



K-Means Clustering of Self-Organizing Maps: An Empirical Study on the Information Content of Self-Classification of Hedge Fund Managers

Marcus Deetz

FOM University of Applied Science, Germany

Abstract: With the implementation of the 2-step approach according to Vesanto & Alhoniemi (2000), this article extends the procedure of visual evaluation of the Kohonen Maps usually chosen in the hedge fund literature for classification with Self-Organizing Maps. It introduces an automated procedure which guarantees a consistent combination of adjacent output units and thus an objective classification. The practical application of this method results in a reduction of the strategy groups specified by the database. This is also accompanied by a significant reduction in the Davies Bouldin Index (DBI) of the SOM partitions. Since a small dispersion within the clusters and large distances between the clusters lead to small DBIs, a minimization of this measure is desired. This significantly better partitioning of SOMs in comparison to the classification of hedge funds into the categorization scheme specified by the database provider can be observed in all examined data samples (robustness analyses). Ultimately, none of the original 23 strategy groups can be empirically validated. Furthermore, no stable classification can be found. Both the number of empirically determined categories (SOM clusters) and the composition of these clusters differ significantly in the subsamples examined. Thus the results essentially confirm the results and conclusions in the literature, according to which the original, self-classified strategy labels of the database providers are misleading and therefore do not contain any information content.

Keywords: Self-Organizing maps, Clustering, Classification, Hedge funds, Style creep

1. Introduction

In contrast to traditional investment funds, which are subdivided according to the asset classes in which they are predominantly invested, there is no uniform classification scheme for hedge funds. There are many approaches to classify hedge funds according to their investment strategy. However, these are characterized by the database providers and are usually neither comparable nor consistent. Empirical studies of different hedge fund databases regularly show that the same hedge fund is managed differently in different databases and that the different databases have little in common. The differences are reflected in the definitions of investment styles and the number of hedge fund categories in which hedge fund managers generally have to classify themselves. The problem is that, given the heterogeneity of hedge fund strategies and styles on the one hand and the heterogeneity of database providers on the other, misclassifications can occur. Besides, there is the danger of the so-called "style creep", according to which fund managers make erroneous entries deliberately or do not report changes in their investment style, for example in order to improve their performance in comparison to the "peer group". The study of Baghai-Wadji et al (2006) shows that this phenomenon is widespread in the hedge fund industry. In particular, the authors point out that the probability of a style change of hedge funds with incorrect self-classification is twice as high as that of funds with a correct self-classification.

The consequences of these deliberate or unintentional classification errors are far-reaching, as they lead to distortions in the analysis of hedge fund strategies and their performance. They also involve the potential risk of making the wrong investment decisions, for example, to construct funds of funds or to add a hedge fund to an existing portfolio. Das & Das (2005) find that different empirical analyses produce different results based on the selected database. Against this background, there is a need for a procedure that avoids the problems associated with the inconsistent and often incomparable strategy classes of database providers and enables an objective classification.

Among others, the empirical studies of Fung & Hsieh (1997), Fung & Hsieh (2004) or Das & Das (2005) therefore attempted to classify the hedge fund universe objectively based on the observed returns of the funds. Self-Organizing Maps, as demonstrated in Maillet & Rousset (2001) or Baghai-Wadji et al. (2006), have proven to be appropriate, especially from the criticism that classification within the framework of linear (return-based) style analyses, such as those used in traditional investment funds, contradicts the dynamic nature of hedge fund strategies.¹ After studies by Ultsch & Vetter (1994), Mangiameli et al. (1996) and Moutarde & Ultsch (2005), among others, demonstrated their superiority to traditional statistical clustering methods, the use of SOMs was also suggested in the studies by Deboeck (1998) and Das & Das (2005).² As a result, the studies come to a smaller number of hedge fund classes and reveal a high degree of misclassification compared to the specifications by the (various) data providers.

Motivated by these findings, this article is devoted to Self-Organizing Maps. In various studies, this approach has already demonstrated its superior ability for objective classification. The contribution of this article lies in the extension of the procedure usually chosen in the hedge fund literature. As Poddig & Sidorovitch (2001) point out, a trained SOM map (Kohonen layer) usually does not yet provide an exact representation of the discovered structures in the input data. To overcome this problem, the 2-stage procedure according to Vesanto & Alhoniemi (2000) is implemented. Afterward, the SOMs are trained and in a second step, partitioning is derived using the K-means Clustering Algorithm. Based on the taxonomy determined in this way, the quality of the self-classification of hedge fund managers is investigated.

The results of the empirical study can be summarised as follows: As in comparable studies, the number of derived strategy groups (SOM partitions) is significantly smaller than the predetermined number given by the database. This is also accompanied by a significant reduction in the Davies Bouldin Index (DBI) of the SOM partitions. Since a small dispersion within the clusters and large distances between the clusters lead to small DBIs, a minimization of this measure is desired. This significantly better partitioning of SOMs than the classification of hedge funds into the categorization scheme specified by the database provider can be observed in all investigated data samples (robustness analyses). Ultimately, none of the original 23 strategy groups can be empirically validated. Furthermore, no stable classification can be found. Both the number of empirically determined categories (SOM clusters) and the composition of these clusters differ significantly in the subsamples examined. These results confirm the results and conclusions of Maillet & Rousset (2001), Das & Das (2005) and Baghai-Wadji et al. (2006), according to which the original, self-classified strategy designations of database providers are misleading and therefore regularly contain no information content. The article is structured as follows. The next section introduces the Self-Organizing Maps and shows the advantages of clustering the Kohonen layer with the k-means algorithm as an interpretation aid. The third section presents the data and the structure of the empirical study and describes the implementation of the 2-stage approach according to Vesanto & Alhoniemi (2000). Furthermore, the empirical results including the robustness analyses are discussed. The fourth section concludes the article with a presentation of the most relevant results.

2. Methodological Framework

2.1 Self-Organizing Maps

Kohonen developed the Self-Organizing Maps in the 1980s.³ They belong to the unsupervised learning methods (neural networks) and thus have the advantage of uncovering unknown structures in the data to be investigated without prior information. The SOMs have the ability to map high-dimensional feature spaces into an output space of a lower dimension while preserving the inherent structure of the original input data (topology preservation).

The network architecture of a SOM consists of only one input and one output layer. This is a two-tier feed-forward network, where the input layer is used to read external information into the network. The number of input units is determined by the dimension of the feature space to be mapped. Compared to other network types, SOMs have some distinctive features. Thus, the units of the input layer are first completely connected to the units of the output layer. Besides, this is the real peculiarity; the units on the output layer (Kohon layer) have a firmly defined order relation, which indicates the spatial neighborhood among each other. These fixed neighborhood relationships of the output units enable the topology to be maintained in the dimension-reducing mapping of the input space. Furthermore, as the studies Baghai-

¹ Connor & Lasarte (2003) conclude that any attempt to develop a formal classification system for hedge fund strategies is limited in particular by the fact that the strategies are subject to a continuous development process.

² See, e.g., Seiler (2009) or Bares et al. (2001) on the application of traditional clustering methods in connection with hedge funds.

³ See Kohonen (1981), Kohonen (1982a), Kohonen (1982b) and Kohonen (1982c).

Wadji et al (2006), Das & Das (2005), Maillet & Rousset (2001) and Deboeck (1998) point out, a metric is implemented, which allows quantifying distances between the output units.

SOMs are trained in an iterative process. In each step t , a training vector $x(t)$ is selected from the input data according to a probability function and the distances d_i to the weight vectors (also prototype vectors) $w_i(t) = [w_{i1}; \dots; w_{in}]$, where n corresponds to the dimension of the input vector, of all output units on the Kohon layer are determined.⁴ To calculate the distance, the Euclidean measure d is proposed by Kohonen (1997) and Poddig & Sidorovitch (2001):

$$ed_i = \|x(t) - w_i(t)\|_2 := \sqrt{\sum_{j=1}^n (x_j(t) - w_{ij}(t))^2} \quad (1)$$

The goal is to determine the Best Matching Unit wBMU (BMU), which has the smallest distance to the training vector $x(t)$:

$$\|x(t) - w_{BMU}(t)\| = \min\{\|x(t) - w_i(t)\|\} \quad (2)$$

Once the BMU, also known as the Winning Unit, has been determined, the actual learning process begins: the weights of the BMU and the weights of the neighboring units within a certain radius are modified so that they adapt to the direction of the selected input vector $x(t)$ according to equation (3).

$$w_i(\text{new}) = w_i(\text{old}) + \alpha(t)h_{BMU_i}(t)[x(t) - w(t)] \text{ with} \quad (3)$$

$\alpha(t)$: learning rate, $0 \leq \alpha(t) \leq 1$

$h_{BMU_i}(t)$: neighbourhood function

In this process, the Best Matching Unit "learns" the most. Which of the remaining output units change with which intensity is determined by the neighborhood function $h_{BMU_i}(t)$. Both the distance between the BMU and the adjacent output units U_i and the radius mentioned above are decisive. Poddig & Sidorovitch (2001) name the Gaussian bell curve, which has its maximum and its center of symmetry at the BMU, as a suitable neighborhood function:

$$h_{BMU_i} = \exp\left[-\frac{\|BMU - U_i\|^2}{2\sigma^2(t)}\right] \quad (4)$$

Among others, Kerling (1997), Kohonen (1997) and Vesanto & Alhoniemi (2000) propose a monotonously decreasing function for both the radius $\sigma(t)$ and for the learning rate $\alpha(t)$. This initially results in fast and coarse development of the network, which is refined further and further in the following process.

After the training phase has been completed, whereby the studies of Kohonen (1997) and Poddig & Sidorovitch (2001) recommending 10,000 to 100,000 learning steps, the output units are distributed over (hyper) space. Finally, the input space has been segmented into individual catchment areas (Voronoi elements) of respective output units. The weight vectors of the output units represent the centers of these Voronoi elements and can be understood as cluster centroids of their catchment areas, as Kerling (1997) pointed out. The determination of the weights of a SOM is based on the minimization of the following error function:

$$E = \sum_{k=1}^N \sum_{i=1}^M h_{BMU_i} \|x_k - w_i\|^2 \quad (5)$$

This minimization is done in the iterative procedure according to equation (3), which is also known as online or pattern-per-pattern weight adjustment procedure and can be understood as a form of stochastic learning.

The advantages of the SOMs compared to a conventional k-Means cluster analysis lies in the ability mentioned above to map high-dimensional feature spaces into an output space of a lower dimension while maintaining the inherent structure of the original input data (topology preservation/dimension reduction). By maintaining the topology, the relationships in the original characteristic space are retained in the (dimension-reduced) mapping. Thus, neighboring units on the output

⁴ Output units are represented by the weight vectors $w_i(t)$.

layer correspond to similar (data) clusters in the original characteristic space. Baghai-Wadji et al. (2006) see a practical advantage of SOMs over traditional cluster methods in their ability to deal with return histories of varying lengths.⁵ Ultimately, the SOM algorithm is a partitioning process. However, the mentioned topology maintenance leads to continuity in the mapping, so that no precise number of clusters can be specified after the end of the training phase. The exact determination of cluster boundaries requires further procedures.

2.2 Clustering of Self-Organizing Maps

Typically, the dimensionally reduced mapping of a SOM is displayed on a two-dimensional map. One way to determine the number of clusters is therefore to visualize this trained Kohonen layer ("Clustering via Visualization"). The simplest approach is to label the trained map. The training data is presented to the SOM again and the respective BMUs are marked according to the corresponding input data. An advanced approach is the so-called "Unified Distance Matrix" (U-Matrix). The U-matrix adds the Euclidean distances between each output unit and its direct neighbors as a third dimension on the Kohonen layer. This facilitates the manual formation of clusters.⁶

Ultimately, the visualization techniques can only assist in the subsequent "manual" cluster formation. This is a tedious process that does not guarantee a consistent summary of neighboring output units. To overcome this weakness, Vesanto & Alhoniemi (2000) propose an "automated" approach, as shown in Figure 1.

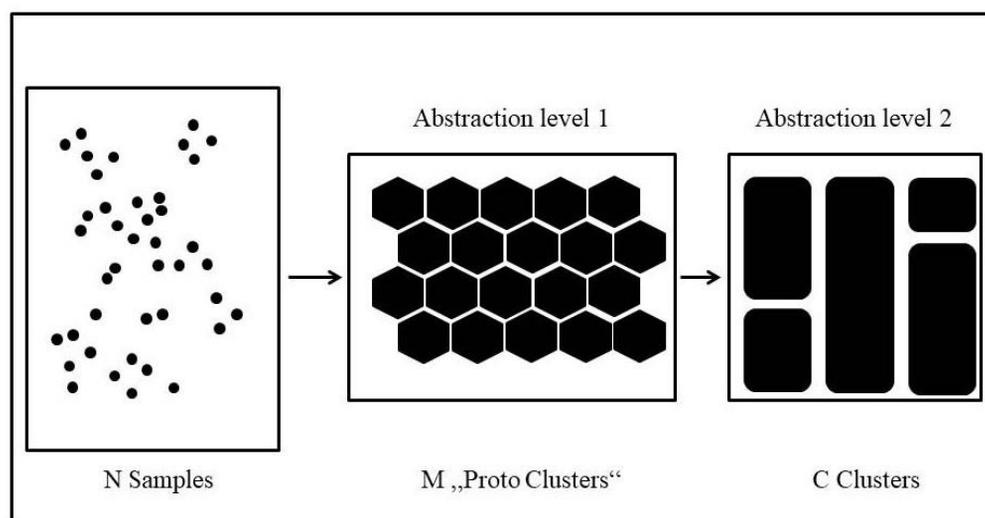


Figure 1: 2-step approach according to Vesanto & Alhoniemi (2000). Source: Vesanto & Alhoniemi (2000), p. 588.

In the first step, the Self-Organizing Maps reduce the dimensions of the original feature space while preserving the inherent structure of the original data entered. The set of prototype vectors (abstraction level 1) created after the end of the training phase usually exceeds the number of expected clusters. However, the prototype vectors can be seen as "proto-clusters", which can be combined in a second step to form the actual clusters (abstraction level 2). Vesanto & Alhoniemi (2000) use the k-Means clustering method, which can be outlined by the following five steps as discussed in MacQueen (1967):

- 1 Determination of the number of clusters
- 2 Random initialization of the cluster centers
- 3 Calculation of partitioning/classification
- 4 Calculation of the new cluster centers
- 5 Repeat steps 3 to 5 until the partitioning remains unchanged or the algorithm has converged

⁵ Everitt et al (2001), for example, recommend the elimination of incomplete data a series within the framework of a traditional cluster analysis. Consequently, only hedge funds with equal return histories could be analyzed, thus leading to a significant reduction of the available data.

⁶ Detailed descriptions of the U-Matrix can be found, e.g. in Ultsch (1993). For further visualization techniques discussed in the literature, such as the P and U* matrix, see Ultsch (2003). For the Adaptive Coordinates method see Merkl & Rauber (1997) and the sources mentioned there.

If the "true" number of clusters, as in the present study, is a priori unknown, the algorithm presented above can be repeated for a set of different cluster numbers. The "best" partitioning is then determined using a validity index. In connection with SOMs, Vesanto & Alhoniemi (2000) use the Davies-Bouldin Index (DBI), which relates the compactness within the clusters to the separation between the clusters.⁷ With $S_c(Q_k)$ as the mean distance (averaged over all objects N_k) in cluster Q_k to the cluster center c_k and $d_{ce}(Q_k, Q_l)$ as the distance between the clusters Q_k and Q_l , the DBI is calculated as follows:

$$DBI(C) = \frac{1}{C} \sum_{k=1}^C \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \quad \text{with} \quad (6)$$

$$S_c = \frac{\sum_i \|x_i - c_k\|}{N_k}$$

$$d_{ce} = \|c_k - c_l\|$$

The authors see the advantage of their two-step approach over direct clustering of the data set in the reduction of noise as well as in the reduction of computational effort. The prototype vectors ("proto-clusters") are local averages of the data and therefore less sensitive to random changes compared to the original data. Ultimately, the negative influence of outliers on the clustering to be carried out in the second step is reduced.⁸

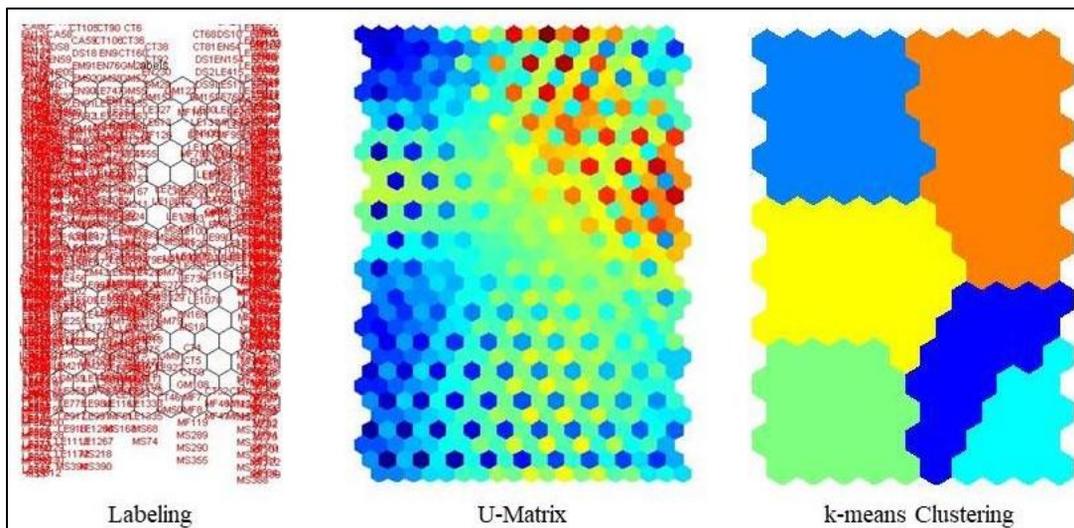


Figure 2: Evaluation of a trained SOM map

By overcoming the weaknesses of visualization techniques mentioned above, this two-stage process ultimately enables an objective classification of the original data while at the same time preserving the advantages of SOMs over traditional methods. Figure 2 shows by way of example a trained SOM map and illustrates the determination of clusters comparing the labeling approach to the U-matrix procedure and the k-Means clustering method. As shown in Figure 2, only k-Means clustering allows a precise determination of the number of clusters. It can also be seen that the "labeling" approach is unsuitable for determining the clusters, especially for larger amounts of data, and that the U-matrix method leaves a great deal of scope for interpretation.

3. Empirical Study

3.1 Research Object and Data

In this study, a systematic approach to classify (single) hedge funds are derived. The empirical studies by Maillet & Rousset (2001), Das & Das (2005) or Baghai-Wadji et al. (2006) have shown that SOMs are the appropriate tool for such an endeavor. In addition, studies by Ultsch & Vetter (1994), Mangiameli et al (1996) or Moutarde & Ultsch (2005) show that SOMs can better classify noisy data sets compared to traditional classification approaches and require less additional information. In previous studies on the classification of hedge funds with SOMs, the analysis was carried out using the

⁷ See Davies & Bouldin (1979). For alternative validity indices see Bezdek (1998), or Milligan & Cooper (1985).

⁸ For further information on noise reduction in connection with SOM, see, e.g. Kerling (1997).

visualization techniques criticized above, which leave room for interpretation in cluster formation and thus do not guarantee a consistent combination of adjacent output units. By applying the Vesanto & Alhoniemi (2000) two-step approach, the present study extends the approach chosen so far in the hedge fund literature and ensures an impartial, consistent classification that is free of scope for interpretation. This classification approach is used to answer the question of the quality and stability of the information provided by hedge funds about their investment style. Based on the study of Baghai-Wadji et al (2006), the derived classifications (SOM partitions) are compared to the given classifications of the database provider in order to measure quality and stability. (SOMs-)Classifications are determined for the entire period as well as for three subperiods and compared with the classifications of the database provider, which is based on the hedge funds managers' self-declarations. With this procedure three questions can be answered, which allow conclusions on the existence of a stable classification:

1. Did the classification process, based on self-declaration of hedge funds managers result in homogeneous strategy groups and is this classification stable in time?
2. Is there an alternative, time-stable hedge funds strategy grouping to the given database categorization?
3. Do the classifications determine based on the SOMs differ for the entire period and the subperiods in terms of composition and the number of strategy groups, so that the existence of a time-stable categorization can be negated?

For the empirical analysis, monthly returns from the Lipper-TASS database are used. Against the background of the survivorship problem⁹This database offers the advantage over many other providers as it contains not only active funds but also track records of hedge funds that have been eliminated from the market (so-called defunct). The data set comprises 2863 hedge funds and CTAs with different monthly returns and spans the period from 31.01.1999 to 31.12.2008.¹⁰ In order to avoid the backfilling bias¹¹, the first 24 return observations of all hedge funds are deleted from the sample following the procedure of Capocci & Hübner (2004) or Fung & Hsieh (2000) and funds that have a return history of less than 24 months are removed.

Table 1: Number of hedge funds in the different data samples grouped by strategy

Strategy	Identifier	Total sample 31.01.1994 - 30.04.2009	Subsample 1 31.05.2004 - 30.04. 2009	Subsample 2 31.05.1999 - 30.04. 2004	Subsample 3 31.01.1994 - 30.04. 1999
Convertible Arbitrage	CA	92	46	48	21
Distressed Securities	DS	67	42	23	19
Discretionary	DY	43	36	14	3
Event Driven Multi-Strategy	ED	73	45	36	16
Equity Long Only	EL	57	40	25	4
Emerging Markets	EM	187	119	81	27
Equity Market Neutral	EN	84	57	30	3
Equity Long/Short	ES	687	344	299	173
Fixed Income Arbitrage	FA	64	31	26	4
Fixed Income	FI	39	27	6	18
Fixed Income - MBS	FM	30	15	18	4
Global Macro	GM	86	46	29	23
Merger Arbitrage	MA	49	14	29	11
Multi-Strategy	MS	74	71	19	0
Relative Value Multi-Strategy	RV	35	23	15	3
Short Bias	SB	19	6	11	4
Sector	SC	106	47	2	4
Single Strategy	SS	14	14	46	22
Systematic	SY	183	173	68	23
Others ¹	OT	21	14	14	0
Total		2010	1210	839	382

¹ Due to the small number (<10) hedge funds of the strategies "Capital Structure" (CS), "Option Arbitrage" (OA), "Other Relative Value" (OV) and "Regulation D" (RD) are grouped into the strategy "Others".

Eventually, 2010 hedge funds from 23 strategy groups remain in the sample, which has a consistent return history of at least 24 observations corrected for backfilling bias. A potential distortion by the survivorship bias is counteracted since

⁹ On Survivorship Bias see, e.g. Seiler (2009).

¹⁰ In the further course of this study, the CTAs are regarded as hedge funds.

¹¹ On Backfilling Bias see, e.g. Fung & Hsieh (2000).

the data set contains not only 810 living but also 1200 so-called defunct hedge funds. In addition to the overall sample, three subsamples are formed to test for time stability of the classification. Table 1 shows the periods of the data samples including the number of hedge funds (classified according to the categorization of the database provider) contained therein.¹²

3.2 Implementation of the 2-step Approach

In order to implement the two-stage classification approach according to Vesanto & Alhoniemi (2000), the model-specific "degrees of freedom" of the SOMs must be specified. As with any neural network, this includes the specification of the network architecture. In section 2.1 SOMs are described as two-tier feedforward networks with a fixed arrangement of units on the output layer. Due to this particular architecture, the network specification is limited to determining the number of output units and their arrangement on the output layer, as discussed in Blackmore & Mikkulainen (1993) or Fritzke (1994). The topologies are based on the heuristics documented by Vesanto et al (2000).

Further degrees of freedom exist in the choice of the learning rate, the neighborhood function hBMU, the neighborhood radius and the number of learning steps t. Up to the authors' knowledge, there are no known methods which allow an optimal determination of these parameters. Therefore, the specification is made based on empirical findings. Following the study by Baghai-Wadji et al (2006), the learning rate is initialized with 0.06 and the neighborhood radius with 11. Following Kohonen (1997), the radius of the neighborhood is determined as a linearly decreasing function and the adaptation of the learning rate is calculated based on an inverse monotonously decreasing function depending on the learning steps t in each case. According to Kohonen (1997), the number of learning steps t is 500 times the number of output units. The Gaussian bell curve is selected as a neighborhood function based on the properties discussed in Section 3.1.

In addition to the degrees of freedom listed here, the stochastic nature of the learning process, which depends on the sequence of the selected training vectors as well as on the initialization of the output units, requires multiple repetitions of the training of the topologies listed above with different parameterizations and initializations. In order to subsequently identify the "adequate" SOM realization, Kohonen (1997) discusses the mean quantization error (mQuant) and the topology error (TopoErr). The mean quantization error, which is linked to the error function of the Self-Organizing Maps (equation 5), results from the mean Euclidean distance between all training vectors with their best matching units. The topology error evaluates the quality of a SOM realization by the continuity of the mapping from the input space to the trained SOM map. According to Kaski & Lagus (1996), this measure can be operationalized as a percentage of input vectors whose best matching units are not adjacent to their second best matching units.

In the second stage of the classification approach, the "adequate" SOM map identified in the first stage is clustered to determine the cluster boundaries objectively. As explained in section 2.2, Vesanto & Alhoniemi (2000) use the k-Means algorithm, which is also used in this study. Due to the sensitivity of this algorithm to the initial initialization, this procedure is performed 1000 times with random start values. In each of these 1000 passes, a set of different cluster numbers is tested and the number of clusters for "best" partitioning is determined using the Davies-Bouldin index.¹³ Ultimately, this results in a distribution of cluster numbers. To determine the final partitioning, the k-Means cluster method is applied again using the identified SOM partition, with the difference that the modal value of the determined distribution of cluster numbers is specified as the "true" value. For a reason mentioned above, the procedure is performed 1000 times and the sum of the mean Euclidean distances over all clusters (mEUKLID) is determined as a quality criterion in each iteration:

$$mEUKLID = \frac{1}{C} \sum_{k=1}^C \left(\frac{\sqrt{\sum_i (x_i - c_k)^2}}{N_k} \right) \quad (7)$$

The iteration with the lowest mEUKLID is identified in the present study as the final partitioning of the "adequate" SOM map determined in the first stage.

¹² Characterization of the individual strategy groups is given in Deetz (2013)

¹³ See section 2.2 for further details on the Davies Bouldin Index. According to Vesanto et al. (2000), the maximum number of clusters to be tested is as follows: $\sqrt{\text{Number of funds}} = \sqrt{2010} \cong 45$.

According to Kohonen (1997) and Ritter (1997), the map of a specific topology, which occurs particularly frequently, is identified as typical.

3.3 Empirical Evidence

In accordance with the studies of Das & Das (2005) and Baghai-Wadji et al (2006), the categories determined on the basis of the SOM show a significantly lower number of classes

Table 2: Number of categories and DBIs of the database provider compared to the empirically determined SOM classifications for all examined data samples

Sample	Number of categories		DBI	
	Database	SOM	Database	SOM
Total	23	14	2.51	0.66
Sub1	23	8	3.98	0.62
Sub2	23	12	4.00	0.62
Sub3	18	10	3.24	0.61

compared to the original classification scheme.¹⁴ This is also accompanied by a significant reduction in the Davies Bouldin Index (DBI) of the SOM partitions. Since a small dispersion within the clusters and large distances between the clusters lead to small DBIs, a minimization of this size is desired.¹⁵ This significantly better partitioning of SOMs than the classification of hedge funds into the categorization scheme specified by the database provider can be observed in all examined data samples. Table 2 summarizes the results.

Table 3 compares the final partitioning of the two-level classification approach (rows) with the original database classification (columns) of the entire data sample. Following the studies of Fung & Hsieh (1997), Brown & Goetzmann (1997) and Baghai-Wadji et al. (2006), the designation (column "Name" in Table 3) of the empirically derived

SOM partitions are based upon the prevailing strategy of the original classification. In particular, the study of Baghai-Wadji et al (2006), according to which hedge funds from the original strategy group(s) must make up at least 40% of the funds allocated to a SOM cluster, is followed. With this procedure, the quality of the original classification that is based upon self-declaration of the hedge fund managers is assessed. Furthermore, it can be determined whether the categorization specified by the database provider already represents a consistent classification scheme.

As the application of the naming convention described in Table 3 shows, the empirically determined SOM classification differs significantly from the original database grouping thereby showing significantly better DBIs - as recorded in Table 2.

Ultimately, none of the original 23 strategy groups can be empirically validated. This result confirms the results and conclusions of Maillet & Rousset (2001), Das & Das (2005) and Baghai-Wadji et al. (2006), according to which the original self-classification-based strategy designation ("label") is misleading and therefore regularly has no information content. In the studies of Brown & Goetzmann (1997) and Baghai-Wadji et al. (2006) so-called "style gaming" is discussed as one of the reasons for misleading labeling. According to this, hedge fund managers make incorrect entries regarding their strategy at the database provider on purpose or deliberately do not display style changes in order to improve their performance in comparison to the peer group. On the other hand, there is the argument given in Section 1 that by exploiting its possibilities, especially against the background of the less restrictive legal requirements, it is not the nature of hedge funds to pursue only one (fix) strategy, but to act dynamically on the markets. As Eling (2006) points out in this context, each hedge fund ultimately represents an individual strategy. This can be a mix of the known strategy

¹⁴ In these studies, the evaluations of the SOMs are based on the visualization techniques criticized in Section 2.2. However, the partitions identified in this study are all based on the two-stage classification approach introduced by Vesanto & Alhoniemi (2000) in Section 2.2. In the further course of this study, the term "SOM clusters" refers to cluster determined on the basis of the two-level classification approach, unless otherwise mentioned.

¹⁵ The DBI relates the compactness within the clusters to the separation between the clusters. See equation 6 in Section 2.2.

groups, but its composition can also vary over time. Fixed (strategy-specific) labeling is therefore not compatible with the dynamic nature of a hedge fund.

Table 3: Percentage allocation of hedge funds to the SOM clusters based upon their original classification for the total data sample

Total sample: 31.01.1994 - 30.04.2009												
Cluster	Name ¹	CA	DS	DY	ED	EL	EM	EN	ES	FA	FI	FM
Cl1	ES&DS&EM	11,20	12,45	0,83	8,71	0,41	12,45	1,66	18,67	6,22	6,22	2,07
Cl2	SY	0,00	0,00	6,78	0,00	1,69	2,54	0,00	0,85	0,00	0,00	0,00
Cl3	ES&CA&EN	12,72	3,89	1,06	4,24	0,00	2,47	10,95	19,08	9,19	2,12	4,95
Cl4	ES&EN&SY	3,64	0,40	7,29	1,21	0,00	1,62	14,57	20,24	6,07	2,43	3,64
Cl5	EM	0,00	0,75	0,00	1,49	2,99	53,73	0,00	29,85	0,00	0,00	0,00
Cl6	ES	1,37	2,74	0,68	4,79	12,33	2,74	0,68	56,85	0,68	0,00	0,00
Cl7	ES	2,94	3,53	0,00	2,94	8,24	0,59	0,00	61,76	2,35	0,59	0,00
Cl8	ES	1,21	1,21	0,61	1,82	7,27	12,12	0,61	53,94	0,00	0,00	0,61
Cl9	ES	5,31	5,31	0,88	12,39	0,00	1,77	0,88	44,25	1,77	4,42	0,00
Cl10	ES	1,85	3,70	0,00	0,93	3,70	15,74	0,00	61,11	0,00	0,93	0,00
Cl11	ES	2,22	0,00	0,00	2,22	0,00	11,11	1,11	57,78	1,11	1,11	0,00
Cl12	ES&EM	0,00	1,41	0,00	2,82	0,00	18,31	0,00	39,44	0,00	0,00	0,00
Cl13	SY	1,43	1,43	2,86	1,43	2,86	1,43	4,29	12,86	0,00	0,00	1,43
Cl14	ES&DY&GM&SY	0,00	0,00	12,96	0,00	0,00	5,56	11,11	27,78	0,00	7,41	0,00
Total ²		92	67	43	73	57	187	84	687	64	39	30

Total sample: 31.01.1994 - 30.04.2009 (continued)												
Cluster	Name ¹	GM	MA	MS	RV	SB	SC	SS	SY	Others ⁴	Sum(%)	Total ³
Cl1	ES&DS&EM	2,49	3,32	5,39	2,90	0,00	2,49	0,83	0,41	1,24	100	241
Cl2	SY	5,08	0,85	0,85	0,00	0,00	0,00	0,00	81,36	0,00	100	118
Cl3	ES&CA&EN	2,47	9,89	3,18	5,30	0,00	1,77	0,71	3,53	2,47	100	283
Cl4	ES&EN&SY	8,50	2,83	4,05	0,81	6,88	1,21	2,02	9,31	3,24	100	247
Cl5	EM	4,48	0,00	0,75	0,00	0,00	4,48	1,49	0,00	0,00	100	134
Cl6	ES	0,68	0,00	4,11	0,68	0,00	8,90	0,68	1,37	0,68	100	146
Cl7	ES	1,18	1,18	3,53	0,00	0,59	10,00	0,00	0,59	0,00	100	170
Cl8	ES	1,82	0,00	1,21	0,00	0,00	17,58	0,00	0,00	0,00	100	165
Cl9	ES	0,00	0,88	3,54	7,08	0,88	7,08	0,00	1,77	1,77	100	113
Cl10	ES	1,85	0,93	2,78	0,93	0,00	3,70	0,93	0,93	0,00	100	108
Cl11	ES	5,56	0,00	11,11	1,11	0,00	4,44	0,00	1,11	0,00	100	90
Cl12	ES&EM	7,04	0,00	9,86	0,00	0,00	11,27	0,00	9,86	0,00	100	71
Cl13	SY	21,43	1,43	0,00	0,00	0,00	0,00	1,43	45,71	0,00	100	70
Cl14	ES&DY&GM&SY	12,96	0,00	3,70	0,00	0,00	5,56	0,00	12,96	0,00	100	54
Total ²		86	49	74	35	19	106	14	183	21		2010

Following Bagahai-Wadji et al. (2006) this cross table compares the determined SOM clusters (rows) to the original categorization scheme of the database provider (columns). The results are given as a percentage.

¹SOM based cluster names according to Bagahai-Wadji et al. (2006): Name is based upon the current strategy (at least 40% of the funds allocated to a SOM cluster) of the original classification.

²Total number of hedge funds in their respective categories given by the database provider.

³Total number of hedge funds in their respective categories determined by the SOM.

⁴Composition "Others": 3 CS, 8 OA, 5 OV and 5 RD. See Table 1 for an explanation of these hedge funds strategy abbreviations.

3.4 Robustness Analysis

In the next step, the stability of the presented results of the overall sample is analyzed. Therefore the hedge fund data is divided into the three subsamples listed in Table 1. Subsequently, the SOM partitioning is determined empirically based on the two-stage classification approach.

As explained at the beginning of this section and documented in Table 2, the resulting SOM partitions in the subsamples also have a significantly smaller number of classes compared to the original classification, with significantly lower DBIs. Thus, the empirically determined SOM classifications represent the qualitatively better partitions. Similar to the presentation of the results of the complete sample in Table 3, the results from the subsamples are prepared in cross tables in such a way that the determined SOM clusters (rows) are compared with the respective original database groupings of the subsample (columns).

Cross-tables 4, 5 and 6 show that the number of SOM clusters is significantly lower than the original number of classes, but varies over time.

Table 4: Percentage allocation of hedge funds to the SOM clusters based upon their original classification for the total data subsample 1

		Subsample 1: 31.05.2004 - 30.04.2009										
Cluster	Name ¹	CA	DS	DY	ED	EL	EM	EN	ES	FA	FI	FM
C11	ES	2,76	3,45	0,00	6,90	17,24	2,07	1,38	42,76	0,00	0,00	0,00
C12	EM&ES	0,00	0,00	0,69	0,69	4,14	38,62	0,00	38,62	0,00	0,00	0,00
C13	SY	0,00	0,00	3,60	0,00	2,88	2,88	0,00	2,88	0,00	0,00	0,00
C14	ES&SY&EN	3,39	0,85	10,17	1,27	0,42	1,69	13,14	16,10	3,81	2,97	2,54
C15	ES&EM&EN	9,27	8,29	1,46	5,37	0,00	11,22	10,73	22,93	3,41	5,37	2,44
C16	ES&DS&FA	8,70	9,32	0,62	8,70	1,24	3,73	0,62	29,19	9,32	4,35	0,00
C17	ES	0,00	1,04	1,04	1,04	1,04	16,67	0,00	64,58	0,00	0,00	0,00
C18	ES&MS&SY	1,20	2,41	1,20	6,02	1,20	8,43	1,20	33,73	0,00	2,41	4,82
Total ²		46	42	36	45	40	119	57	344	31	27	15

		Subsample 1: 31.05.2004 - 30.04.2009 (continued)										Sum(%)	Total ³
Cluster	Name ¹	GM	MA	MS	RV	SB	SC	SS	SY	Others ⁴	Sum(%)	Total ³	
C11	ES	0,69	0,00	7,59	0,69	0,00	11,72	0,69	2,07	0,00	100	145	
C12	EM&ES	2,07	0,00	2,07	0,00	1,38	7,59	1,38	2,76	0,00	100	145	
C13	SY	3,60	0,00	1,44	0,00	0,00	0,00	0,72	82,01	0,00	100	139	
C14	ES&SY&EN	11,44	0,42	5,93	1,27	0,85	2,12	2,54	13,98	5,08	100	236	
C15	ES&EM&EN	0,49	4,39	5,85	3,90	0,00	0,98	1,46	2,44	0,00	100	205	
C16	ES&DS&FA	0,62	2,48	7,45	6,83	0,00	4,35	0,00	1,24	1,24	100	161	
C17	ES	2,08	0,00	6,25	0,00	2,08	2,08	1,04	1,04	0,00	100	96	
C18	ES&MS&SY	7,23	0,00	13,25	0,00	0,00	3,61	0,00	13,25	0,00	100	83	
Total ²		46	14	71	23	6	47	14	173	14		1210	

Following Bagahai-Wadji et al. (2006) this cross table compares the determined SOM clusters (rows) to the original categorization scheme of the database provider (columns). The results are given as a percentage.

¹SOM based cluster names according to Bagahai-Wadji et al. (2006): Name is based upon the prevailing strategy (at least 40% of the funds allocated to a SOM cluster) of the original classification.

²Total number of hedge funds in their respective categories given by the database provider.

³Total number of hedge funds in their respective categories determined by the SOM.

⁴Composition "Others": 1 CS, 8 OA, 3 OV, and 2 RD. See Table 1 for an explanation of these hedge funds strategy abbreviations.

As can be seen from Tables 1 and 2, subsamples 1 and 2 contain 23 (identical) pre-defined/original strategy groups. In Subsample 3, on the other hand, only 18 strategies exist. However, the number of empirically determined SOM clusters in subsample 3 is higher than in subsample 1, so the variations in the number of SOM clusters over time do not appear to be related to the underlying classification scheme specified by the database provider. Instead, it can be assumed that the original classification with the significantly worse DBIs is bloated and contains categories that are not based on (objectively) measurably discriminating features. This is also supported by the fact that none of the original 23 strategy groups can be identified in any of the subsamples applying the naming convention according to Baghai-Wadji et al. (2006).¹⁶ The comparison of the SOM clusters designations in Tables 4, 5 and 6 shows that not only the number of SOM clusters

Table 5: Percentage allocation of hedge funds to the SOM clusters based upon their original classification for the total data subsample 2

Subsample 2: 31.05.1999-30.04.2004												
Cluster	Name ¹	CA	DS	DY	ED	EL	EM	EN	ES	FA	FI	FM
C11	ES	1,14	1,14	0,00	1,14	10,23	10,23	1,14	46,59	0,00	0,00	1,14
C12	SY	0,00	0,00	4,17	0,00	2,08	0,00	0,00	6,25	0,00	2,08	0,00
C13	CA&ES&FA	17,65	2,94	3,68	4,41	0,00	1,47	8,09	16,91	13,24	0,00	10,29
C14	ES&SY&CA	10,37	6,67	0,74	5,93	0,00	6,67	0,74	20,00	2,96	1,48	2,22
C15	ES	1,23	1,23	0,00	0,00	3,70	18,52	0,00	55,56	2,47	0,00	0,00
C16	ES	3,45	0,00	1,72	1,72	5,17	34,48	0,00	46,55	0,00	0,00	0,00
C17	EN&ES	2,86	0,00	0,00	0,00	0,00	2,86	31,43	28,57	0,00	0,00	0,00
C18	ES	4,92	9,84	0,00	18,03	0,00	6,56	0,00	45,90	0,00	1,64	0,00
C19	SY	3,03	0,00	15,15	3,03	0,00	0,00	6,06	12,12	0,00	0,00	0,00
C110	ES	0,00	0,00	0,00	1,47	0,00	8,82	4,41	69,12	1,47	1,47	0,00
C111	ES	1,89	3,77	0,00	9,43	13,21	16,98	0,00	43,40	0,00	0,00	0,00
C112	ES	0,00	0,00	0,00	4,65	4,65	13,95	2,33	48,84	2,33	2,33	0,00
Total ²		48	23	14	36	25	81	30	299	26	6	18

Subsample 2: 31.05.1999-30.04.2004 (continued)												
Cluster	Name ¹	GM	MA	MS	RV	SB	SC	SS	SY	Others ⁴	Sun(%)	Total ³
C11	ES	2,27	0,00	0,00	0,00	0,00	22,73	1,14	1,14	0,00	100	88
C12	SY	12,50	25,00	4,17	0,00	2,08	0,00	0,00	41,67	0,00	100	48
C13	CA&ES&FA	2,94	0,00	2,21	3,68	6,62	0,00	0,74	2,21	2,94	100	136
C14	ES&SY&CA	2,22	8,89	5,19	5,19	0,00	3,70	0,00	16,30	0,74	100	135
C15	ES	2,47	0,00	1,23	1,23	1,23	8,64	0,00	0,00	2,47	100	81
C16	ES	1,72	1,72	0,00	0,00	0,00	3,45	0,00	0,00	0,00	100	58
C17	EN&ES	11,43	5,71	0,00	0,00	0,00	5,71	0,00	2,86	8,57	100	35
C18	ES	1,64	0,00	0,00	0,00	0,00	8,20	0,00	0,00	3,28	100	61
C19	SY	12,12	0,00	0,00	0,00	0,00	0,00	0,00	48,48	0,00	100	33
C110	ES	2,94	0,00	2,94	1,47	0,00	2,94	0,00	2,94	0,00	100	68
C111	ES	0,00	1,89	3,77	0,00	0,00	1,89	0,00	1,89	1,89	100	53
C112	ES	0,00	2,33	4,65	2,33	0,00	4,65	0,00	4,65	2,33	100	43
Total ²		29	29	19	15	11	46	2	68	14		839

Following Bagahai-Wadji et al. (2006) this cross table compares the determined SOM clusters (rows) to the original categorization scheme of the database provider (columns). The results are given as a percentage.

¹SOM based cluster names according to Bagahai-Wadji et al. (2006): Name is based upon the prevailing strategy (at least 40% of the funds allocated to a SOM cluster) of the original classification.

²Total number of hedge funds in their respective categories given by the database provider.

³Total number of hedge funds in their respective categories determined by the SOM.

⁴Composition "Others": 5 CS, 1 OA, 3 OV and 5 RD. See Table 1 for an explanation of these hedge funds strategy abbreviations.

¹⁶ The strategies ES and SY can be identified in all three subsamples, but they are used inflationary to name the SOM clusters determined.

Fluctuates over time, but also their compositions. The SOM clusters in the subsample differ significantly and are not stable, in a way that the SOM partitions of the subsamples are mostly independent of each other. There is no stable alternative SOM partitioning of the hedge funds examined for the original classification. In summary and against the background of the findings, the three questions formulated at the beginning of Section 3.1 regarding the validity of the original classification or the existence of alternative partitioning and, if existing, the stability of one of these categorizations can be answered as follows.

Table 6: Percentage allocation of hedge funds to the SOM clusters based upon their original classification for the total data subsample 3

Subsample 3: 31.01.1994-30.04.1999												
Cluster	Name ¹	CA	DS	DY	ED	EL	EM	EN	ES	FA	FI	FM
CI1	ED&DS&ES	5,26	21,05	0,00	36,84	0,00	0,00	0,00	21,05	0,00	5,26	5,26
CI2	EM&FA	0,00	3,77	0,00	1,89	0,00	35,85	0,00	20,75	22,64	0,00	0,00
CI3	ES&CA	17,39	7,25	0,00	4,35	0,00	2,90	1,45	31,88	7,25	1,45	2,90
CI4	ES	0,00	0,00	1,35	0,00	1,35	0,00	0,00	60,81	0,00	0,00	0,00
CI5	ES&SY	7,32	2,44	4,88	0,00	0,00	9,76	2,44	31,71	0,00	4,88	0,00
CI6	ES	0,00	20,00	0,00	10,00	10,00	5,00	0,00	50,00	0,00	0,00	0,00
CI7	ES	10,53	2,63	0,00	2,63	2,63	2,63	2,63	68,42	0,00	0,00	0,00
CI8	ES	3,13	6,25	0,00	3,13	0,00	0,00	0,00	65,63	3,13	0,00	0,00
CI9	ES	0,00	0,00	0,00	3,70	0,00	0,00	0,00	66,67	0,00	0,00	0,00
CI10	ES&MA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	33,33	0,00	0,00	11,11
Total ²		21	19	3	16	4	27	3	173	18	4	4

Subsample 3: 31.01.1994-30.04.1999 (continued)												
Cluster	Name ¹	GM	MA	MS	RV	SB	SC	SS	SY	Others ⁴	Sum(%)	Total ³
CI1	ED&DS&ES	0,00	0,00	0	0,00	0,00	5,26	0,00	0,00	0	100	19
CI2	EM&FA	11,32	0,00	0	0,00	3,77	0,00	0,00	0,00	0	100	53
CI3	ES&CA	2,90	13,04	0	1,45	0,00	2,90	0,00	2,90	0	100	69
CI4	ES	6,76	0,00	0	0,00	2,70	13,51	0,00	13,51	0	100	74
CI5	ES&SY	9,76	0,00	0	2,44	0,00	2,44	0,00	21,95	0	100	41
CI6	ES	5,00	0,00	0	0,00	0,00	0,00	0,00	0,00	0	100	20
CI7	ES	0,00	0,00	0	0,00	0,00	7,89	0,00	0,00	0	100	38
CI8	ES	3,13	0,00	0	3,13	0,00	12,50	0,00	0,00	0	100	32
CI9	ES	14,81	0,00	0	0,00	0,00	3,70	3,70	7,41	0	100	27
CI10	ES&MA	0,00	22,22	0	0,00	0,00	0,00	33,33	0,00	0	100	9
Total ²		23	11	0	3	4	22	4	23	0		382

Following Bagahai-Wadji et al. (2006) this cross table compares the determined SOM clusters (rows) to the original categorization scheme of the database provider (columns). The results are given as a percentage.

¹SOM based cluster names according to Bagahai-Wadji et al. (2006): Name is based upon the prevailing strategy (at least 40% of the funds allocated to a SOM cluster) of the original classification.

²Total number of hedge funds in their respective categories given by the database provider.

³Total number of hedge funds in their respective categories determined by the SOM.

⁴Composition "Others": 0 CS, 0 OA, 0 OV and 0 RD. See Table 1 for an explanation of these hedge funds strategy abbreviations.

The original categorization based on the information provided by the hedge fund managers in the strategy groups specified by the database provider cannot be empirically validated. Both concerning the number of classes and the composition of the strategy groups, the SOM partitioning differs from the original partitioning in all data samples or subperiods examined, with significantly smaller DBIs. In addition, no stable alternative categorization can be determined

for the entire period. Each subsample has its independent SOM partitioning. These results confirm the argument put forward in the introduction that it is not in the nature of hedge funds to pursue only one strategy, but to make full use of their much more extensive scope of opportunity and to act dynamically on the markets compared to traditional investment funds. Thus, as Eling (2006) explains, each hedge fund ultimately represents an individual strategy, which in turn is difficult to compare or to classify.

4. Conclusion and Outlook

There is currently no generally accepted classification of hedge funds. The database providers characterize the existing approaches and, according to the results of various empirical studies, are neither consistent nor comparable and show a high degree of misclassification.

With Self-Organizing Maps, this paper deals with a method that overcomes the problems of traditional regression-based classification procedures that can arise in connection with the often non-linear return-risk structures of hedge funds and enables an objective taxonomy. With the implementation of Vesanto & Alhoniemi's (2000) two-step approach, this study extends the problematic procedure of visual evaluation of the SOM maps usually chosen in the hedge fund literature, which does not always provide an exact representation of the discovered structures and require further interpretation aids. Also, the applications of the Kohonen rule for determining the number of learning cycles and the heuristics documented by Vesanto et al. (2000) for determining the topology contribute to a further objectification of the classification approach. In the empirical study, the application of the 2-stage SOM-based classification method led to a reduction from 23 to 14 hedge fund categories in the period from January 31, 1994, to April 30, 2009. However, this partitioning has not proved to be stable as shown in the robustness analyses. Both concerning the number of classes and the composition of the strategy groups, the SOM partitioning differs from the original partitioning in all data samples or subperiods examined, with significantly smaller DBIs. In addition, no stable alternative categorization can be determined for the entire period of the hedge funds examined.

The findings on the "true" affiliation of the individual hedge funds and their investment behavior gained using the 2-stage SOM-based classification approach presented here are essential for various reasons. For example, they can help to avoid non-diversified portfolios, which is particularly important in the construction of funds where risk is spread through a mix of styles. Furthermore, they enable the construction of benchmarks that can be used in the context of style analysis, e.g., for performance attribution or risk management. Maillet & Rousset (2001) already point out the possibility of extracting benchmarks or indices using the Kohonen algorithm for selecting the funds. In this context, these results can contribute to the creation of diversified portfolios. Furthermore, the 2-step classification approach introduces and implements a technique that can be used for the automated extraction of benchmarks and style indices. With regard to further research, the present results can be used to investigate the extent of "style creep" mentioned in the introduction. Furthermore, findings on the motivation of this stylistic change problem can be analyzed in more detail by further investigating the statistical properties of the centroid in the next step.

References

- Agarwal, V. & Naik, N. (2000), on taking the 'alternative' route: risks, rewards, style and performance persistence of hedge funds', *Journal of Alternative Investments*, Vol. 2, pp.6-23. [Crossref](#)
- Baghai-Wadji, R., El-Berry, R., Klocker, S. & Schwaiger, M. (2006), Changing investment styles: style creep and style gaming in the hedge fund industry, *Journal of Intelligent Systems in Accounting, Finance and Management*, Vol.14, pp. 157-177. [Crossref](#)
- Bares, P., R., G. & Gyger, S. (2001), Style consistency and survival probability in the hedge funds industry, Working Paper, Swiss Institute of Banking. [Crossref](#)
- Bezdek, J. (1998), Some new indexes of cluster validity, *IEEE Transactions on Systems, Man and Cybernetics Part B*, Vol. 28, pp. 301-315. [Crossref](#)
- Blackmore, J. & Miikkulainen, R. (1993), Incremental grid growing: encoding high-dimensional structure into a two-dimensional feature map, in *Proceedings of ICNN'93*; IEEE
- Brown, S. & Goetzmann, W. (1997), Mutual fund styles, *Journal of Financial Economics*, Vol.43, No. 3, pp. 373-399. [Crossref](#)

- Capocci, D. & Hübner, G. (2004), Analysis of hedge fund performance, *Journal of Empirical Finance*, Vol. 11, pp. 55-89. [Crossref](#)
- Connor, G. & Lasarte, T. (2003), An introduction to hedge fund strategies, Research Paper, London School of Economics.
- Das, N. & Das, R. (2005), Hedge fund classification technique using Self-Organizing feature Map neural network, Working Paper, Financial Management Association International.
- Davies, D. & Bouldin, D. (1979), A cluster separation measure, *IEEE Transactions on pattern analysis and machine intelligence PAMI-1*, pp. 224-227. [Crossref](#)
- Deboeck, G. (1998), Picking mutual funds with Self-Organizing Maps, in G. Deboeck & T. Kohonen, eds, *Visual Explorations in Finance with Self-Organizing Maps*, Springer, London, pp. 39-58. [Crossref](#)
- Deetz, M. (2013), Zur Persistenz in der Performance von Hedge-Fonds: Eine empirische Untersuchung unter Berücksichtigung klassifikationsinduzierter Probleme, PhD thesis, University of Bremen, Verlag Dr. Kovač.
- Eling, M. (2006), Hedgefonds-Strategien und ihre Performance, Ph.D. thesis, University of St. Gallen.
- Everitt, B., Landau, S. & Leese, M. (2001), *Cluster analysis*, 4 edn, Arnold, London.
- Fritzsche, B. (1994), Growing cell structures - a Self-Organizing network for unsupervised and supervised learning, *Neural Network*, Vol. 7, pp. 1441-1460. [Crossref](#)
- Fung, W. & Hsieh, D. (1997), Empirical characteristics of dynamic trading strategies: the case of hedge funds, *Review of Financial Studies*, Vol. 10, No. 2, pp. 275-302. [Crossref](#)
- Fung, W. & Hsieh, D. (2000), Performance characteristics of hedge funds and commodity funds: natural vs spurious biases, *Journal of Quantitative and Financial Analysis*, Vol. 35, No. 3, pp. 291-307. [Crossref](#)
- Fung, W. & Hsieh, D. (2004), 'Extracting portable alphas from equity long/short hedge funds', *Journal of Investment Management* 2(4), 1(19).
- Kaski, S. & Lagus, K. (1996), Comparing Self-Organizing Maps, in von der Malsburg, C. and von Seelen, W. and Vorbrügge, J. and Sendhoff, B., ed., *Proceedings of ICANN'96, International Conference on Artificial Neural Networks*, pp. 809-814. [Crossref](#)
- Kerling, M. (1997), *Moderne Konzepte in der Finanzanalyse. Markthypothesen, Renditegenerierungsprozesse und Modellierungswerkzeuge*, PhD thesis, University of Bremen, Uhlenbruch.
- Kohonen, T. (1981), Automatic formations of topological maps in a Self-Organizing system, in E. Oja & O. Simula, eds, *Proceedings of the 2nd Scandinavian Conference on Image Analysis*, pp. 214-220.
- Kohonen, T. (1982a), Clustering, taxonomy and topological maps of patterns, in *Proceedings of the Sixth International Conference on Pattern Recognition*, Silver Spring, MD (IEEE Computer Society), pp. 114-118. [Crossref](#)
- Kohonen, T. (1982b), Self-organized formation of topologically correct feature maps, *Biological Cybernetics* Vol. 43, pp. 59-69. [Crossref](#)
- Kohonen, T. (1982c), A simple paradigm for the self-organization of structured feature maps, competition and cooperation in neural nets, in S. Amari & M. Arbib, eds, *Lecture Notes in Biomathematics*, Vol. 45, Berlin, pp. 248-266. [Crossref](#)
- Kohonen, T. (1997), *Self-Organizing Maps*, Springer, Berlin.
- MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, in Lecam, L. & Neyman, J. eds, *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281-297.
- Maillet, B. & Rousset, P. (2001), Classifying hedge funds with Kohonen maps: a first attempt, Working Paper, Social Science Research Network (SSRN). [Crossref](#)
- Mangiameli, P., Chen, S. & West, D. (1996), A comparison of SOM neural network and hierarchical clustering methods, *European Journal of Operational Research*, Vol. 93, pp. 402-417. [Crossref](#)
- Merkl, D. & Rauber, A. (1997), Alternative ways for cluster visualization in Self-Organizing Maps. in: *proceedings of WSOM'97, workshop on Self-Organizing Maps*, Discussion Paper, Helsinki University of Technology.
- Milligan, G. & Cooper, M. (1985), An examination of procedures for determining the number of clusters on a data set, *Psychometrika*, Vol. 50, No. 2, pp. 159-179. [Crossref](#)
- Moutarde, F. & Ultsch, A. (2005), U*f clustering: a new performant cluster-mining method based on segmentation of Self-Organizing Maps, workshop on Self-Organizing Maps (WSOM'2005), Paris, France.

- Poddig, T. & Sidorovitch, I. (2001), Künstlich Neuronale Netze: Überblick, Einsatzmöglichkeiten und Anwendungsprobleme, in Hippner, H., Küsters, U., Meyer, M. & Wilde, K. eds, Handbuch Data Mining im Marketing, Vieweg, pp. 363-402.
- Seiler, K. (2009), Phasenmodelle und Investmentstilanalyse von Hedge- und Investmentfonds, Ph.D. thesis, University of Bremen.
- Ultsch, A. (1993), Self-Organizing neural networks for visualization and classification, in Opitz, O., Lausen & Klar, R., eds, Information and classification: concepts, methods and applications, Springer, Berlin, pp. 307-313. [Crossref](#)
- Ultsch, A. (2003), U*-matrix: a tool to visualize clusters in high dimensional data, Research Paper 36, University of Marburg.
- Ultsch, A. & Vetter, C. (1994), Self-Organizing-Maps versus statistical clustering methods: a benchmark, Research Paper 0994, FG Neuroinformatik und Künstliche Intelligenz, University of Marburg.
- Vesanto, J. & Alhoniemi, E. (2000), Clustering of Self-Organizing Map, IEEE Transactions on Neural Networks Vol.11 No. 3, pp 586-600. [Crossref](#)
- Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankangas, J. (2000), SOM toolbox for matlab 5, Technical Report, Report A57, Helsinki University of Technology.
- Zimmermann, H. (1994), Neuronale Netze als Entscheidungskalkül, in Rehkugler, H. & Zimmermann, H., eds, Neuronale Netze in der Ökonomie, Frankfurt School Verlag, pp. 1-87.