Multivariate Classification of Business Phases

Claus Weihs[*]
Fachbereich Statistik
Universität Dortmund

Michael C. Röhl
Fachbereich Statistik
Universität Dortmund

Winfried Theis
Fachbereich Statistik
Universität Dortmund

April 1999

Abstract

We propose multivariate classification as a statistical tool to describe business cycles. These cycles are often analyzed as a univariate phenomenon in terms of GNP or industrial net production ignoring additional information in other economic variables. Multivariate classification overcomes these limitations by reducing dimension in a way suitable for human perception. Based on a four phase scheme (upswing, upper turning point, downswing, lower turning point) we demonstrate the potential of classification methods by determining the important economic variables (stylized facts) for the German business cycle.

KEY WORDS: Business cycle, classification, dimension reduction, simulated annealing, transition structure

[*]e-mail: weihs@amadeus.statistik.uni-dortmund.de

1 Introduction

Although business cycles are mainly treated as a univariate phenomenon in the literature, it is realized to be actually a *multivariate* one. Often business cycles are tackled by univariate time series methods ignoring the interplay of the different economic variables. This article shows the potential of a multivariate technique which takes this interrelation into account: classification. There have been essentially two approaches in the literature concerned with business phase classification: Meyer and Weinberg (1975a,b) analyzed the American business cycle and Heilemann and Münch (1996) the German one. Both use a well–known classification procedure called linear discriminant analysis (LDA) which is briefly described in the third section. This procedure has some limitations. In this paper we enlarge and systematize the preceding studies. Special emphasis is put on complexity reduction and graphical representation.

Recently, Diebold and Rudebusch (1996) discuss two main aspects of the old definition of Burns and Mitchell (1946) of the business cycle: the comovement of important economic variables and the partition of the cycle in different phases, which are assigned to different economic regimes. Because of the assumed comovements in the newer literature models with latent factors are discussed (see e.g. Stock and Watson, 1991). Because of the regime-switching assumption models with latent states were developed (see e.g. Hamilton, 1989, Filardo, 1994). Therefore, Diebold and Rudebusch propose a model, in which a latent factor depends on the latent state of the economy, which is actually a combination of the recent interpretations of the ideas of Burns and Mitchell.

Both, the idea of switching regimes and of latent factors are also in the focus of multivariate classification methods. Indeed, oberservations, i.e. time periods, are classified and the classes can be thought of as different economic regimes, and classification methods typically identify latent factors, the values of which are as different as possible in the different classes, i.e. business phases. Unfortunately, these latent factors are most of the time difficult to interpret, since they are linear combinations of the observed economic factors. Moreover, the time dependency of the transition from one phase to another is completely ignored. Indeed, the business phase of the next time period can only be the phase of the current period or the 'next' phase in the defined course of phases.

The aim of this paper is to demonstrate how to utilize multivariate classification methods to model business cycles. In particular, we will demonstrate how to interpret the latent variables properly, and how to introduce transition restrictions into classification methods. The method will be applied to analyse the German business cycle. One aim is to identify important 'stylized facts' (Lucas, 1983) for this business cycle. The data used is the same as in HEILEMANN/MÜNCH elongated back to 1955/4. In particular, we use the same four phase scheme as Heilemann and Münch with phases called (in their natural course of appearance) 'upswing', 'upper turning points', 'downswing', and 'lower turning points'.

The paper is organized as follows. The classification procedure will be described in sections 2–4. In section 5 the analysed data will be introduced. In sections 6–7 the classification results for this data will be discussed. Section 8 gives the conclusions of our study.

2 Classification of Business Cycle Phases

Classification deals with the allocation of objects to g predetermined groups (or classes), indexed by $G = \{1, 2, \cdots, g\}$. In our application the objects will be time periods (quarters), the groups above business cycle phases. The variables considered to be important for discriminating between the groups can be continuous (GNP, consumption etc.) or discrete (number of firms, number of inhabitants etc.). In the following we concentrate on, say d, continuous variables with values in a portion B of the d–dimensional real space ($B \subseteq \mathbb{R}^d$).

Based on some preclassified objects (learning sample) a classification rule is learned, a rule incorporating the information inherent in the learning sample. Future objects will then be classified by this classification rule.

In order to construct the classification rule the information given in the learning sample is typically "encoded" in terms of a real valued density function $p : B \to \mathbb{R}$ with $\int_B p(x)\,\mathrm{d}x = 1$. The group densities $p_i(x)$ $(1 \leq i \leq g)$ are estimated for each group with the help of the learning sample. New objects with variables vector x are classified to group i if $p_i(x) > p_j(x)$ $(j \neq i)$. The goodness of classification depends on the class of densities we use. Often one uses densities with a small number of parameters in order to facilitate estimation. A popular density is the normal one: only the mean vector and a measure of interrelation – the covariance matrix – have to be specified.

The goal is to choose a classification rule *which minimizes the misclassification error (error rate)* of new objects. To achieve this, the misclassification error has to be specified. There are mainly two ways:

1. In the first variant (model dependent) we estimate the group densities based on the whole learning sample. Then the "overlap" in B of the different group densities $p_i(x)$ determines the misclassification error (see figure 1). This overlap can be evaluated by integration.

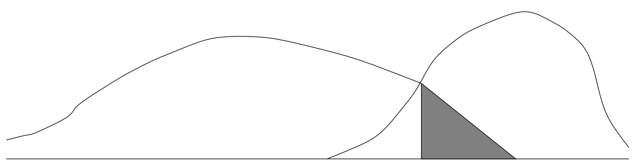

Figure 1: The error rate of the left group is gray shaded.

2. A completely different approach (variant II, modelfree) is to use the data more extensively. Here, we construct the classification rule by learning its unknown parameters, e.g. corresponding to densities, from only one part of the sample (training sample). With the other part (test sample) the misclassification error is determined. This approach is data intensive but more reliable in error rate determination than integration in 1 because now the data "speaks for itself" whereas the densities in 1 may not be appropriate, thus leading to a false misclassification error. This splitting procedure is called *crossvalidation*.

   If the data set is "small enough", an extreme form of crossvalidation, so–called *leave one out*, can be justified (see, e.g., Weiss and Kulikowski (1991)). Suppose there are n objects in the learning sample. Then we calculate the classification rule with $n-1$ objects and classify the left out one. This is done for all the n objects, thereby arriving at n different classifications with m errors altogether, say. The reported misclassification error is then m/n. This procedure is necessarily computer intensive because we have to construct the rule n times.

Note that the misclassification rate is related to future objects. If we would construct the classification rule with the whole learning sample and then classify the same sample (and no future objects), with the constructed rule, this would give the so–called *apparent* misclassification error rate. Of course, this error rate is optimistically biased because we use the same data for construction and validation.

Further note the difference between the classification rule and the way one computes the misclassification error. Even when the densities do not reflect the right model for the data, they can be used to construct a classification rule. Moreover, even if the classification rule is constructed by the model dependent approach, the misclassification error should be estimated modelfree, i.e. by means of crossvalidation, in order to judge the rule in "real life" and not in the model world, and in order to be able to compare different classification rules.

In procedure 1 apriori information can easily be incorporated. If we know how probable group i is apriori, i.e. before having seen any data, denote π_i the apriori probability for group i, then the classification rule reads as follows: allocate an object to group i if $\pi_i p_i(x) > \pi_j p_j(x)$ $(j \neq i)$. In business cycle applications such apriori probabilities are even time dependent since the phases have, at least in principle, to follow the pattern $\ldots g \to 1 \to 2 \to \ldots \to g \to \ldots$ being allowed not to change from one time period to the next. If one does not allow any exceptions to this rule, this can be modelled by setting the apriori probabilities $\pi_j(t+1) = 0$ for all $j \neq \mathit{phase}(t)$ or $\mathit{successor}(t)$, $\mathit{phase}(t)$ being the business cycle phase realized in time period t and $\mathit{successor}(t)$ being equal to $\mathit{phase}(t)+1$, if $\mathit{phase}(t) \neq g$, and equal to 1, otherwise. $\pi_j(t+1)$, $j = \mathit{phase}(t), \mathit{successor}(t)$, can be estimated based on the relative frequencies the corresponding phases are appearing in the learning sample. Using apriori probabilities this way, only the allowed group densities are taken into account, and from these the phase with highest probability is selected.

E.g., when we are in phase 2 in time period t, then the phase in the next time period can only be 2 or 3. Therefore, we should only compare $p_2(x)$ and $p_3(x)$ taking also into account the apriori probability for the two phases. This is realized by comparing $\pi_2 p_2(x)$ and $\pi_3 p_3(x)$ and taking the phase with the bigger value. We call this procedure "classification with exact transitions" because there will not be any forbidden transitions in contrast to the standard classification procedure. This incorporation of apriori information makes classification much more realistic for business cycle applications. Indeed, in our learning sample no (in the above sense) forbidden transition appeared.

3 Dimension Reduction

In order to find the best classification rule typically dimension reduction is used to separate structure from noise. This is accomplished by reducing the dimension from d to d', say. There are mainly two possible strategies. We can reduce the number of variables, that is select d' from d (variable selection). Or we take a linear projection from d–space onto d'–space using a $d' \times d$ projection matrix (linear dimension reduction). But often it is difficult to interpret the coefficients of such a projection matrix. A nonlinear projection is also possible but often infeasible, except when there is clear advice what kind of nonlinearity has to be modelled for the data.

In variable selection one may proceed as follows. For any subset of variables, a classification rule is constructed using densities $p_i(x)$ $(1 \leq i \leq g)$ and its performance (misclassification error) is evaluated modelfree by leave one out. This can be either done by exhaustive search or by genetic algorithms (Siedlecki and Sklanski (1989)) which implement a more intelligent strategy inspired by evolution. With more than 20, say, variables, exhaustive search is not feasible. With variables selection the important variables are detected directly and no further interpretation is needed.

The simplest group densities $p_i(x)$ are normal. In this paper three different normal models are tried.

1. Estimate for each group an individual covariance matrix (quadratic discriminant analysis, QDA, see e.g. McLachlan (1992)). This is the most general model. Unfortunately, for groups with small sample size the estimate of the covariance matrix and thus the density estimates may not be reliable in high dimensional space because of the big number of parameters to be estimated.

2. Estimate a joint covariance matrix for all groups (linear discriminant analysis, LDA, see e.g. McLachlan (1992)).

3. Estimate for each group an individual but diagonal covariance matrix, i.e. set the covariances equal to zero apriori (diagonal discriminant analysis, DDA).

In linear dimension reduction, i.e. dimension reduction by linear projection, we proceed as follows. First, we estimate the group densities in the original d–space. Then, the densities are projected on a d' dimensional subspace and the overall error rate is evaluated by integration. By means of direct minimization of the error rate we choose that subspace which minimizes the overall error. Simulated Annealing (Bohachevsky et al. (1986)) is used as the optimization algorithm. When the data set is "small", we rely on LDA or DDA in the original space. But after the optimal subspace is selected and the data is projected into it, less parameters have to be estimated, and QDA is applied. Finally, the error rate is evaluated by leave one out applied to time periods. Other possibilities might be to leave one business phase of a cycle out, or even one whole business cycle.

Having found the optimal subspace, we are mainly interested in the interpretation of the projected original data in order to learn about conditions under which the different groups are realized. One way to interprete the projection is to partition the optimal subspace into regions corresponding to the different classification groups. The borders of such regions are constructed by means of equating the projected group densities $p_{R,i}$ i.e. by analyzing the equations $p_{R,i}(x) = p_{R,j}(x)$ for all $i \neq j$. The densities $p_{R,i}$ are generated by applying a projection matrix R to the original group densities p_i. Having constructed the borders, the problem is to determine under what economic conditions which of the regions is projected into, i.e. which business cycle phase is realized. What we need is an interpretation of the borders of the regions by means of the original economic variables. An example for this is given in section 7.2.

The algorithm to determine the optimal subspace will be briefly discussed in the next section.

4 Optimization Procedure

The algorithm optimizes the entries in the $e \times d$–matrix R, i.e. $R \in \mathcal{M}(e, d)$ which is used for the projection of the original d–dimensional space on a lower dimensional space of dimension e. More specifically, we minimize the misclassification rate $f(r)$:

$$\begin{aligned} \text{Minimize}\, f: \quad \mathcal{M}(e,d) &\rightarrow \quad \mathbb{R}^+ \\ r &\mapsto \sum_{i=1}^{g} p(i) \int_{B_i} \pi_{R,i}(x)\, \mathrm{d}x, \end{aligned} \tag{1}$$

where r is the vector of entries in R, π_i is the apriori density of group i, $B_i = \{x \mid \exists j : p_{R,i}(x) < p_{R,j}(x)\}$, and $p_{R,i}$ is the i–th projected group density. The integration is performed by numerical quadrature procedures.

As optimization procedure we use an implementation of the simulated annealing algorithm based on a routine in Numerical Recipes in C (Press et al (1992)). The basis of this routine is the search algorithm of Nelder and Mead (cp. also Press et al. (1992)) which encloses the optimum by shrinking simplices. The shrinking is proportional to a so–called "temperature" sequence $T_n \longrightarrow 0$, $n \to \infty$, $\quad T_n > 0$. Therefore, as T_n approaches zero, the movements have to be more and more local and the algorithm converges to the next optimum.

For fixed T_n the algorithm generates a Markov chain with transition function $q(r_i, r_p)$ corresponding to a stochastic version of the moves of the NELDER AND MEAD algorithm. The transition function $q(r_i, r_p)$ specifies the probability of a transition from vector r_i to vector r_p.

The trial vector r_p is accepted with probability $\exp(-(f(r_p) - f(r_i))/T_n)$. In this way matrices R leading to a decrease of misclassification rate $f(R)$ are accepted in any case, but also matrices increasing the error rate are accepted with some probability. This is the reason, why simulated annealing is able to overcome local optima and thus avoids the selection of multiple starting points. This makes simulated annealing useful even in lower dimensions.

After a number of steps in the markov chain, the temperature will be decreased, $T_{n+1} = \alpha T_n$ $(0 < \alpha < 1)$, and a new chain will be created.

This computerintensive method achieves minimal misclassification error if adequately implemented.

5 The data set

The data set consists of 13 economic variables listed in table 1 and 157 quarterly observations from 1955/4 to 1994/4 (price index base=1991, y=yearly growth rates) (see Heilemann and Münch (1996)). These economic variables were selected out of more than 120 variables as possible candidates for stylized facts for business cycle characterization.

We use a four phase scheme (upswing (1), upper turning point (2), downswing (3) and lower turning point (4)) for the German business cycle. The 157 observations are classified based on economic and heuristic considerations. This data set is the learning sample introduced in the foregoing section. The quarters are distributed among the four phases as follows:

- upswing (1): 59 observations
- upper turning point (2): 24 observations
- downswing (3): 47 observations
- lower turning point (4): 27 observations

For a detailed characterization in economic terms cp. Heilemann and Münch (1996).

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 as, 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: Economic variables.

The four phases have, at least in the data set, a definite transition structure: $4 \to 1 \to 2 \to 3 \to 4 \to 1$ (where it is allowed to stay in the same phase). This structure will be utilized for classification (cp. section 2).

In the rest of the paper we apply the ideas of sections 2 and 3 to this data set. More details are given in Röhl (1998).

6 Variable Selection

In order to select the variables really characterizing the business cycle out of the 13 variables presented in the last section, for each possible subset of variables we construct a classification rule by exhaustive search using the densities $p_i(x)$ $(1 \leq i \leq g)$ estimated by QDA, LDA or DDA from the data (see section 3) and evaluate the performance (misclassification error) modelfree by leave one (period) out (see sections 2, 3).

Figure 2 shows the results of the different methods. The error rate displayed is the smallest leave one out error rate one can obtain with a subset of n variables $(1 \leq n \leq 13)$.

The best performance has QDA. LDA and DDA show a stagnation of the error rate for $n > 7$ and then the error rate even rises. For QDA the smallest error rate is reached for $n = 9$. This indicates that not more than 9 variables are needed. To use too many variables in a subset might lead to unstable parameter estimates and overfitting. This results in a higher (and possibly unreliable) error rate.

To get an impression what variables are involved in the best subsets, for QDA and

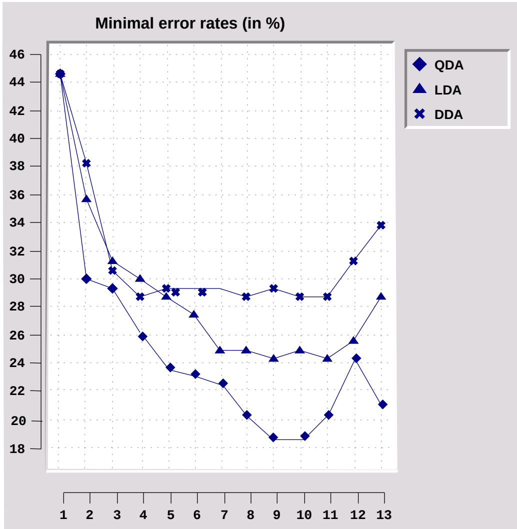

Figure 2: Comparison of the three normal models.

LDA table 2 shows the best eight subsets with two and four variables, respectively. QDA and LDA are put next to each other to facilitate direct comparison.

Rank	*Var*1	*Var*2	cv–error	*Var*1	*Var*2	*Var*3	*Var*4	cv–error	Model
1	*L*	*LC*	29.94	*Y*	*LC*	*PY*	*RS*	26.11	*QDA*
	L	*LC*	35.67	*Y*	*LC*	*PY*	*RS*	29.94	*LDA*
2	*L*	*PY*	36.94	*C*	*L*	*IE*	*PY*	26.75	*QDA*
	L	*PC*	36.31	*L*	*LC*	*PY*	*PC*	29.94	*LDA*
3	*IE*	*LC*	37.58	*L*	*IE*	*LC*	*RS*	27.39	*QDA*
	L	*PY*	39.49	*L*	*IE*	*LC*	*PC*	30.57	*LDA*
4	*C*	*PY*	40.76	*L*	*IE*	*LC*	*PY*	27.39	*QDA*
	*M*1	*LC*	40.12	*GD*	*L*	*IE*	*PY*	30.57	*LDA*
5	*Y*	*PY*	40.76	*Y*	*IE*	*LC*	*RS*	27.39	*QDA*
	Y	*LC*	40.76	*L*	*LC*	*PY*	*RL*	31.21	*LDA*
6	*C*	*LC*	40.76	*C*	*IE*	*LC*	*RS*	27.39	*QDA*
	Y	*RS*	40.76	*L*	*IE*	*LC*	*PY*	31.21	*LDA*
7	*IE*	*PY*	40.76	*L*	*IE*	*LC*	*PC*	28.03	*QDA*
	IE	*LC*	40.76	*GD*	*L*	*IE*	*LC*	31.21	*LDA*
8	*M*1	*LC*	41.40	*Y*	*C*	*IE*	*PY*	28.03	*QDA*
	L	*RS*	41.40	*L*	*LC*	*PY*	*RS*	31.21	*LDA*

Table 2: Ranking of subsets with 2 and 4 variables.

The first column shows the rank. The label "cv–error" denotes the (crossvalidated) leave one out error. Even when QDA and LDA have distinctly different error rates, the best variable combinations are similiar and the first rank is even identical. This shows that the importance of the different variables is somewhat "robust" against model variation. This simplifies interpretation.

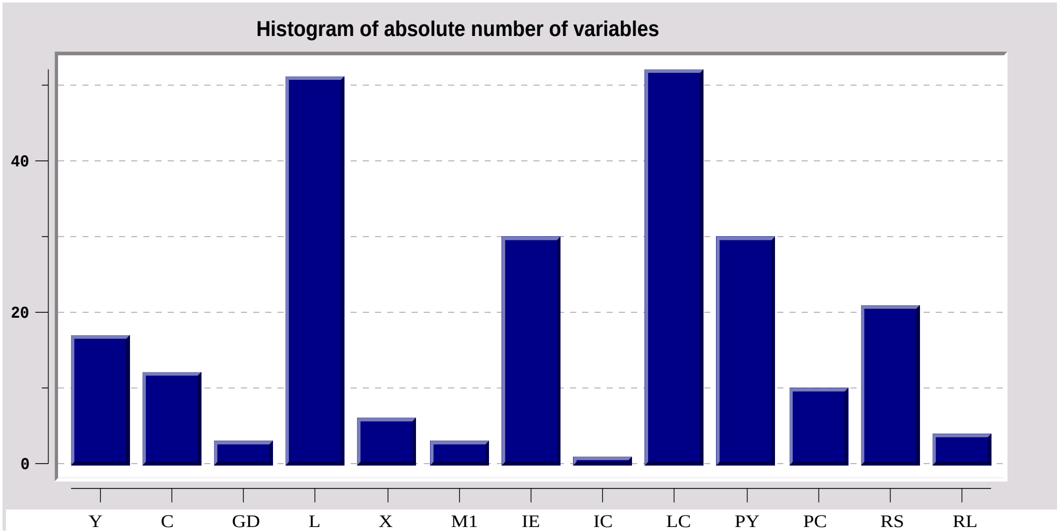

Figure 3: Importance of the variables in the data set from 55/4 – 94/4.

To get a more general perspective, we summarize table 2, the respective tables for $n = 1, 3$ variables, and the corresponding outcomes for the model DDA in a histogram (figure 3). The total number of variables reported this way is

$$\sum_{n=1}^{4} \text{number of models} \times \text{number of ranks} \times n = 3 \times 8 \times (1 + 2 + 3 + 4) = 240.$$

Note that the histogram is not a listing of individual variables generated by a univariate procedure. On the contrary, these variables are chosen by a multivariate procedure, i.e. in presence of all the other variables.

The dominance of variables L (wage and salary earners) and LC (unit labour cost) is obvious, followed by IE and PY. Of no importance are GD, X, M1, IC and RL.

7 Linear Dimension Reduction

7.1 Error rate minimization

In this section, we follow the procedure indicated in section 3 for dimension reduction by linear projection. First, we determine that two dimensional subspace by direct minimization in which the overall error rate is minimal under the assumption of normal densities with identical covariance structure in the groups. After the plane is selected, the data set is projected into it and a QDA is performed in the two dimensional space. Finally, the error rate is evaluated by leave one out.

In the original space densities are fixed according to LDA, and not to DDA, since LDA showed a better performance for more than five variables (figure 2). But how many variables should we take? In section 6, for LDA we mentioned the stagnation of the error rate from seven variables on. Therefore it suffices to use seven variables. Which variables out of the 13 should we take? On the one hand, we can take the best seven variables of figure 3 (variables Y, C, L, IE, LC, PY and RS), on the other hand the seven variables (C, L, X, IE, PY, RS and RL) figure 2 is based on. The two sets differ only in two variables, but the first subset contains the most important variables L and LC, so we prefer this one.

For the optimal two dimensional projected plane the boundaries of the four phases, determined by $p_{R,i}(x) = p_{R,j}(x)$ $(i \neq j)$, are displayed in figure 4. Note, that each group density $p_{R,i}(x)$ has its own covariance matrix because we performed QDA in the plane.

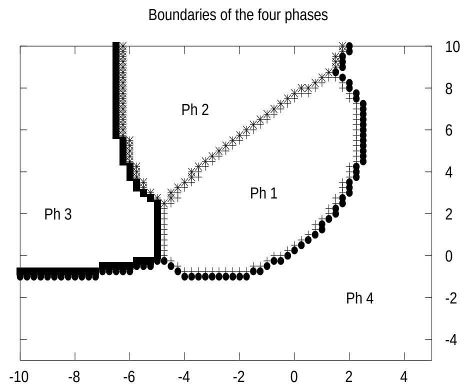

Figure 4: Splitting of the optimal projection plane into the four phases.

The leave one out error is 26.8%, compared to the variable combination L and LC (29.9%, table 2) only a decrease of 3% points. But the variables L and LC can be easily interpreted, linear combinations of seven variables not. This relies on the fact that small weights do not necessarily mean small importance and vice versa (Rencher (1992)).

7.2 Comparison with Combination L and LC

Figure 5 visualizes the boundaries of the group densities $p_i(x)$, having performed QDA in the plane spanned by the variables L and LC.

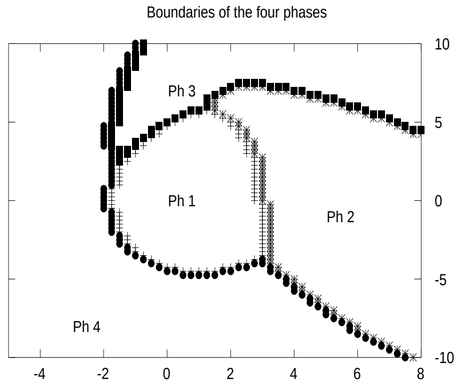

Figure 5: Splitting of the projection plane into the four phases. The horizontal axis corresponds to variable L (wage and salary earners) and the vertical axis to variable LC (unit labour cost).

The upswing is quite symmetric about the origin (phase 1). Small absolute values belong therefore to an upswing, more extreme ones lead to a different phase. The upper turning point (phase 2) can only be reached if the growth rate of the employment exceeds 3% and at the same time the growth rate of unit labour cost does not rise above 5%. Because then we are faced, even if employment is reduced by 1%, with the transition to downswing (phase 3). A change of employment smaller than -2% belongs to the lower turning point (phase 4), no matter how unit labour cost behaves. Even an increase of employment of 4% compared to last year's corresponding quarter and a reduction of unit labour cost about -7% does not alter the situation: we stay at phase 4. This shows how such figures can be easily interpreted to discuss different economic constellations.

A comparison of figure 5 with the 90 degrees clockwise rotated figure 4 leads to the conclusion, that the horizontal axis in figure 4 is mainly explained by unit labour cost and the vertical one by wage and salary earners. Thus, the partition resulting from linear dimension reduction can easily be interpreted by means of variables L and LC.

7.3 Transition Structure of the Phases

The phases have to follow the pattern $4 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ (or they stay at the same phase). Standard QDA classification does not take this into account, but assumes that the observations are independent. Let us, thus, incorporate the transition structure in the classification rule for the left out observation as indicated in section 2. Such "classifications with exact transitions" do not produce any forbidden transitions at all.

Table 3 compares the performance of the two procedures for the three data sets:

data set 1 Variables L, IE, LC and PY,

data set 2 Variables Y, LC, PY and RS,

data set 3 Variables L and LC.

and the original, so–called all variables, data set with all 13 variables.

Data set 1 was chosen because variables L, IE, LC, PY are the most important four variables concerning to variable selection (cp. fig. 3) and is the fourth best selection in table 2. Data set 2 was chosen because variables Y, LC, PY, RS build the best predicting group of four variables in the whole time period. Data set 3 was chosen because variables L, LC build the best predicting group of two variables in the whole time period. Table 3 clearly indicates the better performance of the classification with exact transitions not only concerning the number of forbidden transitions but also concerning the classification error. Note however that table 3 also seems to indicate that the proposed variables selection is not useful at all since the error rates are smallest for the all variables set. This is further discussed in the next section.

error rates and transitions	data set 1	data set 2	data set 3	all var.
error rate standard classification	27.3%	26.1%	29.9%	21.0%
error rate with exact transitions	17.2%	17.8%	19.1%	16.6%
No. of forbidden trans. st. class.	11	8	16	10
No. of forbidden trans. exact trans.	0	0	0	0

Table 3: cv–Error and number of forbidden transitions of the different procedures.

7.4 Forecasts

Next, we consider a prognostic/forecast situation. We construct the classification rule by means of QDA without the last complete cycle plus the very last 3 quarters (82/2 – 94/4) and classify the left out quarters afterwards. This reduces the learning sample to 106 quarters (phase 1: 32, phase 2: 18, phase 3: 38 and phase 4: 21). The quarters in the test period (phase 1: 29, phase 2: 6, phase 3: 9 and phase 4: 7) are classified by standard QDA and classification with exact transitions.

error rates in the forecasts	data set 1	data set 2	data set 3	all variables
error rate standard classification	23.5%	62.7%	29.4%	60.8%
error rate with exact transitions	25.4%	62.7%	19.6%	56.9%

Table 4: Error rates in the forecast situation

Also in this situation the superiority of the exact transitions is clear. E.g., in the case of data set 3 we reduce the number of errors in phase 2 from 3 to 0, in phase

3 from 8 to 6 and in phases 1 and 4 they stay the same. Overall, we reduce the number of errors from 15 to 10 and do not have any forbidden transitions.

Note that the surprisingly bad performance of data set 2 as reported in table 4 can be interpreted as a strong hint that GNP (Y) and short term interest rate (RS) are less significant for the characterization of the business phase in the test period than in the learning period before. This might also cause the bad performance of the all variables set, and indicates that careful variables selection might be important at least for prediction. Note that this is not discovered by leave-one-out resampling, which is, at least in our example, much too optimistic.

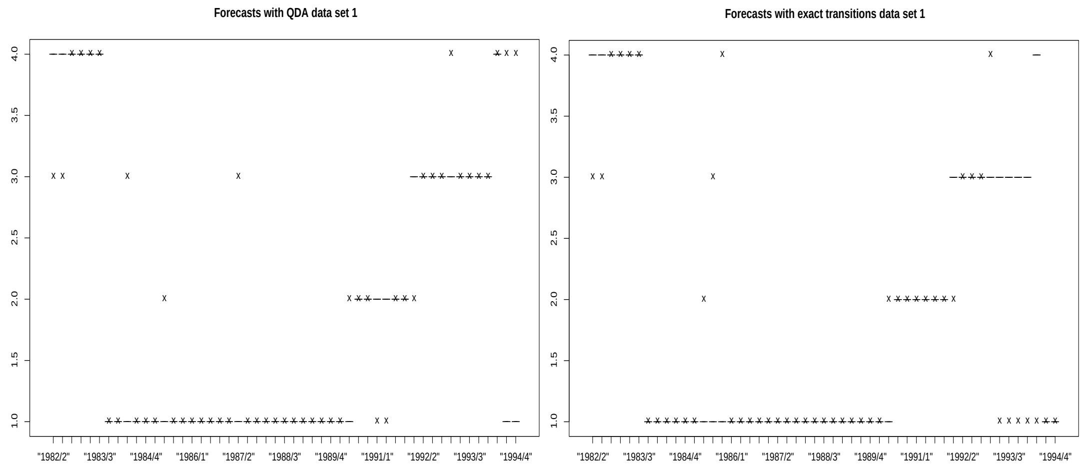

Figure 6: Comparison of the forecasts to the realized phases based on data set 1; Left: Forecasts with QDA, Right: Forecasts with exact transitions

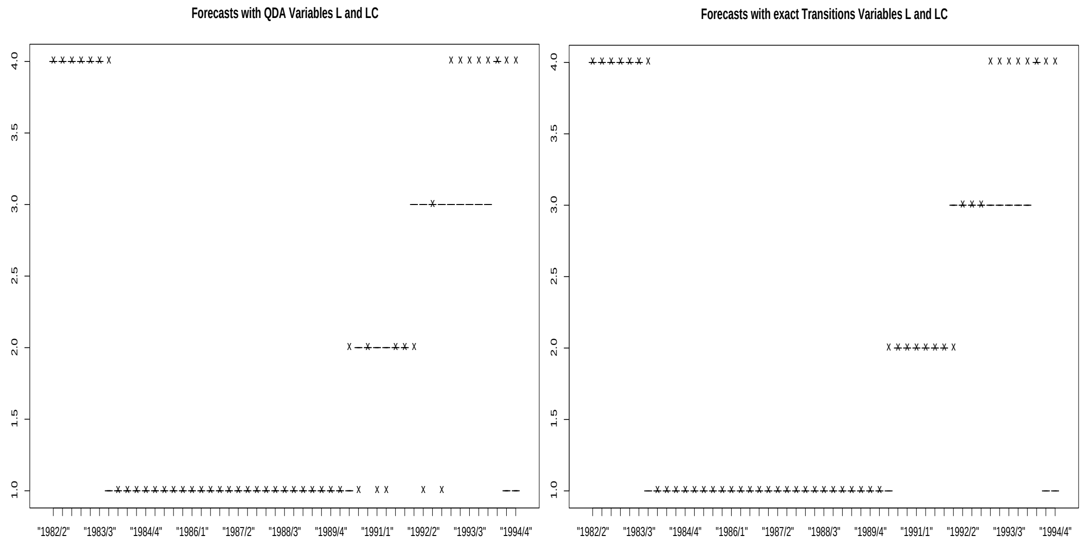

Figure 7: Comparison of the forecasts to the realized phases based on data set 3 ; Left: Forecasts with QDA, Right: Forecasts with exact transitions

In figure 7 the realized phases are plotted against time as lines, and forecasts based on variables L and LC are indicated as crosses. Figure 6 shows forecasts based on data set 1. The false predictions around 1986 in figure 6 can be economically explained by an interrupted upswing in 1986/1 (Tichy (1994)). Note that if in the case of exact transitions a phase is left once, one has to go through the whole cycle to return to this phase. This explains the short cycle around 1986. The four variables of data set 1 thus seem to be more sensible to changes in the state of economy than the two variables L and LC in data set 3. In data set 1 the errors in the different phases are, by exact transitions, reduced from 6 to 4 in upswing, 2 to 0 in the upper turning point and are increased from 2 to 6 in downswing and from 2 to 3 in the lower turning point. So the difference of 2% in table 4 results from one additional error.

8 Conclusion

Classification as a multivariate statistical tool has clearly proved its ability to characterize business cycle phases utilising dimension reductions with optimal forecast ability. The consideration of the transition structure of the phases not only guaranteed sensible classification but even improved the error rate.In an example problem, mainly four of 13 economic variables proved to adequately characterize classification into four business cycle phases.

Acknowledgment

This work has been supported by the Collaborative Research Centre "Reduction of Complexity in Multivariate Data Structures" (SFB 475) of the German Research Foundation (DFG). Also we would like to thank Prof. Dr. A. Kneip and Prof. Dr. U. Heilemann for their helpful comments.

References

I. O. Bohachevsky, M. E. Johnson, M. L. Stein (1986), Generalized Simulated Annealing for Function Optimization, *Technometrics*, **28**, 3, 209-217.

A. F. Burns, W. C. Mitchell (1946), Measuring business cycles , NBER, New York

F. X. Diebold, G. D. Rudebusch (1996), Measuring business cycles: a modern perspective, *The Review of Economics and Statistics* **78**, 67-77

A. J. Filardo (1994), Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics* **12**, 299-308

J. D. Hamilton (1989), A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357-384

U. Heilemann and H. J. Münch (1996), West German Business Cycles 1963-1994: A Multivariate Discriminant Analysis, *CIRET–Conference in Singapore, CIRET–Studien 50.*

R. E. Lucas (1983), Understanding business cycles, *in: Studies in Business-Cycle Theory*, Cambridge, Mass. und London, 215-240

G. J. McLachlan (1992), *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons .

J. R. Meyer and D. H. Weinberg (1975a), On the classification of economic fluctuations, *Explorations in Economic Research*, **2**, 167-202 .

J. R. Meyer, D. H. Weinberg (1975b), On the usefulness of cyclical taxonomy. Paper presented at the *12th CIRET Conference, 25-27th* June 1975

A. C. Rencher (1992), *Interpretation of Canonical Discriminant Functions, Canonical Variates, and Principal Components*, The American Statistician , **46**, 3, 217-225 .

M. C. Röhl (1998), Computerintensive Dimensionsreduktion in der Klassifikation, *Reihe: Quantitative Ökonomie Band* **95**, Josef Eul Verlag, Germany.

W. Siedlecki und J. Sklansky (1989), A note on genetic algorithms for large–scale feature selection, *Pattern Recognition Letters*, **10**, 335-347 .

J. H. Stock, M. W. Watson (1991), A probability model of the coincident economic indicators, in: K. Lahiri, G.H. Moore (Eds.), *Leading economic indicators: new approaches and forecasting records*, Cambridge University Press, 63-89

S. M. Weiss, C. A. Kulikowski (1991), Computer systems that learn, Morgan Kaufmann, San Francisco