Stability of multivariate representation of business cycles over time

Claus Weihs*†, Ursula Garczarek*[1]

*Fachbereich Statistik and SFB 475, Universität Dortmund,
D-44221 Dortmund, Germany
†weihs@statistik.uni-dortmund.de

May 8, 2002

Abstract

In order to replace the univariate indicators standard in the literature (cp. [Opp96]) by a multivariate representation of business cycles, the relevant 'stylized facts' are to be identified which optimally characterize the development of business cycle phases. Based on statistical classification methods we found that, somewhat surprisingly, only two variables, 'wage and salary earners' and 'unit labor costs', are able to characterize the German business cycle not only the most stable over all sub-cycles but also with a quite reasonable error rate.

Keywords: business cycle, prediction, multivariate characterization, classification, leave-one-cycle-out, cross validation

1 Introduction

In order to replace the univariate indicators standard in the literature (cp. [Opp96]) by a multivariate representation of business cycles, statistical classification methods were applied to quarterly after-war data of the German economy classified into four business classes called upswing, upper turning points, downswing, and lower turning points. The aim was to find multivariate models of 'stylized facts' with maximum predictive power, i.e. with maximum ability predicting the correct business cycle phase from the state of the economy. In order to maximize predictive power, the cross-validation methods standard in statistical analysis[WK91] were adapted to business cycle analysis by replacing techniques like leave-one(-observation)-out- or 10-fold-cross-validation by the so-called double-leave-one-cycle-out analysis. This way, we looked for those 'stylized facts' being best able to characterize the

[1]This work has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

business cycle over the whole time period available. This cross-validation particularly produces classification rules for each individual business cycle, thus allowing for the assessment of the stability of the multivariate characterization in the six business cycles available in the data. The results give a somewhat unexpected insight into the German economy: the two variables 'wage and salary earners' and 'unit labor costs' play a stable dominant role in the characterization of business cycles.

The organization of the paper is as follows. In Section 2 the data is introduced on which the analysis is based upon. In Section 3 a mathematical problem formulation is given. Section 4 briefly introduces the classification methods used in the paper and the kinds of classification rules resulting from them. In Section 5 the double-leave-one-cycle-out cross-validation method is developed. Section 6 gives the results of the classification methods, and Section 7 discusses the results from an economic standpoint. Section 8 concludes the paper.

2 Data

The data set consists of 13 so-called 'stylized facts' cp. [Luc87] for the (West-) German business cycle and 157 quarterly observations from 1955/4 to 1994/4 (price index base is 1991). The stylized facts (and their abbreviations) are real-gross-national-product-gr (Y), real-private-consumption-gr (C), government-deficit (GD), wage-and-salary-earners-gr (L), net-exports (X), money-supply-M1-gr (M1), real-investment-in-equipment-gr (IE), real-investment-in-construction-gr (IC), unit-labor-cost-gr (LC), GNP-price-deflator-gr (PY), consumer-price-index-gr (PC), nominal short term interest rate (RS), and real long term interest rate (RL). The abbreviation 'gr' stands for growth rates relative to last year's corresponding quarter.

We base our analyses on the data preparation in [HM96] where the selection of the above 'stylized facts' out of more than 100 available variables of the German economy is described, as well as the assignment of one of four business cycle phases to each quarter from 1955/4 to 1994/4. The phases of the used 4-phase business cycle scheme are called 'upswing' (up), 'upper turning points' (utp), 'downswing' (down), and 'lower turning points' (ltp). This classification was supposed to be the 'correct' classification for the purpose of our study.

3 Classification of Business Cycle Phases

The multivariate characterization of business cycles can mathematically be described as a multivariate classification rule.

Classification deals with the allocation of objects to, say, G predetermined groups (or classes). In our application the objects will be time periods (quarters), the groups the business cycle phases. For each object, variables $X_{k,t}$, $k = 1, ..., K$, $t = 0, ..., T$ considered to be important for discriminating between the groups are assumed to be observable at all time periods. Such variables can be continuous (GNP, consumption etc.) or discrete (number of firms, number of inhabitants etc.). In the following, moreover, we assume that the vector of these variables $\vec{X}_t$ has at each time period $t \in \mathbb{N}$ values in a portion $\mathbf{B}$ of the K-dimensional real space $\mathbf{B} \in \mathbb{R}^K$. Based on some pre-classified objects (training sample) a classification rule is learned incorporating the information inherent in the training.

We then classify a member of a sequence of future objects at time periods $t = t_o + h, h = 1, ..., H$ based on all observations $\vec{x}_{t_o}$,$\vec{x}_{t_o+1}$,...,$\vec{x}_t$ up to and at that time period, and the last known state s_{t_0}.

In order to construct the classification rule, the information given in the training sample is typically "encoded" in terms

(1) of the unknown parameters of an assumed conditional distribution of the $X_{k,t}$, $k = 1, ..., K$, for objects belonging to one of the groups at time period t given all information from the past, and

(2) in terms of some parameters for the probability to be in any of the groups at time period t given all information from the past.

The product of the evidence from the current observation in (1) and the a-priori probabilities for the current class in (2) result in a probability estimate for the current class.

Since the classes are known a-priori, all observations related to one class can be used for the estimation of the parameters. New objects with observed variables vector $\vec{X}_t = \vec{x}_t$ are classified to group $g \in \{1, ..., G\}$ if the estimated probability of this group is highest given all available information from the past. The goodness of classification depends on the class of distributions we use. Often one uses distributions with a small number of parameters in order to facilitate estimation. Therefore, typically strong independence assumptions are made about time-dependencies. Additionally, one uses popular densities like the normal one: only the mean vector and a measure of interrelation - the covariance matrix - have to be specified.

If there is a choice between different classification rules, the goal is to choose that classification rule which minimizes the misclassification error (error rate) of new objects.

4 Classification methods and classification rules

The compared classification methods include classical standard procedures like Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA). A recently developed method [WRT99] based on a projection pursuit algorithm selects optimal linear combinations of the original variables by a leave-one-observation out cross validation procedure. This method is combined with both, LDA and QDA. All these methods learn parameters of the conditional distribution of variables given phases, as if all observations in the training sample belonging to a certain phase were an i.i.d. sample from this distribution.

Another more modern method is a Continuous Dynamic Bayesian Network with a certain 'rake'-structure, tailored for classification in dynamic domains, named "CRAKE" in [SW99]. CRAKE represents a certain markov regime switching model, c.p. [Kro97].

Only CRAKE models directly time-dependencies in the conditional distribution of variables given business phases. To be able to take advantage of the knowledge about the cyclical structure of the succession of phases, for the other methods we added the structure of a hidden Markov model. That means, we model a first order Markov chain for the succession of phases, and the distribution of variables is independent of the past given the current phase. This idea was introduced by [KÖ98]. Details are given in [SW01].

All these methods are applied either to all the variables mentioned in Section 2, or to certain subsets discussed later.

The classification rules corresponding to the different classification methods all lead to different partitions of the corresponding space of predictor variables in regions of assumed highest evidence from the predictor variables for each of the phases. LDA, as well as LDA-MEC1 both derive partitions of the space of involved variables with linear borders where each subregion is related to one and only one business cycle phase. QDA, as well as QDA-MEC1 both derive partitions with nonlinear borders. CRAKE partitions the space of the involved variables and the lagged variables. The projections of the inter-variable partitions in the space of the non-lagged variables have non-linear borders, and resemble very much those of QDA. The intra-variable borders in the space of predictor variables and their predecessors are also non-linear.

For a more detailed discussion of borders corresponding to classification rules of various classification methods cp., e.g., [WBS93].

5 Double leave-one-cycle-out cross validation

The development of an optimal classification rule should be related to the optimization of predictive power since a rule, once developed, should optimally 'predict' the groups (classes, business cycle phases) of future objects (time periods). For the maximization of predictive power cross-validation methods are standard in statistics [WK91]. Typical variants are leave-one(-observation)-out cross-validation, and 10-fold cross-validation. In the former variant one observation is left out in order to be predicted by a classification rule derived from the other observations. In the latter case the observations are partitioned into 10 equally sized parts, predicting one part by a rule derived from the observations in the other 9 parts.

Obviously, both methods do not relate cross-validation to the structure of our data, i.e. to business cycles. Indeed, what would be most adequate here is to use a 6-fold cross-validation but not with equal sized parts. Instead, the parts should be equal to the business cycles observed in the data. If one then leaves out one business cycle, one could test whether this cycle is predictable by means of information from the other cycles. Note, however, that as an objection to this method one might argue that time structure is partly ignored because information from later cycles is used to predict the left out cycle. This drawback was accepted, though, because of the lack of enough data for deriving a reliable classification rule for early cycles if only past cycles would be allowed for training.

We use leave-one-cycle-out (l1co) cross-validation for the determination of error rates for a classification method. We first leave out each whole business cycle once. This is the outer l1co loop. The data from the other 5 cycles is then used to derive a 'best' classification rule for these cycles. All methods have intrinsic definitions of what is 'best': 'best' according to theoretical predictive power according to certain distributional assumptions is basic to LDA, QDA, and CRAKE, where MEC1 additionally finds a 'best' rule with respect to a leave-one-observation out error. Additionally, we analyzed variants of these methods including model selection steps: Variable selection for LDA, QDA, and CRAKE, dimension selection for MEC1. In order to judge the predictive power of potential rules, we re-apply (double!) leave-one-cycle-out to the 5 cycles, the inner l1co loop. We derive a classification rule for each group of 4 cycles, and test this rule on the left-out 5th cycle. The mean of the corresponding 5 error rates, called the mean l1co training error, gives the predictive power of the classification method on this group of 5 business cycles. The classification rule derived from the data of the 5 cycles is then applied to the left-out 6th cycle giving the so-called prediction error.

The difference of the mean l1co training error and the prediction error is used as a rough measure for the similarity of the test set and the training sets, a negative sign indicating problems with extrapolation from training sets to test set. The prediction error itself is a measure for the quality of the derived classification rule for the test set. In comparing different classification methods the minimum prediction error indicates the most adequate rule. The mean of the prediction errors characterize the overall predictive power of the classification method.

Note that with this method we particularly derive so-called 'local', cycle specific, measures of predictive power which reflect 'local' properties of cycles. Thus, we are able to assess the stability of rules over the different cycles: We say a rule is stable in it's structure, if the best variables or the best dimension does not change too much on the six training sets. And we say a rule is stable in its quality, if prediction errors and mean l1co errors are stable.

6 Classification results and resulting models

6.1 Linear discriminant analysis

Classical linear discriminant analysis was performed in two variants: using all 13 variables to classify the current phase (LDA-all) and with variable selection of the best two variables from these 13 (LDA-best2), based on a leave-one-cycle-out procedure on the training set. Additionally, two variants of the introduced projection pursuit algorithms are LDA-procedures: LDA-MEC1-2D minimizes the leave-one-observation-out error of a two-dimensional linear combination of all 13 variables that is then used as new input variables in LDA. LDA-MEC1-bestD selects additionally the best dimension (among 1-8) of such a linear combination in a leave-one-cycle-out procedure.

The prediction error rates for LDA based on all variables were unacceptable, at least for the first four cycles (Table 2). Looking for the two most important variables was motivated by results of [Röh98] and [WRT99]. Astonishing enough, the two best predicting variables for each individual business cycle were always the same with LDA: LC and L, i.e. 'unit labor costs' and 'wage and salary earners'. Unfortunately, also for these two variables the prediction errors were unacceptably high for cycles 2 and 4, namely 50% and 67% (Table 2). On the other hand, for cycles 1 and 3 the improvement by avoiding overfitting by reducing the number of involved variables was high. For cycle 2 there appears to exist better variables (combinations) since the error rate of LDA-best2 was even worse than of LDA-all. And indeed, LDA-MEC1-2D found a better set of two directions in the 13 dimensional space with the same prediction error rate as LDA-all. The weights of the original variables on these directions were found as indicated in Table 1.

	Dir1	Dir2	Dir1 standardized	Dir2
IE	62	-10	7	-1
C	-148	26	-53	9
Y	53	104	18	35
PC	-181	143	-96	76
PY	624	-618	357	-353
IC	10	7	1	1
LC	-131	5	-38	1
L	131	603	74	341
M1	-77	-28	-16	-6
RL	560	-421	371	-278
RS	-414	165	-162	65
GD	-115	-131	-41	-47
X	105	32	50	15

Table 1: LDA-MEC1-2D‘s directions ($*10^3$) of best linear combinations on cycle 2

Note that one has to standardize these weights by the variables' standard deviations (cp. Table 1), at least, in order to interpret them! Moreover, note that the best number of dimensions found by LDA-MEC1-bestD was 3 for cycle 2 giving the same error rate of 44% as LDA-all and LDA-MEC1-2D. Also note that the relatively high dimensions 5 and 6 found to be best for cycles 1 and 3 gave much worse predictions than, e.g., LDA-best2. This is a strong argument against such high dimensions. Also, in the mean LDA-best2 gave the best prediction results.

Looking at the similarity of test sets and training sets measured by the difference of mean training error and prediction error (cp. Table 4) in the mean similarity is high. However, the individual error rates are most of the times lower for prediction than in training.

Though LDA-best2 shows a high structural stability, it has no high stability in its absolute performance: prediction errors range from 18% to 67%. This is also reflected in Table 2 that shows that the similarity of cycles is pretty low from the perspective of LDA-best2.

6.2 Quadratic discriminant analysis

Like LDA, we performed classical quadratic discriminant analysis without and with variable selection (QDA-all and QDA-best2), and with two projection pursuit variants (QDA-MEC1-2D and QDA-MEC1-bestD). Inspired by the results of LDA-best2, we looked additionally at quadratic discriminant analysis based on 'unit labor costs' and 'wage and salary earners', only (QDA-LC,L).

The results were qualitatively similar (cp. Table 6) to those of the linear analysis. QDA-all delivered unacceptable prediction errors, QDA-LC,L was best in the mean, QDA-MEC1-2D was able to improve QDA-LC,L only in two cycles, namely cycles 3 and 5, and QDA-MEC1-bestD never improved QDA-MEC1-2D. One has to mention, though, that the best

all	best2	MEC1-2D	MEC1-bestD
Error on test cycles			
0.78	0.33 (LC,L)	0.56	0.72 (D=5)
0.44	0.50 (LC,L)	0.44	0.44 (D=3)
0.41	0.18 (LC,L)	0.35	0.41 (D=6)
0.67	0.67 (LC,L)	0.67	0.67 (D=1)
0.28	0.25 (LC,L)	0.41	0.41 (D=2)
0.27	0.21 (LC,L)	0.31	0.48 (D=1)
Mean error			
0.47	0.36	0.46	0.52

Table 2: LDA's prediction errors

all	best2	MEC1-2D	MEC1-bestD
Mean l1co-error on training set			
0.39	0.35	0.41	0.33
0.54	0.32	0.52	0.47
0.49	0.41	0.55	0.51
0.44	0.29	0.48	0.45
0.54	0.34	0.50	0.50
0.51	0.42	0.51	0.45
Mean of mean l1co-errors			
0.49	0.35	0.50	0.45

Table 3: LDA's mean training errors

number of dimensions found by QDA-MEC1-bestD was always smaller than 4, i.e. never as high as 5 and 6 as found by LDA-MEC1-bestD, and is thus structurally more stable. QDA-best2, though is less stable than LDA-best2, as LC and L were chosen only for cycles 3,4, and 5, whereas for cycles 1 and 2 the variables L and RS were chosen, and for cycle 6 none of the variables L or LC was chosen, but Y and RS. The most important result, though, is that QDA-LC,L was only in cycle 2 better than LDA-LC,L. Overall, only on cycles 2 and 5 any QDA-procedure could outperform LDA-LC,L. The QDA-MEC1-2D and QDA-MEC1-bestD results on cycle 5 show the existence of a two-dimensional combination of all variables that has about the same mean l1co training error as LDA-LC,L (42% compared with 41%) and a much better performance in predicting cycle 5 (19% compared with 50%).

The corresponding directions can be characterized by the weights of the standardized variables as in Table 5.

all	best2	MEC1-2D	MEC1-bestD
Difference of prediction and training errors			
0.38	-0.01	0.15	0.40
-0.10	0.18	-0.08	-0.03
-0.08	-0.23	-0.20	-0.09
0.22	0.38	0.19	0.22
-0.26	-0.09	-0.09	-0.09
-0.24	-0.21	-0.20	0.03
Mean of difference			
-0.01	0.00	-0.04	0.07
Range of difference			
0.65!!	0.61	0.39	0.49

Table 4: LDA's differences of errors

	Dir1	Dir2	Dir1 standardized	Dir2 standardized
IE	-61	63	-7	7
C	242	60	87	21
Y	-75	304	-26	103
PC	398	172	249	108
PY	-425	-750	-253	-447
IC	-3	7	-0	1
LC	170	127	51	38
L	-160	340	-94	202
M1	125	-60	29	-14
RL	-576	-174	-394	-119
RS	389	41	170	18
GD	183	-323	71	-125
X	-81	190	-38	90

Table 5: QDA-MEC1-2D's directions ($*10^3$) of best linear combinations on cycle 5

Moreover, note that QDA prediction errors are nearly as often lower than the corresponding training errors as vice versa (cp. Table 8).

6.3 Continuous RAKE-Method

The CRAKE model is a Markov switching vector autoregressive model of first order, abbreviated as MS-VAR(1) according to [Kro97], with a certain covariance structure. More precisely, for given phase $s_t \in \{1, ..., S\}$ the vector $\vec{x}_t \in I\!R^K$ is generated by a first-order vec-

all	LC,L	best2	MEC1-2D	MEC1-bestD
		Error on test cycles		
0.50	0.50	0.61 (L,RS)	0.56	0.56 (2D)
0.75	0.38	0.69 (L,RS)	0.50	0.50 (2D)
0.47	0.29	0.29 (LC,L)	0.24	0.59 (3D)
0.75	0.67	0.67 (LC,L)	0.67	1.00 (1D)
0.72	0.41	0.41 (LC,L)	0.19	0.19 (2D)
0.31	0.23	0.46 (Y,RS)	0.33	0.33 (2D)
		Mean error		
0.58	0.41	0.52	0.42	0.53

Table 6: QDA's prediction errors

all	LC,L	best2	MEC1-2D	MEC1-bestD
		Mean l1co-errors on training sets		
NaN	0.53	0.48	0.45	0.45
0.50	0.45	0.44	0.38	0.38
0.67	0.42	0.42	0.53	0.51
0.60	0.40	0.40	0.46	0.43
0.64	0.48	0.48	0.42	0.42
NaN	0.48	0.39	0.52	0.52
		Mean of mean l1co-errors		
0.60	0.46	0.44	0.46	0.45

Table 7: QDA's mean training errors. The inner leave-one-cycle-out error of QDA-all can not be calculated for cycles 1 and 6, because for that purpose - among others - one would have to learn parameters on a training set without these two cycles. On this set, though, there are only 13 observations on the UTP-group, which is not enough for the learning.

tor autoregressive model with diagonal covariance matrices, such that for each $k, k = 1, ..., K$, we get a model equation

$$x_{k,t}(s_t) = \mu_k(s_t) - \beta_k(s_t)x_{k,t-1} + u_{k,t}$$

with $u_{k,t}$, $k = 1, ..., K$, $t = 1, ..., T$, being independently normally distributed, $u_{k,t} \quad \mathcal{N}(0, \sigma_k)$ given $s_t, t = 1, ..., T$. For the generating process of states we assume - just like in the hidden Markov model - a first order Markov chain.

The CRAKE method we tested in variants with

(1) all variables, (CRAKE-all)

(2) with a model selection for the best two variables (CRAKE-best2), and

(3) with variables LC and L only (CRAKE-LC,L).

all	LC,L	best2	MEC1-2D	MEC1-bestD
Difference of training and prediction error on cycles				
NaN	-0.03	0.13	0.10	0.10
0.25	-0.08	0.24	0.12	0.12
-0.20	-0.12	-0.12	-0.30	0.08
0.14	0.27	0.27	0.21	0.57
0.08	-0.08	-0.08	-0.23	-0.23
NaN	-0.26	0.07	-0.19	-0.19
Mean of difference				
0.07	-0.05	0.09	-0.05	0.07
Range of difference				
0.45*	0.52	0.39	0.51	0.81!!

Table 8: QDA's differences of errors

This surprisingly leads to a clear improvement of the result for cycle 4 from 67% error for LDA-LC,L to 42% for CRAKE-LC,L (cp. Tables 2 and 9).

The corresponding CRAKE model looks as follows:

$$1stphase: \quad LC_t = 0.81 + 0.63LC_{t-1}, \quad L_t = 0.44 + 0.82L_{t-1}$$
$$2ndphase: \quad LC_t = 2.86 + 0.21LC_{t-1}, \quad L_t = 0.99 + 0.59L_{t-1}$$
$$3rdphase: \quad LC_t = 1.21 + 0.80LC_{t-1}, \quad L_t = -0.27 + 0.93L_{t-1}$$
$$4thphase: \quad LC_t = -1.14 + 0.96LC_{t-1}, \quad L_t = -0.48 + 0.78L_{t-1}$$

Concerning the mean prediction error, the CRAKE method based on all variables was best (Table 9). This method was also the overall best for cycle 2, but sharing the performance of exactly 0.375% prediction errors with QDA-LC,L. Like with LDA and QDA the selection of LC,L is quite stable though the CRAKE model is substantially different from QDA and LDA: From the 78 possible combinations of two out of 13 variables, the pair LC,L was selected 4 of 6 times by CRAKE-best2. And any time another pair was selected (cycles 3 and 6) the performance on the left-out cycle decreased.

Concerning similarity, only cycle 6 is problematic for CRAKE-LC,L (cp. Table 11). Additionally, the absolute prediction error of CRAKE-LC,L on the 4th cycle, on which all other methods have high difficulty, is lowest among all models (cp. Tables 2, 6, 9). This confirms the impression that in LC and L one finds a stable cross-cycle information about the interplay of stylized facts and phases.

6.4 Comparison of classification rules

Finally, the overall best prediction error rates are compared to the error rates obtained by our "standard method", i.e. LDA-LC,L (cp. Table 12). Obviously, LDA-LC,L is only clearly suboptimal for cycles 2 and 4. Even better, best models for these cycles cycle 4 are also based on variables LC and L only. Moreover, cycles 1,2, and 4 can be predicted clearly less exact than cycles 3,5, and 6.

Overall, from the modeling standpoint variables LC and L are clearly the most important

all	LC,L	best2
Error on test cycles		
0.56	0.44	0.44 (LC,L)
0.38	0.44	0.44 (LC,L)
0.47	0.47	0.94 (C,PY)
0.58	0.42	0.42 (LC,L)
0.25	0.47	0.47 (LC,L)
0.30	0.40	0.52 (RS,GD)
Mean error		
0.42	0.44	0.54

Table 9: CRAKE's prediction errors

all	LC,L	best2
Mean l1co-errors on training sets		
0.44	0.38	0.38
0.44	0.40	0.40
0.41	0.51	0.48
0.37	0.34	0.34
0.48	0.36	0.46
0.44	0.61	0.45
Mean of mean l1co-errors		
0.43	0.43	0.40

Table 10: CRAKE's mean training errors

for the characterization of the German business cycle. In this respect, the outcome of our analysis is astonishing stable over time and models.

7 Economic implications

Surely, there might be other models delivered by other classification methods leading to even better predictions than in our study. From our analysis, however, extreme 'multivariate' dimension reduction to only two characteristics of the German business cycle, namely L ('wage and salary earners') and LC ('unit labor costs'), appears to be well reasonable. This is true even 'locally', i.e. for each individual business cycle of the German Federal Republic. Thus, from an economic standpoint one might have to stress that business cycle development in Germany was mainly dependent on (the growth rates of) the number of employees and on labor costs.

all	**LC,L**	**best2**
Difference of training and prediction error		
0.12	0.07	0.07
-0.06	0.03	0.03
0.07	-0.04	0.46
0.21	0.08	0.08
-0.22	0.11	0.11
-0.14	-0.22	0.07
Mean of difference		
-0.00	0.01	0.14
Range of difference		
0.44	0.33	0.42

Table 11: CRAKE's differences of errors

LDA-LC,L	overall best	Model
0.33	0.33	LDA-LC,L
0.50	0.38	CRAKE-all, QDA-LC,L
0.18	0.18	LDA-LC,L
0.67	0.42	CRAKE-LC,L
0.25	0.19	QDA-mec1-two
0.21	0.21	LDA-LC,L

Table 12: Comparing LDA-LC,L with best models according to prediction error

8 Conclusion

In order to even better support our findings, there is need for a method selecting BEST PREDICTING classification rules for the different cycles out of the data, taking into account the other cycles because of the obvious interrelation of different cycles, and because of lack of observations.

References

[HM96] Ullrich Heilemann and Heinz J. Münch. West german business cycles 1963-1994: A multivariate discriminant analysis. In *CIRET–Conference in Singapore, CIRET–Studien 50*, 1996.

[KÖ98] Lasse Koskinen and Lars-Erik Öller. A hidden markov model as a dynamic bayesian classifier, with an application to forecasting business-cycle turning points. Technical report, National Institute of Economic Research, 1998. 59.

[Kro97] Hans-Martin Krolzig. *Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis.* Springer, Berlin, 1997.

[Luc87] Robert E. Lucas. *Models of business cycles.* Basil Blackwell, 1987.

[Opp96] Karl Heinrich Oppenländer. *Konjunkturindikatoren.* R. Oldenbourg Verlag, München, 2 edition, 1996.

[Röh98] Michael C. Röhl. *Computerintensive Dimensionsreduktion in der Klassifikation.* Josef Eul, Lohmar, 1998.

[SW99] Ursula M. Sondhauss and Claus Weihs. Dynamic bayesian networks for classification of business cycles. Technical report, SFB 475, University of Dortmund, 1999. 17/99.

[SW01] Ursula M. Sondhauss and Claus Weihs. Incorporating background knowledge for better prediction of cycle phases. Technical report, SFB 475, University of Dortmund, 2001. 24/01.

[WBS93] Claus Weihs, W. Baumeister, and H. Schmidli. Classification methods for multivariate quality parameters. *Journal of Chemometrics*, 7:131–142, 1993.

[WK91] S. M. Weiss and C. A. Kulikowski. *Computer Systems that Learn.* Morgan Kaufmann, San Francisco, 1991.

[WRT99] Claus Weihs, Michael C. Röhl, and Winfried Theis. Multivariate classification of business phases. Technical report, SFB 475, University of Dortmund, 1999. 26/99.