

ARTICLE TYPE

Implémentation d'un stratégie Smart Beta et Back-Test

EL KAMLI Adam

Institut National de Statistique et d'Économie Appliquée

Author for correspondence: EL KAMLI Adam, Email: aelkamli@insea.ac.ma.

Abstract

Cet article examine la conception et la performance d'une stratégie smart beta multifactorielle appliquée à l'indice MASI 20, un indice de référence clé du marché boursier marocain. En combinant plusieurs facteurs d'investissement tels que la valeur, la taille, le momentum, la faible volatilité et la qualité, l'étude vise à construire un portefeuille capable de générer des rendements ajustés au risque supérieurs par rapport aux approches traditionnelles pondérées par la capitalisation. En utilisant des données historiques des constituants du MASI 20, nous mettons en œuvre des techniques systématiques de construction de portefeuille basées sur des facteurs et évaluons leurs performances à travers des tests rétroactifs. Les résultats mettent en évidence les avantages potentiels de l'intégration de stratégies multifactorielles, notamment une diversification accrue, une réduction des baisses et une amélioration des ratios de Sharpe. En outre, nous analysons la robustesse de la stratégie dans différents contextes de marché et évaluons les coûts et le turnover liés aux rééquilibrages. Cette recherche contribue à l'enrichissement des travaux sur les stratégies smart beta dans les marchés émergents et offre des perspectives pratiques aux gestionnaires d'actifs cherchant à optimiser les résultats d'investissement sur le marché financier marocain.

Keywords: Smart Beta, Modèle à plusieurs facteurs, Facteurs Fama-Frensh, Backtesting, Analyse de régression

1. Introduction

Au cours des 50 dernières années, l'investissement dans les indices boursiers a révolutionné l'industrie de l'investissement. Plus récemment, l'évolution des préférences des investisseurs et les avancées technologiques ont conduit au développement de nouveaux produits et stratégies basés sur les indices, tels que les fonds négociés en bourse (ETF) smart beta et l'investissement direct. Ces produits basés sur les indices gagnent en popularité parallèlement au passage des investisseurs vers une personnalisation accrue. Cependant, l'introduction de ces nouveaux produits d'investissement sur indice a suscité une certaine confusion quant à ce qui constitue réellement « l'investissement indiciel », compte tenu de la distinction traditionnelle entre gestion « active » et « passive ».

L'objectif principal des modèles factoriels est de comprendre les moteurs des prix des actifs. De manière générale, la logique derrière l'investissement factoriel repose sur le fait que la performance financière des entreprises dépend de certains facteurs, qu'ils soient latents et inobservables ou liés à des caractéristiques intrinsèques (comme les ratios comptables, par exemple). En effet, comme le formule (Cochrane 2011), la première question essentielle est : *"which characteristics really provide independent information about average returns?"* Répondre à cette question aide à comprendre la structure transversale des rendements et peut ouvrir la voie à leur prédiction.

Théoriquement les modèles factoriels linéaires sont un cas spécial du modèle d'évaluation par arbitrage de (Ross 1977), qui assume que le rendement d'un actif n peut être modéliser

comme une combinaisons linéaires de facteurs f_k :

$$r_{t,n} = \alpha_n + \sum_{k=1}^K \beta_{n,k} f_{t,k} + \epsilon_{t,n}$$

On introduit les hypothèse économétriques usuel sur les modèles linéaires à savoir $E(\epsilon_{t,n}) = 0$, $Cov(\epsilon_{t,n}, \epsilon_{t,m}) = 0$, $Cov(f_n, \epsilon_n) = 0$. Si ces facteurs existent, alors ils sont en contradiction avec le modèle le plus célèbre en gestion de portefeuille, le MEDAF (Modèle d'Evaluation des actifs financiers), ou le seul moteur des actifs est le portefeuille du marché (F. 1964), (Lintner), (J. 1966). Pour cette raison on appelle les facteurs des *anomalies*. Depuis la publication dual de (Eugene F. Fama 1992) et (Eugene F. Fama 1993) les preuves empiriques de ces anomalies sont de plus en plus publiés.

2. Sélection des facteurs

Comme présenter dans l'introduction ces facteurs sont issus d'une anomalie par rapport au modèle du MEDAF, et par conséquent le choix des facteurs et dans un premier lieu une observation des anomalies. Et donc, la première étapes dans la construction de notre stratégie est de décider : *qu'elle sont les anomalies présente dans le marché de Casablanca ?*

Afin de cerner ces anomalies on va chercher leurs impacts en construisant deux portefeuilles extrêmes et calculer la différence de leurs rendements, comme ca on peut quantifier l'anomalie qui se présente souvent par un rendement supérieur de l'un des portefeuille par rapport à l'autre d'où la signifiicativité de la différence.

2.1 Construction des facteurs

Les facteurs les plus connus dans ce sens sont les 5 facteurs de Fama-French présenter ci dessous :

1. **Taille** (*SMB = small minus big firms*)
2. **Valeur** (*HML = High minus Low firms*)
3. **Momentum** (*WML = Winners minus Losers*)
4. **Profitabilité** (*RMW : Robust minus weak*)
5. **Investissement** (*CMA = Conservative minus aggressive*)
6. **'Risque' inférieur** (*BAB : Betting Against Beta*)

Dans le cas de Maroc ce genre de donnée n'est malheureusement pas disponible et par conséquent on est dans l'obligation de les calculer soit même.

J'ai collecté le cours historique des trois dernière années (quand ils existent) de 74 entreprises cotés^a dans le marché à partir du siteweb de la bourse de Casablanca^b, ensuite j'ai copié les ratios comptables des années (2023, 2022, 2021) dans un fichier excel nommé (*Fun_data.xlsx*).

La deuxième étape est de catégoriser les actions en trois groupes selon le ratio **B/M** qui est un ratio comptable qui exprime le rapport entre la valeur comptable et la valeur du marché de l'entreprise et il est choisie dans notre contexte comme proxy de la valeur de l'entreprise.

1. Small : $B/M < \text{Quantile}(30\%)$
2. Mid : $\text{Quantile}(30\%) < B/M < \text{Quantile}(70\%)$
3. Big : $B/M > \text{Quantile}(70\%)$

Ensuite on catégorise davantage les actions en deux groupes additionel par rapport à leurs tailles selon leurs capitalisation du marché.

1. Small Cap : $\text{Market_Cap} < \text{Cap_Moyenne}$
2. Big Cap : $\text{Market_Cap} > \text{Cap_moyenne}$

Les données comptables sont publiées annuellement dans le siteweb de la Bourse de Casablanca et par conséquent cette catégorisation se repète chaque année.

Cette catégorisation nous permet de construire 6 portefeuilles extrêmes.

- SS : Small & Small-Cap
- MS : Mid & Small-Cap
- BS : Big & Small-Cap
- SB : Small & Big-Cap
- MB : Mid & Big-Cap
- BB : Big & Big-Cap

Les facteurs de taille et de valeurs^c sont calculer alors :

$$\begin{cases} SMB = \left(\frac{R_{SS} + R_{SB}}{2} \right) - \left(\frac{R_{BS} + R_{BB}}{2} \right) \\ HML = \left(\frac{R_{SB} + R_{MB} + R_{BB}}{3} \right) - \left(\frac{R_{SS} + R_{MS} + R_{BS}}{3} \right) \end{cases}$$

a. La bourse de Casa copte 77 entrerises cotés selon le dernière rapport de 2023

b. casablanca-bourse.com

c. On peut choisir de prendre le MASI Mid & Small Cap comme proxy des entrprise de petite taille.

Ou R_{SS} est exemple le rendement du portefeuille équi pondéré des actions de catégorie (SS).

Ensuite on calcul le facteur de **Momentum** mais cette fois on utilise les rendements des 12 derniers mois comme indicateur, pour ensuite construire deux portefeuilles équi pondérée. L'un avec les 21 premiers entreprises en terme de momentum et l'autre des derniers 21 entreprises.

Pour la **Profitabilité** on suit la même démarche en prennat le ROE (*Return On Equity*) comme proxy et en cotégorisant selon les quantil de 30 et 70 pourcent pour obtenir deux portefeuilles extrêmes ($> 70\%$ et $< 30\%$).

Le dernier facteur qu'on va utiliser est le taux de rendement excédentaire du marché, qui est le facteur initial du modèle du MEDAF. Le calcul ici est simple ou on soustrait le taux sans risque du rendement du MASI l'indice du marché de la bourse de Casa qu'on va prendre comme proxy de marché. Cependant, le choix du taux sans risque est un choix important à prendre (taux de bons de trésors, taux interbancaire, ...). Dans cette étude j'ai pris le taux de bon de trésors du Maroc de maturité 10 ans cependant la qualité et a fiabilité des donnée reste à considérer car j'ai importer l'historique su site web *investing*^d. Le calcul des ces facteurs est enregistré dans une nouvelle base de donnée nommée '*Factors.csv*'.

On peut voir dans la figure 6 l'évolution quotidienne des facteurs calculés.

Ce travail est fait en Python dans le script nommé **Factors.py**.

2.2 Détection des anomalies

Pour la détection des anomalies on va utiliser une méthode proposé dans l'article (Fama 1973), ou on va analyser la structure transversal des rendements des actifs en suivants les deux étapes suivantes :

1. Régression de chaque action sur les facteurs :

$$r_{t,n} = \alpha_n + \sum_{k=1}^K \beta_{n,k} f_{t,k} + \epsilon_{t,n}$$

2. Régression transversale des rendements des actions sur les facteurs faite jours par jours :

$$r_{t,n} = \gamma_{t,0} + \sum_{k=1}^K \gamma_{t,k} \hat{\beta}_{n,k} + \epsilon_{t,n}$$

Ces deux régression nous pemrmettent de caluler la significativté de chaque facteur en calculant la statistique suivante qui suit une loi de Student et qui est basé sur notre estimation des γ appelée '*Risk Premia*' car elles représentent la significativité de l'exposition au facteur:

$$t_k = \frac{\frac{1}{T} \sum_{t=0}^T \hat{\gamma}_{t,k}}{\hat{\sigma}_k / \sqrt{T}}$$

Dans un premier lieu les régression on été faites sur les données quotidienne, mais la significativité globale des facteurs

d. investing.com

étaient négligeable cependant les rendement mensuelle montre autrement ce qui est logique vu que l'effet des facteurs et ces anomalies sont très rarement observer au niveau quotidien. Ceci sont les p-value de chaque facteurs :

Table 1. P-values de chaque facteurs

Taille	Valeurs	Rendement exc	Profitabilité	Momentum
0.9117	$8.417 \times 10^{-10*}$	0.8051	1.646^{-9*}	$1.370 \times 10^{-7*}$

P-valeurs de chaque facteur

* Significative à un seuil de 5%

Comme on peut voir les facteurs de Valeurs, Profitabilité et Momentum sont les facteurs les plus significative dans le marché Marocain avec un niveau de confiance de 95%.

Ceci implique que le marché marocain est sensible aux entreprise présentant des fortes preuve de profitabilité et de valeurs mais aussi de dynamisme.

On trouve de la figure 7 l'évolution des 'Risk Premia' ou on peut observer que les facteur non significative présente une grande volatilité avec des fluctuations permanentes au niveau quotidien.

Le Code de cette partie à été fait en Python dans le fichier source **Anomalies_facteurs.py**

3. Préparation des données

La première partie des données a préparer est l'historique des prix. Puisque notre stratégie s'intègre dans un contexte de stratégie SmartBeta multifactoriel, on va choisir les actions des 20 entreprises qui composent l'indice MASI 20.

A l'addition de ces prix on doit ajouter pour chaque anomalie découverte un facteur pour la quantifier.

3.1 Valeur

Les actions dont les prix sont faibles par rapport à leur valeur fondamentale ont tendance à générer des rendements supérieurs à ceux d'un indice pondéré par la capitalisation. Les facteurs de valeur reflètent cette corrélation et sont conçus pour émettre des signaux d'achat pour les actifs sous-évalués et relativement bon marché, ainsi que des signaux de vente pour les actifs surévalués. Ainsi, au cœur de toute stratégie de valeur se trouve un modèle qui estime la juste valeur ou la valeur fondamentale d'un actif. La juste valeur peut être définie comme un niveau de prix absolu, un écart par rapport à d'autres actifs ou une fourchette dans laquelle un actif devrait se négocier.

Afin de capturer la valeur de l'entreprise on va utiliser les facteurs suivants :

- B/M : Book to Market ratio.
- Rendement des dividendes.

3.2 Taille

Les actions de petites capitalisations ont historiquement tendance à générer des rendements supérieurs à ceux des grandes capitalisations, reflétant une prime de taille. Les facteurs de taille exploitent cette relation et sont conçus pour privilégier les investissements dans les entreprises de plus petite taille, qui

sont souvent perçues comme ayant un potentiel de croissance plus élevé ou des inefficacités de marché exploitables. Ces facteurs envoient généralement des signaux d'achat pour les actifs de petite capitalisation et des signaux de vente pour les actifs de grande capitalisation. Au cœur d'une stratégie basée sur le facteur de taille se trouve un modèle qui classe les actifs en fonction de leur capitalisation boursière, en mettant l'accent sur les opportunités offertes par les entreprises de moindre envergure.

Afin de capturer la taille de l'entreprise on va utiliser :

- Capitalisation.

Certe la taille s'est avéré non significative comme facteur cependant on va l'ajouter dans notre modèle pour une deuxième vérification par la suite.

3.3 Momentum

L'investissement basé sur le momentum est l'une des stratégies factorielles les plus établies, soutenue par des preuves quantitatives depuis les travaux de (Narasimhan Jegadeesh 2011) sur le marché boursier américain. Cette approche suit l'adage : la tendance est votre amie ou laissez vos gagnants courir. Les facteurs de momentum sont conçus pour prendre des positions longues sur les actifs ayant bien performé, tout en prenant des positions courtes sur ceux ayant affiché de faibles performances sur une certaine période. Clifford Asness, fondateur du fonds spéculatif AQR, qui gère 200 milliards de dollars, a récemment présenté des preuves des effets du momentum à travers huit classes d'actifs et marchés différents

Les facteurs utilisés dans cet étude sont :

- Accélérateur de prix : L'accélération des prix calcule le gradient de la tendance des prix (ajusté pour la volatilité) en utilisant une régression linéaire sur les prix quotidiens pour une période longue (6mois) et une période courte (1 mois).
- RSI : Le RSI compare l'ampleur des variations récentes des prix entre les actions pour identifier celles qui sont en situation de surachat ou de survente.
- Momentum du prix : Ce facteur calcule le rendement total sur un nombre donné de jours de bourse précédents. Dans la littérature académique, il est courant d'utiliser les 12 derniers mois, à l'exception du mois le plus récent, en raison d'un effet de retournement à court terme qui est fréquemment observé.
- 52-Week-High : Ce facteur utilise la différence en pourcentage entre le prix le plus récent et le prix le plus élevé des 52 dernières semaines.

3.4 Profitabilité

Les facteurs de qualité peuvent indiquer une surperformance, car des fondamentaux supérieurs tels qu'une rentabilité soutenue, une croissance régulière des flux de trésorerie, un endettement prudent, un faible besoin de financement sur les marchés de capitaux ou un faible risque financier soutiennent la demande pour les actions et appuient le prix de ces entreprises à long

terme. Du point de vue de la finance d'entreprise, une entreprise de qualité gère souvent son capital avec soin et réduit le risque de surendettement ou de surcapitalisation. Les facteurs utilisés dans cet étude sont :

- ROE : est un indicateur financier qui mesure la rentabilité d'une entreprise par rapport à ses capitaux propres.
- Payout : est un indicateur financier qui mesure la proportion des bénéfices nets d'une entreprise qui est distribuée sous forme de dividendes aux actionnaires.
- BPA : est un indicateur financier qui mesure la part du bénéfice net d'une entreprise attribuable à chaque action en circulation.

3.5 Volatilité

L'anomalie de la faible volatilité est une énigme empirique qui va à l'encontre des principes de base de la finance. Le modèle d'évaluation des actifs financiers (CAPM) et d'autres modèles de tarification des actifs affirment que le risque plus élevé devrait générer des rendements plus élevés, mais dans de nombreux marchés et sur de longues périodes, le contraire a été observé, avec des actifs moins risqués surpassant leurs homologues plus risqués.

Les facteurs utilisés dans cet étude sont :

- Vol : La volatilité annualisé sur 1 mois .
- Volatilité implicite : La volatilité issu de la modélisation du volatilité par le modèle GARCH ^e

4. Modélisation des facteurs

Une fois qu'on a la base de donnée de tout les actions, leurs rendements et leurs facteurs on va dans un premier lieu construire deux *labels* de plus qui sont les rendement forward de 1, 3 et 6 mois on va voir le rendement future de chaque action pendant le mois prochain qu'on va nommée **target_1m** et le rendement passé de 1, 3 et 6 mois noté **R1M**.

Avant de commencer la modélisation, on va donc étudier la corrélation linéaire entre nos variable qu'on va visualiser dans le graphe 1

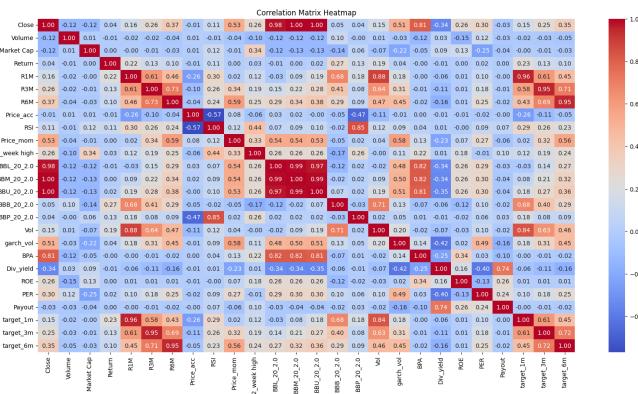


Figure 1. Matrice de corrélation entre nos variables

e. On va utiliser un GARCH(1,1) avec $\mu = 0$ car c'est le modèle plus optimale pour la majorité des actions au Maroc.

Comme on peut voir en générale la corrélation entre les variable n'est pas très forte avec l'exception de quelque variable qui présente de forte corrélation.

La démarche ici est d'éliminer les relations qui présente une corrélation de valeur absolue supérieur à 0.7, cependant vu notre utilisation de la régression élastique on va laisser ces variables.

4.1 Régression linéaire

Le premier modèle qu'on va utiliser est le simple modèle de la régression multiple linéaire.

On va diviser notre base de donnée en une base d'entraînement et de test avec comme date de séparation 2023-06-14 (2,5 ans d'entraînement 1 semestre de Test).

On applique alors la régression et on trouve les résultats suivants :

OLS Regression Results						
Dep. Variable:	y	R-squared:	0.917			
Model:	OLS	Adj. R-squared:	0.917			
Method:	Least Squares	F-statistic:	3239.			
Date:	mer., 08 janv. 2025	Prob (F-statistic):	0.00			
Time:	21:29:20	Log-Likelihood:	15545.			
No. Observations:	6782	AIC:	-3.104e+04			
Df Residuals:	6758	BIC:	-3.088e+04			
Df Model:	23					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.0073	0.004	1.971	0.049	3.86e-05	0.014
x1	-4.168e-09	2.41e-09	-1.731	0.083	-8.89e-09	5.51e-10
x2	1.978e-14	1.37e-14	1.440	0.150	-7.14e-15	4.67e-14
x3	0.0625	0.019	3.312	0.001	0.026	0.100
x4	0.8325	0.008	105.485	0.000	0.817	0.848
x5	-0.1604	0.007	-22.856	0.000	-0.174	-0.147
x6	-0.0021	0.001	-1.842	0.066	-0.004	0.000
x7	-0.1373	0.008	-16.336	0.000	-0.154	-0.121
x8	-0.0003	7.3e-05	-4.681	0.000	-0.000	-0.000
x9	-0.0002	0.000	-0.438	0.661	-0.001	0.001
x10	0.0001	2.56e-05	4.389	0.000	6.22e-05	0.000
x11	-6.587e-06	3.75e-06	-1.756	0.079	-1.39e-05	7.69e-07
x12	2.358e-08	2.98e-07	0.079	0.937	-5.61e-07	6.08e-07
x13	6.635e-06	3.65e-06	1.820	0.069	-5.11e-07	1.38e-05
x14	2.658e-05	6.4e-05	0.415	0.678	-9.88e-05	0.000
x15	0.0141	0.002	6.764	0.000	0.010	0.018
x16	1.0986	1.010	1.088	0.277	-0.880	3.078
x17	-0.0048	0.011	-0.427	0.669	-0.027	0.017
x18	-1.849e-05	1.76e-05	-1.052	0.293	-5.29e-05	1.6e-05
x19	0.0007	0.001	0.966	0.334	-0.001	0.002
x20	2.957e-05	4.83e-05	0.612	0.540	-6.51e-05	0.000
x21	9.581e-05	3.86e-05	2.481	0.013	2.01e-05	0.000
x22	-2.679e-05	3.43e-05	-0.782	0.434	-9.39e-05	4.04e-05
x23	0.1907	0.007	28.433	0.000	0.178	0.204
x24	0.0027	0.001	2.348	0.019	0.000	0.005
Omnibus:	359.708	Durbin-Watson:	2.083			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1160.707			
Skew:	0.202	Prob(JB):	9.03e-253			
Kurtosis:	4.986	Cond. No.	9.11e+22			

Figure 2. Résultat de la régression linéaire

Si on calcule le Hit ratio qui est la probabilité que la prédiction et le résultats de la base de test ont le même signe on trouve :

$$Hit = 0.8692$$

et le MSE affiche :

$$MSE = 0.0305$$

Cependant, on peut voir par ce graphe que 15 variables sur 24 ne sont pas statistiquement différentes de 0, une chose qui est

causer par la colinéarité qui existe entre les différentes variables. Une solution à ce problème de colinéarité est de pénaliser les variables qui présente ce phénomène et dans des cas extrêmes voir les éliminer. Pour faire ça on va utiliser la régression élastique.

4.2 Régression élastique

La régression élastique (Elastic Net Regression) est une technique de régularisation utilisée pour traiter des problèmes de régression avec des ensembles de données ayant un grand nombre de variables indépendantes, particulièrement dans des contextes où la multicollinéarité est présente. Elle combine les pénalités L1 du Lasso (Least Absolute Shrinkage and Selection Operator) et L2 de la régression Ridge, offrant ainsi à la fois la sélection de variables et la réduction des coefficients.

La fonction de coût de la régression élastique est exprimée comme suit :

$$\min_{\beta_0, \beta} \left(\frac{1}{2n} \sum_{i=1}^n (y_i - \beta_0 - X_i \beta)^2 + \alpha \|\beta_1\| + \lambda \|\beta_1\|^2 \right)$$

Où :

- y_i est la variable dépendante (la cible),
- X_i est le vecteur des variables explicatives,
- β_0 est l'ordonnée à l'origine (l'intercept),
- β est le vecteur des coefficients de régression,
- n est le nombre d'observations,
- $\|\beta_1\|$ est la norme L1 des coefficients (pénalité Lasso),
- $\|\beta_1\|^2$ est la norme L2 des coefficients (pénalité Ridge),
- α et λ sont des hyperparamètres qui contrôlent respectivement l'importance de la pénalité L1 et L2.

On continuant avec la même date de séparation pour la base d'entraînement et de test on va appliquer la régression élastique avec les paramètres $\alpha = 0.1$ et $\lambda = 0.1$ ces deux valeurs sont les résultats d'une cross-validation qui a été faite sur le modèle avec différentes valeurs de α et λ .

Avec cette regression on a une diminution du Hit Ratio qui devient :

$$Hit = 0.8070$$

Cette diminution cependant vient avec une augmentation significative du MSE qui devient :

$$MSE = 0.312$$

Ce modèle nous permet aussi une sélection des facteurs selon leurs significative et leurs colinéarité car dans le back test sur chaque période d'entraînement, le modèle va sélectionner les facteurs les plus intéressants dans cette période en annulant ceux qui ne présente plus une information de plus (colinéaire).

Dep. Variable:	y	R-squared:	0.628			
Model:	OLS	Adj. R-squared:	0.628			
Method:	Least Squares	F-statistic:	1040.			
Date:	dim., 12 janv. 2025	Prob (F-statistic):	0.00			
Time:	15:20:59	Log-Likelihood:	10468			
No. Observations:	6782	AIC:	-2.091e+04			
Df Residuals:	6770	BIC:	-2.083e+04			
Df Model:	11					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
const	-0.2552	0.004	-59.076	0.000	-0.264	-0.247
x1	-3.999e-09	5.07e-09	-0.789	0.430	-1.39e-08	5.94e-09
x2	-1.339e-13	2.52e-14	-5.312	0.000	-1.83e-13	-8.45e-14
x3	0.0050	5.77e-05	86.442	0.000	0.005	0.005
x4	0.0004	4.34e-05	9.994	0.000	0.000	0.001
x5	-5.083e-05	7.63e-06	-6.659	0.000	-6.58e-05	-3.59e-05
x6	-2.76e-06	2.96e-07	-9.321	0.000	-3.34e-06	-2.18e-06
x7	4.531e-05	7.27e-06	6.235	0.000	3.11e-05	5.96e-05
x8	0.0007	0.000	6.488	0.000	0.000	0.001
x9	6.285e-05	1.93e-05	3.260	0.001	2.51e-05	0.000
x10	0.0003	9.85e-05	3.498	0.000	0.000	0.001
x11	0.0004	4.97e-05	8.226	0.000	0.000	0.001
x12	-6.861e-06	2.86e-05	-0.240	0.810	-6.29e-05	4.92e-05
=====						
Omnibus:	1125.913	Durbin-Watson:	0.552			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	9394.433			
Skew:	0.552	Prob(JB):	0.00			
Kurtosis:	8.659	Cond. No.	9.17e+22			
=====						

Figure 3. Résultat de la régression élastique sur la totalité de la base d'entraînement

On peut voir que cette fois ci à l'exception de la seconde variable tous les variables sont statistiquement significative et que les hypothèses de la régression sont bien vérifiées. Comme on peut voir la régression a annulé 12 facteurs et même si le R^2 est inférieur à celui de la régression linéaire mais ceci vient de fait que la régression élastique va utiliser moins de facteurs.

Le code de la section 3 et 4 on été faite en Python dans le fichier **Factor_investing.py**.

5. Back testing

Pour le Backtesting on peut prendre des fenêtres glissantes ou à chaque fois on décide de notre base d'entraînement avec de tester et de calculer la performance, ou on peut à chaque étapes inclure la base d'entraînement précédente de tel sorte qu'à la fin on va entraîner notre modèle sur toute la base.

Dans notre étude on va choisir le premier choix, car le deuxième présente un biais pour les données récentes qui eux seront entraîner sur la totalité des données contrairement aux premières dates, de plus le débat sur l'utilisation d'un long historique des données ne permet pas la prévision de court ou future fluctuation, cependant cette deuxième approche présente l'avantage de permettre au modèle de s'entraîner sur différentes conditions de marché.

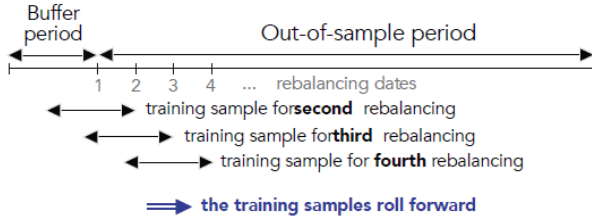


Figure 4. Backtesting selon la première approche

Un deuxième choix à prendre sont la fréquence de rééquilibrage et l'horizon sur lequel l'étiquette est calculée. Il n'est pas évident qu'ils doivent être égaux, mais leurs choix doivent être logiques. Il peut sembler pertinent de s'entraîner sur une étiquette à 12 mois (qui capture les tendances à plus long terme) et d'investir mensuellement ou trimestriellement. Cependant, il semble étrange de faire l'inverse, c'est-à-dire de s'entraîner sur les mouvements à court terme (mensuels) et d'investir sur un horizon long.

Puisque dans notre étude on va travailler avec le label de **target_1m** on va prendre comme horizon 1 mois et la fréquence de rééquilibrage 1 ans, et si on note par :

- Δ_h : la période entre deux rééquilibrage.
- Δ_s : la taille de la base d'entraînement.
- Δ_l : l'horizon du label

Alors la taille de la base d'entraînement doit être égale à $\Delta_s + \Delta_l$ et pour chaque moment t la base d'entraînement doit s'arrêter à $t - \Delta_l$ comme ça le label est calculé jusqu'à t . Une fois

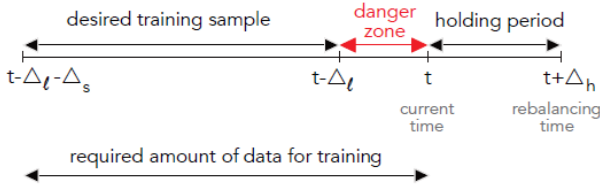


Figure 5. Visualization du choix de la taille de l'entraînement

ces deux choix sont pris on va procéder pour chaque base d'entraînement de calculer la performance d'un portefeuille équilibré (considéré comme référence ou Benchmark) en suite on va calculer les poids du portefeuille basé sur les facteurs en choisissant comme poids la proportion des éléments supérieure à la médiane des prédictions par le même modèle.

5.1 Turnover

La mise à jour de la composition du portefeuille n'est pas gratuite. En règle générale, le coût total d'un rééquilibrage à l'instant t est proportionnel à $C_t = \sum_{n=1}^N |\Delta w_{t,n}| c_{t,n}$ ou $\Delta w_{t,n}$ est le changement de position pour l'actif n et $c_{t,n}$ est la commission correspondante. Cette dernière quantité est souvent difficile à prédire, il est donc courant d'utiliser une proxy qui dépend, par exemple, de la capitalisation boursière (les grandes actions ayant des parts plus liquides et nécessitant ainsi des frais plus faibles) ou des spreads acheteur-vendeur (des spreads plus

petits signifient des frais plus faibles). Comme approximation de premier ordre, il est souvent utile de calculer le turnover moyen.

$$\text{Turnover} = \frac{1}{T-1} \sum_{t=2}^T \sum_{n=1}^N |w_{t,n} - w_{t-1,n}|$$

Où $w_{t,n}$ sont les poids désirés et $w_{t-1,n}$ sont les poids avant le rééquilibrage.

Dans la suite les coûts de transaction peuvent être approximer par un multiple du turnover fois un proxy des coûts (médiane des coûts dans la cross section des actions).

Vu le manque de ces informations on va se contenter de calculer uniquement le turnover.

5.2 Métriques de performance

1. **Rendement moyen** : C'est le rendement moyen de la stratégie.
2. **Volatilité** : C'est l'écart-type du rendement de la stratégie est il représente le risque.
3. **Ratio de Sharp** : Rendement ajusté au risque.
4. **Var_5** : Value at risk à 5%.
5. **Turnover**

5.3 Résultats

Les résultats de notre backtest sont les suivants : On peut

Table 2. Résultats du Backtesting

	Rendement	Volatilité	Sharp	Var 5	Turnover
Benchmark ^a	0.013223	0.094425	0.140034	-0.069087	0.000077
Elastique	0.056733	0.186608	0.304024	-0.043089	0.000417

a équilibrée

observer par ces résultats que notre stratégie performe mieux qu'un simple portefeuille équilibré, en effet :

- Le rendement moyen de notre stratégie est de 5% contrairement à celui de 1%.
- Notre stratégie est plus volatile avec une volatilité de 0.18 contre 0.013.
- Cependant le rendement ajusté au risque Sharp Ratio de notre stratégie est supérieure à celui du benchmark.
- Le turnover du portefeuille équilibré doit être nul, cependant ce n'est pas le cas car la division sous python donne différentes approximations pendant le backtesting, on effectue 1/20 est parfois calculé comme 0.055555555, 0.0536 ou 0.05 ce qui a causé le turnover à augmenter, cependant ceci n'affecte pas la comparaison et on peut voir que en moyenne notre stratégie ne nécessite pas beaucoup de changement avec un turnover moyen de 0.00417.
- Pour 95% des cas notre stratégie ne dépassera pas en terme de perte 6.19%.

Les résultats de ce backtest semblent réalistes même si le petit rendement, car ce backtest a été fait de telle sorte à éviter tout overfitting en éliminant la crossvalidation de la régression élastique

sur chaque base d'entraînement. Ou l'adoption de stratégie simpliste avec des seuils qui peuvent augmenter le overfitting.

6. Quelques considérations

Cette étude est loin d'être suffisante pour adopter la stratégie, en effet il existe encore plusieurs considérations à prendre et des choix à faire :

- Le manque des données fondamentaux et des facteurs ajoute une couche de complexité supérieure et par conséquent augmente la marge des erreurs.
- Le manque de fiabilité des données historique des taux sans risque peut être la cause principale de l'écartement du facteur du rendement excédentaire.
- Seul la régression linéaire, Ridge, Lasso et Elastique ont été utilisés, et ils sont tous des modèles linéaires, et par conséquent une comparaison avec un modèle non linéaire est nécessaire.
- On pourrait ajouter des filtres *Kalman* pour éliminer le bruit de nos facteurs.
- Des considérations de coût de transaction, de période de rééquilibrage et du label à calculer doivent être révisés.
- Optimisation des calculs qui prennent un grand temps ($\sim 1h$) même si la base est plus ou moins petite (11 602 rows x 30 colonnes).
- D'autres considérations doivent être prises par rapport au calcul pour éviter les problèmes de virgule, comme le cas du turnover du portefeuille équilibré.
- Il existe d'autres facteurs à explorer notamment les facteurs issus d'une analyse sentimentale du marché et les *data-driven factors* issus d'une analyse des composantes principales appliquée sur les rendements des actions choisies.

7. Conclusion

Pour conclure notre étude est un très bon début pour la construction d'un modèle multifactoriel afin de répliquer l'indice MASI 20.

Ce travail n'est pas encore achevé, il nécessite d'autres considérations qui peuvent être prêtes d'ici février.

Code et Data l'organisation du code est la suivante :

1. **Factors.py** : ce fichier calcule les facteurs et il donne comme résultats :
 - la base de données *Price.csv* : contient les prix quotidiens des 74 actions utilisées.
 - la base de données *Returns_fact.csv* : contient les rendements quotidiens des 74 actions utilisées.
 - la base de données *Factors.csv* : contient l'évolution quotidienne des facteurs calculés.
 - Figure *Facteurs.png* : Visualisation de l'évolution quotidienne des facteurs.
2. **Anomalies_facteurs.py** : ce fichier permet la sélection des facteurs et il donne :

- *regression_summaries_daily.txt* et *regression_summaries_month.txt* : des fichiers text contenant les résultats de la régression de chaque action sur les facteurs (quotidien et mensuel)
 - *Crossection_regression.txt* *Crossection_regression_month.txt* : Contient les résultats de la régression de la régression Fama-McBeth (quotidien et mensuel).
 - Figure *daily_factor_premia.png* : figure de l'évolution quotidienne des risk premia.
 - Afficher les p-values de chaque risk premia
 - Base de données *Ret_mon.csv* : qui contient le rendement mensuel de chaque action.
 - Base de données *Fact_mon.csv* : qui contient les facteurs mensuels.
 - *Factor_competition.txt* et *Factor_competition_monthly.txt* : contient la régression de chaque facteur sur les autres pour l'étude de la compétition entre ces facteurs (au niveau quotidien et mensuel).
3. **Factor_investing.py** : ce fichier contient le reste du calcul et la back testing et à part l'affichage des résultats il donne :
 - Figures *Boxplot.png* le box plot de quelques facteurs importants, *Corr_matrix.png* la matrice de corrélation

References

- Cochrane, John H. 2011. Presidential Address: Discount Rates. *Journal of Finance* 131. <https://doi.org/10.1111/j.1540-6261.2011.01671.x>.
- Eugene F. Fama, Kenneth R. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 32 (June): e004. <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>.
- . 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 32 (February): e004. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5).
- F., Sharp W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 32 (September): e004. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>.
- Fama, MacBeth. 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 32 (May).
- J., Mossin. 1966. Equilibrium in a capital asset market. : *Journal of the Econometric Society* 32 (October): e004. <https://doi.org/10.2307/1910098>.
- Narasimhan Jegadeesh, Sheridan Titman. 2011. Momentum. 32 (August): e004. <https://doi.org/10.2139/ssrn.1919226>.
- Ross, Stephen A. 1977. THE CAPITAL ASSET PRICING MODEL (CAPM), SHORT-SALE RESTRICTIONS AND RELATED ISSUES. *Journal of Finance* 32: e004. <https://doi.org/10.1111/j.1540-6261.1977.tb03251.x>.

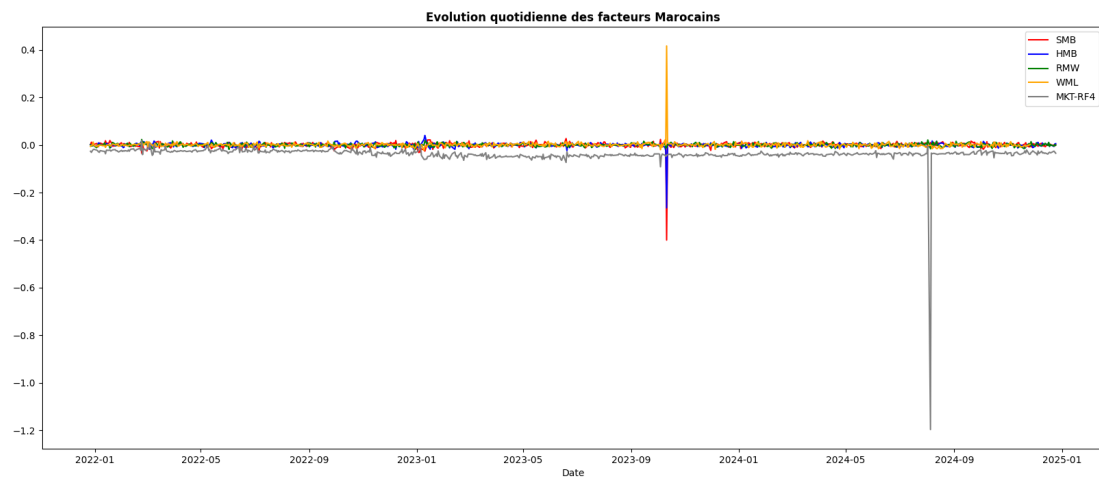


Figure 6. Evolution quotidienne des facteurs de marché

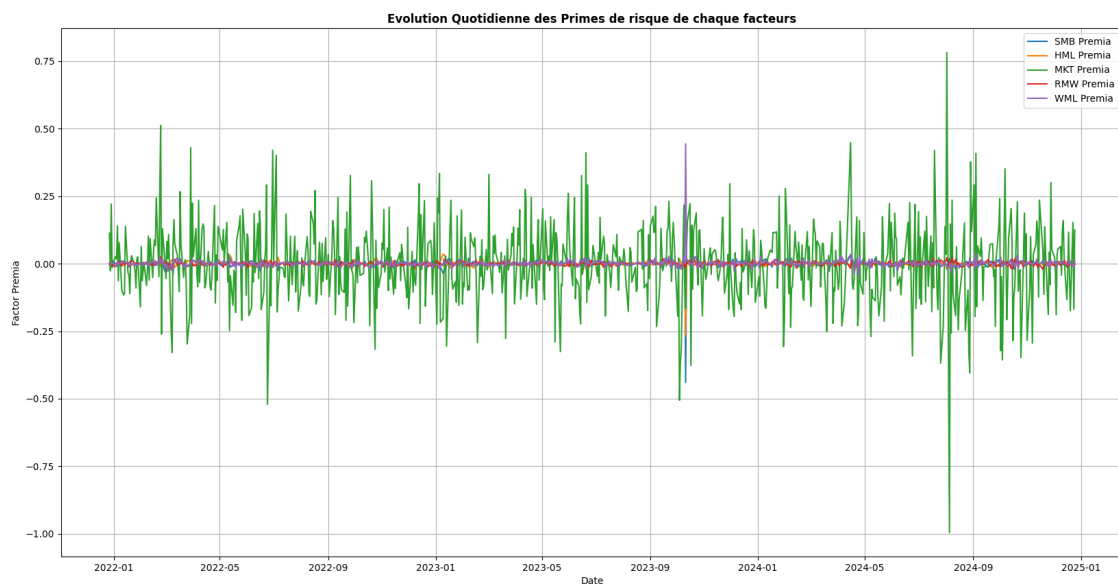


Figure 7. Evolution quotidienne des risk premia