

Alpha forecasting in factor investing

Discriminating the informational content of firm characteristics

B. Bordachar, P. de Surville et A. Vicini

Master 272

Université Dauphine, Paris, France

3 février 2023



Table des matières

- 1** Présentation de l'article
 - Restitution dans la littérature
 - Objectifs et méthodes d'alpha forecasting
 - Résultats théoriques
- 2** Réplication du papier
 - Traitement et construction de la base de données
 - Simulation
 - Backtest
- 3** Extensions
 - Stratégie avec changement de régime
 - Simulation simplifiée dans 4 faits stylisés
- 4** Conclusion

Plan

1 Présentation de l'article

- Restitution dans la littérature
- Objectifs et méthodes d'alpha forecasting
- Résultats théoriques

2 Réplication du papier

3 Extensions

4 Conclusion

Plan

- 1** Présentation de l'article
 - Restitution dans la littérature
 - Objectifs et méthodes d'alpha forecasting
 - Résultats théoriques

Restitution dans la littérature

Question : *"Qu'est-ce qui explique les rendements des actions et comment les prédire ?"*

■ **Théorie Moderne du Portefeuille de Markowitz (1952) :**

- Introduction du concept de diversification.
- Suppose que les investisseurs achèteront toujours l'actif qui présente un rendement optimal par rapport à son niveau de risque.

■ **CAPM de Treynor, Sharpe, Lintner et Mossin (1960s) :**

- Lien entre rendement et risque systématique.
- $E(R_{actif}) = R_F + \beta_{actif} \cdot [E(R_M) - R_F]$ où $[E(R_M) - R_F]$ représente la prime de risque du marché.
- Limites du modèle : contraintes, biais et preuves à l'encontre de l'efficacité du marché (Wojciechowski et Thompson, 2006).

■ **Extension du CAPM par E. Fama et K. French (1990s) :**

- Inclusion de facteurs systématiques.
- Sensibilité d'une action à un facteur peut augmenter le rendement attendu.

Restitution dans la littérature

■ Investissement passif :

- Initialement concentré sur la capture du bêta du marché par le biais d'indices pondérés en fonction de la capitalisation boursière.
- Evolution avec Gander et al (2012) puis Lo (2016) pour finalement s'associer à un investissement facile et à moindre coût.

■ Investissement factoriel :

- **Morningstar** (1990) : cherche à expliquer les différences de style entre les gestionnaires d'actions en classant leurs portefeuille en neuf catégories.
- **Russel** (2000s) : premier à véritablement proposer de l'investissement factoriel
- Indices de style : bon marchés, simples, clairs et transparents avec surperformance potentielle.
- Popularisation de l'investissement factoriel mais questionnement sur sa rentabilité ("Is Smart Beta really Smart" (2014) de Malkiel, "How Can Smart Beta Go Horribly Wrong?" (2016) de Arnott, Beck, Kalesnik)

Plan

1 Présentation de l'article

- Restitution dans la littérature
- Objectifs et méthodes d'alpha forecasting
- Résultats théoriques

Objectifs et méthodes d'alpha forecasting

- **Objectifs** : Améliorer les stratégies d'investissement factoriel grâce à l'utilisation de la valeur informationnelle de caractéristiques d'entreprises.
- **Méthodes d'alpha forecasting** :
 - Comparer les prévisions d'alphas selon deux approches : Grinold et Kahn (2000) avec une approche d'allocation naïve.
 - **Grinold et Kahn (2000)** :
 - . relie la volatilité résiduelle, les signaux bruts des rendements des actifs et le coefficient d'information pour prédire le rendement résiduel : $\alpha = \sigma * IC * Signaux$
 - . définissent l'alpha prévisionnel d'un titre comme la différence entre les rendements attendus conditionnels et inconditionnels :
 $\alpha_{GK,i,t} = E(R_{i,t}^* | g_{i,t-1}) - E(R_{i,t}^*)$ avec $R_{i,t}^*$ le rendement de l'actif i au temps t et g le vecteur de dimension $(M \times 1)$ des signaux
- **MacKinlay et Pàstor (2000)** : Poids du portefeuille optimal orthogonal proportionnels aux alphas des entreprises

Plan

1 Présentation de l'article

- Restitution dans la littérature
- Objectifs et méthodes d'alpha forecasting
- Résultats théoriques

Résultats théoriques

■ Portefeuilles multifactoriels :

- Atteignent les meilleures performances, le rendement de l'approche de Grinold et Kahn (GK) étant supérieur à l'approche naïve z-score (NZ).
- **Long-only (LO)** : Le rendement du portefeuille multifactoriel LO de GK surperforme largement le marché et surperforme le portefeuille LO NZ.
- **Long-short (LS)** : Le rendement du portefeuille multifactoriel LS de GK surperforme largement le marché et surpasse en fin de période le portefeuille LS NZ. CI positifs mais dispersion faible : pas de différence significative entre l'approche GK et NZ.

■ Portefeuilles monofactoriels : aucune des deux approches n'est en mesure de surperformer de manière significative le portefeuille de référence.

■ Conclusion : Possibilité de surperformer l'indice de référence avec des signaux et CI significatifs et proches de leur vraie valeur, l'estimation des CI étant le principal déterminant de la performance.

Plan

1 Présentation de l'article

2 Réplication du papier

- Traitement et construction de la base de données
- Simulation
- Backtest

3 Extensions

4 Conclusion

Plan

2 Réplication du papier

- Traitement et construction de la base de données
 - Création et traitement des données
 - Statistiques descriptives et inférences
- Simulation
 - Simulation d'un nouvel univers
 - Construction des portefeuilles
 - Estimation des IC, alphas réalisés et IR
- Backtest
 - Performance des portefeuilles multifactoriels
 - Performance des portefeuilles multi et mono-factoriels
 - Portefeuilles quantiles

Création et traitement des données

- Echantillon de données de l'Europe Stoxx 600 (SXXP)
 - ▶ fréquence mensuelle de janvier 2008 à janvier 2023
 - ▶ téléchargement via l'API Bloomberg
- Taux obligataire 3-mois gouvernement euro comme taux sans risque
- Sélection des caractéristiques basée sur les facteurs du MSCI¹
 - ▶ Réduction à 3 ans des fenêtres glissantes (EPSGT, SPSGT et EVar)
 - ▶ Volatilité des rendements mensuels pour le momentum

Correction des biais

Utilisation de la composition historique de l'indice

Remplacement des données fondamentales au moment de leur publication

1. MSCI Global Investable Market Value and Growth Index Methodology (2017)

Création et traitement des données

Factor	Characteristic	Definition
Growth	Earnings Growth 1 Year (EPSG1Y)	$EPSG1Y_t = \text{YoY EPS Growth}_t$
	Earnings Growth 3 Years Average (EPSG3Y)	$EPSG3Y_t = \frac{1}{3} \sum_{t=1}^3 \text{YoY EPS Growth}_t$
	Internal Growth Rate (IGR)	$IGR_t = \frac{\text{TTM EPS}_t}{\text{BVPS}_t} - \frac{\text{TTM DPS}_t}{\text{BVPS}_t}$
	Earnings Growth Trend (EPSGT)	$EPSGT_t = \frac{\beta_{EPS}}{\frac{1}{3} \sum_{t=1}^3 \text{EPS}_t }$
	Sales Growth Trend (SPSGT)	$SPSGT_t = \frac{\beta_{SPS}}{\frac{1}{3} \sum_{t=1}^3 \text{SPS}_t }$
Value	Dividend Yield (DivYld)	$\text{DivYld}_t = \frac{\text{TTM DPS}_t}{P_t}$
	Earnings to Price (EtP)	$\text{EtP}_t = \frac{\text{TTM EPS}_t}{P_t}$
	Book to Price (BtP)	$\text{BtP}_t = \frac{\text{BVPS}_t}{P_t}$
	Cash Flow to Enterprise Value (CFOtEV)	$\text{CFOtEV}_t = \frac{\text{TTM CFO}_t}{\text{EV}_t}$
Quality	Return on Equity (RoE)	$\text{RoE}_t = \frac{\text{TTM EPS}_t}{\text{BVPS}_t}$
	Debt to Equity (DtE)	$\text{DtEY}_t = \frac{\text{TD}_t}{\text{BV}_t}$
	Earnings Variability (EVar)	$\text{EVar}_t = \sqrt{\frac{1}{3} \sum_{t=1}^3 (\text{YoY EPS Growth}_t - \overline{\text{YoY EPS Growth}}_t)^2}$
Momentum	6 Months Price Momentum (Pmom6M)	$\text{n-Months Price Momentum}_t = \frac{(\frac{P_{t-1}}{P_{t-n-1}} - 1) - R_{f,t}}{\sigma_t}$
	12 Months Price Momentum (Pmom12M)	

Création et traitement des données

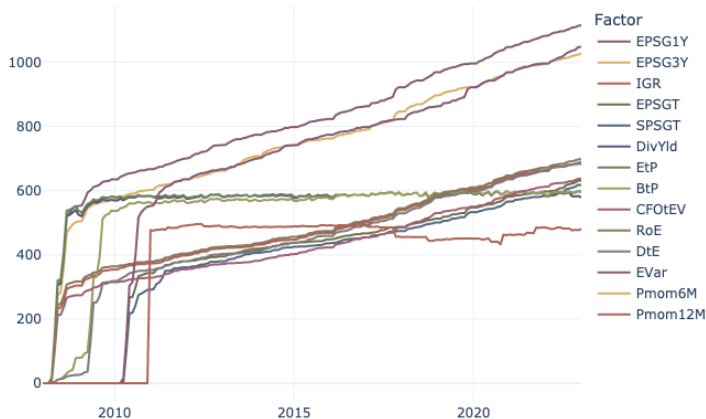


Figure – Nombre de données par date et facteur

Statistiques descriptives et inférences

Factor	Characteristic	Mean	SE	0.025-Qt.	0.975-Qt.
Growth	EP SG1Y	-0.087	0.624	-1.018	1.429
	EP SG3Y	-0.241	0.411	-0.909	0.703
	IGR	1.514	0.961	0.168	3.936
	EP SGT	0.885	0.796	-0.280	2.838
	SP SGT	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
	CF OtEV	-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

Table – Moyenne des IC et résultats du bootstrap

Statistiques descriptives et inférences

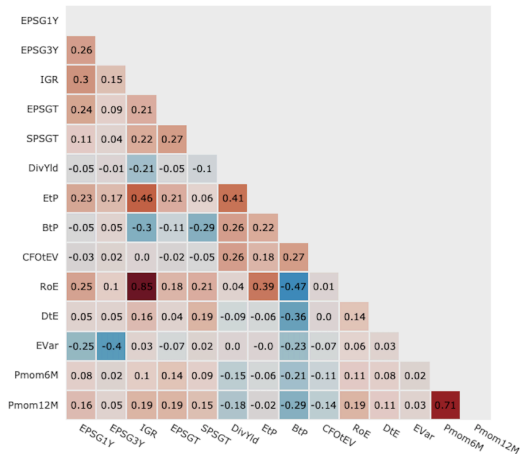


Figure – Correlation moyenne entre les caractéristiques des entreprises

Plan

2 Réplication du papier

- Traitement et construction de la base de données
 - Création et traitement des données
 - Statistiques descriptives et inférences
- Simulation
 - Simulation d'un nouvel univers
 - Construction des portefeuilles
 - Estimation des IC, alphas réalisés et IR
- Backtest
 - Performance des portefeuilles multifactoriels
 - Performance des portefeuilles multi et mono-factoriels
 - Portefeuilles quantiles

Simulation d'un nouvel univers

Modèle de génération des simulations à facteur linéaire conditionnel :

- $R_{i,t}^* = \alpha_{i,t} + \beta_i R_{B,t}^* + \varepsilon_{i,t}$
 - ▶ $\alpha_{i,t} = \sigma_{\sigma_{i,t}} k_{CS,t} C_{CS,t}^{-1} z_{i,t-1}$
 - ▶ $\varepsilon_t \sim N(0, \psi)$
 - ▶ $z_{i,t-1} \sim N(0, C_{CS})$
 - ▶ $R_{B,t}^* \sim N(\mu_B, \sigma_B)$

avec :

- k_{CS} (vecteur des IC) et C_{CS} (matrice de corrélation des signaux) sont tirés aléatoirement parmi l'ensemble des observations.
- ψ est la matrice de covariance des résidus de la regression des rendements observés contre les rendements du benchmark.
- μ_B et σ_B sont les maximum likelihood estimates des rendements observés du marché.

Simulation d'un nouvel univers

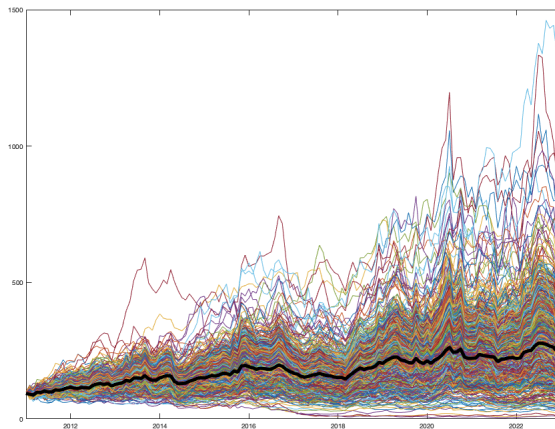


Figure – Exemple de performance d'univers simulé (avec benchmark en noir)

Simulation d'un nouvel univers

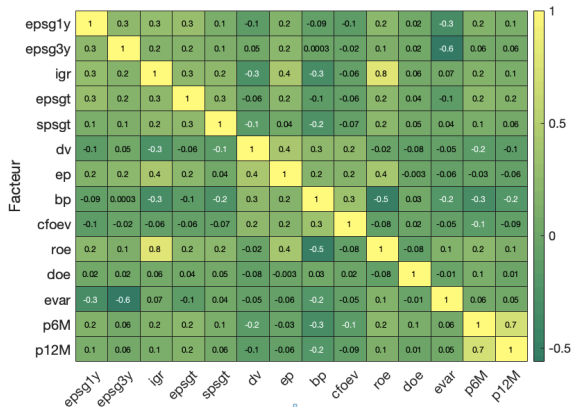


Figure – Corrélation moyenne des Z-scores simulés

Construction des portefeuilles

Méthode d'estimation des α_{GK} et α_{NZ} :

- $\hat{\alpha}_{GK,i,t} = \hat{k} \hat{C}^{-1} z_{i,t-1}$
 - ▶ $\hat{k} = (\text{IC}_1 \dots \text{IC}_M)$
 - ▶ $\text{IC}_m = \frac{1}{59} \sum_{t=1}^{59} \text{Corr} [\theta_t, z_{m,t-1}]$
 - ▶ $\hat{C} = \frac{1}{60} \sum_{t=1}^{60} C_t$
- ▶ avec θ_t rendements résiduels de la regression temporelle des rendements des actions contre les rendements du benchmark
- $\hat{\alpha}_{NZ,i,t} = \sum_{m=1}^M z_{i,m,t-1}$

Méthode de pondération des actifs proportionnelle aux alphas :

- $a_A = \frac{\alpha}{\mathbb{1}'\hat{\alpha}}$
 - ▶ Portefeuille short : $a_{short} = \frac{\hat{\alpha}^-}{\mathbb{1}'\hat{\alpha}^-}$
avec $\hat{\alpha}^- = \max(-\hat{\alpha}, 0)$
 - ▶ Portefeuille long : $a_{long} = \frac{\hat{\alpha}^+}{\mathbb{1}'\hat{\alpha}^+}$
avec $\hat{\alpha}^+ = \max(\hat{\alpha}, 0)$

Estimation des IC, alphas réalisés et IR

Mesure de la précision des \hat{IC}

- Biais : $\overline{(IC_m - \hat{IC}_m)}$
- Précision : $\overline{(IC_m - \hat{IC}_m)^2}^{0.5}$

Calcul des alphas réalisés à partir des estimations :

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

Calcul des coefficients d'informations :

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

Estimation des IC, alphas réalisés et IR

	IC obs.	IC est.	Biais	Précision	IC obs. (>0)	IC est. (>0)	DHR
EPSPG1Y	-0.009	-0.003	-0.012	0.015	49.4 %	56.6%	38.2%
EPSPG3Y	-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%
EPSPGT	0.005	0.009	-0.004	0.011	55.3%	77.1%	78.6%
SPSPGT	-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%
CFOtEV	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%

Table – Résultats des estimations des coefficients d'information pour 1000 simulations

Estimation des IC, alphas réalisés et IR

			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

Table – Résultats des simulations d'alpha (en pourcentage) et d'IR des portefeuilles long-short GK et NZ

Plan

2 Réplication du papier

- Traitement et construction de la base de données
 - Création et traitement des données
 - Statistiques descriptives et inférences
- Simulation
 - Simulation d'un nouvel univers
 - Construction des portefeuilles
 - Estimation des IC, alphas réalisés et IR
- Backtest
 - Performance des portefeuilles multifactoriels
 - Performance des portefeuilles multi et mono-factoriels
 - Portefeuilles quantiles

Performance des portefeuilles multifactoriels

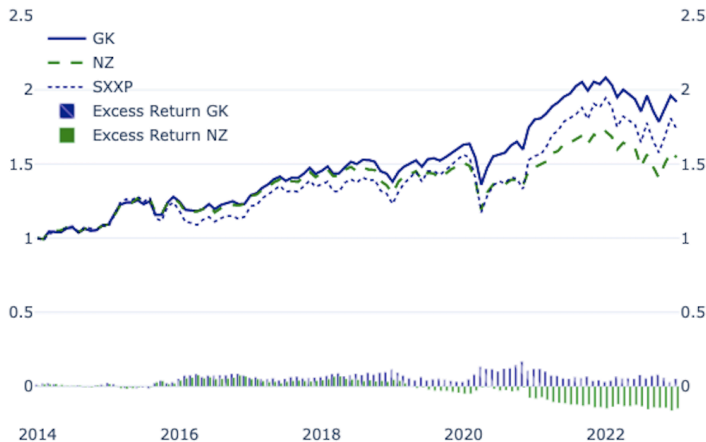


Figure – Performance cumulée des portefeuilles multifacteurs long-only

Performance des portefeuilles multifactoriels

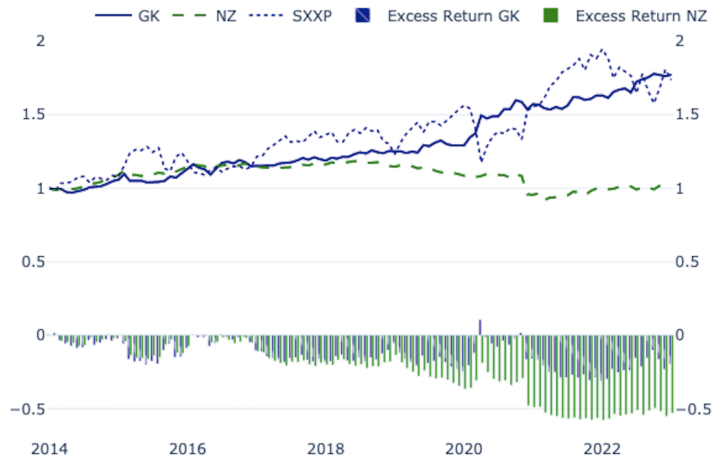


Figure – Performance cumulée des portefeuilles multifacteurs long-short

Performance des portefeuilles multi et mono-factoriels

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

Table – Résultats du backtest

Performance des portefeuilles multi et mono-factoriels

		Alpha (%)
Multif.	Long-Only GK	2.94
	Long-Short GK	6.42
Growth	Long-Only GK	2.96
	Long-Short GK	4.49
Value	Long-Only GK	2.08
	Long-Short GK	2.99
Quality	Long-Only GK	-0.49
	Long-Short GK	1.10
Momentum	Long-Only GK	0.04
	Long-Short GK	1.55

Table – Alphas du portefeuille GK contre le portefeuille NZ

Portefeuilles quantiles

- Construction de portefeuilles long-only dans lesquels les alphas prévus sont transformés pour ne recevoir que des valeurs alpha positives :

$$\alpha\text{-score}_i = \begin{cases} 1 + \hat{\alpha}_i & \text{si } \hat{\alpha}_i \geq 0 \\ (1 - \hat{\alpha}_i)^{-1} & \text{sinon} \end{cases}$$

- Alpha et ratio d'information pour les quantiles portfolios :

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

Plan

1 Présentation de l'article

2 Réplication du papier

3 Extensions

- Stratégie avec changement de régime
- Simulation simplifiée dans 4 faits stylisés

4 Conclusion

Plan

3 Extensions

- Stratégie avec changement de régime
- Simulation simplifiée dans 4 faits stylisés

Variation de la dispersion des IC

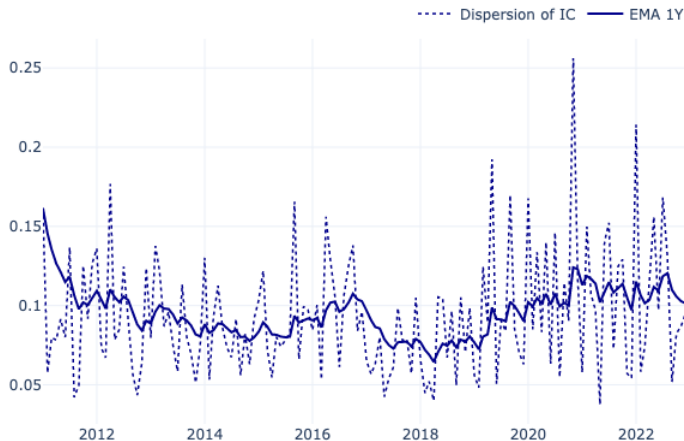


Figure – Dispersion des coefficients d'information

Modèle de la stratégie

- Objectif : Améliorer la performance en tenant compte de la variation de la dispersion des IC dans le temps
- Stratégie avec changement du modèle de prédiction

$$\hat{\alpha}_t = \begin{cases} \hat{\alpha}_{GK,t} & \text{si } \sigma_{IC} \geq \lambda \\ \hat{\alpha}_{NZ,t} & \text{sinon} \end{cases}$$

avec :

$$\begin{aligned} \hat{\alpha}_{GK,t} &= \hat{k} \hat{C}^{-1} z_{t-1} \\ \alpha_{NZ,t} &= \sum_{m=1}^M z_{m,t-1} \end{aligned}$$

- Estimation du λ par optimisation séquentielle à l'aide d'arbres de décision entre 2014 et 2018

Performance de la stratégie

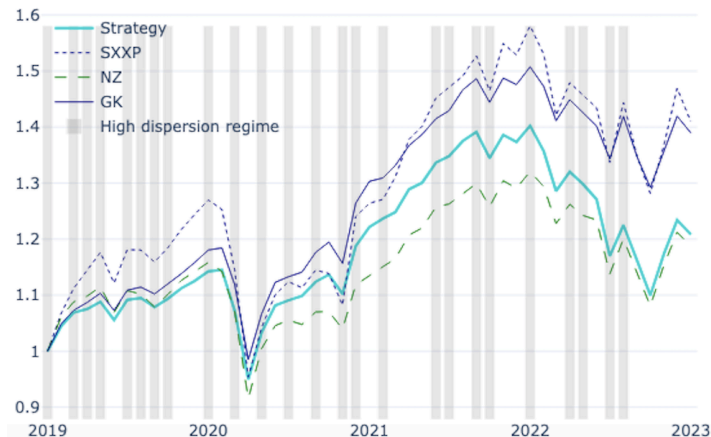


Figure – Backtest stratégie de changement de méthode

Plan

3 Extensions

- Stratégie avec changement de régime
- Simulation simplifiée dans 4 faits stylisés

Simulation simplifiée dans 4 faits stylisés

But : Mettre en lumière les différences de performances des stratégies Top-Down, Bottom-Up NZ et Bottom-Up GK.

- Définition des différentes configurations d'IC et de matrice de corrélation pour deux signaux x et y :

- ▶ Faible corrélation :

$$\text{Corr}[x, y] = 0.1.$$

- ▶ Forte corrélation :

$$\text{Corr}[x, y] = 0.9.$$

- ▶ Forte dispersion d'IC :

$$IC(x) = 0.03 \text{ et } IC(y) = 0.01.$$

- ▶ Faible dispersion d'IC :

$$IC(x) = 0.03 \text{ et } IC(y) = 0.01.$$

- Simulation des performances et calcul des alphas réalisés des stratégies dans 4 faits stylisés :

- ▶ Faible corrélation & Faible dispersion d'IC.

- ▶ Forte corrélation & Faible dispersion d'IC.

- ▶ Faible corrélation & Forte dispersion d'IC.

- ▶ Forte corrélation & Forte dispersion d'IC.

Simulation simplifiée dans 4 faits stylisés

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

Table – Résultats des alpha réalisés des 3 approches sur une simulation simplifiée²

2. Méthode tirée de *Bottom-up versus top-down factor investing : an alpha forecasting perspective* de Zurek & Heinrich (2020)

Plan

- 1 Présentation de l'article
- 2 Réplication du papier
- 3 Extensions
- 4 Conclusion**