

# PROJET DE GESTION QUANTITATIVE

Benoît Bordachar  
Paul de Surville  
Alexis Vicini

Date  
Soutenance orale  
le 4 *Janvier 2023*

Alpha forecasting in factor investing : discriminating between the informational  
content of firm characteristics

Lars Heinrich et Martin Zurek

# Table des matières

<b>1</b>	<b>Présentation de l'article</b>	<b>1</b>
1.1	Restitution dans la littérature . . . . .	1
1.2	Objectifs . . . . .	1
1.3	Méthodes d'alpha forecasting . . . . .	2
1.4	Résultats . . . . .	2
<b>2</b>	<b>Réplication du papier</b>	<b>3</b>
2.1	Traitement et construction de la base de données . . . . .	3
2.1.1	Création des facteurs . . . . .	3
2.1.2	Statistiques descriptives et inférence . . . . .	4
2.2	Simulation . . . . .	5
2.2.1	Simulation d'un nouvel univers . . . . .	5
2.2.2	Construction des portefeuilles . . . . .	5
2.2.3	Estimation des IC, alpha réalisé et IR . . . . .	6
2.3	Résultats du backtest . . . . .	7
2.3.1	Performance des portefeuilles multifactoriels . . . . .	8
2.3.2	Métrique des performances des portefeuilles mono et multifactoriels . . . . .	9
2.3.3	Quantiles portfolios . . . . .	10
<b>3</b>	<b>Extension</b>	<b>11</b>
3.1	Estimation avec changement de régime . . . . .	11
3.2	Simulation simplifiée dans 4 faits stylisés . . . . .	12
<b>4</b>	<b>Conclusion</b>	<b>13</b>
	<b>Références</b>	<b>14</b>

# 1 Présentation de l'article

## 1.1 Restitution dans la littérature

En finance, la prédiction de la performance future d'un fonds d'investissement est cruciale pour les investisseurs et les gestionnaires de portefeuilles. Cependant, trouver un moyen fiable et précis de prévoir les rendements peut être un défi. Si les actions constituent l'une des classes d'actif les plus échangées, une question persiste : qu'est-ce qui explique les rendements des actions et comment les prédire ?

En 1952, Harry Markowitz a exposé sa théorie financière de « diversification efficiente » au travers de la Théorie Moderne du Portefeuille, précédant le développement du Capital Asset Pricing Model (CAPM) dans les années 1960, reliant les rendements au risque en considérant qu'à l'équilibre, un portefeuille efficient est une combinaison d'un actif sans risque et d'un portefeuille optimal. Malgré son succès, le modèle présente des limites évidentes en raison de contraintes et biais, conduisant au constat que le modèle n'est que théorique et que l'explication des rendements ne se limite pas à la sensibilité au marché. Dans « Market truths : theory versus empirical simulations » (2006), Wojciechowski et Thompson fournissent des preuves solides contre l'hypothèse d'un marché efficient.

Par la suite, des chercheurs tels que Barr Rosenberg, Eugene Fama et Kenneth French ont étendu le CAPM dans les années 1990 pour inclure des facteurs systématiques importants dans l'explication des rendements en révélant que la sensibilité d'une action à la prime de valeur et à la prime de taille augmente également le rendement attendu. Jusqu'à présent concentré sur la capture du bêta du marché par le biais d'indices pondérés en fonction de la capitalisation boursière, la notion d'investissement passif a considérablement évolué, notamment avec Gander et al (2012) ou Lo (2016), pour passer d'un type d'investissement dans un indice à la notion d'investissement facile et à moindre coût.

Considéré comme le premier à véritablement proposer de l'investissement factoriel, Russell propose en 1996 les indices « Style Boxes », tirés des boîtes de style de Morningstar. Il cherche à expliquer les différences de style entre les gestionnaires d'actions en classant leurs portefeuilles en neuf catégories. Russell contribue ainsi à un changement nécessaire de l'industrie financière en instaurant les indices de style comme des véhicules d'investissement capables de fournir des surperformances potentielles en restant bon marchés, simples, clairs et transparents.

## 1.2 Objectifs

Au travers de cet article, Heinrich et Zurek étudient les moyens d'améliorer les stratégies d'investissement factoriel en comparant une approche basée sur les prévisions d'alphas définis par Grinold et Kahn dans « Active Portfolio Management » (2000) [1] avec une approche d'allocation naïve. Le but de Heinrich et Zurek est ainsi d'examiner comment l'utilisation de la valeur informationnelle de différentes

caractéristiques d'entreprise peut être appliquée dans les stratégies d'investissement factoriel.

### 1.3 Méthodes d'alpha forecasting

Dans l'article «Active Portfolio Management» de Grinold et Kahn (2000), les auteurs définissent l'alpha de deux manières :

- Optique prospective (ex-ante), l'alpha est une prévision du rendement résiduel.
- Optique rétroactive (ex-post), l'alpha est la moyenne des rendements résiduels réalisés.

Ils énoncent également que la prévision prend les signaux bruts des rendements des actifs et les transforment en prévisions affinées, permettant de relier un score standardisé de moyenne nulle et d'écart-type unitaire à une prévision de rendement résiduel sous forme d'alpha, établissant ainsi :  $\alpha = \sigma * IC * signaux$ , avec  $\sigma$  la volatilité résiduelle, IC le coefficient d'information correspondant à la corrélation entre les signaux et les rendements.

Heinrich et Zurek s'en inspirent en définissant l'alpha prévisionnel d'un titre comme étant la différence entre les rendements attendus conditionnels et inconditionnels. Ils proposent l'expression du alpha suivante pour l'actif i au temps t :

$$\alpha_{GK,i,t} = E(R_{i,t}^* | g_{i,t-1}) - E(R_{i,t}^*)$$

avec :  $R_{i,t}^*$  le rendement de l'actif i au temps t et  $g_{i,t-1}$  le vecteur de dimension (M x 1) des signaux .

Nous garderons dans la suite la méthode de calcul des poids utilisée par L. Heinrich et M. Zurek, tirée des travaux de MacKinlay et Pástor (2000) [2]. Les poids du portefeuille optimal orthogonal sont proportionnels aux alphas des entreprises tels que  $a_A = \frac{\alpha}{1'\alpha}$ . Les stocks avec des alphas prédits positifs sont assignés au portefeuille long  $a_{long} = \frac{\hat{\alpha}^+}{1'\hat{\alpha}^+}$  avec  $\hat{\alpha}^+ = \max(\hat{\alpha}, 0)$  et ceux avec des alphas prédits négatifs au portefeuille short  $a_{short} = \frac{\hat{\alpha}^-}{1'\hat{\alpha}^-}$  avec  $\hat{\alpha}^- = \max(-\hat{\alpha}, 0)$

### 1.4 Résultats

Heinrich et Zurek présentent que dans les données du Eurostoxx 600, les portefeuilles multifactoriels atteignent les meilleures performances, l'approche de Grinold et Kahn (GK) offrant un rendement supérieur dans le portefeuille multifactoriel long-only à celui de l'approche naïve (NZ) et bien supérieur au rendement de l'indice de référence. Sur le portefeuille multifactoriel long-short, l'approche de Grinold et Kahn surperforme largement le marché et légèrement l'approche naïve à la fin de la période. Cependant, comme la plupart des coefficients d'information sont positifs mais que la dispersion de ces coefficients d'information est plutôt faible, il n'y a pas de différence significative entre l'approche GK et NZ. Dans les portefeuilles monofactoriels, les résultats proposés par les auteurs placent les portefeuilles avec un petit nombre de coefficients d'information moyens significatifs, aucune des deux approches n'est en mesure de surperformer de manière significative le portefeuille de référence.

En conclusion, Heinrich et Zurek présentent au travers de leurs résultats qu'il est possible de surperformer l'indice de référence grâce à la construction de portefeuilles multifactoriels basés sur une approche de Grinold et Kahn ou naïve z-score, à condition que les coefficients d'information soient statistiquement significatifs et que les signaux soient proches de leur vraie valeur. Les auteurs ajoutent que l'estimation des coefficients d'information constitue le principal déterminant de la performance, et justifient donc l'intérêt de réduire l'erreur d'estimation.

## 2 Réplication du papier

### 2.1 Traitement et construction de la base de données

Nous avons investigué nos résultats pour un jeu de données constitué sur le Stoxx Europe 600 (SXXP). Cet indice est habituellement utilisé comme proxy du marché actions européen. En guise de taux d'intérêt sans risque, nous avons utilisé le taux 3-mois des obligations gouvernementales européennes.

#### 2.1.1 Création des facteurs

Les prix sont ajustés des opérations sur titres (OST). Les facteurs que nous avons pris sont les mêmes que ceux calculés par L. Heinrich et M. Zurek, qui correspondent à des facteurs proposés par le MSCI [4]. Néanmoins pour étendre la période du backtest, nous avons réduit les fenêtres de calcul des facteurs concernés à 3 ans au lieu de 5. De plus, pour les facteurs momentum, nous avons pris la volatilité des rendements mensuels plutôt qu'hebdomadaires, de par la limite accordée par Bloomberg dans la récupération de données historiques. Les données fondamentales ainsi que les prix sont exprimés en euro.

Pour éviter la création d'un biais du survivant dans notre backtest, nous avons collecté les données en utilisant les compositions historiques de l'indice. Nous avons récupéré les données en utilisant l'API Bloomberg. Le format des requêtes passées est le suivant :

```
= @BQL.Query("
    get(dropna(px_last(fill=PREV, currency=EUR, ca_adj=FULL)))
    for(filter(members('SXXP Index Index',
        dates=range(start=2012-01-01, end=2023-01-30, frq=M))))
    preferences(ANCHORRELDATESBY=BROADCAST)
", "showheaders=false")
```

Pour s'assurer que les prévisions ne sont pas faussées par des valeurs aberrantes, les caractéristiques des entreprises sont winsorisées au-dessus du 95e percentile et au-dessous du 5e percentile.

Ensuite, pour annuler le biais *look-ahead*, nous avons téléchargé les dates de publications pour ensuite remplacer les données au moment où elles étaient réellement disponibles.

### 2.1.2 Statistiques descriptives et inférence

Dans le tableau 1, on indique les moyennes des séries temporelles de 116 estimations transversales de coefficients d'information. Pour quantifier les incertitudes dans les inférences, nous calculons les erreurs standard (SE) bootstrap et les intervalles de confiance à 95% sur 100 000 rééchantillons. L'écart-type de l'échantillon des IC moyens, sur l'ensemble des bootstraps, est utilisé comme estimation de l'erreur-type. Pour obtenir les limites inférieures et supérieures des intervalles de confiance pour les IC moyens, les percentiles 2,5% et 97,5% des moyennes des échantillons bootstrap sont calculés.

Factor	Characteristic	Mean	SE	0.025-Qt.	0.975-Qt.
Growth	EPSG1Y	-0.087	0.624	-1.018	1.429
	EPSG3Y	-0.241	0.411	-0.909	0.703
	IGR	1.514	0.961	0.168	3.936
	EPSGT	0.885	0.796	-0.280	2.838
	SPSGT	1.014	0.967	-0.223	3.576
Value	DivYld	0.153	0.874	-1.403	2.028
	EtP	0.769	0.661	-0.435	2.149
	BtP	-1.770	1.459	-5.282	0.450
	CFOtEV	-0.158	1.029	-1.891	2.135
Quality	RoE	1.796	1.080	0.170	4.404
	DtE	1.420	0.866	-0.511	2.879
	EVar	1.010	0.679	-0.306	2.345
Momentum	Pmom6M	1.625	1.433	-1.032	4.592
	Pmom12M	2.610	1.502	0.452	6.338

**TABEAU 1** – Moyenne des IC et résultats du bootstrap

La majorité des signaux ont un IC moyen positif. De plus, les caractéristiques de certains facteurs présentent des IC moyens très différents. Cependant, à la différence des résultats trouvés dans le papier, il n'y a que le Pmom12M qui soit vraiment significatif.

La figure 1 montre une carte thermique des moyennes des séries chronologiques des corrélations croisées entre les caractéristiques des entreprises pour chaque échantillon de données. On observe qu'un grand nombre de coefficients de corrélation sont inférieurs à 20%, la plupart des caractéristiques ne sont pas fortement corrélées. Cela indique que la prise en compte de plusieurs caractéristiques d'entreprises bénéficie des effets de diversification. Les résultats soulignent que la prise en compte de la matrice de corrélation des signaux peut améliorer la performance, mais la principale source de contribution à la performance semble se trouver dans la prise en compte des IC.

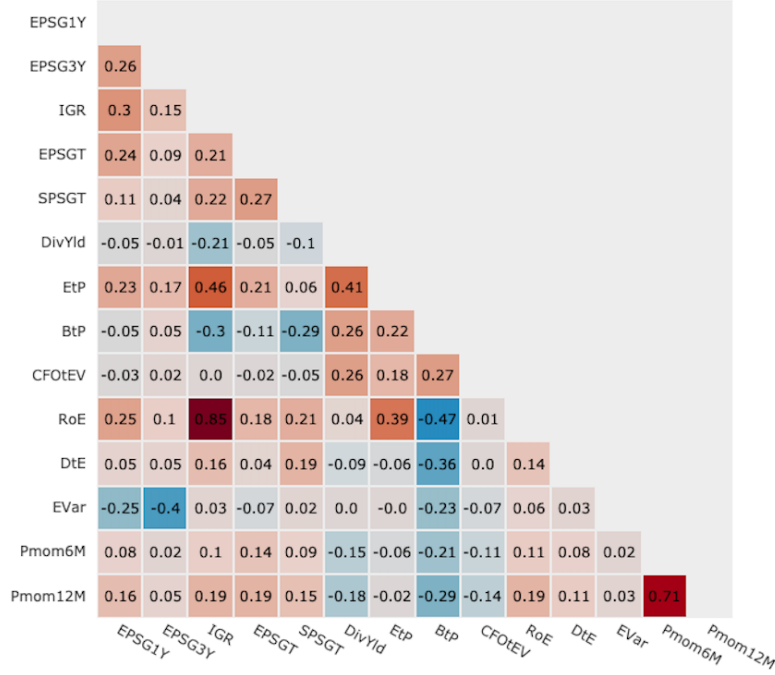


FIGURE 1 – Correlation moyenne entre les caractéristiques des entreprises

## 2.2 Simulation

### 2.2.1 Simulation d'un nouvel univers

Dans cette partie, nous allons simuler un univers de rendements d'actions, benchmark et Z-score afin d'analyser dans un environnement "robuste", sans facteur de risque manquant qui pourrait influencer les conclusions de nos backtests. Les paramètres du modèle sont obtenus à partir de nos observations empiriques. Les données sont générées à partir du modèle suivant :

$$R_{i,t}^* = \alpha_{i,t} + \beta_i R_{B,t}^* + \varepsilon_{i,t} \quad (1)$$

où :  $\alpha_{i,t} = \sigma_{\sigma_{i,t}} k_{CS,t} C_{CS,t}^{-1} z_{i,t-1}$  avec  $z_{i,t-1} \sim N(0, C_{CS})$

$\varepsilon_t \sim N(0, \psi)$

$R_{B,t}^* \sim N(\mu_B, \sigma_B)$

Une fois notre simulation générée, nous allons procéder à la construction de nos portefeuilles. Afin d'avoir un échantillon important, nous allons réaliser nos estimations sur 1000 simulations.

### 2.2.2 Construction des portefeuilles

Pour chaque simulation, nous allons calculer à chaque période d'observations nos estimations de  $\alpha_{GK}$  et  $\alpha_{NZ}$  ainsi que les  $\hat{IC}$  et les  $\hat{C}$ , puis les poids dans les portefeuilles de ces stratégies respectives.

Pour notre estimation, nous utilisons les moyennes sur le temps avec une période roulante de 60 mois. Nos estimations des  $\alpha_{GK}$  et  $\alpha_{NZ}$  sont calculés à partir de l'équation suivante tirées de la méthode d'alpha forecasting :

$$\hat{\alpha}_{GK,i,t} = \hat{k} \hat{C}^{-1} z_{i,t-1} \quad (2)$$

où :  $\hat{k} = (\hat{IC}_1 \dots \hat{IC}_M)$  avec  $\hat{IC}_m = \frac{1}{59} \sum_{t=1}^{59} \text{Corr}[\theta_t, z_{m,t-1}]$  et  $\hat{C} = \frac{1}{60} \sum_{t=1}^{60} C_t$

$$\hat{\alpha}_{NZ,i,t} = \sum_{m=1}^M z_{i,m,t-1} \quad (3)$$

Nous obtenons donc un jeu de donnée similaire statistiquement aux observations. La moyenne des corrélations entre les caractéristiques des entreprises est bien similaire à celle donnée dans le papier. Une fois nos  $\alpha_{GK}$  et  $\alpha_{NZ}$  estimés, on calcule les poids de chaque actif proportionnellement à leur estimation de  $\alpha$  comme défini précédemment. Ce modèle de construction de portefeuille sera le même dans notre backtest.

### 2.2.3 Estimation des IC, alpha réalisé et IR

Ensuite, pour mesurer la capacité de nos  $\hat{IC}$  à estimer les vrais IC de nos observations, nous calculons le biais et la précision pour chaque simulation. Le biais est l'écart moyen des estimations par rapport au vrai coefficient d'information, i.e  $(\hat{IC}_m - IC_m)$ . La précision est calculée comme l'erreur quadratique moyenne (la racine carrée de l'écart quadratique moyen), i.e  $\left(\hat{IC}_m - IC\right)^{20.5}$ . Comme dans le papier, nous obtenons une précision de nos estimations assez faibles. On évalue alors le nombre de fois où le signe du vrai coefficient d'information et celui de notre estimation sont les mêmes, c'est le DHR ("direction hit ratio"). On calcule aussi le pourcentage de coefficient d'information observés et estimés positifs. On obtient sur nos 1000 simulations les résultats présentés dans le Tableau 2.

On remarque que nos estimations sont relativement proches des coefficients d'information observés. L'ordre de grandeur de nos biais est similaire à celui obtenu dans le papier : de l'ordre du millièrme. Concernant notre précision, on constate que nos estimations sont légèrement moins précises, notre ordre de grandeur est du centième, contre plusieurs centièmes pour le papier. Enfin, les pourcentages de coefficient d'information observés et estimés positifs sont relativement proches dans nos simulations à ceux du papier. Seul les direction hit ratio sont plus élevés dans nos estimations. On remarque que lorsque le coefficient d'information d'un signal est élevé, son DHR est aussi plus élevé. Ce résultat est intéressant dans notre comparaison des méthodes GK et NZ. En effet, même si la DHR moyenne d'un portefeuille GK est inférieure à celle d'un NZ, l'approche GK fournira une meilleure prévision globale en pondérant davantage les signaux qui ont une meilleure probabilité de prendre la bonne direction.

Pour évaluer la performance des portefeuilles GK et NZ, on calcule leurs alpha réalisés et leurs ratios d'information à partir des pondérations estimées et des vrais alpha issus de la simulation à partir des formules ci-dessous. La performance obtenue représente donc bien la "vraie" performance et non



	IC obs.	IC est.	Biais	Précision	IC obs. (>0)	IC est. (>0)	DHR
EP SG1Y	-0.009	-0.003	-0.012	0.015	49.4 %	56.6%	38.2%
EP SG3Y	-0.004	-0.003	-0.001	0.007	44.7%	37.7%	66.5%
IGR	0.013	0.019	-0.006	0.015	54.1%	81.9%	89.0%
EP SGT	0.005	0.009	-0.004	0.011	55.3%	77.1%	78.6%
SP SGT	-0.001	0.016	-0.017	0.023	52.9%	79.9%	11.3%
DivYield	0.003	0.003	0.000	0.012	45.9%	55.8%	57.9%
EtP	0.006	0.010	-0.004	0.011	54.1%	74.8%	83.2%
BtP	0.001	-0.017	0.018	0.027	50.6%	27.5%	19.5%
CF OtEV	0.005	0.000	0.005	0.015	48.2%	49.4%	49.8%
RoE	0.005	0.016	-0.011	0.019	50.6%	77.0%	84.8%
DtE	0.011	0.013	-0.002	0.016	51.8%	72.3%	80.3%
EVar	0.008	0.008	0.000	0.009	54.1%	72.4%	81.3%
Pmon12M	0.012	0.030	-0.018	0.026	50.6%	86.2%	94.9%
Pmon6M	0.014	0.019	-0.005	0.019	54.1%	76.8%	85.4%

**TABLEAU 2** – Résultats des estimations des coefficients d’information pour 1000 simulations

une performance estimée. On obtient les résultats suivants sur nos 1000 simulations présentés dans le Tableau 3 :

Alpha réalisé :  $\alpha_{A,GK,t} = a'_{A,GK,t-1}\alpha_t$  et  $\alpha_{A,NZ,t} = a'_{A,NZ,t-1}\alpha_t$

Information réalisé :  $IR_{A,GK,t} = \frac{\alpha_{A,GK,t}}{\sqrt{a'_{A,GK,t-1} \psi a_{A,GK,t-1}}}$  et  $IR_{A,NZ,t} = \frac{\alpha_{A,NZ,t}}{\sqrt{a'_{A,NZ,t-1} \psi a_{A,NZ,t-1}}}$

Dans un premier temps, on remarque que les valeurs des alpha moyens et des coefficients obtenus sont très élevées. Cela s’explique puisque nous travaillons dans un univers où il n’y a pas de facteur de risque manquant et où les variations des rendements reposent uniquement sur les hypothèses du modèle. On obtient des valeurs d’alpha réalisé cohérentes avec celles obtenues dans le papier pour le STOXX 600. Notre portefeuille multifactoriel superforme bien les mono-factoriels et la méthode GK se révèle bien plus efficace que la méthode NZ même si les deux méthodes sont capables de générer de l’alpha positif. Par ailleurs, on constate que les facteurs comprenant des caractéristiques fondamentales avec un coefficient d’information élevé (comme le momentum dans nos simulations) performant mieux. On note par ailleurs que nos ratios d’information sont aussi plus élevés que ceux obtenus dans le papier. Cela peut s’expliquer par le fait que dans le papier, les auteurs utilisent une estimation réduite de la matrice de corrélation  $\psi$ . Ils indiquent que leurs ratios d’information sont sous-estimés par rapport aux ratios d’information réels. On considère que donc que cela explique nos différences.

### 2.3 Résultats du backtest

Le cadre du backtest est choisi pour ressembler au comportement d’investissement réaliste d’un investisseur institutionnel, dont l’objectif principal est de surpasser un portefeuille de référence sous-jacent pondéré par les capitalisations. Cela inclut la prise en compte des contraintes de long-only, des coûts et fréquences de rééquilibrage couramment utilisées et des ensembles de données représentatifs.

			Mean	SE	0.025-Qt.	0.975-Qt.
Factor	Méthod					
Multifactor	$\overline{\alpha_A}$	NZ	3.34	1.77	-0.53	6.68
		GK	8.38	2.00	5.35	12.98
	$\overline{IR_A}$	NZ	4.32	2.27	-0.07	8.33
		GK	10.84	2.61	6.82	16.55
Growth	$\overline{\alpha_A}$	NZ	2.11	2.00	-2.14	5.73
		GK	4.27	1.80	1.00	7.86
	$\overline{IR_A}$	NZ	2.72	2.60	-2.64	7.48
		GK	5.49	2.32	1.23	9.76
Value	$\overline{\alpha_A}$	NZ	-0.17	2.25	-4.87	3.72
		GK	5.01	1.89	2.04	8.94
	$\overline{IR_A}$	NZ	-0.25	2.97	-6.23	4.94
		GK	6.52	2.49	2.63	11.37
Quality	$\overline{\alpha_A}$	NZ	2.71	2.95	-2.64	8.43
		GK	4.27	2.29	0.65	9.15
	$\overline{IR_A}$	NZ	3.46	3.81	-3.65	10.88
		GK	5.49	2.92	0.85	11.99
Momentum	$\overline{\alpha_A}$	NZ	3.08	2.30	-1.23	7.25
		GK	4.89	1.83	1.73	8.77
	$\overline{IR_A}$	NZ	4.03	2.94	-1.96	9.5
		GK	6.84	2.37	2.17	11.45

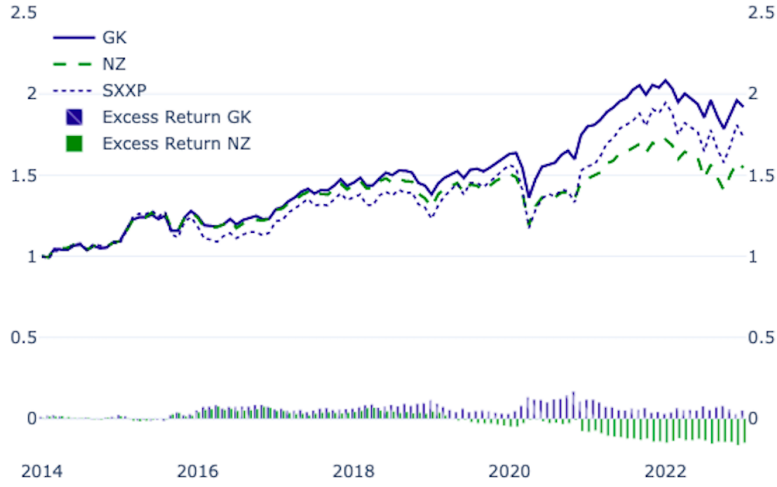
**TABLEAU 3** – Résultats des simulations d’alpha (en pourcentage) et d’IR des portefeuilles long-short GK et NZ

Nous supposons que les signaux sont disponibles à la fin de chaque mois. Ainsi, le portefeuille est rééquilibré au début du mois suivant. Pour rester aussi proche que possible d’un portefeuille investissable, nous prenons en compte les coûts de transaction causée par le rééquilibrage du portefeuille. Pour quantifier le volume pertinent en termes de coûts, le taux de rotation du portefeuille ( $PTR_t = \sum_{n=1}^N |a_{i,t-1} - a_{i,t}|$ ) est utilisé similairement au papier initial. Le PTR détermine le pourcentage du portefeuille qui entraîne des coûts de transaction. Dans le backtest, nous gardons le montant des coûts à dix points de base par volume négocié. Cette valeur comprend la commission de négociation et une taxe de négociation. Les coûts calculés sont délimités le jour où ils surviennent. Par conséquent, ces coûts entraînent une réduction directe du rendement le jour du rééquilibrage.

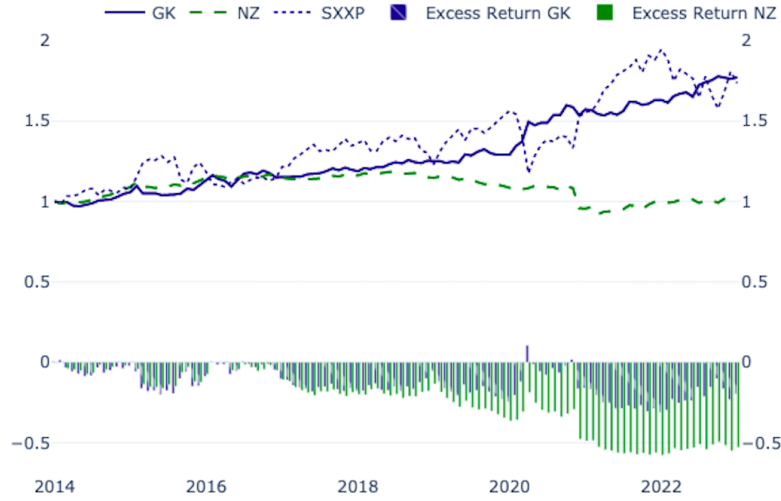
Conformément à notre étude de simulation, nous utilisons une fenêtre glissante de 3 ans, ce qui correspond à 36 observations mensuelles. L’intervalle d’évaluation hors échantillon commence en 2014. Les titres sont pondérés proportionnellement à leurs prévisions d’alpha (1.3), et les prévisions sont effectuées comme dans les formules (2) et (3) mais sur une période de 3 ans.

### 2.3.1 Performance des portefeuilles multifactoriels

Nous analysons d’abord la performance de l’allocation de titres des portefeuilles en utilisant tous les composants de l’indice sont sélectionnés. On observe que le portefeuille GK long-only surperforme



**FIGURE 2** – Performance cumulée des portefeuilles multifacteurs long-only



**FIGURE 3** – Performance cumulée des portefeuilles multifacteurs long-short

le Europe Stoxx 600 quand le naive-zscore sous-performance sur toute la période. Néanmoins, le portefeuille long-short GK peine face à l'indice malgré une bonne performance pendant la crise du covid (voir les figures 2 et 3).

### 2.3.2 Métrique des performances des portefeuilles mono et multifactoriels

Pour examiner la performance des portefeuilles, on a calculé des statistiques (alphas, ratios d'information et de Sharpe) face à l'indice pondéré par les capitalisations. Pour examiner leur performance par rapport aux portefeuilles de référence associés, pondérés en fonction de la capitalisation, les paramètres alpha annualisés, les ratios d'information et les ratios de Sharpe pour chaque stratégie factorielle sont présentés dans le tableau 4. Le calcul des statistiques suit la méthode décrite par L. Heinrich et M. Zurek.

Les trois mesures de performance conduisent à des résultats cohérents dans la plupart des cas

Factor	Method	Alpha (%)	Information Ratio	Sharpe Ratio
Multif.	Long-Only NZ	0.05	0.017	0.511
	Long-Short NZ	0.89	0.155	0.169
	Long-Only GK	2.84	1.111	0.769
	Long-Short GK	8.07	1.432	1.125
Growth	Long-Only NZ	-0.58	-0.172	0.452
	Long-Short NZ	-1.34	-0.205	-0.014
	Long-Only GK	2.08	0.727	0.690
	Long-Short GK	4.03	0.968	1.152
Value	Long-Only NZ	-0.72	-0.136	0.444
	Long-Short NZ	-1.06	-0.137	0.081
	Long-Only GK	0.84	0.224	0.559
	Long-Short GK	4.29	0.538	0.371
Quality	Long-Only NZ	1.27	0.546	0.668
	Long-Short NZ	2.68	0.460	0.122
	Long-Only GK	0.84	0.308	0.601
	Long-Short GK	2.91	0.444	0.271
Momentum	Long-Only NZ	2.08	0.419	0.627
	Long-Short NZ	4.53	0.434	0.187
	Long-Only GK	1.81	0.397	0.611
	Long-Short GK	4.94	0.521	0.318

**TABLEAU 4** – Résultats du backtest

lorsque l'on compare les approches GK et NZ. Une exception peut être trouvée dans les portefeuilles LS momentum où une forte volatilité conduit à un sharpe ratio plus faible, bien que l'alpha soit élevé.

Dans la plupart des cas, la performance multifactorielle GK semble être bénéfique par rapport aux portefeuilles monofactoriels. Pour les portefeuilles multifactoriels LS, on peut noter que les approches GK et NZ ont toutes les deux pu obtenir des paramètres alpha significatifs dans presque tous les échantillons.

### 2.3.3 Quantiles portfolios

En outre, nous analysons les effets de sélection pour le portefeuille multifactoriel en séparant les composants de l'indice en portefeuilles quantiles.

En théorie et en pratique, il y a une tendance à construire des sous-portefeuilles où le portefeuille long contient le quantile supérieur et le portefeuille court contient le quantile inférieur des titres. Pour étudier l'impact de cette approche de sélection, nous séparons le montant total des titres de l'indice en fonction de leurs prévisions d'alpha en cinq portefeuilles quantiles. Pour démontrer que les alphas des portefeuilles diminuent dans les portefeuilles du quantile inférieur, nous construisons des portefeuilles long-only dans lesquels les alphas prévus doivent être transformés pour ne recevoir que des valeurs

Method	Quantile	Alpha (%)	Information ratio
NZ	Q5	-2.52	-0.410
	Q4	0.96	0.343
	Q3	0.36	0.105
	Q2	-1.44	-0.475
	Q1	-0.96	-0.291
GK	Q5	-0.12	-0.023
	Q4	2.04	0.701
	Q3	-0.72	-0.268
	Q2	-0.72	-0.277
	Q1	-5.04	-1.245

**TABLEAU 5** – Alpha et ratio d’information ratio pour les quantiles portfolios

alpha positives :

$$\alpha\text{-score}_i = \begin{cases} 1 + \hat{\alpha}_i & \text{if } \hat{\alpha}_i \geq 0 \\ (1 - \hat{\alpha}_i)^{-1} & \text{otherwise} \end{cases} \quad (4)$$

Les portefeuilles extérieurs, ainsi que les portefeuilles GK des deuxième et quatrième quantiles, affichent une légère surperformance, cela indique que la méthode GK est plus appropriée pour sélectionner les titres dans les deuxième et quatrième quantiles que l’approche NZ, ce qui serait bénéfique pour les portefeuilles avec un nombre élevé de composants.

### 3 Extension

#### 3.1 Estimation avec changement de régime

D’après les résultats trouvés par L. Heinrich et M. Zurek, l’approche GK est moins performante dans les modèles factoriels à faible dispersion des coefficients d’information. La dispersion des IC n’étant pas constante dans le temps, cela impliquerait qu’il y ait des périodes où il faudrait dévier d’une approche équipondérée et d’autres s’en contenter.

Dès lors, on a expérimenté une stratégie de trading basée sur cette idée. Pour estimer la dispersion des coefficients d’information on utilise l’écart-type. Si bien que notre alpha prédit en  $t$  pour  $t+1$  est égal à :

$$\hat{\alpha}_t = \begin{cases} \hat{\alpha}_{GK,t} & \text{if } \sigma_{IC} \geq \lambda \\ \hat{\alpha}_{NZ,t} & \text{otherwise} \end{cases} \quad (5)$$

avec  $\hat{\alpha}_{GK,t} = \hat{k} \hat{C}^{-1} z_{t-1}$  et  $\alpha_{NZ,t} = \sum_{m=1}^M z_{m,t-1}$

Pour estimer le paramètre  $\lambda$ , nous avons utilisé une optimisation séquentielle à l’aide d’arbres

de décision entre 2014 et 2018. Nous avons testé la méthode de calcul du  $\lambda$  (valeur fixe ou moyenne glissante), sa valeur dans le cas d'une valeur fixe et la taille de la fenêtre dans le cas d'une moyenne mobile exponentielle. Dans notre cas, il est ressorti que le recours à une moyenne glissante de 23 périodes maximisait notre gain. Nous avons ensuite testé la stratégie entre 2019 et aujourd'hui, qui ne réussit pas à surperformer ni l'indice ni le GK. Il semble donc que la prédiction d'un alpha avec la méthode GK seulement au moment de forte dispersion des IC n'est pas bénéfique.

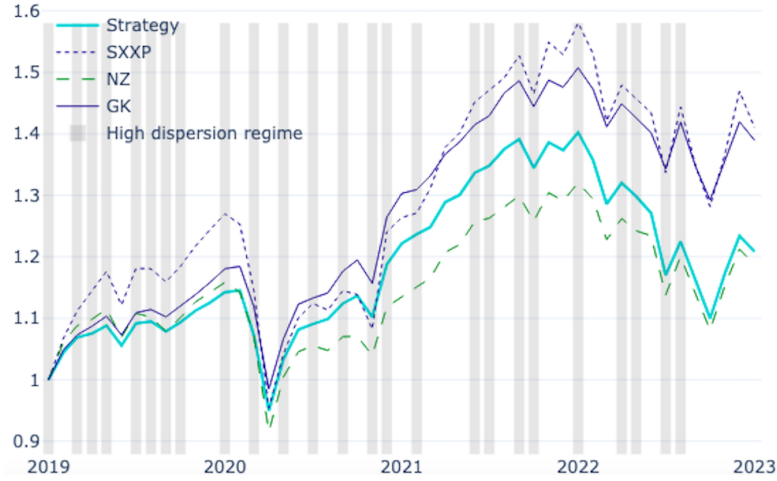


FIGURE 4 – Backtest stratégie de changement de méthode

### 3.2 Simulation simplifiée dans 4 faits stylisés

Dans cette partie, nous allons étudier les performances de 3 approches différentes d'allocation d'actifs : Top-Down (TD), Bottom-Up GK (BU GK) et Bottom-Up NZ (BU NZ). L'approche Top-Down consiste à construire des portefeuilles single-factor sur chaque signal, puis à les regrouper dans un portefeuille multifactoriel. A l'inverse, dans une approche Bottom-Up, on regroupe directement l'ensemble des signaux dans un unique portefeuille multifactoriel. Les approches Bottom-Up ont donc l'avantage de prendre en considération les interactions entre les signaux. La méthode naïve (BU NZ) permet d'ajouter ou d'annuler les signaux entre eux et la méthode GK poursuit cette intégration des effets d'interaction en les prenant aussi en compte dans l'estimation de la prévision de alpha. De plus, comme nous l'avons vu précédemment, la robustesse de la méthode GK par rapport à la méthode NZ dépend de la dispersion des coefficients d'information. L'approche de GK permet d'obtenir des meilleures performances que l'approche de NZ lorsque les coefficients d'information sont dispersés.

Par conséquent, nous avons décidé, pour illustrer ces différences, de calculer les alphas réalisés de ces 3 approches dans une simulation simplifiée avec seulement deux signaux  $x$  et  $y$  dans 4 faits stylisés :

- Faible corrélation des signaux et faible dispersion des coefficients d'information (FC et FD).
- Faible corrélation des signaux et haute dispersion des coefficients d'information (FC et HD).
- Haute corrélation des signaux et faible dispersion des coefficients d'information (HC et FD).
- Haute corrélation des signaux et haute dispersion des coefficients d'information (HC et HD).

Scénario	TD (%)	NZ BU (%)	GK BU (%)
Faible corrélation & Faible dispersion d'IC	2.79	3.76	3.75
Forte corrélation & Faible dispersion d'IC	1.79	2.42	2.75
Faible corrélation & Forte dispersion d'IC	2.83	2.91	2.89
Forte corrélation & Forte dispersion d'IC	1.87	1.91	4.56

**TABLEAU 6** – Résultats des alpha réalisés des 3 approches sur une simulation simplifiée

Pour x et y faiblement corrélés on pose :  $Corr[x, y] = 0.1$

Pour x et y fortement corrélés on pose :  $Corr[x, y] = 0.9$

Une faible dispersion des IC signifie que x et y ont un IC de 0.01

Une forte dispersion des IC signifie que x possède un IC de 0.03 et y de 0.01

Ces résultats sont cohérents avec les définitions statistiques des constructions de chaque méthode étudiée. Ils sont aussi confirmés par ceux obtenus par Lars Zurek et Lars Heinrich dans leur papier [3] dans le paragraphe *Simplified simulation exemple* où ils réalisent la même simulation pour comparer les trois approches d'allocation d'actifs.

## 4 Conclusion

Les résultats de performance issus de simulations et de backtests hors échantillon indiquent que les portefeuilles factoriels, qui prennent en compte de multiples caractéristiques d'entreprises, sont capables de surperformer de manière significative leurs portefeuilles de référence. Par rapport à la sélection de titres avec des z-scores naïfs, nous avons étudié les gains de performance causés par l'utilisation d'un modèle de prévision linéaire de l'alpha, permettant de prendre en compte les coefficients d'information entre les différentes caractéristiques des entreprises.

En extension du papier, nous avons exploré une stratégie combinant les deux approches en estimant un alpha via un modèle simple à changement de régime entre les deux méthodes de calcul. Il apparaît que cela ne génère pas de surperformance, signifiant alors le peu d'intérêt à dévier du portefeuille GK dans le cas du STX. Nous avons aussi apporté une analyse supplémentaire dans laquelle nous avons mis en lumière les différences de performance dans 4 faits stylisés bien particuliers. Ils ont une nouvelle fois illustré l'importance de la corrélation des signaux et de la dispersion des coefficients d'informations dans la prévision d'alphas.

## Références

- [1] GRINOLD ET KAHN. *Bottom-up versus top-down factor investing : an alpha forecasting perspective*. <https://www.studocu.com/row/document/institute-of-finance-and-economics/finance/active-portfolio-management/4766023>. 2000.
- [2] MACKINLAY, A.C., PÁSTOR. « Asset pricing models : implications for expected returns and portfolio selection ». In : *Rev. Financ. Stud.* (2000), p. 883-916.
- [3] MARTIN ZUREK ET LARS HEINRICH. *Bottom-up versus top-down factor investing : an alpha forecasting perspective*. <https://link.springer.com/article/10.1057/s41260-020-00188-9.pdf>. 2016.
- [4] MSCI. *Factors Definition*. [https://www.msci.com/eqb/methodology/meth\\_docs/MSCI\\_Aug16\\_GIMIVGMethod.pdf](https://www.msci.com/eqb/methodology/meth_docs/MSCI_Aug16_GIMIVGMethod.pdf).