Stability of multivariate representation of business cycles over time*

Claus Weihs, Ursula Garczarek
Fachbereich Statistik, Universität Dortmund
{weihs,garczarek}@statistik.uni-dortmund.de

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 subcycles but also with a quite reasonable error rate.

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 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 we introduce the data, the problem, the classification methods under consideration, and the way to compare them. Section 3 gives the results of the classification methods, and Section 4 discusses the results from an economic standpoint and concludes the paper.

*This work has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

2 Design of Comparison

2.1 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.

2.2 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.

2.3 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 we consider most adequate here is to use a 6-fold cross-validation but not with equal sized parts. Instead, the parts are equal to the business cycles observed in the data. Leaving out one business cycle, one can test whether this cycle is predictable by means of information from the other cycles. We use this 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: Selection of the best two predictors for LDA (L2B), QDA (Q2B), and CRAKE (C2B). We select the best projection in 2 dimensions for MEC1 (M2D), and the best projection in the best number of dimensions for MEC1 (MBD). In order to judge the predictive power of potential rules, we re-apply (double!) l1co 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 basic models of LDA, QDA, and CRAKE based on all predictors are called Lall, QAll, and CAll.

The prediction error 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.

3 Classification results and resulting models

3.1 Linear discriminant analysis

Lall	L2B	M2D	MBD
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 1: LDA's prediction errors

The prediction error rates for LDA based on all variables were unacceptable, at least for the first four cycles (Table 1). 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 1). 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 L2B was even worse than of LAll. And indeed, LM2D found a better set of two directions in the 13 dimensional space with the same prediction error rate as LAll. Moreover, note that the best number of dimensions found by LMBD was 3 for cycle 2 giving the same error rate of 44% as LAll and LM2D. 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., L2B. This is a strong argument against such high dimensions. Also, in the mean L2B gave the best prediction results.

Though L2B 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 1 that shows that the similarity of cycles is pretty low from the perspective of L2B.

3.2 Quadratic discriminant analysis and CRAKE

Inspired by the results of L2B, we looked additionally to those variants of QDA and CRAKE based only on 'unit labor costs' and 'wage and salary earners', QLcL and CLcL. The results of QDA were qualitatively similar (cp. Table 2) to those of the linear analysis.

QDA					CRAKE		
all	LC,L	2B	M2D	MBD	all	LC,L	2B
Error on test cycles							
0.50	0.50	0.61 (L,RS)	0.56	0.56 (2D)	0.56	0.44	0.44 (LC,L)
0.75	0.38	0.69 (L,RS)	0.50	0.50 (2D)	0.38	0.44	0.44 (LC,L)
0.47	0.29	0.29 (LC,L)	0.24	0.59 (3D)	0.47	0.47	0.94 (C,PY)
0.75	0.67	0.67 (LC,L)	0.67	1.00 (1D)	0.58	0.42	0.42 (LC,L)
0.72	0.41	0.41 (LC,L)	0.19	0.19 (2D)	0.25	0.47	0.47 (LC,L)
0.31	0.23	0.46 (Y,RS)	0.33	0.33 (2D)	0.30	0.40	0.52 (RS,GD)
Mean error							
0.58	0.41	0.52	0.42	0.53	0.42	0.44	0.54

Table 2: QDA's and CRAKE's prediction errors

QAll delivered unacceptable prediction errors, QLcL was best in the mean, QM2D was able to improve QLcL only in two cycles, namely cycles 3 and 5, and QMBD never improved QM2D. One has to mention, though, that the best number of dimensions found by QMBD was always smaller than 4, i.e. never as high as 5 and 6 as found by LMBD, and is thus structurally more stable. Q2B, though is less stable than L2B, 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 QLcL was only in cycle 2 better than LLcL. Overall, only on cycles 2 and 5 any QDA-procedure could outperform LLcL. The QM2D and QMBD 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 LLcL (42% compared with 41%) and a much better performance in predicting cycle 5 (19% compared with 50%).

Surprisingly CLcL leads to a clear improvement of the result for cycle 4 from 67% error for LLcL to 42% for CLcL (cp. Tables 1 and 2).

Concerning the mean prediction error, the CRAKE method based on all variables was best (Table 2). This method was also the overall best for cycle 2, but sharing the performance of exactly 0.375% prediction errors with QLcL. 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 C2B. And any time another pair was selected (cycles 3 and 6) the performance on the left-out cycle decreased.

Concerning the absolute prediction error of CLcL on the 4th cycle, on which all other methods have high difficulty, is lowest among all models (cp. Tables 1, 2). This confirms the impression that in LC and L one finds a stable cross-cycle information about the interplay of stylized facts and phases.

4 Conclusion

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. 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.

Literaturverzeichnis

[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, New York, 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.

[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.