ACSEG'2002 proceedings - Connectionist Approaches in Economics and Management Sciences
Boulogne-sur-Mer (France), 21-22 November 2002, pp. 201-212

Classification of investment funds by self-organizing maps

P. Cardon[1], A. Lendasse[1], V. Wertz[1], E. de Bodt[2], M. Verleysen[3]

[1]Université catholique de Louvain, CESAME, 4 av. G. Lemaître
B-1348 Louvain-la-Neuve, Belgium, {lendasse, wertz}@auto.ucl.ac.be.
[2]Université catholique de Louvain, IAG, 1 pl. des Doyens,
B-1348 Louvain-la-Neuve, Belgium, debodt@fin.ucl.ac.be et
Université Lille 2, ESA, Place Deliot, BP 381, F-59020 Lille, France
[3]Université catholique de Louvain, Département d'Electricité, 3 pl. du Levant,
B-1348 Louvain-la-Neuve, Belgium, verleysen@dice.ucl.ac.be.

An investment fund (or mutual fund) is an investment structure collecting money coming from individuals and investing according to preestablished objectives. Professional managers decide of the investment strategy and assets selection in the name of those who invest in the fund, buying and selling placements, such as cash holdings, bonds and shares.

The investment strategy announced for funds (through publicity, stock market information, etc.) is crucial information for the public, in order to choose among the variety of investment funds. Unfortunately, all managers do not necessarily follow the announced strategy [1]: the performances of a specific fund being traditionally compared with those of other funds which claim to follow a similar investment strategy, one of the reasons of the divergence between announced and implemented strategy can be a deliberated act of the manager, who has an obvious interest to have its fund compared to others ones with weaker performances, whatever is the real investment strategy. This makes difficult the task of the investor, who cannot trust the public information about the fund anymore and consequently is not able to estimate the real risk he/she takes when investing in the fund.

Classifications exist to provide supplementary information to the investors. Nevertheless, some studies showed that specific funds are wrongly classified [2, 3]. In this work, we establish a classification of investment funds, based only on the fund returns, without using any information from the fund manager. In that purpose, we do not use the fund returns time series, but rather features measuring the sensitivity of these returns with regards to a set of market indexes, as proposed by Sharpe [4] under the name of *style analysis*. To build an independent classification based on these features, we use Kohonen self-organizing maps. The study is illustrated on the CRSP (Center of Research in Security Prices) database from the Chicago University. In order to compare the results, we use a reference classification from the ICDI (Investment Company Data, Inc.) and Standard & Poor's Fund Services.

We model the problem into the following system of equations:

$$R_i(t) = b_{1i}F_1(t) + b_{2i}F_2(t) + \cdots + b_{ni}F_n(t) + e_i(t)$$

where R_i is the return of fund i, b_{ki} is the sensitivity of fund i return with regard to the indexes returns F_k, and e_i includes all factors not explained by the set of market indexes.

The above analysis is valid only when each indicator rate of return cannot be perfectly explained as a function of other ones [3]. This limitation is the well-known *multi-colinearity* problem. In order to solve it, it is thus necessary to replace the set of indicators F_j with another set of less-correlated indicators G_k ($1 \le k \le m$, $m < n$). In that purpose, we will use a classical statistical technique, the Principal Component Analysis (PCA) (see for example Saporta [5]).

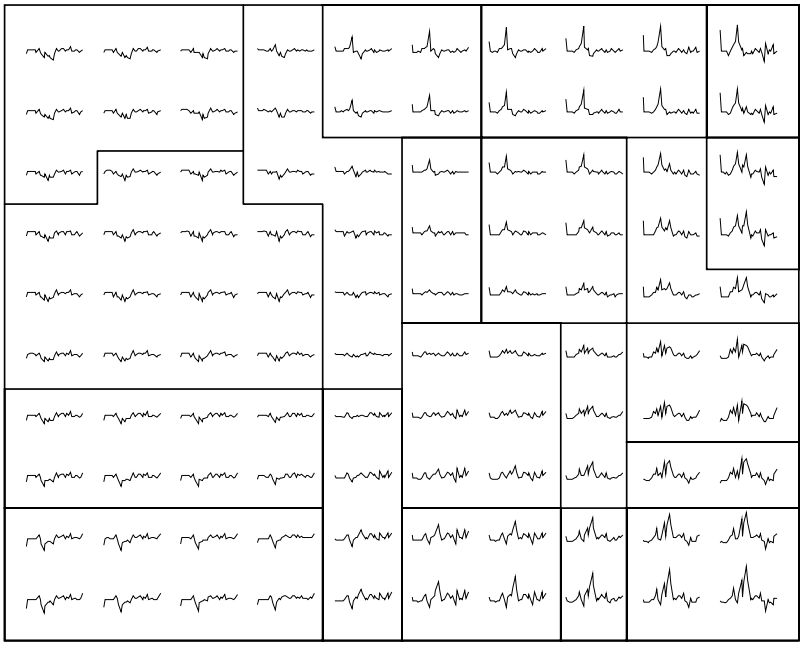
Thanks to the PCA, we are able to substitute the F_j indicators with the new set of G_k indicators in the original equations, leading to

$$R_i(t) = c_{1i}G_1(t) + c_{2i}G_2(t) + \cdots + c_{mi}G_m(t) + e_i(t).$$

Note that, from the c_{ki} coefficients obtained by solving this last system, it is possible to compute new b'_{ji} coefficients similar to the n b_{ji} ones, i.e. regression coefficients with respect to the original indices F_i, by inverting the PCA.

The sensitivity vectors b'_{ji} are first classified according to Kohonen self-organizing maps [6]. Next, the Ward algorithm [5, 7] is used to further reduce the number of classes obtained using the Kohonen classification. Working in two steps rather than using Kohonen or Ward as single classification, makes it possible to combine the advantages of both methods: visualization properties and quantization performances of Kohonen maps, and ability to work with a low number of classes with Ward's algorithm.

The following figure shows the representative vectors (codewords, or centroids) of 100 classes resulting from the Kohonen classification, grouped in 20 macro-classes using Ward's algorithm [8]. Each curve represents the set of b'_{ji} coefficients from the representative vector of one class.



The classification obtained with the proposed method is compared to the reference classification, using a contingency table. From a classification point of view only, the intra-class inertia has been calculated for both classifications: we obtain 0.07 for our method and 0.13 for the reference classification (i.e. a performance increase of about 50%!). The advantages of a more coherent classification should enable investors to take into account a different source of information, complementary to the traditionally used reference classifications.

REFERENCES

1. Investment Company Institute, *Fundamentals*, 1999, http://www.ici.org/pdf/fm-v8n1.pdf.
2. Kim T.-H., Stone D. and Tomas M., *Mutual fund objective misclassification.* Journal of Economics and Business, 2000. **52**: p. 309-323.
3. Dibartolomo D. and Witowski E., *Mutual fund misclissification : Evidence based on style analysis.* Financial Analyst Journal, 1997: p. 32-43.
4. Sharpe F. W., *Asset Allocation: management style an performance measurement*. Journal of Portfolio Management, 1992. **46**: p. 7-19.
5. Saporta G., *Probabilités, Analyse des Données et statistiques.* 1990, Paris: Editions Technip.
6. Kohonen, T., *Self-organizing maps*. Springer series in information sciences ; 30. 1995, Berlin: Springer,. xv, 362.
7. Ward J. H., *Hierarchical grouping to optimize an objective function.* Journal of the American Statistical Association, 1963. **58**: p. 236-244.
8. Cottrell M., Girard B. and Rousset P, *Forecasting of Curves using a ohonen classification.* Journal of Forecasting, 1998. **17**: p. 5-6.

ACSEG'2002 proceedings - Connectionist Approaches in Economics and Management Sciences
Boulogne-sur-Mer (France), 21-22 November 2002, pp. 201-212

Classification de fonds communs d'investissement par cartes auto-organisées

P. Cardon[1], A. Lendasse[1], V. Wertz[1], E. de Bodt[2], M. Verleysen[3]

[1]Université catholique de Louvain, CESAME, 4 av. G. Lemaître
B-1348 Louvain-la-Neuve, Belgium, {lendasse, wertz}@auto.ucl.ac.be.
[2]Université catholique de Louvain, IAG, 1 pl. des Doyens,
B-1348 Louvain-la-Neuve, Belgium, debodt@fin.ucl.ac.be et
Université Lille 2, ESA, Place Deliot, BP 381, F-59020 Lille, France
[3]Université catholique de Louvain, Département d'Electricité, 3 pl. du Levant,
B-1348 Louvain-la-Neuve, Belgium, verleysen@dice.ucl.ac.be.

Résumé. Les classements effectués régulièrement par la presse financière concernant les performances des fonds communs de placement constituent un enjeu de première importance. Dans ce contexte, un risque sérieux de distorsion entre la stratégie affichée et la stratégie effectivement mise en œuvre par les gestionnaires existe. Ces derniers peuvent effet avoir intérêt à être classé dans une catégorie de fonds qui ne reflètent par la réalité de leur stratégie afin de se comparer favorablement aux performances des autres fonds repris dans la catégorie en question. Notre travail montre comment une classification indépendante, basée sur les taux de rentabilité des fonds, peut être construite, le but étant de donner aux investisseurs une information complémentaire, plus objective. La classification des fonds se base sur des caractéristiques extraites de leurs taux de rentabilité, à savoir les sensibilités par rapport aux taux de rentabilité d'indices boursiers de référence (l'analyse des styles de gestion introduite par Sharpe [1]). La classification proposée est comparée à une classification de référence (ICDI et S&P). L'analyse des différences entre les résultats obtenus et, en particulier, de leur origine est éloquente.

1. Introduction

Un fonds commun de placement (ou fonds mutuel) est un investissement qui met en commun des sommes d'argent provenant d'autres individus et qui investit l'argent selon des objectifs préétablis. Les gestionnaires professionnels prennent des décisions d'investissement au nom de ceux qui investissent dans les fonds, en achetant et en vendant des placements, tels que des titres du marché monétaire, des obligations et des actions. Concrètement, le fonds commun de placement est un outil extrêmement utile pour l'investisseur moyen permettant de combiner diversification et investissement thématique. En plaçant son argent dans un fonds, il espère bien sûr faire fructifier le capital investi.

Dans [2] on estime qu'en 1999, 44.4 millions de ménages aux états Unis (plus de 45\%) possédaient des fonds communs. Aujourd'hui plus de 766 milliards de dollars sont investis dans des fonds communs. Si ces fonds connaissent beaucoup de succès, c'est parce qu'ils sont abordables, faciles à acheter et à vendre et présentent des éléments de diversification (la plupart des fonds mutuels sont composés de dizaines de placements différents). Pour l'investisseur moyen, l'investissement dans les fonds mutuels est un moyen facile et abordable pour obtenir l'accès aux placements qui seraient en d'autres cas disponibles uniquement pour les grandes sociétés ou pour les plus fortunés. Parce qu'ils détiennent habituellement de 50 à 100 placements différents, les fonds mutuels offrent un degré de diversification qu'un individu pourrait difficilement atteindre par lui-même.

La stratégie d'investissement annoncée pour un fonds (à travers la publicité, les informations de bourse disponibles, etc.) est un élément d'information capital aux yeux du public, en vue de choisir parmi l'ensemble des fonds disponibles. Compte tenu des enjeux, il existe un incitant fort pour les gestionnaires de fonds à ne pas suivre la stratégie de gestion annoncée [2]. Les gestionnaires peuvent ainsi chercher à être comparés à des fonds qui, en réalité, suivent des stratégies très différentes (par exemple, caractérisées par une prise de risque plus faible) afin d'afficher de manière aussi systématique que possible des performances supérieures. Ceci pose un problème à l'investisseur qui ne pourra plus se fier à la documentation du fonds et par conséquent ne pourra pas estimer le risque qu'il prend en investissant dans le fonds.

Des classifications existent pour apporter un complément d'information aux investisseurs ; néanmoins, des études ont montré que certains fonds y étaient mal classés [3, 4]. Dans ce travail, nous établirons une classification des fonds communs d'investissement, basée uniquement sur l'évolution du taux de rentabilité des fonds, et sans utiliser la moindre information provenant de la stratégie annoncée par le gestionnaire. Pour ce faire, nous utiliserons non pas les taux de rentabilité bruts des fonds mais bien des caractéristiques mesurant la sensibilité de ces taux par rapport aux taux de rentabilité d'indices financiers connus, en prenant appui sur l'analyse des styles de Sharpe [1]. Pour établir une classification indépendante de ces caractéristiques, nous utiliserons les propriétés des cartes auto-organisées de Kohonen.

L'extraction de ces caractéristiques fera l'objet de la section 2, analyse de styles. La section 3 détaillera les données utilisées ainsi que les résultats obtenus par l'analyse de style sur celles-ci. La section 4 présentera la classification effectuée sur base des cartes de Kohonen et la comparera à une classification existante. L'étude sera illustrée sur la base de données du CRSP (Center of Research in Security Prices) de l'Université de Chicago. La classification de référence utilisée a été établie par l'ICDI (Investment Company Data, Inc.) et le Standard & Poor's Fund Services.

2. Analyse de styles

Dans cette étude, nous voulons classer les fonds communs d'investissement en se basant uniquement sur l'évolution historique de leurs taux de rentabilité. Utiliser les valeurs brutes de ces taux serait néanmoins une erreur. Par exemple, un fonds d'actions suédoises pourrait très bien avoir une évolution historique de son taux de rentabilité fort semblable à l'évolution de l'indice Nikei, sans pour cela que le gestionnaire du fonds n'ait investi dans des actions japonaises.

Il est dès lors indispensable de baser la classification sur un certain nombre d'indicateurs statistiques qui caractérisent le style de gestion effectué par le gestionnaire, caractéristiques qui seront davantage discriminantes pour la classification recherchée. L'extraction de ces caractéristiques est ici basée sur les travaux de W.F. Sharpe [1].

Afin d'extraire ces caractéristiques, nous allons tenter d'expliquer le taux de rentabilité d'un fonds *i* comme une combinaison linéaire des taux de rentabilité de différents marchés de référence dans lequel le gestionnaire a pu investir. Par exemple, si le gestionnaire de fonds a investi la totalité de l'argent sur un certain marché, nous trouverons une forte corrélation entre le taux de rentabilité du fonds et celui de ce marché. En pratique, le taux de rentabilité d'un marché sera représenté par celui d'un indice, par exemple le Nasdaq-Composite pour le marché des actions technologiques.

Nous modéliserons le problème de la façon suivante

$$R_i = b_{1i}F_1 + b_{2i}F_2 + \cdots + b_{ni}F_n + e_i \quad (1)$$

où R_i est le taux de rentabilité du fonds i, b_{ki} représente la sensibilité du fonds i par rapport à l'indice de référence F_k et e_i reprend tout ce qui n'a pas été expliqué par les indices de référence. Dans ce modèle nous supposerons que les différents e_i sont non-corrélés entre eux.

Le modèle (1) est similaire au modèle de régression linéaire. Ce modèle est constitué de deux parties (2) : le *style* qui est la portion de l'équation 1 permettant d'expliquer le taux de rentabilité du fonds comme celui d'un portefeuille passif, sans sélection d'actifs spécifiques, qui investit sur différents marchés, et la *sélection* qui est la partie dynamique du fonds, due à sa gestion active.

$$R_i = \underbrace{b_{1i}F_1 + b_{2i}F_2 + \cdots + b_{ni}F_n}_{\text{Style}} + \underbrace{e_i}_{\text{Sélection}}. \quad (2)$$

Si le taux de rentabilité du fonds est inférieur au taux de rentabilité du portefeuille passif, cela implique que le fonds a été mal géré. En effet, un investissement fixe sans aucune gestion dans les indices F_k représenté par les sensibilités b_{ij}, serait plus rentable.
En ajoutant la contrainte

$$\sum_j b_{ij} = 1, \quad (3)$$

nous pouvons interpréter les coefficients b_{ij} comme les pourcentages d'un portefeuille investi dans chaque type de marché; ceci constitue le *style* d'un fonds.

L'équation (1) doit être vérifiée à tous les instants t pour lesquels R_i est calculé. Nous obtenons donc l système d'équations (4):

$$R_i(t) = b_{1i}F_1(t) + b_{2i}F_2(t) + \cdots + b_{ni}F_n(t) + e_i(t) \quad (4)$$

avec $1 \le t \le T$. Sachant que les $e_i(t)$ sont inconnus, il existe une infinité de solutions au système (4) (pour autant que $T>n$) ; la solution retenue sera celle qui minimise la somme des carrés des $e_i(t)$ [1], afin de mettre en évidence la relation liant le taux de rentabilité R_i au style.

Cette solution peut néanmoins comprendre certains termes b_{ij} négatifs. Ceci implique une corrélation négative entre le taux de rentabilité du fonds et celui de l'indice, ce qui, en termes financiers, est une vente à découvert. Le gestionnaire peut vendre, à un prix fixé à l'avance, une promesse de vente dans le futur pour des actifs qu'il ne possède pas. Il s'engage donc à acheter un actif, à un prix fixé, pendant une certaine période. Si le prix de l'actif monte, l'acheteur de la promesse exercera son droit d'achat et le gestionnaire sera obligé d'acheter les actifs sur le marché à un prix plus élevé qu'il ne les vend. Il fera donc de la perte et ceci explique la corrélation négative. Dans le cas où le prix descend, le détenteur n'exercera pas son droit d'achat car il pourra acheter l'actif à moindre prix sur le marché et le gestionnaire empochera la prime de vente. Pour un grand nombre de fonds, ce type de pratique n'est pas permis pour les gestionnaires (ou fait l'objet de limites très strictes), car il peut amener à des prises de risque très importantes, ce qui n'est pas nécessairement l'objectif d'un fonds commun. Nous n'envisagerons pas ce cas et par conséquent nous imposerons également que

$$b_{ij} \ge 0, \forall i, j\,. \quad (5)$$

Finalement, le modèle obtenu est celui formé par les équations (4), sous les contraintes (3) et (5).

La solution de ce système relève des techniques classiques d'algèbre linéaire (voir par exemple [5]). Néanmoins, nous pouvons mentionner une spécificité liée à l'utilisation de ces techniques dans notre problème spécifique.

L'analyse de style présentée ci-dessus n'a de sens que lorsque les taux de rentabilité des indices ne peuvent pas être expliqués en fonction des autres indices sélectionnés [4]. Ce problème est bien connu dans le cadre général de la régression sous le nom de multi-collinéarité. Intuitivement ceci veut dire que si les indices que nous utilisons sont fortement corrélés alors leur influence individuelle peut ne pas ressortir lors de la régression. Par exemple, si nous considérons deux indices identiques F_1 et F_2 et que nous tentons d'expliquer le taux de rentabilité du fonds en fonction de ces indices: $y = b_0 + b_1 F_1 + b_2 F_2$, alors les coefficients b_1 et b_2 ne sont pas déterminés de manière univoque. En effet, comme $F_1 = F_2 = F$ nous pouvons écrire $y = b_0 + (b_1 + b_2) F$. Toutes les paires (b_1, b_2) telles que $b_1 + b_2 = c$ où c est une constante sont des solutions.

Nous pouvons généraliser ce cas en considérant que les indices ne sont pas identiques mais fortement corrélés. La source du problème est la variance du coefficient b_k ; cette variance mesure la précision de l'estimation du coefficient b_k lorsque les entrées du système (indices F_j) varient légèrement. Si la variance de b_k est grande, les estimations présenteront une grande déviation par rapport à la vraie valeur du paramètre que nous tentons d'estimer.

En pratique, la colinéarité, c'est à dire la corrélation entre les différents indices boursiers, est très importante. Il est donc nécessaire de remédier à ce problème afin d'obtenir des estimateurs fiables, c'est à dire à variance faible, des coefficients b_{ij} de la régression. Dans le cas contraire, la classification opérée sur les b_{ij} serait sans intérêt.

Pour remédier à ce problème, il faut donc remplacer l'ensemble des indices F_j par un autre ensemble d'indices G_k ($1 \le k \le m$, $m < n$) qui soient moins corrélés. A cette fin, nous utiliserons une technique statistique classique, l'Analyse en Composantes Principales (PCA) (voir par exemple [6]).

La PCA consiste en une projection linéaire d'un ensemble de données de dimension n sur un sous-espace de dimension m, $m<n$. Les axes de ce sous-espace sont choisis de telle manière à conserver le plus possible l'information contenue dans les données de départ. Cette information est mesurée par la variance des données, qu'il convient donc de maximiser après projection. Il est bien connu que ceci est équivalent à minimiser l'erreur de reconstruction, c'est à dire le déplacement des données dans l'espace de départ après que celles-ci aient été projetées et reconstruites à partir de la projection.

Grâce à la PCA, nous disposons donc d'un nouvel ensemble d'indices G_k qui peuvent être substitués aux indices F_j dans le système d'équations (4) pour obtenir

$$R_i(t) = c_{1i} G_1(t) + c_{2i} G_2(t) + \cdots + c_{mi} G_m(t) + e_i(t). \qquad (6)$$

Notons qu'à partir des coefficients c_{ki} obtenus en résolvant le système (5), il est possible de calculer un ensemble de coefficients b'_{ji} similaires aux n coefficients b_{ji}, c'est à dire des coefficients de régression par rapport aux indices originaux F_i, en opérant la transformation inverse de la PCA.

Il faut néanmoins mentionner que les indices G_k n'ayant pas la signification « physique » qu'avaient les indices F_j, le raisonnement qui nous a conduit ci-dessus à imposer une contrainte de positivité sur les coefficients b_{ji} ne tient plus lorsqu'on parle des coefficients c_{ki}. En utilisant une projection par PCA, nous perdons donc l'avantage des contraintes de positivité imposées par l'équation (5). En contrepartie, les nouveaux coefficients b'_{ji} obtenus auront une variance plus faible, ce qui permettra une classification fiable des fonds d'investissement.

3. Données et résultat de l'analyse de styles

L'analyse de style décrite plus haut a été effectuée sur des fonds d'investissement provenant de la base CRSP (Center of Research in Security Prices) de l'Université de Chicago.
Le choix de 33 indices F_i a été effectué par similitude avec le raisonnement de Sharpe [1] : croissance, petite/moyenne/grande capitalisation et taux intérêt (long/moyen/court terme), ceci pour les les principales régions du monde ayant une influence considérable sur les marchés financiers : Etats Unis d'Amérique, Europe (Royaume Uni, France et Allemagne) et Japon. Le nombre élevé (33) d'indices choisis est justifié par la possibilité d'utiliser une projection (PCA) supprimant de manière automatique et objective les colinéarités. Ces 33 indices seront détaillés dans la table 2.

Afin de pouvoir effectuer l'analyse de style, nous devons disposer des taux de rentabilités des fonds et des indices pour la même période (dans notre cas janvier 1993 à décembre 1998, soit 72 points de mesures mensuelles). Nous devons aussi disposer d'une classification de référence pour ces fonds. Au total, 5822 fonds ont été analysés.

Dans la base de données CRSP, une stratégie est attribuée à chaque fonds ; ceci constitue notre classification de référence. Celle-ci a été établie par l'ICDI (Investment Company Data, Inc.) en collaboration avec le Standard & Poor's Fund Services. Au total 24 stratégies types ont été choisies par cette institution ; celles-ci sont reprises au tableau 1.

Stratégie		
Aggressive Growth	Government Securities	Taxable Money Market
Balanced	International Equities	High Quality Municipal Bonds
High Quality Bonds	Income	Option Income
High Yield Bonds	Long-Term Growth	Precious Metals
Global Bonds	Tax-Free Money Market	Sector Funds
Global Equity	Gov Securities Money Market	Special Funds
Growth & Income	High Quality Municipal Bonds	Total Return
Ginnie Mae Funds	Single-State Municipal Bonds	Utility Funds

Table 1 : ensemble des 24 stratégies adoptées par l'ICDI et le Standard & Poor.

Une PCA a été effectuée sur les 33 indices boursiers F_j ; 17 indices après projection G_k ont été retenus, ce qui correspond à une conservation de 99.88% de la variance initiale des indices F_i. La résolution du système (6) nous permet alors de calculer les sensibilités c_{ki}. Comme décrit dans la section précédente, la transformation inverse de la PCA nous permet de revenir dans l'espace des indices initiaux F_i, et donc de calculer les sensibilités b'_{ji}. Un exemple de telles sensibilités, pour le fonds Vanguard/Trustees Equity Fund : US Portfolio (également étudié par Sharpe [1]), est donné au tableau 2.

4. Classification

Les 5822 vecteurs de sensibilités b'_{ji} (de dimension 33) sont classés par l'utilisation successive des cartes auto-organisées de Kohonen et de la classification hiérarchique de Ward [7].

Les cartes de Kohonen [8] effectuent une double opération. Tout d'abord, elles résument l'information contenue dans un ensemble important de données (dans notre cas 5822 vecteurs) ; cette opération porte le nom de quantification vectorielle. Le résultat de la quantification vectorielle est un ensemble discret de vecteurs appelés centroïdes (dont le

Indice	Nom de l'indice	Estimateur b'_j
0		0.0052
1	Dow Jones 30	0.1156
2	Lehman Brothers' US Credit Bond Index	0.0322
3	Lehman Brothers' Intermediate-term Government Bond Index	0.0165
4	Lehman Brothers' Long-term Government Bond Index	0.0448
5	Lehman Brothers' Mortgage-Backed Securities Index	0.0206
6	Lehman Brothers' 1-3 month Treasury Bill	0.0001
7	Nasdaq 100	0.1558
8	Salomon Brothers' Non-US Government Bond Index	0.0055
9	S&P400 Medium Capitalization	0.1042
10	S&P500	0.1139
11	S&P500 Sharpe / BARRA Growth Index	0.1123
12	S&P500 Sharpe / BARRA Value Index	0.1046
13	S&P600 Small Capitalization	0.1032
14	FTSE100	0.0690
15	FTSE250 Growth	0.0101
16	FTSE250 Value	0.0174
17	FTSE Small Capitalization	0.0044
18	UK Bank Bills 3 month	-0.0027
19	UK 2 year Government Index	0.0336
20	UK 10ear Government Index	-0.0200
21	TOPIX100	0.0326
22	TOPIX400 Medium Capitalization	-0.0267
23	Japan Gensaki Bond Reference 3 month	0.0038
24	Japan Benchmark 2 year Government Index	-0.0048
25	Japan Benchmark 10 year Government Index	-0.0605
26	CAC40	-0.0057
27	SBF250	0.0558
28	France Money Market 3 month	0.0357
29	France Benchmark 2 year Government	-0.0012
30	France Benchmark 10 year Government	0.0123
31	DAX30	0.0581
32	Germany Money Market 3 month	0.0016
33	Germany Money Market 10 year	-0.0508

Table 2 : ensemble des indices boursiers et des sensibilités b'_j pour le fonds Vanguard/Trustees Equity Fund : US Portfolio

nombre est faible par rapport au nombre de données initiales). A chacun des centroïdes est associé un certain nombre de données initiales, selon la règle du plus proche voisin. Les zones de l'espace ainsi associées à chaque centroïde (appelées zones de Voronoï) forment la partition de la classification.

Par rapport à d'autres méthodes de quantification vectorielle, les cartes de Kohonen réalisent une seconde opération. Celle-ci, connue sous le nom de propriété de conservation de la topologie, permet de représenter les centroïdes dans un tableau à une ou deux dimensions (souvent deux), de telle manière que deux centroïdes proches dans le tableau sont également proches dans l'espace des données. Cette propriété permet de représenter les centroïdes dans le tableau (appelé carte de Kohonen) ; deux centroïdes proches seront donc semblables.

La figure 1 représente une carte de Kohonen comprenant 100 centroïdes. Chacun de ceux-ci est un vecteur de dimension 33, dont les composantes sont représentées sous la forme d'une courbe de 33 points (ordonnés de façon arbitraire selon leur numérotation). Ces 100 centroïdes proviennent de la quantification des 5822 vecteurs de sensibilités b'_{ji}.

ACSEG'2002 proceedings - Connectionist Approaches in Economics and Management Sciences
Boulogne-sur-Mer (France), 21-22 November 2002, pp. 201-212

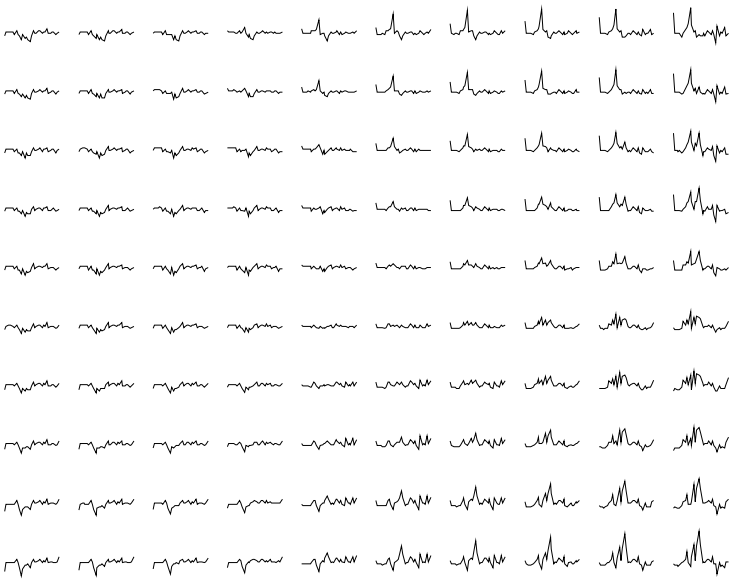


Figure 1 : carte de Kohonen construite sur les 5822 vecteurs de sensibilités b'_{ji}.

Pour réduire le nombre de classes, l'algorithme de Ward [6, 9] est utilisé sur ces centroïdes. L'algorithme de Ward est un algorithme de classification hiérarchique. A chaque étape de l'algorithme deux centroïdes sont regroupés selon un critère de distance.

Définissons l'inertie I_l ($1 \le l \le P$ = nombre de classes) d'une classe (zone de Voronoï) comme étant la moyenne des distances au carré entre les données de cette classe et son centroïde (l'inertie est donc une estimation de la variance des données autour du centroide, à l'intérieur d'une classe). L'inertie intra-classes I_W est définie comme :

$$I_W = \frac{1}{N} \sum_{l=1}^{P} N_l I_l \ , \qquad (7)$$

où N_l est le nombre de données associées à la classe l et N le nombre total de données (5822 dans notre cas).

L'algorithme de Ward fusionne à chaque étape les deux classes qui feront le moins augmenter l'inertie I_W. Cette opération est répétée un certain nombre de fois, jusqu'à l'obtention du nombre désiré de classes. A chaque itération, il est possible de mesurer l'augmentation de l'inertie intra-classe I_W, ce qui permet de se fixer un critère d'arrêt de l'algorithme. Les classes ainsi obtenues par regroupement seront appelées macro-classes.

L'algorithme de Ward va regrouper préférentiellement des centroïdes similaires. Pour rappel, les propriétés topologiques des cartes de Kohonen font que des centroïdes similaires se retrouvent proches sur la carte. Les macro-classes ainsi obtenues forment donc des zones homogènes sur la carte. La figure 2 représente le résultat de l'algorithme de Ward avec 20 macro-classes, sur les centroïdes de la figure 1.

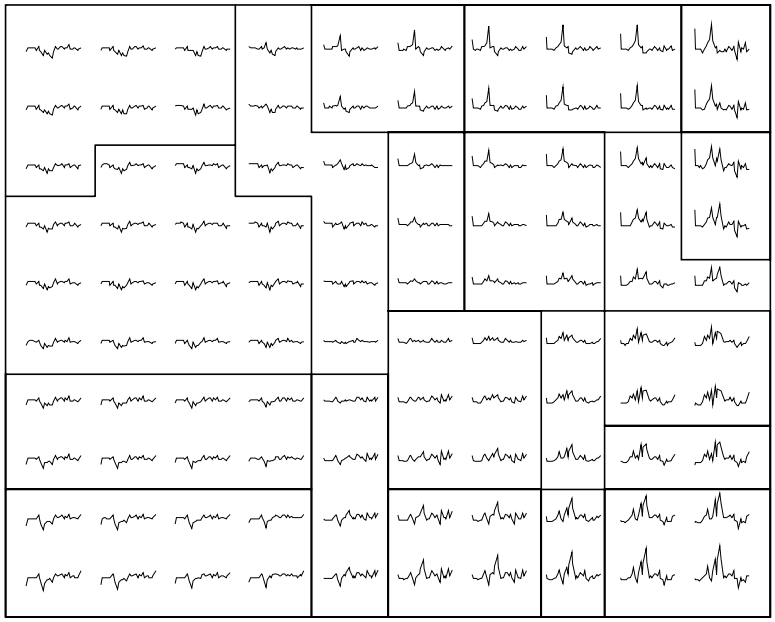


Figure 2 : macro-classes résultant de l'algorithme de Ward appliqué aux centroïdes de la figure 1.

La classification obtenue peut être comparée à celle de référence. Dans le premier cas (Kohonen + Ward), 20 classes résultent de la procédure. Dans le second, 24 classes ont été établies par l'ICDI et le S&P. On peut donc établir le tableau de contingence entre ces deux classifications ; la table 3 donne, pour chacun des macro-classes C (Kohonen+Ward), le pourcentage de fonds appartenant aux classes S (S&P). A titre d'exemple, la figure 3 montre, pour deux catégories spécifiques de la classification S&P (S16 –MS- Single-State Municipal Bond Funds et S12 –LG- Long-Term Growth Funds), le pourcentage de fonds classés dans chaque catégorie de notre classification.

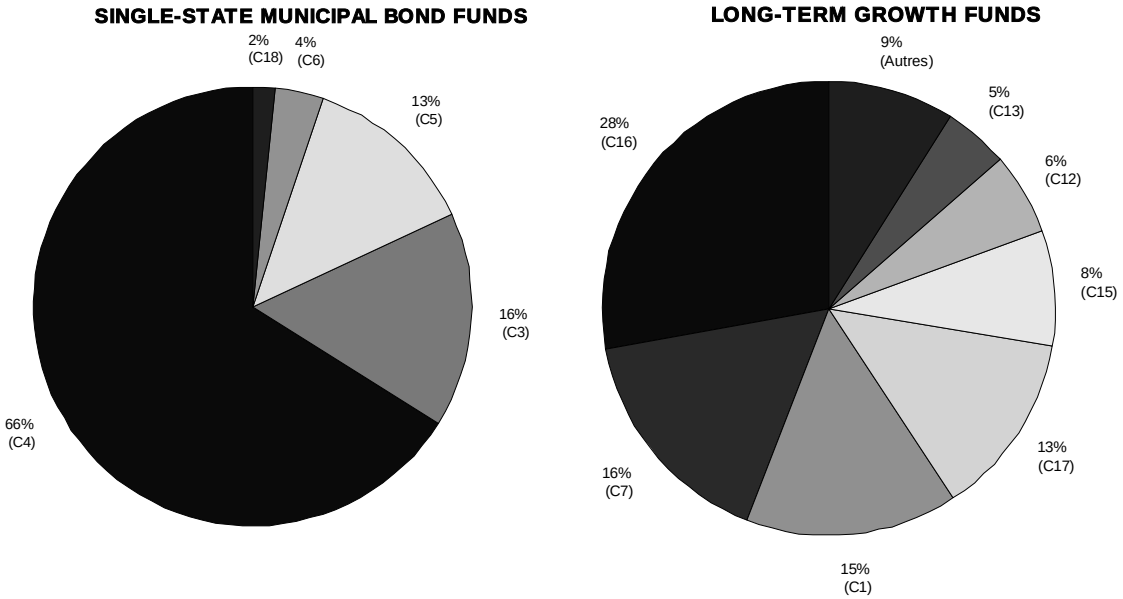


Figure 3 : pourcentage de fonds classés dans chaque catégorie de notre classification, pour deux catégories spécifiques du S&P..

%	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
S1 (AG)	3.8	0.0	0.0	0.0	0.0	0.6	41.4	72.1	2.8	0.0
S2 (BL)	4.1	40.1	0.3	0.0	0.0	0.5	0.0	0.0	5.5	0.0
S3 (BQ)	0.0	0.0	33.0	6.8	18.5	16.9	0.0	0.0	0.4	0.0
S4 (BY)	0.0	0.6	0.0	0.2	2.0	15.1	0.0	0.0	0.0	5.5
S5 (GB)	0.0	0.6	1.3	0.7	10.3	8.8	0.0	0.0	0.0	7.4
S6 (GE)	0.6	0.6	0.0	0.0	0.0	0.3	0.4	2.3	29.3	33.3
S7 (GI)	38.1	7.5	0.0	0.0	0.0	0.2	1.7	1.5	3.6	0.0
S8 (GM)	0.0	0.0	6.3	1.1	10.9	6.5	0.0	0.0	0.0	0.0
S9 (GS)	0.0	0.0	21.5	2.8	20.0	10.3	0.0	0.0	0.0	0.0
S10 (IE)	0.0	0.6	0.8	0.0	0.0	1.7	0.4	0.0	35.9	35.2
S11 (IN)	17.9	11.6	0.3	0.0	0.0	0.5	0.4	0.0	2.4	1.9
S12 (LG)	27.4	1.2	0.0	0.0	0.0	0.3	44.0	20.2	4.7	0.0
S13 (MF)	0.0	0.0	0.0	0.0	0.0	7.5	0.0	0.0	0.0	0.0
S14 (MG)	0.0	0.0	0.0	0.0	0.0	7.9	0.0	0.0	0.0	0.0
S15 (MQ)	0.0	0.0	7.7	19.6	13.6	6.3	0.0	0.0	0.0	0.0
S16 (MS)	0.0	0.0	27.2	66.2	23.1	6.5	0.0	0.0	0.0	0.0
S17 (MT)	0.0	0.0	0.0	0.0	0.0	8.1	0.0	0.0	0.0	0.0
S18 (MY)	0.0	0.0	0.8	2.4	1.2	0.2	0.0	0.0	0.0	0.0
S19 (OI)	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0
S20 (PM)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S21 (SF)	2.9	0.0	0.0	0.0	0.0	0.0	11.7	3.9	2.4	0.0
S22 (SP)	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0
S23 (TR)	3.5	18.6	0.8	0.2	0.4	1.5	0.0	0.0	11.8	16.7
S24 (UT)	1.7	18.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0

%	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
S1 (AG)	0.0	9.4	64.1	26.2	27.5	1.7	5	0.7	2.0	0.5
S2 (BL)	0.0	0.0	0.4	2.5	0.7	0.7	28.6	13.0	0.0	0.0
S3 (BQ)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	0.0	0.0
S4 (BY)	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.7	0.0	0.5
S5 (GB)	6.3	3.8	1.6	8.7	0.7	0.0	0.4	4.8	0.0	0.0
S6 (GE)	1.0	4.7	0.0	18.8	4.9	0.2	2.0	4.8	0.0	19.1
S7 (GI)	3.2	24.5	1.6	2.5	8.5	45.9	12.9	2.1	0.0	0.0
S8 (GM)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S9 (GS)	0.0	0.0	0.0	0.0	0.0	0.4	0.0	8.9	0.0	0.0
S10 (IE)	81.0	0.9	1.2	17.5	0.0	1.0	0.0	1.4	89.8	78.6
S11 (IN)	0.0	6.6	0.0	0.0	6.3	3.4	2.4	0.0	0.0	0.0
S12 (LG)	4.2	34.9	11.7	11.3	35.9	42.5	32.6	0.0	4.1	0.0
S13 (MF)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S14 (MG)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S15 (MQ)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	0.0	0.0
S16 (MS)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.6	0.0	0.0
S17 (MT)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S18 (MY)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0
S19 (OI)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S20 (PM)	0.0	0.0	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S21 (SF)	1.1	9.5	17.4	11.3	0.0	2.2	2.4	0.0	4.1	0.9
S22 (SP)	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0
S23 (TR)	0.0	1.0	0.0	0.0	14.1	1.5	13.7	17.0	0.0	0.0
S24 (UT)	3.2	4.7	0.4	0.0	0.0	0.5	0.0	16.4	0.0	0.4

Table 3 : tableau de contingence entre la classification de Kohonen+Ward (20 classes) et celle de référence.(24 classes).

5. Comparaison des classifications et conclusions

Notre travail présente une méthode objective de classification de fonds d'investissement. Celle-ci se base uniquement sur l'évolution du taux de rentabilité des fonds, sans utiliser la

moindre information provenant de la stratégie annoncée par le gestionnaire, cette dernière pouvant être influencée par les objectifs propres du gestionnaire, désirant de classer son fonds parmi d'autres ayant des performances moindres.

Les résultats de notre classification conduit à des résultats différents que ceux d'une classification de référence, à savoir celle de l'ICDI et du S&P. D'un strict point de vue de la classification, l'inertie intra-classe I_W (7) a été calculée pour chacune des deux classifications : 0.07 pour notre classification et 0.13 pour la classification de référence (soit un gain de près de 50% !), malgré un nombre de classes légèrement inférieur dans notre cas. Notre classification regroupe donc clairement mieux les fonds semblables, et est donc plus cohérente.

La table 3 et la figure 3 montrent de plus des différences relativement importantes entre les deux classifications. Ceci peut être dû aux raisons suivantes.

- La classification de référence pourrait se baser sur des méthodes de classification moins sophistiquées. En particulier, notre expérience a révélé que l'utilisation d'un seul des deux algorithmes (Kohonen et Ward) donnait de moins bons résultats (en termes d'inertie intra-classe I_W) que leur combinaison, le premier pour une première classification grossière, le second pour un affinement en un nombre réduit de classes.
- La stratégie du gestionnaire pourrait, dans un certain nombre de cas, avoir été modifiée au cours de la période étudiée. Les deux classifications, se basant sur des périodes différentes, pourraient donc amener des conclusions différentes.
- La classification de référence pourrait se baser, dans une trop grande mesure, sur des informations communiquées par les gestionnaires. Dans ce cas , notre classification permettrait d'avoir un avis plus objectif sur la stratégie réelle de gestion du fonds.

Dans tous les cas, les avantages d'une classification plus cohérente devraient permettre aux investisseurs de prendre en compte une information différente, complémentaire à celle des classifications de référence habituellement utilisées.

Remerciements

Michel Verleysen est Maître de Recherches du Fonds National de la Recherche Scientifique (FNRS) belge. Une partie des résultats présentés dans ce papier a été financée par le Pôle d'Attraction Inter-Universitaire (PAI), initié par l'état belge, ministère des sciences, des technologies et de la culture. La responsabilité scientifique est laissée aux auteurs.

Références

1. Sharpe F. W., *Asset Allocation: management style an performance measurement*. Journal of Portfolio Management, 1992. **46**: p. 7-19.
2. Investment Company Institute, *Fundamentals*, 1999, http://www.ici.org/pdf/fm-v8n1.pdf.
3. Kim T.-H., Stone D. and Tomas M., *Mutual fund objective misclassification.* Journal of Economics and Business, 2000. **52**: p. 309-323.
4. Dibartolomo D. and Witowski E., *Mutual fund misclassification : Evidence based on style analysis.* Financial Analyst Journal, 1997: p. 32-43.
5. Liew C. K., *Inequality constrained least squares estimation.* Journal of the American Statistical Association, 1976. **71**: p. 746-751.
6. Saporta G., *Probabilités, Analyse des Données et statistiques.* 1990, Paris: Editions Technip.
7. Cottrell M., Girard B. and Rousset P, *Forecasting of Curves using a Kohonen classification.* Journal of Forecasting, 1998. **17**: p. 5-6.
8. Kohonen, T., *Self-organizing maps*. Springer series in information sciences ; 30. 1995, Berlin: Springer,. xv, 362.
9. Ward J. H., *Hierarchical grouping to optimize an objective function.* Journal of the American Statistical Association, 1963. **58**: p. 236-244.