Lo A., Mamaysky H., Wang J., "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation", *The Journal of Finance*, Vol. 55, No. 4, 2000.

Markovitz H., *Portfolio Selection: Efficient Diversification of Investments*, John Wiley, New York, 1959.

Mordecai Kurtz, Motolese Maurizio, "Endogenous Uncertainty and Market Volatility", Working paper, Stanford University, 1999.

McCullough B., Renfro C., "Benchmarks and Software Standards: A Case Study of GARCH Procedures", *Journal of Economic and Social Measurement*, 1999.

Nelson D.B., "Stationarity and persistence in the GARCH(1,1) model", *Econometric Theory* Vol. 6, 1990, pp. 318-334.

Refenes A.P. (ed), *Neural Networks in the Capital Markets*, John Wiley & Sons, New York, 1995.

Sharpe, W. F., "Captial asset prices: A theory of market equilibrium under conditions of risk", *Journal of Finance*, 19, 1964, pp. 425-442.

Scholes M., Williams J., "Estimating Betas from Nonsynchronous Data", *Journal of Financial Economics*, 5, 1977, pp. 309-327.

Sharpe W., "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance*, 19, 1964, pp. 425-442

Sullivan R., Timmerman A. White H., "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap", *The Journal of Finance*, Vol. 54, No. 5, 1999, pp. 1647-1691.

Volterra V., *Theory of Functionals and of Integro-Differential Equations*, Dover, New York, 1959.

Wang K. Q., "Asset Pricing with Conditioning Information: A New Test", *Journal of Finance*, forthcoming.

White H. Maximum Likelihood Estimation of Mispecified Models, *Econometrica*, 53, pp. 1-16.

Self-Organizing Feature Maps for the Classification of Investment Funds

A. Lendasse*— P. Cardon* — V. Wertz* — Eric de Bodt**
M. Verleysen***

* *Université catholique de Louvain, CESAME*
4 av. G. Lemaître
B-1348 Louvain-la-Neuve

{Lendasse ; wertz}@auto.ucl.ac.be

** *Université catholique de Louvain, IAG*
1, pl. des Doyens
B-1348 Louvain-la-Neuve

debodt@fin.ucl.ac.be

*** *Université catholique de Louvain, Département d'Electricité*
3 pl. du Levant
B-1348 Louvain-la-Neuve

verleysen@dice.ucl.ac.be

ABSTRACT. *The ranking of investment funds regularly carried out by the financial press is a major concern for investors. In this context, there exists a serious risk of distortion between the announced strategy and the actual strategy implemented by the managers of the funds. Indeed these managers could try to have the funds they manage classified in a category that does not reflect the reality of their strategy, in order to be favourably compared to the performances of other funds in the same category. Our work shows how an independent classification based on the style analysis, originally introduced by Sharpe in 1992 (Sharpe 1997), can be built. This leads to a classification of the funds based on characteristics extracted from their rates of returns and less prone to any manipulation. The proposed classification is compared to a reference classification (ICDI and S&P). The analyses of the differences between the obtained results and, in particular, of their origin speak for themselves.*

KEYWORDS: *Classification, Mutual Fund, Investment Funds, Kohonen Maps, Ward.*

1. Introduction

An investment fund (or mutual fund) is an investment structure collecting money coming from investors 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.

Concretely, the mutual fund is a very useful tool for non experienced investor that allows combining diversification and thematic investment. Of course, in investing his money in these funds, this non experienced investor hopes to increase the invested capital.

In (Sharpe, 1997) it is estimated that 44.4 millions of US families (more than 45%) own mutual funds. Today, more than 766 billions of dollars are invested in mutual funds. If these funds are very popular, this is because they are accessible, easy to buy and to sell (they are liquid) and well diversified (the majority of these funds are made of dozens of different placements). For the non experienced investor, the investment in a mutual fund is then an easy and accessible way to access to placements that should be in other cases only accessible to professionals.

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 (Sharpe, 1997): 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 the real investment strategy is (a classical case of moral hazard situation, well-know from economists). This makes the task of the investor difficult, 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 additional information to the investors. Nevertheless, some studies showed that specific funds are wrongly classified (Investment Company Institute, 1999; Kim, 2000). In this work, we establish a classification of investment funds, based only on correlations between funds' returns and a set of market indexes, without using any information from the fund manager. This classification should allow one to detect the real investment strategy of 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 (Sharpe, 1997) under the name of *style analysis*. To build an independent classification based on these features, we use Kohonen self-organizing maps.

The extrasset of characteristics will be developed in section 2, style analysis. Section 3 will detail the data that are used and the results with style analysis. Section 4 will present the classification done by Kohonen Maps and will compare it to an existent classification.

The study is illustrated using 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.

2. Style analysis

To avoid any influence of funds' managers on the generated classification, we base our work only on the historic evolution of their rate of return. We do not perform the classification on the return values themselves, a procedure which would be clearly problematic. Indeed, a fund of Swedish assets could have an historic evolution of its rate of return which is close to the evolution of the Nikei index, without any investments in Japanese assets in this fund, simply by chance or because of very sample specific conditions. Hence, we build the classification using statistical indicators that characterize the management style performed by the manager. This extras set of characteristics is based on the works of W.F. Sharpe (Sharpe, 1997) and allows to really take into account the nature of the manager's investment strategy.

Following the Sharpe style analysis, we express the fund's rate of returns i as a linear combination of a set of reference market indexes, where the fund could have been invested. The intuition behind this procedure is the correlation that should exist between the fund's rate of returns and the market indexes rates of return where the fund is invested. For example, if the fund manager has invested the large part of its money in US stock, we should find a large correlation between the fund's rate of return and US stock market indexes (such as the S&P500, the DJ30, etc.).

This idea leads to the following model:

$$R_i = b_{1i}F_1 + b_{2i}F_2 + \cdots + b_{ni}F_n + e_i \qquad [1]$$

where R_i is the rate of return of the fund i, b_{ki} represents the sensitivity of the fund i with respects to the reference index F_k and e_i summarizes everything that has not been explained by the reference indices. In this model, we will suppose that the different e_i are not cross-sectional correlated.

Model [1] is similar to a linear regression problem. This model is made of two parts [2]: the *style* which is the part of equation 1 that express the return as a linear combination of the set of reference indexes and the *selection* term which is the dynamic part of the fund, due to its active management (reflecting stock picking).

$$R_i = \underbrace{b_{1i}F_1 + b_{2i}F_2 + \ldots + b_{ni}F_n}_{style} + \underbrace{e_i}_{selection} . \qquad [2]$$

If we add the constraint

$$\sum_j b_{ij} = 1, \qquad [3]$$

we can interpret the b_{ij} coefficients as the percentage invested in each king of market. This is the style of the fund.

Equation [1] can be estimated on historical data (under the standard assumption of stationary process). This leads to the following reformulation:

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

with $1 \le t \le T$ and t representing the time subscript. Estimation could be realized by ordinary least square. This would however potentially lead to some negative coefficients b_{ij}. Such a result would imply a negative correlation between the fund's rate of return and the index one which appears in case of short sales. A short sale is the operation by which an investor is selling, at a price determined today, an asset that it does not posses (the investor has to borrow it). He will then have to buy back the asset at a future date in order to reimburse the lender. Such a strategy, justified by anticipation of declining future prices, is clearly highly risky and strictly regulated (in many case, forbidden) for public investment funds. To take this fact into account, we will also impose

$$b_{ij} \ge 0, \forall i, j \qquad [5]$$

Finally, the model obtained is that formed by the equations [4], under the constraints [3] and [5].

The solution of this system uses the traditional techniques of linear algebra (see for example (Liew, 1976)).

The style analysis presented above makes sense only when the rates of return of the reference set of indexes are not too much correlated to each others (Dibartolomo *et al.*, 1997). This problem is well-known within the general framework of the regression under the name of multi-colinearity. If the correlations are too strong, it leads to unstable coefficients' estimates: the variance of the estimated coefficient b_k becomes very high. As this variance measures the precision of the coefficients' estimates, it does mean there is a high risk of great deviation compared to the true value of the parameter which we try to estimate.

In practice, colinearity, *i.e.* the correlation between the various indexes of security prices, is very important. It is thus necessary to solve this problem in order to obtain reliable estimators, *i.e.* small variance, of the b_{ij} regression coefficients. Otherwise, the classification operated on b_{ij} would be without interest.

To solve this problem, it is thus necessary to replace the whole set of the indices F_j by another set of indices G_k ($1 \le k \le m$, $m < n$) which is less correlated. For this purpose, we will use a traditional statistical technique, the Principal Components Analysis (PCA) (see for example (Saporta, 1990)).

The PCA consists in a linear projection of a set of data of dimension n on a subspace of dimension m, $m<n$. The axes of this subspace are selected in such manner as to keep as much as possible the information contained in the initial data. This information is measured by the variance of the data. It is well-known that this is equivalent to minimizing the reconstruction error, *i.e.* the displacements of the data in the initial space after those were projected and reconstructed from the projection.

Thanks to the PCA, we thus have a new set of indices G_k who can be substituted for the indices F_j in the system of equations [4] to obtain

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

Let us note that starting from the coefficients c_{ki} obtained by estimating the system [6], it is possible to express b_{ji} coefficients, i.e. coefficients of the regression with respect to the initial indices F_i, by operating the reverse PCA transformation.

It should nevertheless be mentioned that the G_k indices have lost the economic significance of the F_j indices. Specifically, the reasons which led us to impose a positivity constraint on the b_{ji} coefficients does not hold any more when one deals with the c_{ki} coefficients. By using a projection by PCA, we thus loose the positivity constraints imposed by equation [5]. On the other hand, the new b_{ji} coefficients will have a smaller variance, which will allow a more reliable classification of the investment funds.

3. Data and result of the style analysis

The style analysis described above was carried out on investment funds coming from the database CRSP (Center of Research in Security Prices) of the University of Chicago.

The choice of 33 indices F_i was carried out by similarity with the reasoning of Sharpe (1997): growth, small/middle/big capitalization and interest rate (long/middle/court term), this for the main areas of the world having a considerable influence on the financial markets: The United States of America, Europe (United indices is justified by the possibility of using a projection (PCA) in order to remove the colinearities in an automatic and objective way. These 33 indices are detailed in Table 2.

In order to be able to carry out the style analysis, we must have the rates of return of the funds and the indices for the same period (in our case January 1993 to December 1998, *i.e.* 72 points of monthly measurements). We must also have a classification of reference for these funds. In total, 5822 funds were analyzed.

In the CRSP database, a strategy is allotted to each fund; it constitutes our reference classification. This one was established by the ICDI (Investment Company Data, Inc.) in collaboration with the Standard & Poor' S Fund Services. In total, 24 standard strategies were chosen by this institution; they are reproduced in Table 1.

Table 1. *The 24 strategies adopted by the ICDI and the Standard & Poor*

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

A PCA was carried out on the 33 indexes of security prices F_j ; after projection, 17 indices G_k were retained, which corresponds to keeping 99.88% of the initial variance of the F_i indices.

Equation [6] then enables us to calculate the c_{ki} sensitivities. As described in the preceding section, the reverse PCA transformation enables us to go back to the space of the initial F_i indices, and thus to express the sensitivities b_{ji}. An example of such sensitivities, for the funds Vanguard/Trustees Equity Fund: US Portfolio (also studied by Sharpe (Sharpe, 1997)), is given in Table 2.

4. Classification

The 5822 vectors of sensitivities b_{ji} (of dimension 33) are classified by the successive use of the Kohonen Maps and the hierarchical classification of Ward (Cottrell *et al.*, 1998).

The Kohonen Maps (Kohonen, 1995) carry out a double operation. First of all, they summarize the information contained in an important dataset (in our case 5822 vectors); this operation is named vector quantization. The result of the vector quantization is a small set of vectors called centroids (their number is small compared to the number of initial data). To each centroid a certain number of initial data is associated, according to the rule of nearest neighbour. The zones of space thus associated to each centroid (called Voronoï zones) form the partition of the classification.

Table 2. *Set of indexes of security prices and sensitivities bj for the funds Vanguard/Trustees Equity Fund: US Portfolio*

Index	Index name	b_j *estimator*
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

Compared to other vector quantization methods, the Kohonen Maps carry out one second operation. This one, known by the name of topology conservation, makes it possible to represent the centroids in a table with one or two dimensions (often two), in such a way that two close centroids in the table are also close the data space.

Figure 1 represents a Kohonen Map with 100 centroids. Each one of those is a vector of dimension 33, whose components are represented as a curve of 33 points (ordered arbitrarily according to their classification). These 100 centroids come from the quantization of the 5822 vectors of sensitivities b_{ji}.

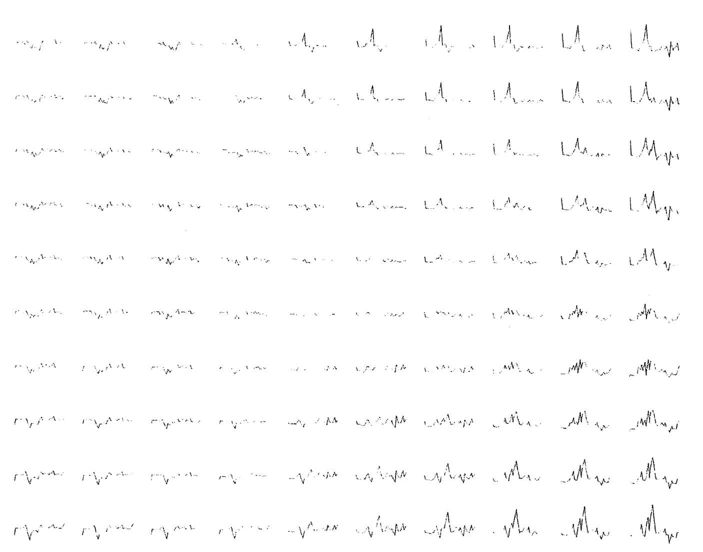

Figure 1. *Kohonen map built on the 5822 vectors of sensitivities bji*

To reduce the number of classes, the Ward algorithm (Saporta, 1990; Ward, 1963) is used on these centroids. The Ward algorithm is a hierarchical classification algorithm. At each stage of the algorithm two centroids are merged according to a distance criterion.

Let us define the inertia I_l ($1 \leq l \leq P$ = a number of classes) of a class (Voronoï zone) as being the average of the squared distances between the data of this class and its centroid (the inertia is thus an estimate of the variance of the data around the centroid, inside a class). The intraclass inertia I_W is defined as:

$$I_w = \frac{1}{N}\sum_{l=1}^{P} N_l I_l \qquad [7]$$

where N_l is the number of data associated with class l and N the total number of data (5822 in our case).

The Ward algorithm merges at each stage the two classes which will give the smallest increase to the interclass inertia I_W. This operation is repeated until obtaining the desired number of classes. At each iteration, it is possible to measure the intraclass inertia increase, which makes it possible to set a stopping criterion for the algorithm. The classes thus obtained by merging will be called macro-classes.

The Ward algorithm will merge the similar centroids preferentially. Recall that the topological properties of the Kohonen Maps are such that similar centroids are found close together on the Map. The obtained macro-classes thus form homogeneous zones on the chart. Figure 2 represents the result of the Ward algorithm with 20 macro-classes, on the centroids of figure 1.

The obtained classification can be compared with the reference one. In the first case (Kohonen + Ward), 20 classes result from the procedure. In the second, 24 classes were established by the ICDI and the S&P. One can thus draw up the table of contingency between these two classifications; Table 3 gives, for each macro-class C (Kohonen+Ward), the percentage of funds belonging to the classes S (S&P). As an example, Figure 3 shows, for two specific categories of the S&P classification (S16 - ms Single-State Municipal Jump Funds and S12 - LG-Long-Term Growth Funds), the percentage of funds classified in each category of our classification.

Our classification leads to results different than those from a reference classification, namely that of the ICDI and S&P. From a strict point of view of classification, intraclass inertia I_W (7) was calculated for each of the two classifications: we obtained 0.07 for our classification and 0.13 for the reference classification (*i.e.* a gain of almost 50%!), in spite of a slightly lower number of classes in our case. Our classification thus gathers clearly better the similar funds, and is thus more coherent.

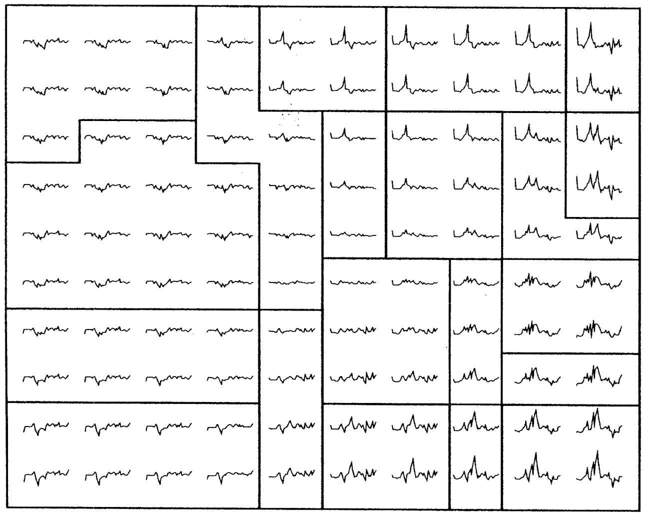

Figure 2. *Macro-classes resulting from the Ward algorithm applied to centroids of Figure 1*

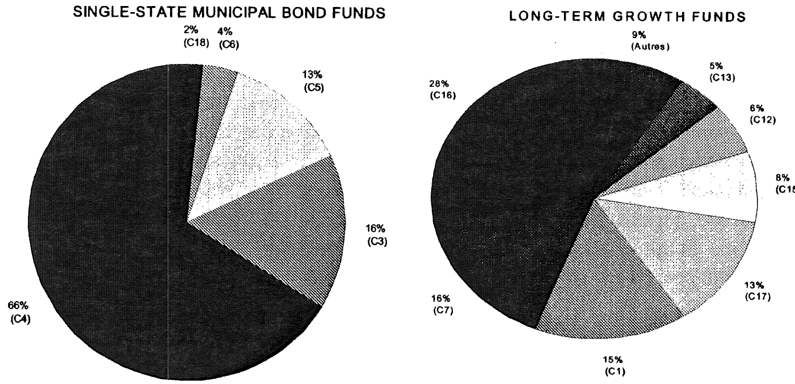

Figure 3. *Percentage of funds classified in each category of our classification, for two specific categories of the S&P*

Table 3. *Table of contingency enters the classification of Kohonen + Ward (20 classes) and the reference one (24 classes)*

%	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 and Figure 3 show moreover the relatively important differences between the two classifications. This can be explained by the following reasons:

– The reference classification could be based on less sophisticated classification methods. In particular, our experiment revealed that the use of only one of the two algorithms (Kohonen and Ward) gave worse results (in terms of intraclass inertia I_W) than their combination, the first for a first detailed classification, the second for a refinement in a reduced number of classes.

– In a certain number of cases, the strategy of the manager could be modified during the period of study. Two classifications, being based on different periods, could thus lead to different conclusions.

– The reference classification could be based, to a large extend, on the information communicated by the managers. In this case, our classification would make it possible to have a more objective opinion on the management strategy of the funds.

In all cases, the advantages of a more coherent classification should make it possible to the investors to take into account different information, in addition to that of reference classifications usually used.

5. Conclusions

Our work presents an objective classification method for investment funds. It is based only on the evolution of the rate of profitability of the funds, without using information coming from the strategy announced by the manager, the latter being potentially influenced by the proper objectives of the manager, wishing to classify his fund among others having weaker performances.

We test our propositions on a dataset composed by the monthly returns of 5822 investment funds, spanning the period 1993-1998. Our results clearly show that the standard reference classification (the Investment Company Data Inc. classification in our case) is clearly suboptimal. The more coherent classification that we obtain is the first step to a more objective ranking of funds by performance. To test the moral hazard hypothesis (are funds' managers manipulating the information that they provide on their investment strategy), it would be of interest to see if the misclassified funds are the over-performing ones.

Acknowledgements

Michel Verleysen is Senior research associate at the Belgian National Fund for Scientific Research (FNRS). The work of A. Lendasse and V. Wertz is supported by the Interuniversity Attraction Poles (IAP), initiated by the Belgian Federal State, Ministry of Sciences, Technologies and Culture. The scientific responsibility rests

6. Bibliography

Cottrell M., Girard B. and Rousset P, "Forecasting of Curves using a Kohonen classification", *Journal of Forecasting*, 17, 1998, p. 5-6.

Dibartolomo D. and Witowski E., "Mutual fund misclassification: Evidence based on style analysis", *Financial Analyst Journal*, 1997, p. 32-43Sharpe F. W., "Asset Allocation: management style an performance measurement", *Journal of Portfolio Management*, 46, 1992, p. 7-19.

Investment Company Institute, Fundamentals, 1999, http://www.ici.org/pdf/fm-v8n1.pdf.

Kim T.-H., Stone D. and Tomas M., "Mutual fund objective misclassification", *Journal of Economics and Business*, 52, 2000, p. 309-323.

Kohonen, T., *Self-organizing maps*, Springer series in information sciences, 30, Springer, Berlin, 1995.

Liew C. K., "Inequality constrained least squares estimation", *Journal of the American Statistical Association*, 71, 1976, p. 746-751.

Saporta G., *Probabilités, Analyse des Données et statistiques*, Editions Technip, Paris, 1990.

Ward J.H., "Hierarchical grouping to optimize an objective function", *Journal of the American Statistical Association*, 1963.

European Journal of Economic and Social Systems

Volume 17 - No. 1-2/2004

Neural Networks

Guest Editor

Patrice Gaubert

Hermes

Lavoisier

11, rue Lavoisier – F-75008 Paris
ISSN 1292-8895
ISBN 2-7462-0909-8
Head of publication : Patrick Fenouil

Lavoisier
14 rue de Provigny – F-94236 Cachan cedex
tél. : 33 (0)1 47 40 67 00
fax : 33 (0)1 47 40 67 02
ejess@lavoisier.fr

http://www.e-revues.com
http://ejess.e-revues.com
