2ème Année ENSAE

Mémoire du projet de statistiques appliquées, sujet 88

MENTION CONFIDENTIEL

Prédiction temps réel des fluctuations des actions avec un modèle de multirégression et recherche d'améliorations de ce modèle par méthodes de clustering

Samy Mekkaoui-Simon Depaule-Eros Giovanetto-Alex Jin

Sous la Supervision de Monsieur Vivien Bouche

Année Universitaire 2021-2022

Sommaire

Remerciement

Pour la rédaction de ce mémoire, nous voulions remercier notre encadrant Vivien Bouche dont la bienveillance et l'expertise nous ont permis de mener à bien notre projet.

1 Introduction

Le métier de quant ou d'analyste quantitatif consiste à rechercher et appliquer de nouveaux modèles pour le pricing de produits financiers. En effet, son but va être de construire une stratégie de portefeuille qui lui permettra de générer des profits.

Par ailleurs, il existe de nombreuses stratégies de portefeuille classiques telles que le modèle développé par Markowitz dans les années 1950 expliquant la manière dont les investisseurs rationnels utilisent la diversification des actifs afin de minimiser le risque de leur portefeuille. Ce dernier a notamment servi de base à l'un des modèles phares de la finance : le Modèle CAPM (*Capital Asset Pricing Model*) qui permet d'estimer le taux de rentabilité d'un actif en fonction de son risque systématique. Cette notion qui est au centre des modèles de finance reste dure à évaluer.

Nous pouvons également évoquer des modèles de pricing d'options plus récents tels que le fameux modèle de Black-Scholes qui postule une dynamique gaussienne pour les rendements de l'action. De manière plus précise, en notant S_t la valeur de l'actif, nous obtenons

$$S_t = S_0 \, e^{(\mu - \frac{\sigma^2}{2})t + \sigma \times \sqrt{dt}\Phi} \tag{1}$$

avec :

1. μ : la tendance de l'action
2. $\phi \sim \mathcal{N}(0,1)$
3. σ : la volatilité de l'action

Dans ce projet, nous allons adopter une approche purement statististique. Nous nous intéresserons à un modèle prédictif contrairement au modèle de Black-Scholes où la prédiction est impossible. Pour ce faire, nous aurons l'occasion de travailler sur une base de données contenant les rendements d'actions américaines et nous testerons un modèle de régression linéaire sur ceux-ci.

2 Notre base de données

2.1 Construction de la base de données

Pour notre projet, nous avons travaillé sur les actions du S&P 500 pour la période 2006-2021. Avant de construire notre base de données nous devions nous assurer que les actions que nous allions selectionner allaient rester cotées tout au long de la période d'étude. Nous avons donc supprimer celles qui auraient pu rentrer ou sortir de l'indice au cours de cette période. Néanmoins, cela introduit un biais de survie car les actions que l'on exclut sont celles qui ne performent pas suffisamment. Pour ce faire, nous avons fait du web-scrapping à partir de la page Wikipedia qui référence toutes les actions faisant partie du S&P 500.

Une fois cette liste faite, il fallait télécharger le cours de ces actions. Nous avons décidé d'utiliser la variable *"Adjusted Close"* que nous considérerons comme étant le prix de référence pour nos actions. Nous avons fait ce choix car cet indicateur était plus approprié que le prix à la fermeture puisqu'il nous permet de nous affranchir des problèmes liés aux versements des dividendes et au *splitting* des actions.

Nous avons importé nos données via le site Yahoo Finance étant donné que Python est bien adapté pour ce type de manipulations. En effet, via le module "pandas−datareader" nous pouvons directement importer les prix de clôture ajustés de chaque action et les mettre sous forme d'un *DataFrame* qui servira de base de données initiale.

Nous obtenons donc une base de données de 307 colonnes, correspondant au nombre d'actions et de 3991 lignes, correspondant au nombre de jours ouvrés boursiers pour notre période d'étude, comme le montre la figure ci-dessous.

Date	GD	NSC	AEP
04/01/2006	0.000087	0.002299	-0.001607
05/01/2006	-0.012335	-0.016203	-0.005646
06/01/2006	0.022371	-0.016234	0.015517
⋮	⋮	⋮	⋮
11/09/2021	0.001184	-0.000141	-0.000963

Tableau 1 : Une partie de notre base de données pour 3 actions et 4 dates

Par la suite, nous noterons pour une action i, à la date t : $\mathrm{P}_t^{(i)}$ son prix ajusté à la fermeture. Une fois notre base de données créée il a fallu définir de nouvelles variables qui nous seront utiles par la suite.

Tout d'abord, la variable **"rendement"** notée : X_t^i qui se définit comme :

$$\mathrm{X}_t^i = \ln\left(\frac{\mathrm{P}_t^{(i)}}{\mathrm{P}_{t-1}^{(i)}}\right) \tag{2}$$

Puis une variable nommée **"rendement nette des performances du marché"** se définissant comme :

$$Y_t^{(i)} = X_t^{(i)} - \frac{1}{N}\sum_{j=1}^{N} X_t^{(j)} \tag{3}$$

visant à isoler les variations de l'action de la tendance générale.

Une fois ces variables crées il a fallu nettoyer notre base de données, c'est-à- dire vérifier à la fois qu'il n'y ait pas trop de données manquantes ainsi que de valeurs de rendements extrêmes.

Concernant le premier point, notre base de données présentait des données manquantes pour seulement trois jours, nous avons donc raisonablement décidé de supprimer ces trois jours manquants de notre *DataFrame*.

Concernant le second point, nous avons décidé de borner nos rendements à plus ou moins 15% afin d'éviter d'éventuels sauts de rendements entre deux dates consécutives, ces derniers pouvant nuire à notre modèle. Nous expliquerons par la suite les problèmes que cela peut engendrer.

2.2 Statistiques descriptives sur notre base de données

Voici la performance de l'indice du S&P 500 pour notre période d'étude :

FIGURE 1 – Evolution de l'indice du S&P 500

Nous pouvons observer deux chutes notables du cours du S&P 500 : la première étant liée à la crise financière de 2008, la seconde correspondant à celle de la crise sanitaire du covid. Ces crises entraînent des valeurs extrêmes de rendements, nous illustrerons ce phénomène via la manipulation consistant à borner ces dernières.

Au cours de ce projet, nous souhaitons donner un poids sensiblement égal à chaque donnée historique. Comme nous l'expliquerons par la suite, étant donné que l'échantillon utilisé pour chaque regression regroupe les données sur cinq ans, si certaines d'entre elles appartiennent à l'une des deux crises, les coefficients estimés donneraient alors trop d'importance à ces dernières. Cela illustre donc l'importance de borner nos rendements à plus ou moins 15 %.

A titre informatif, nous avons résumé dans le tableau ci-dessous le nombre de fois que nous avons utilisé cette manipulation chaque année.

Année	Nombre de *clip* réalisés sur une année	% du nombre de *clip* par année
2006	19	0.024756
2007	22	0.028550
2008	723	0.930849
2009	470	0.607518
2010	16	0.020681
2011	27	0.034900
2012	20	0.026059
2013	15	0.019389
2014	9	0.011633
2015	19	0.024559
2016	23	0.029730
2017	19	0.024657
2018	22	0.028550
2019	25	0.032315
2020	87	0.627004
2021	11	0.016588

Tableau 2 : Statistiques descriptives sur le nombre de *clip* réalisés

Nous observons que le nombre de fois où l'on borne les rendements des actions est maximal pour les années 2008 et 2020, cela confirme donc bien le phénomène expliqué précédemment.

3 Modèle théorique et premiers résultats

3.1 Expression du modèle théorique

Dans ce mémoire, nous utiliserons le modèle suivant où N représente le nombre d'actions et T le nombre de jours sur lequel nous travaillerons :

$$\forall i \in [\![1\ ;\ \mathrm{N}]\!], \quad \sum_{l=1}^{5} \mathrm{Y}_{t+l}^{(i)} = \beta_{1,t}^{(i)} \mathrm{X}_t^{(i)} + \frac{\beta_{2,t}^{(i)}}{\sqrt{5}} \sum_{l=1}^{5} \mathrm{X}_{t-l}^{(i)} + \frac{\beta_{3,t}^{(i)}}{\sqrt{21}} \sum_{l=1}^{21} \mathrm{X}_{t-l}^{(i)} + \frac{\beta_{4,t}^{(i)}}{\sqrt{256}} \sum_{l=1}^{256} \mathrm{X}_{t-l}^{(i)} + \epsilon_t^{(i)} \tag{4}$$

où :

1. $\mathrm{Y}_t^{(i)} := \mathrm{X}_t^{(i)} - \frac{1}{\mathrm{N}} \sum_{j=1}^{\mathrm{N}} \mathrm{X}_t^{(j)}$ représente le rendement net de celui du marché.
2. $\mathrm{R}_t^{(i)} := \sum_{l=1}^{5} \mathrm{Y}_{t+l}^{(i)}$ représente la performance de l'action sur les cinq prochains jours nette du marché
3. $\mathrm{X}_{1,t}^{(i)} := \mathrm{X}_t^{(i)} = ln(\frac{\mathrm{P}_t^{(i)}}{\mathrm{P}_{t-1}^{(i)}})$ représente le rendement de l'action i à la date t.
4. $\mathrm{X}_{2,t}^{(i)} := \frac{1}{\sqrt{5}} \sum_{l=1}^{5} \mathrm{X}_{t-l}^{(i)}$ représente la somme normalisée des rendements de l'action i sur les cinq derniers jours par rapport à la date t. Cela est représentatif de la performance de l'action sur la dernière semaine boursière.
5. $\mathrm{X}_{3,t}^{(i)} := \frac{1}{\sqrt{21}} \sum_{l=1}^{21} \mathrm{X}_{t-l}^{(i)}$ représente la somme normalisée des rendements de l'action i sur les 21 derniers jours par rapport à la date t. Cela est représentatif de la performance de l'action sur le dernier mois boursier.
6. $\mathrm{X}_{4,t}^{(i)} := \frac{1}{\sqrt{256}} \sum_{l=1}^{256} \mathrm{X}_{t-l}^{(i)}$ représente la somme normalisée des rendements de l'action i sur les 256 derniers jours par rapport à la date t. Cela est représentatif de la performance de l'action sur la dernière année boursière.
7. $\epsilon_t^{(i)}$ représente le résidu de la régression linéaire associée

Nous poserons :

— $\mathrm{X}_t = (\mathrm{X}_{1,t}^{(i)}, \mathrm{X}_{2,t}^{(i)}, \mathrm{X}_{3,t}^{(i)}, \mathrm{X}_{4,t}^{(i)})^\top$ qui représente le vecteur des variables explicatives

— $\beta = (\beta_{1,t}^{(i)}, \beta_{2,t}^{(i)}, \beta_{3,t}^{(i)}, \beta_{4,t}^{(i)})^\top$ qui représente le vecteur des coefficients

La théorie de la régression linéaire nous donne alors :

$$\beta = \mathbb{E}(\mathrm{X}\mathrm{X}^\top)^{-1} \mathbb{E}(\mathrm{X}\mathrm{Y}) \tag{5}$$

On a par ailleurs un estimateur consistant de β donné par $\widehat{\beta} = (\frac{1}{n} \sum_{l=1}^{n} \mathrm{X}_{t-l} \mathrm{X}_{t-l}^\top)^{-1} (\frac{1}{n} \sum_{l=1}^{n} \mathrm{X}_{t-l} \mathrm{Y}_{t-l})$ où n représente le nombre de données sur les 5 dernières années

Ainsi, nous supposerons un modèle linéaire où les rendements nets d'une action peuvent être prédits grâce aux rendements normalisés de cette dite action sur le jour, la semaine, le mois et l'année précédente.

Notons dans la régression linéaire la présence des facteurs de normalisation $\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{21}}$ et $\frac{1}{\sqrt{256}}$. L'intérêt de cette normalisation est de faciliter la comparaison à la fois qualitative et quantitative des coefficients β de notre régression. En effet, il est nécessaire d'avoir un écart type similaire entre nos variables explicatives pour pouvoir comparer nos coefficients. Calculons la variance d'une des variables explicatives sans le facteur de normalisation :

$$\mathbb{V}(\sum_{l=1}^{5} X_{t-l}^{(i)}) \approx \sum_{l=1}^{5} \mathbb{V}(X_{t-l})^{(i)} \approx 5\mathbb{V}(X_{t-l}^{(i)}) \tag{6}$$

sous hypothèse que

— les covariances entre les rendements sont négligeables

— les variances des rendements journaliers sont supposées égales

Cela illustre la nécessite de normaliser par le facteur $\frac{1}{\sqrt{5}}$

D'après notre modèle, pour chaque action, nous obtiendrons autant de régressions linéaires que de jours de données. Nous avons choisi d'effectuer nos régressions en utilisant comme échantillon les données des cinq années précédant la date t de notre régression. Autrement dit, nous ne pourrons pas obtenir les régressions associées aux dates antérieures à 2011. Les premières regressions que nous obtiendrons seront celles de l'année 2012 car la création de la variable explicative $X_{4,t}^{(i)}$ nécessite un an de données pour être créée.

A partir de ce modèle, nous obtiendrons $\forall i \in [\![1\ ;\ N]\!], \forall t \in [\![1\ ;\ T]\!], \hat{R}_t^{(i)} = \sum_{l=1}^{5} \hat{Y}_{t+l}^{(i)}$: la valeur prédite de la somme des rendements de l'action i sur les cinq prochains jours. Notre position à la date t pour une action i sera caractérisée par la formule suivante : $P_t^{(i)} = \frac{1}{5} \sum_{l=1}^{5} \hat{R}_{t-l}^{(i)}$. Elle correspond à une moyenne mobile sur cinq jours de $\hat{R}_t^{(i)}$, et cela nous permet d'obtenir une estimation de la performance d'une action i à la date t. Dès lors, nous pouvons obtenir nos gains, notés par la suite $G_t^{(i)}$, définis par la formule suivante : $G_t^{(i)} = P_t^{(i)} X_{t+1}^{(i)}$. Les gains journaliers de notre portefeuille sont alors caractérisés comme suit : $G_t = \sum_{i=1}^{N} G_t^{(i)}$. Notons "G$ains$" le vecteur des gains journaliers. Notre but sera alors de maximiser la quantité suivante : $SR = \frac{\hat{\mu}_{Gains}}{\hat{\sigma}_{Gains}}$ où $\hat{\mu}_{Gains}$ représente la moyenne empirique du vecteur G$ains$ et $\hat{\sigma}_{Gains}$ son écart-type empirique. La quantité SR introduite représente le ratio de Sharpe, notion que nous définirons à la partie suivante.

3.2 Notion du ratio de Sharpe

3.2.1 Définition du ratio de Sharpe

Comme expliqué précédemment, notre objectif est de construire une stratégie qui nous permet de maximiser nos gains tout en minimisant notre risque. Pour ce faire, nous baserons notre étude sur le critère du ratio de sharpe qui se définit ainsi :

$$SR = \frac{\mu}{\sigma} \tag{7}$$

où

— μ représente l'espérance de rendement du portefeuille d'actions que l'on choisira

— σ sa variance en supposant le *"Free Rate Risk"* nul ici.

Afin d'annualiser le ratio de Sharpe, on multipliera ce dernier par $\sqrt{256} = 16$ où 256 représente le nombre de jours ouvrés boursiers dans l'année.

Par ailleurs, nous allons considérer cet indicateur sous un point de vue microéconomique. Nous pouvons distinguer trois cas :

1^{er} cas : $SR < 0$ signifie que le portefeuille sous-performe un placement sans risque et donc il n'est pas logique d'investir dans un tel portefeuille.

$2^{ème}$ cas : $0 < SR < 1$ signifie que l'excédent de rendement par rapport au taux sans risque est plus faible que le risque pris.

$3^{ème}$ cas : $SR > 1$ signifie que le portefeuille surperforme un placement sans risque et donc il génère une plus forte rentabilité.

Ainsi plus le ratio de Sharpe est élevé plus le portefeuille est performant. Savoir le calculer est utile car si deux portefeuilles ont le même rendement, ainsi il est conseillé, les agents étant averses au risque, d'opter pour celui avec un ratio de Sharpe supérieur : cela signifie que la volatilité du rendement ce portefeuille est plus faible.

3.2.2 Lien entre ratio de Sharpe et test statistique

Plaçons nous dans le cas où nous disposons de $X_1, X_2, ..., X_n$ n variables i.i.d suivant une distribution gaussienne de paramètres μ et σ^2. Considérons alors $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$ qui sont respectivement les estimateurs sans biais de la moyenne et de la variance. Considérons maintenant la statistique de test $t_0 = \sqrt{n}\frac{\hat{\mu}-\mu_0}{\hat{\sigma}}$. Cette dernière suit alors une distribution dite de Student à n-1 degrés de libertés et de paramètre de non centralité de $\delta = \sqrt{n}\frac{\mu-\mu_0}{\sigma}$.

Nous allons maintenant nous intéresser au ratio de Sharpe. L'estimateur standard du ratio de Sharpe est : $\widehat{SR} = \frac{\hat{R}}{\hat{\sigma}}$ en supposant le *"Free Rate Risk"* nul ici. En considérant le ratio de Sharpe annualisé, nous retrouvons bien l'estimateur que nous calculons : $\widehat{SR}_a = 16 \times \frac{\hat{R}}{\hat{\sigma}}$. Par ailleurs, nous voyons que sous l'hypothèse où les gains suivent une loi normale de paramètres μ et σ^2 et sont i.i.d, nous pouvons retrouver

la *t-test* en effectuant l'opération $\sqrt{N} \times \widehat{SR} = \sqrt{T} \times \widehat{SR}_a$ avec N le nombre d'observations et T le nombre d'années.

Mais alors, il est possible de faire des tests d'hypothèses associés au ratio de Sharpe notamment celui-ci : Sous l'hypothèse $H_0 : \mathbb{E}(R) = \mu_0$ contre l'hypothèse alternative $H_1 : \mathbb{E}(R) > \mu_0$ où R représente un jour de gain, nous pouvons regarder la t-test t_0 et rejeter l'hypothèse H_0 si t_0 est supérieur au quantile d'ordre 1-α d'une t-distribution à n-1 degrés de libertés.

Nous observons un lien entre le ratio de Sharpe et la *t-test* : plus la *t-test* est grande, moins on pourra rejeter l'hypothèse que les gains ont une moyenne strictement positive. Cela explique donc qu'un ratio de Sharpe élevé implique des gains importants via ce lien direct avec la t-test (sous certaines hypothèses restrictives).

3.3 Premiers résultats obtenus à l'aide de la stratégie initiale

Premièrement, nous appelerons la stratégie initiale : la stratégie Benchmark. La mise en place de cette stratégie nous a permis d'obtenir des résultats notamment sur la significativé des coefficients ainsi que sur les R^2 des régressions. Voici ci-dessous les résultats pour une action et pour trois dates :

Name	Beta 1	Beta 2	Beta 3	Beta 4	Significativité à 5% des Betas	R^2
GD	0.091067	0.216356	0.033549	-0.053314	True | True | False | False	0.030168
GD	0.092194	0.217905	0.032790	-0.051963	True | True | False | False	0.030401
GD	0.092253	0.217984	0.032853	-0.051905	True | True | False | False	0.030562

Tableau 3 : Résultat de nos régressions linéaires pour une action et trois dates

Il est intéressant de noter les très faibles valeurs des R^2 qui semblent suggérer que la relation linéaire existant entre les rendements futurs et passés est très faible. D'autre part, les faibles valeurs des R^2 rendent très compliquée l'analyse des coefficients β. En effet, étant donné le faible pouvoir explicatif du modèle, il n'est pas simple de pouvoir affirmer que les coefficients estimés ont une réelle significativité statistique.

Par ailleurs, la valeur du ratio de Sharpe associée au portefeuille issu de notre stratégie initiale est de 0.74. Cette valeur signifie selon les critères microéconomiques définis plus haut que les gains associés à notre portefeuille ne nous couvrent pas assez du risque que nous prenons. Nous allons chercher à améliorer notre modèle en se basant sur le Benchmark que nous avons défini plus haut.

La partie suivante sera donc consacrée à l'amélioration de ce modèle de base. Pour cela, nous allons notamment utiliser des méthodes de clustering.

4 Amélioration des résultats obtenus sur le Benchmark en utilisant des stratégies basées sur le clustering

Après avoir obtenus les résultats de base à l'aide du portefeuille d'investissement Benchmark, nous allons désormais mettre en place des stratégies visant à améliorer ce plan initial.

En effet, le but de notre projet est de maximiser le ratio de Sharpe de notre portefeuille d'investissement. Ainsi, nous nous attacherons à minimiser l'écart-type des rendements de notre portefeuille, tout en essayant de maximiser ces derniers : cela nous permettra d'obtenir le ratio de Sharpe le plus grand possible. Pour ce faire, nous allons utiliser le clustering : nous allons « filtrer » les actions sur lesquelles nous allons investir en fonction de différents critères que nous allons étudier dans cette partie.

Les indicateurs que nous avons jugés utiles afin de distinguer les actions sont :

1. **La qualité de prédiction de notre modèle via l'utilisation du R^2 :**
   En effet, pour une régression donnée à la date t pour une certaine action (i), cette grandeur représente la part de variance du rendement net de cette action sur les cinq jours suivants, expliquée par notre modèle. Soit, autrement dit, en termes mathématiques :

$$R^2 = \frac{\mathbb{V}(\widehat{R}_t)}{\mathbb{V}(R_t)} \text{ avec } \widehat{R}_t \text{ la prédiction de } R_t \tag{8}$$

   Ainsi nous aurions tendance à penser que plus le R^2 de notre régression est élevé, plus notre modèle prédit bien les variations du rendement de l'action (i). L'idée de cette stratégie est donc de sélectionner les prédictions issues des régressions ayant le R^2 le plus élevé, de ce fait nous ne conserverions ainsi que les actions les mieux fittées par le modèle.

2. **La volatilité de l'action sur les 100 jours précédant la date de notre régression :**
   Elle représente la dispersion des rendements de l'action par rapport leur moyenne sur les 100 jours précédents. C'est ainsi que cette grandeur permet de mesurer le risque associé à un investissement sur cette action. Étant donné que notre but est de maximiser le ratio de Sharpe, grandeur dépendant négativement de la volatilité, nous aurions ainsi tendance à sélectionner les actions les moins volatiles, quitte à perdre en rendement. Nous allons donc vérifier cette hypothèse.

3. **La sensibilité des calculs à une erreur : le conditionnement de la matrice $X^\top X$:**
   Ce paramètre caractéristique d'une matrice donne une idée de la manière dont l'erreur relative augmente lors d'une opération. Ainsi, plus une matrice est mal conditionnée (conditionnement est grand), plus les calculs deviennent sensibles à une éventuelle erreur. De ce fait, comme expliqué dans la partie 3.1, étant donné que le calcul des coefficients de la régression dépend directement de la matrice $X^\top X$ (équation n°5), un mauvais conditionnement de cette dernière serait responsable d'une instabilité sur nos coefficients, ce que nous essayons d'éviter.
   Cette erreur de spécification de la matrice pourrait trouver sa source dans un arrondi, un mauvais web scraping etc…, mais également et plus probablement elle pourrait être due au fait que les

valeurs calculées le sont toujours à une incertitude statistique près. L'idée de cette stratégie était ainsi de sélectionner les actions dont le conditionnement de la matrice $X^\top X$ serait le plus faible possible afin d'avoir le minimum d'erreurs possibles sur les calculs.

4. **La significativité à** 5% **des coefficients de notre régression :**
   En effet, pour la régression d'une action donnée, à une date donnée, si le coefficient associé à une variable explicative de notre modèle n'est pas significatif à 5%, nous ne pouvons pas rejeter l'hypothèse nulle stipulant que cette variable explicative n'influe pas sur la variable cible. Ainsi prendre en compte l'effet de la variable explicative non significative ne va faire que « bruiter » notre prédiction, ce dont nous voulons nous affranchir. De ce fait, l'idée de cette stratégie était de ne conserver que les prédictions provenant d'une régression dont tous les coefficients seraient significatifs à 5% afin de limiter au maximum l'erreur de prévision, si l'on suppose la véracité de notre modèle.

4.1 Méthode de protection de portefeuille

Afin de comprendre comment nous allons protéger notre portefeuille par rapport au Benchmark, il faut tout d'abord comprendre le concept suivant : le « bêta » (β) qui est une mesure du risque d'un actif. Le bêta est un coefficient utilisé afin de mesurer la volatilité d'un investissement sur un certain marché, dans notre cas le Benchmark, par rapport à celle de l'ensemble du marché. Autrement dit, cela revient à calculer dans quelle mesure le cours de notre portefeuille fluctue en comparaison aux mouvements du Benchmark. Comment est calculé le bêta ? De cette manière :

$$\beta = \frac{\text{Cov}(R_{Stratégie Clustering}, R_{Benchmark})}{\mathbb{V}(R_{Benchmark})} \tag{9}$$

avec

— $R_{Stratégie Clustering}$ qui représente le rendement de notre portefeuille « clusterisé »

— $R_{Benchmark}$ qui représente le rendement de notre stratégie Benchmark.

Il est donc possible d'obtenir cette valeur en faisant la régression linéaire des rendements de notre portefeuille sur les rendements du portefeuille Benchmark. Plus précisément, et en prenant les mêmes notations que précédemment, nous obtenons la formule de la régression :

$$R_{Stratégie Clustering} = \beta R_{Benchmark} + \epsilon \quad \text{avec} \quad \mathbb{E}(R_{Benchmark}\epsilon) = 0 \tag{10}$$

Nous pouvons maintenant expliquer la théorie de protection du portefeuille « clusterisé ». Celle-ci consiste à se concentrer sur les résidus de la régression précédente (Formule n°10) : nous allons ainsi obtenir les rendements résiduels. Ainsi, les positions que nous allons adopter avec cette stratégie combinant à la fois le clustering et la « protection » sont déterminées par la matrice suivante :

$$Position_{Stratégie\text{"}Clustering+protection\text{"}} = Position_{StratégieClustering} - \beta Position_{Benchmark} \tag{11}$$

Avec

— $Position_{Stratégie«Clustering+protection»}$, la matrice des positions avec la stratégie « Clustering et protection »

— $Position_{StratégieClustering}$, celle avec la stratégie Clustering

— $Position_{Benchmark}$, celle avec la stratégie Benchmark

— β qui est défini de la même manière que dans la formule n°9

Pour plus de clarté et à des fins pédagogiques, nous allons désormais travailler avec le vecteur des gains à la place de celui des rendements. Quels sont les plus-values et/ou les pertes amenées par ce travail sur les résidus ?

Premièrement, nous allons expliquer l'impact que cela va avoir sur la volatilité de notre portefeuille. Calculons la variance de notre nouveau portefeuille :

$$\begin{aligned}\mathbb{V}(Gains_{StratégieClustering}) &= \mathbb{V}((Gains_{StratégieClustering} - \beta Gains_{Benchmark}) + \beta Gains_{Benchmark}) \\ &= \mathbb{V}(Gains_{StratégieClustering} - \beta Gains_{Benchmark}) + \mathbb{V}(\beta Gains_{Benchmark})\end{aligned}$$

Nous avons ici utilisé l'indépendance entre les résidus de la régression linéaire et la variable explicative, ici $Gains_{Benchmark}$. Ainsi, nous obtenons :

$$\begin{aligned}\mathbb{V}(Gains_{Stratégie"Clustering+protection"}) &= \mathbb{V}(Gains_{StratégieClustering}) - \beta^2 \mathbb{V}(\beta Gains_{Benchmark}) \\ &\leq \mathbb{V}(Gains_{StratégieClustering})\end{aligned}$$

De ce fait, travailler sur le portefeuille des résidus nous permet d'utiliser la corrélation entre notre stratégie de clustering et celle sur le Benchmark afin de « protéger » notre portefeuille clusterisé par rapport au marché et ainsi réduire sa volatilité. Comme nous l'avons vu précédemment, cela est à notre avantage puisque, toutes choses égales par ailleurs, une réduction de la volatilité de notre portefeuille fait augmenter le ratio de Sharpe.

Deuxièmement, regardons l'impact que cela va avoir sur nos gains par rapport à la stratégie utilisant seulement le clustering. Calculons l'espérance de gains de notre nouveau portefeuille :

$$\begin{aligned}\mathbb{E}(Gains_{Stratégie"Clustering+protection"}) &= \mathbb{E}(Gains_{Stratégie"Clustering"}) - \beta Gains_{Benchmark}) \\ &= \mathbb{E}(Gains_{Stratégie"Clustering"}) - \beta \mathbb{E}(Gains_{Benchmark}) \\ &\leq \mathbb{E}(Gains_{Benchmark}) \text{ si } \beta \geq 0\end{aligned}$$

Ainsi, il est clair que lorsque le paramètre β est positif, la stratégie combinant « protection » et clustering est coûteuse en termes de rendement puisque l'espérance de gain est plus faible, en comparaison avec la stratégie basée sur un clustering simple. Nous verrons par la suite que comme attendu, le β obtenu sera toujours positif quel que soit la stratégie mise en place, ce qui est cohérent puisque ces dernières ont toutes comme point de départ le Benchmark.

En conclusion, cette méthode de protection génère deux effets opposés sur notre portefeuille d'investissement clusterisé. D'une part, l'aspect « protection » entre en jeu en faisant baisser la volatilité de notre portefeuille initial : cela provoque une augmentation du ratio de Sharpe. D'autre part, cette méthode est coûteuse en rendement, ce qui va faire baisser le ratio de Sharpe. Nous verrons donc, dans la pratique, quel effet l'emporte et donc nous pourrons analyser l'utilité de cette méthode dans notre cas.

4.2 Résultats obtenus

Afin d'évaluer la pertinence de nos indicateurs dans le but de maximisation du ratio de Sharpe, nous avons discriminé les actions en deux parties, pour chaque date : le premier groupe étant celui composé des actions ayant leur indicateur supérieur à la médiane et le second de celles dont l'indicateur est inférieur à la médiane. Pourquoi avoir choisi de répartir nos actions en seulement deux groupes pour chaque paramètre ?

Cela va nous permettre à la fois de profiter d'une espérance de gains plus élevée, qu'avec une division en un plus grand nombre de groupes, mais également d'une réduction du risque non systématique de notre portefeuille. En effet, ce dernier représente le risque que la valeur d'un placement change en raison de facteurs qui sont propres à cet investissement, et non au marché en général. Ainsi, il est possible de réduire ce risque spécifique grâce à la diversification (contrairement au risque de marché qui lui, existe quoiqu'il arrive). De ce fait, conserver un nombre suffisant d'actions permet d'exploiter le phénomène de diversification du portefeuille dont nous venons de parler. Dans une partie suivante, nous mettrons en place cette stratégie de manière plus poussée en utilisant notamment la théorie moderne du portefeuille de Markowitz.

Nous allons maintenant présenter les résultats obtenus pour les différentes stratégies dans le tableau suivant :

Stratégies	Ratio de Sharpe	Sharpe Résiduel
Coefficients β ayant les plus grands R^2	0.666	-0.401
Coefficients β ayant les plus faibles R^2	0.808	0.409
Coefficients β significatifs à 5%	0.754	0.175
Conditionnement de la matrice de la régression	0.754	0.231
50% des actions les plus volatiles	0.523	-0.986
50% des actions les moins volatiles	1.246	1.006
Combinaison des méthodes de sélection des actions les moins volatiles et avec les R^2 les plus faibles	0.732	0.514

Tableau 4 : Tableau regroupant les ratio de Sharpe des stratégies testées

Effectuons l'analyse de nos résultats. Tout d'abord, nous pouvons remarquer que les stratégies de clustering que nous avons implémentées ont été utiles à l'amélioration de la stratégie initiale Benchmark. En effet, le « filtrage » des actions suivant les indicateurs de volatilité et de R^2 font augmenter considérablement le ratio de Sharpe de notre portefeuille « clusterisé », par rapport au Benchmark. Nous avons donc jugé utile de tester la stratégie visant à ne conserver que les actions étant à la fois les moins volatiles et dont les régressions associées présentaient les R^2 les plus faibles.

Comme attendu, conserver seulement les actions les moins volatiles va nous permettre de diminuer considérablement la volatilité de notre portefeuille. Nous pourrions, malgré tout, nous attendre à ce que cela se fasse au grand détriment de l'espérance de gain et ainsi que cela soit à l'origine d'une diminution du ratio de Sharpe. Cependant nous obtenons avec cette stratégie le meilleur ratio de Shape de notre projet avec une valeur de 1.246. Cette stratégie surperforme largement toutes les autres que nous avons testées, notamment la stratégie inverse qui visait à miser sur les actions les plus volatiles. Cette observation est liée à un phénomène connu sous le nom de *Low-volatility anomaly*. Ce dernier, basé sur des observations empiriques, stipule que les actions à faible volatilité ont un rendement plus élevé que les actions à forte volatilité dans la plupart des marchés. Il s'agit d'un exemple d'anomalie boursière puisqu'il contredit la prédiction centrale de nombreuses théories financières, dont notamment le modèle CAPM, selon laquelle une prise de risques doit être rémunérée par un rendement plus élevé.

Discutons désormais des effets de la stratégie de « protection » dans la pratique. Comme l'attestent les résultats regroupés dans le tableau n°4, la « protection » dégrade systématiquement le ratio de Sharpe du portefeuille initial. Ceci montre donc que l'effet de réduction de l'espérance de gain l'emporte sur celui réduisant la volatilité du portefeuille initial, d'où une diminution globale du ratio de Sharpe. Cette stratégie s'est donc avérée peu concluante dans notre cas.

Pour conclure, nous avons représenté sur un graphique les courbes des gains cumulés de nos trois meilleures stratégies, d'après le tableau n°4, ainsi que ceux du Benchmark. Il est intéressant de noter la forte volatilité du portefeuille Benchmark en comparaison aux trois autres stratégies. Les trois meilleurs portefeuilles ont tous un ratio de Sharpe supérieur à celui de la stratégie initiale Benchmark. Les dénominateurs communs de nos trois stratégies sont : des gains inférieurs à celui de la stratégie Benchmark mais une volatilité beaucoup plus faible. De ce fait, nous aurions tendance à penser qu'en pratique, la minimisation de la volatilité importe plus que la maximisation des gains dans la maximisation du ratio de Sharpe.

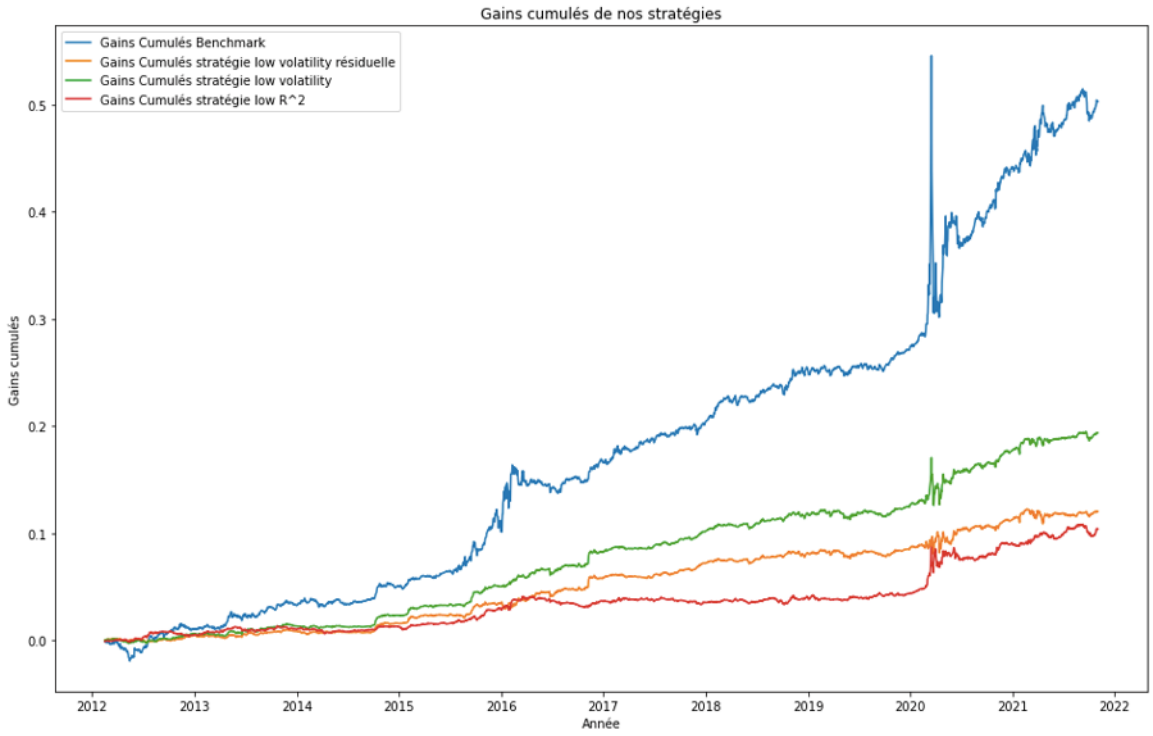

FIGURE 2 – Gains cumulés de nos 3 meilleures stratégies et ceux de la stratégie Benchmark

5 Mise en place de nouvelles régressions

Nous nous intéressons désormais à deux variantes de la régression linéaire : la régression Lasso et la régression Ridge. Ces deux méthodes mènent à deux estimateurs ayant des propriétés intéressantes et complémentaires.

5.1 Régression Lasso

La régression Lasso permet de filtrer les variables explicatives.Ainsi cette méthode permet de faire face au problème d'*overfitting*. L'estimateur Lasso est solution du problème : $\min_{\beta \in \mathbb{R}^d} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1$.
La pénalité en norme 1 permet de contrôler la valeur des coefficients en introduisant un biais qui réduit leur valeur ; la pénalité étant proportionnelle à λ, plus λ est grand, plus le nombre de variables dont les coeffecients sont mis à zéro sera élevé. Cet effet peut se voir dans l'expression analytique de $\hat{\beta}^{Lasso}$ dans le cas de données orthonormales (*i.e.* $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_p$), la coordonnée j de l'estimateur de la régression Lasso s'écrivant : $\hat{\beta}_j^{Lasso} = \text{sgn}(\hat{\beta}_j^{Lasso}) \cdot \max(0, |\hat{\beta}_j^{Lasso}| - \lambda)$.

Pour des questions de temps de calcul, nous avons d'abord effectué plusieurs régressions Lasso pour dix actions afin de déterminer la valeur de l'hyperparamètre λ qui laisse une, deux, trois et quatre variables non nuls. Ces dix variables ont été sélectionnées dans un intervalle formé par les déciles de volatilité pour avoir un échantillon plus représentatif. Cependant, pour les valeurs testées, nous observons que le nombre

de variables dont le coefficient est non nul peut varier selon l'action ou le temps ; nous regardons donc l'évolution de la valeur moyennée sur les actions et sur le temps de chacun des coefficie3ts (Figure 3) et sélectionnons les valeurs de λ qui nous semblent pertinentes.

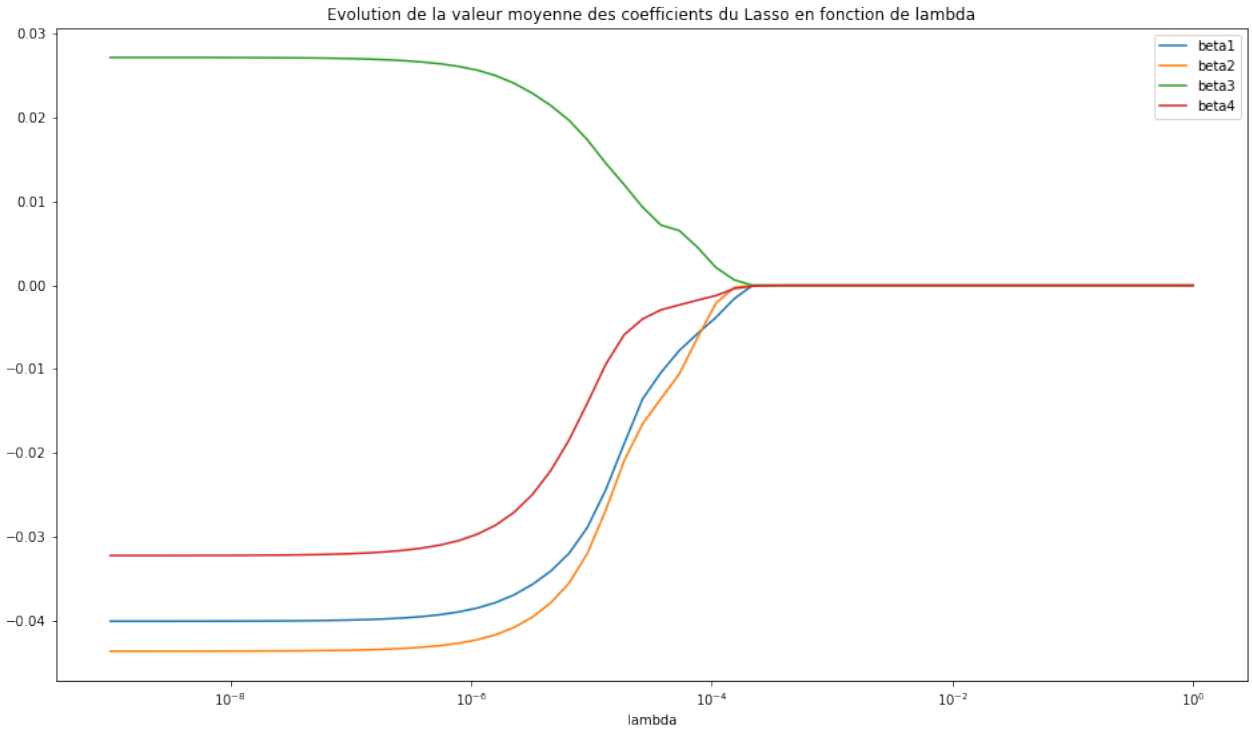

FIGURE 3 – Evolution de la valeur moyenne des coefficients du Lasso en fonction de lambda

Pour ces valeurs de λ, nous effectuons une régression Lasso pour l'ensemble de nos actions. Nous utilisons les coefficients obtenus de deux manières : tels quels et en les utilisant pour sélectionner les variables explicatives (celles avec un coefficient non nul) pour effectuer une régression linéaire classique. Dans les deux (Tableau 5 et Tableau 6) , et pour tous les paramètres λ retenus, nous obtenons des performances moindres en comparaison avec le Benchmark, ce qui suggère l'absence d'*overfitting* dans notre modèle initial. Un inconvénient de la régression Lasso est qu'en présence de fortes corrélations entre les variables explicatives importantes pour la prédiction, le Lasso n'en sélectionnera qu'une seule.

5.2 Régression Ridge

La régression Ridge permet d'estimer les coefficients lorsque les variables explicatives sont fortement correlées, ce qui compense le défaut de la régression Lasso.

Lambda	Perf pf réel	SR	Perf ptf residuel	SR residuel
0.000001	0.966766	0.435625	0.482931	0.227801
0.000003	0.983169	0.440476	0.492225	0.231051
0.00001	1.004818	0.443218	0.533681	0.245350
0.000032	0.808804	0.351087	0.521943	0.229858
0.000100	0.261324	0.128674	0.445388	0.221007
0.000158	0.430460	0.210332	0.349943	0.171230

Tableau 5 : Résultats de nos régressions linéaires Lasso

Lambda	Perf pf réel	SR	Perf ptf residuel	SR residuel
0.000001	0.958573	0.433020	0.478899	0.226335
0.000003	0.977692	0.441022	0.487476	0.230510
0.00001	0.924568	0.413534	0.430430	0.201810
0.000032	0.721387	0.313503	0.360605	0.160383
0.000100	-0.101495	-0.047057	-0.080296	-0.037231
0.000158	0.245530	0.137656	0.505446	0.289156

Tableau 6 : Résultats de nos régressions linéaires avec sélection par le Lasso

L'estimateur est solution du problème : $\min_{\beta \in \mathbb{R}^d} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$. Contrairement à la régression Lasso, nous disposons toujours d'une expression analytique pour l'estimateur : $\hat{\beta}^{Ridge} = (X^T X + \lambda I_p)^{-1} X^T y$. La régression Ridge est bien adaptée dans le cas où le problème est mal posé ($X^T X$ est mal conditionnée voire n'est pas inversible) : en effet, nous pouvons observer un terme additionnel dans l'estimateur Ridge par rapport à l'OLS, qui permet de d'améliorer le conditionnement de la matrice.

De plus, la régression Ridge mène à un estimateur biaisé mais dont la variance est plus faible, ce qui peut amener à une meilleure performance (meilleur ratio de Sharpe).

Nous avons appliqué la même méthodologie que précédemment pour choisir différents λ mais sans tenir compte du critère de la nullité des coefficients (Figure 4) puisque la régression Ridge, contrairement à la régression Lasso ne met les coefficients à zéro, mais les fait tendre vers zéro. En effet, la pénalité en norme 2, favorise les solutions dont les normes sont petites comme dans le Lasso mais sans les annuler (on ne peut donc pas réduire le nombre de variables explicatives avec ce modèle).

Les résultats obtenus dans les régressions Ridge sont également moins bons que ceux obtenus dans le cadre d'une régression linéaire classique (Tableau 7).

En conclusion, toutes les variables explicatives utilisées pour obtenir le modèle Benchmark sont pertinentes dans la régression linéaire.

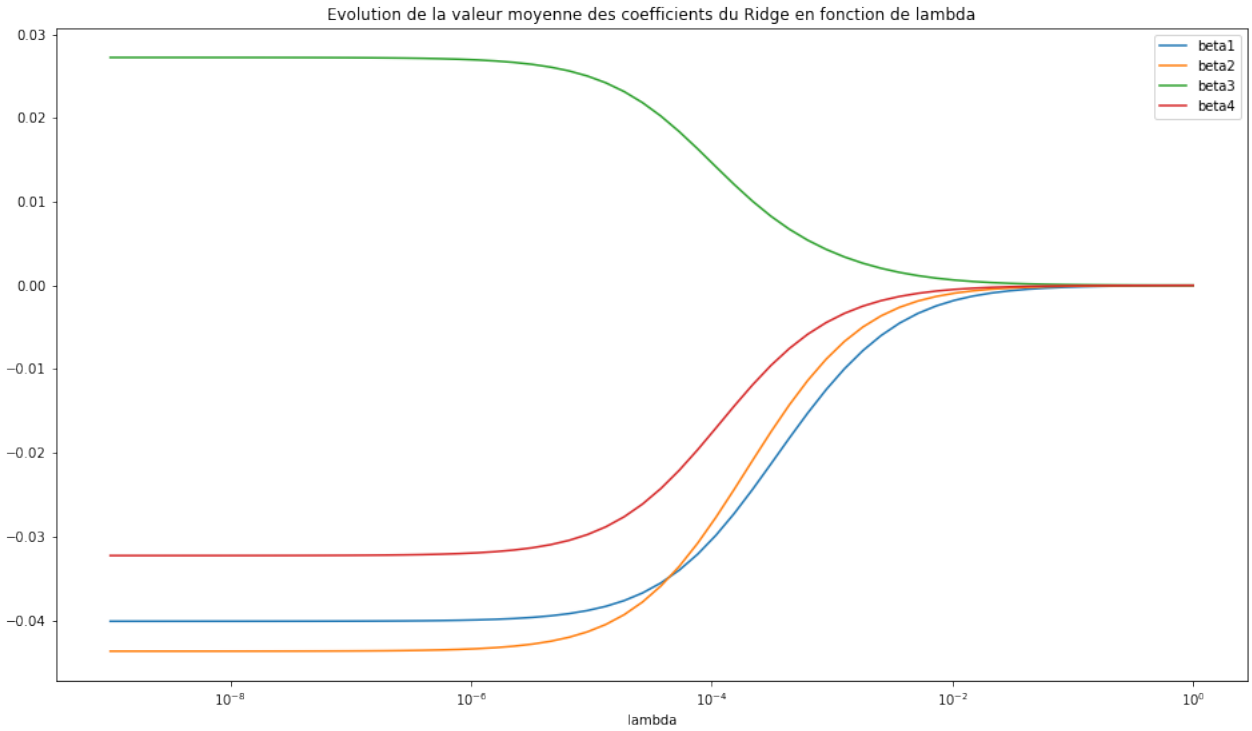

FIGURE 4 – Evolution de la valeur moyenne des coefficients du Ridge en fonction de lambda

lambda	Perf ptf réel	SR	Perf ptf residuel	SR residuel
0.00001	0.972496	0.438561	0.497865	0.234627
0.0001	0.952037	0.423020	0.480554	0.222708
0.000316	0.905253	0.399347	0.446532	0.204857
0.001	0.847717	0.373104	0.408296	0.186229
0.01	0.798998	0.350937	0.380177	0.172440
0.025119	0.800330	0.351336	0.383554	0.173817

Tableau 7 : Résultats de nos régressions linéaires Ridge

6 Amélioration de notre modèle par la théorie moderne du portefeuille de Markowitz

6.1 Explication de la stratégie

Nous allons dans cette partie nous intéresser à une amélioration de notre modèle à travers la théorie moderne du portefeuille de Markowitz. En effet, cette théorie explique comment des investisseurs ra-

tionnels utilisent la diversification des actifs afin de minimiser le risque de leur portefeuille. Elle stipule également qu'il existe un portefeuille optimal que nous allons désormais caractériser.

Pour ce faire, nous allons noter C_t la matrice de variance-covariance des rendements associée à la date t. Pour la calculer, nous avons considéré à une date t les données des cinq années qui précèdent cette dernière pour calculer la matrice de variance-covariance. Ainsi, comme pour les régressions linéaires, nous possédons autant de matrices de variance-covariance que de dates à partir de 2017. D'après la théorie de Markowitz, la position optimale à la date t est donnée par $P'_t = C_t^{-1} P_t$ où P_t représente la position que nous avons obtenue pour la stratégie *"Low Volatility"*. Le ratio de Sharpe associé à la stratégie *"Low Volatility"* sur les données depuis 2017 est de 1.05. Cependant, lors de l'implémentation de cette nouvelle position, notre ratio de Sharpe ne s'améliore pas.

6.2 Instabilité des matrices de variance-covariance et amélioration du modèle

Cette performance moindre du ratio de Sharpe est due, en pratique, à l'instabilité des matrices de variance-covariance et notamment de leurs faibles valeurs propres. En effet, augmenter les valeurs propres les plus faibles évite à la matrice C_t^{-1} de prendre des coefficients très élevés étant donné qu'elle est semblable au sens matriciel à sa matrice diagonale qui contient les inverses des valeurs propres de la matrice C_t. Ainsi, nous évitons de mettre une position trop forte sur ces actions assurant une meilleure diversification du portefeuille au sens de Markowitz.

Afin de modifier ces faibles valeurs propres dans les matrices de variance-covariance, nous allons utiliser le théorème spectral. En effet, les matrices C_t étant des matrices symétriques positives que nous supposerons définies positives, nous avons d'après le théorème spectral $C_t = O_t^\top D_t O_t$ où O_t est une matrice orthogonale qui caractérise les directions des vecteurs propres et D_t est une matrice diagonale contenant les valeurs propres de C_t. Nous allons artificiellement modifier les valeurs de propres de C_t en modifiant la matrice D_t. Nous noterons $\lambda_{1,t}, \lambda_{2,t}, .., \lambda_{N,t}$ les valeurs propres rangées dans l'ordre décroissant. Pour ce faire, nous avons fait le choix suivant : les $\sqrt{N}$ valeurs propres les plus grandes de la matrice ne sont pas modifiées et les $N - \sqrt{N}$ autres sont mises égales à leur moyenne. Cela nous permet d'avoir un contrôle sur les faibles valeurs propres. Notons $\tilde{C}_t$ la matrice de variance-covariance ainsi modifiée. Ce rapport des valeurs propres pour chaque matrice de variance-covariance représente également une mesure du conditionnement de la matrice (notée κ) associée à la $\|\|_2$ matricielle. En effet, C_t étant une matrice symétrique, nous savons que $\|C_t\|_2 = \rho(C_t)$ où $\rho(A) = \max_i |\lambda_i|$ où les λ_i représentent les valeurs propres de la matrice A. Dans notre cas, $\rho(C_t) = \lambda_{1,t}$ et $\rho(C_t^{-1}) = \lambda_{N,t}$. Ainsi, nous obtenons $\kappa(C_t) = \dfrac{\|C_t\|_2}{\|C_t^{-1}\|_2} = \dfrac{\lambda_{1,t}}{\lambda_{N,t}}$ qui est donc la mesure de dispersion que nous avons choisie. Nous avons alors représenté sur le graphique ci-dessous en échelle logarithmique l'allure du rapport des valeurs propres $(\dfrac{\lambda_{1,t}}{\lambda_{N,t}})$ pour les matrices C_t (en orange) et les matrices $\tilde{C}_t$ (en bleu).

Ce graphique confirme donc bien que notre modification des valeurs propres a rendu les matrices de variance-covariance plus "stables". Par ailleurs, nous nous sommes intéressés à la dispersion de ces

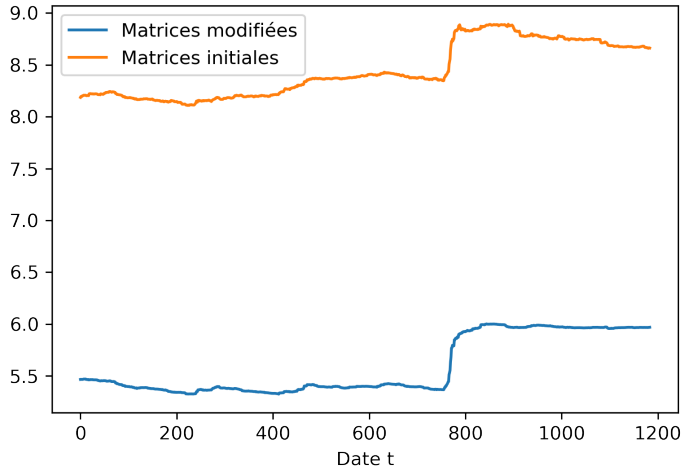

FIGURE 5 – Evolution du rapport des valeurs propres pour différentes matrices de variance-covariance en échelle logarithme

rapports de valeurs propres dans les deux cas. Nous trouvons pour les C_t initiales un écart-type de 1292 et pour les $\tilde{C}_t$ modifiés un écart-type de 82.

Quel impact cette évolution de la théorie de Markowitz a-t-elle sur le ratio de Sharpe ? Les nouvelles positions caractérisées par : $\tilde{P}_t = {\tilde{C}_t}^{-1} P_t$ où P_t représente la position que nous avons obtenue pour la stratégie *"Low Volatility"*, nous ont permis d'obtenir un ratio de Sharpe de 1.20 contre 1.05 avec la stratégie précédente. Ceci nous semble cohérent puisque selon la théorie du portefeuille optimal de Markowitz, une meilleure diversification est censée nous assurer une variance du portefeuille plus faible et donc avec des gains similaires une augmentation du ratio de Sharpe. Nous traçons égalemement ci-dessous l'allure des gains cumulés pour la position associée à la théorie de Markowitz améliorée.

A titre informatif, d'autres transformations de la matrice de variance-covariance existent. En effet, nous pouvons pondérer de manière convexe les matrices C_t en considérant :

$C_{t,\alpha} = (1-\alpha)C_t + \alpha \frac{Tr(C_t)}{N} I_N$ où N représente le nombre d'actions et $\alpha \in [0,1]$.

Cela permet également d'éviter des valeurs propres trop faibles en les pondérant par leur moyenne. Ainsi, cette technique est similaire à la nôtre puisque son but est également d'avoir une maîtrise sur les valeurs propres des matrices de variance-covariance C_t.

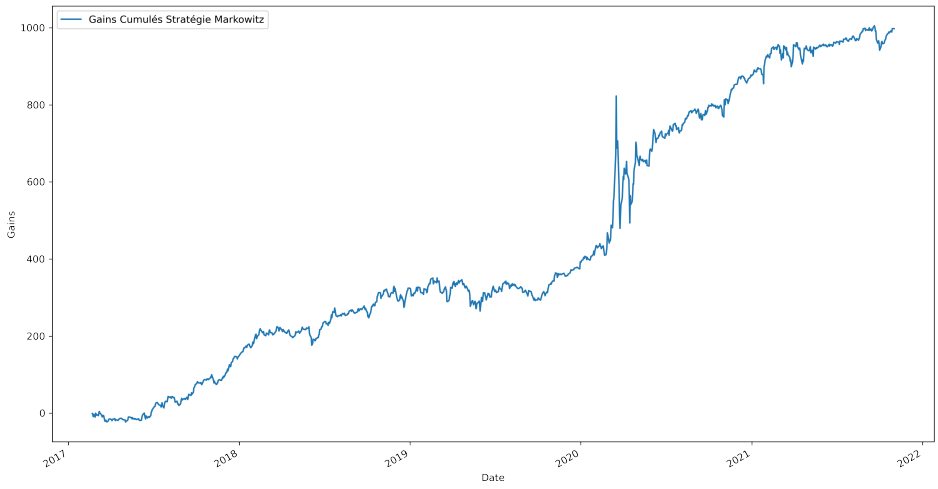

FIGURE 6 – Gains cumulés liés à la stratégie de portefeuille Markowitz amélioré

7 Conclusion

Ce projet visant à construire un portefeuille d'actions maximisant l'indicateur du ratio de Sharpe, se basait sur une prédiction en temps réel des fluctuations des actions de l'indice SP 500. Pour se faire, nous avons utilisé un modèle de régression linéaire multiple afin de créer une stratégie automatique de trading. Toujours dans l'optique d'une maximisation de la performance du portefeuille issu de notre stratégie, nous avons mis en place des méthodes de clustering, basées sur des indicateurs tels que la qualité de prédiction de notre modèle ou la volatilité des actions.

Il s'est avéré que la stratégie la plus performante, à l'issue de nos différentes tentatives, était celle visant à n'investir que sur les 50% des actions les moins volatiles sur les cent jours précédant la date de notre investissement. Ce phénomène est connu sous la dénomination *Low- volatility anomaly*. Cette stratégie permettait d'obtenir un ratio de Sharpe de 1.05 sur la période allant de l'année 2017 à l'année 2022.

Toujours en quête d'améliorations, nous avons décidé d'implémenter une stratégie basée sur la théorie moderne du portefeuille, développée par l'économiste américain Harry Markowitz, que nous avons par la suite appliquée à notre portefeuille le plus performant jusqu'alors. La diversification prônée par cette théorie, nous a permis de réduire une nouvelle fois le risque associé à notre investissement et ainsi d'obtenir un ratio de Sharpe de 1.20, ce dernier étant supérieur à celui la stratégie précédente, sur la même période.

Il existe néanmoins, de possibles améliorations concernant le modèle que nous avons utilisé. Ces dernières pourraient par exemple se baser sur l'utilisation de forêts aléatoires ou de réseaux de neurones afin notamment de gagner en finesse de sélection.

8 Bibliographie

[1] : Jarrow, R. A., Murataj, R., Wells, M. T., Zhu, L. *The Low-volatility Anomaly and the Adaptive Multi-Factor Model*. SSRN Electronic Journal. (2021)

[2] : Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning*. (2001)

[3] : Youssef Louraoui. *Markowitz-asset-allocation-model*. (2022)

[4] : Nop Sopipan. *Forecasting the financial returns for using multiple regression based on principal component analysis*. (2013)

2ÈME ANNÉE ENSAE

NOTE DE SYNTHÈSE DU PROJET DE STATISTIQUES APPLIQUÉES

MENTION CONFIDENTIEL

Prédiction temps réel des fluctuations des actions avec un modèle de multirégression et recherche d'améliorations de ce modèle par méthodes de clustering

---

Samy Mekkaoui-Simon Depaule-Eros Giovanetto-Alex Jin

Sous la Supervision de Monsieur Vivien Bouche

Année Universitaire 2021-2022

1 Présentation du projet et de ses objectifs

L'objectif de notre projet est de construire un modèle de multirégression afin de prédire les actions. Pour ce faire, nous avons travaillé sur les actions du S&P 500 pour la période 2006-2021. Avant de construire notre base de données nous devions nous assurer que les actions que nous allions selectionner allaient rester cotées tout au long de la période d'étude. Nous avons donc supprimer celles qui auraient pu rentrer ou sortir de l'indice au cours de cette période. Par ailleurs, nous avons du effectuer différentes modifications sur notre base de données notamment borner les rendements à plus ou moins 15% afin de minimiser les effets marginaux. A l'issue de la construction de notre base de données, nous étions donc en mesure de mettre en place notre modèle que nous expliciterons à la partie suivante.

2 Construction du modèle initial

Dans ce mémoire, nous utilisons le modèle suivant où N représente le nombre d'actions et T le nombre de jours sur lequel nous travaillerons :

$$\forall i \in [\![1\ ;\ N]\!], \quad \sum_{l=1}^{5} Y_{t+l}^{(i)} = \beta_{1,t}^{(i)} X_t^{(i)} + \frac{\beta_{2,t}^{(i)}}{\sqrt{5}} \sum_{l=1}^{5} X_{t-l}^{(i)} + \frac{\beta_{3,t}^{(i)}}{\sqrt{21}} \sum_{l=1}^{21} X_{t-l}^{(i)} + \frac{\beta_{4,t}^{(i)}}{\sqrt{256}} \sum_{l=1}^{256} X_{t-l}^{(i)} + \epsilon_t^{(i)} \tag{1}$$

où :

1. $Y_t^{(i)} := X_t^{(i)} - \frac{1}{N} \sum_{j=1}^{N} X_t^{(j)}$ représente le rendement net de celui du marché.
2. $R_t^{(i)} := \sum_{l=1}^{5} Y_{t+l}^{(i)}$ représente la performance de l'action sur les cinq prochains jours nette du marché
3. $X_{1,t}^{(i)} := X_t^{(i)} = ln(\frac{P_t^{(i)}}{P_{t-1}^{(i)}})$ représente le rendement de l'action i à la date t.
4. $X_{2,t}^{(i)} := \frac{1}{\sqrt{5}} \sum_{l=1}^{5} X_{t-l}^{(i)}$, $X_{3,t}^{(i)} := \frac{1}{\sqrt{21}} \sum_{l=1}^{21} X_{t-l}^{(i)}$ et $X_{4,t}^{(i)} := \frac{1}{\sqrt{256}} \sum_{l=1}^{256} X_{t-l}^{(i)}$ représente respectivement la somme normalisée des rendements de l'action i sur les cinq derniers jours, le dernier mois et la dernière année boursière rapport à la date t.
5. $\epsilon_t^{(i)}$ représente le résidu de la régression linéaire associée

Ainsi, dans ce modèle nous avons effectué autant de régressions que de jours de tradings (T) et de nombre d'actions (N) avec $R_t^{(i)}$ comme variable d'intérêt et $X_t = (X_{1,t}^{(i)}, X_{2,t}^{(i)}, X_{3,t}^{(i)}, X_{4,t}^{(i)})^\top$ qui représente le vecteur des variables explicatives.

A partir de ce modèle, nous obtenons $\forall i \in [\![1\ ;\ N]\!], \forall t \in [\![1\ ;\ T]\!], \hat{R}_t^{(i)} = \sum_{l=1}^{5} \hat{Y}_{t+l}^{(i)}$: la valeur prédite de la somme des rendements de l'action i sur les cinq prochains jours. Notre position à la date t pour une action i qui signifie la part que nous investissons dans une action i à la date t sera caractérisée par la formule suivante : $P_t^{(i)} = \frac{1}{5} \sum_{l=1}^{5} \hat{R}_{t-l}^{(i)}$.

Dès lors, nous pouvons obtenir nos gains, notés par la suite $G_t^{(i)}$, définis par la formule suivante : $G_t^{(i)} = P_t^{(i)} X_{t+1}^{(i)}$. Les gains journaliers de notre portefeuille sont alors caractérisés comme suit : $G_t = \sum_{i=1}^{N} G_t^{(i)}$. Notons "G$ains$" le vecteur des gains journaliers. Notre but sera alors de maximiser la quantité suivante : $SR = \frac{\hat{\mu}_{Gains}}{\hat{\sigma}_{Gains}}$ où $\hat{\mu}_{Gains}$ représente la moyenne empirique du vecteur G$ains$ et $\hat{\sigma}_{Gains}$ son écart-type empirique.

La quantité SR introduite représente le ratio de Sharpe, une notion fondamentale dans notre projet puisque cette dernière nous renseigne sur la qualité d'une stratégie. En effet, une stratégie nous permettant d'obtenir un ratio de Sharpe superieur à 1 sera considérée comme viable financièrement. C'est cet indicateur que nous chercherons constamment à maximiser.

Pour notre stratégie initiale que nous appellerons par la suite "Stratégie Benchmark" nous obtenons un ratio de Sharpe de 0.74 ce qui signifie que les gains associés à notre portefeuille ne nous couvrent pas assez du risque que nous prenons. Nous allons donc chercher à améliorer notre modèle en nous basant sur le Benchmark que nous avons défini plus haut.

3 Amélioration du modèle par des méthodes de clustering

Dans le but d'améliorer notre stratégie, nous avons utiliser des méthodes de "clustering" c'est à dire que nous avons « filtrer » les actions sur lesquelles nous allons investir en fonction de différents indicateurs.

Les indicateurs que nous avons jugés utiles afin de distinguer les actions sont :

1. **La qualité de prédiction de notre modèle via l'utilisation du R^2 :**

   En effet, pour une régression donnée à la date t pour une certaine action (i), cette grandeur représente la part de variance du rendement net de cette action sur les cinq jours suivants, expliquée par notre modèle. Nous aurions tendance à penser que plus le R^2 de notre régression est élevé, plus notre modèle prédit bien les variations du rendement de l'action (i).

2. **La volatilité de l'action sur les 100 jours précédant la date de notre régression :**

   Elle représente la dispersion des rendements de l'action par rapport à leur moyenne sur les 100 jours précédents. Étant donné que notre but est de maximiser le ratio de Sharpe, grandeur dépendant négativement de la volatilité, nous aurions ainsi tendance à sélectionner les actions les moins volatiles.

3. **La sensibilité des calculs à une erreur : le conditionnement de la matrice $X^\top X$:**

   Ce paramètre caractéristique d'une matrice donne une idée de la manière dont l'erreur relative augmente lors d'une opération. Ainsi, plus une matrice est mal conditionnée (conditionnement est grand), plus les calculs deviennent sensibles à une éventuelle erreur.

4. **La significativité à 5% des coefficients de notre régression :**
   En effet, pour la régression d'une action donnée, à une date donnée, si le coefficient associé à une variable explicative de notre modèle n'est pas significatif à 5%, nous ne pouvons pas rejeter l'hypothèse nulle stipulant que cette variable explicative n'influe pas sur la variable cible.

Le tableau ci-dessous donne les résultats des ratios de Sharpe que nous obtenons pour nos différents clusters. Finalement, la meilleure stratégie est celle qui consiste à miser sur les actions les 50% les moins volatiles. Ce résultat est connu en finance sous le nom de *Low Volatility Anomaly*.

Stratégies	**ratio de Sharpe**
β ayant les plus grands R^2	0.666
β ayant les plus faibles R^2	0.808
β significatifs à 5%	0.754
Conditionnement de la matrice de la régression	0.754
50% des actions les plus volatiles	0.523
50% des actions les moins volatiles	1.246
Actions les moins volatiles + celles avec les R^2 les plus faibles	0.732

Tableau 1 : Tableau regroupant les ratio de Sharpe des stratégies testées

4 Utilisation de la méthode des alphas résiduels

Afin d'améliorer une nouvelle fois notre stratégie, nous avons utilisé la méthode dite des alpha résiduels. Cette dernière consiste à combiner la stratégie Benchmark avec les différentes stratégies de clustering que nous avons énumérés afin de « protéger » notre portefeuille clusterisé par rapport au marché et ainsi réduire sa volatilité pouvant nous permettre d'augmenter notre ratio de Sharpe.

A l'issue de cette partie, la stratégie que nous retenons, c'est à dire celle maximisant le ratio de Sharpe est celle associée à la *Low Volatility*.

5 Stratégie liée au portefeuille optimal de Markowitz

Dans cette partie, nous allons nous intéresser à la mise d'une nouvelle stratégie : la stratégie du portefeuille optimal de Markowitz. Cette théorie explique comment des investisseurs rationnels utilisent la diversification des actifs afin de minimiser le risque de leur portefeuille. Elle stipule également qu'il existe un portefeuille optimal que nous allons désormais caractériser. D'après cette théorie, la position optimale à la date t est donnée par $P'_t = C_t^{-1} P_t$ où P_t représente la position que nous avons obtenue pour la stratégie *"Low Volatility"* et C_t représente la matrice de variance-covariance des rendements à la date t. Cependant, le ratio de Sharpe associé à cette stratégie ne s'améliore pas. Cela s'explique notamment par l'instabilité des matrices de variance-covariance dûe aux faibles valeurs propres de C_t. La mise en

place d'une nouvelle stratégie pour contrôler les faibles valeurs propres de C_t nous permet alors d'obtenir un ratio de Sharpe de 1.20 contre 1.05 avec la seule stratégie *Low Volatility*. Nous traçons égalemement ci-dessous l'allure des gains cumulés pour la position associée à la théorie de Markowitz améliorée.

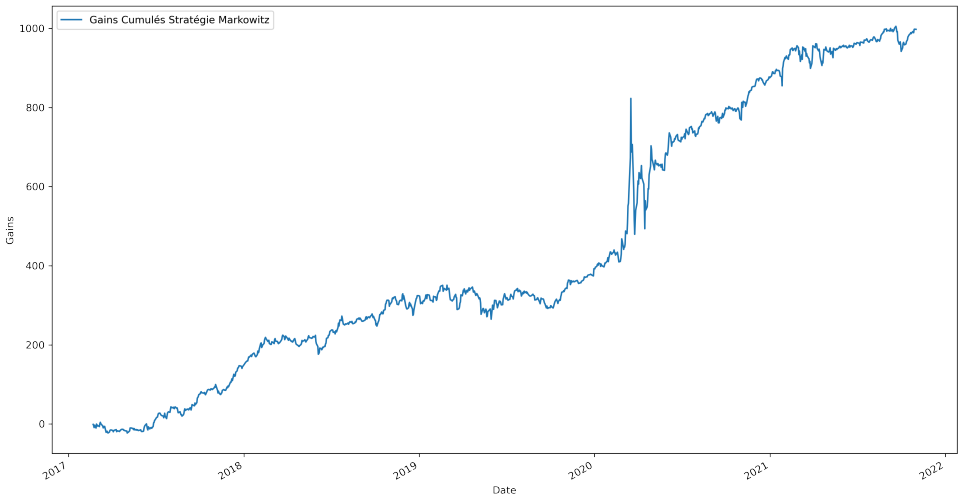

FIGURE 1 – Gains cumulés liés à la stratégie de portefeuille Markowitz amélioré

6 Bilan final du projet

A l'issue de notre projet, il s'est avéré que la stratégie la plus performante, à l'issue de nos différentes tentatives, était celle visant à n'investir que sur les 50% des actions les moins volatiles sur les cent jours précédant la date de notre investissement. Ce phénomène est connu sous nom de *Low- volatility anomaly*. Cette stratégie permettait d'obtenir un ratio de Sharpe de 1.05 sur la période allant de l'année 2017 à l'année 2022.

Toujours en quête d'améliorations, nous avons décidé d'implémenter une stratégie basée sur la théorie moderne du portefeuille, développée par l'économiste américain Harry Markowitz, que nous avons par la suite appliquée à notre portefeuille le plus performant jusqu'alors. La diversification prônée par cette théorie, nous a permis de réduire une nouvelle fois le risque associé à notre investissement et ainsi d'obtenir un ratio de Sharpe de 1.20, ce dernier étant supérieur à celui la stratégie précédente, sur la même période.