Clustering of Business Cycles in Optimal Directions found by SIR and DAME

Claudia Becker[1], Winfried Theis[2]

[1] SFB 475 "Komplexitätsreduktion in multivariaten Datenstrukturen"
Universität Dortmund, D–44221 Dortmund
e-mail: cbecker@statistik.uni-dortmund.de

[2] Graduate College "Applied Statistics"
Universität Dortmund, D–44221 Dortmund
e-mail: theis@statistik.uni-dortmund.de

Abstract

We investigate the combination of the dimension reduction methods SIR (Li, 1991) and DAME (Gather et al., 2001) with fuzzy–clustering to validate a given classification. We consider certain economic variables which are assumed to contain the information relevant to determine the current phase of economics. For a period of about 40 years, observations of these variables are available, together with an experts' judgement about the corresponding business cycle phase. We show that the combination of dimension reduction and fuzzy–clustering leads to a classification reflecting the experts' opinion better than other classification methods. Moreover, the proposed method can be used also in high–dimensional situations where other procedures are no longer applicable.

KEY WORDS: Business cycles, discrimination, SIR, DAME, cluster analysis, fuzzy–clustering

1 Introduction

We consider the problem of classifying economic data into business cycle phases. The long-term objective is to find a method by which we can predict the current business cycle phase based on actual economic data. The basis for our investigation is a data set of quarterly observations of several selected economic variables from about 40 years. Moreover, the data contain an experts' judgement about the business cycle phase corresponding to each quarter. The data set and the four phases are described in more detail in Section 4.

The data have already been investigated with respect to the given objective by two different approaches. The first approach consists of directly clustering the economic data, trying to rediscover the four business cycle phases found by the experts (Theis and Weihs, 1999). In the second approach, a certain subset of the given variables is extracted which is in some sense optimal for predicting the experts' classification (Weihs, Röhl and Theis, 1999). Both approaches lead to fair results, but are less suitable for being applied to high–dimensional data sets. For this reason, we investigate another approach, where we try to reduce the dimension of the data set of economic variables in a first step, followed by a clustering of the dimension–reduced data. As mentioned before, the long–term objective is to predict the current business cycle based on the current economic data. The given data set contains the past information about economics and business cycles. We also include the experts' information in the dimension reduction procedure. To the reduced data, we then apply a fuzzy–clustering method which already turned out to be suitable in the first approach described above (Theis and Weihs, 1999). If the data themselves contain appropriate information about the business cycle phases, and if the experts' classification is congruent with the information in the data, we should be able to rediscover the given business cycles by clustering the dimension–reduced data. In this case, the main conclusion would be that the chosen data can indeed be used to decide upon business cycle phases. Hence, the combination of directed dimension reduction and (undirected) fuzzy–clustering can be regarded as a tool for validating given classifications which were obtained by at least partially subjective methods. Together with this, we get a suitable projection of the high–dimensional data and a corresponding classification which allows for determining the actual business cycle phase.

The paper is organized as follows. Section 2 is dedicated to the dimension reduction methods SIR and DAME which are able to take a certain dependence structure of the data into account and hence are suitable for including the experts' judgement into the process of reducing the dimensionality. In

Section 3 the clustering algorithm is explained, including the choice of the distance measure and the measure of separateness, which is used for the comparison of different clusters. Finally, in Section 4 the data are briefly described and the results are collected.

2 SIR and DAME

The classification of business cycles can also be seen as a regression problem. The four distinct phases represent certain classes of response and thus correspond to a discrete univariate response variable Y with values in $\{1, 2, 3, 4\}$. The economic variables can be seen as explanatory variables $X_1, \ldots, X_p$, and we assume that there exists some functional relationship between $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ and Y. We adopt the idea of Li (1991) for a dimension reduction in this setting. He assumes that

$$Y = f(\boldsymbol{\beta}_1^T \boldsymbol{X}, \ldots, \boldsymbol{\beta}_K^T \boldsymbol{X}, \varepsilon), \tag{1}$$

i.e., the response Y depends on $X_1, \ldots, X_p$ only via the linear combinations $\boldsymbol{\beta}_1^T \boldsymbol{X}, \ldots, \boldsymbol{\beta}_K^T \boldsymbol{X}$. A dimension reduction from p to K is achieved, if $K \ll p$. Here, $\boldsymbol{X} = (X_1, \ldots, X_p)^T$, $\boldsymbol{X}, \varepsilon$ are stochastically independent, and $\boldsymbol{\beta}_i \in \mathbb{R}^p$, $i = 1, \ldots, K$. Such an assumption is well motivated in our problem, as it surely can be assumed that there are interdependencies between the economic variables.

The so-called effective dimension reduction (edr) directions $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ span the edr space $\mathcal{B}$. We assume that $\mathcal{B}$ equals the intersection of all possible edr spaces, the so-called central dimension reduction subspace (Cook, 1994, 1996, 1998a,b), and we assume further that this central subspace exists.

The sliced inverse regression (SIR) method of Li does not aim at estimating the function f itself, but it is designed as a means to estimate the space $\mathcal{B}$ in which the functional relationship takes place. In a second step then f can be estimated in the reduced space (see Becker, 2001). Here, we apply SIR in the sense that we try to find the subspace in which a certain structure of the data (in the form of fuzzy–clusters) manifests. In the second step we will not try to estimate f but to re-reveal this structure in the reduced space. This is just a slightly different view to the same situation, as we may interpret the result of a fuzzy–clustering procedure as a discrimination rule

by taking the maximum of the memberships. We may then also be able to use the SIR directions instead of the whole set of p variables (in our case we have $p = 13$) to predict a current business phase.
Let $(y_i, \boldsymbol{x}_i^T)^T$, $i = 1, \ldots, n$, $\boldsymbol{x}_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$, be a given dataset according to (1), then the SIR procedure consists of:

1. Standardizing: $\boldsymbol{z}_i = \widehat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$, $i = 1, \ldots, n$, where $\widehat{\boldsymbol{\Sigma}} = \sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T/n$, $\overline{\boldsymbol{x}} = \sum_{i=1}^n \boldsymbol{x}_i/n$.

2. Slicing: Split $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ into H slices S_h, $h = 1, \ldots, H$, according to the order of the corresponding values of $y_1, \ldots, y_n$; let $n_h = |S_h|$.

3. Calculating slice means: $\widehat{\boldsymbol{m}}_h = \sum_{S_h} \boldsymbol{z}_i/n_h$, $h = 1, \ldots, H$.

4. (Weighted) principal component analysis for the slice means: $\widehat{\boldsymbol{SIR}} = \sum_{h=1}^H n_h \widehat{\boldsymbol{m}}_h \widehat{\boldsymbol{m}}_h^T/n$ with eigenvalues $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_p$ and respective normalized eigenvectors $\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_p$.

5. Estimating the K edr directions $\boldsymbol{\beta}_i$: $\widehat{\boldsymbol{\beta}}_i = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\eta}}_i$, $i = 1, \ldots, K$.

To estimate K, we follow Li's (1991) suggestion of performing successive tests of hypotheses $H_0^j : K = j$ vs. $H_1^j : K > j$, starting with $j = 0$. We take $\widehat{K}$ to be the value of j, for which H_0^j is not rejected for the first time. The test statistic used in each of the aforementioned tests is

$$t_j := n(p-j)\overline{\lambda}_{(p-j)}\,,$$

where $\overline{\lambda}_{(p-j)}$ denotes the mean of the $(p-j)$ smallest eigenvalues of $\widehat{\boldsymbol{SIR}}$, and H_0^j is rejected if t_j exceeds an appropriate quantile of the χ^2 distribution with $(p-j)(H-j-1)$ degrees of freedom. For details see Li (1991, p. 321).
In the original work of Li (1991), it is assumed that $Y \in \mathbb{R}$. But in the second step of SIR only the ordered values of Y are used to categorize the corresponding $\boldsymbol{X}$ observations. Thus, the SIR procedure can easily be applied to the case of a discrete response Y as it is done here (also see Cook and Lee, 1999). Chen and Li (2001) show how SIR can be used in the context of linear discriminant analysis. The slices S_h can be chosen in the natural way given by the categories of the response. Hence, we take $H = 4$ as given by the business classification, and select the slices according to the

Y categories. We come up with $\widehat{K} = 3$ which is also the maximum value we can test for if $H = 4$.

For our analysis we do not take into account the structure of the data which is given by the time-series aspect. We just consider the $\boldsymbol{X}$ values as if they were i.i.d. observations. This is surely justified to get a first impression of how SIR (and DAME, see below) may be helpful in the context of the classification of business cycles. Further work will be concerned also with the dynamical aspect given by the time-series structures (for first approaches cf. Becker and Fried, 2001, Becker et al., 2001).

The SIR procedure is not robust against outliers in the $\boldsymbol{X}$-space (Gather et al., 2001b). A straightforward approach to robustify SIR is given in Gather et al. (2001a), where Li's basic idea is maintained, while all classical estimators are replaced by suitable robust versions. The resulting dimension adjustment method (DAME) then proceeds similar to SIR. In the first step of the procedure, the estimators $\overline{\boldsymbol{x}}$ and $\widehat{\boldsymbol{\Sigma}}$ are replaced by S-estimators $\boldsymbol{T}_1, \boldsymbol{C}_1$ of location and covariance. These estimators are calculated as described by Davies (1987), and are based on a modified biweight function as introduced by Rocke (1996) with parameters chosen to achieve maximum breakdown point and an asymptotic rejection probability of 0.1. The second step of DAME equals the second step of SIR. In step 3, the slice means $\widehat{\boldsymbol{m}}_h$ are replaced by $\boldsymbol{T}_{2,h}$, the L_1- or spatial medians within the hth slice. In the principal component analysis of the fourth step, instead of $\widehat{\boldsymbol{SIR}}$, we use an estimator $\widehat{\boldsymbol{DAME}} = \boldsymbol{C}_2(\widehat{M})$, where the set $\widehat{M}$ contains all estimated locations $\boldsymbol{T}_{2,h}$, and $\boldsymbol{C}_2$ denotes a projection pursuit covariance estimator according to Li and Chen (1985), based on the univariate robust scale estimator RCQ_α introduced by Rousseeuw and Croux (1993) with α chosen to be 0.5. The last step of DAME consists of estimating the edr directions:

$$\widehat{\boldsymbol{\beta}}_i = (\boldsymbol{C}_1^{-1}\widehat{\boldsymbol{\eta}}_i)/\sqrt{\widehat{\boldsymbol{\eta}}_i^T \boldsymbol{C}_1^{-1}\widehat{\boldsymbol{\eta}}_i}\,, i = 1, \ldots, K\,.$$

For detailed comparisons of the performance of SIR and DAME see Gather et al. (2001a,b).

3 Fuzzy–clustering and choice of distance

As described in Theis and Weihs (1999) and Theis, Vogtländer and Weihs (1999), the experts' classification of the data cannot be easily reproduced by means of standard statistical methods like, for example, time series analysis procedures or k–means clustering. Therefore fuzzy–clustering based on euclidean distances has been used in the search for the number of groups, which can be found empirically. This was combined with a variable selection by a greedy–search–algorithm. As this dimension reduction did not lead to a "proper" clustering, a modified distance measure was introduced, which turned out to model the differences between business cycle phases better than the euclidean distance (Theis and Weihs, 1999). Hence, we also apply this improved combination of fuzzy–clustering and variable selection here.

Let $S = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, $\boldsymbol{x}_i \in \mathbb{R}^p$, be an observed set of data. Fuzzy–clustering does not divide these data into well–separated groups but assigns every point a probability (membership) to belong to a certain group. That means, instead of constructing a partition $\mathcal{P} = \{P_1, \ldots, P_k\}$ of the data set S, $S \subset \dot{\bigcup}_v P_v$ $(P_i \cap P_j = \emptyset, \forall i \neq j)$, fuzzy–$k$–means–clustering constructs a covering $\mathcal{C} = \{C_1, \ldots, C_k\}$, $S \subset \bigcup_v C_v$. Each element $\boldsymbol{x}_i \in S$ does not belong definitely to a set C_v, but $\boldsymbol{x}_i$ may belong to several of the C_vs. For each observation $\boldsymbol{x}_i$, this can be seen as well as an estimation of k membership functions $u_v : \mathbb{R}^p \to [0, 1]$ with $\sum_{v=1}^{k} u_v(\boldsymbol{x}_i) = 1$, $i = 1, \ldots, n$. The number k of clusters has to be chosen beforehand, motivated by knowledge of the problem at hand or by hints from descriptive analysis.

Figure 1 shows on the left hand side a typical partition generated by k–means–clustering and on the right hand side a data set for the fuzzy approach and the sort of groups constructed by it.

Points $\boldsymbol{x}_i$ lying in overlapping regions get memberships $u_v(\boldsymbol{x}_i) =: u_{iv}$ smaller than 1 to belong to a specific group C_v, whereas points lying in only one group get a membership of 1 to belong to this group and 0 for all other groups. Hence, fuzzy–clustering leads to the same result as the k–means approach if there are well separated groups of data points.

The membership values of the n observations in a data set can be combined in the so–called membership matrix $\boldsymbol{U}$:

$$\boldsymbol{U} := (u_{iv})_{i=1,\ldots,n;v=1,\ldots,k} \,.$$

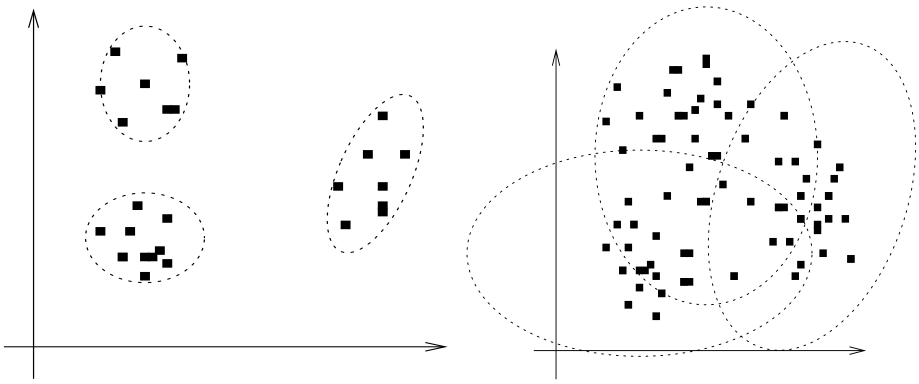

Figure 1: Difference between hard partition and fuzzy–partition

We use fuzzy–k–means clustering as implemented in the R/S–function FANNY (Kaufman, Rousseeuw, 1992, pp. 164–197). The membership values of the observations are obtained as solutions of minimizing the following term:

$$\sum_{v=1}^{k} \frac{\sum_{i,j=1}^{n} u_{iv}^2 u_{jv}^2 d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)}{2\sum_{i=1}^{n} u_{iv}^2},$$

where $d(\boldsymbol{x}, \boldsymbol{y})$ denotes a suitably chosen distance measure.

To judge the goodness of "separation" into the groups, we use the normalized Dunn–coefficient (Kaufman and Rousseeuw, 1992, p. 187)

$$\tilde{F}_k(\boldsymbol{U}) = \frac{kF_k(\boldsymbol{U}) - 1}{k-1},$$

where F_k denotes the usual Dunn–coefficient,

$$F_k(\boldsymbol{U}) := \sum_{v=1}^{k} \sum_{i=1}^{n} \frac{u_{iv}^2}{n}.$$

The normalized coefficient $\tilde{F}_k$ takes values between 0 for no partition (total fuzziness) and 1 for a hard partition.

To perform fuzzy–clustering, an appropriate distance measure is needed. The usually chosen euclidean distance does not fit very well here. Words like "upswing" or "downswing", which characterize certain business cycle phases, describe directions of development. In Theis and Weihs (1999) a new distance was developed to model this fact. The data points are normalized with the euclidean norm because this reduces the information in the observations to the direction in p–dimensional space and therefore the directions of two points relative to the origin can be compared by the euclidean distance of these points. This leads to the distance $d(\boldsymbol{x},\boldsymbol{y}) := \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} - \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \right\|$ (see Figure 2).

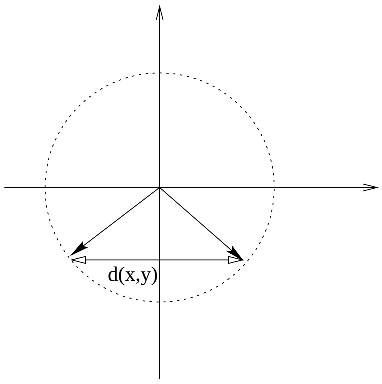

Figure 2: Illustration of distances of normalized data

4 Results

The data set considered here consists of $p = 13$ so-called stylized facts for the german business cycle listed in Table 1 and $n = 157$ quarterly observations from 1955/4 to 1994/4 (price index base=1991, y=yearly growth rates).

These 13 variables have been selected by Heilemann and Münch (1996) from a total of 120 variables. The experts' belief is that these 13 variables contain all information necessary to classify the observations into four business cycle phases called "upswing" (1), "upper turning point phase" (2), "downswing" (3) and "lower turning point phase" (4) as described in Heilemann and Münch (1996).

In the following, we compare the results of various analyses of these data. We investigate two general classification methods, namely quadratic discriminant analysis and fuzzy–clustering . These methods are applied to the original data (the observations of the 13 stylized facts), to the projected data according to the directions given by SIR and DAME, and to a special selection of three of the original variables. This latter selection is that

Abbr.	variable
Y	GNP, real (y)
C	Private consumption, real (y)
GD	Government deficit, percent of GNP
L	Wage and salary earners (y)
X	Net exports, percent of GNP
M1	Money supply M1 (y)
IE	Investment in equipment, real (y)
IC	Investment in construction, real (y)
LC	Unit labour cost (y)
PY	GNP price deflator (y)
PC	Consumer price index (y)
RS	Short term interest rate, nominal
RL	Long term interest rate, real

Table 1: The 13 Stylized Facts

combination of three stylized facts which yields the best result in the fuzzy–clustering classification. To gain these optimal variables, an extensive search algorithm has to be performed. The total number of 13 stylized facts out of which the optimal three have to be found, is surely close to the upper limit for the effective use of this algorithm. Hence, especially with regard to the possible application to higher–dimensional data, alternative procedures are sought for. In the following section, we first describe how the dimension reduced spaces are composed of the original variables.

4.1 Results of dimension reduction

The outcome of applying SIR and DAME to the data set are reduced spaces of dimension three. Since we choose the natural four slices given by the experts' classification, this is at the same time the maximum dimension we can test for. Hence, an optimal reduced space may even consist of more directions, but it is impossible to find them with this type of method. We restrict the detailed discussion exemplary to the outcome of SIR. The reduced space is spanned mainly by linear combinations of PY, L, RL (first direction), C, PY, RS (second direction), and PY, RL, LC (third direction). Seen over all directions together, the variables contributing most to the reduced space are the GNP price deflator (PY), the long and short term interest rates (RL, RS) and the private consumption (C). This selection of main influential variables is at first sight rather different from the optimal selection with respect to the goodness of fuzzy–clustering (L, IE, PC). Nevertheless, there are some dependencies between the choices. First of all, obviously variable L (wage and salary earners) is chosen in both selections. Second, from various analyses of the data set the variables PY and PC seem to be exchangeable. Methods for variable selection tend to choose the one or the other of them without showing any obvious pattern of choice. Hence, one of the main influential variables in dimension reduction can also be found in the variables best for fuzzy–clustering. From the rest of the variables adding mainly to the directions for dimension reduction, we can see that SIR can be interpreted as a classification method to some extent. This is due to the fact that the set of variables chosen is similar to the sets chosen when determining the variables optimal for classification of the data. For example, the set of three variables which are best for quadratic discriminant analysis of the data contains the variables L, LC, and RL, which are all contained in the projections found by SIR (for a more detailed discussion on selecting optimal variables for discrimination see Weihs, Röhl and Theis, 1999). In the following section, we discuss the results of a quadratic discriminant analysis of the various selected data sets.

4.2 First approach to classification: quadratic discriminant analysis

The simplest approach to classify the given data into four phases or classes would be to directly perform a discriminant analysis. Here, we choose quadratic discriminant analysis because it was the best method found by Weihs, Rhl and Theis (1999). The results are compared here for the original data as well as for the projected (SIR, DAME) and selected ones (optimal choice). We report (see Table 2) the leave–one–out crossvalidation error rates for all these data sets. It is obvious that working with the projected space found by SIR decreases the error rate while the decrease gained by

DAME is only slight. The three variables selected to be optimal for the clustering perform badly when used for the classification. This is not surprising since they are not the optimal 3–dimensional set for discriminating between the different phases, and as stated in Weihs, Röhl and Theis (1999) at least four variables are needed to reach a cv–error similar to the other cv–errors reported here. Good cv-error-rates in the classical LDA or QDA with variable selction are around 22-24%. Although the use of the projected data according to SIR yields a clearly better error rate than the use of the original data set, the result may be improved by applying a different classification procedure. Since fuzzy–clustering turned out to be an appropriate choice in this context (Theis and Weihs, 1999), we follow this approach in the next section.

Data Set	cv–error
Original	24.8
SIR 3D	17.8
DAME 3D	23.5
Variables L,IE,PC	38.8

Table 2: Cross–validated errors of QDA in the different data sets

4.3 Second approach to classification: fuzzy–clustering

As before, the classification is performed for the four versions of the data, where we now apply a fuzzy–clustering procedure as described in section 3. The selection of the three variables optimal with respect to fuzzy–clustering is done by a greedy–search algorithm, which deletes each variable once and applies FANNY to the new data set. The variable for which the Dunn–coefficient is increased most in this step, is discarded from the set of possible optimal variables and is therefore deleted from the data set for the rest of the algorithm, and the search is applied to the resulting data. The usual approach would be to proceed until there is no further increase in the Dunn–coefficient. For the sake of comparability with the results of the dimension reduction by SIR and DAME, we continue until only 3 variables are left and thus use the algorithm here to choose an "optimal" set of 3 variables. As mentioned before, the greedy–search algorithm has the important drawback of being an exhaustive search and would not be feasible in the case of more variables in the original data set.

The number k of classes is determined by calculating the Dunn–coefficient for the possible choices $k = 4$, $k = 3$, $k = 2$, starting with $k = 4$ classes like in the experts' classification. For most versions of our data, a clustering with visibly non–zero Dunn–coefficient is reached for a number of two classes. Thus, we choose $k = 2$. Table 3 reports the normalized Dunn–coefficients for the choice $k = 2$ for all versions of our data. We see that, with the optimal variables, we reach the largest value of the Dunn–coefficient, hence there is a way to achieve better separated clusters than with the projected data from SIR or DAME. But the investigations of the following Section 4.4 show that these clusters are not as close to the experts' classification. It can also be seen from the table, that using the new distance in FANNY leads to larger values of the Dunn–coefficient, hence this appears to be much more appropriate for our problem than using the euclidean distance. As we have seen before, it adds to the interpretability of the results.

We can interpret the two classes found as the two "main" phases, namely upswing and downswing. Although the "turning–point" phases, which can be seen as a state of the economy in between the "main" phases, are not identified directly, they are inherently found by looking at the membership value of each observation. This idea is illustrated in Figure 3. The normalized observations are represented as points on the circle in the centre. Assume that these data points move around the circle in the given direction, representing the cyclic movement through the four business phases. The ellipses stand for the main phases upswing and downswing which are identified by the clustering procedure. Then it is obvious that the memberships should be high in the main phases and decrease toward the edges, and with two groups it is possible to interpret memberships around 0.5 as points in between the two phases. So this is the best we can get if we cannot identify all four groups directly. Plotting the memberships in one of the groups over time we expect to find times of high memberships alternated with low membership and in between some points with membership near 0.5, belonging to the respective turning–point phase. The type (upper or lower turning point) can be deduced from the time-related context.

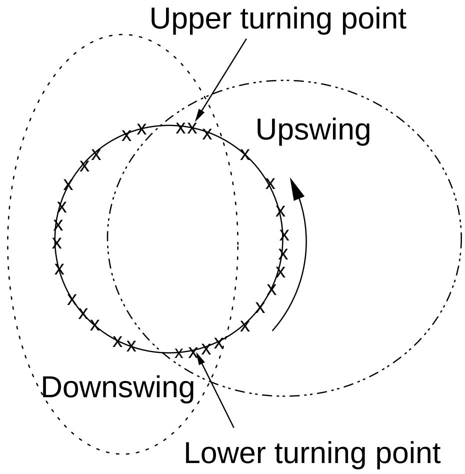

Figure 3: Idea of two groups with the new distance

In the next subsection we compare the memberships in the clusters with the classification into the four phase scheme of Heilemann and Münch.

4.4 Comparison of classifications

In Figures 4, 5, and 6 the membership values of the observations with respect to one of the found clusters (standing for "upswing") is plotted in the course of time. The 4–phase–scheme of Heilemann and Münch is depicted as an additional line in the plots. To make comparison easy the highest value of the line stands for upswing, the lowest for downswing, the lower in between for lower turning–point phase and the higher in between for upper turning–point phase. The heights chosen for the line are to some extent arbitrary, they are chosen in a way that the "middle" values (standing for

Data set	$\tilde{F}_D(U)$ eucl. Dist.	$\tilde{F}_D(U)$ new Dist.
Original data	4.6629367e-15	4.440892e-16
SIR 3D	1.052491e-13	0.1082308
DAME 3D	2.087219e-14	0.08069364
Variables L,IE,PC	0.2773638	0.5204419

Table 3: Normalized Dunn–coefficient for the different projections of the data set and the different distances

the turning point phases) lie around 0.5, corresponding to our interpretation of the membership values.

Figures 4 and 5 correspond to the results of reducing the dimension with SIR and DAME, respectively, and then applying the fuzzy–clustering. Figure 6 corresponds to selecting the three "best" variables by the greedy–search algorithm and applying the fuzzy-clustering procedure to the observations of these variables. Obviously, the clustering is more fuzzy in Figures 4 and 5 than in Figure 6, where the membership values are closer to 1 (definite membership in the class) or 0 (definitely not belonging to the class), respectively. Nevertheless there are some advantages. These can be seen especially at the left-hand side of Figure 5. Instead of being clustered mainly to one group as in Figure 6, the observations are divided between the groups with a pattern which looks similar to the 4–phase–scheme of the experts. The greater fuzziness may be caused by some of the badly discriminating variables, which are deleted by greedy–search but are used by the optimal projections. Considering the discussed interpretation of memberships around 0.5 in our current context, a certain amount of fuzziness is asked for to get some ideas about the turning–point phases.

As in previous investigations (Theis and Weihs (1999)), again the cycle at the beginning of the seventies is a difficult area. As pointed out there, this is the time of the first oil–crisis, and the corresponding observations were found to be a third group in the original data set.

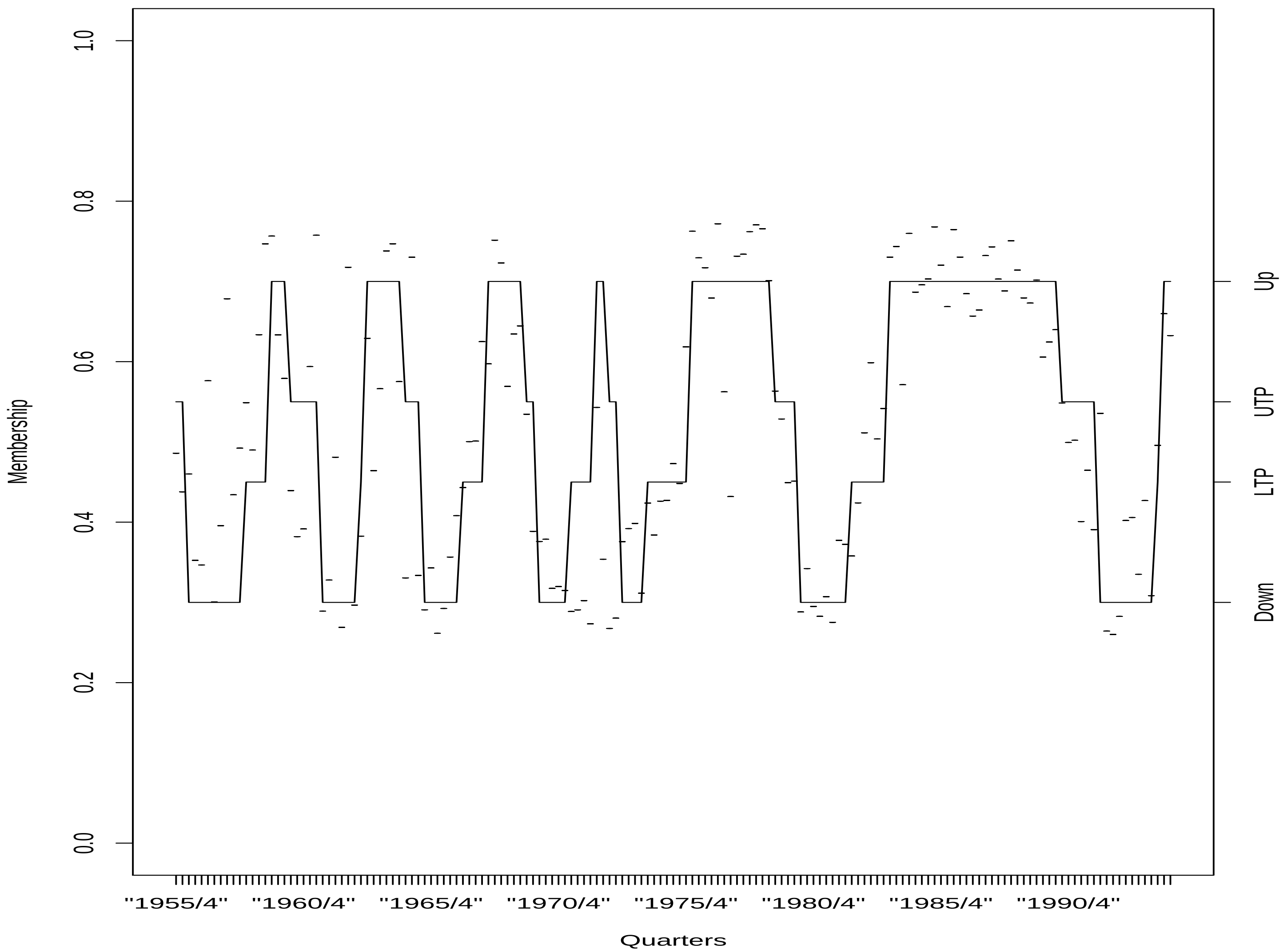

Figure 4: Membership in group #2 in clustering for SIR 3D using the new distance compared to 4–phase–scheme

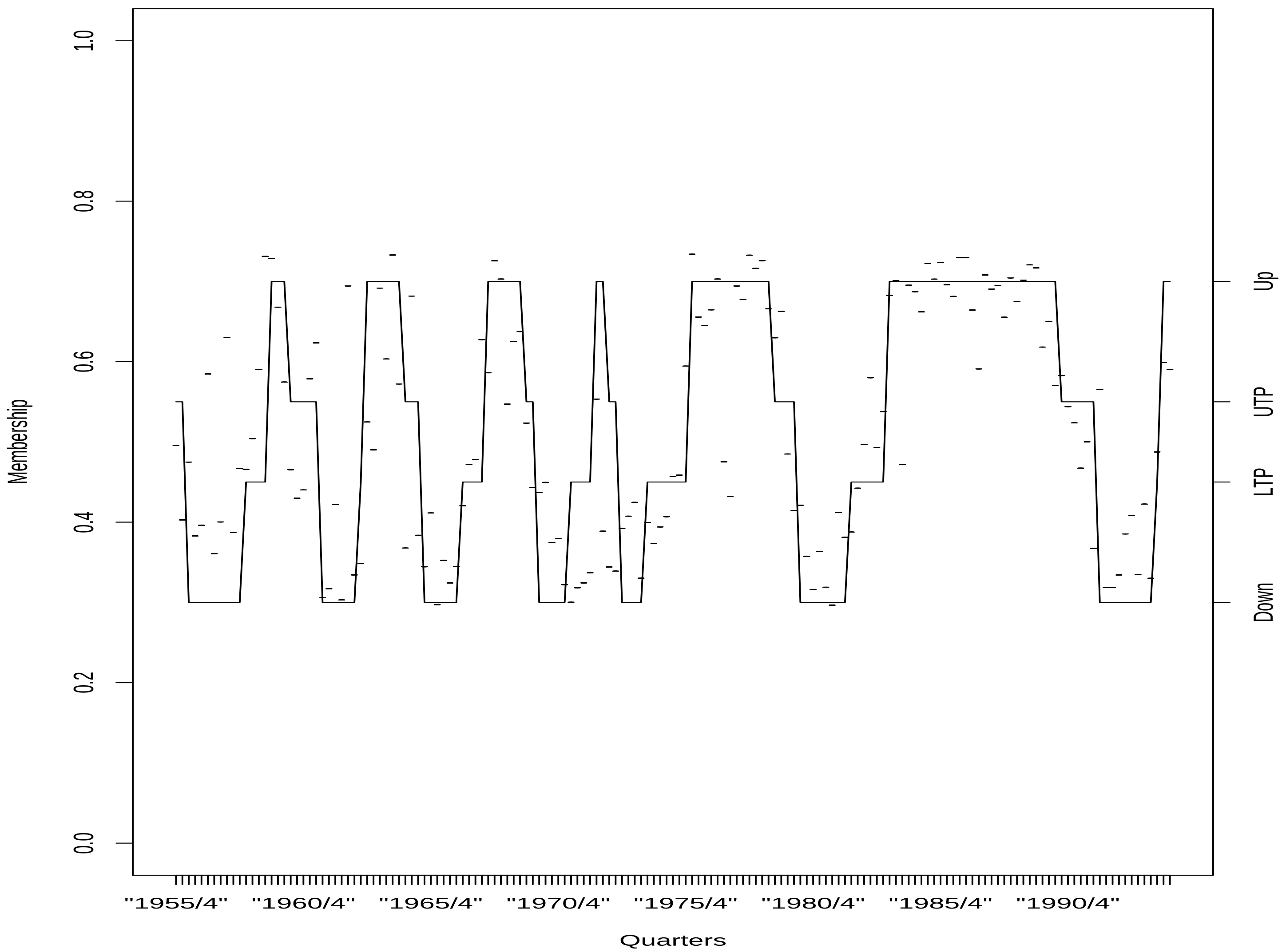

Figure 5: Membership in group #2 in clustering for DAME 3D using the new distance compared to 4–phase–scheme

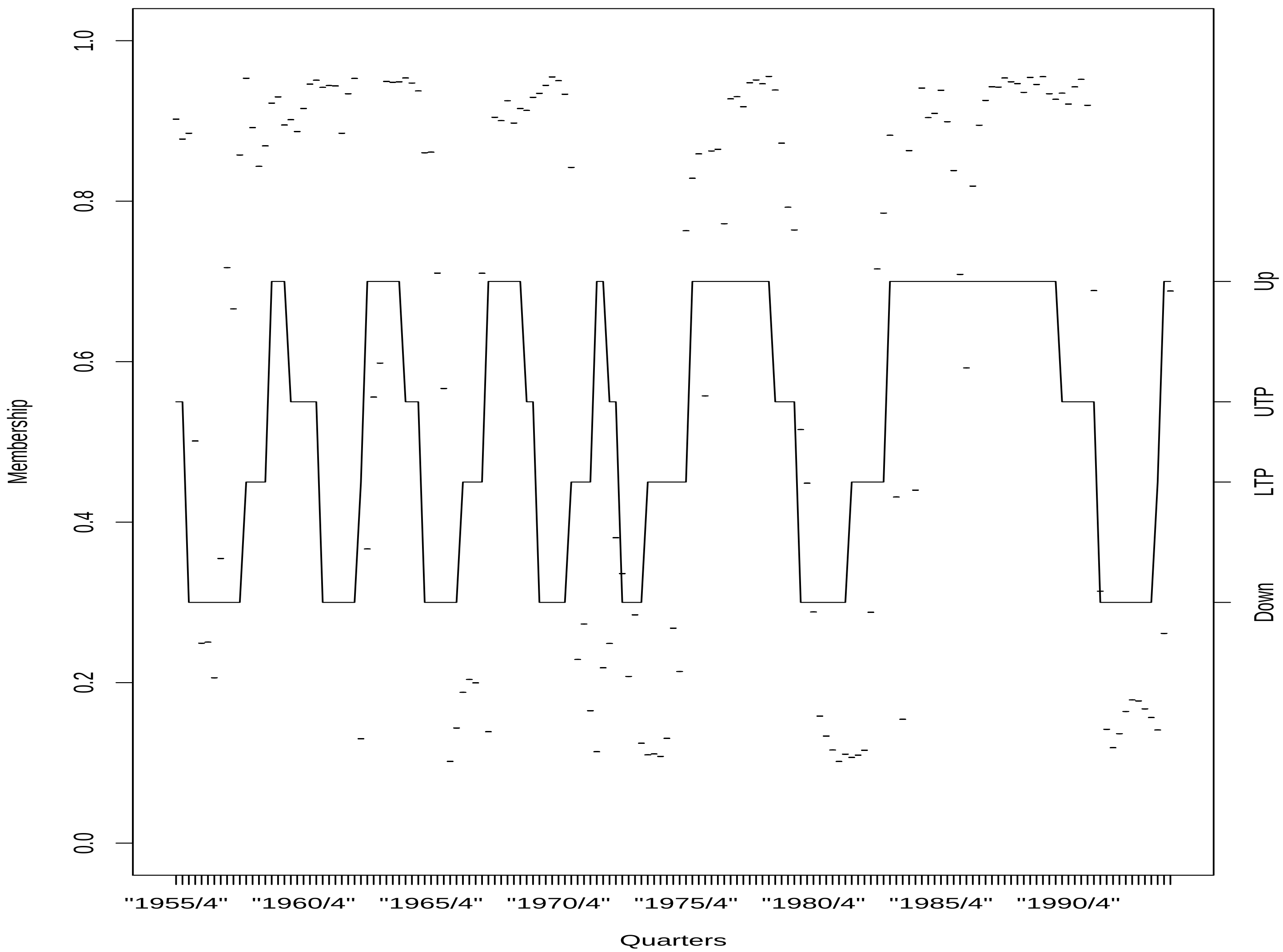

Figure 6: Membership in group #1 in clustering for best 3 variables using the new distance compared to 4–phase–scheme

5 Conclusions

The aim of our investigation was to check whether a given classification of a 13–dimensional dataset can be reproduced by a combination of dimension reduction and fuzzy–clustering. Using a compound method of this type was motivated by results of former analyses where the application of a dimension reduction method as a pre–step led to improved results (Weihs, Röhl and Theis, 1999). This is true here as well. Comparing the results of the fuzzy-clustering using the 3–dimensional projections by SIR and DAME with the results when using the original data, a better separation can be reached by the former methods. This is reflected in an increase of the Dunn–coefficient. Although using the variables selected by the greedy–search leads to even better separated groups, the clustering of the projected data can be interpreted in a way which is much more similar to the expert's classification. Implicitly, we are even able to distinguish three different classes, corresponding to the two main phases "upswing" and "downswing" and to an in–between phase including both turning–point phases. This is not possible with the clustering of the data from the optimally chosen variables. Hence, the combination of dimension reduction and fuzzy–clustering seems to be a good alternative to the time–expensive search for optimal variables as performed by the greedy–search algorithm. It can be expected that using a larger set of variables in the data set to start with will improve the classification. To include a larger number of variables does not pose a problem for the dimension reduction methods, since they are designed for exactly such a situation. On the other hand, it will barely be possible to apply the greedy–search algorithm to such a larger number of variables. Moreover, also the dimension reduction methods offer the possibility to choose an "optimal" set of variables instead of linear combinations of the variables (see Chen and Li , 1998, for details). To conclude, the proposed method seems to be quite promising for checking a given classification, especially with respect to high–dimensional problems where the greedy–search algorithm is not feasible.

Acknowledgments

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of Complexity in Multivariate Data Structures", and Graduate College "Applied Statistics") is gratefully acknowledged.

References

Becker, C. (2001) Dimension Adjustment Methods, in: *Proceedings of the 53rd Session of the International Statistical Institute*, Seoul, Republic of Korea, http://134.75.100.178/ isi2001/data/237.pdf.

Becker, C., and Fried, R. (2001) Sliced Inverse Regression for High-dimensional Time Series, Technical Report **14/2001**, SFB 475, Universität Dortmund.

Becker, C., Fried, R., and Gather, U. (2001) Applying Sliced Inverse Regression to Dynamical Data, in: Kunert, J., and Trenkler, G. (eds.), *Mathematical Statistics with Applications in Biometry, Festschrift in Honour of Siegfried Schach*, Eul-Verlag, Köln, 201–214.

Chen, C.-H., and Li, K.-C. (1998) Can SIR be as Popular as Multiple Linear Regression? *Statist. Sinica*, **8**, 289–316.

Chen, C.-H., and Li, K.-C. (2001) Generalization of Fisher's Linear Discriminant Analysis via the Approach of Sliced Inverse Regression, *J. Korean Statist. Soc.*, **30**, 193–218.

Cook, R.D. (1994) On the Interpretation of Regression Plots, *J. Amer. Statist. Assoc.*, **89**, 177–189.

Cook, R.D. (1996) Graphics for Regressions With a Binary Response, *J. Amer. Statist. Assoc.*, **91**, 983–992.

Cook, R.D. (1998a) Principal Hessian Directions Revisited, *J. Amer. Statist. Assoc.*, **93**, 84–100.

Cook, R.D. (1998b) Regression Graphics: Ideas for Studying Regressions through Graphics, *Wiley, New York*.

Cook, R.D., and Lee, H. (1999) Dimension Reduction in Binary Response Regression, *J. Amer. Statist. Assoc.*, **94**, 1187–1200.

Davies, P.L. (1987) Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices, *Ann. Statist.*, **15**, 1269–1292.

Gather, U., Hilker, T., and Becker, C. (2001a) A Robustified Version of Sliced Inverse Regression, in: Fernholz, T.L., Morgenthaler, S., and Stahel, W. (eds.), *Statistics in Genetics and in the Environmental Sciences*, Birkhäuser, Basel, 147–157.

Gather, U., Hilker, T., and Becker, C. (2001b) A Note on Outlier Sensitivity of Sliced Inverse Regression, *Preprint*.

Heilemann, U., and Münch, H.J. (1996) "'West German Business Cycles 1963-1994: A Multivariate Discriminant Analysis"', Paper presented at the 1995 CIRET-Conference in Singapore, CIRET-Studien 50, München.

Kaufman, L., and Rousseeuw, P.J. (1990) Finding Groups in Data, *Wiley, New York*.

Li, K.-C. (1991) Sliced Inverse Regression for Dimension Reduction (with discussion), *J. Amer. Statist. Assoc.*, **86**, 316–342.

Li, G., and Chen, Z. (1985) Projection–Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo, *J. Amer. Statist. Assoc.*, **80**, 759–766. Correction: *J. Amer. Statist. Assoc.*, **80**, 1084.

Rocke, D.M. (1996) Robustness Properties of S–Estimators of Multivariate Location and Shape in High Dimension, *Ann. Statist.*, **24**, 1327–1345.

Rousseeuw, P.J., and Croux, C. (1993) Alternatives to the Median Absolute Deviation, *J. Amer. Statist. Assoc.*, **88**, 1273–1283.

Theis, W., and Weihs, C. (1999) Clustering techniques for the detection of Business Cycles, Technical Report **40/1999**, SFB 475, Universität Dortmund.

Theis, W., Vogtländer, K., and Weihs, C. (1999) Descriptive Study of the RWI data set, Technical Report **45/1999**, SFB 475, Universität Dortmund.

Weihs, C., Röhl, M.C., and Theis, W. (1999) Multivariate Classification of Business Phases, Technical Report **26/1999**, SFB 475, Universität Dortmund.