

Anisoara ABABII, Gaoussou DIAKITE, Eunice KOFFI
Suivi par Rania KAFFEL

Projet Finance Quantitative

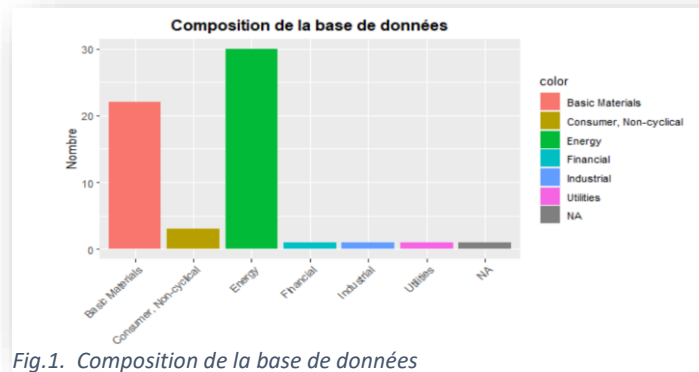
Master 2 MoSEF 2022-2023, Paris 1 Panthéon Sorbonne

Table des matières

Partie 1 : DESCRIPTIF DES DONNEES	2
1.Nettoyage, sélection des données et statistique descriptive.....	2
2.Performance cumulée des titres (base 100).	4
3.Corrélation des titres financiers	4
4.Indicateurs synthétiques du risque.....	6
PARTIE 2 : CLUSTERING	9
1.Clustering 1 : CAH basé sur les ROIC	9
2.Clustering 2: K-means basé sur les returns.....	12
PARTIE 3 : MEAN-VARIANCE	13
1.Représentation de l'ensemble des actions sur un même plan rendement volatilité.	13
2.Représentation de l'ensemble des portefeuilles qu'on peut constituer.....	14
3.Le portefeuille à variance minimale	15
4.Ratio de Sharpe, la rentabilité et volatilité du portefeuille équi pondérant.	17
PARTIE 4 : PORTFOLIO CONSTRUCTION AND MONITORING	17
1.Arbre CART	18
2.ARIMA.....	20

Partie 1 : DESCRIPTIF DES DONNEES

Nous allons, dans cette partie, évoquer une analyse empirique qui consiste à réaliser une étude comparative statique des séries des rendements de l'indice « S5ENRS Index » et de différents 56 titres desquels nous disposons. Pour la partie statique de l'étude, nous allons analyser la distribution des rendements des différents titres et tester quelques hypothèses sur ces rendements. Ensuite, nous étudierons la relation entre ces indices via leur corrélation ainsi que par quantile (VaR). Enfin, nous mesurerons et comparerons, grâce à divers indicateurs, la performance financière de ces actions.



1. Nettoyage, sélection des données et statistique descriptive

Dans le premier temps, dans cette sous-section, nous voulons mettre l'accent sur le fait qu'il existe différents critères utilisés pour un investisseur lors de la sélection de son titre / de son portefeuille comme : choix sur le meilleur ratio rendement/risque, ratio Sharpe /Treynor, le bêta, les frais courants, l'investissement minimum, la maximum « drawdown », etc.. Dans le cadre de notre étude des Returns, l'une des premières difficultés a été l'imputation des valeurs manquantes pour les titres qui disposent de moins de 50% de NaN (valeurs non-renseignées). De ce fait, en ayant 2 actions qui disposaient de plus de 70% de NaN, nous avons décidé de les supprimer du fait qu'aucun investisseur n'aura intérêt à investir dans un titre qui apporte si peu d'information (ce sont : DOW US et CTVA US). D'un point de vue financière, il reste très risqué d'investir dans des titres « à l'aveugle », sans disposer du maximum d'information pour en pouvoir attirer le profit dans le délai le plus court possible. Donc, plus précisément, NTR CT et NTR US sont les deux titres pour lesquels nous avons décidé de faire une imputation par la médiane sur une période de 3 mois (de 09/2017 à 11/2017 inclus). Nous avons opté pour une imputation par la médiane car, les valeurs extrêmes sont moins sensibles à ce type d'imputation. Et vu aussi que nous sommes face à une analyse des séries temporelles, elle (l'imputation par la médiane) reste une meilleure représentation de la majorité des valeurs de la variable, comparaison à la moyenne.

Dans un second temps, il est important de tester si les rendements suivent une distribution normale. Pour cela nous avons effectué un test de normalité de Jarque-Bera¹, test qui est généralement utilisé pour tester si une série chronologique de rendements financiers est normalement distribuée ou pas. Ce test se base alors sur la valeur de la Skewness et de Kurtosis, qui sera appliqué aux rendements de chaque indice. La Skewness et Kurtosis sont deux mesures statistiques qui nous permettent d'évaluer deux dimensions essentielles du risque : l'asymétrie de la distribution des rendements (Skewness) et les événements extrêmes (Kurtosis/ coefficient d'aplatissement)². La conclusion tirée à partir de ce test nous la voyons aussi pour les représentations des histogrammes, ils étant centrés à 0 pour une bonne partie des titres, et pour l'autre non.

L'hypothèse testée de JB sont :

¹ Le test est implémenté dans R à l'aide de la fonction `jarque.bera.test()` de la bibliothèque `tseries`.

² A noter qu'une distribution Normale possède une Skewness égale à 0 et une Kurtosis égale à 3.

$$\left\{ \begin{array}{l} \mathbf{H0:} \text{ Return}_i \sim N(\mu, \sigma^2), \quad \text{où encore tester si} \\ \text{l'échantillon a les propriétés d'asymétrie et d'aplatissement d'une distribution normale.} \\ \mathbf{H1:} \text{ Return}_i \text{ ne suit pas une loi Normale} \end{array} \right.$$

Dans le cas des rendements boursiers : *NEM US, ABX CT, TTE FP, COP US, CVE CT, ALA CT, WLK US, GLEN LN, MPC US, WY US, ET US, VNOM UW, SUN US, PBA US, OXY US, OKE US, PXD US, TRGP US, SLB US, DVN US, HES US, MRO US, WMB US, CTRA US, APA US, KMI US, EQT US, HAL US, FANG US* et le benchmark *S5ENRS Index*, la valeur de la p_value du test de JB est inférieure à 0,05, cela signifie que les rendements ne sont pas distribués normalement, ce qui suggérerait que le cours de l'action ne suit pas une marche aléatoire et que la série de rendements pourrait avoir des modèles prévisibles.

Tandis que pour les titres : *IP US, UPM FH, XOM US, VLO US, NUE US, FMC US, FCX US, ADM US, PKX US, BHP US, TECK/B CT, RIO US, WIL SP, MNDI LN, AAL LN, MOS US, PSX US, WRK US, AA US, MTS SQ, NTR CT, NTR US, CVX US, BKR US, EOG US*, la valeur p est supérieure à 0,05, cela signifie que les rendements sont distribués normalement, ce qui suggère que le cours de l'action suit une marche aléatoire et qu'il n'y a pas de schémas prévisibles dans la série de rendements.

A partir d'une fonction ACF (Autocorrélation Fonction), nous remarquerons que les séries sont auto-corrélées puisque la corrélation entre les observations à des temps différents est significative. Autrement dit, les observations à des temps différents sont liées les unes aux autres.

Pour la description statistique des titres (dans cette partie), nous avons fait une hypothèse préliminaire de considérer que chaque titre représente un portefeuille. Donc, à partir de cette optique, nous avons réalisé une analyse sommaire (Min, Max, Moyenne, Variance, Volatilité, Kurtosis, Skewness) par portefeuille (c'est-à-dire, par titre) sur une période de 2017-2022.

equities <tbl>	Minimum <tbl>	Maximum <tbl>	Moyenne <tbl>	Variance <tbl>	Volatilité <tbl>	Kurtosis <tbl>	Skewness <tbl>
AA US Equity - Basic Materials	-55.5876	54.0248	1.4041	369.8376	19.2312	0.5997	-0.0495
AAL LN Equity - Basic Materials	-27.0928	26.0929	1.9244	110.5664	10.5151	-0.1155	-0.2656
ABX CT Equity - Basic Materials	-19.4732	41.1102	0.6453	114.0671	10.6802	2.1145	0.9820
ADM US Equity - Consumer, Non-cyclical	-15.2208	15.0542	1.5308	44.0963	6.6405	-0.0072	-0.5295
ALA CT Equity - Utilities	-39.4070	33.6272	0.7707	114.6954	10.7096	3.6078	-0.1858
APA US Equity - Energy	-83.2263	213.8315	4.1496	1133.4052	33.6661	23.0360	3.7342
BHP US Equity - Basic Materials	-15.8096	17.1178	1.5459	67.0650	8.1893	-0.6328	-0.2527
BKR US Equity - Energy	-34.7421	32.8571	0.3261	202.7025	14.2374	-0.3260	0.1712
BP/ LN Equity - Energy	-17.4222	32.5626	0.3407	67.4887	8.2152	2.5450	0.6820
COP US Equity - Energy	-36.3899	38.2250	2.2204	156.5806	12.5132	1.8058	0.3612

1-10 of 56 rows

Previous 1 2 3 4 5 6 Next

Fig.2. Statistique descriptive des actions sur la période 08/2017-08/2022

En termes générales, les titres sont considérés comme plus volatils lorsqu'ils ont des fluctuations de prix plus importantes et plus fréquentes, exactement comme dans notre cas. Cela signifie qu'ils peuvent augmenter ou diminuer de valeur rapidement et de manière imprévisible. Etant face à des actions, elles sont donc plus volatiles que les obligations ou autres actifs financiers. Cela peut signifier que les investisseurs ont tendance potentiellement à gagner ou perdre de l'argent plus rapidement sur les actions que sur d'autres types d'investissements, mais ils peuvent également être plus risqués.

Dans ce contexte d'analyse des titres financiers, la skewness et le kurtosis sont des métriques qui permettent de caractériser la distribution des rendements des titres financiers, et qui peuvent communiquer des informations sur le risque associé à ces titres. Ensuite une skewness positive nous a donné l'indication que les rendements supérieurs à la moyenne sont plus fréquents que les rendements inférieurs à la moyenne, et une skewness négative indique le contraire. Nous avons également observé le Kurtosis des titres. Alors, le kurtosis supérieure à 3 indique que les rendements extrêmes sont plus fréquents que prévu par une distribution normale, ce qui peut indiquer un risque plus élevé pour les investisseurs.

2. Performance cumulée des titres (base 100).

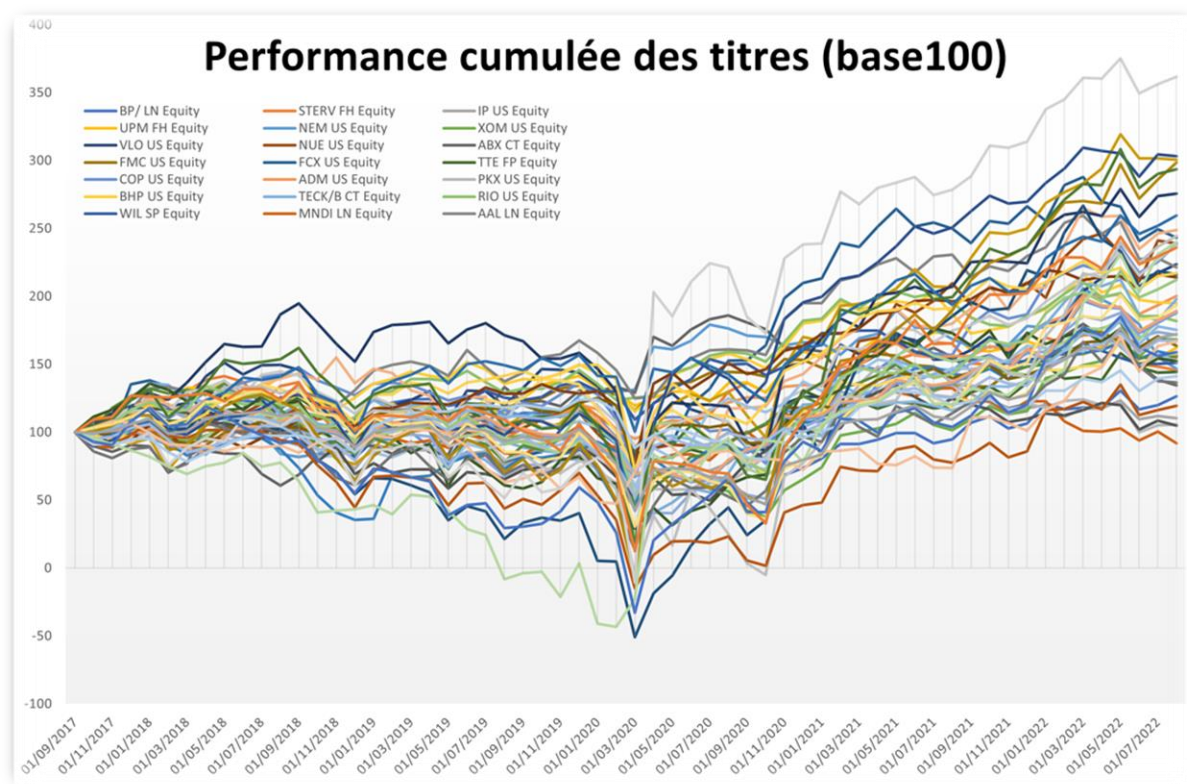


Fig.3. Performance cumulée des titres (base 100) (Graphique réalisé en EXCEL)

D'après le graphique obtenu plus haut, nous pouvons faire la conclusion que la crise de la COVID-19 a eu un impact important sur les rendements des titres financiers au monde entier. Lorsque la pandémie a commencé à se propager à l'échelle mondiale en 2020, les marchés boursiers ont connu une forte volatilité, vu des baisses très importantes, comme nous pouvons le remarquer également sur la représentation graphique. Beaucoup de secteurs économiques ont été fortement touchés par les fermetures d'entreprises et par les restrictions imposées par les états, ce qui a entraîné comme conséquence une dégradation de confiance des investisseurs. Les actions des entreprises liées aux énergies, ont été fortement impactées, avec des baisses importantes de rendements, alors que les actions des entreprises liées à l'industrie ont plutôt bien résisté. Donc, au niveau mondial, la crise de la COVID-19 a eu un impact négatif sur les rendements de la plupart des titres financiers, en raison de la réduction de la confiance des investisseurs et de la forte incertitude économique. Cependant, certains secteurs ont été moins touchés que d'autres. A partir de la fin d'année 2020, les rendements des titres ont commencer à augmenter. Alor, comme nous pouvons observer sur le graphique, en s'appuyant sur une base 100, le titre qui a réussi à rattraper le plus vite possible la crise et se manifester en première sur la dernière année est APA US, comparaison à SLB US qui reste en dernière par rapport au rendement des autres actifs financiers (les deux actions appartiennent au secteur Energie).

3. Corrélation des titres financiers

La matrice de corrélation entre les titres financiers réalisée est utilisée pour mesurer la relation linéaire entre deux titres, étant utilisée comme outil d'aide à la décision d'investissement. Les actions avec un coefficient de corrélation de 1 indiquent une corrélation parfaite positive, c'est-à-dire que les deux titres varient de manière similaire dans le temps, tandis qu'un coefficient de corrélation de -1 indique une corrélation parfaite négative, c'est-à-dire que les deux titres varient de manière inverse l'un de l'autre. Un coefficient de corrélation de 0 indique qu'il n'y a pas de corrélation entre les deux titres.

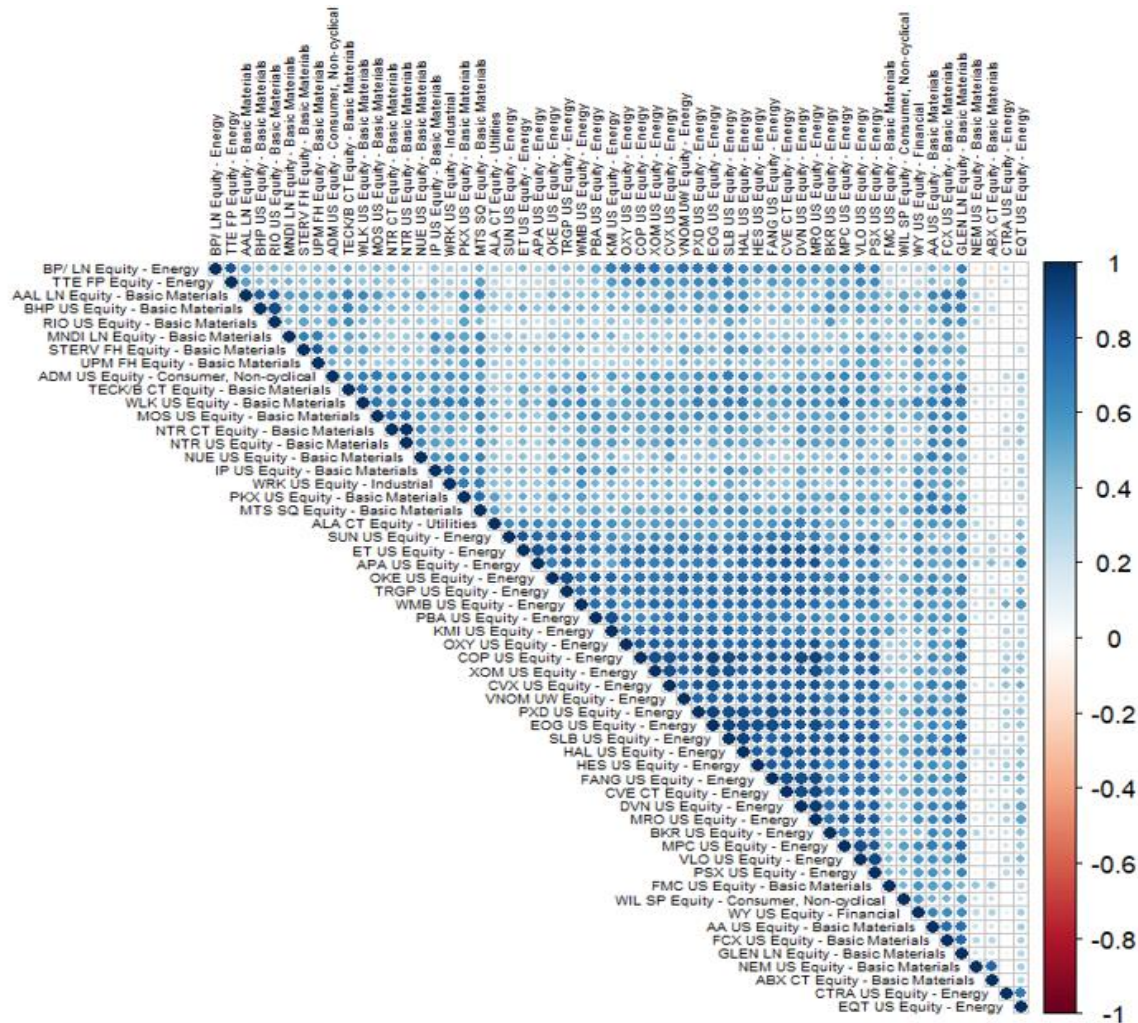


Fig.4. Matrice de corrélation des titres

Pour conclure, les actions qui appartiennent au même secteur sont très corrélées entre elles. Nous distinguons uniquement 4 titres financiers qui ont un coefficient de corrélation faible par rapport aux autres (NEM US-Basic Materials (1921), ABX CT - Basic Materials (1983), CTRA US – Energie (1925), EQT US – Energie (1894)). Cela peut être dû à l'année d'apparition de ces entreprises, elles étant des entreprises qui datent après la première guerre mondiale (1914-1918).

4. Indicateurs synthétiques du risque

	Sharpe_ratio <dbl>	Treynor_ratio <dbl>	Calmar_ratio <dbl>	Sortino_ratio <dbl>	VaR_hist <dbl>	Max_drawdowns <dbl>
BP/ LN Equity - Energy	0.03709607	-1.20980227	1.834373e-04	0.06157247	-13.7668	-1.857583e+03
STERV FH Equity - Basic Materials	0.05188998	-1.11215106	8.315754e-06	0.07941640	-15.3464	-6.300405e+04
IP US Equity - Basic Materials	-0.03391992	-3.12306755	-4.837773e-08	-0.04784963	-12.6739	-4.685962e+06
UPM FH Equity - Basic Materials	0.11900928	-0.35810264	5.791805e-04	0.17127960	-12.6632	-1.612608e+03
NEM US Equity - Basic Materials	0.08653154	-2.30988433	1.270885e-03	0.15109233	-11.4248	-6.515713e+02
XOM US Equity - Energy	0.10044684	-0.06233229	1.062104e-08	0.15566521	-14.0551	-9.346614e+07
VLO US Equity - Energy	0.13075201	0.72863522	1.944109e-04	0.24627999	-19.9502	-9.857077e+03
NUE US Equity - Basic Materials	0.16140153	1.67747275	4.401823e-06	0.30862788	-15.8821	-4.537786e+05
ABX CT Equity - Basic Materials	0.05705127	-2.98544550	3.000677e-05	0.11859182	-13.8337	-2.150581e+04
FMC US Equity - Basic Materials	0.12493189	-0.48934278	4.949052e-06	0.19706128	-11.8128	-1.799558e+05

Fig.5 Indicateurs synthétiques du risque

- Ratios de Sharpe

Pour mesurer le rendement ajusté au risque d'investissement, nous avons opté pour le ratio de Sharpe (en se servant de la fonction du logiciel R *SharpeRatio*). Ainsi, ce ratio est calculé sur toute la période étudiée, en divisant le rendement excédentaire d'un investissement (le rendement de l'investissement moins le taux sans risque (ici égal à 0.036) par l'écart type des rendements de l'investissement.

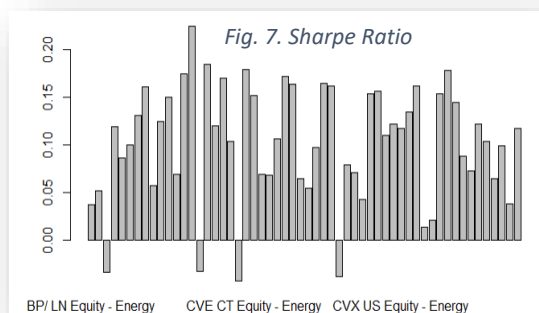
Après une analyse plus minutieuse, nous remarquons que parmi les 56 titres financiers, il y a uniquement 4 titres

Equities with a negative Sharp Ratio :			
{r}			
ratios_2[ratios_2[,1] < 0,]			
sharp.IP US Equity - Basic Materials -0.0339	sharp.PKX US Equity - Basic Materials -0.0330	sharp.MNDI LN Equity - Basic Materials -0.0427	sharp.WRK US Equity - Industrial -0.0382

Fig. 6. Les titres financiers avec un ratio de Sharp négatif

qui ont un ratio de Sharp négatif (trois en représentant les matériels basiques et un, l'industrie).

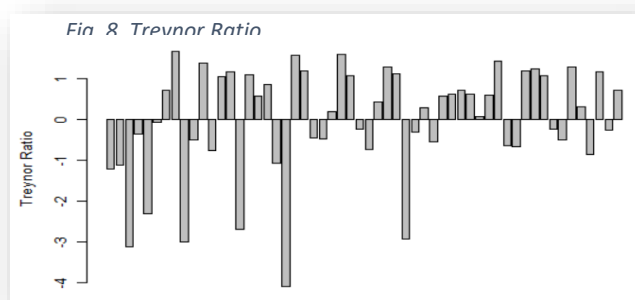
On ne peut surtout pas considérer que les ratios desquels nous disposons sont bons parce qu'ils sont tous



inférieurs à 1. Nous nous sommes conduits sur le fait qu'un ratio de Sharpe de plus +1 est généralement considéré comme un bon ratio, car il indique que l'investissement offre un niveau de rendement relativement élevé pour le niveau de risque pris, c'est-à-dire que plus le ratio est élevé, plus le risque est élevé et plus le rendement potentiel est élevé (donc un meilleur rendement ajusté au risque). Donc, même si la majorité des actions individuelles ont un rendement plus élevé de 0, elles ont également un écart type plus élevé comparaison aux autres.

- Ratios de Treynor

Le ratio Treynor se ressemble beaucoup à celui de Sharpe, juste que pour son calcul, il est utilisé le Beta du portefeuille comme dénominateur au lieu de l'écart-type et il représente la mesure de la performance d'un investissement ou d'un portefeuille qui prend en compte le risque de l'investissement. Dans notre analyse, nous l'utilisons pour évaluer le rendement d'un portefeuille (titre individuel) par rapport à la quantité de risque systématique que le portefeuille prend, qui est mesuré par le bêta du portefeuille/ou d'investissement³. Nous avons remarqué que les titres provenant du secteur de l'Energie ont un ratio compris entre -0,5 et 1,5. Comme précédemment, un ratio plus élevé est généralement considéré comme meilleur, car il indique que le portefeuille a généré des rendements plus élevés pour un niveau de risque donné. Également, on insiste sur le fait qu'un ratio Treynor élevé signifie que le portefeuille a bien performé par rapport à l'indice de référence SP500. Prenons le cas de l'action NUE US qui enregistre tout au long de la période un ratio Treynor égal à 1,67. Cela signifie que les rendements ajustés au risque du portefeuille sont 1,67 fois supérieurs à ceux de l'indice de référence. Pour cette analyse exploratoire, nous utilisons TreynorRatio pour comparer les performances de différents investissements que nous disposons.



- Calmar Ratio

Le ratio de Calmar évalue l'efficacité des fonds en comparant le rendement moyen par rapport à son indice de référence à sa perte la plus grande sur la période donnée. Il ne peut être interprété que s'il y a une surperformance positive. Comme nous pouvons le remarquer sur le graphique, que ces ratios pour nos actions sont autour de 0.001. De manière visible, les titres pris individuellement sur une période si large ne surperforment pas et qui sont un signe que nous indiquent que c'est encore risqué d'investir dans ces titres

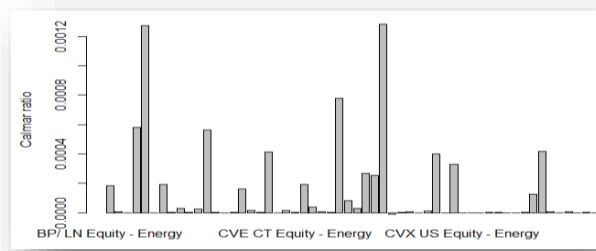


Fig.9. Calmar Ratio

séparément. La solution est plutôt d'avoir un portefeuille de plus en plus diversifié et de faire de la diversification et de réduire la corrélation sur certains titres (référence aussi à la figure 4).

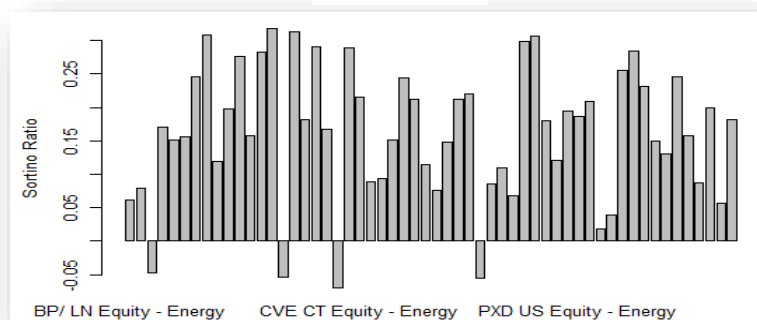
- Sortino ratio

³ Le ratio Treynor est calculé comme le rendement excédentaire du portefeuille sur le taux sans risque, divisé par le bêta du portefeuille. Le bêta mesure la sensibilité des rendements d'un investissement aux variations du marché global. (Nous avons utilisé le package PerformanceAnalytics pour estimer la bêta).

Le ratio de Sortino est un indicateur de performance, et plus précisément de la performance d'un portefeuille ou d'une stratégie d'investissement qui mesure la performance par rapport au risque. Il se concentre sur les pertes sous performantes et non sur les gains. En comparaison aux ratios précédents, le ratio de Sortino est calculé en divisant le rendement moyen du portefeuille par l'écart à la baisse, qui est l'écart type des rendements du portefeuille en dessous du rendement cible. Par exemple, un ratio de 0,151 (NEM US) signifie que le portefeuille ou la stratégie a généré un rendement de 0,151 pour chaque unité de risque prise, parce que plus le ratio de Sortino est élevé, plus le rendement est élevé pour chaque unité de risque prise.

Dans ce cas, l'actif financier a un rendement moyen de 0,151 fois l'écart à la baisse, ce qui suggère que l'actif en question génère des rendements relativement élevés tout en encourant un risque de baisse relativement faible.

Fig. 10. Sortino



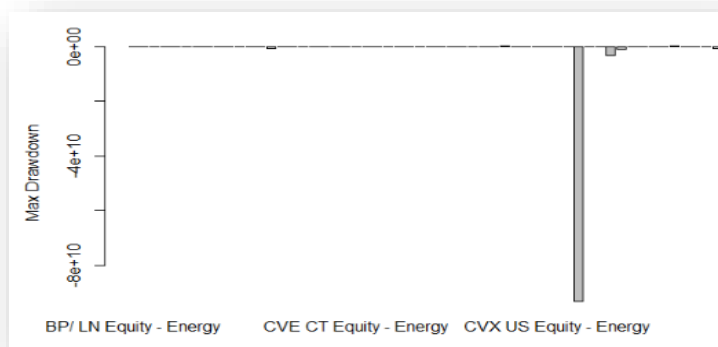
Nous avons insisté sur le calcul de ce ratio car il appartient à la catégorie des ratios asymétriques de rendement/risque relatifs, où la méthode utilisée pour mesurer le risque ne tient compte que d'une partie de la volatilité (asymétrie), à la différence du ratio de Sharpe qui est symétrique.

D'après le graphique, il y a plus des titres financiers dont leurs ratios Sortino ont une valeur supérieure à 0. Cela signifie que leur rendement est supérieur au seuil de risque choisi. En d'autres termes, pour chaque unité de risque prise, ces titres financiers génèrent un rendement positif (titres du secteur Energie, Basic Matériels). Un ratio de Sortino élevé peut indiquer que les investissements dans ces titres financiers sont relativement plus sûrs et peuvent offrir un meilleur rendement pour chaque unité de risque prise par rapport à d'autres investissements. Toutefois, il est important de prendre en compte d'autres facteurs pour évaluer la performance et la viabilité d'un investissement.

- **Max drawdown**

Pour l'analyse de nos actions, nous avons également opté pour le calcul du ratio Max drawdown, parce qu'il représente la perte la plus importante subie par un fonds sur la période déterminée. Dans ce cas, il sera pas mal à mesurer l'écart entre le point le plus haut et le plus bas de la période, en pourcentage (%). C'est pour cette raison que nous avons choisi à évaluer la performance d'un titre, en surveillant le drawdown respectif sur différentes périodes (2017-2022).

Fig. 11. Max Drawdown



- *La valeur à risque (VaR)*

La valeur à risque (VaR) est une mesure du risque de perte d'un investissement ou d'un portefeuille. Il représente la perte maximale qu'un investissement devrait subir sur un horizon temporel donné avec un niveau de confiance donné. Nous utilisons la VaR pour gérer et savoir quel serait le risque de notre portefeuille. La méthode de calcul de la VaR, que nous avons choisi est la méthode historique, qui est basée sur la distribution historique des rendements de l'investissement et ne tient pas compte d'éventuelles évolutions futures du marché. Pour la calculer nous avons pris en compte la série chronologique de rendements pour l'investissement et nous avons estimé la distribution de ces rendements. Prenons un exemple : un VaR de -13,76 (BP/LN Equity) signifie qu'il existe un certain niveau de confiance (généralement 95%) que la perte potentielle du portefeuille ne dépassera pas -13,76. Il est également important de garder à l'esprit qu'un nombre négatif dans la VaR indique que le portefeuille perd de l'argent, une VaR de -13,76 signifie qu'il y a 5 % de chances de perdre 13,76 % de la valeur du portefeuille en investissant dans cet actif.

PARTIE 2 : CLUSTERING

Dans cette partie, nous travaillons principalement avec les bases de données **Returns** et **ROIC**. Il s'agit de rapprocher les titres et de les regrouper dans des classes afin de comprendre leur similarité et différences les uns des autres.

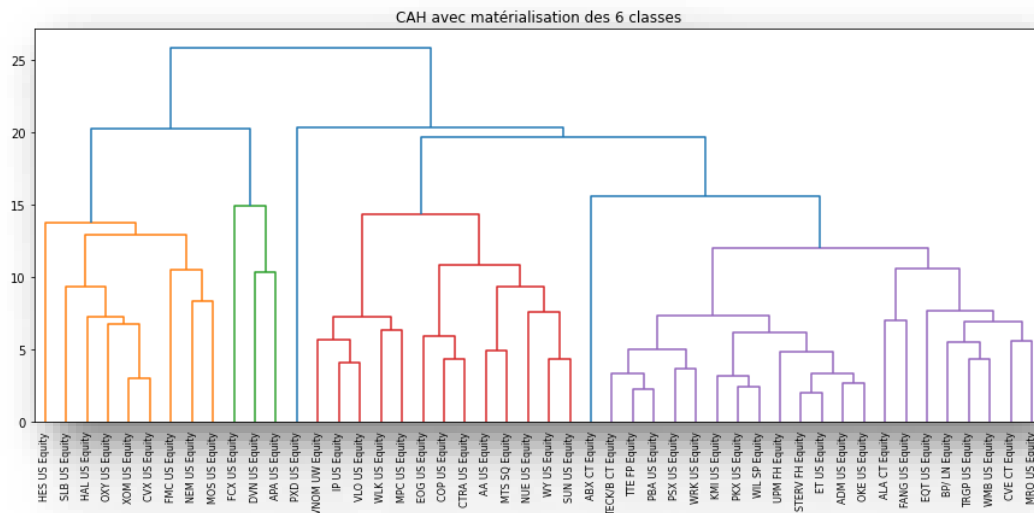
1. Clustering 1 : CAH basé sur les ROIC

La base ROIC contient la **rentabilité trimestrielle** de Septembre 2010 à Juin 2022. Nous avons en lignes les dates et en colonnes les titres.

Pour commencer, nous supprimons les titres pour lesquels nous avons plus de 50% d'informations sont manquantes. Puis, nous imputons les autres valeurs manquantes par la médiane. La configuration de la table, ne nous permettant pas de réaliser notre analyse simplement, nous allons la transposer. Nous obtenons finalement, une base de données avec 48 titres et 48 colonnes.

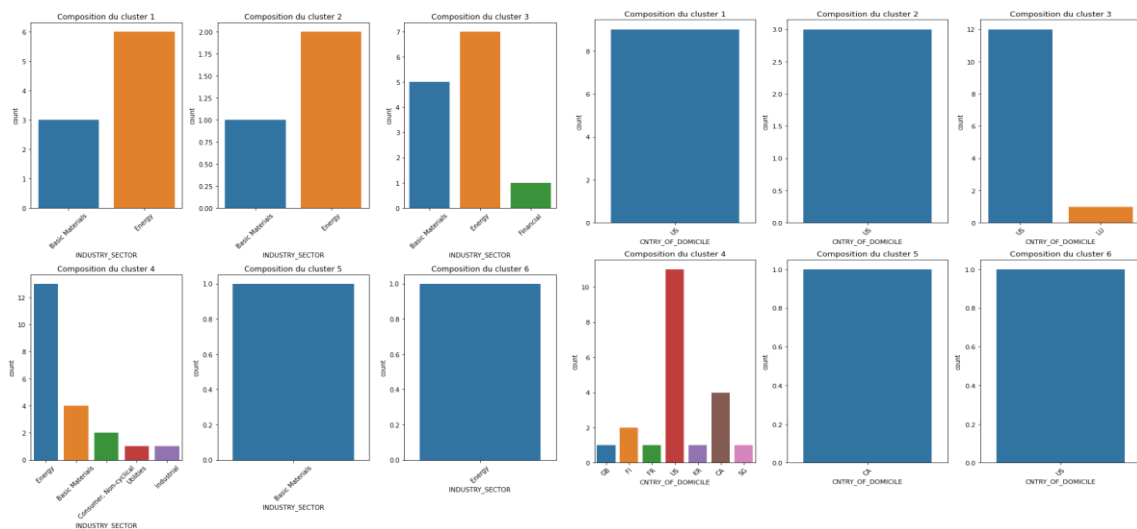
A cette base de données, nous appliquons une méthode de Classification Ascendante Hiérarchique.

Après analyse, nous décidons de répartir nos titres dans 6 clusters.



Nous pouvons constater sur ce graphique, une répartition inégale du nombre de titres par clusters. En effet, nous avons des clusters qui comportent un seul élément. Pour mieux comprendre la répartition des titres dans les clusters, nous allons effectuer quelques analyses.

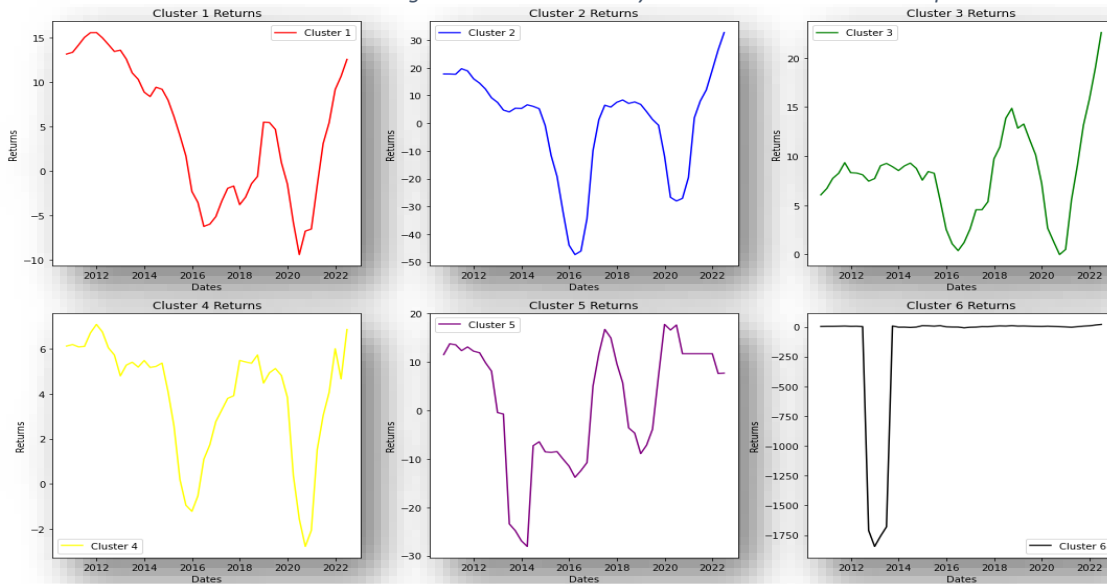
Figure 12 : Constitution des clusters



Dans tous les portefeuilles, à l'exception du portefeuille 5, nous avons la présence de titres d'énergie. Nous pouvons qualifier le cluster 4 de portefeuille diversifié puisqu'il intègre des titres de différents secteurs, domiciliés dans différents pays avec des monnaies différentes.

Le ROIC du cluster 3 est toujours positif sur la période considérée.

Figure 13: rendement moyen du cluster au cours du temps



Nous avons des similitudes concernant les titres des différents clusters :

- Les effets de la crise COVID sont visibles sur les courbes. On a un pic à la baisse.

Nous constatons une différence entre les différents clusters à plusieurs niveaux :

- Au niveau de l'amplitude des variations : On remarque que les différents clusters ont des titres qui sont plus ou moins volatiles.
- Au niveau des échelles : Certains clusters contiennent des titres avec des rentabilités plus hautes que d'autres
- Au niveau des variations : Sur certaines années, les pics sont *corrélés* et sur d'autres clusters, on a ces mêmes variations avec un retard

A présent, essayons d'analyser la contribution des variables à la construction de la partition.

Pour vérifier si une variable quantitative influence la partition ou non, on va considérer la classe de chaque individu comme une variable qualitative et tester la liaison entre la variable qualitative la variable quantité. Cette méthode est connue sous le nom de ANOVA, ou *analyse de la variance à 1 facteur*. Il s'agit de vérifier si on peut considérer les moyennes conditionnelles d'une variable sur chaque classe comme égales.

Au seuil de 5%, on rejette l'hypothèse H0 selon laquelle les ROIC du 3^e trimestre 2010 n'ont pas d'influence

Figure 14 : Exemple de résultats de test ANOVA

2010-09-30 00:00:00					
	df	sum_sq	mean_sq	F	PR(>F)
cluster	1.0	398.943069	398.943069	18.205601	0.000098
Residual	46.0	1008.007453	21.913206	NaN	NaN
2010-12-31 00:00:00					
	df	sum_sq	mean_sq	F	PR(>F)
cluster	1.0	348.753968	348.753968	14.103775	0.000485
Residual	46.0	1137.474358	24.727703	NaN	NaN

sur la construction des classes. On rejette donc qu'il y'a indépendance entre les valeurs du ROIC de ce trimestre et la variable qualitative. On dit que cette variable caractérise la partition. Cela n'est pas vrai pour plusieurs

autres trimestres pour lesquelles le test est significatif. Ces trimestres sont les trimestres de **2012-09-30 à 2022-06-30**.

Les premiers trimestres sont donc les seuls qui contribuent le plus à la construction de la partition.

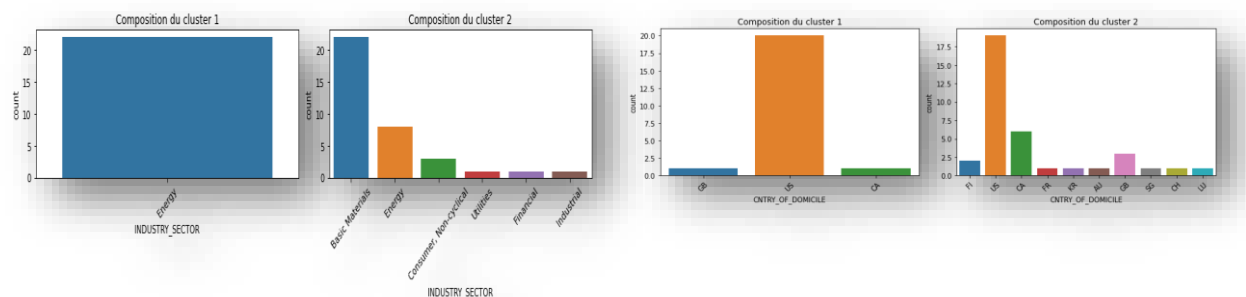
2. Clustering 2: K-means basé sur les returns

La base de données Returns contient la rentabilité mensuelle des titres du 31 Octobre 2019 au 30 Septembre 2022.

Nous avons en lignes les titres et en colonnes les dates. Après avoir imputé les valeurs manquantes par la médiane, nous supprimons de la base la ligne correspondant aux informations relatives à l'indice de référence. Elle contient finalement 58 titres et 60 colonnes.

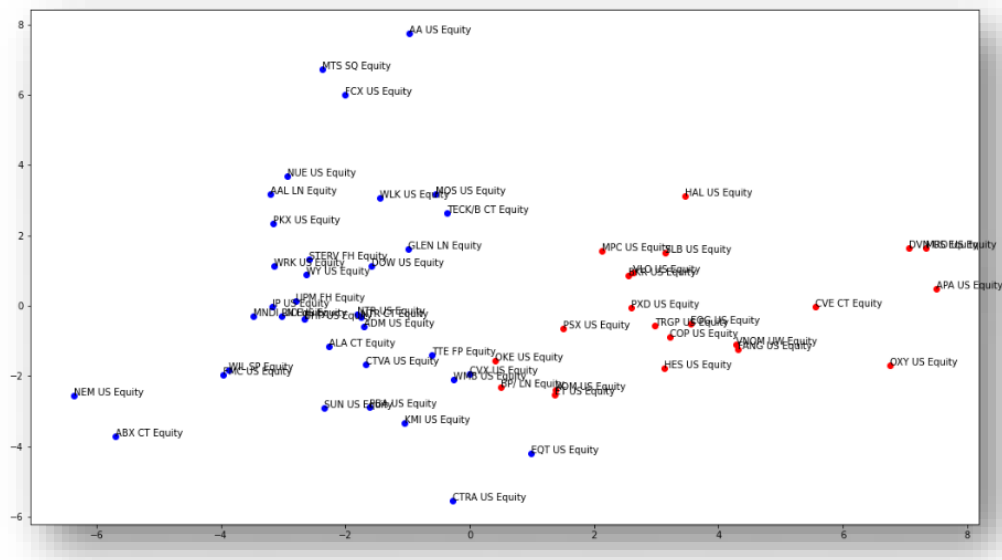
A cette base de données, nous appliquons une méthode de K-means pour regrouper les titres en deux Clusters. Les résultats obtenus sont présentés ci-dessous.

Figure 15 : Composition des clusters



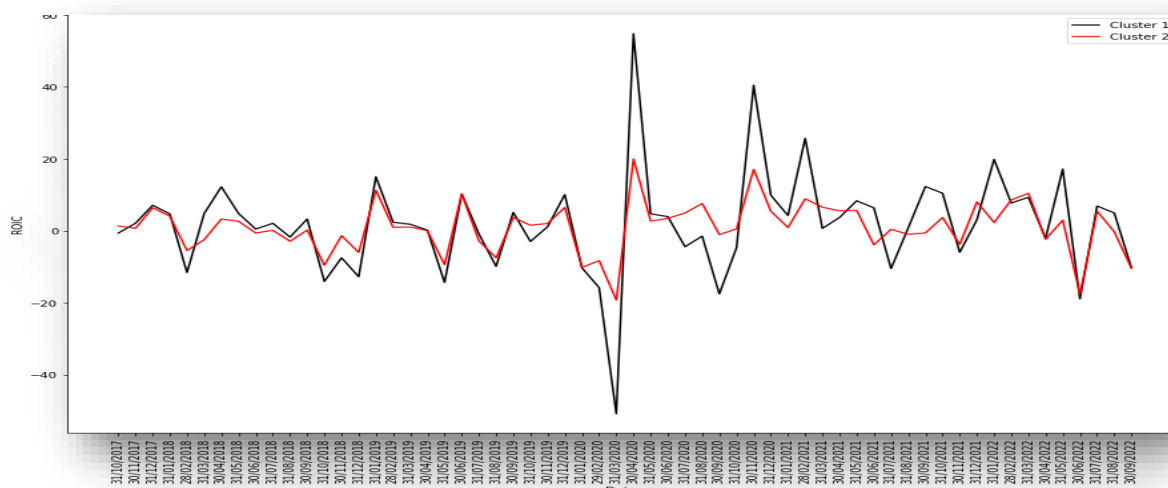
Nous constatons que le premier cluster est uniquement composé de titres du secteur de l'énergie tandis que le deuxième est diversifié.

Figure 16 : Représentation des clusters dans un plan à deux dimensions



Nous constatons que nous arrivons à bien différencier les deux clusters sur le plan formés des deux composantes principales issues de l'ACP.

Figure 17 : Returns moyennes en fonction des clusters



Nous constatons sur ce graphique que les titres du cluster 1 sont plus volatiles que les titres du cluster 1. Cependant les deux ROIC sont en moyenne nuls. Le cluster 1 est un portefeuille constitué de titres du secteur de l'énergie uniquement. C'est le portefeuille le plus volatile. Il a connu de fortes perturbations lors de la crise covid.

Dans tous les portefeuilles construits avec la CAH ou le K-means, les résultats nous montrent l'instabilité des titres de l'énergie et la sensibilité de ceux-ci face à un choc.

Par ailleurs, un portefeuille diversifié permet de diversifier de répartir les risques sur tous les titres.

PARTIE 3 : MEAN-VARIANCE

1. Représentation de l'ensemble des actions sur un même plan rendement volatilité.

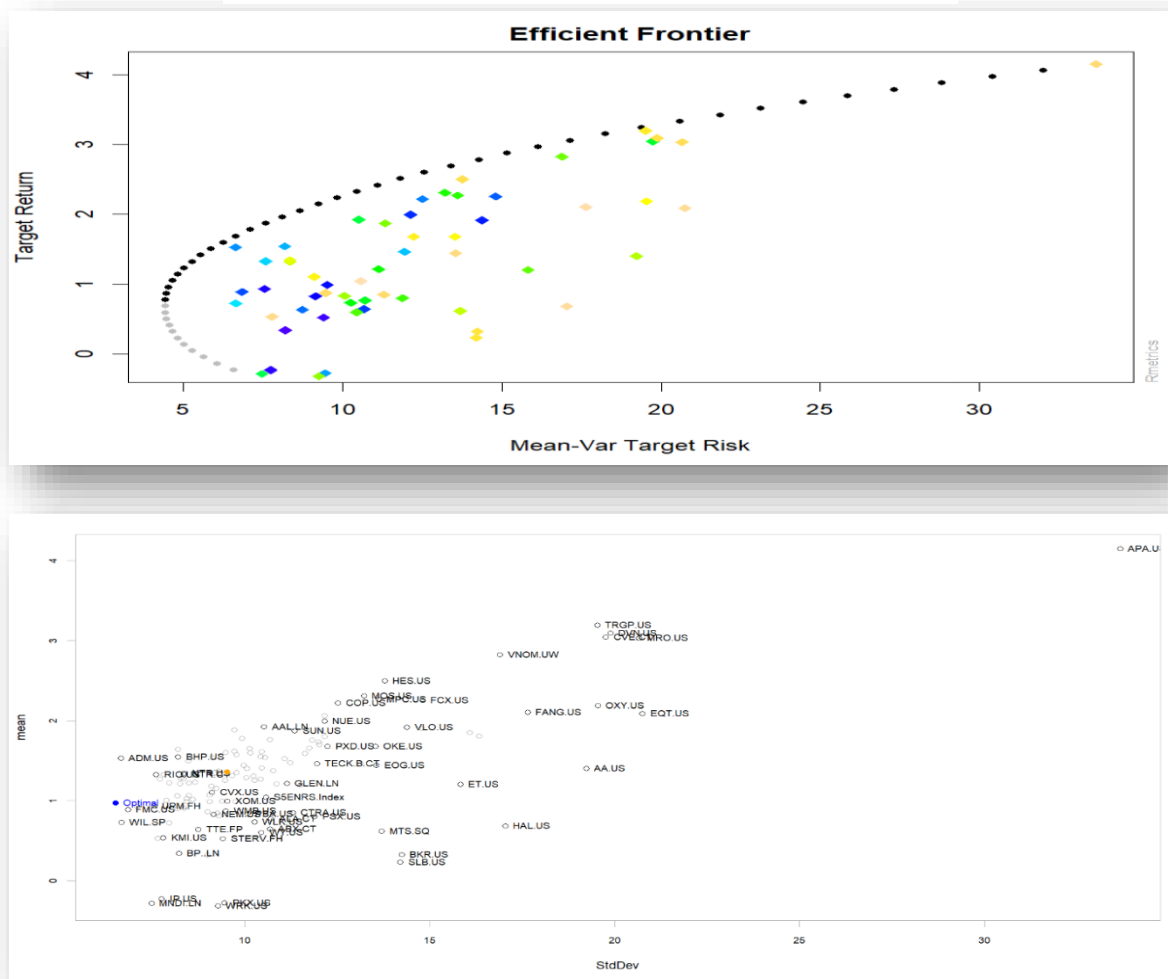
Un investisseur, dans un monde d'investissement suivant parfaitement la théorie de H. Markowitz doit s'intéresser au rendement espéré d'un portefeuille mais également au risque associé à ce dernier.

Dans cette section, on va représenter l'ensemble de nos titres sur un plan rendement volatilité. Pour se faire, on a utilisé la base de données Returns qui contient les rendements mensuels de nos actions.

Sur l'axe des abscisses on a la volatilité de nos actions et l'axe des ordonnées nous donne le rendement de ces actions. Pour la meilleure interprétabilité de ce graphique, on l'affiche avec le nom des stocks.

Dans ce graphique, il est clairement observable que, dans le cadre de la moderne portfolio théorie de Markowitz,

Fig. 18. L'ensemble des actions sur un plan rendement - volatilité



certains stocks sont considérés comme étant efficaces s'ils disposent d'un niveau de risque minimal pour un certain niveau de rendement ou encore s'ils disposent d'un rendement maximal pour un certain niveau de risque donnée. En d'autres termes, la frontière efficace, regroupe l'ensemble des actions optimales pour tout niveau de risque.

Prenant un exemple : Sur le graphique ci-dessus, les stocks encadrés en haut du graphique (HES.US, MOS.US.) et ceux du bas (BKR.US, SLB.US.) pour un même risque de 15% l'investisseur aura deux choix, un portefeuille qui a un rendement espéré de 2.5% et un deuxième qui a un rendement espéré de 0.5%. S'il est rationnel il choisira celui du premier, car pour un même risque donné ce portefeuille a un rendement plus élevé que le deuxième.

2. Représentation de l'ensemble des portefeuilles qu'on peut constituer

L'ensemble des portefeuilles que l'on peut constituer dépendent du rendement et du niveau de risque associé à chaque action. On peut utiliser la théorie moderne de la diversification pour sélectionner des actions qui ont des rendements non corrélés, ce qui permet de réduire le risque global du portefeuille. Cependant, il est important de noter que même avec une bonne diversification, il existe toujours un certain niveau de risque inhérent aux investissements en actions.

3. Le portefeuille à variance minimale

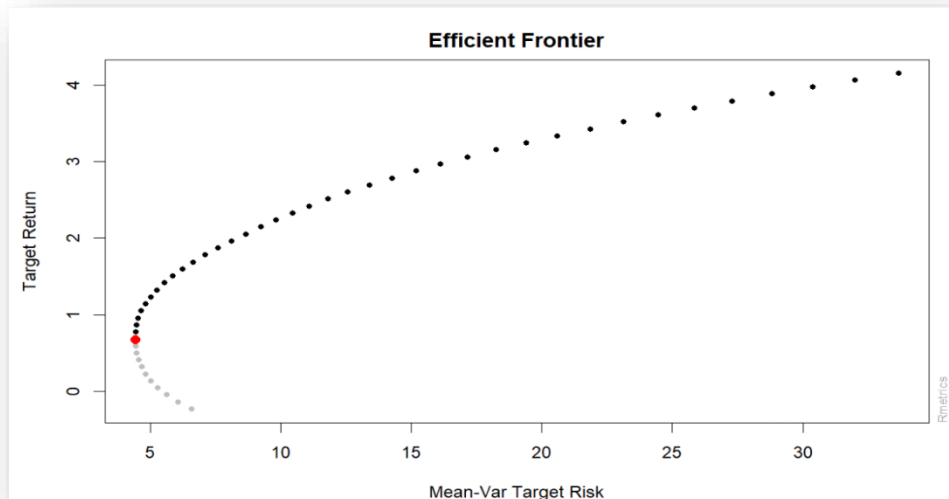
The figure consists of three vertically stacked plots illustrating portfolio optimization results.

Top Plot: Efficient Frontier
 The y-axis is labeled "Target Return" (ranging from 0 to 4) and the x-axis is labeled "Mean-Var Target Risk" (ranging from 5 to 30). The plot shows a series of points representing different portfolios, with a blue line indicating the Efficient Frontier. A legend on the right lists various assets, including BP.LN, STERV.FH, IP.US, UPM.FH, NEM.US, XOM.US, VLO.US, NUE.US, ABX.CT, FMC.US, FCX.US, TTE.FP, COP.US, ADM.US, PKX.US, BHP.US, TECK.B.CT, RIO.US, WIL.SP, MNDL.LN, and AAL.LN.

Middle Plot: Efficient Frontier
 The y-axis is labeled "Target Return[mean]" (ranging from 0 to 4) and the x-axis is labeled "Target Risk[Sigma]" (ranging from 0 to 30). The plot shows a series of points representing different portfolios, with a blue line indicating the Efficient Frontier. A legend on the right lists various assets, including BP.LN, STERV.FH, IP.US, UPM.FH, NEM.US, XOM.US, VLO.US, NUE.US, ABX.CT, FMC.US, FCX.US, TTE.FP, COP.US, ADM.US, PKX.US, BHP.US, TECK.B.CT, RIO.US, WIL.SP, MNDL.LN, and AAL.LN.

Bottom Plot: Weight
 The y-axis is labeled "Weight" (ranging from 0.0 to 1.0) and the x-axis is labeled "Target Return" (ranging from -0.226 to 4.15). The plot shows a stacked area chart representing the weights of various assets across the Efficient Frontier. A legend on the right lists various assets, including BP.LN, STERV.FH, IP.US, UPM.FH, NEM.US, XOM.US, VLO.US, NUE.US, ABX.CT, FMC.US, FCX.US, TTE.FP, COP.US, ADM.US, PKX.US, BHP.US, TECK.B.CT, RIO.US, WIL.SP, MNDL.LN, and AAL.LN.

Le concept du portefeuille à variance minimale a été introduit par Harry Markowitz dans son modèle de portefeuille moderne. Ce modèle est fondé sur l'hypothèse que les investisseurs cherchent à minimiser le risque tout en maximisant le rendement. Pour atteindre cet objectif, le modèle utilise une méthode mathématique pour déterminer la répartition optimale des investissements dans différents actifs afin de minimiser, la variance du portefeuille, c'est-à-dire la mesure de la volatilité des rendements d'un portefeuille tout en maximisant le rendement attendu. Le point rouge sur le graphique nous donne ce portefeuille à variance minimale.



Target Return and Risks (sorties R):

Mean	Cov	CVar	VaR
0.6785	4.4268	8.4089	5.8734

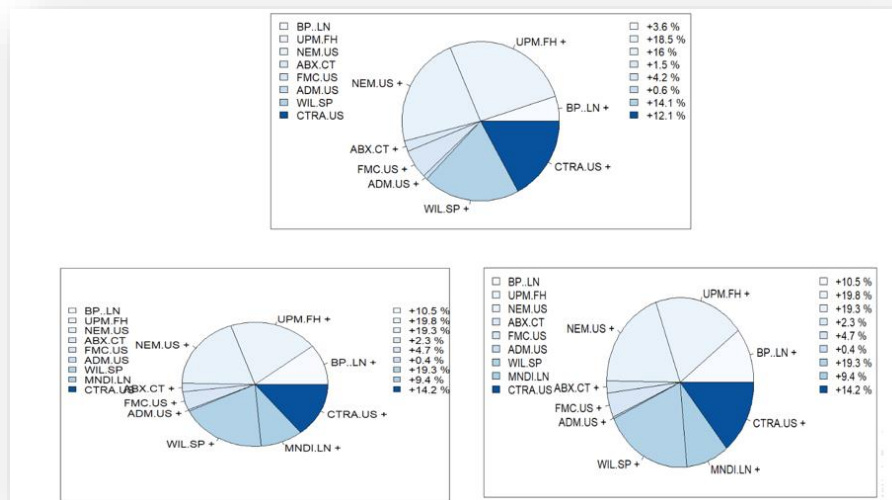
Title: MV Minimum Variance Portfolio
Estimator: covEstimator
Solver: solveRquadprog
Optimize: minRisk
Constraints: LongOnly

Fig. 20. Portefeuille à variance

Portfolio Weights:							
BP..LN	STERV.FH	IP.US	UPM.FH	NEM.US	XOM.US	VLO.US	NUE.US
0.1054	0.0000	0.0000	0.1985	0.1932	0.0000	0.0000	0.0000
ABX.CT	FMC.US	FCX.US	TTE.FP	COP.US	ADM.US	PKX.US	BHP.US
0.0227	0.0469	0.0000	0.0000	0.0000	0.0040	0.0000	0.0000
TECK.B.CT	RIO.US	WIL.SP	MNDI.LN	AAL.LN	CVE.CT	ALA.CT	WLK.US
0.0000	0.0000	0.1934	0.0943	0.0000	0.0000	0.0000	0.0000
GLEN.LN	MOS.US	MPC.US	PSX.US	WY.US	ET.US	VNOM.UW	SUN.US
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WRK.US	PBA.US	AA.US	MTS.SQ	NTR.CT	NTR.US	OXY.US	OKE.US
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CVX.US	PXD.US	TRGP.US	SLB.US	BKR.US	DVN.US	HES.US	MRO.US
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WMB.US	CTRA.US	APA.US	EOG.US	KMI.US	EQT.US	HAL.US	FANG.US
0.0000	0.1416	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Covariance Risk Budgets:							
BP..LN	STERV.FH	IP.US	UPM.FH	NEM.US	XOM.US	VLO.US	NUE.US
0.1054	0.0000	0.0000	0.1985	0.1932	0.0000	0.0000	0.0000
ABX.CT	FMC.US	FCX.US	TTE.FP	COP.US	ADM.US	PKX.US	BHP.US
0.0227	0.0469	0.0000	0.0000	0.0000	0.0040	0.0000	0.0000
TECK.B.CT	RIO.US	WIL.SP	MNDI.LN	AAL.LN	CVE.CT	ALA.CT	WLK.US
0.0000	0.0000	0.1934	0.0943	0.0000	0.0000	0.0000	0.0000
GLEN.LN	MOS.US	MPC.US	PSX.US	WY.US	ET.US	VNOM.UW	SUN.US
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WRK.US	PBA.US	AA.US	MTS.SQ	NTR.CT	NTR.US	OXY.US	OKE.US
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CVX.US	PXD.US	TRGP.US	SLB.US	BKR.US	DVN.US	HES.US	MRO.US
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WMB.US	CTRA.US	APA.US	EOG.US	KMI.US	EQT.US	HAL.US	FANG.US
0.0000	0.1416	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Tous les titres à qui un coefficient nul est attribué ne font pas partie de notre portefeuille à variance minimale. En effet, l'algorithme considère que ces titres ont une forte volatilité et donc il leur attribut un poids nul.

Ci-dessous nous représentons la construction de trois portefeuilles possibles avec les titres qui les composent. Le graphique nous renseigne également sur les poids attribués de chaque titre dans les portefeuilles.



4. Ratio de Sharpe, la rentabilité et volatilité du portefeuille équipondérant.

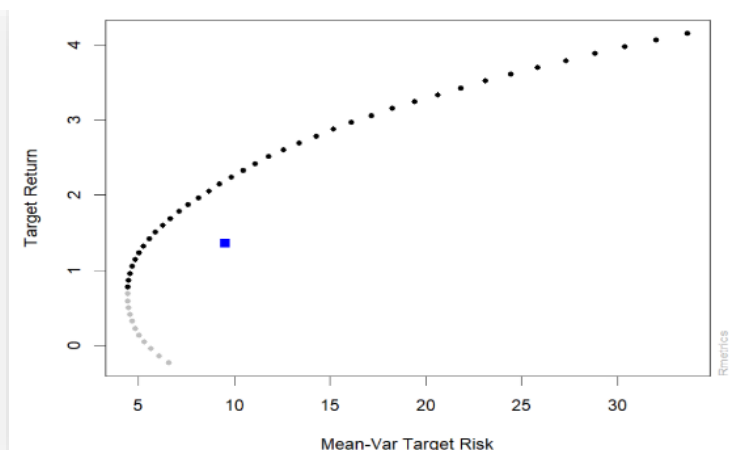
Le portefeuille équipondéré est un type de portefeuille dans lequel les investissements sont répartis de manière égale sur différents actifs. Cela signifie que chaque actif reçoit une allocation de capital similaire dans le portefeuille, plutôt que d'être pondéré en fonction de sa performance passée ou de son potentiel futur.

L'objectif d'un portefeuille équipondéré est souvent de maximiser la diversification en répartissant les investissements sur un grand nombre d'actifs différents. Cela peut aider à réduire le risque global du portefeuille en limitant l'exposition à un seul actif ou secteur d'industrie. Sur le graphique ci-dessous, le point bleu

```

Rendement cumulé? Volatilité? Ratio de Sharpe
portfolio.returns -8.5439e+58 22.11172 0.1158459

```



correspond à notre portefeuille équipondéré en termes de volatilité on a 22% et un ratio de Sharpe de 0,1158.

PARTIE 4 : PORTFOLIO CONSTRUCTION AND MONITORING

Pourrions-nous battre le S&P 500 en ne sélectionnant que les meilleures actions disponibles ?

Il y a 3 principaux problèmes à résoudre pour cette partie :

- Comment construire des entités à partir de données de séries chronologiques ?
- Comment choisir une méthode d'apprentissage automatique appropriée pour la sélection de titres ?
- Comment évaluer la rentabilité future des titres que contiennent notre portefeuille ?

Pour répondre à ces questions nous commençons par sélectionner certains titres en fonctions de leurs caractéristiques. Ensuite nous avons fait une estimation future des rentabilités des actifs sélectionnés pour évaluer la pertinence de leur intégration dans notre portefeuille.

1. Arbre CART

Pour commencer nous avons d'abord choisi les variables pertinentes pour la construction de notre stratégie. Les variables utilisées sont :

- Le score ESG
- La valeur moyenne des entreprises
- Le rendement moyenne sur la période considérée
- La volatilité des actifs

Pour créer les étiquettes de sortie, nous avons calculé le ratio de Sharpe, puis nous l'avons comparé avec le ratio de Sharpe de l'indice de référence. Si la valeur de ratio de Sharpe de l'action est supérieure à celle de S&P 500 ($=0.096$) alors la variable cible (target) vaut 1, sinon 0.

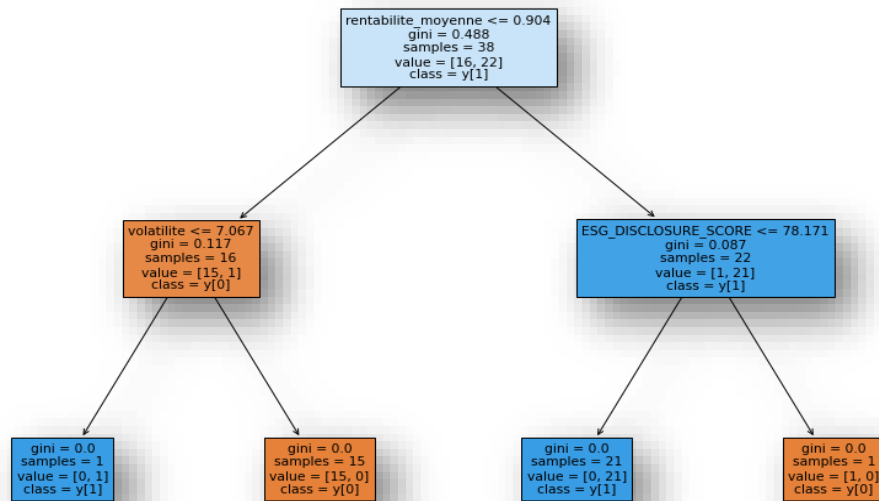
Nous avons opté pour un arbre de décision de Monte Carlo (CART ou Decision Tree) parce qu'il est considéré comme un outil essentiel pour la prise de décision financière. Dans notre cas, il permet de modéliser et de prévoir les différentes possibilités de résultats financiers dans un environnement incertain en utilisant une approche de simulation probabiliste. Pour la construction de notre analyse, la CART nous permet d'évaluer les options d'investissement et d'évaluer les risques associés à notre portefeuille. Aussi, il est capable de prendre en compte les incertitudes et les variables complexes.

- *Le résultat de l'arbre CART*

La variable la plus importante est la rentabilité qui représente la racine de l'arbre. Les deux nœuds-fils à partir desquels nous obtenons les 4 classes pures sont la volatilité et le score ESG, qui segmentent la racine. Pour la décision de sélection de titres on se base uniquement sur la racine. Alors, dans notre portefeuille, nous avons pris en compte que les titres qui ont une rentabilité supérieure à 0.904.

Finalement, les 21 titres sélectionnés sont : *UPM FH* , *XOM US* , *VLO US* , *FCX US* , *ADM US*, *TECK/B CT*, *RIO US*, *AAL LN*, *GLEN LN*, *MPC US*, *VNOM UW*, *SUN US*, *AA US*, *NTR CT*, *NTR US*, *CTVA US*, *OKE US*, *CVX US*, *PXD US*, *HES US*, *APA US*, *EQT US*.

Fig. 21. Arbre CART



Après le résultat obtenu, nous avons vu que les actifs que constituent le portefeuille nous permettent de battre l'indice de référence en termes de volatilité, de rendement et de ratio de Sharpe.

	Cumulative Return	Volatility	Sharpe Ratio
Portefeuille	108.016091	10.030899	0.175883
Indice de reference	62.773752	10.482659	0.096371

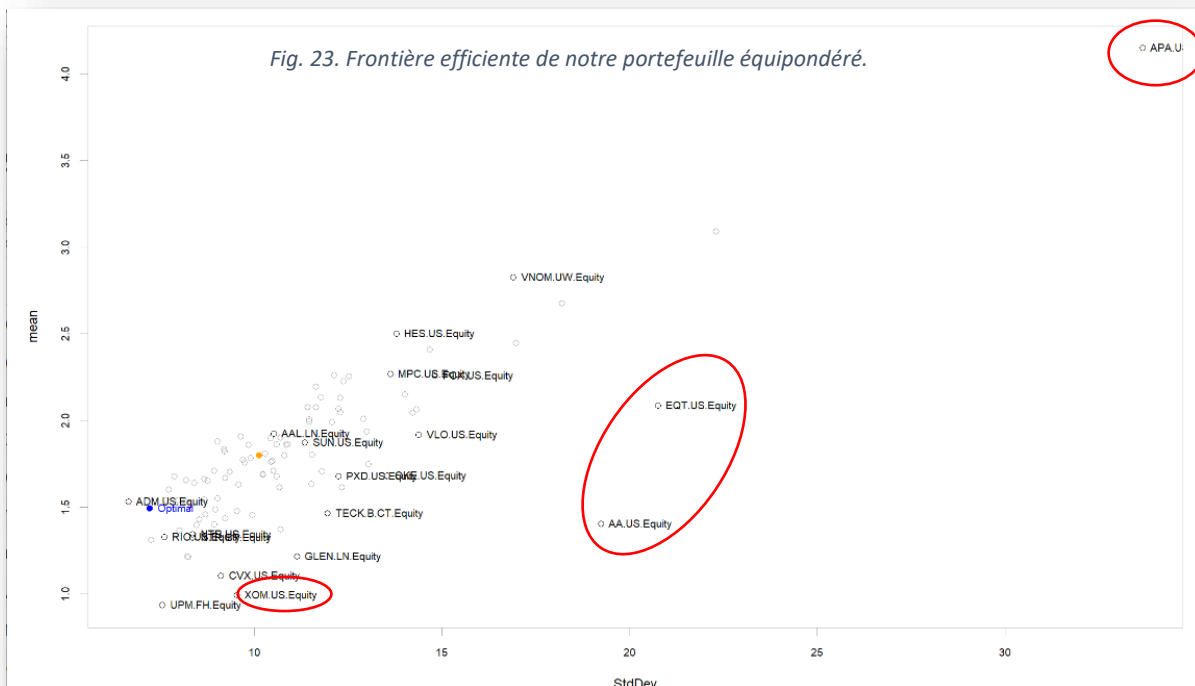
Parmi les actions sélectionnées, nous distinguons des titres qui sont plus volatiles dans notre portefeuille, cependant pour limiter le risque, on souhaite retirer ces actions. Pour se faire nous affichons la frontière d'efficacité avec les actifs pour arbitrer.

Fig. 22. Statistique descriptive des titres sélectionnés

	mean	std	min	25%	50%	75%	max
UPM FH	0,933991	7,545553	-18,17581	-3,39147	1,360282	6,047841	16,86093
XOM US	0,992708	9,52452	-26,18585	-4,474271	1,752606	5,669578	24,13793
VLO US	1,916323	14,380837	-31,53208	-7,380648	2,168741	9,631668	41,78483
FCX US	2,257247	14,80671	-32,22892	-6,552729	1,161181	11,886195	36,2069
ADM US	1,530795	6,640502	-15,2208	-2,060386	1,682716	6,519455	15,05417
TECK/B CT	1,465628	11,947684	-26,15643	-5,37931	1,485447	11,322245	21,1171
RIO US	1,327506	7,598793	-16,939	-3,45835	1,581724	6,643353	16,73522
AAL LN	1,924395	10,515054	-27,09281	-5,163959	1,627077	10,100153	26,09288
GLEN LN	1,2162	11,131331	-37,76344	-5,351308	0,235152	7,598809	40,20381
MPC US	2,268121	13,617243	-50,18979	-5,670125	3,250091	9,448904	35,8171
VNOM UW	2,826379	16,897002	-62,81548	-4,525402	1,729549	11,48441	61,48788
SUN US	1,872835	11,34422	-42,13836	-2,647764	1,117662	7,595023	58,69565
AA US	1,404068	19,231163	-55,5876	-8,88448	-0,297857	12,06686	54,02477
NTR CT	1,323647	8,353728	-18,13541	-4,596429	0,489496	7,200656	22,88344
NTR US	1,343627	8,352845	-17,52752	-4,315707	0,713526	7,22252	23,19484
OKE US	1,681389	13,531913	-67,31115	-4,452633	2,472129	7,749736	41,86948
CVX US	1,104674	9,107432	-22,36983	-3,414786	0,817818	4,929788	27,29795
PXD US	1,678758	12,234556	-42,39267	-6,955068	2,302094	8,224861	27,3129
HES US	2,500313	13,778557	-40,30188	-5,895348	4,265447	8,910882	46,06607
APA US	4,149615	33,666084	-83,22633	-9,206559	3,377876	11,316628	213,8315
EQT US	2,087414	20,747757	44,49541	-7,596112	0,306793	9,302984	106,3649

A travers la frontière efficace, nous voyons que nos actions se comporte bien sur un plan volatilité-risque. Par exemple, le titre APA US a un grand rendement mais également une grande volatilité. Pour minimiser le risque

pris, nous excluons cette action. Aussi pour deux titres d'une même volatilité, nous ne garderons que celui qui a le plus grand rendement, étant donné qu'on est averse au risque. Nous supprimerons également les titres corrélés entre eux pour limiter le risque global.



	Cumulative Return	Volatility	Sharpe Ratio
Portefeuille	108.016091	10.030899	0.175883
Indice de reference	62.773752	10.482659	0.096371
Portefeuille (après la suppression des titres plus risqués)	100.548448	8.862007	0.185038

Finalement, nous aboutissons à surperformer l'indice SP500.

2. ARIMA

Les entreprises d'investissement, les fonds spéculatifs et même les particuliers utilisent des modèles financiers pour mieux comprendre le comportement du marché et réaliser des investissements et des transactions rentables. C'est pour cette raison que nous continuerons une dernière analyse de nos actifs financiers avec un modèle ARIMA.

Après l'étape de sélection des actifs financiers rentables pour notre portefeuille, nous avons fait une analyse de séries chronologiques et des prévisions pour le marché boursier. La base de données utilisée dans cette partie de modélisation ARIMA est la base Returns. Cette étape comprend également la suppression des colonnes dont nous n'avons pas eu besoin, la vérification des valeurs manquantes, et le regroupement des returns par date. Notre objectif de base dans cette analyse de séries chronologiques est d'utiliser le modèle ARIMA pour prédire la valeur future et la comparer avec le modèle SARIMAX. L'une des parties importantes de l'analyse de séries chronologiques à l'aide de python est le package *statsmodel*. Grâce à ça nous avons choisi les titres suivant étant comme le plus pertinents pour le portefeuille final, en ne prenant pas trop de risque dans le futur. La liste des titres qui compose le portefeuille final est : **VLO US** , **FCX US** , **ADM US** , **TECK/B CT** , **RIO US** , **AAL LN** , **GLEN LN** , **VNOM UW** , **SUN US** , **AA US** , **PXD US** , **HES US**.

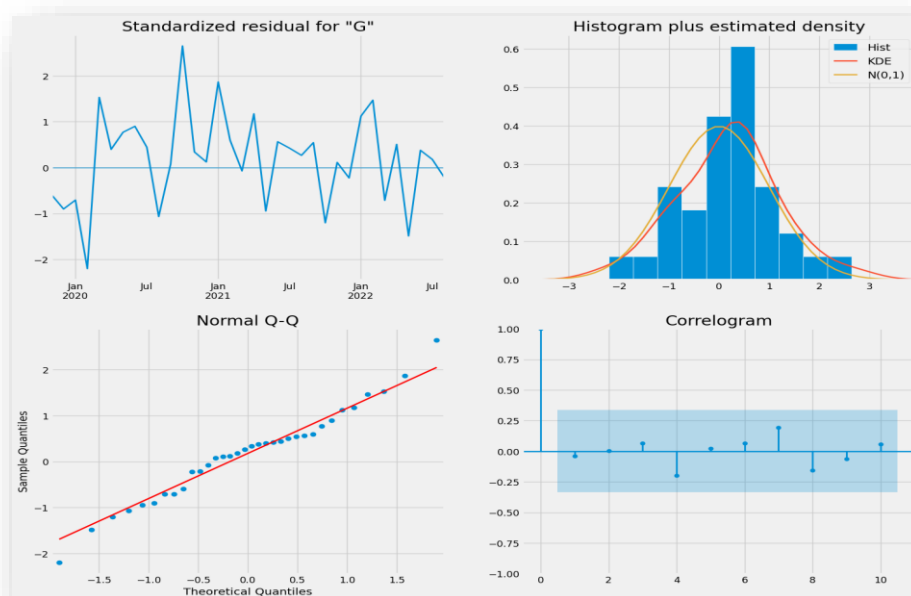
De plus, certains modèles de séries temporelles apparaissent distinctifs lorsque nous traçons un modèle de saisonnalité pour les titres financiers. Cela nous amène également à visualiser nos données à l'aide d'une méthode appelée décomposition de séries chronologiques. Celle-ci nous permet de décomposer nos séries chronologiques en trois composantes distinctes : tendance, saisonnalité et bruit.

Ensuite, les graphiques (Seasonal decomposition of : actif_x) montrent clairement que le returns des actifs financiers sont des séries chronologiques instable, ainsi que leurs saisonnalités évidentes. Pour précision, la tendance nous fournit des informations sur le mouvement de hausse et de baisse des données au cours du temps sur une longue période ; tandis que la saisonnalité stipule des indications sur la variance saisonnière (par exemple une baisse de la performance des actifs pendant la période de Covid). Ensuite, les bruits représentent des pics et des creux à intervalles aléatoires.

- Etape de sélection des paramètres pour le modèle de série chronologique ARIMA.

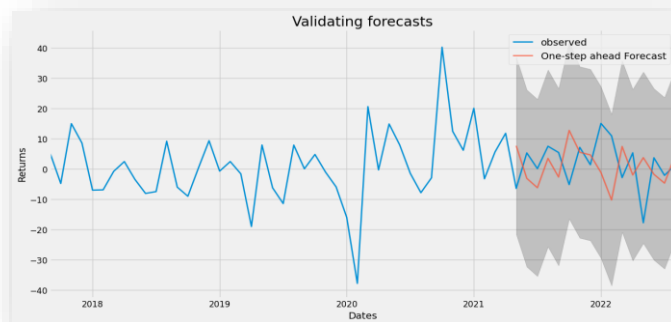
Les modèles ARIMA sont désignés par la notation ARIMA (p, d, q). Ces trois paramètres tiennent compte de la saisonnalité, de la tendance et du bruit dans les données. Nous avons généré les combinaisons possibles de ces paramètres et nous avons sélectionné d'après le critère statistique le AIC minimum. Un exemple des graphiques des résidus standardisé, Histogramme, graphiques de quantiles, et du corrélogramme est fourni dans l'application. (Ici prenons l'exemple de GLEN LN Equity - Basic Materials).

- Etape de validation des prévisions :

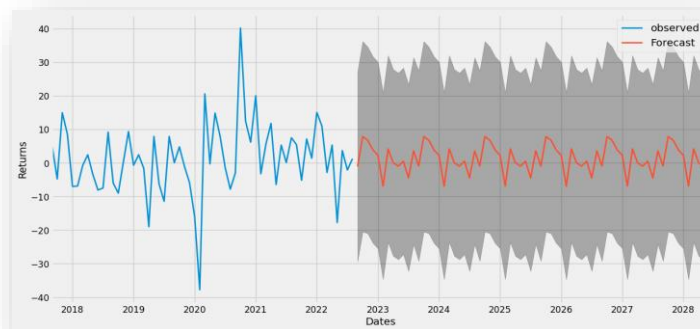


Pour nous aider à comprendre l'exactitude de nos prévisions, nous comparons le return des actions prévues au return réel de la série temporelle, et nous fixons des prévisions pour qu'elles commencent le 2021-05-31 jusqu'à la fin des données.

Le tracé linéaire montre les valeurs observées par rapport aux prévisions de prévisions glissantes. Dans l'ensemble, nos prévisions s'alignent très bien sur les vraies valeurs, montrant une tendance à la (hausse /baisse/ stable, indépendamment des titres retenus et du secteur) qui commence dès le début de l'année.



L'erreur quadratique moyenne (RMSE) nous indique que notre modèle a été en mesure de prévoir la performance d'actif dans l'ensemble de test (The Mean Squared Error of our forecasts is 109.04 The Root Mean Squared Error of our forecasts is 10.44).



En conclusion, nous pouvons dit que notre modèle a clairement capturé la saisonnalité des returns proches. Alors que nous prévoyons plus loin dans l'avenir, il est naturel que nous devenions moins confiants dans nos valeurs. Cela se reflète dans les intervalles de confiance générés par notre modèle, qui augmentent à mesure que nous nous éloignons dans le futur. Également, les crises et les chocs enregistrés au cours de temps (Covid, Guerre en Ukraine) affecte nos prévisions et l'IC.

Grâce à notre stratégie nous avons abouti à avoir un portefeuille qui contient 12 actifs financiers, qui pris ensemble nous permet d'avoir à la fois un portefeuille diversifié et avoir une rentabilité supérieure à toutes les autres obtenus avant (> 109). Finalement nous avons atteint l'objectif de battre l'indice de référence.

	Cumulative Return	Volatility	Sharpe Ratio
Portefeuille	108.016091	10.030899	0.175883
Indice de reference	62.773752	10.482659	0.096371
Portefeuille (après la suppression des titres plus risqués)	100.548448	8.862007	0.185038
Portefeuille finale (après modèle ARMA)	109.602229	10.045966	0.178251