce projet vise à trouver les actifs qui sont liés de façon systémique afin de pouvoir faire ressortir du portefeuille les actifs qui ont le plus d'impact sur les autres. Une fois la topologie du réseau faite, nous définirons les indicateurs de clustering afin de savoir comment va se diffuser l'information et voir quels actifs sont les plus systémiques. Les actifs sélectionnés seront alloués (equally weighted) sur une période temps donnée.

#### **3.0.2 Importation des données**

Dans un premier temps, nous avons récupéré les données d'un grand nombre d'entreprises via yahoo finance. Ces données sont composées des données de fermeture, qu'on transformera par la suite en rendement. Les entreprises figurant dans les bases de données sélectionnées sont répertoriées dans la liste SP 500 de l'année 2023. Le SP 500 (Standard Poor's 500) est un indice boursier qui mesure la performance des 500 plus grandes entreprises cotées en bourse aux États-Unis. Cet indice est largement reconnu comme un indicateur important de la santé globale du marché boursier américain. Nous avons sélectionné près de 200 entreprises dans le but d'analyser et comprendre la performance des entreprises leaders dans différents secteurs au cours de l'année 2023.

### 3.0.3 Statistiques descriptives

Nous avons choisi de garder les données entre début 2007 et fin 2023, dans le but de prendre en compte plusieurs cycles financiers afin de tester si notre modèle considèrerait bien les actifs systématiques. Les variables (qui sont quantitatives) ont été décrites sous forme de moyenne, médiane, 1er et 3ème quartiles, minimum et maximum.

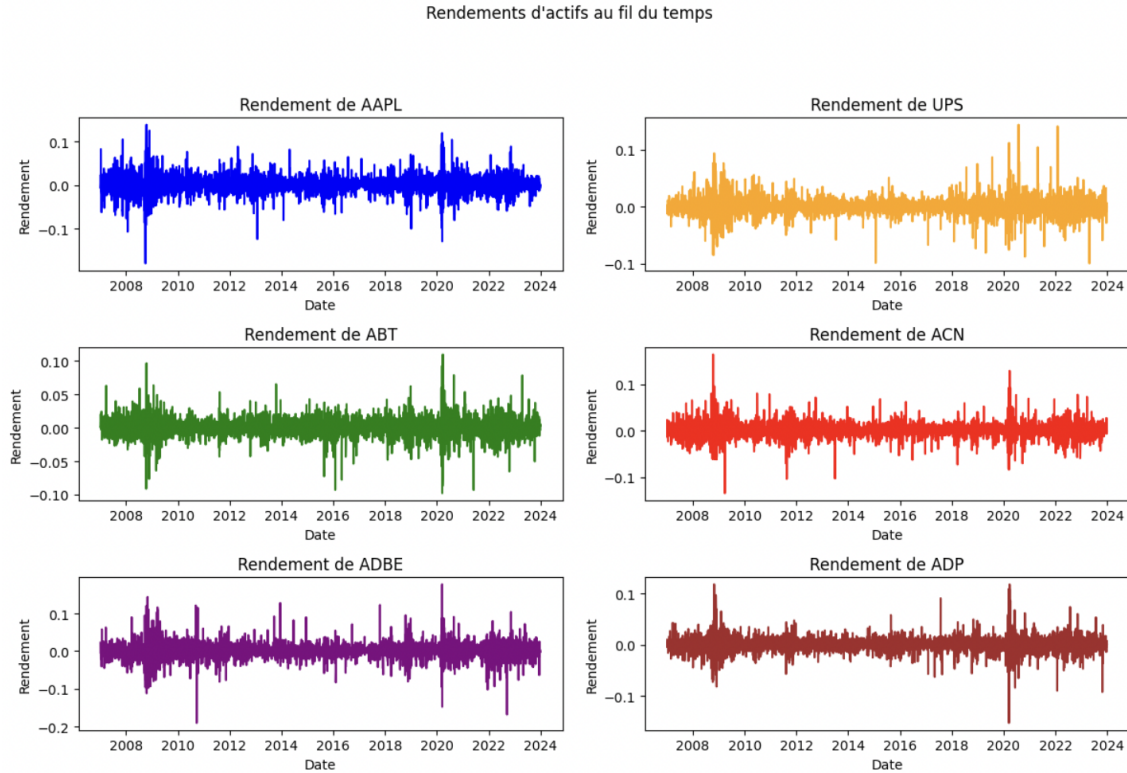


Figure 1: Représentation des rendements de 6 actifs du S&P500 au cours du temps

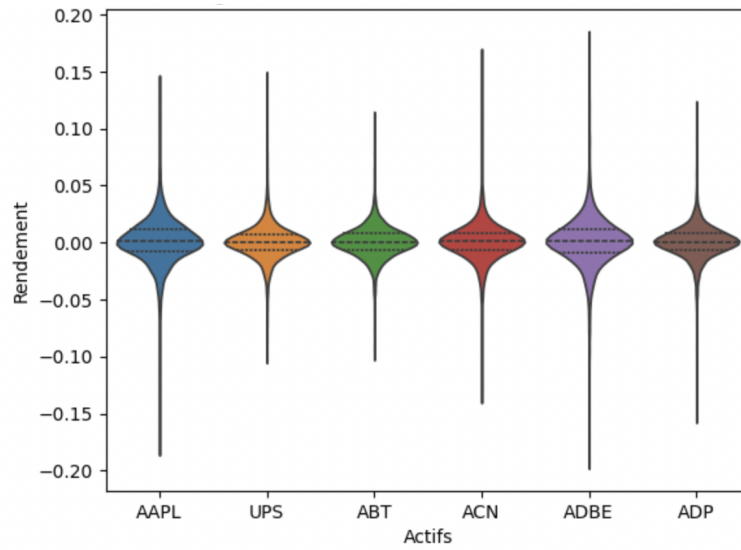


Figure 3 : Diagramme en violon du rendement de 6 actifs du S&P500

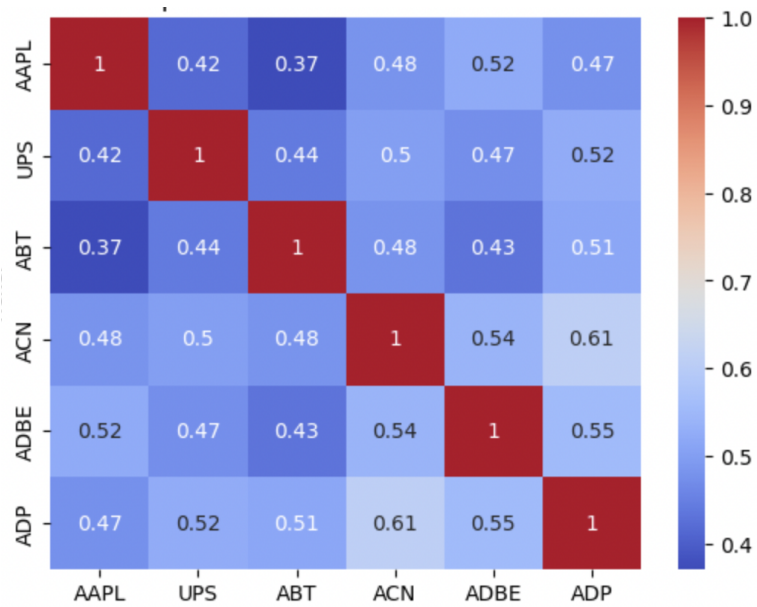


Figure 2: Matrice de corrélation de 6 actifs du S&P500



### 3.1 Création du réseau

La théorie des graphes permet d'étudier les graphes et leurs applications dans divers contextes. Dans notre cas, un graphe est une représentation visuelle de l'ensemble de nos actifs, appelés nœuds, qui sont reliés par des lignes, appelées arêtes. Nous appliquons le test de Granger entre la série temporelle  $i$  et  $j$  de chaque actif à l'aide de la fonction "grangercausalitytests" de la bibliothèque Statsmodels.

#### 3.1.1 Causalité à la Granger

La causalité à la Granger est une méthode statistique qui permet de déterminer si une série temporelle ( $X$ ) est une cause de l'autre série temporelle ( $Y$ ) ou inversement. Cette méthode évalue si l'information d'une série temporelle améliore les prédictions d'une autre série temporelle par rapport aux prédictions obtenues sans cette information. Si l'ajout de l'autre série améliore de manière significative les prévisions, cela suggère que la série ajoutée influence la série cible. En d'autres termes, la causalité de Granger aide à identifier si le passé de actifs impacte le présent d'autres actifs. L'approche la plus simple consiste à utiliser la spécification autorégressive d'un vecteur AR bivarié. Pour se faire, on suppose une longueur de décalage autorégressive  $p$  et estimer l'équation sans restriction par MCO.

$$x_t = c_1 + \sum_{i=1} \alpha_i x_{t-i} + \sum_{i=1} \beta_i y_{t-i} + u_t$$

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

On a ici l'hypothèse nulle  $H_0$  : «  $Y$  ne cause pas  $X$  au sens de Granger » et l'alternative  $H_1$  : «  $Y$  cause  $X$  au sens de Granger ». On continue en effectuant un test de Fisher de l'hypothèse nulle en estimant l'équation restreinte suivante également par MCO :

$$x_t = c_t + \sum_{i=1} \lambda_i x_{t-i} + \epsilon_t$$

Après avoir fait cela, on peut alors comparer la somme des résidus au carré des deux modèles :

$$RSS_1 = \sum_{t=1} \hat{u}_t^2 \quad RSS_0 = \sum_{t=1} \hat{e}_t^2$$

Ces hypothèses peuvent être testées à l'aide du test de Fisher. On obtient la statistique de test suivante:

$$F = \frac{T(RSS_0 - RSS_1)}{RSS_1}$$

Le rejet de l'hypothèse nulle impliquera une causalité réciproque entre  $X$  et  $Y$ . Dans le cadre d'un test de causalité à la Granger, il ne faut pas omettre le fait que les variables  $Y$  doivent nécessairement être stationnaire. Dans le cadre de ce projet, le problème ne se pose pas étant donné que nous travaillons sur des

rendements.

### 3.1.2 Matrice d'Adjacence

Après avoir calculé tous nos liens entre les actifs grâce à la méthode de Granger, nous les représentons par une matrice d'adjacence. Cette matrice d'adjacence représente les liens entre les différents éléments d'un réseau. Elle est utilisée afin de décrire les relations entre les différents nœuds du réseau, qui représentent ici nos actifs. On établit un seuil de significativité au dessous duquel la variable passe de 0 à 1. Pour se faire, on crée une fonction qui retourne la matrice d'adjacence de causalité de Granger, où les valeurs 1 indiquent une causalité statistiquement significative entre les séries correspondantes et les valeurs 0 indiquent l'absence de causalité statistiquement significative. Cette matrice est de taille  $N \times N$ , où  $N$  est le nombre de nœuds du réseau, qui ici est de taille  $147 \times 147$ . On a donc 1 dans la cellule de la matrice si un actifs  $i$  est lié à un autre  $j$ , et 0 s'ils ne le sont pas.

Ticker	AAPL	ABT	ACN	ADBE	ADP	ADSK	AEP	AFL	AIG	ALGN
Ticker										
AAPL	0	1	1	0	1	1	1	1	1	0
ABT	1	0	1	1	1	1	0	1	0	0
ACN	0	1	0	0	1	0	0	0	0	0
ADBE	1	1	1	0	1	0	1	1	1	0
ADP	1	1	0	0	0	0	1	0	0	1
ADSK	0	0	0	0	1	0	1	0	0	0
AEP	0	0	0	0	1	0	0	0	0	0
AFL	0	0	1	0	1	0	0	0	1	1
AIG	1	1	1	1	1	1	1	1	0	0
ALGN	0	0	0	0	0	0	0	0	0	0

Table 3 : Matrice d'adjacence binaire entre 10 compagnies basé sur la causalité de Granger.

## 3.2 Méthode de Clustering

Après avoir émis les liens entre les actifs et nos entreprises à travers une matrice d'adjacence et une représentation par graphe, il était pertinent pour nous de créer des clusters qui permettront de regrouper nos actifs afin d'obtenir le portefeuille le plus rentable possible. Les indicateurs de clustering vont pouvoir nous dire comment se diffuse l'information et quels sont les actifs les plus systémiques.

En théorie des graphes, le coefficient de clustering d'un graphe est une mesure du regroupement des nœuds dans un réseau. Plus précisément, ce coefficient est la probabilité que deux nœuds soient connectés sachant qu'ils ont un voisin en commun (triangle entre 3 actifs). En d'autre terme, cela revient à détecter des sous-ensembles de nœuds qui sont fortement interconnectés dans le réseau. Cette mesure fournit une indication sur la tendance d'un nœud à former des "groupes" de nœuds voisins. Plus le coefficient de clustering est élevé, plus le nœud forme des triangles orientés, indiquant une certaine cohésion dans son voisinage. Il est basé sur le nombre de triangles fermés (trois nœuds connectés) dans le voisinage du nœud. Un seuil de clustering a été établi, permettant de filtrer les relations significatives et de ne retenir que les actifs fortement liés entre eux.

Afin de créer notre cluster, nous nous sommes basés sur l'article de Giorgio Fagiolo (Clustering in Complex Directed Networks) écrit en 2007. L'article explore différentes méthodes de clustering pour les réseaux dirigés, notamment l'utilisation de la matrice de transition de Markov et de la matrice d'adjacence afin d'identifier les clusters. Il étudie les méthodes de clustering dans les réseaux dirigés complexes, c'est-à-dire les réseaux où les liens ont une direction (réseaux orientés) et où les nœuds peuvent avoir des relations de dépendance et de causalité. Il montre également comment ces méthodes peuvent être utilisées pour étudier les propriétés topologiques des réseaux, comme la centralité ou la fragmentation, et comment elles peuvent être utilisées pour identifier des patterns dans les données.

Dans le contexte des réseaux orientés, où les relations sont représentées par des arêtes orientées, le calcul du coefficient de clustering prend en compte les degrés d'entrée et de sortie des nœuds. Dans les réseaux orientés, les arêtes sont orientées (flèches) et les relations de voisinage ne sont pas nécessairement symétriques. Le degré d'entrée du nœud  $i$  représente le nombre d'arêtes pointant vers  $i$  (c'est-à-dire les flèches pointant vers le nœud  $i$ ), tandis que le degré extérieur du nœud  $i$  est défini lui comme le nombre d'arêtes provenant de  $i$  (c'est-à-dire provenant du nœud  $i$ ). On modélise cela :

$$d_i^{in} = \sum_{j \neq i} a_{ji} = (A^T)_i \mathbf{1}$$
$$d_i^{out} = \sum_{j \neq i} a_{ij} = (A)_i \mathbf{1},$$

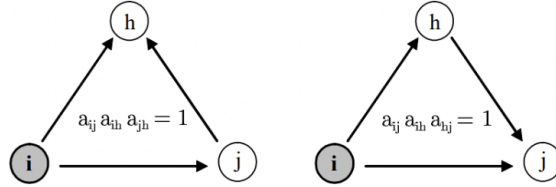
La matrice d'adjacence binaire  $A$  est utilisée, où chaque entrée  $A_{ij}$  est égale à 1 si une arête orientée existe de  $i$  vers  $j$ , et 0 sinon. La transposée de  $A$  est notée  $A^T$ ,  $(A)_i$  représente la  $i$ -ème ligne de  $A$  qu'on multiplie au vecteur colonne  $\mathbf{1}$  à  $N$  dimensions :  $(1, 1, \dots, 1)^T$ . Le degré total d'un nœud est simplement la somme de ses degrés d'entrée et de sortie. Dans le contexte de notre projet, cela suggère que nous utilisons ces concepts pour évaluer les liens entre les différents actifs financiers, en calculant le nombre de connexions (liens) de chaque actif avec les autres actifs de la base.

$$d_i^{tot} = d_i^{in} + d_i^{out} = (A^T + A)_i \mathbf{1}.$$

Si aucune auto-interaction n'est présente, on calcule le nombre d'arêtes bilatérales entre  $i$  et ses voisins (c'est-à-dire le nombre de nœuds  $j$  pour lesquels il existe à la fois une arête  $i \rightarrow j$  et une arête  $j \rightarrow i$ ) avec :

$$d_i^{\leftrightarrow} = \sum_{j \neq i} a_{ij} a_{ji} = A_{ii}^2.$$

Chaque coefficient de clustering sera défini comme le rapport entre le nombre de triangles réellement formés par  $i$  (les triangles impactés) et le nombre total de triangles qu'il aurait pu éventuellement former (nombre de triangles maximum du réseau). Pour trouver le nombre de triangle réellement formés par  $i$ , on calcule le produit matriciel triple qui permet de compter les 3-cycles dans le réseau. Dans un réseau dirigé, un 3-cycle est un motif où chaque nœud est connecté aux deux autres par des arêtes dirigées, ce qui permet alors de mesurer la connectivité locale entre les actifs.



La sélection des actifs joue donc un rôle clé dans la constitution des clusters. Les actifs retenus ont été ceux dont les coefficients de clustering étaient inférieurs au seuil pré-défini, indiquant une forte connectivité locale. Cette démarche a permis de mettre en évidence des groupes d'actifs partageant des relations de causalité similaires. Un coefficient élevé suggère une forte interconnexion entre les voisins du nœud, tandis qu'un coefficient bas peut indiquer une structure plus dispersée.

### 3.3 Simulation : Achat / revente

#### 3.3.1 Construction

La topologie du réseau faite et les indicateurs de clustering établis nous permettent de modéliser notre portefeuille efficient. Nous définissons les paramètres clés tels que la période de début et de fin de calcul, la taille

de la fenêtre d'investissement, la période pendant laquelle nous souhaitons investir, et un seuil de clustering. La fenêtre d'investissement, représentée par une période de 22 jours, nous permet d'analyser les données historiques des actifs sur cette période pour prendre de meilleures décisions d'investissement. L'objectif est d'obtenir le portefeuille le plus rentable possible pour les entreprises.

Pour se faire, on utilise la causalité de Granger et les coefficients de clustering pour sélectionner les actifs appropriés à investir. Nous conservons les actifs en fonction des coefficients de clustering, en ne retenant que ceux dont le coefficient est inférieur à 0,1, indiquant une faible corrélation avec les autres actifs du groupe. Cette sélection minutieuse vise à diversifier le portefeuille d'investissement en écartant les actifs trop liés entre eux, pouvant causer de réel dégâts si l'un d'entre eux venait à baisser considérablement.

La période d'investissement est ensuite définie, représentant les 10 jours suivant la fenêtre d'analyse. Nous répartissons le capital uniformément entre les actifs sélectionnés, calculons les rendements individuels de chaque actif sur la période d'investissement, et ajustons le capital total en conséquence.

### 3.3.2 Résultats

Cette approche itérative de simulation d'achat-revente nous permet de suivre l'évolution du capital au fil du temps, reflétant les résultats de nos décisions d'investissement. La méthode offre ainsi une perspective dynamique sur la performance du portefeuille dans un environnement financier changeant.

Le périmètre temporel étant très large, nous avons décidé de ne regarder l'impact que des périodes où des crises financières se sont déroulés : la crise de 2008 et du COVID-19. Nous cherchons à voir comment notre modèle agira face à ces crises. Nous commencerons par regarder les résultats sur une période de calcul de 22 jours (1 mois en trading) puis, en fonction des résultats obtenue, nous investirons dans les 10 jours suivants. Un intervalle de 30 jours nous permettra d'évaluer la capacité de notre modèle à s'ajuster rapidement aux conditions de marché changeantes. L'investissement dans les 10 jours suivants après la période de calcul peut être justifié par notre volonté de voir les prises de décisions rapides.

De plus, pour enrichir notre analyse, nous avons décidé d'ajouter le calcul du ratio de Sharpe. En effet, les périodes que nous avons sélectionnées sont caractérisées par une volatilité significative et des incertitudes de marché. Le ratio de Sharpe nous permettra donc d'évaluer dans quelle mesure notre modèle parvient à générer un rendement ajusté au risque, fournissant ainsi des insights cruciaux sur sa robustesse pendant des périodes de crises financière.

1. 2008-2009 : 22/08 jours

Date investissement	Nb actifs portefeuille	Ratio Sharpe	Capital fin période (\$)
2008-04-20 2008-04-30	9	-0.02	1055.82
2008-05-30 2008-06-09	4	-0.39	1003.55
2008-07-09 2008-07-19	4	-0.51	1028.53
2008-08-18 2008-08-28	3	0.34	1057.30
2008-09-27 2008-10-07	9	-0.25	928.55
2008-11-06 2008-11-16	12	0.18	917.13

Table 1: Résultats de la simulation achat/revente sur une période de calcul de 22 jours et d'investissement de 8 jours.

La période allant d'Août 2008 à Janvier 2009 a été témoin de l'une des crises financières les plus dévastatrices. Les répercussions ont été ressenties à l'échelle mondiale, entraînant d'importantes pertes pour les investisseurs. La simulation d'achat-revente pour cette période a reflété la sévérité de la crise avec une forte baisse des rendements pour l'année 2008.

Suite à la mise en place de notre stratégie, on peut constater que le Ratio de Sharpe a été négatif pour la majorité des périodes d'investissement. Cela signifie que le rendement ajusté au portefeuille est inférieur au taux sans risque, indiquant que l'investisseur aurait pu mieux performer s'il avait investi dans des actifs sans risque. Cependant, compte tenu du contexte de l'époque, ces résultats étaient attendus. Les Ratios de Sharpe positifs indiquent néanmoins que nous avons réussi à limiter certaines pertes.

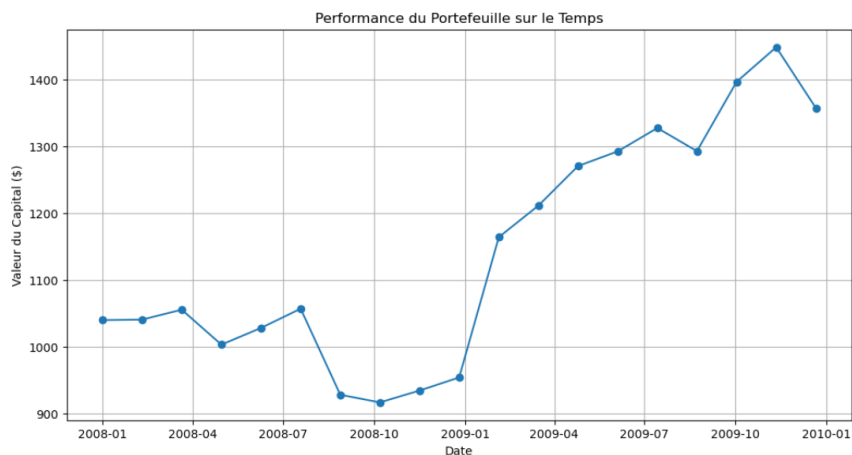


Table 1: Évolution du capital après investissement pour la période 2007 à 2008 du S&P500.

2. 2020-2022 : 22/08 jours

<b>Date investissement</b>	<b>Nb actifs portefeuille</b>	<b>Ratio Sharpe</b>	<b>Capital fin période (\$)</b>
2020-03-11 2020-03-21	926.56	7	-0.19
2020-04-20 2020-04-30	947.16	12	0.23
2020-05-30 2020-06-09	965.90	7	-0.11
2020-07-09 2020-07-19	1018.08	29	0.15
2020-08-18 2020-08-2	1023.07	5	0.18
2020-09-27 à 2020-10-07	1053.54	7	0.02

*Table 1: Résultats de la simulation achat/revente sur une période de calcul de 22 jours et d'investissement de 8 jours en 2020-2021.*

La période de 2020 à 2022 a été marquée par une réponse économique de la part des gouvernements et des banques centrales face aux défis posés par la pandémie de COVID-19. La politique monétaire, dirigée par les banques centrales, a joué un rôle crucial dans la stabilisation des marchés financiers et la stimulation de l'économie. La Réserve fédérale américaine, par exemple, a adopté des politiques monétaires accommodantes, abaissant les taux d'intérêt à des niveaux historiquement bas pour encourager les emprunts et les investissements. Le S&P 500, indicateur phare du marché boursier américain, a bénéficié de ces interventions politiques, enregistrant une forte croissance des rendements au cours de cette période. Malgré la récession économique de cette période, le rendement de nos actifs n'a cessé de croître jusqu'à aujourd'hui, avec une valeur de portefeuille dépassant les 1250\$ après janvier 2022.

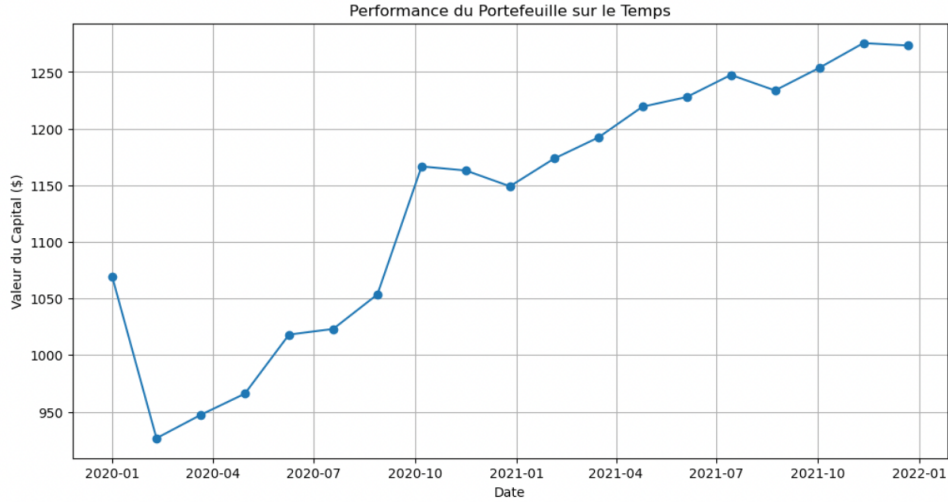


Table 2: Évolution du capital après investissement pour la période 2020 à 2021 du S&P500.

L'indice de performance est faible ou négatif. Il est plus bas lorsque nous diversifions moins notre portefeuille, ce qui correspond à une prise de risque plus importante.

## 4 Deuxième approche : Méthode

L'utilisation de méthodes de Machine Learning peuvent s'avérer être très pertinent dans notre cas. Selon nous, elles permettent de traiter efficacement des données financières, pouvant identifier les modèles complexes et également détecter les relations non linéaires entre différents actifs.

Pour approfondir notre analyse, nous avons donc opté pour une seconde approche pour estimer la causalité entre les rendements des actifs. Cette démarche repose sur l'utilisation d'un modèle de machine learning, en l'occurrence le Random Forest, pour explorer et exploiter les relations temporelles au sein des séries chronologiques financières. Nous avons créé une fonction qui prend en entrée les données de rendement financier. En appliquant la méthode du Random Forest à chaque actif, nous avons calculé l'importance des caractéristiques (actifs retardé en variables explicatives) de chaque modèle associé respectivement à un actif spécifique pour construire une matrice d'adjacence, représentant ainsi la causalité entre chaque rendement des actifs dans un contexte temporel. Pour cela, la méthode "features\_importances\_" du RandomForestRegressor a été utilisée. Il convient de noter que nous avons pris soin de normaliser les scores d'importance des caractéristiques afin de permettre une comparaison plus intuitive entre différentes caractéristiques.

Dans cette méthode, nous avons commencé par calculer quels étaient les actifs dont le passé impacte le plus le présent d'un autre. Après avoir fait cela, nous souhaitons garder les actifs avec les liens les moins forts. Nous avons donc établi un faible seuil de 0,1 afin de ne sélectionner que les actifs en dessous de ce dernier, et



donc ne garder que ceux faiblement corrélés.

#### 1. 2008-2009 : 22/08 jours

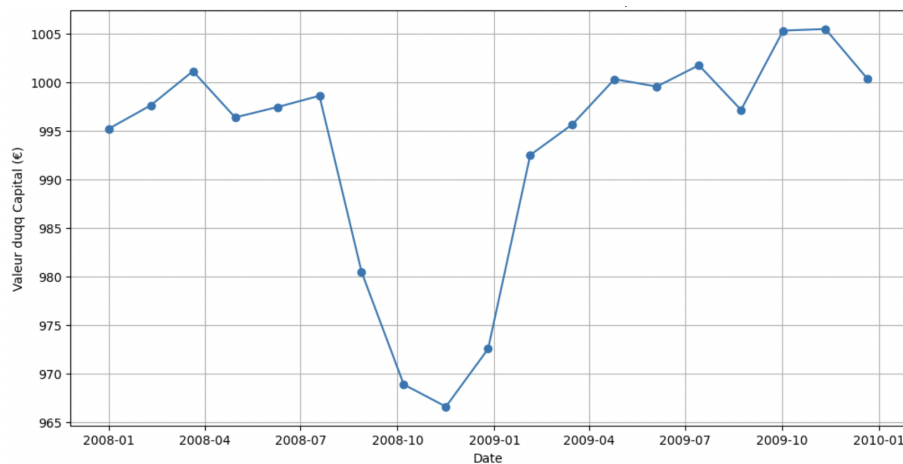


Table 3: Évolution du capital après investissement pour la période 2008 à 2010 du S&P500.

#### 2. 2020-2022 : 22/08 jours

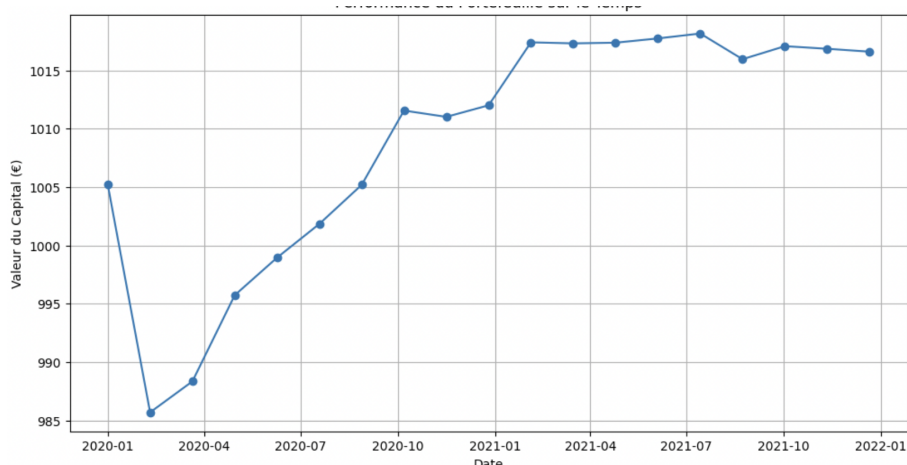


Table 4: Évolution du capital après investissement pour la période 2020 à 2022 du S&P500.

Les résultats de cette méthode semblent être très similaire à la première testé. En effet, les périodes de baisse pour les deux méthodes sont identiques. Cependant, durant les deux périodes de crise, le premier modèle perd beaucoup plus d'argent lors du pique de baisse que le second, mais regagne également beaucoup plus. En effet, on voit par exemple que durant le mois d'Octobre 2008, le capital vaut environ 910\$ contre 970\$ pour le Random Forest. Le même constat est fait pour la période du COVID-19, ou le pique de perte de Mars 2023 est représenté par un capital de 920\$ pour la première méthode et 985\$ pour la deuxième, sans regagner tout autant après.

Le premier modèle est donc plus "risqué" que le second, car il réagit plus sensiblement aux chocs exogènes. Un investisseur averse au risque favorisera donc plutôt la deuxième méthode.

## 5 Conclusion

Notre projet est axé sur la sélection d'actifs les moins systémiques dans la gestion de portefeuille en utilisant deux approches distinctes. La première méthode, basée sur le calcul de la causalité de Granger, a montré des performances supérieures en termes de gains, mais également une sensibilité accrue en période de crise par rapport à la méthode de causalité avec la régression random forest, qui présentait une légère réduction du risque pendant ces périodes difficiles. Ces deux approches, bien que différentes, ont montré des similitudes en raison de l'utilisation de la théorie des graphes dans le calcul du coefficient de clustering.

## References

### 1. Clustering in Complex Directed Networks

Author: Giorgio Fagiolo

Source: Physical Review E. - September 2007

Url : *(PDF) Clustering in Complex Directed Networks (researchgate.net)*

### 3. Measurement of Linear Dependence and Feedback between Multiple Time Series

Author : John Geweke

Source : Journal of the American Statistical Association - 1982

### 4. Investigating Causal Relations by Econometric Models and Cross-spectral Methods

Author : C. W. J. Granger

Source: Econometrica , Aug., 1969, Vol. 37, No. 3 (Aug., 1969), pp. 424-438

Url: <https://www.jstor.org/stable/1912791>

### 5. Portfolio Selection

Author : Harry Markowitz.

Source : Journal of Finance,, 77-91.(1952).

Url: [https://www.jstor.org/stable/2975974?metadata=1&no\\_toc\\_contents](https://www.jstor.org/stable/2975974?metadata=1&no_toc_contents)

### 6. Performance Measurement in a Downside Risk Framework

Author : Frank A. Sortino and Lee N. Price

Source : The Journal of Investing Fall 1994

Url : <https://joi.pm-research.com/content/3/3/59>