

Gestion Quantitative - Réplication de l'étude de Fays, Hübner et Lambert 2018

Robin Gallego, Leslie Selamme, Marie-Lidwine Tardif

April 2020

Abstract

Ce document présente les travaux de Fays, Hübner et Lambert (2018), analyse leur étude et propose une réplication de leur modèle. Nous retrouvons, comme les auteurs de l'étude répliquée, qu'un tri séquentiel avec post-conditionnement sur la variable de contrôle et avec des seuils sur l'échantillon entier apporte plus d'avantages que les tris traditionnels. Les différentes conclusions restent vérifiées sur les six dernières années, que Fays, Hübner et Lambert n'avaient pas prises en compte.

Contents

1	Introduction	3
2	Présentation du papier de recherche de Fays, Hübner et Lambert	3
2.1	Les objectifs	3
2.2	Revue de littérature	3
2.3	Rappel de leurs résultats	3
3	Réplication de l'étude de Fays, Hübner et Lambert	5
3.1	Réplication de la partie 3 : Construction de portefeuille en utilisant un tri indépendant ou séquentiel	5
3.2	Réplication de la Partie 4 : Différents choix pour trier les actions	5
3.3	Réplication de la partie 5 : Diversification du portefeuille et les effets sur les biais sur la construction des facteurs	10
3.4	Réplication de la partie 6 : Tri triple pour créer 3x3x3 portefeuilles	13
4	Extension	16
4.1	Distributions des entreprises parmi les caractéristiques des 2x3 portefeuilles	16
4.2	Régression des 2x3 et 3x3 portefeuilles sur les facteurs de Fama French sur la période récente	17
5	Conclusion	19

1 Introduction

“Factoring Characteristics into Returns: A Clinical Approach to Fama-French Portfolio Decomposition” a été publié en juin 2018 par l’EDHEC Business School par Boris Fays doctorant à l’université de Liège, Georges Hübner, professeur de finance à l’université de Liège et à l’université de Maastricht et Marie Lambert, chercheuse à l’EDHEC-Risk Institute.

Dans ce papier, nous tâchons de répliquer leur étude afin de retrouver leurs conclusions. Dans un deuxième temps, nous étendons les différentes analyses aux dernières données disponibles.

2 Présentation du papier de recherche de Fays, Hübner et Lambert

2.1 Les objectifs

De nombreux modèles expliquent le rendement des actifs à travers divers facteurs, notamment le modèle à 3 facteurs de Fama-French.

Le modèle de Fama French à 3 facteurs a été développé en 1992 par Eugene Fama et Kenneth French. Ce modèle permet d’élargir le CAPM en ajoutant au facteur de risque de marché, des facteurs tel que la taille des entreprise et la valeur de leurs actions. Le modèle prend donc en compte que les small cap côté en bourse ont une capitalisation boursière faible et des rendements majoritairement plus importants. Cela va permettre de construire des portefeuilles délimités avec les points de ruptures de la Bourse de New York.¹

Cet article démontre l’importance de la méthode de sélection des actifs pour la création de ses facteurs. Les données mensuelles sélectionnés s’étalent sur une période de 51 ans (1963-2014) et concernent tous les actions du NYSE, de l’AMEX et du NASDAQ issus de la fusion des bases de données CRSP et COMPUSTAT. Ces données ont été sélectionnées pour pouvoir comparer les résultats à ceux du modèle de Fama French.

2.2 Revue de littérature

Fays, Hübner et Lambert appuient principalement leurs travaux sur les recherches des auteurs suivants : Fama and French (1993), Carhart (1997). En effet, ils se servent du portefeuille à trois facteurs de Fama and French (1993) tout au long de leur étude, comme comparaison avec les différentes méthodes de construction de facteurs qu’ils présentent. Carhart (1997) a pour objectif principal de montrer que la persistance dans les rendements des fonds communs de placement est presque entièrement expliquée par des facteurs communs. Fays, Hübner et Lambert se servent principalement de son facteur momentum et vérifient que leurs résultats restent robustes avec cet ajout.

En 1997, une autre étude que celle de Carhart apporte un résultat très important pour le papier de recherche que nous discutons ici. En effet, Daniel and Titman (1997) montrent que ce sont plutôt les caractéristiques à partir des quelles les facteurs ont été créés, et non pas la structure de covariance des rendements comme Fama and French (1993) le pensaient, qui sont responsables du phénomène de mouvements liés entre les actions dites *value* et *growth*.

Fays et al. (2018) illustrent l’importance de l’enjeu auquel ce papier répond, avec les travaux de Kogan and Tian (2015) qui sont parmi les premiers à montrer que la performance des modèles à plusieurs facteurs est dépendante de la méthode de construction de ces facteurs.

2.3 Rappel de leurs résultats

Une première étude consiste à appliquer une méthode de tri séquentiel (trier par un facteur à la fois) là où Fama & French construisent leurs portefeuilles en réalisant un tri indépendant (ils déterminent une médiane et deux percentile pour réaliser leurs études). Cette méthode permet de prendre en compte la

¹Le point de rupture est la valeur en dollars du plafond du marché qui délimite les 70% supérieurs du reste.

corrélation non nulle entre les caractéristiques des facteurs (le ratio 'Book to market equity' pour le facteur Value et 'market value of Equity' pour le facteur Size). D'une part, le tri Post-conditioning (S-post) consiste à trier d'abord sur la variable à pricer puis sur la variable de contrôle, d'autre part le tri Pre-conditioning trie d'abord sur la variable de contrôle puis sur la variable à pricer. Afin de voir les effets des différentes méthodes de tri, les auteurs construisent les portefeuilles puis les régressent face aux facteurs de Fama & French et Carhart afin de comprendre quels en sont les drivers.

Le tri S-post n'est pas impacté par la corrélation entre les variables, puisque le second tri est fait au sein de chaque bucket. Ainsi la corrélation entre SMB S-post et HML S-post se rapproche de 0. Par ailleurs, le S-post va donner un poids plus faible au portefeuille large cap et donc augmenter l'exposition au small cap.

L'utilisation d'un tri séquentiel donne des effets différents du tri indépendant. Un tri séquentiel permet d'abord de réduire les effets des facteurs externes. En utilisant le tri S-Post, la corrélation entre les facteurs SMB et HML nouvellement calculés passe de -23% à 4%. Cependant cette méthode ne fait que réallouer les portefeuilles facteurs suivant la variable de contrôle. En comparaison, le tri S-Pre crée des portefeuilles facteurs pour chaque niveau de contrôle, portefeuilles qui sont ensuite agrégés. En utilisant cette méthode, SMB n'est plus exposé au facteur HML de Fama & French.

Dans un second temps, les auteurs ont comparé l'apport de points d'arrêt basés sur l'univers NYSE avec ceux basés sur l'univers total (AMX, NYSE, NASDAQ). En effet, les points d'arrêt pour les capitalisations boursières (Size) varient suivant les conditions de marché et utiliser l'échantillon entier permet notamment d'avoir une meilleure définition de la catégorie 'Large Cap'. Finalement, l'utilisation de l'échantillon entier pour les points d'arrêt va permettre d'équilibrer le nombre de composants dans chaque catégorie, alors que les points d'arrêts de NYSE créaient de grandes disparités en nombre de composants entre les portefeuilles. En effet, en utilisant NYSE, Fama & French cherchent à équilibrer les portefeuilles en termes de capitalisation boursière totale.

Enfin, les auteurs réitèrent l'analyse de la partie en construisant des portefeuilles qui se basent sur l'univers total, puis les régressent contre les facteurs Value, Size et Momentum. Ils en concluent qu'utiliser l'univers total rend les facteurs SMB et HML indépendants du facteur Momentum.

Un tri symétrique (3x3) est ensuite testé afin de pondérer de la même manière chaque facteur de risque. Un tri symétrique, combiné à l'utilisation de 'Name breakpoints', diminue l'exposition de HML au facteur Size (après le pré-conditionnement par HML).

Enfin des tests se basant sur deux variables avec une corrélation négative parfaite sont réalisés. De ces tests les auteurs voient que le tri symétrique produit un fort déséquilibre entre les sous-portefeuilles alors qu'utiliser les name breakpoints dans un tri séquentiel améliore la diversification. Par ailleurs, le tri indépendant ne permet pas de retrouver l'effet Size lorsque l'on utilise un tri symétrique. Ils observent aussi qu'un tri dépendant peut être appliqué à des variables très corrélées sans introduire de biais.

Ensuite, des tris 3x3x3 sont réalisés en utilisant le facteur momentum. Les auteurs observent alors que travailler en 3 dimensions rend les facteurs différents de ceux de Fama French. Les tri avec pre et post-conditionnement impliquent des relations différentes entre les effets size et value dans le cadre d'un tri 3x3x3.

Finalement, les auteurs concluent sur l'apport du facteur momentum qui permet d'améliorer le pricing dans presque tous les cas (S-pre, name breakpoints, tri symétrique).

3 Réplication de l'étude de Fays, Hübner et Lambert

Depuis quelques années, nous assistons à une démultiplication des facteurs créés. Dans ce contexte, il est intéressant et même primordial de se rendre compte des enjeux existants dans le choix de la méthodologie de construction du facteur, avant même le choix du facteur en lui-même. Le travail de Hübner, Fays et Lambert (2018) est particulièrement complet et apporte des résultats importants sur ce sujet. Pour avoir le recul nécessaire par rapport à leurs résultats, nous avons cherché à répliquer entièrement leur étude, afin de bien comprendre leur façon de procéder et pour repérer d'éventuelles limitations.

3.1 Réplication de la partie 3 : Construction de portefeuille en utilisant un tri indépendant ou séquentiel

Dans cette partie, les auteurs présentent la différence entre les deux sortes de tri, sur un portefeuille 2x3 de Fama et French. Nous ne répliquons pas les Figure 1, 2, 3 et 4 qui présentent les étapes de tri et les résultats théoriques sur l'allocation des stocks. En revanche, nous répliquons la table 1 qui présente les effets de prix entre les différentes méthodes de construction.

Table 1 : Reproduction de la Table 1 de l'étude

	post HML	post SMB	pre HML	pre SMB
(alpha, Coeff)	0.00	0.00	0.00	0.00
(alpha, t-stats)	0.71	1.00	0.12	1.96
(hml, Coeff)	0.98	0.19	0.94	-0.01
(hml, t-stats)	88.93	37.37	90.25	-1.05
(mktrf, Coeff)	-0.03	0.02	-0.01	0.01
(mktrf, t-stats)	-4.33	6.00	-2.17	1.22
(smb, Coeff)	0.21	1.00	0.00	0.94
(smb, t-stats)	21.62	214.41	0.36	157.34
(umd, Coeff)	-0.00	-0.01	-0.01	-0.01
(umd, t-stats)	-0.18	-2.58	-2.12	-1.94
R2	0.94	0.99	0.94	0.98

Nous retrouvons donc bien des coefficients très proches de ceux de Fays, Hübner et Lambert. Et ce pour chaque méthode. Nous voyons bien que le HMLS-POST est long du facteur size de Fama French et est indépendant d'une stratégie momentum. Nous retrouvons donc la même conclusion : l'utilisation d'un tri indépendant, préconditionné sur la variable à valoriser, modifie les facteurs de risque car les seuils sont ajustés pour prendre en compte la corrélation entre les caractéristiques de tri.

3.2 Réplication de la Partie 4 : Différents choix pour trier les actions

La figure 5 ci-dessous montre clairement les avantages de l'utilisation des "name breakpoints" (seuils définis sur l'ensemble de l'échantillon) par rapport aux "NYSE breakpoints". En effet, les seuils calculés seulement sur les actions de la bourse de New York changent beaucoup plus au cours du temps (cela est particulièrement flagrant pour la médiane des capitalisations boursières). L'utilisation des name breakpoints permet donc de construire des portefeuilles dont la répartition va rester relativement stable dans le temps.

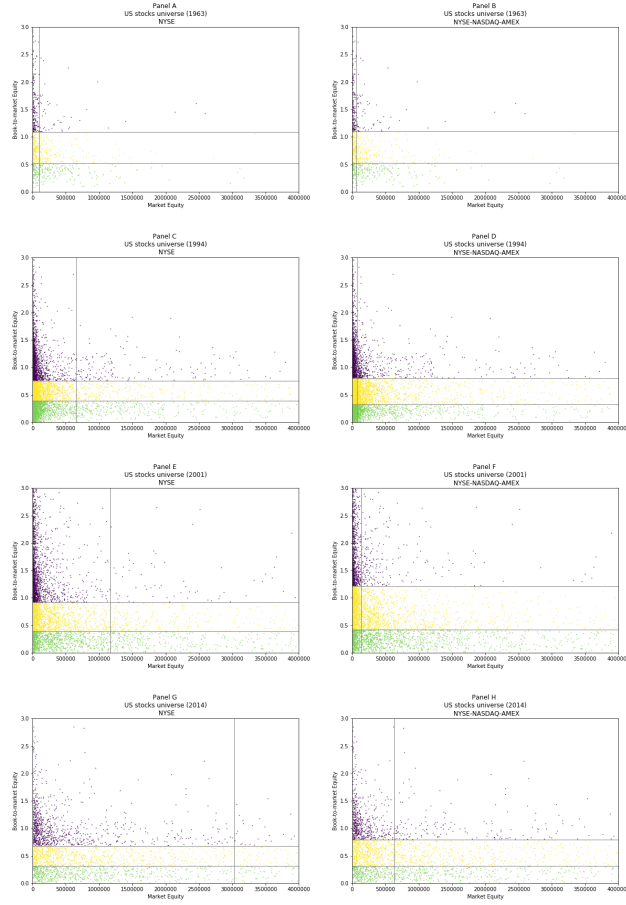


Figure 1: Réplication de la Figure 5

Nous répliquons maintenant la figure 6, qui permet de visualiser la répartition des entreprises constituant les différents portefeuilles. Nous voyons très clairement que lors d'un tri indépendant, on trouve une répartition du nombre d'actions qui évolue dans le temps, avec un biais sur les actions qui pourraient être considérées comme "value" et se retrouvent en fait en "growth". Ce biais est plus ou moins marqué en fonction des conditions de marché. Ici, on retrouve qu'il est particulièrement marqué à la fin des années 90, comme noté par Fays et al. (2018) ainsi que Chan et al. (2009).

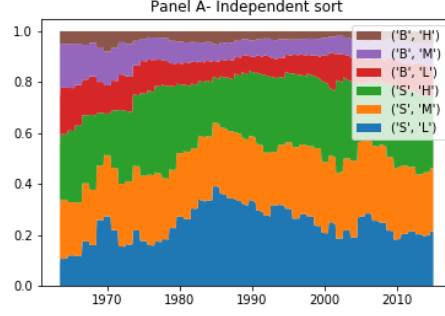


Figure 2: Réplication de la Figure 6 - Panel A : Tri indépendant

Table 2 : Description du nombre d'actions par portefeuille de la figure 2

szport bmport	S			B		
	L	M	H	L	M	H
mean	957.40	986.84	1,034.78	353.63	294.10	136.41
std	464.89	398.30	416.15	112.34	55.34	39.51
min	60.00	122.00	168.00	120.00	128.00	40.00
max	1,791.00	1,947.00	2,053.00	796.00	428.00	220.00

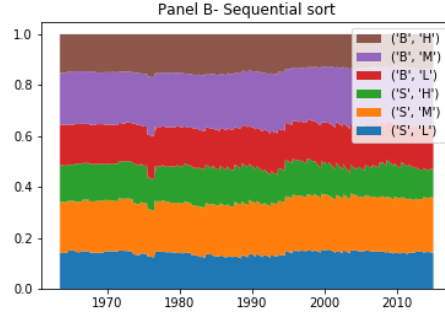


Figure 3: Réplication de la Figure 6 - Panel B : Tri séquentiel

Table 3 : Description du nombre d'actions par portefeuille de la figure 3

szport bmport	S			B		
	L	M	H	L	M	H
mean	528.08	783.63	502.61	593.84	818.46	536.37
std	197.83	299.13	177.27	213.89	301.68	180.30
min	95.00	120.00	91.00	100.00	133.00	99.00
max	993.00	1,446.00	910.00	1,008.00	1,392.00	856.00

En revanche, avec un tri séquentiel, les répartitions des actions entre les différents portefeuilles restent homogènes au cours du temps. Chaque portefeuille est composé d'un nombre d'actions similaire. En revanche, la composition est déséquilibrée en terme de capitalisation.

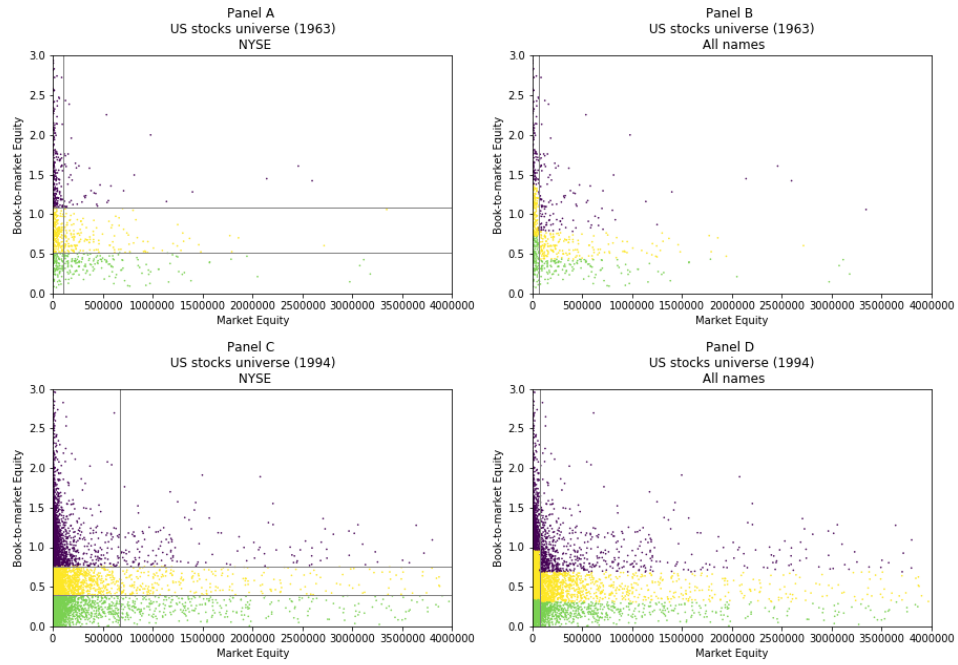


Figure 4: Réplication de la Figure 7

Nous réalisons maintenant les régressions des 2x3 portefeuilles sur les facteurs de Fama et French.

Table 4 : Réplication de la table 2 de l'étude. Régressions sur les 2x3 portefeuilles, avec les deux méthodes de calcul de seuils (NYSE et "all names") et les deux méthodes de tri (S-Pre et S-Post)

	NYSE breakpoints		names breakpoints	
	post SMB	pre SMB	post SMB	pre SMB
(alpha, Coeff)	0.00	0.00	-0.00	-0.00
(alpha, t-stats)	1.00	1.96	-0.44	-0.38
(hml, Coeff)	0.19	-0.01	0.13	-0.07
(hml, t-stats)	37.37	-1.05	5.04	-2.89
(mktrf, Coeff)	0.02	0.01	-0.06	-0.08
(mktrf, t-stats)	6.00	1.22	-3.54	-4.87
(smb, Coeff)	1.00	0.94	1.26	1.19
(smb, t-stats)	214.41	157.34	52.34	53.75
(umd, Coeff)	-0.01	-0.01	-0.01	-0.01
(umd, t-stats)	-2.58	-1.94	-0.53	-0.90
R2	0.99	0.98	0.82	0.84

	NYSE breakpoints		names breakpoints	
	post HML	pre HML	post HML	pre HML
(alpha, Coeff)	0.00	0.00	0.00	0.00
(alpha, t-stats)	0.71	0.12	5.70	4.50
(hml, Coeff)	0.98	0.94	0.93	0.96
(hml, t-stats)	88.93	90.25	45.74	51.51
(mktf, Coeff)	-0.03	-0.01	-0.09	-0.02
(mktf, t-stats)	-4.33	-2.17	-6.61	-1.92
(smb, Coeff)	0.21	0.00	0.06	-0.08
(smb, t-stats)	21.62	0.36	3.25	-4.49
(umd, Coeff)	-0.00	-0.01	0.01	-0.01
(umd, t-stats)	-0.18	-2.12	0.98	-1.04
R2	0.94	0.94	0.81	0.84

Pour chacune de ces régressions, nous retrouvons les mêmes coefficients que Fays, Hübner et Lambert pour les facteurs SMB et HML de Fama French. La plus grande différence se trouve dans le coefficient du facteur HML de Fama French dans la régression du SMB S-PRE, avec un coefficient de -0.07 au lieu du -0.102 trouvé par les auteurs. Concernant les t-stats, nous retrouvons des valeurs très similaires en valeur absolue, mais nous remarquons des différences de signe. Concernant le facteur marché, nos résultats paraissent plus volatiles, avec des différences de signe dans les coefficients pour le facteur marché dans les régressions sur le facteur SMB (en revanche dans les régressions du facteur HML les coefficients sont très proches des résultats des auteurs). Tous nos coefficients trouvés pour l'alpha sont beaucoup plus proches de 0.

Enfin, nous regardons les régressions des 3x3 portefeuilles.

Table 5 - Réplication de la table 3 de l'étude. Régressions sur les 3x3 portefeuilles, avec les deux méthodes de calcul de seuils (NYSE et "all names") et les deux méthodes de tri (S-Pre et S-Post)

	NYSE breakpoints		names breakpoints	
	post SMB	pre SMB	post SMB	pre SMB
(alpha, Coeff)	0.00	0.00	0.00	0.00
(alpha, t-stats)	0.15	0.72	0.68	0.19
(hml, Coeff)	0.27	-0.04	0.14	-0.10
(hml, t-stats)	23.02	-3.76	3.24	-2.37
(mktf, Coeff)	0.01	-0.01	-0.15	-0.16
(mktf, t-stats)	1.83	-2.21	-5.49	-5.89
(smb, Coeff)	1.23	1.18	1.37	1.36
(smb, t-stats)	117.66	131.25	35.72	36.70
(umd, Coeff)	-0.01	-0.01	-0.04	-0.04
(umd, t-stats)	-0.76	-1.57	-1.34	-1.35
R2	0.96	0.97	0.68	0.70

	NYSE breakpoints		names breakpoints	
	post HML	pre HML	post HML	pre HML
(alpha, Coeff)	0.00	0.00	0.00	0.00
(alpha, t-stats)	2.26	0.29	5.85	4.33
(hml, Coeff)	0.96	0.90	0.86	0.83
(hml, t-stats)	72.14	61.75	35.54	38.10
(mktf, Coeff)	-0.05	-0.01	-0.08	-0.03
(mktf, t-stats)	-6.08	-1.32	-5.09	-1.94
(smb, Coeff)	0.15	-0.04	-0.00	-0.10
(smb, t-stats)	12.60	-2.93	-0.10	-5.12
(umd, Coeff)	0.01	-0.02	-0.01	0.01
(umd, t-stats)	1.43	-1.87	-0.35	0.72
R2	0.91	0.88	0.72	0.75

Pour conclure sur ces deux tableaux (régressions des facteurs avec les deux méthodes de tri séquentiel, avec un tri asymétrique puis un tri symétrique) :

- Nous retrouvons bien le fait que les facteurs SMB et HML sont indépendants du facteur momentum pour les deux méthodes de tri, que ce tri soit symétrique ou non.
- De plus, nous retrouvons bien un ajustement vers le bas des facteurs HML et SMB pour le tri qui démarre avec la variable à valoriser et qui utilise tous les noms de bourse pour les seuils de tri (S-Pre, name breakpoints). Cet ajustement correspond à la correction de l'effet cross-section : le risque de size dans le facteur HML est diminué et vice-versa.
- Les coefficients obtenus restent très similaires entre le symétrique et le non-symétrique, mais on verra ensuite que le symétrique apporte une meilleure diversification.
- Il y a néanmoins une différence notable par rapport aux résultats des auteurs : Nous ne pouvons pas confirmer que les facteurs créés avec le tri séquentiel et les seuils de l'ensemble de l'échantillon apportent plus d'information que les facteurs traditionnels de Fama French. En effet, les coefficients des alphas de nos régressions restent à 0 ou extrêmement proche.
- De plus, les auteurs concluent que la méthode S-Pre permet de réduire l'impact des conditions de marché sur le facteur size. Cependant, cela n'est pas du tout flagrant dans nos résultats. Nous avons bien retrouvé que l'utilisation d'un tri séquentiel avec name breakpoints au lieu d'un tri indépendant avec NYSE breakpoints le permet (voir Figure 3) mais nous ne voyons pas de différence notable entre la méthode S-Post et la méthode S-Pre dans les tables 2 et 3 ci-dessus.

3.3 Réplication de la partie 5 : Diversification du portefeuille et les effets sur les biais sur la construction des facteurs

Dans cette partie, les auteurs adaptent les conclusions de l'Appendix B de l'article de Cochrane 'Asset Pricing as a Function of Characteristics'. Cochrane conclue que dans la répartition des portefeuilles selon Fama-French, le Sharpe Ratio (ici le ratio rendement/risque) augmente en fonction de la séparation des portefeuilles. Le Sharpe ratio, est proportionnel à la statistique t pour le rendement moyen du portefeuille d'écarts tel que :

$$\frac{E(R^{ei} - R^{ej})}{\sigma(R^{ei} - R^{ej})} = \frac{E(f)}{\sigma(f)} \frac{b(C_i - C_j)}{\sqrt{b^2(C_i - C_j)^2 + 2 * \frac{\sigma_\epsilon^2}{N} \frac{E(f)^2}{\sigma(f)^2}}} \quad (1)$$

Où C correspond aux caractéristiques (tel que la taille ou le book to market). À mesure que la différence entre les caractéristiques i et j augmente, nous retrouvons le Sharpe ratio classique. Cependant cette

différence ne va pas augmenter le ratio de Sharpe continuellement. Cochrane conclue que la séparation plus rigoureuse des portefeuilles permet d'obtenir la moyenne ou l'alpha voulue de 1% par mois, mais n'augmentera pas le Sharpe ratio ou les t-stats. Dans notre article, les auteurs rajoutent les indices i et j pour le nombre d'actions (N) dans la branche longue et courte des portefeuilles de différence, car comme nous l'avons vu, ils peuvent être différents. L'équation devient alors :

$$\frac{E(R^i - R^j)}{\sigma(R^i - R^j)} = \frac{E(f)}{\sigma(f)} \frac{b(C_i - C_j)}{\sqrt{b^2(C_i - C_j)^2 + \frac{\sigma_e^2}{N_i} \frac{E(f)^2}{\sigma(f)^2} + \frac{\sigma_e^2}{N_j} \frac{E(f)^2}{\sigma(f)^2}}} \quad (2)$$

Les auteurs définissent l'écart du facteur SMB tel que $SMB = \frac{1}{3}[(R_{SL} - R_{BL}) + (R_{SM} - R_{BM}) + (R_{SH} - R_{BH})]$, ainsi que des écarts de portefeuille parfaitement corrélés pour les différentes caractéristiques. La formule devient

$$\frac{E(SMB)}{\sigma(SMB)} = \frac{E(f)}{\sigma(f)} \frac{\sum_{i=L,M,H} b(C_{Si} - C_{Bi})}{\sum_{i=L,M,H} \sqrt{b^2(C_{Si} - C_{Bi})^2 + \frac{\sigma_e^2}{N_{Si}} \frac{E(f)^2}{\sigma(f)^2} + \frac{\sigma_e^2}{N_{Bi}} \frac{E(f)^2}{\sigma(f)^2}}} \quad (3)$$

Avec cette formule les auteurs veulent s'assurer d'une bonne diversification des portefeuilles, en effet avec un tri plus spécifique, le nombre d'actions des portefeuilles diminuerait.

En pratique, nous allons reproduire les résultats tel que la corrélation entre les matrices soient égales à -1. Pour cela nous allons remplacer le ratio book-to-market par l'opposé de la capitalisation pour avoir cette corrélation négative parfaite. Nous allons dans un premier temps trier le facteur SMB avec la capitalisation boursière et dans un second temps le facteur HML avec l'opposé de la capitalisation, Nous affichons ensuite la moyenne, l'écart type, le minimum et le maximum, comme effectué dans l'article. Nous prenons la période allant de Juillet 1963 à Décembre 2014.

Table 6 : Réplication de la table 4. Distribution des actions dans les portefeuilles triés sur des caractéristiques corrélées.

		SL	SM	SH	BL	BM	BH	ML	MM	MH
Panel A: Independent - 2x3 - NYSE	Mean	0	477	2625	445	348	0	0	0	0
	Std	0	141	1102	96	89	0	0	0	0
	Min	0	164	413	223	156	0	0	0	0
	Max	0	822	4865	716	604	0	0	0	0
Panel B: Independent - 2x3 - Name	Mean	0	765	1139	1231	759	0	0	0	0
	Std	0	273	425	445	261	0	0	0	0
	Min	0	195	281	291	188	0	0	0	0
	Max	0	1310	2118	2036	1257	0	0	0	0
Panel C: Independent - 3x3 - NYSE	Mean	0	1	2625	444	0	0	1	823	0
	Std	0	1	1102	96	0	0	1	230	0
	Min	0	0	413	222	0	0	0	318	0
	Max	0	9	4865	716	1	0	7	1426	0
Panel D: Independent - 3x3 - Name	Mean	0	1	1139	1229	0	0	2	1523	0
	Std	0	1	425	447	0	0	2	534	0
	Min	0	0	281	288	0	0	0	382	0
	Max	0	4	2118	2036	0	0	9	2550	0
Panel E: Dependent - 2x3 - NYSE (First sort on HML)	Mean	233	475	2104	211	350	520	0	0	0
	Std	53	142	912	42	88	201	0	0	0
	Min	113	162	277	110	158	135	0	0	0
	Max	383	822	3979	333	604	910	0	0	0
Panel E: Dependent - 2x3 - NYSE (First sort on SMB)	Mean	338	656	2105	212	313	269	0	0	0
	Std	99	245	913	42	73	69	0	0	0
	Min	121	176	277	111	153	117	0	0	0
	Max	595	1138	3980	333	527	460	0	0	0
Panel F: Dependent - 2x3 - Name (First sort on HML)	Mean	596	762	558	635	761	581	0	0	0
	Std	206	274	206	240	260	221	0	0	0
	Min	147	192	135	144	191	146	0	0	0
	Max	964	1310	1053	1083	1257	1065	0	0	0
Panel F: Dependent - 2x3 - Name (First sort on SMB)	Mean	566	775	560	635	788	569	0	0	0
	Std	203	291	205	240	272	194	0	0	0
	Min	142	196	135	144	193	144	0	0	0
	Max	968	1407	1054	1084	1278	942	0	0	0
Panel G: Dependent - 3x3 - NYSE (First sort on HML)	Mean	143	305	1771	125	202	281	176	317	572
	Std	33	96	767	24	50	104	38	86	255
	Min	68	101	205	63	95	78	91	124	128
	Max	228	524	3304	200	355	507	287	547	1125
Panel G: Dependent - 3x3 - NYSE (First sort on SMB)	Mean	281	571	1771	125	176	143	202	317	306
	Std	104	255	767	24	38	33	50	86	96
	Min	78	128	205	63	91	68	95	124	101
	Max	507	1125	3304	200	287	228	355	547	524
Panel H: Dependent - 3x3 - Name (First sort on HML)	Mean	354	460	325	387	459	347	490	604	466
	Std	121	168	119	151	156	132	174	210	176
	Min	89	118	77	86	112	86	116	153	116
	Max	565	803	629	673	742	636	812	1008	859
Panel H: Dependent - 3x3 - Name (First sort on SMB)	Mean	347	466	325	387	490	354	459	604	460
	Std	132	176	119	151	174	121	156	210	168
	Min	86	116	77	86	116	89	112	153	118
	Max	636	859	629	673	812	565	742	1008	803

- Nous constatons que la méthode de tri indépendante entraîne de fortes disparités entre les portefeuilles. Par exemple, le panel A (tri indépendant 2x3 prenant en compte tout l'univers des breakpoints) a une moyenne de 0 actions pour le portefeuille SL et une moyenne de 2625 actions pour le portefeuille SH.
- Les portefeuilles ayant un tri dépendant sur tout l'univers des breakpoints ont une répartition plus uniforme, et particulièrement pour les portefeuilles symétriques (3x3).

Nous avons calculé dans le tableau ci-dessous l'écart type de la moyenne des répartitions des porte-

feuilles. Ce tableau confirme bien nos conclusions où le Panel H a le plus faible écart type et donc une homogénéisation de la répartition de ces portefeuilles.

Table 7 : *Ecart type des moyennes des portefeuilles pour les panel E, F, G et H de la table 6.*

	Panel E	Panel F	Panel G	Panel H
SD	575	87	422	82

Table 8 : *Réplication de la table 5. Mesure du biais dans la méthode de construction de facteur.*

	Nb Obs	Mean (in %)	Std (in %)	Sum	Min	Max	Correl
Panel A: Independent - 2x3 - NYSE / HML	618.00	0.33	3.95	2.02	-0.19	0.26	1.00
Panel A: Independent - 2x3 - NYSE / SMB	618.00	0.19	2.48	1.21	-0.15	0.14	0.98
Panel B: Independent - 2x3 - Name / HML	618.00	0.40	4.94	2.46	-0.19	0.32	1.00
Panel B: Independent - 2x3 - Name / SMB	618.00	0.23	3.15	1.41	-0.12	0.22	0.98
Panel C: Independent - 3x3 - NYSE / HML	618.00	0.33	3.96	2.02	-0.19	0.26	1.00
Panel C: Independent - 3x3 - NYSE / SMB	618.00	0.20	4.46	1.22	-0.19	0.26	0.81
Panel D: Independent - 3x3 - Name / HML	618.00	0.40	4.94	2.46	-0.19	0.32	1.00
Panel D: Independent - 3x3 - Name / SMB	618.00	0.27	5.96	1.68	-0.19	0.32	0.77
Panel E: Dependent - 2x3 - NYSE / SMB	618.00	0.26	3.56	1.60	-0.17	0.22	1.00
Panel E: Dependent - 2x3 - NYSE / HML	618.00	0.19	2.58	1.17	-0.15	0.16	0.88
Panel F: Dependent - 2x3 - Name / SMB	618.00	0.35	4.57	2.15	-0.19	0.26	1.00
Panel F: Dependent - 2x3 - Name / HML	618.00	0.24	3.31	1.46	-0.14	0.20	0.86
Panel G: Dependent - 3x3 - NYSE / SMB	618.00	0.29	3.53	1.79	-0.16	0.21	1.00
Panel G: Dependent - 3x3 - NYSE / HML	618.00	0.26	3.47	1.63	-0.16	0.22	0.93
Panel H: Dependent - 3x3 - Name / SMB	618.00	0.47	4.68	2.89	-0.18	0.28	1.00
Panel H: Dependent - 3x3 - Name / HML	618.00	0.38	4.54	2.33	-0.19	0.25	0.92

Dans un second temps nous avons étudié comme dans l'article, les corrélations des portefeuilles qui devraient théoriquement être parfaitement corrélées et avoir de même statistiques.

- Pour le tri indépendant 2x3, nous retrouvons la même forte corrélation entre les facteurs SMB et HML, et une forte disparité entre les deux moyennes et les écarts-type que les auteurs de l'article. Cependant contrairement au tableau original, nous avons trouvé des corrélations élevées entre les deux facteurs du tri indépendant 3x3, mais avec toujours des différences importantes entre les moyennes et les écarts type. Nous ne pouvons pas conclure comme dans l'article une répartition inégale des actions dans les portefeuilles pour la méthode de tri indépendant, néanmoins nous pouvons comparer avec les méthodes de tri séquentiel, pour déterminer la méthode qui va introduire le moins de biais.
- Pour le tri séquentiel 2x3, pour les deux univers nous avons des corrélations élevées mais toujours des disparités entre les statistiques premières. Le tri séquentiel 3x3 est la méthode avec les plus fortes corrélations ainsi que de plus grandes similitudes entre les statistiques calculées.

Les conclusions des tableaux 4 et 5 se rejoignent en nous montrant que le Panel H (tri séquentiel 3x3 S-pre) est la méthode de tri la plus adéquate pour le tri de caractéristiques corrélées, avec une répartition des portefeuille homogène et une corrélation importante entre les facteurs SMB et HML.

Nous pouvons conclure que le tri séquentiel 3x3 est la méthode la plus efficace pour trier des caractéristiques ayant une corrélation importante. Même si nous ne retrouvons pas des corrélations aussi significatives que les auteurs de l'article, la méthode de tri séquentiel 3x3 reste celle avec le moins de biais sont introduits.

3.4 Réplication de la partie 6 : Tri triple pour créer 3x3x3 portefeuilles

Les auteurs vont dans cette partie, rajouter au tri séquentiel le facteur momentum. Ce facteur a été introduit par Carhart en 1997 au modèle de Fama French. Le facteur momentum se calcule en prenant en

compte les récentes sur-performances (Up) ou sous-performances des actions (Down) pour les investissements futures, en continuant d'investir dans les actions qui ont surperformé et de vendre les actions ayant sous performé. Nous avons choisis de prendre un an comme période temporelle historique en enlevant le dernier mois passé.

Le momentum facteur est définis par le rendement moyens des deux portefeuilles à haut rendement antérieur moins le rendement moyen des deux portefeuilles à faible rendement antérieur tel que :

$$UMD = \frac{1}{2}(SH + BH) - \frac{1}{2}(SL + BL) \quad (4)$$

Pour l'intégrer à notre tri, nous trions en premier le momentum, en répartissant en fonction du 30th percentile et du 70th percentile. La suite du tri se fait selon la méthode choisie comme nous l'avons vu dans les parties précédentes pour les méthodes de tri séquentiel 3x3.

Table 9 : Réplication de la table 6. Régressions de portefeuilles 3x3x3 avec les seuils déterminés sur l'ensemble de l'échantillon (name breakpoints) à comparer avec les tables 4 et 5 de ce papier.

	names breakpoints			
	post HML	post SMB	pre HML	pre SMB
(alpha, Coeff)	0.00	-0.01	0.01	-0.00
(alpha, t-stats)	0.56	-1.08	2.56	-0.72
(hml, Coeff)	1.14	0.17	0.85	-0.04
(hml, t-stats)	5.69	0.77	5.03	-0.19
(mktrf, Coeff)	0.19	-0.23	0.14	-0.34
(mktrf, t-stats)	1.14	-1.32	1.00	-1.96
(smb, Coeff)	-0.15	1.48	-0.09	1.45
(smb, t-stats)	-0.75	6.65	-0.51	6.60
(umd, Coeff)	-0.13	0.08	-0.28	0.10
(umd, t-stats)	-0.82	0.46	-2.09	0.60
R2	0.75	0.79	0.76	0.81

Nous remarquons tout d'abord que les R2 de nos régressions se rapprochent des R2 que nous avons obtenus pour les régressions sur les facteurs 3x3 avec les name breakpoints. Nous nous serions cependant attendu à des R2 encore plus faibles, puisque ces facteurs s'écartent encore plus des facteurs de Fama French.

Pour ce qui est des effets observés plus tôt, ils se retrouvent dans cette table 9 :

- Les facteurs ne sont pas, ou très peu, affectés par l'effet momentum.
- On retrouve un coefficient proche de 0 du HML pour le pre SMB et un coefficient négatif du SMB pour le pre HML : avec 3x3x3 portefeuilles, conditionner le tri en premier sur la variable de controle entraîne une relation négative entre les effets size et value plus forte que pour les 3x3, voir une indépendance (indépendance du pre SMB par rapport au facteur value, comme trouvé par les auteurs).

Nous avons par la suite cherché à répliquer la dernière table de l'étude : la table 7. Cette table illustre les résultats obtenus précédemment en présentant les erreurs de valorisation sur 10x10 portefeuilles (Size et Value) en fonction des méthodes de construction de facteurs. En suivant les indications des auteurs, nous avons pu récupérer ces portefeuilles sur le site de Kenneth French. Cependant, nous ne sommes pas sûrs d'utiliser les données de ces portefeuilles de la même manière que les auteurs, car nous n'obtenons pas du tout les bons résultats concernant les moyennes des valeurs absolues des coefficients et t-statistiques des alphas.

Table 10 : Essai de réplcation de la table 7 - Panel A (modèle à 3 facteurs). Moyennes des valeurs absolues des alphas des régressions des 10x10 portefeuilles (size et value) sur les différents facteurs créés : en haut, les 4 facteurs 2x3 ainsi que le tri indépendant de Fama French, en bas, les 4 facteurs 3x3.

2x3	NYSE - independent	NYSE - post	names - post	NYSE - pre	names - pre
Alpha	0.42	0.42	0.41	0.41	0.38
t-stat	4.15	4.12	3.83	4.05	3.65
R2	0.81	0.81	0.78	0.81	0.79
	3x3	NYSE - post	names - post	NYSE - pre	names - pre
Alpha	0.42	0.40	0.43	0.38	
t-stat	4.13	3.53	4.23	3.47	
R2	0.80	0.77	0.80	0.77	

Table 11 : Essai de réplcation de la table 7 - Panel B (modèle à 3 facteurs). Moyennes des valeurs absolues des alphas des régressions des 10x10 portefeuilles (size et value) sur les différents facteurs créés : en haut, les 4 facteurs 2x3 ainsi que le tri indépendant de Fama French, en bas, les 4 facteurs 3x3.

2x3	NYSE - independent	NYSE - post	names - post	NYSE - pre	names - pre
Alpha	0.44	0.44	0.44	0.42	0.39
t-stat	4.25	4.20	3.96	4.08	3.66
R2	0.81	0.81	0.79	0.81	0.79
	3x3	NYSE - post	names - post	NYSE - pre	names - pre
Alpha	0.44	0.42	0.44	0.40	
t-stat	4.25	3.60	4.27	3.53	
R2	0.80	0.77	0.81	0.77	

Nos résultats nous semblent malheureusement inexploitable car nous obtenons des moyennes des valeurs absolues d’alphas bien supérieures aux chiffres obtenus par les auteurs. Nous retrouvons cependant la même dynamique que les auteurs puisque nous voyons bien que les alphas diminuent en utilisant les seuils calculés sur l’échantillon entier, et en utilisant un pre-conditionnement sur la variable de contrôle.

4 Extension

Dans cette partie, nous tâchons de confirmer les résultats trouvés précédemment avec les dernières données disponibles : de Décembre 2014 à Décembre 2019.

4.1 Distributions des entreprises parmi les caractéristiques des 2x3 portefeuilles

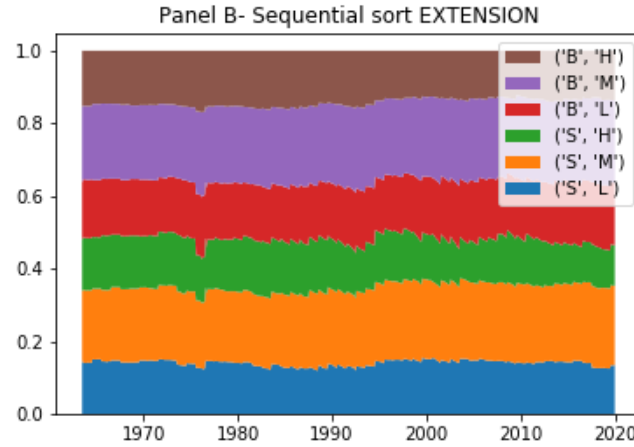


Figure 5: Tri séquentiel effectué avec les "name breakpoints", jusqu'en 2020

Sur la figure 5, nous remarquons que la répartition des entreprises entre les différents portefeuilles reste bien homogène dans le temps, même sur la période plus récente, lorsqu'un tri séquentiel est utilisé.

4.2 Régression des 2x3 et 3x3 portefeuilles sur les facteurs de Fama French sur la période récente

Nous étendons la régression des 2x3 portefeuilles sur la période allant de Décembre 2014 à Décembre 2019. Nous retrouvons des résultats très similaires à ceux obtenus sur la période de 1963 à 2014.

	NYSE breakpoints		names breakpoints	
	post SMB	pre SMB	post SMB	pre SMB
(alpha, Coeff)	-0.00	0.00	0.00	0.00
(alpha, t-stats)	-0.73	1.01	0.35	1.53
(hml, Coeff)	0.20	0.07	0.14	-0.05
(hml, t-stats)	12.46	4.16	1.99	-0.99
(mktrf, Coeff)	0.01	0.01	-0.14	-0.12
(mktrf, t-stats)	0.96	0.60	-2.89	-3.77
(smb, Coeff)	0.96	0.90	1.17	1.03
(smb, t-stats)	65.89	56.71	18.58	24.28
(umd, Coeff)	-0.01	0.04	-0.07	-0.03
(umd, t-stats)	-0.55	2.78	-1.28	-0.90
R2	0.99	0.98	0.88	0.92

	NYSE breakpoints		names breakpoints	
	post HML	pre HML	post HML	pre HML
(alpha, Coeff)	-0.00	-0.00	-0.00	0.00
(alpha, t-stats)	-0.09	-1.62	-0.11	0.12
(hml, Coeff)	0.84	0.77	0.88	0.81
(hml, t-stats)	23.81	23.25	15.42	11.16
(mktrf, Coeff)	-0.09	-0.12	-0.14	-0.14
(mktrf, t-stats)	-3.71	-5.24	-3.70	-2.77
(smb, Coeff)	0.15	-0.00	-0.09	-0.08
(smb, t-stats)	4.69	-0.07	-1.68	-1.27
(umd, Coeff)	-0.08	-0.11	-0.06	-0.15
(umd, t-stats)	-3.20	-4.49	-1.37	-2.79
R2	0.95	0.95	0.88	0.82

Nous renouvelons l'exercice pour les 3x3 portefeuilles. Là encore, les coefficient et t-stats obtenus sont très proches des résultats précédents.

	NYSE breakpoints		names breakpoints	
	post SMB	pre SMB	post SMB	pre SMB
(alpha, Coeff)	0.00	0.00	-0.00	0.00
(alpha, t-stats)	0.63	1.62	-0.25	0.03
(hml, Coeff)	0.24	0.03	-0.13	-0.20
(hml, t-stats)	5.92	1.08	-1.07	-1.88
(mktrf, Coeff)	-0.05	-0.05	-0.22	-0.22
(mktrf, t-stats)	-1.62	-2.40	-2.71	-3.00
(smb, Coeff)	1.14	1.11	1.00	1.05
(smb, t-stats)	30.19	40.18	9.02	10.51
(umd, Coeff)	-0.03	0.01	-0.22	-0.14
(umd, t-stats)	-1.05	0.54	-2.41	-1.78
R2	0.95	0.97	0.63	0.68

	NYSE breakpoints		names breakpoints	
	post HML	pre HML	post HML	pre HML
(alpha, Coeff)	0.00	-0.00	0.00	0.00
(alpha, t-stats)	1.04	-0.54	1.41	1.68
(hml, Coeff)	0.85	0.78	0.75	0.79
(hml, t-stats)	21.40	15.24	6.59	8.54
(mktf, Coeff)	-0.14	-0.11	-0.21	-0.17
(mktf, t-stats)	-5.04	-3.20	-2.65	-2.72
(smb, Coeff)	0.03	-0.05	-0.27	-0.12
(smb, t-stats)	0.81	-1.06	-2.60	-1.42
(umd, Coeff)	-0.09	-0.11	-0.15	-0.02
(umd, t-stats)	-3.06	-2.86	-1.71	-0.31
R2	0.94	0.89	0.64	0.70

Nous remarquons que sur cette période, l'effet de couverture du risque du facteur externe est beaucoup plus évident pour les facteurs HML et SMB lors d'un tri S-Post avec les name breakpoints. Par exemple, en 3x3, pour le HML S-Post, le coefficient du SMB est de -0.27, et pour le HML S-Pre il est de -0.12. Sur les 50 ans précédents, on avait bien un coefficient négatif comme cela pour le S-Pre, mais le coefficient était beaucoup moins négatif pour le S-Post. On trouve la même chose pour les 2x3 portefeuilles.

5 Conclusion

En répliquant l'étude, nous nous sommes aperçus que les données nécessaires étaient très facilement accessibles. Que cela soit sur le site de Kenneth French ou sur le site du WRDS. La difficulté était donc plutôt de comprendre dans le moindre détail la façon dont les auteurs ont traité ou regroupé la donnée. Nous avons cependant obtenu des résultats très satisfaisants pour les tris de portefeuilles (voir graphiques de la réplique de la partie 4). Pour les régressions, nous avons également obtenu des coefficients majoritairement très proches des coefficients trouvés par les auteurs. Une interrogation reste pour les t-stats, puisque les nôtres s'éloignent souvent des résultats attendus. De plus, notre travail comporte une limite importante puisque nous n'avons pas pu répliquer entièrement le tableau 7, qui permet d'appliquer les conclusions précédentes à un exercice concret de valorisation sur 100 portefeuilles.

Au vu de nos résultats, nous pouvons corroborer les conclusions de Fays et al. (2018) avec confiance sur les points suivants :

- Le tri qui démarre avec la variable à valoriser et qui utilise tous les noms de bourse pour les seuils de tri (S-Pre, name breakpoints) entraîne bien un ajustement vers le bas des facteurs HML et SMB afin de couvrir le risque exogène : le risque du facteur size dans le facteur HML est diminué, et vice-versa. Cet effet est encore plus évident sur les six dernières années.
- Le tri séquentiel symétrique (3x3 plutôt que 2x3) apporte une meilleure diversification des portefeuilles. Cela est important pour éviter le problème de facteur biaisé, qui est provoqué par une répartition inégale entre les différents portefeuilles lors d'un tri indépendant sur les caractéristiques qui sont corrélées entre elles.
- Le tri séquentiel 3x3 S-pre est la méthode de tri la plus adéquate pour le tri de caractéristiques corrélées, avec une répartition des portefeuilles homogène et une corrélation importante entre les facteurs SMB et HML.
- Toujours avec ce tri S-pre, pour 3x3x3 portefeuilles, on retrouve une relation négative entre les effets size et value plus forte que pour les 3x3, voir une indépendance (indépendance du pre SMB par rapport au facteur value, comme trouvé par les auteurs).

Nous sommes moins confiants pour les points suivants :

- Nous ne pouvons pas confirmer que les facteurs créés avec le tri séquentiel et les seuils de l'ensemble de l'échantillon apportent plus d'information que les facteurs traditionnels de Fama French. En effet, les coefficients des alphas de nos régressions restent à 0 ou extrêmement proche.
- De plus, les auteurs concluent que la méthode S-Pre permet de réduire l'impact des conditions de marché sur le facteur size. Cependant, cela n'est pas du tout flagrant dans nos résultats. Nous avons bien retrouvé que l'utilisation d'un tri séquentiel avec name breakpoints au lieu d'un tri indépendant avec NYSE breakpoints le permet (voir Figure 3) mais nous ne voyons pas de différence notable entre la méthode S-Post et la méthode S-Pre dans les tables 2 et 3 ci-dessus.

Pour conclure, nous retrouvons, comme les auteurs de l'étude répliquées, qu'un tri séquentiel avec post-conditionnement sur la variable de contrôle et avec des seuils sur l'échantillon entier apporte plus d'avantages que les tris traditionnels. Il serait souhaitable d'étendre l'étude sur un modèle autre que le modèle à quatre facteurs size, value, momentum, comme les auteurs l'ont proposé.

References

- Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, 1997. doi: 10.1111/j.1540-6261.1997.tb03808.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1997.tb03808.x>.
- L Chan, S Dimmock, and J Lakonishok. Performance: Issues and evidence. *Rev. Financ. Stud.*, 22(11): 4554–4599, 2009.
- Kent Daniel and Sheridan Titman. Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance*, 52(1):1–33, 1997. doi: 10.1111/j.1540-6261.1997.tb03806.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1997.tb03806.x>.
- Eugene Fama and Kenneth French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. URL <https://EconPapers.repec.org/RePEc:eee:jfinec:v:33:y:1993:i:1:p:3-56>.
- Boris Fays, Georges Hübner, and Marie Lambert. Factoring characteristics into returns: A clinical approach to fama-french portfolio decomposition. 2018.
- Leonid Kogan and Mary H. Tian. Firm characteristics and empirical factor models: A model-mining experiment. *FRB International Finance Discussion Paper*, 2015. doi: 10.2139/ssrn.2182139. URL <https://ssrn.com/abstract=2182139>.