RWI : Schriften

Ullrich Heilemann and Claus Weihs (Eds.)

Classification and Clustering in Business Cycle Analysis

Heft 79

RWI ESSEN

Duncker & Humblot · Berlin

RWI : Schriften
Heft 79

Schriftleitung: Prof. Dr. Christoph M. Schmidt, Ph.D.
Redaktionelle Bearbeitung: Joachim Schmidt

Ullrich Heilemann and Claus Weihs (Eds.)

Classification and Clustering in Business Cycle Analysis

DOI https://doi.org/10.3790/978-3-428-52425-9 | Generated on 2025-02-22 16:18:59

RWI : Schriften

Heft 79

Ullrich Heilemann and Claus Weihs (Eds.)

Classification and Clustering in Business Cycle Analysis

RWI ESSEN

Duncker & Humblot · Berlin

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Fotoprint: Berliner Buchdruckerei Union GmbH, Berlin
Printed in Germany

ISSN 0720-7212
ISBN 978-3-428-12425-1

Gedruckt auf alterungsbeständigem (säurefreiem) Papier
entsprechend ISO 9706 ♾

Internet: http://www.duncker-humblot.de

Preface

The analysis of cyclical macroeconomic phenomena is an important field of econometric research. In the recent past, research interests have de-emphasized quantitative forecasting exercises and have addressed the qualitative diagnosis of the relative stance of the economy regarding "upswing", "recession", or "boom" periods, i. e. the classification of the state of the economy into a limited number of discrete states. In this context the principal challenge is to reduce the multifaceted and sometimes abundant quantitative information about the business cycle to such qualitative statements in an efficient way. For more than six years this task was the focus of the project "Multivariate determination and analysis of business cycles" within the SFB 475 "Reduction of complexity in multivariate data structures", funded by the German Research Foundation (DFG).

The necessity for complexity reduction is, of course, not unique to business cycle analysis but is studied in many fields and in a number of ways. This broad interest in the reduction of problem dimensionality and in the appropriate combination of data and of theory caused the RWI Essen and the Statistical Department of the University of Dortmund in January 2002 to hold a workshop at the RWI Essen where the findings of this and similar projects were presented and discussed. The present publication collects revised versions of the papers presented at this workshop. Although the workshop took place some five years ago, these papers mark an importment juncture in the development of business cycle research.

The RWI Essen thanks the editors of the volume and the authors for their contributions and the other participants for lively discussions. Particular thanks have to go to Joachim Schmidt who supervised the production of the volume and to Anette Hermanowski and Claudia Lohkamp whose support of the production of the manuscripts and of the organisation of the workshop was indispensable.

Essen, November 2006

Rheinisch-Westfälisches Institut
für Wirtschaftsforschung

Christoph M. Schmidt

Contents

Ullrich Heilemann and Claus Weihs

Introduction

"Interest in business cycles is itself subject to a wavelike movement, waxing during and after periods of turbulence and depression, waning in periods of substantial growth" (Zarnowitz 1992: 20). After decades of rather moderate interest, in the 1980s the empirical analysis of business cycles[1] experienced, if not a renaissance, at least enhanced attention. This holds for both business economists and academia, not only for traditional fields of interest like economic indicators but also for more basic fields like measuring and structuring business cycles. The causes for the renewed interest are manifold: The huge expansion of supply of economic data, the progress that has been made in handling large amounts of data, and the improved possibilities of analysing these data with parametric and non-parametric methods with results delivering good approximations to "judgemental results" (Harding, Pagan 2006). Other reasons may be found in the subject itself; for example, some disappointment with the empirical capabilities of new theories of the business cycle such as "real business cycles" or the various disequilibrium approaches. A role might have been played by the impression that forecasting failures of macroeconometric models and of other methods were due to the negligence of cyclical aspects. Another factor might have been the "fabulous decade" of the U.S. in the 1990s, and, ironically enough, the even if only for a short period re-animated debate about "the end of the business cycle". Both did not only help re-establish the belief in macroeconomic policy but also re-directed analytical focus to the business cycle, its changes, and its endurance. As a usual consequence, the classification of U.S. cycles by the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER) experienced during this time an interest that it seldom enjoyed before, motivating the establishment of similar institutions elsewhere.

One of the strands where modern methods of statistical analysis in combination with the very much enlarged data supply seemed to be promising was

[1] Here no difference is made between "business cycle" and "growth cycle". For a detailed elaboration of the differences see the paper below by Zarnowitz (2006).

the closer study of "What happens during business cycles?" (Mitchell 1951); that is, the business cycle itself. A very fruitful starting point for such studies is located in the mid 1970s. In a series of common papers *John Meyer* and *David Weinberg* (M/W) had shown how multivariate discriminant analysis could be used to split the business cycle in a number of related stages[2] along the lines suggested by *Haberler* in the 1930s (Meyer, Weinberg 1975a, 1975b, 1976). The extended scheme distinguished not only between periods of "upswing" and "downswing", but also within these phases. With the help of multivariate discriminant analysis it was possible (1) to establish a multivariate quantitative picture of (2) a four-phase "modern" business cycle (actually, of a growth cycle) and thereby (3) examine the contribution of various variables to this picture. As M/W made clear, such a scheme would, firstly, put business cycle analysis on a more solidly founded base and thereby improve the dating of business cycles. The "multivariate approach" would also be broader than the very production oriented NBER understanding of the phenomenon. It would enhance the role played by important variables and fields ("stylized facts"), their change over time, etc. Secondly; the scheme would be useful for diagnostic and perhaps even for forecasting purposes. All of these objectives had, to some degree, been on the agenda of Mitchell (1951) and of Burns/Mitchell (1947) for a long time, but the computational possibilities were lacking.

Of course, a number of criticisms of cycle classification and of that suggested by M/W can be brought forward. The schemes by Mitchell and by Burns/Mitchell were more empirically guided and more inductively developed than deduced from theory, which is not a surprise. *Wesley Mitchell* had his roots in the *German Historical School* and in *American Institutionalism*, and hence induction. M/W can also be criticised on the operational level of their work. Part of this critique is linked to multivariate discriminance analysis and the straightforward interpretation of its results or the list of variables they used; another part of criticism is simply related to the small number of periods of recessions in the post WW II era, which means a very limited sample. Some of these arguments proved to be negligible and at least when judged by the results of a number of classification exercises for the U.S. and for Germany, the results were impressive and promising (Heilemann, Münch 1999). Other findings of M/W such as their quantitatively, multivariate based four-phase scheme were then regarded as too ambitious. The .two new phases – "Demand Pull" and "Stagflation" – were seen as overly tied to actual experiences (Zarnowitz 2006), an assessment that soon would prove to be right.

Despite the fact that M/W had widely presented their results (e. g., at the NBER and at the 1975 CIRET Conference in Stockholm), their scheme and their method had been largely ignored at the time and later. But as recent

[2] In this paper "phases" and "stages" are used synonymously.

work in the Collaborative Research Centre "Reduction of Complexity in Multivariate Data Structures" (SFB 475) by the present authors and their colleagues documents, there is ample evidence to presume that potential as well as problems of this approach well deserve closer examination. For this reason a number of scholars with different perspectives and backgrounds were asked to examine and explore their view of the territory of business cycle analysis or of the M/W approach in particular. These views were presented at a workshop held at the RWI in Essen, Germany, from January 31 to February 1, 2002. The present volume collects these papers. The presentation by H. M. Krolzig ("Markov Switching models of the German business cycle") is missing. Readers interested in this presentation should contact the author.

The workshop was opened by *Victor Zarnowitz*, not only the doyen of business cycle research but also a scholar very much in the centre of the recovery of business cycle analysis in the U.S. during the 1990s. Given the widespread confusion about the subject of economic fluctuations and the various perspectives of its analysis, his paper ("Phases and Stages of Recent U.S. Business Cycles") prepared the stage very well. The oldest, simplest, and probably still most widely shared concept of business cycles is that of a sequence of expansion and contraction in the level of overall economic activity (Burns, Mitchell 1947). Real GDP, the most comprehensive measure of the latter, was not available in the needed form at the time, and a weighted average of the principal monthly coincident indicators – broad measures of production, employment, real income, and sales – were used instead. This "classical" business cycle consists of two phases: expansion and contraction. The longer growth trend is included. This paper asks what the most common stages are of each of these phases. It presents five, the first three for expansion, the last two for contraction: (1) Recovery (from trough to the intersection with trend); (2) Rise (movement above trend to the growth cycle peak); (3) Slowdown; (4) Downturn; (5) Decline (continuation from the downward movement below trend to the trough). The scheme is empirically applied to the United States in the post-WW II period.

The second paper by *Ullrich Heilemann* and *Heinz Josef Münch* ("The U.S. Business Cycle and its Phases 1948–2000") illustrates the possibilities and the limits of a concurrent scheme proposed by M/W. The authors classify the U.S. business cycles between 1948-5 and 2000-12 into a four-phase scheme (Recession, Recovery, Demand Pull, and Stagflation). Similar to M/W they use 20 variables suggested by economic theory and general experience in business cycle analysis. They also refer to the variables employed in the NBER two-phase scheme, but also like with M/W, their general approach is more of the "growth cycle" than of the "business cycle" type. Compared with the original M/W-scheme the authors find some changes in the four-phase cycle pattern and some shifts as to the classification meaning of the variables employed. Given that the sample examined covers more than 50 years, of which 25 lie outside M/W's, this should not be surprising. More impressive is the fact that, by and large, the four-phase scheme and the influence of its constituting variables seem to be rather stable and that the changes found are to a large degree explicable.

An important question in this, as well as in any other empirical analysis of the business cycle, is its sensitivity as to the methods and variables employed for classification. *Claus Weihs* and *Ursula Garczarek* ("Stability of Multivariate Representation of Business Cycles over Time") examine this question in their analysis of the M/W approach. They start from the M/W-scheme modified by Heilemann/Münch (1999) using 13 macroeconomic variables as classifiers and test it for the period 1955-4 to 1994-4. They repeat the classification by substituting linear discriminant analysis (LDA) by quadratic discriminant analysis (QDA), as well as modern methods based on projection pursuit algorithms and a continuous dynamic Bayesian network with a certain "rake" structure. The predictive power of these methods is tested by cross-validation. Since these approaches do not usually take the time structure of business cycles into account, the authors employ a specially fitted double-leave-one-cycle-out cross-validation. In addition, they test the sensitivity of the classification power of the complete set of 13 variables by systematically reducing this number and find that the Number of wage and salary earners and of Unit labour costs are the most important variables/classifiers in all cycles and for almost all classification methods. With regard to its meaning for business cycle analysis these are important findings, backing previous findings by Heilemann/Münch (1999) and Röhl (1998).

Quite another view on sensitive questions of cycle analysis present *Marlene Amstad* and *Bernd Schips* in their paper on „Wachstumsfluktuationen, Zykluskonzepte und konjunkturelle Wendepunkte" [Fluctuations of Growth, Concepts of the Cycle, and Turning Points]. They start by characterizing business cycles as latent phenomena, which are best described by multiple times series. While some studies of this topic use single time series to describe business cycles, others employ multiple times series, sometimes aggregated. Of course, seasonal as well as trend effects have to be modelled or corrected for. The authors discuss methods of seasonal correction such as recursive filters like Census X11 and X12, and various trend correction methods. With Swiss data from 1980 to 2001 they illustrate the consequences of the various options of these methods for the emerging picture of the cycle. The adjusted series illustrate the consequences of the various methods employed. The authors show that trend "spoiled" approaches lead to models of the cycle that are valid only for a short time period – findings that for many make it easier to accept the growth cycle concept. In this context also the problem of highly correlated predictors deserves attention. It may be overcome by adding data from

"Business and Consumer surveys" to hard data. This data being more readily available and not revised have some advantages over hard data, as the authors conclude. It may, however, be noted that the inclusion of Ifo-survey data in the set of classification performance of the M/W-scheme for West Germany did not improve its explanatory accuracy and raised some doubts about their power as indicators (Heilemann, Münch 2001).

Bernd Lucke and *Malte Knüppel* approach the topic of business cycle analysis also from the data side; however, they do this in a rather unique methodical context. In their paper on "Unternehmensgrößenklassen im ifo-Konjunkturtest: eine Burns-Mitchell-Analyse" [Company Size in the ifo–Business Cycle Test: A Burns-Mitchell-Analysis] the authors examine data of the Ifo-test, grouped into various classes according to company size. The business situation is measured by rates of growth of real GDP and the cycle phases are classified with the help of a method developed by Burns/Mitchell (1947) at the NBER. The Burns-Mitchell method, until recently erroneously much neglected, is informative and delivers descriptive statistics to supplement business cycle analysis. Especially the transformation of calendar time into cycle time, reflecting the asymmetry of cycle phases, adds considerably to our understanding of the phase structure of business cycles. The paper also emphasizes the importance of disaggregated information by illustrating that the informational content of the ifo data varies considerably with the classes of company size. Most importantly for the behaviour of small and medium sized companies during the business cycle is their financial status. However, this finding still has to be embedded into a general analysis of the cyclical behaviour of companies of this type.

For many observers as well as analysts, their understanding of business cycle classification, interdependencies of phases, role of classifiers, etc. is often hampered by a host of statistical and technical prerequisites. *Katharina Morik* and *Stefan Rüping* offer in their paper "An Inductive Logic Programming Approach to the Classification of Phases in Business Cycles" a workbench for knowledge acquisition and data analysis. With the help of "Inductive Logic Programming (ILP)" they classify business cycles by model relations on the basis of time or value intervals. Note that "continuous" variables are not captured by ILP therefore its processing has to be preceded by discretisation. ILP is supported by inspecting "learned rules", not only with respect to their coverage, accuracy, and redundancy, but also regarding consistency (i.e., logical contradictions). The paper describes the workbench MOBAL, its learning algorithm RDT, the pre-processing of data, and the first (and encouraging) results for the German business cycle. The multivariate nature of ILP and the automatic selection of most relevant variables is one way to approach the business cycle problem. Restricting the search for classification rules to rules with two variables each, although some variables were more important prominent than others in the rules, all the 13 variables of the classification by M/W or by Heilemann/Münch (1999) were of significant influence in the classification process. Not surprisingly, the explanatory accuracy of a two-phase scheme ("upswing", "downswing") was clearly higher than that of the four-phase scheme.

Ursula Garczarek and *Claus Weihs take* up this possibility of further reducing the number of classifiers ("Univariate Characterization of the German Business Cycle 1955–1994"). In order to find simple univariate characterizations of business cycle phases, they use the empirical framework of their previous paper (Weihs, Garczarek 2006) and of Heilemann/Münch (1999). By com-paring the graphs of the stylized facts with the track of the business cycle they drop five of the 13 variables as univariate indicators. Furthermore, after an inspection of the correlation matrix and the corresponding scatterplots, they conclude that the relevant factor dimension is even lower than eight. With the help of parallel box plots, 18 rules are identified for the characterization of the four cycle phases of which only 11 are used by a higher order rule discriminating between several applicable rules. The mean prediction error rate, by applying the double-leave-one-cycle-out cross-validation, is 54%, which lies between that of quadratic discriminant analysis (58%) and LDA (47%).

A proper understanding of business cycle analysis and of classification in particular, requires being aware of the reduction of complexity achieved by discriminant (or cluster) analysis. There is always a trade-off between "too much" and "not enough" reduction with both putting the diagnostic (or prognostic) usefulness of classification at risk. *Claudia Becker* and *Winfried Theis* ("Combining Dimension Reduction and Fuzzy-clustering: An Application to Business Cycles") explore possibilities, limits, and their risks. They combine dimension reduction and fuzzy clustering to mimic an expert's classification of business cycle phases. Starting from the 13 selected variables of the M/W approach as used by Heilemann/Münch (1999) and Weihs/Garczarek (2006) "expert's classification" is used to develop a "supervised learning" type of projection by "Sliced Inverse Regression" (SIR) into a lower dimensional space. In a second step, ignoring the information of the expert's classification, fuzzy-clustering applied to the dimension reduced data is compared to the clusters of the best clustering set of three original variables. In both cases only two classes could be distinguished. While the separation of the clusters based on SIR is not as good as when using an optimal subset of variables, clusters based on SIR much better reflect the expert's classification. The two main phases can – again! – be interpreted as "Upswing" and "Downswing", whereas areas of uncertainty (between the clusters) can be interpreted as in-between phases ("Upper Turning Point" and "Lower Turning Point").

"Self-Organizing Maps" (SOMs) are a type of artificial neural network that implement unsupervised learning (clustering). The paper by *Gabriela*

Guimarães ("Self-organizing Maps for Time Series Analysis") discusses ways to apply SOMs to time series data like those constituting business cycles, since traditional SOMs do not take time into account. The paper starts with a short introduction to SOMs and its standard generating algorithm. To incorporate time into SOMs several approaches exist, with each of having particular properties. The remainder of the paper concentrates on three types of modified SOMs. In the first type, time is introduced during pre- and post-processing, while the basic SOM algorithm is left unchanged. The second type modifies the learning rule (estimation procedure) in order to take temporal dependencies into account; the third type varies the topology of the SOM. This is done by introducing SOMs with feedback and by using SOMs with a previously established hierarchical structure. It is shown that the usefulness of each of these types depends on the particular application, and none is generally superior. Best results are obtained by using a combination of these methods. All in all, it seems promising to apply SOMs in the field of business cycle analysis where classification procedures often ignore the consecutive character of cycle phases and cycles themselves.

In the concluding paper, *Victor Zarnowitz* complements his introducing paper by looking for changes in the business cycle and their explanations ("Modern Trends and Their Effects on International Business Cycles"). International integration of markets for products, labour, and capital (globalization) represents the most widely observed, debated, pervasive, and important economic and financial developments in the world today. Zarnowitz discusses briefly the main advantages of globalization to the economies that open up to it, developed and developing, rich and poor, labour and capital intensive, but points also at its own difficult problems, particularly when combined with concurrent technological progress and institutional changes. The advent of the new information age marks another modern trend of extraordinary importance. A question of particular interest is the role of these new trends in shaping international business cycles. Will the greater speed of flows across the borders accelerate the transmission, and hence increase the synchronization of business cycles as well? The question is pertinent and undoubtedly important and though early raised still open. The paper stresses that other factors have major roles in this context as well, notably the effects of the prevailing monetary regime. Zarnowitz discusses how business cycles resembled each other and differed in three eras: (1) Under the gold standard before WW I, business cycles were well synchronized, mostly substantial, and at times of serious consequences; but economic activity tended to prosper and inflation was contained. (2) The interwar period, 1919–39, was prone to deflation and depression, least successful and stable. (3) The post-WW II era avoided the turbulences of the interwar period. It did so by establishing a gold-standard rule until 1971. This system of more or less fixed exchange rates proved efficient for a rather long time. Later, the economies switched to a system of flexible exchange rates (on other secular changes see, for example, Gordon 1986).

Even though all papers presented at the workshop had a common subject, they are rather difficult to summarize. Without doing too much harm to each of them, at least six general results seem to be notable:

– First, cyclical classification in general and that of M/W in particular appear to deliver or impose a rather stable pattern on post WW II cycles. This seems to be valid with the four-phase scheme structuring the cycle, as with respect to the choice of variables constituting the business cycle. No doubt, the original M/W-scheme underwent considerable changes as to the set of classifying variables or as to the label of some phases ("Demand-Pull", "Stagflation"). The multivariate or multi-dimensional character of the business cycles that M/W had been emphasising was obviously not much affected by these changes. But these changes are probably not much greater than the other explanations or descriptions the business cycle had to submit.[3] We still are not very far beyond the platitude that "all cycles look alike but some more" and all generalisations about business cycles or any kind of "reference cycle" can be found only at the considerable loss of correspondence in the single case.

– Second, the role played by the various classifiers changes over time and the patterns of these changes seem to be unclear. Sensitivity studies showed that usually a large share of total explanation comes from less than a handful of variables. However, ignoring the rest of the variables would mean a sacrifice of complexity (admittedly a very subjective criterion) for simplicity. However, caution is always advised with interpreting statistical ascriptions of influence to the variables used as representing economic importance. Given the considerable changes the German or the U.S. economy experienced during the last 50 years, in the present case, shifts of weight or influence of the classifiers appear as surprisingly small.

– Third, once again, LDA delivered results that were hardly outperformed by more sophisticated procedures.

– Fourth, a number of tests of the scheme applying very different methods and approaches to the problem of classification did not contradict the findings of multivariate discriminant analysis. Of course, most of all they demonstrated the usefulness of these approaches, but they also added to the understanding and interpretation of the classification scheme and its working.

[3] Most of the studies pointing at changes in economic fluctuations deal only with changes of their widths or depths and do this usually on a highly aggregated level. They rarely try to relate these changes to the many institutional, structural or behavioural changes the authors are well aware of.

– Fifth, the classification scheme examined here seems to relax the old debate between the business cycle-perspective and the growth cycle-perspective (Zarnowitz 1992: 203ff.) – at least on the level discussed here.
– Finally the most important conclusion even if not all too new: all contributions demonstrated in one way or another that the phenomenological approach of Burns/Mitchell can be reconciled to some degree with the analytical approach. "Reduction of complexity in multivariate data structures" has its price but thanks to modern methods of data analysis it may be lower than Burns/Mitchell and their followers might have feared.

As with all empirical studies, new data for different samples, countries and levels of aggregation will put the scheme discussed here to test. Multivariate analyses' hunger for data is well known and the post WW II era with its 10 cycles and less than 100 months of recession in the U.S. and less than 30 in Germany leave only limited room for estimation, testing, or "learning" – for "business cycles" even less than for "growth cycles". It seems more important to examine in depth the many shifts or changes to be observed. The scheme and the methods presented at this workshop offer a broad variety of promising approaches and results. They may touch or transcend traditional lines of quantitative business cycle analysis. But it should become clear that their potentials in the field are well worth further exploration.

References

Burns, A.F. and W.C. Mitchell (1946), Measuring business cycles. NBER Studies in Business Cycles 2. NBER, New York.

Gordon, R.J. (1986), Introduction: continuity and change in theory, behaviour, and methodology. In R.J. Gordon (ed.), The American business cycle – Continuity and change. NBER-Studies in Business Cycles 25. Chicago: University of Chicago Press, 1–33.

Harding, D. and A. Pagan (2006), Measurement of Business Cycles. University of Melbourne, Department of Economics Working Paper 966.) University of Melbourne.

Heilemann, U. and H.J. Münch (1999), Classification of West German Business Cycles 1955–1994. *Jahrbücher für Nationalökonomie und Statistik* 219: 632–656.

Heilemann, U. and H.J. Münch (2001), Ifo Data and the German Business Cycle – Results of Multivariate Discriminant Analysis 1961–1997. *ifo-Studien* 47: 41–63.

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

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

Meyer, J.R. and D.H. Weinberg (1976), On the classification of economic fluctuations: an update. *Explorations in Economic Research* 3: 584–85.

Mitchell, W.C. (1951), What happens during business cycles – A progress report. NBER, Cambridge, MA.

Röhl, M.C. (1998), *Computerintensive Dimensionsreduktion in der Klassifikation*. Lohmar: Josef Eul.

Weihs, C. and U. Garczarek (2006), Stability of Multivariate Representation of Business Cycles over Time. In this volume.

Zarnowitz, V. (1992), Business cycles – theory, history, indicators, and forecasting. NBER Studies in Business Cycle Research 27. Chicago and London: University of Chicago Press.

Zarnowitz, V. (2006), Phases and Stages of Recent U.S. Business Cycles. In this volume.

Victor Zarnowitz

Phases and Stages of Recent U.S. Business Cycles

1. Introduction

Business cycles are sequential developments with many diverse features but one common characteristic: they consist of recurrent expansions and contractions in the level of total economic activity. The latter concept is not one-dimensional but embraces many interrelated economic variables. It is best thought of as co-movement: tendency to grow and fluctuate together of production, employment, real income, sales, prices, monetary aggregates etc. in many sectors and regions of the economy (e.g. Zarnowitz 1985).

These movements are imperfectly correlated as the variables differ in cyclical sensitivity and timing, in part systematically, but they affect the economy as a whole in most of its aspects. At least after the fact, with most of the important comprehensive data in, there is usually a prevailing agreement that a boom or a recession had indeed occurred.

Yet there is little that is clearly repetitive, periodic, and predictable about business cycles, and they are indeed poorly predicted. The reason is that the economy's structure and dynamics change and interact, and that outside events and policies affect the outcomes. Some of the change is gradual, some is fast. There are policy reactions as well as initiatives; both have various constraints – economic, financial, and political – but both can benefit from learning. So can private decision-making.

Economic and other variables undergo shifts due to unexpected shocks, which influence business cycles in ways that are elusive. Still, important self-sustaining or endogenous elements undeniably exist in business cycles and it is mainly their study that promises progress in analysis and forecasting. Some prominent economists have worked in this difficult area with impressive results. They succeeded in describing what happens during expansions and contractions, and how one phase gives way to the other. These advances led to the acquisition of considerable tested knowledge of what caused specific fluctuations of the past.

As defined and observed, the business cycle has a standard two-phase breakdown: a trough-to-peak expansion and a peak-to-trough contraction. Further subdivisions or classifications, however, are more complicated. Of the diverse suggested taxonomies, none has been generally accepted. However, much of the work has been revealing, and the resulting terms such as "recession," "depression," "revival," "boom," etc., are much in use.

In this paper, I revisit the question of whether a multiple-stage conception of the business cycle is meaningful and applicable. The starting point is the distinction between business cycles and growth cycles, based on the classical decomposition of time series into trend, cyclical, seasonal, and irregular variations. The "contraction" (better known as recession, or when particularly severe, depression) suspends, without suppressing, the generally prevailing tendency of the market economy to grow. A recession ends at some point below the economy's longer-term upward trend. The initial stage of an expansion is a rebound upward back to that trend. This is sometimes treated as a separate phase of "recovery." It was frequently during the recoveries that the highest rates of growth in total economic activity were achieved. As the expansion proceeded, growth measured from a rising base tended to decline.

As the economy regains and then exceeds its pre-recession peak, the recovery gives way to an above-trend growth phase. When sufficiently pronounced, this part of the expansion is informally called a "boom." The levels of general economic activity are now high, and it may be increasingly difficult to maintain high growth of output and employment without generating strong upward pressures on prices and wages. The last stage of expansion is often marked by a slowdown in activity that eventually deepens into a recession.

This, I believe, is the simplest breakdown of the expansion phase into stages, but how applicable is it empirically, and what, if any, are its theoretical implications? Do recessions have meaningful stages and if so, what are they? How do the results, based on post-World War II U.S. economic history, vary over time and with the characteristics of the cycles covered? What conclusions can be drawn?

2. Growth Cycles and Business Cycles

Some substantial slowdowns merely interrupt economic expansions, i.e., they are followed by a re-acceleration of growth rather than by a collapse into a period of negative growth. They have been treated like recessions in countries that experienced long periods of high economic growth, e.g., Germany and Japan in the first 20–25 years after World War II, and some developing

Table 1

US Growth Cycles and Business Cycles
1948 to 2000; durations of cycles and their phases

Growth cycles			Durations in months of growth			Business cycles			Durations of months of business cycles and phases		
Peaks (P) and Troughs (T)			Cycles and Phases			Peaks (P) and Troughs (T)					
P	T	P	P to T	T to P	P to P	P	T	P	P to T	T to P	P to P
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Jan–48	Oct–49	Jan–51	21	15	36	Nov–48	Oct–49	Jul–53	11	45	56
Jan–51	Jul–52	Mar–53	18	8	26						
Mar–53	Aug–54	Feb–57	17	30	47	Jul–53	May–54	Aug–57	10	39	49
Feb–59	Apr–58	Jan–60	14	21	35	Aug–57	Apr–58	Apr–60	8	24	32
Jan–60	Feb–61	Apr–62	13	14	27	Apr–60	Feb–61	Dec–69	10	106	116
Apr–62	Jan–64	Mar–66	21	26	47						
Mar–66	Oct–67	Aug–69	19	22	41						
Aug–69	Nov–70	Nov–73	15	36	51	Dec–69	Nov–70	Nov–73	11	36	47
Nov–73	Apr–75	Mar–79	17	47	64	Nov–73	Mar–75	Jan–80	16	58	74
Mar–79	Jul–80	Jul–81	16	12	28	Jan–80	Jul–80	Jul–81	6	12	18
Jul–81	Dec–82	Sep–84	17	21	38	Jul–81	Nov–82	Jul–90	16	92	108
Sep–84	Jan–87	Jan–89	28	24	52						
Jan–89	Dec–91	Jan–95	35	37	72	Jul–90	Mar–91	Mar–01	8	120	128
Jan–95	Jan–96	Jun–00	12	53	65						
Mean			18.8	26.1	44.9				10.7	59.1	69.8
Median			17.0	23.0	44.0				10.0	45.0	56.0
Standard Deviation			6.2	13.2	14.7				3.4	38.0	39.1

National Bureau of Economic Research

economies. Chronologies of "growth cycles," that is, fluctuations in detrended measures of total economic activity, capture these slowdowns and speedups.

Table 1, columns 1–6, shows the timing and duration of phases for the fourteen complete peak-to-peak U.S. growth cycles since 1948 determined according to the deviations from trend of the Coincident Index (CI) of The Conference Board. CI combines four monthly indicators of current economic conditions: total nonfarm employment; real personal income less transfer payments; real manufacturing, retail and wholesale trade series; and the index of industrial production (listed in descending order in terms of comprehensiveness of coverage). The trend is based on a centered 75-month moving average and its extrapolations at the beginning and end of the series; it is made to connect the midpoints of smoothed phases in the deviations of the series from its moving average. As such it is flexible and nonlinear, known as "phase average trend" (PAT) and long used internationally in identifying and dating growth cycles (notably by OECD). PAT goes back to the work of the National Bureau of Economic Research (NBER) in the 1970s (Boschan, Ebanks 1978).

Table 1 also includes a parallel listing of the turning dates and durations for the nine complete peak-to-peak U.S. business cycles since 1948 (columns 7-12). These dates having been determined by the NBER, there is a high degree of consistency between the growth cycle and the business cycle chronologies used here. (NBER based its dating largely on the four monthly indicators that are the components of CI).

Not every slowdown deepens into a recession, but every recession implies a slowdown *a fortiori*. Hence there are more slowdowns than recessions and, by the same token, more growth cycles than business cycles covered (e.g. 14 vs. 9 in the period 1948–2001). Clearly, growth cycles tend to be not only shorter but also far less asymmetrical than business cycles. While post-WWII business cycle expansions were on the average about 5 times longer than contractions, growth-cycle rises were only 1.4 times longer than declines.

Table 1 shows what is behind this fact. The business expansions of 1949–53, 1982–90 and 1991–2000, were each interrupted by one significant slowdown, while that of 1961–69 was interrupted by two. These slowdowns gave rise to additional declines in the detrended aggregates when compared with the original declines in the levels of the same series. Also, on all but two occasions, growth cycle peaks preceded business cycle peaks (cf. cols. 1 and 7), that is, slowdowns occurred before recessions. On the other hand, growth cycle troughs mostly coincided with, or lagged slightly, business cycle troughs (cf. cols. 2 and 8). This is because most of the postwar recoveries were quick or "V-shaped." The one notable exception was the sluggish recovery of 1991, where the growth cycle trough (December) lagged significantly behind the business cycle trough (March).

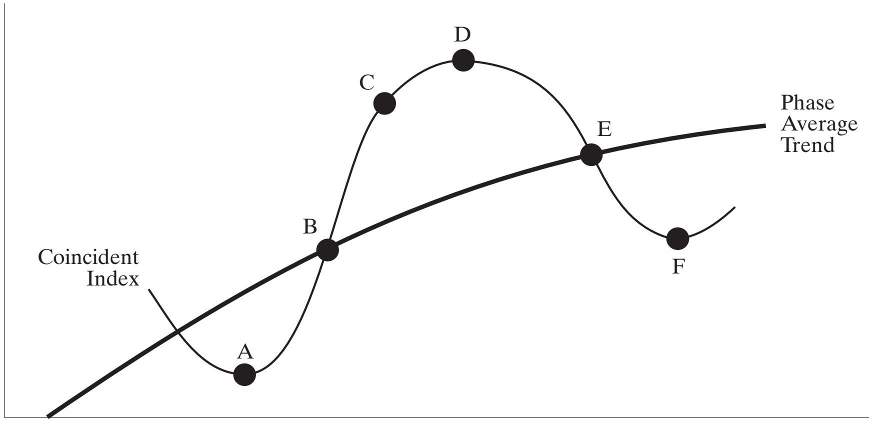
3. Stages of Expansion and Contraction

Business expansions vary greatly in duration, amplitude, and persistence; business contractions vary less. The shortest of the postwar expansions lasted one year, the longest ten years; the range of the recessions was from six to sixteen months (Table 1). Four of the ten expansions were interrupted by slowdowns, while all of the contractions were essentially continuous. Given this diversity, can the cycle phases be usefully subdivided into some further stages with similar characteristics?

The answer is a qualified yes. Many past classifications proved to be only temporarily valid, probably because they were overly ambitious.[1] Results that have more general validity across time and space can be obtained by using the trend-cycle decomposition applied to real time series.

[1] For example, the effort to derive a chronology of business-cycle stages by cross-classifying the cyclical movements of quantity and price variables suffers from the fact that "demand-pull" and "stagflation" phases exist only at some times and not at other times (Meyer, Weinberg 1975), even in the generally inflationary period of 1947–73, let alone in longer periods that include deflationary trends (see on this also the contribution of Heilemann, Münch 2006).

Figure 1
Stages of Expansion and Contractions



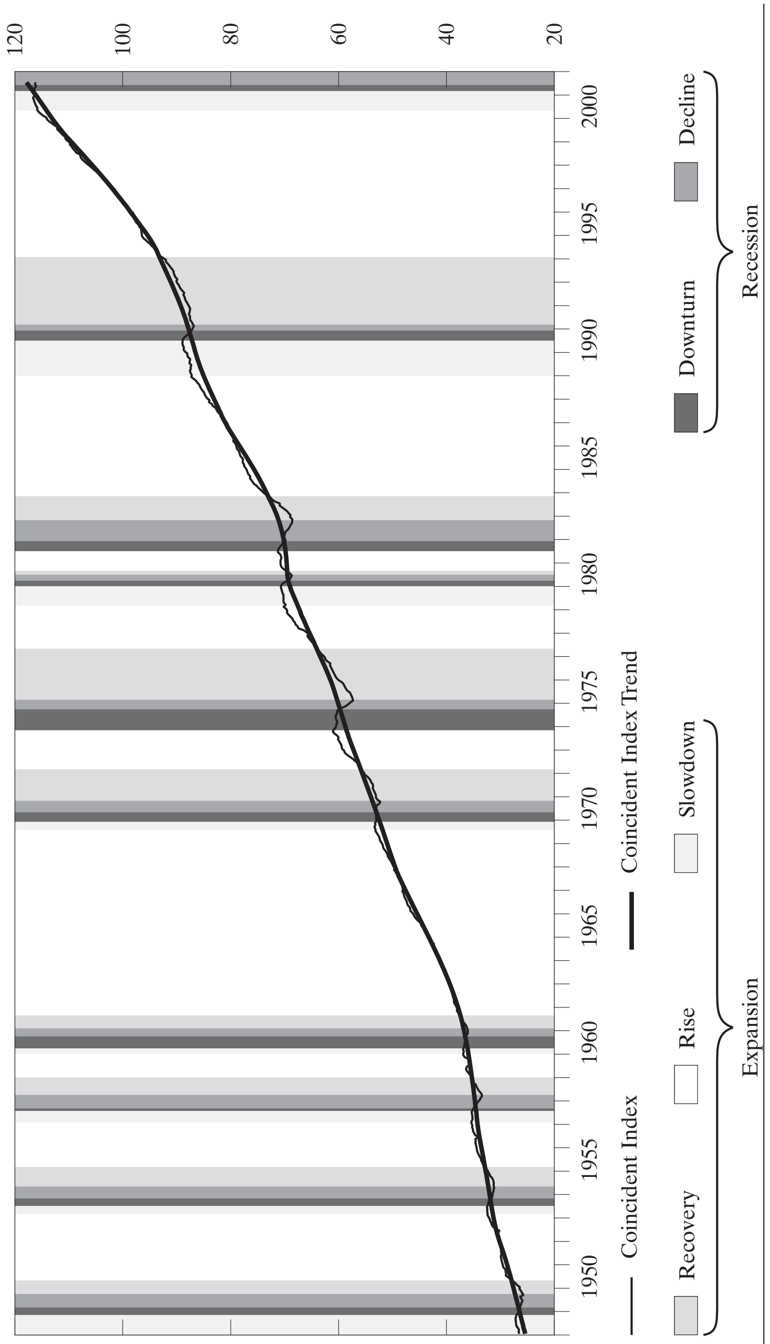
Stages:
A: Businnes Cycle Through
B: Rising Through Trend
C: Growth Cycle Peak
D: Business Cycle Peak
E: Falling Through Trend
F: Business Cycle Through

Phases:
AB: Recovery
AD: Expansion
BC: Rise
CD: Slowdown
DE: Downturn
DF: Contraction (Recession)
EF: Decline

One must begin with a simplified schematic representation of the underlying ideas. Consider a comprehensive time series or index measuring economic activity or its diverse interrelated aspects (e.g., GDP or CI). Take the prototypical cycle in economic activity to be a smoothed sinusoidal movement around its long growth trend. As depicted in Figure 1, the trend goes approximately through the midpoints of the cyclical declines in the so measured aggregate economic activity (like PAT does empirically for CI). Thus the series rises from trough to peak and falls from peak to trough, crossing the trend from below and from above, respectively; accordingly its troughs (peaks) lie below (above) the trend. Now divide this movement into recovery (rebound from trough to trend), boom or rise (rising above trend to the largest distance from trend, the growth cycle peak), slowdown (declining positive growth), downturn (negative growth bringing CI down to the trend) and decline (continuation of the downward movement below trend to the new trough. The first three of the above steps constitute the expansion; the last two, the contraction or recession.

It is apparent that the scheme incorporates one basic assumption, namely that the economy is growing in each business cycle. Thus each peak is higher than the peak in the preceding cycle. This is normally the case, and exceptions to

Figure 2
U.S. Business Cycles – Stages of Expansion and Contraction with Growth Cycle Slowdowns
1949 to 2001; 1996 = 100



this rule are very rare. One can conceive of a very depressed period during which growth ceased and one or even more cycles occurred with expansions consisting only of (perhaps incomplete) recoveries. Indeed, examples of such stagnant periods exists in the U.S. and foreign economic histories. But they point to anomalous conditions that call for major reforms and/or end in major cataclysms: economic or financial crises; institutional or policy failures, and combinations thereof.[2] Such conditions were absent in the postwar United States economy (The short expansion of July 1980 – July 1981 ended just slightly higher than it began (Figure 2). But this turbulent period of high inflation and interest rates was short and initiating an era of lower and more stable prices and higher and more stable growth.)

Another assumption underlying Figure 1 is that the last stage of expansion is a slowdown during which the gains continue but diminish. This has been so in most but not all recent U.S. business cycles.

Of the stages of expansion, the first and third have the familiar and self-explanatory names of recovery and slowdown, respectively, but the second, which denotes mid-expansion growth rising above the trend, has not. If growth is strong and maintained or rising, this stage may properly be called a "boom." But this popular term is not always appropriate because in some cycles growth in this stage is only moderate (often less than in the recovery stage), and may also be discontinuous. "Rise" is preferred as more neutral and more accurate. It is in this stage that economic activity rises above the highest past levels and begins to register net gains.

The literature offers no good names for the two recession stages. Schumpeter (1939) called the above-trend part a "recession," and the below-trend decline a "depression;" the former movement was toward a new equilibrium, the latter away from it. But the terms acquired a different meaning, depression denoting a very severe, recession a milder, more common contraction. Therefore, to note the importance and "nearness" of the adverse directional change, let us call the first part of contraction a "downturn." And, to match the "rise" above the trend, let us call the second part a "decline."

These are again simple descriptive labels. Some classifications use economic processes associated with recessions or depressions and refer to stages of "liquidation" (of excess inventories and debts) or "absorption" (of excess

---

2 In the *Great Depression*, the upward trend of the U.S. economy was temporarily suspended, and the 1937 peak failed to exceed the 1929 peak by the best available measures of overall economic activity. Another contraction, shorter but deep and starting from already depressed levels, came in 1937–38. The definite end to the long economic disaster, which spread globally, arrived only with World War II. – In *Japan*, the 1990s brought a sharp decline in the growth trend, from high to flat, and a related sharp increase in cyclicality. There were two complete recessions and a third one under way, and the economy was stagnant and plagued by deflation.

capacities). These processes are real and important enough but their temporal allocation within the cycle presents difficulties: they may start in slowdowns and extend beyond contractions. Historically important, too, but more episodic, were various other processes or events such as tight money policies, financial crises and crunches, oil price shocks, and stock market bubbles – but clearly none of these can be helpful here. Neither can the older emphasis on deflation, although recent developments in Japan remind us of deflation's continued depressant power.

4. Applicability to U.S. Postwar Expansions

Table 2 shows that it is possible to fit the above scheme to actual data representing the evolution of U. S. economy in the past half-century, but that this requires sufficient flexibility in recognition of the diversity of the individual business cycles and growth cycles covered. No single multi-stage classification can fit the variety of cyclical experience without modifications.

Table 2

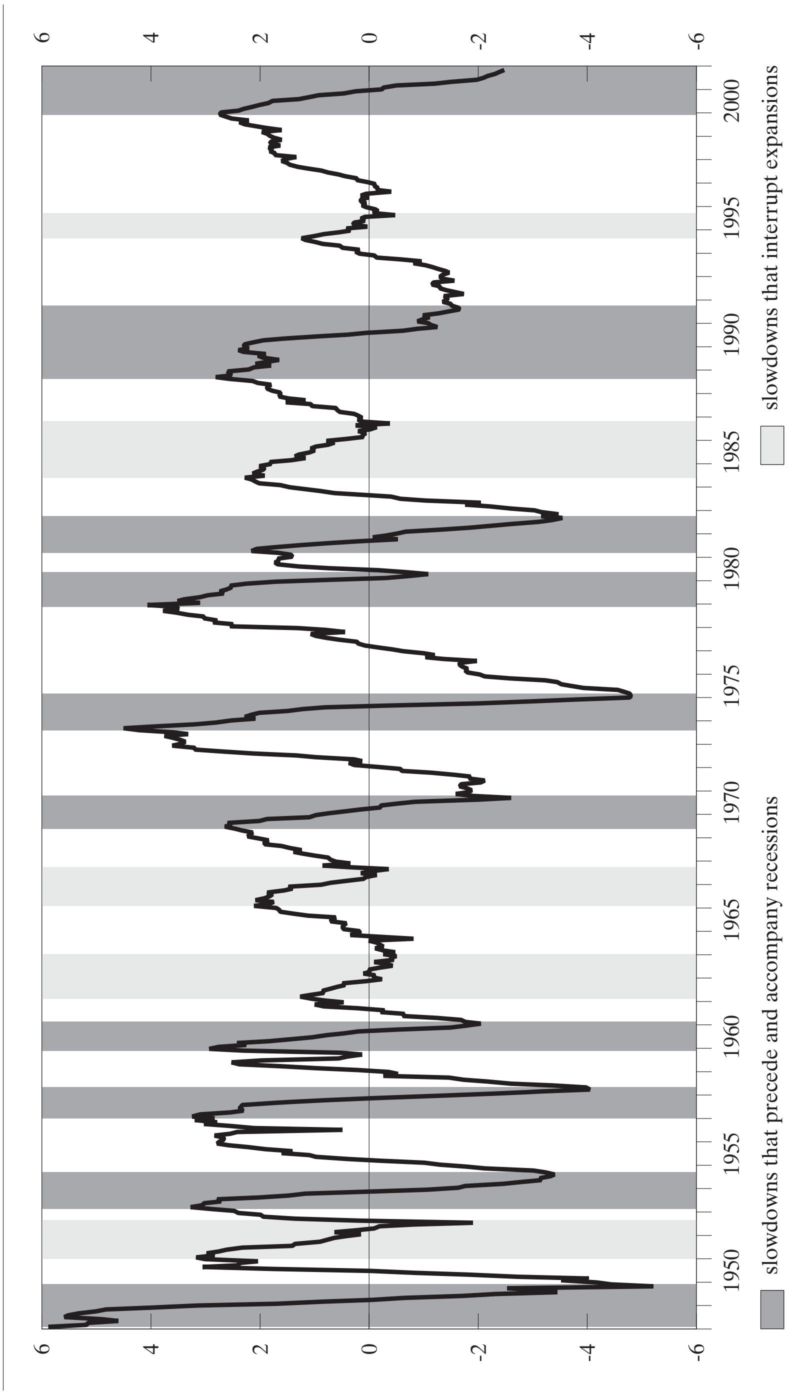
Stages of Expansion and Contraction in U.S. Economy
1948 to 2001

Line	Expansion (from BCT to BCP)			Recession (from BCP to BCT)	
	Recovery	Rise	Slowdown	Downturn	Decline
	(from BCT to PAT)	(sustained growth above PAT)	(from GCP to BCP)	(from BCP to PAT)	(from PAT to BCT)
	(1)	(2)	(3)	(4)	(5)
1			1/48–11/48 (10)	11/48–3/49 (4)	3/49–10/49 (7)
2	10/49–5/50 (7)	5/50–3/53 (34)	3/53–7/53 (4)	7/53–11/53 (4)	11/53–5/54 (6)
3	5/54–3/55 (10)	3/55–2/57 (23)	2/57–8/57 (6)	8/57–10/57 (3)	10/57–4/58 (5)
4	4/58–1/59 (9)	1/59–1/60 (12)	1/60–4/60 (3)	4/60–10/60 (6)	10/60–2/61 (4)
5	2/61–9/61 (7)	9/61–8/69 (95)	8/69–12/69 (4)	12/69–5/70 (5)	5/70–11/70 (6)
6	11/70–3/72 (16)	3/72–11/73 (18)	none	11/73–10–74 (11)	10/74–3/75 (5)
7	3/75–5/77 (25)	5/77–3/79 (22)	3/79–1/80 (10)	1/80–4/80 (3)	4/80–7/80 (3)
8	7/80–9/80 (2)	9/80–7/81 (10)	none	7/81–12/81 (5)	12/81–11/82 (11)
9	11/82–11/83 (12)	11/83–1/89 (62)	1/89–7/90 (18)	7/90–12/90 (5)	12/90–3/91 (3)
10	3/91–2/94 (35)	2/94–5/00 (75)	5/00–3/01 (10)	3/01–6/01 (3)[a]	6/01–11/01 (5)[a]

		Duration of stages, in month				
		Recovery	Rise	Slowdown	Downturn	Decline
11	Mean	13.8	39.0	8.1	4.9	5.5
12	Median	10	23	8	4.5	5
13	Standard deviation	10.5	30.7	5.0	2.4	2.3
14	Range	2–35	10–75	3–18	3–11	3–11
15	Average percent of phase	22.7	64.0	13.3	47.1	52.9

Source: National Bureau of Economic Research. – BCT: business cycle trough; BCP: business cycle peak; PAT: phase average trend; GCP: growth cycle peak. Numbers in parenthesis represent stage durations in months. – [a]Tentative. Based on the PAT updated through April 2002 and on the assumption that the recession ended in November 2001.

Figure 3
U.S. Current Condition Index (U.S. Coincident Index)
1949 to 2001; Deviations from Phase Average Trend in %



But the deviations from the suggested pattern can themselves by readily identified, dated and explained, as it is done in Table 2.

The TCB composite index of coincident indicators (CI), a monthly, seasonally adjusted series, 1948–2001, provides a broad measure of current economic conditions that serves as our database. CI is plotted in Figure 1, along with its trend (PAT). The proposed breakdown by stages is also based on the NBER dates of business cycle peaks and troughs, and on the growth cycle turning points, which refer to the deviations of CI from PAT. The two chronologies have a high degree of consistency: one relates to levels and the other to detrended values, but both have common sources, data, and methodology.

As shown in Table 2 (column 3) and Figure 2, all but two of the ten post-WW II expansions in the U.S. ended in slowing growth that turned negative. The three-year expansion of the early seventies ended abruptly in November 1973, a month after the Organization of Petroleum Exporting Countries (OPEC) imposed an oil embargo on trade partners of Israel, starting a sharp climb in oil prices. The extremely short (one-year) expansion of 1980–81 also ended abruptly, which may be attributed to the uniquely high and rising real interest rates. But such sharp downturns were exceptions in the past half-century and due to special factors. (In contrast, crises were more frequent in earlier times, and some theorists perceived sharp peaks to be common.[3])

Table 2 lists only the slowdowns that represent the end stages of expansions. But, as discussed above, growth-cycle contractions (slowdowns) occurred also in 1951–52, 1962–63, 1966–67, 1984–86, and 1995, in each case interrupting a long rise. Figure 3 shows these rises and interruptions in light grey.

Figure 3 shows the deviations from the trend (PAT) of the TCB current conditions index (U.S. coincident index). Whereas the slowdowns that precede and accompany recessions show large and long declines below PAT, the slowdowns that merely interrupt expansions do not: they barely dip briefly below PAT. This is an important distinction analytically, which may prove to be some value in practical application.

Recoveries, rises, and final slowdowns averaged about 14, 39, and 8 months, respectively. The corresponding standard deviations are large, matching the great dispersion of business cycle expansions (Table 1). On the average,

[3] Keynes (1936: 314) wrote of "the phenomenon of the *crisis* – the fact that the substitution of a downward for an upward tendency often takes place suddenly and violently, whereas there is, as a rule, no such sharp turning-point when an upward is substituted for a downward tendency." But abrupt declines occur mostly in crises that, never very frequent, have become more far and between than ever. In the post-WW II era, the downturns seem to have become more gradual or rounded, the upturns more abrupt or sharper.

recoveries account for about 22 percent, rises for 65 percent, and slowdowns for about 13 percent of the expansion durations (Table 2 for detail).

Figure 2 confirms that the 1990s witnessed the longest recovery as well as the longest rise of all postwar expansions. The 1960s and the 1980s had quick and short recoveries and also long but more uneven rises with longer intervening slowdowns. The developments in the 1950s, 1970s, and early 1980s show greater cyclicality: as many as six short sequences of expansion and contraction, with recoveries more pronounced relative to rises and slowdowns mostly short.

5. Applicability to U.S. Postwar Contractions

The nine complete business cycle contractions in 1949–91 were generally short, and the recession of 2001 looks now short, too. One may well question, then, the need to divide these episodes any further. However, each recession, presumably brings the economy down below its local growth trend for some time: The PAT, like most other trend estimates that are useful for growth cycle analysis, meets this condition (Zarnowitz, Ozyildirim 2002). Hence, experimentally, the two recession stages are distinguished here.

The downturns ranged from 3 to 11 months in duration, and so did the declines below the trend. On average, the two stages of recession had about equal length, a little over 5 months. But the recessions of 1948–49, 1953–54, 1957–58 and, especially, 1981–82 had relatively short downturns and long declines, whereas the reverse was true of the recession of 1960–61, 1973–75, and 1990–91. The recessions of 1969-70 and 1980 were about equally divided (Table 2).

As shown in Figure 2 and Table 2, the first four post-World War II business cycles (1948–61) had steep downturns and declines followed by similarly fast recoveries. The recessions lasted 8–11 months, the recoveries 7–10 months. Later the downturns and declines grew more diverse, the recoveries considerably longer. The recession of 1969–70 was average in duration, shallow, and about evenly divided between downturn and decline; it was succeeded by a mostly slow recovery about twice as long (16 months) as the average of the four earlier ones. The recessions of 1973–75 and 1981–82 were the most severe and the longest of the set (16 months each) but the former had a shorter, steeper, and deeper decline than the latter. The recovery back to the PAT lasted 25 months in 1975–77 and only 12 months in 1982–83. Finally, the recession of 1980 and 1990-91 both involved only short and comparatively small declines in economic activity as measured by percent changes in CI or real GDP, but the former was followed directly by a short and small rise, the latter by a long (35 month) recovery along the trend (Figure 2).

6. An Extension to the Last Recession

In practice, the most important question arising from empirical business cycle research is almost invariably this: What does the past tell us about the current and future economic conditions? As shown above, the U.S. economy passed in 1991-2000 through a recovery-rise-slowdown sequence much like the previous long expansions of the 1960's and 1980's. Indeed, in all three expansions, the rise can very properly be called a boom, and in the 1990's the boom was the longest and the interrupting growth-cycle slowdowns were the shortest (Figure 2).

This was written in May 2000, when there were rising signs that the 2001 recession had come to an end. According to the NBER, the downturn started in March 2001; CI started declining in December 2000, but this reflected mainly the earlier slump in manufacturing. From June to November 2001, CI declined below the trend; thereafter, it rose in each of the following five months through April 2002. Real GDP declined only in Q3 2001, but real net domestic product (NDP) declined in Q1, Q2, and Q3; both rose in the last quarter of 2001 and in the first quarter of 2002.

But the most recent and future data are still uncertain and subject to revisions. From the historical perspective and as a matter of reasonable caution, it would have been be premature to date the end of recession and beginning of recovery. The NBER Dating Committee[4] was therefore waiting for more information. However, I shall tentatively assume here that the recession ended in November 2001. This is done only for the sake of an experimental answer to a hypothetical question: If the recession did end in November, and no substantial revisions of the data occurred, how would this cycle compare with its predecessors?

The experiment required an updating of the PAT; previously, we worked with a trend for 1948–2000, now the trend had to be extended through April 2002. The computerized updating might have changed the trend sufficiently to alter the dates in Table 2 that are based in part on the turning points in the deviations of CI from its PAT. But actually the new trend turned out to be identical with the old one, except only for some slight tilting down of the most recent values (after 1997). Moreover, the new PAT estimate looks very reasonable, running below CI in the boom years 1997–2000 and crossing to above CI in 2001 (Figure 2). Still, the reader should appreciate that the main difficulty with all trend estimates is that they are weakest at the end of the period. This is so because here the slope of the trend must be estimated beyond the reach of the available data (Zarnowitz 1992: 204ff.).

---

4 The author is a member of the NBER Business Cycle Dating Committee.

For all these cautions, it is safe to state that, (assuming it is over) the recession was among the mildest on record in terms of GDP and CI as a whole. Of the components of this index, industrial production had an early and substantial contraction; real business (manufacturing and trade) sales also an early but a shallow one; real personal income merely flattened, and briefly; but employment still shows a gradual decline. In terms of corporate profits, stock prices, and output of high-tech and some other manufacturing industries, the latest recession was anything but short and mild.

As I review this paper at present (July 2006), the date still evidence that the recession that ended in November 2001 is still happily the last one recorded.

7. Conclusion

As shown in this paper, there is a strong suggestion from the data that recessions with well articulated declines below the growth trend are followed by sharp and quick rebounds. These recoveries usually raise the level of economic activity above the trend promptly.[5] In contrast, shallow recessions with small and short declines are succeeded by slower and longer recoveries; here activity may for some time be rising parallel to the trend but below it, as in 1971 and 1991-93. There are always forces of growth at work in the economy, and the rises and declines away from the upward trend are effectively countered by the equilibrating slowdowns and recoveries toward the trend.

Another lesson to be drawn here, is that it is possible to divide business cycle expansions and contractions into sequences of stages that are reasonably consistent and interestingly related. The key is to combine the proper measures relating to the major fluctuations in levels and deviations from trend of aggregate economic activity. However, this is probably not the only interesting approach to the classification of business cycles by phases and stages. Also while relatively transparent conceptually, it is quite difficult empirically like all methods that require estimations of macroeconomic trends that tend to vary over time with cyclical and other forces.

References

Boschan, C. and W.W. Ebanks (1978), The Phase-Average Trend: A New Way of Measuring Growth. In American Statistical Association (ed.), *Proceedings of the Business and Economic Statistics Section*. Washington, DC.

Heilemann, U. and Münch, H. J. (2006), Classification of U.S. Business Cycle Stages. In this volume.

[5] Note again the recessions and recoveries of 1948–50, 1953–55, 1957–58, 1960–61, and 1981–83 in Figure 2. The 1973–75 recession, with its short and sharp decline, had initially also a strong rebound as an aftermath but here the recovery slowed and lengthened later in 1976.

Keynes, J.M. (1936), The general theory of employment, interest, and money. London: McMillan.

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

Schumpeter, J.A. (1939), Business cycles: a theoretical, historical and statistical analysis of the capitalist process. Vol. 1. New York: McGraw-Hill.

Zarnowitz, V. (1985), Recent work on business cycles in historical perspective. *Journal of Economic Literature* 23: 523–580.

Zarnowitz, V. (1992), Business cycles – theory, history, indicators, and forecasting. NBER Studies in Business Cycle Research 27. Chicago and London: University of Chicago Press.

Zarnowitz, V. and A. Ozyildirim (2002), Time Series Decompensation and Measurement of Business Cycles, Trends and Growth Cycles. NBER Working Paper 8736. NBER, Cambridge, MA.

Ullrich Heilemann and Heinz Josef Münch

The U.S. Business Cycle and its Phases 1948–2000

1. Introduction

This paper is about the decomposition of the business cycle of the U.S. economy. The decomposition of the business cycle into different but related segments, stages, or phases is by definition an essential characteristic of business cycles: "A cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle" (Burns, Mitchell 1946: 3). Next, defining the business cycle, "cyclical taxonomy" serves two purposes. First, it offers a deeper understanding of its subject. To know "what happens during business cycles" (Mitchell 1951) or "the different economic process from stage to stage" (Burns, Mitchell 1946: 29) may help to find out the "stylised facts" to be explained and identify the forces that generate them. Second, if the phases can clearly be separated and relationships within them and between them are stable, this should help with forecasting in general and to identify in time pathological states of the economy in particular.

While cyclical taxonomy is, almost by definition, as old as business cycle analysis, the now prevailing foundations were laid by the seminal contributions of Mitchell (1927), Spiethoff (1955 [1925]), Haberler (1937), and Burns/Mitchell (1946). While Haberler' s summing up of business cycle theories led him to the currently standard four-phase scheme, Burns/Mitchell adhered to the two-phase scheme ("Expansion" and "Contraction"), comprising nine stages.[1] Both approaches had analytical orientations, although rather different ones. Forecasting intentions played no great role in them. The advent of macroeconomics after WW II and its dominance soon after greatly reduced the interest in business cycle research and in the Burns/Mitchell approach. For a long time its use was confined to the *National Bureau of Economic Research* (NBER) and its analyses of business cycles and growth cycles (Zarnowitz

[1] For definition and details see e.g., Zarnowitz 2006b.

DOI https://doi.org/10.3790/978-3-428-52425-9 | Generated on 2025-02-22 16:18:59

1992: 316ff.; Zarnowitz 2006b); however, univariate derivates of a simple two-phase scheme were in wide use to mark the upswings/downswings of GDP growth.

Burns/Mitchell, grounded in the German historical school and in inductive reasoning, were well aware that taxonomy can serve only as a starting point for a theoretical analysis. But even 40 years later, the taxonomic approach is still criticised for being only phenomenally oriented, presenting "associations" and not "causes" (Auerbach 1986), the same arguments that were made in the Koopmans/Vining-debate. The recent heightened interest in and discussion of "stylised facts" of the business cycle seems to have softened this criticism. (Harding, Pagan 2006).

The nevertheless unchallenged authority of the Burns/Mitchell scheme was not much contested until 1975, when John Meyer and Daniel Weinberg (M/W) (1975a, b, 1976) presented a new scheme to classify U.S. business cycles. But despite the results of their approach, which were promising in many directions, and its even more promising analytical possibilities, M/W's contribution remained unnoticed[2]. Eckstein/Sinai (1986) while on a similar track in their article about a five-stages scheme in 1984[3] did not even mention it.[4]

The starting point of M/W's approach was the NBER's dating of business cycles. It was changed in four respects: First, the two-phase scheme was augmented to a four-phase scheme.[5] It includes "Recession", "Recovery", and the two "modern" phases "Demand Pull" and "Stagflation". Second, the very strong real or production orientation of the Burns/Mitchell-scheme (Burns, Mitchell 1946: 20) was broadened according to the post-WWI experience. Beyond real GDP it included 18 variables from spheres like employment, prices, money, government, and foreign trade. Third, the selection of the variables included and their weight to separate the various phases or cycles was achieved by using multivariate linear discriminant analysis (LDA).

2 As *Victor Zarnowitz* had revealed in a personal communication in January 2002, M/W's approach was discussed at the NBER – Meyer was then president of NBER (also M/W 1976) but did not find much interest. The main reason for this was M/W's by and large ignorance of the distinction between business cycles and growth cycles. Other factors were their transition of the production sphere (see also Zarnowitz 2006b).

3 The work on the scheme, established in the early 1980s for 1945–4 to 1982–4, had started in the mid 1970s. Its dating of the cycles followed the NBER chronic of business cycles, but with its distinction of five stages, the authors tried to pay more tribute to the role of finance market conditions ("credit crunch", "reliquidation") during recession and recovery.

4 M/W (1975a: 167) express their appreciation of *Otto Eckstein*'s (and others) for their helpful comments. In 1980 Eckstein had also commented on Heilemann's M/W scheme based examination of the forecasting accuracy of the DRI model of the U.S. economy (Heilemann 1982).

5 Equally could be argued that the 9-stages scheme was reduced to a 4 stages scheme. Here "stages" and "phases" are used synonymously.

Finally, the traditional concept of "business cycles" was relaxed in the direction of "growth cycles" to be able to be applied on e.g. the output of macroeconometric models. The final scheme (stages and variables) and the classification power of its variables were successfully tested, not only for the (then) passed five U.S. post-WW II cycles but also for pre-WW II cycles. In various updates and extensions by M/W and the present authors (for the U.S.: Heilemann 1982; Heilemann, Münch 2004; for Germany: M/W 1975b; Heilemann, Münch 1999, 2001), the scheme proved to still be surprisingly successful, despite the fact that the sample period had almost doubled.

The overall encouraging results for the M/W scheme suggest its (new) re-examination. The present paper does so for the U.S. economy for the period 1948 to 2000. Although the scheme and some of its classification performances appear remarkably stable, a number of tests also point towards important changes in the nature and causes of U.S. post-WW II cycles – or of M/W's "modern view" of the cycle. M/W's idea of the modern cycle was dominated by the growth/inflation experience, or trade-off, as perceived in the early 1970s, reflecting policy, most of all fiscal policy interventions and their immediate fiscal consequences and stabilizing economic fluctuations. Effects that had not played significant roles in Burns/Mitchell analyses were confirmed also, for example, by Zarnowitz/Moore (1986) in the framework of the NBER business cycle approach. The present paper will suggest that the variables and the scheme used by M/W will need some reworking. Though it is beyond the possibilities of this paper to do this, some suggestions will be made.

Section 2 describes in some detail M/W's four-phase scheme and the data employed, and it presents the results of our replication of M/W's results up to 1973 (a short description of this can be found in the Appendix). Section 3 reports on the extension of M/W's classification up to 2000 and discusses some economic implications of the results. In the light of the present results, Section 4 reflects on the methodical efficiency of reductionistic approaches to investigate macroeconomic fluctuations and makes some suggestions for future research.

2. Re-examining Meyer/Weinberg

The phases of the M/W scheme had been defined as a "first approximation," as follows (M/W 1975a: 172f.): (1) Recession: a period of some duration in which total aggregate activity actually declines somewhat from previous peak levels and is reasonably widely diffused throughout the economy. (2) Recovery: the early expansion out of a recession and a state of economic affairs in which everything is "going well" – unemployment is declining, prices are relatively stable, productivity is rising, and total output is expanding. (3) Demand-Pull Inflation: the classic inflationary situation, in which "too much money chases

Table 1

Classification of U.S. Business Cycles into a 4-stage Scheme
1948-5 to 2000-12

	Cycle[1]	Starting months of …			
		Recovery	Demand-Pull	Stagflation	Recession
1	1948–5 to 1949–10 (18)	…	…	1948–5 (7)	1948–12 (11)
2	1949–11 to 1954–7 (57)	1949–11 (8)	1950–7 (6)	1951–1 (34)	1953–11 (9)
3	1954–8 to 1958–4 (45)	1954–8 (7)	1955–3 (30)	–	1957–9 (8)
4	1958–5 to 1961–1 (33)	1958–5 (25)	–	–	1960–6 (8)
5	1961–2 to 1970–11 (118)	1961–2 (51)	1965–5 (31)	1967–12 (25)	1970–1 (11)
6	1970–12 to 1975–3 (52)	1970–12 (25)	1973–1 (21)	–	1974–10 (6)
7	1975–4 to 1980–9 (66)	1975–4 (39)	1978–7 (12)	–	1979–7 (15)
8	1980–10 to 1982–12 (27)	1980–10 (6)	1981–4 (6)	–	1981–10 (15)
9	1983–1 to 1991–11 (107)	1983–1 (15)	1984–4 (43)	1987–11 (36)	1990–11 (13)
10	1991–12 to 2000–12 (109)	1991–12 (109)	…	…	…
	All				
	1948–5 to 2000–12 (632)	285	149	102	96

Sources: Meyer/Weinberg (1948–5 to 1973–9), and authors' computations (1973–10 to 2000–12). – [1]Cycle/phase length in parentheses.

too few goods." The forces of recovery are somehow allowed to achieve too much force or pull, with production forced up to capacity constraints, prices rising, rates of productivity improvement declining, etc. (4) Stagflation: a situation of stagnation at a high level of activity mixed with price inflation. The strains of demand-pull perhaps recede, and total monetary expansion diminishes; however, prices and wages continue to increase, perhaps because of catch-up effects due to sectoral imbalances created during the preceding demand-pull inflation or because productivity does not improve enough to stabilise wage cost.

As laid out before, M/W's scheme differs from the standard four-phase scheme (e.g., Haberler (1963 [1937]: 257ff.) mainly by re-defining "upswing" and "downswing," complementing "economic activity" by "inflation" and other characteristics of the modern cycle as essential qualities of the business cycle, and applying LDA to detect the classifying variables.

The estimation of the discriminant functions requires an a priori classification. M/W started with the two-stage NBER classification of the period from February 1947 to September 1973. Demand-Pull and Stagflation were separated – from Upswing and Recession, respectively – by "common economic sense" augmented by general knowledge of "recent business cycle history" (M/W 1975a: 175)[6]. The thus derived *a priori* classification of the sample period was then classified with Bayesian LDA based on 20 variables. Boundary months

6 For overviews over the various U.S. cycles, see the corresponding articles in Glasner 1997 or Zarnowitz 1992: 20ff.

Table 2

Average Values of Classifying Variables
1948–5 to 2000–12

Variable		Stage[1] Recovery	Demand-Pull	Stagflation	Recession	All
Real GNP[2]	a	4.20	5.20	4.94	–0.10	3.95
	b	3.59	4.30	3.37	–0.28	3.17
	c	3.48	4.03	3.37	–0.82	3.08
	d	3.84	4.51	4.39	–0.32	3.45
Unemployment rate	a	5.76	4.18	3.29	5.47	4.79
	b	6.34	6.79	5.43	7.63	6.50
	c	6.01	7.02	5.43	8.16	6.37
	d	6.10	5.33	4.04	6.56	5.66
Index of unit labor cost, private economy[2]	a	–0.49	2.14	4.98	4.03	2.04
	b	1.53	1.80	2.32	5.76	2.26
	c	0.28	0.68	2.32	3.92	1.08
	d	0.70	2.49	4.04	5.42	2.38
Govt. surplus or deficit as per cent of GNP[2]	a	–0.22	–0.19	0.15	–0.28	–0.14
	b	–1.97	–2.90	–2.36	–2.87	–2.32
	c	–1.64	–3.44	–2.36	–3.53	–2.33
	d	–1.26	–1.34	–0.73	–1.58	–1.24
GNP price deflator[2]	a	2.06	3.28	4.02	2.20	2.81
	b	3.57	4.46	3.66	6.36	4.14
	c	2.57	3.71	3.66	5.15	3.26
	d	2.96	4.13	3.89	4.58	3.63
Prime rate[3]	a	0.07	1.73	1.20	–1.00	0.56
	b	0.46	0.39	0.30	–1.55	0.15
	c	0.52	–0.15	0.30	–2.87	–0.04
	d	0.30	1.19	0.88	–1.60	0.32
Gross govt. expenditures[2]	a	2.36	6.07	23.40	3.39	8.00
	b	3.78	8.64	3.28	9.95	5.54
	c	2.48	8.27	3.28	8.02	4.40
	d	3.20	7.16	16.30	6.94	6.82
Money supply M2[2]	a	7.88	6.37	5.68	3.61	6.37
	b	6.66	8.08	5.02	7.11	6.82
	c	5.19	8.18	5.02	6.66	5.94
	d	7.16	7.05	5.45	5.31	6.58
Money supply M1[2]	a	3.56	3.65	4.62	0.75	3.38
	b	4.64	9.92	3.00	6.18	5.71
	c	4.10	10.41	3.00	5.96	5.42
	d	4.20	6.35	4.05	3.40	4.56
Net exports as per cent of GNP	a	0.23	0.41	0.54	0.83	0.43
	b	–0.66	–0.54	–0.50	–0.01	–0.53
	c	–0.86	–0.66	–0.50	–0.02	–0.67
	d	–0.30	–0.01	0.17	0.41	–0.05
Wholesale price index, industrial[2] commodities only	a	0.87	3.51	3.57	0.96	2.12
	b	3.26	2.97	4.03	7.67	3.91
	c	2.09	1.43	4.03	2.99	2.35
	d	2.29	4.42	3.73	5.43	3.50
Compensation per man-hour[2]	a	4.59	5.96	5.67	2.83	4.89
	b	5.26	4.94	5.09	4.91	5.13
	c	5.36	5.01	5.09	5.09	5.22
	d	4.99	5.42	5.46	3.89	5.00
Average yields on corporate bonds (Moody's)[3]	a	0.12	1.00	0.67	–0.60	0.35
	b	0.03	0.09	–0.25	–0.21	–0.02
	c	0.03	–0.01	–0.25	–1.32	–0.18
	d	0.07	0.68	0.34	–0.45	0.18
Consumer price index[2]	a	1.57	2.54	4.63	2.25	2.58
	b	4.05	5.13	4.67	8.40	4.94
	c	3.23	4.09	4.67	5.83	3.92
	d	3.04	4.21	4.64	5.59	3.96

Table 2 cont.

Average Values of Classifying Variables
1948–5 to 2000–12

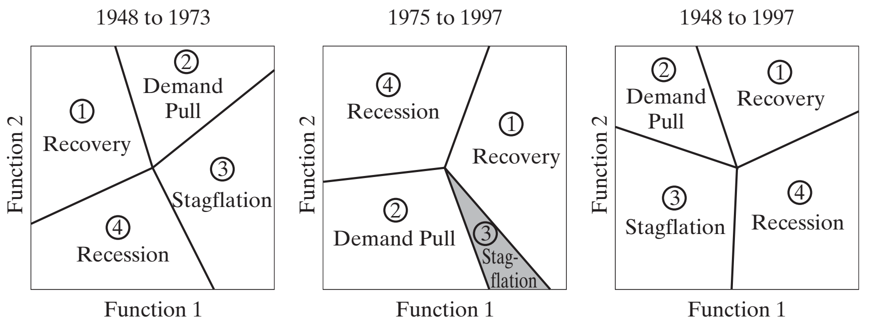
Variable		Stage[1]				
		Recovery	Demand-Pull	Stagflation	Recession	All
Consumer price index, food only[2]	a	0.89	3.16	5.11	1.55	2.47
	b	3.52	5.42	5.15	5.56	4.37
	c	2.73	3.93	5.15	3.84	3.46
	d	2.45	5.15	5.12	3.92	3.74
Output per man-hour[2]	a	3.89	2.84	2.14	1.64	2.90
	b	2.39	1.54	1.09	-0.15	1.72
	c	2.34	1.64	1.09	0.21	1.77
	d	3.00	1.99	1.77	0.73	2.22
N.Y. Stock Exchange composite price index[2]	a	1.23	0.34	0.17	0.33	0.64
	b	0.92	0.87	0.25	1.58	0.93
	c	1.02	0.97	0.25	1.51	0.95
	d	1.05	0.24	0.20	1.11	0.73
Consumer price index, all commodities[2] exept food	a	1.80	2.31	4.52	2.77	2.67
	b	4.13	5.06	4.57	9.00	5.04
	c	3.26	4.13	4.57	6.23	3.97
	d	3.18	3.91	4.54	6.13	4.02
Wholesale price index[2]	a	0.71	3.95	3.44	0.39	2.06
	b	2.91	3.00	4.12	6.36	3.55
	c	1.95	1.26	4.12	2.12	2.15
	d	2.02	4.61	3.68	4.2	3.23

Authors' computations. – [1]a: Results for period 1948–5 to 1973–9, b: 1975–4 to 2000–12, c: 1980–10 to 2000–12, d: 1948–5 to 2000–12. – [2]Changes are against previous year. – [3]Per cent change per month.

between cyclical stages were – in an iterative way – re-assigned according to the LDA classifications. The variables used in the initial LDA were (1) much the same variables the NBER used in its cycle chronic, (2) variables suggested by policy and (3) by historical considerations, (4) variables that figured prominently in macroeconometric models, or (5) variables that had been singled out as particularly sensitive cyclical indicators (M/W 1975a: 176); however, whilst the NBER business cycle dating is based on levels (Zarnowitz 2006b), classification procedures like LDA have to be based on "stationary" data to deliver reasonable results. Therefore, all variables with an underlying trend had to be transformed into changes or differences. The resulting dating of the first six post-WW II cycles and their stages is shown in Table 1. The average values in the four stages "more or less confirm prior expectations in different cyclical stages" (M/W 1975a: 178; Table 2, line a).

M/W had not revealed the complete set of variables they had tested and the criteria for inclusion in their discriminant functions. Looking at the variables today, it is astonishing that they had their data base most of all broadened and not much deepened. The list of disaggregated demand variables (Gross government expenditures, Net exports as percent of GNP) is surprisingly short and does not include, for example, private consumption or business investment.

Figure 1
Stability of the Meyer/Weinberg-Scheme for the U.S. Business Cycle
1948 to 1997



Authors' computations.

As to the number of discriminant functions, eigenvalues and cumulative proportions of "explained" dispersion of variance had led M/W to find two (canonical) discriminant functions as sufficient for classification, especially since they thought this would allow them a straightforward interpretation of results. The first discriminant function differentiates by unemployment, interest rate changes, productivity, and various price deflators, thus separating recessions and recoveries from the two "inflation" periods. "Specifically, high unemployment rates, good productivity gains, negative changes in corporate bond rates, and small to negative price changes will yield a high negative score on this index; opposite conditions will register positively" (M/W 1975a). The second function apparently adds only a little to this differentiation. Mainly the pattern of interest rates helps somewhat in separating the "growth" stages (Recovery, Demand-Pull) from the two "no-growth" periods (1948 to 1973; Figure 1).

To test their scheme, M/W classified periods that had not been used for the estimation of their functions. Results for both "back-casting" – after some modifications of the classifying variables and their periodicity – for the period 1920 to 1951 and for forecasting with the original variable set (1973–10 to 1974–9) were seen as confirmations of the scheme (M/W 1975a: 184ff.). In the present case its power is very much limited: in a technical sense, since the number of observations/cases is for LDA comparatively small; in an economic sense, since the reader may miss a more direct analysis of possible causes of a changed business cycle.[7]

7 Many of the arguments listed in Zarnowitz/Moore (1986), Burns/Mitchell (1946: chapter 10), Gordon (1986) or even Burns/Mitchell (1946) could have been examined. However, one has to concede, all the authors listed and many others were not too ambitious in doing so.

Table 3

Estimation Results for the Standardized Canonical Discriminant Functions[1]
1948–5 to 2000–12

Variable		Coefficients of function[1]			F-Value to enter
		1	2	3	
Real GNP[2]	a	0.70	1.06	0.04	54.6
	b	–0.31	–0.44	0.42	87.8
	c	–0.67	0.36	0.59	91.4
	d	–0.51	0.50	0.13	123.8
Unemployment rate	a	1.12	–0.03	0.24	292.7
	b	0.90	2.18	1.87	19.0
	c	2.79	0.19	1.49	29.4
	d	1.21	0.66	0.38	75.9
Index of unit labor cost, private economy[2]	a	0.20	–0.12	–0.51	64.4
	b	0.11	0.40	0.22	15.8
	c	–0.02	0.21	–0.11	13.8
	d	–0.03	–0.14	0.30	52.5
Govt. surplus or deficit as per cent of GNP[2]	a	–0.10	0.12	0.47	3.6
	b	0.56	1.83	1.94	5.4
	c	2.50	0.36	2.17	16.5
	d	0.78	0.20	0.30	4.1
GNP price deflator[2]	a	0.66	0.80	–1.39	20.8
	b	–1.86	–0.98	–0.92	17.9
	c	–0.85	0.13	–1.20	19.4
	d	–0.08	0.41	0.17	15.4
Prime rate[3]	a	–0.05	0.04	0.00	12.4
	b	–0.09	0.15	0.11	2.8
	c	0.12	0.19	0.18	7.3
	d	–0.02	0.16	–0.10	12.7
Gross govt. expenditures[2]	a	–0.27	–0.25	0.28	33.8
	b	0.33	–1.15	0.51	40.9
	c	–0.46	–0.86	1.31	34.5
	d	–0.06	–0.22	–0.38	34.8
Money supply M2[2]	a	0.28	0.95	0.38	33.1
	b	0.02	0.00	–0.45	7.0
	c	–0.84	0.10	–0.80	15.8
	d	–0.17	0.18	0.09	15.2
Money supply M1[2]	a	–0.46	–0.50	0.24	30.6
	b	0.08	–0.26	0.68	28.4
	c	–0.46	0.03	0.88	25.5
	d	–0.22	–0.09	0.72	15.3
Net exports as per cent of GNP	a	–0.06	0.40	–0.11	12.2
	b	–0.08	0.60	0.41	7.5
	c	0.30	0.46	0.47	8.2
	d	–0.14	–0.37	0.45	21.2
Wholesale price index, industrial[2]	a	–0.46	–0.78	1.03	19.0
	b	–0.94	2.98	2.31	12.1
	c	–0.18	1.23	–0.64	4.6
	d	–0.48	0.91	0.57	12.2
Compensation per man-hour[2]	a	–0.25	0.47	0.15	23.8
	b	0.04	–0.02	0.20	13.6
	c	–0.19	–0.05	0.21	10.1
	d	–0.30	0.10	–0.02	21.1
Average yields on corporate bonds (Moody's)[2]	a	–0.09	0.23	–0.11	10.6
	b	0.01	0.06	0.10	0.2
	c	–0.02	–0.01	0.06	2.2
	d	–0.07	0.12	0.06	5.5
Consumer price index[2]	a	1.24	1.48	1.63	31.4
	b	–0.09	0.24	0.23	29.4
	c	0.04	0.08	0.04	14.8
	d	0.62	0.90	0.16	20.2

Table 3 cont.

Estimation Results for the Standardized Canonical Discriminant Functions[1]
1948–5 to 2000–12

Variable		Coefficients of function[1] 1	2	3	F-Value to enter
Consumer price index, food only[2]	a	–0.87	–0.98	0.78	40.7
	b	2.59	–0.73	–0.71	34.0
	c	0.44	–1.46	0.27	16.0
	d	–0.16	–1.76	–1.38	26.0
Output per man-hour[2]	a	–0.06	–0.25	–0.08	24.7
	b	–0.61	0.48	0.10	57.1
	c	0.58	0.71	–0.34	31.9
	d	0.32	0.18	–0.59	48.1
N.Y. Stock Exchange composite price index[2]	a	–0.07	0.17	0.12	2.4
	b	0.10	0.09	0.20	1.1
	c	0.06	–0.06	0.16	0.8
	d	0.04	0.02	–0.06	3.3
Consumer price index, all commodities[2] except food	a	–1.26	–1.14	0.16	18.3
	b	0.20	–0.06	0.42	12.9
	c	–0.43	0.48	–0.23	21.8
	d	–0.88	–0.06	–0.17	20.8
Wholesale price index[2]	a	0.15	0.31	–1.85	16.5
	b	–0.27	–1.49	–2.27	8.5
	c	0.44	–0.01	0.20	6.3
	d	–0.88	–0.06	–0.17	20.8

Eigenvalues

Function		Eigen-value	% of variance	cumula-tive %	canonical correlation	after function	Wilks' λ	χ^2	df	Signifi-cance
a	1	4.8	70.3	70.3	0.9	1	0.4	920.3	57	0.00
	2	1.4	19.6	89.9	0.8	2	0.3	404.1	36	0.00
	3	0.7	10.1	100.0	0.6		0.6	154.1	17	0.00
b	1	2.2	50.0	50.0	0.8	1	0.1	752.3	57	0.00
	2	1.6	37.7	87.7	0.8	2	0.2	411.8	36	0.00
	3	0.5	12.3	100.0	0.6		0.7	126.0	17	0.00
c	1	3.3	53.0	53.0	0.9	1	0.0	736.0	57	0.00
	2	2.0	31.9	84.9	0.8	2	0.2	401.9	36	0.00
	3	0.9	15.1	100.0	0.7		0.5	151.3	17	0.00
d	1	1.3	56.0	56.0	0.8	1	0.2	1018.3	57	0.00
	2	0.9	36.0	92.0	0.7	2	0.5	492.5	36	0.00
	3	0.2	8.0	100.0	0.4		0.8	108.0	17	0.00

Authors' computations. Eigenvalue: eigenvalues of the discriminant functions in declining order. % of variance: % importance of the discriminant functions. cum %: cumulative importance in relative terms. df: degrees of freedom. For a detailed description of the statistics see Brosius (1989). – [1]a: Results for period 1948–5 to 1973–9, b: 1975–4 to 2000–12, c: 1980–10 to 2000–12, d: 1948–5 to 2000–12. – [2]Changes are against previous year. – [3]Per cent change per month.

Before extending the M/W analysis to 2000, we tried to reproduce their results. Because of revisions, redefinitions, etc. of the data, this is a notoriously burdensome and unfinished exercise. Most of the variables employed by M/W were seasonally adjusted and de-trended by transforming them to annual percentage rates of changes. Because some variables were not available for us in seasonally adjusted form, we decided to de-trend by using change rates against previous year: a simple, usually effective, method of seasonal adjustment. One important consequence of this was (besides only a small number of missing variables) that the start of our analysis was shifted forward by fifteen months. A further difference to M/W is that interest rate changes were calculated as percentage changes per month. The variables used and their average values are listed in Table 2 (line a).

Also different from M/W, we opted for the use of three discriminant functions, but this had no consequences for the parameters of the first two functions. Finally, in all analyses, Money GNP was excluded from the set of classifying variables because it failed to be significant, especially in small samples. Results with considerably smaller sets of variables were, to a large extent, similar to those derived with the 19 variables, but for comparison with M/W we present results for the larger set. All in all, the reproduction of M/W results appears as not too bad for both discriminant functions (Table 3, line a) and accuracy of classification (cumulative percent explanation). According to the F-values to enter, the new results differ most of all with respect to Government Surplus, GNP deflator, and Prime rate, which sink from ranks 4, 5, and 6 to ranks 19, 11, and 15, respectively. Correspondingly, the Consumer price index (excluding food) went up to rank 5. The classification results with the newly estimated discriminant functions improved slightly, and the total error rate fell from 10 percent (31 cases) to 9 percent (29 cases), mainly because the explanation of Demand-Pull stages has been improved. But Stagflation is, like in M/W's final classification, identified only in four and Demand-Pull only in five of the first six post-WW II cycles. Table 6 in the Appendix presents the misclassified periods in M/W's analysis and in our re-estimation. All in all, the economic results are much in line with those described by, for example, Zarnowitz/Moore (1986).

3. Stability of Meyer/Weinberg's Results

While the reproduction of the M/W results leaves room for improvement, it also encourages an updating. There is no doubt that since the 1970s the then "modern" cycle and its phases experienced a number of new changes as to duration, intensity, and volatility but also with respect to the role and influence of the constituting variables, as a number of studies have shown (for example Zarnowitz/More 1986, Gordon 1986 and Gordon (ed.) 1986, Zarnowitz 2006a). Our update of the M/W scheme after 30 years is therefore not only a test of its stability but should also tell something about degree of cyclical change, its forms, and its causes.

The update will be made in two steps. First, we check the classification power of the original discriminant functions estimated in the period 1974/2000; second, starting again with the 1948/73 sample, the functions are re-estimated by consecutively including later cycles.

Because all data were used for re-estimation, no data were available to examine the outside-sample performance of our findings, but a number of leave-one-cycle out-tests should be sufficiently good substitutes.

3.1 A Four-phase Classification for 1973 to 2000

The classification procedure follows M/W (1975a). The (iterative) procedure started again with the NBER classification of 1973/2000 into Recovery/Demand-Pull and Stagflation/Recession. This a priori classification of the new sample period was modified according to the classification results of discriminant functions, estimated for various sample periods. The five new cycles and their phases from 1974 to 2000 are shown in Table 1. As could be guessed from the previous results and from history since 1973, Stagflation is identified only in one of the five new cycles.

To get an impression of the economic essence of this classification, Table 2 (line b) presents average values for the classifying variables. The results are mostly in line with our present understanding of the stylised facts of the U.S. cycle. When compared with averages of the M/W-sample (1948–5 to 1973–9), the levels of some variables (rates of change) are different, but all in all the differences between phases is not much different from that of the M/W-sample.

3.2 New Sample Performance of the M/W Scheme

Estimation results for the new sample show for many variables new ranks of importance (Table 3, line b, F-values to enter). In particular, this is the case for Unemployment and Real GNP, as could already have been presumed from their average values (Table 2). Only four of the 19 variables corroborate their previous ranks: Gross government expenditure, M1, Compensation per man-hour, and Output per man-hour. Most of them are of minor importance in the discriminant functions.

The classificatory meaning of most variables has been reduced, in particular that of the various measures of inflation. The importance of Real GNP and Net exports as a percentage of GNP, both indicators of economic activity, has increased. Though economic interpretations of these results must be cautious, the results seem to underline that inflation-related variables lost discriminating power.

The "explained variance" (Table 3) by the first discriminant function is reduced from nearly 70 percent to 50 percent, corresponding to a doubling of this ratio in the second function from 20 percent to 40 percent. This also confirms the picture emerging from the "F-values to enter".

Table 4

Classification Results for Different Samples
1948–5 to 2000–12

Actual group	Predicted group membership[1]				
	No. of cases	Recovery	Demand-Pull	Stagflation	Recession
			1948–5 to 1973–9		
Recovery	116	105 (90,5)	3 (2.6)	0 (0.0)	8 (6.9)
Demand-Pull	76	1 (1.3)	68 (89.5)	6 (7.9)	1 (1.3)
Stagflation	66	0 (0.0)	3 (4.5)	62 (93.9)	1 (1.5)
Recession	47	2 (4.3)	2 (4.3)	2 (4.3)	41 (87.2)
Total error rate: 9.5%					
			1948–5 to 2000–12		
Recovery	285	207 (72.6)	54 (18.9)	0 (0.0)	24 (8.4)
Demand-Pull	149	20 (13.4)	100 (67.1)	22 (14.8)	7 (4.7)
Stagflation	102	3 (2.9)	16 (15.7)	80 (78.4)	3 (2.9)
Recession	96	2 (2.1)	0 (0.0)	7 (7.3)	87 (90.6)
Total error rate: 25.0%					
			1961–2 to 2000–12		
Recovery	245	207 (84.5)	24 (9.8)	0 (0.0)	14 (5.7)
Demand-Pull	113	5 (4.4)	97 (85.8)	7 (6.2)	4 (3.5)
Stagflation	61	4 (6.6)	0 (0.0)	56 (91.8)	1 (1.6)
Recession	60	1 (1.7)	0 (0.0)	6 (10.0)	53 (88.3)
Total error rate: 13.8%					
			1970–12 to 2000–12		
Recovery	194	160 (82.5)	18 (9.3)	8 (4.1)	8 (4.1)
Demand-Pull	82	7 (8.5)	67 (81.7)	6 (7.3)	2 (2.4)
Stagflation	36	0 (0.0)	0 (0.0)	36 (100.0)	0 (0.0)
Recession	49	2 (4.1)	0 (0.0)	1 (2.0)	46 (93.9)
Total error rate: 14.4%					
			1975–4 to 2000–12		
Recovery	169	145 (85.8)	12 (7.1)	7 (4.1)	5 (3.0)
Demand-Pull	61	2 (3.3)	56 (91.8)	1 (1.6)	2 (3.3)
Stagflation	36	1 (2.8)	0 (0.0)	35 (97.2)	0 (0.0)
Recession	43	2 (4.7)	0 (0.0)	1 (2.3)	40 (93.0)
Total error rate: 10.7%					
			1980–10 to 2000–12		
Recovery	130	119 (91.5)	3 (2.3)	5 (3.8)	3 (2.3)
Demand-Pull	49	1 (2.0)	45 (91.8)	2 (4.1)	1 (2.0)
Stagflation	36	0 (0.0)	1 (2.8)	35 (97.2)	0 (0.0)
Recession	28	0 (0.0)	0 (0.0)	1 (3.6)	27 (96.4)
Total error rate: 7.0%					

Authors' computations. – [1]Share in % in parenthesis.

The total error rate of classification increases over the new sample period to nearly 11 percent. This is somewhat worse than the 9.5% that has been recorded for the 1948/1973-sample (Table 4), mainly a result of a better explanation of the Recoveries in the older sample. For all other phases the new sample performs better.

3.3 Outside Sample Performance

A strong test of the newly estimated M/W scheme is the accuracy of its classification for the complete sample period (1948 to 2000), technically seen an *outside-/within*-sample period test. The overall error rate (not shown here) increases from 9 percent to more than 60 percent, which signals structural breaks or shifts in the overall sample 1948/2000. The deterioration is not evenly distributed over all phases and cycles; particularly bad results are experienced for the two inflation periods. The break in the succession of the phases presented in Figure 1 seems to have happened in the late 1960s, but we should also judge with care. "Ex-post forecasts" in an outside-sample period as long as the sample period caused severe stability problems for, as an illustration, the Permanent Income Hypothesis as well as for the NBER leading indicator. One may even wonder why, given the considerable changes of the U.S. economy, the results here are so remarkably stable. The reader of Burns/ Mitchell's (1946) work might have similar thoughts.

3.4 Stability of the M/W-Scheme in the New Sample

Looking at the phases of U.S. cycles after 1970 as chronicled in Table 1, Stagflation, together with Demand-Pull, major innovations in M/W's "modern business cycle," nearly completely disappears. While three Stagflations were identified in the six cycles examined by M/W, in the following four cycles only one such phase is identified. Although M/W and others before[8] (Bur had not assumed that each cycle must comprise all phases), the near complete vanishing of Stagflation might be seen as a serious challenge for the M/W scheme. A consequence of the missing Stagflation may be seen in the new sequence of phases (Figure 1). While for the M/W and the complete sample period the succession of phases seems to be somewhat "natural" (Recession → Recovery → Demand Pull → Stagflation → Recession → …), in the 1974/2000 period the succession of Demand Pull and Stagflation is reversed.

Compared with the old sample, lengths of cycles and phases have been rather stable. Restricted to full cycles, the average duration is still 62 months (NBER: 63). Recoveries now last 22 months (M/W: 23), Demand-Pull 21 months (22), Stagflations 32 months, (30) and Recessions 9 months (11 months). A similar

8 Burns/Mitchell (1946: 7ff.) are a notable exception.

Table 5

Error Rates of Leave-one-cycle-out Classifications[1]
1948–5 to 2000–12

	Cycle	Recovery	Demand-Pull	Stagflation	Recession	All
1	1948–5 to 1949–10	–	–	0	9	6
2	1949–11 to 1954–7	25	50	82	0	58
3	1954–8 to 1958–4	43	97	–	13	73
4	1958–5 to 1961–1	24	–	–	25	15
5	1961–2 to 1970–11	24	58	56	100	47
6	1970–12 to 1975–3	96	67	–	–	73
7	1975–4 to 1980–9	77	75	–	73	76
8	1980–10 to 1982–12	100	100	–	20	56
9	1983–1 to 1991–12	0	100	100	46	79
10	1992–1 to 2000–12	100	–	–	–	–
Total sample period 1948–5 to 2000–12		27	33	22	13	25

Authors' computations. – [1]Summary of Table 7, Appendix.

picture emerges for the differences between the phases when measured by the relative difference between the averages of the classifying variables.

A similar picture emerges from more explicit stability tests (Table 3c, d). While the classification results for various post-1973 samples continuously improve, as signalled by total error rates (Table 4 for 1980/2000), the stability, as indicated by the "leave one cycle out" results outside the sample period, is continuously worsening (Table 5 and in detail Table 7, Appendix). The verdict may be seen in a milder light when it is taken into account that this test produces for the MW-sample (1948/1973) error rates of more than 40 percent for cycles 2, 3, 5, and 6. Again, given the long time span outside the sample period, the expectation of a stable explanatory accuracy is already for simple empirical considerations not realistic.

Looking for "causes" of this deterioration, a first explanation is found in the unfolding of co-factors, i.e., the average values. The changed levels of most price variables (e.g., GNP price deflator) and of some indicators of economic activity make it clear that M/W's "modern cycle" so far was an episode only.

4. Summary and Conclusions

The re-estimation of M/W's classification over the old sample period and its extension forward to 2000 showed an unexpected stability of the four-phase scheme and its constituting variables. The classificationary performance over the post-1973 period proved to be nearly the same as that for the pre-1973 period, both results being much better than that for the complete sample period. This indicates shifts in the cyclical forces and changes in the importance of classifying variables in the scheme. While the roles of the most important variables such as real GNP, Unemployment rate, Consumer price index, and Money supply seem to have been rather stable, that of Net exports and Consumer prices increased and improved the explanation of the "two inflation phases."

Our findings suggest that M/W's modern scheme of the business cycle was only of temporary or episodic meaning. Whether it can be integrated into a more general, comprehensive system remains to be seen. The results also meet with the findings of Zarnowitz/Moore (1986), emphasizing the overall conformity of the U.S. cycle in the post-WW II period until the early 1980s[9]. Though there was a considerable evolution of the cycle caused by "various structural, institutional, and policy changes" (Zarnowitz, Moore 1986: 572), according to them the most basic characteristics of the cycle remained much the same. The present results suggest that behind this consistency were considerable shifts. For both cognitive and policy these shifts and changes deserve an intensive study. A reworked M/W scheme and multivariate LDA offer perspectives to be further explored.

Of course, the M/W approach is open to the employment of much of the business cycle research performed by Burns/Mitchell or in the NBER tradition. On a smaller, more feasible scale, particularly two avenues for further research seem advisable[10]. First, variables not used in the present discriminant functions should be classified, and their cyclical behaviour should be examined. Promising candidates are, for example, the more than 30 variables classified by Mitchell (1951: 256ff.) or influences examined by Zarnowitz/Moore (1986).[11] Of course, the results are the more convincing the better their inclusion is economically justified. At the same time it should be clear that classificatory usefulness of a variable cannot be the arbiter of cyclical relevance of a variable. A second way opened by the M/W-approach is to look more closely into the various stages. Is the transition from one phase to the next smooth or sudden? Which are the driving forces for transition, and did their role change over time? Much could already be gained by simple sensitivity studies of the discriminant functions. More consistent results would result, of course, in connection with a macroeconometric model explaining the

9 This also a result of a detailed analysis with an updated M/W-scheme for the Clinton era (Heilemann, Münch 2004).

10 The test of additional classification methods appears to have lower priority. Robustness and clarity of linear discriminant analysis make it, in the present context, first choice (Heilemann, Münch 1996).

11 An intresting category borrowed from Biology would be the distinction of allopatric and sympatric variables or categories. The former describing forms or species (Variables!) that do not occur together, the latter describing those that overlap or coincide.

variables of the discriminant function as one of the present authors did using the DRI model of the U.S. economy (Heilemann 1982; similar Eckstein, Sinai 1986).

References

Auerbach, A.J. (1986), Comment to Victor Zarnowitz and Geoffrey H. Moore, Major changes in cyclical behaviour. In R.J. Gordon (1986), 573–575.

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone (1984), *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Brosius, G. (1989), *SPSS/PC+ Advanced Statistics and Tables*. Hamburg and New York: McGraw-Hill.

Burns, A.F. and W.C. Mitchell (1946), Measuring business cycles. NBER Studies in Business Cycles 2. NBER, New York.

Eckstein, O. and A. Sinai (1986), The mechanism of the business cycle in the post-war era. In R. J. Gordon (ed.), 83–105.

Erb, W.-D. (1990), Anwendungsmöglichkeiten der linearen Diskriminanzanalyse in Geographie und Regionalwissenschaft. Schriften des Zentrums für regionale Entwicklungsforschung der Justus-Liebig-Universität Gießen 39. Hamburg: Verlag Weltarchiv.

Glasner, D. (ed.) (1997), *Business Cycles and Depressions – An Encyclopedia*. New York: Garland.

Gordon, R.J. (ed.) (1986), The American business cycle – Continuity and change. NBER Studies in Business Cycles 25. Chicago: University of Chicago Press.

Haberler, G. (1963 [1937]), *Prosperity and Depression*: A Theoretical Analysis of Cyclical Movements. New York: Atheneum [Geneva: Legue of Nations].

Harding, D. and A. Pagan (2006), Measurement of Business Cycles. University of Melbourne, Department of Economics Working Paper 966.) University of Melbourne.

Heilemann, U. (1982), Cyclical Taxonomy and Large Econometric Models. In R.F. Drenick and F. Kozin (eds.), System Modeling and Optimization. Lecture Notes in Control and Information Sciences 38. Berlin: Springer, 721–734.

Heilemann, U. and H.J. Münch (1996), West German Business Cycles 1963–1994: A Multivariate Discriminant Analysis. CIRET-Studien 50. Ifo-Institut, München, 220–250.

Heilemann, U. and H.J. Münch (1999), Classification of West German Business Cycles 1955-1994. *Jahrbücher für Nationalökonomie und Statistik* 219: 632–656.

Heilemann, U. and H.J. Münch (2001), Ifo Data and the German Business Cycle – Results of Multivariate Discriminant Analysis 1961–1997. *ifo-Studien* 47: 41–63.

Heilemann, U. and H.J. Münch (2004), The Clinton era and the U.S. business Cycle: What changed? *European Journal of Management and Public Policy* 3: 16–34.

Heilemann, U. and C. Weihs (2000), Diskriminanzanalyse. In W. Voss et al. (ed.), *Taschenbuch der Statistik*. Leipzig: Hanser, 583–608.

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

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

Meyer, J.R. and D.H. Weinberg (1976), On the classification of economic fluctuations: an update. *Explorations in Economic Research* 3: 584–585.

Mitchell, W.C. (1927), *Business Cycles:* The Problem and its Setting. NBER, New York.

Mitchell, W.C. (1951), *What happens during business cycles* – A progress report. NBER, Cambridge, MA.

Ripley, B.D. (1994), Neural Networks and related methods for classification. *Journal of the Royal Statistical Society* B 56: 409–456.

Sichel, D.E. (1994), Inventories and the three phases of the business cycle. *Journal of Business and Economic Statistics* 12: 2344.

Spiethoff, A. (1955 [1925]), *Die wirtschaftlichen Wechsellagen* – Aufschwung, Krise, Stockung. Band I. Tübingen and Zürich: Mohr/Siebeck and Polygraphischer Verlag AG [“Krisen”. Article in Handwörterbuch der Staatswissenschaften, 4th edition, vol. 6, Jena: Gustav Fischer, 8–91.

Weihs, C. and U. Sondhauss (2000), Business phase classification and prediction: How to compare interpretability of classification methods? SFB 475 Technical Report 24. Department of Statistics, University of Dortmund.

Zarnowitz, V. (1992), Business cycles – theory, history, indicators, and forecasting. NBER Studies in Business Cycle Research 27. Chicago and London: University of Chicago Press.

Zarnowitz, V. (2006a), Modern Trends and Their Effects on International Business Cycles. In this volume.

Zarnowitz, V. (2006b), Phases and Stages of Recent U.S. business Cycles. In this volume.

Appendix

Linear Discriminant Analysis

Modern classification analysis comprises a multitude of procedures for separation of groups and objects. Besides the oldest and most simple technique of linear discriminant analysis (LDA; Heilemann, Weihs 2000), a number of modern procedures such as neural networks (NN; Ripley 1994) and classification trees (TREE; Breiman et al. 1984) have been developed (Weihs, Sondhauss 2000). Their main innovation is the way in which they separate the groups (here: phases of the business cycle) in the multidimensional space. The reasons for applying LDA here are, as in most other studies, its robustness, its particularly large analytical possibilities and its clarity due to the linear character of the discriminant functions (Erb 1990: 5). Given the limited space of this paper, it reports only on LDA results.

The main objective of LDA (and, of course, any other classification method) is to classify objects by a set of independent variables $x_1, \ldots, x_m$ into g given groups,

(1) $$y_i = c_1 x_1 + \ldots + c_m x_m$$

where
y_t: dependent (grouping) variable, with $i = 1, \ldots, g$ (number of groups with $g \geq 2$);
x_j: independent variables, $j = 1, \ldots, m$;
c_j: coefficients.

For n cases, the observations $x_1, \ldots, x_m$ of the m-dimensional criterion are given. The observations of the (n,m)-matrix

(2) $$x = \begin{bmatrix} x_1^{'} \\ \cdot \\ \cdot \\ \cdot \\ x_n^{'} \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

arise from g different groups or classes, and so x can be partitioned into $g(n_k, m)$-submatrices (with $n = n_1 + \ldots + n_g$)

(3) $$x = \begin{bmatrix} x_1 \\ \cdot \\ x_k \\ \cdot \\ x_g \end{bmatrix}$$

with $x_k = (x_{k1}, \ldots, x_{ki}, \ldots, x_{kn_k})'$ containing the observations from group $G_k (k = 1, \ldots, g)$.

In the simple case of two groups, (3) reduces to

(4) $$x_1^{'} = (x_{11}, \ldots, x_{1n_1}) \text{ and } x_2^{'} = (x_{21}, \ldots, x_{2n_2}).$$

By a linear transformation of the m-dimensional vector of observations x to a scalar, the m-dimensional problem becomes a 1-dimensional one:

(5) $$y_i = c_1 x_1 + c_2 x_2 .$$

In LDA, the coefficients (c_j) are estimated in such a way that the values of the discriminant function (5) differ as much as possible *between* the groups, or so that for the discriminant scores the ratio

$$\frac{\text{between-groups sum of squares}}{\text{within-groups sum of squares}} \tag{6}$$

is maximized.

In the general case of $g \geq 2$ groups, a maximum number of $\min(m, g-1)$ discriminant functions can be derived. The first function has the largest ratio of between-groups to within-groups sums of squares. The second function is orthogonal to the first and has the next largest ratio, and so on. Because the coefficients of the different discriminant functions are derived from a classic intrinsic value problem (Erb 1990: 36), special norming conditions have to be set up to achieve unique solutions.

The main questions about classification scheme being asked and answered by LDA are:

- How well do the variables discriminate between given groups?
- Which variables are good discriminators?
- What decision rule should be used for classifying (new) objects?

Table 6

Misclassifications
1948–5 to 1973–9

Year	Month	Phase	M/W	H/M
1948	May	Stagflation	Demand Pull	
	June	Stagflation	Demand Pull	
	November	Stagflation	Demand Pull	Recession
1949	October	Recession		Recovery
	November	Recovery		Recession
	December	Recovery		Recession
1950	June	Recovery		Demand Pull
1951	January	Stagflation		Demand Pull
1953	November	Recession	Stagflation	Stagflation
1954	August	Recovery		Recession
	October	Recovery		Recession
	November	Recovery		Demand Pull
1955	February	Recovery		Demand Pull
	April	Demand Pull		Recovery
1956	February	Demand Pull	Stagflation	
	June	Demand Pull		Recession
	October	Demand Pull	Stagflation	
1957	September	Recession		Demand Pull
	October	Recession		Demand Pull
1958	May	Recovery	Recession	Recession
	June	Recovery		Recession
1959	July	Recovery	Recession	
	August	Recovery	Recession	
1960	March	Recovery	Recession	
	April	Recovery		Recession
	May	Recovery	Recession	Recession
	August	Recession	Recovery	
1961	Jan.	Recession		Recovery
1962	July	Recovery	Recession	
	October	Recovery	Recession	
1964	July	Recovery	Demand Pull	
1965	December	Demand Pull	Stagflation	
1966	February	Demand Pull	Stagflation	
	March	Demand Pull		Stagflation
	April	Demand Pull		Stagflation
	September	Demand Pull	Stagflation	Stagflation
	October	Demand Pull	Stagflation	Stagflation
	November	Demand Pull	Stagflation	Stagflation
	December	Demand Pull		Stagflation
1967	February	Demand Pull	Stagflation	
	March	Demand Pull	Stagflation	
	May	Demand Pull	Stagflation	
	July	Demand Pull	Stagflation	
	December	Stagflation		Demand Pull
1968	February	Stagflation		Demand Pull
	December	Stagflation	Demand Pull	
1969	December	Stagflation	Demand Pull	
1970	January	Recession	Demand Pull	Stagflation
	April	Recession	Demand Pull	
	May	Recession	Demand Pull	
	July	Recession	Recovery	
1973	May	Demand Pull	Recovery	

Meyer/Weinberg's and authors' computations.

Table 7

Leave-one-out Classifications[1]
1948–5 to 2000–12

Actual group	Predicted group membership[2]				
	No. of cases	Recovery	Demand-Pull	Stagflation	Recession
			All		
Recovery	285	207 (72.6)	54 (18.9)	0 (0.0)	24 (8.4)
Demand-Pull	149	20 (13.4)	100 (67.1)	22 (14.8)	7 (4.7)
Stagflation	102	3 (2.9)	16 (15.7)	80 (78.4)	3 (2.9)
Recession	96	2 (2.1)	0 (0.0)	7 (7.3)	87 (90.6)
Total error rate: 25.0%					
			“Leave one out”		
Recovery	285	202 (70.9)	57 (20.0)	0 (0.0)	26 (9.1)
Demand-Pull	149	24 (16.1)	95 (63.8)	22 (14.8)	8 (5.4)
Stagflation	102	3 (2.9)	18 (17.6)	77 (75.5)	4 (3.9)
Recession	96	3 (3.1)	0 (0.0)	7 (7.3)	86 (89.6)
Total error rate: 27.1%					
			Without cycle 1 (1948–5 to 1949–10)[2]		
Recovery	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Demand-Pull	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Stagflation	7	0 (0.0)	0 (0.0)	7 (100.0)	0 (0.0)
Recession	11	0 (0.0)	0 (0.0)	1 (9.1)	10 (90.9)
Total error rate: 5.6%					
			Without cycle 2 (1949–11 to 1954–7)[2]		
Recovery	8	6 (75.0)	0 (0.0)	0 (0.0)	2 (25.0)
Demand-Pull	6	3 (50.0)	(50.0)3	0 (0.0)	0 (0.0)
Stagflation	34	0 (0.0)	28 (82.4)	6 (17.6)	0 (0.0)
Recession	9	0 (0.0)	0 (0.0)	0 (0.0)	9 (100.0)
Total error rate: 57.9%					
			Without cycle 3 (1954–8 to 1958–4)[2]		
Recovery	7	4 (57.1)	0 (0.0)	0 (0.0)	3 (42.9)
Demand-Pull	30	9 (30.0)	1 (3.3)	15 (50.0)	5 (16.7)
Stagflation	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recession	8	0 (0.0)	0 (0.0)	1 (12.5)	7 (87.5)
Total error rate: 73.3%					
			Without cycle 4 (1958–5 to 1961–1)[2]		
Recovery	25	19 (76.0)	0 (0.0)	0 (0.0)	6 (24.0)
Demand-Pull	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Stagflation	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recession	8	2 (25.0)	0 (0.0)	0 (0.0)	6 (75.0)
Total error rate: 24.2%					

Table 7 cont.

Leave-one-out Classifications[1]
1948–5 to 2000–12

Actual group	Predicted group membership[2]				
	No. of cases	Recovery	Demand-Pull	Stagflation	Recession
		Without cycle 5 (1961–2 to 1970–11)[3]			
Recovery	51	39 (76.5)	10 (19.6)	0 (0.0)	2 (3.9)
Demand-Pull	31	0 (0.0)	13 (41.9)	18 (58.1)	0 (0.0)
Stagflation	25	0 (0.0)	14 (56.0)	11 (44.0)	0 (0.0)
Recession	11	5 (45.5)	1 9.1)	5 (45.5)	0 (0.0)
Total error rate: 46.6%					
		Without cycle 6 (1970–12 to 1975–3)[3]			
Recovery	25	1 (4.0)	22 (88.0)	0 (0.0)	2 (8.0)
Demand-Pull	21	0 (0.0)	7 (33.3)	9 (42.9)	5 (23.8)
Stagflation	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recession	6	0 (0.0)	0 (0.0)	0 (0.0)	6 (100.0)
Total error rate: 73.1%					
		Without cycle 7 (1975–4 to 1980–9)[3]			
Recovery	39	9 (23.1)	22 (56.4)	0 (0.0)	8 (20.5)
Demand-Pull	12	0 (0.0)	3 (25.0)	9 (75.0)	0 (0.0)
Stagflation	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recession	15	0 (0.0)	0 (0.0)	11 (73.3)	4 (26.7)
Total error rate: 75.8%					
		Without cycle 8 (1980–10 to 1982–12)[3]			
Recovery	6	0 (0.0)	0 (0.0)	0 (0.0)	6 (100.0)
Demand-Pull	6	4 (66.7)	0 (0.0)	0 (0.0)	2 (33.3)
Stagflation	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recession	15	3 (20.0)	0 (0.0)	0 (0.0)	12 (80.0)
Total error rate: 55.6%					
		Without cycle 9 (1983–1 to 1991–12)[2]			
Recovery	16	16 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Demand-Pull	43	43 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Stagflation	36	36 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recession	13	6 (46.2)	0 (0.0)	0 (0.0)	7 (53.8)
Total error rate: 78.7%					
		Without incomplete cycle 10 (1992–1 to 2000–12)[3]			
Recovery	108	2 (1.9)	28 (25.8)	75 (69.4)	3 (2.8)
Demand-Pull	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Stagflation	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Recession	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Total error rate: 98.1%					

Authors' computations. – [1]Successive elimination of one month from the sample period. – [2]Share in % in parenthesis. – [3]Sample period without corresponding cycle.

Claus Weihs and Ursula Garczarek

Stability of Multivariate Representation of Business Cycles over Time

1. Introduction

In order to replace the univariate indicators standard in the literature (Oppenländer 1996) 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 (Weiss, Kulikowski 1991) 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 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" (Lucas 1987) 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 (Heilemann, Münch 1996) 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 $X_{k,t}, k=1,\dots,K, t=0,\dots,T$ 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 B of the K-dimensional real space $B \in \mathbb{R}^K$. Based on some preclassified 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 $t=t_0+h, h=1,\dots,H$, based on all observations $\vec{x}_{t_0}, \vec{x}_{t_0+1},\dots,\vec{x}_t$ 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 $X_{k,t}, k=1,\dots,K$, 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 $g \in \{1,\dots,G\}$ 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 (MEC1; Weihs et al. 1999) 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 Sondhauss/Weihs (1999). CRAKE represents a certain markov regime switching model (Krolzig 1997).

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 Koskinen/Öller (1998; see Sondhauss, Weihs 2001 for details).

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.[1]

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 (Weiss, Kulikowski 1991). 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

[1] For a more detailed discussion of borders corresponding to classification rules of various classification methods see, e.g., Weihs et al. 1993.

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" concerning theoretical predictive power according to certain distributional assumptions is basic to LDA, QDA, and CRAKE. 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. MEC1 additionally finds a "best" rule with respect to a leave-one- observation out error. 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-

Table 1

LDA-MEC1-2D's Directions ($*10^3$) of Best Linear Combinations on Cycle 2

	Dir1	Dir2	Dir1	Dir2
			Standardized	
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

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 Röhl (1998) and Weihs/Röhl/Theis (1999). 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.

Note that one has to standardize these weights by the variables' standard deviations (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

Table 2

LDA's Prediction Errors

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 3

LDA's Mean Training 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 4

LDA's Differences of Errors

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

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 (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% (Table 2).

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 (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 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

Table 5

QDA-MEC1-2D's Directions (*10^3) of Best Linear Combinations on Cycle 5

	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 6

QDA's Prediction Errors

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 7

QDA's Mean Training Errors[1]

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

[1]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.

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

Table 8

QDA's Differences of Errors

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!!

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 Krolzig (1997), with a certain covariance structure. More precisely, for given phase $s_t \in \{1, \ldots, S\}$ the vector $\vec{x}_t \in IR^K$ is generated by a first-order vector autoregressive model with diagonal covariance matrices, such that for each k, $k = 1, \ldots, K$, 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 $u_{k,t}, k = 1, \ldots, K, t = 1, \ldots, T$, being independently normally distributed, $u_{k,t} \sim N(0, \sigma_k)$ given $s_t, t = 1, \ldots, T$. 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).

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 (Tables 2 and 9).

The corresponding CRAKE model looks as follows:

$$\begin{aligned}
LC_t &= 0.81 + 0.63 LC_{t-1}, & L_t &= 0.44 + 0.82 L_{t-1} & &\text{in phase 1,}\\
LC_t &= 2.86 + 0.21 LC_{t-1}, & L_t &= 0.99 + 0.59 L_{t-1} & &\text{in phase 2,}\\
LC_t &= 1.21 + 0.80 LC_{t-1}, & L_t &= -0.27 + 0.93 L_{t-1} & &\text{in phase 3,}\\
LC_t &= 1.14 + 0.96 LC_{t-1}, & L_t &= -0.48 + 0.78 L_{t-1} & &\text{in phase 4.}
\end{aligned}$$

Table 9

RAKE's Prediction Errors

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 10

CRAKE's Mean Training 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 11

CRAKE's Differences of Errors

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

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:

Table 12

Comparing LDA-LC,L with Best Models According to Prediction Error

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-MEC1-2D
0.21	0.21	LDA-LC,L

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 (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 (Tables 2, 6 and 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 (Table 12). Obviously, LDA-LC,L is only clearly suboptimal for cycles 2 and 4. Even better, best models for these cycles 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 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 labor costs.

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

Heilemann, U. and H.J. Münch (1996), West German Business Cycles 1963–1994: A Multivariate Discriminant Analysis. In Selected Papers Submitted to the CIRET-Conference in Singapore. CIRET-Studien 50. Ifo, München, 219–250.

Koskinen, L. and L.-E. Öller (1998), A hidden markov model as a dynamic Bayesian classifier, with an application to forecasting business-cycle turning points. NIER Working Paper 59. National Institute of Economic Research, Stockholm.

Krolzig, H.-M. (1997), *Markov-Switching Vector Autorregressions.* Modelling, Statistical Inference and Application to Business Cycle Analysis. Berlin et al.: Springer.

Lucas, R.E. (1987), *Models of business cycles.* Oxford: Basil Blackwell.

Oppenländer, K.H. (1996), *Konjunkturindikatoren*. 2nd ed. München: Oldenbourg.

Röhl, M.C. (1998), *Computerintensive Dimensionsreduktion in der Klasifikation*. Lohmar: Eul.

Sondhauss, U.M. and C. Weihs (1999), Dynamic bayesian networks for classification of business cycles. SFB 475 Technical report 17/99. University of Dortmund.

Sondhauss, U.M. and C. Weihs (2001), Incorporating background knowledge for better prediction of cycle phases. SFB 475 Technical report 24/01. University of Dortmund.

Weihs, C., W. Baumeister and H. Schmidli (1993), Classification methods for multivariate quality parameters. *Journal of Chemometrics* 7: 131–142.

Weihs, C., M.C. Röhl and W. Theis (1999), Multivariate classification of business phases. SFB 475 Technical report 26/99. University of Dortmund.

Weiss, S.M. and C.A. Kulikowski (1991), *Compter Systems that Learn*. San Francisco, CA: Morgan Kaufmann.

Marlene Amstad und Bernd Schips

Wachstumsfluktuationen, Zykluskonzepte und konjunkturelle Wendepunkte[1]

1. Einleitung

Der Begriff „Konjunktur“ ist bereits recht alt und steht im Allgemeinen für die Beobachtung des gesamtwirtschaftlichen Auf und Ab anhand von verschiedenen Indikatoren. Konjunktur wird also überwiegend als ein erst durch die Beobachtung von mehreren Indikatoren zu erschließendes und somit als latentes Phänomen verstanden. Die empirisch orientierte und auf wirtschaftspolitische Maßnahmen fokussierte Konjunkturforschung bedarf jedoch einer umsetzbaren und allgemein verständlichen Operationalisierung des Phänomens Konjunktur. Die empirische Konjunkturanalyse beschränkt sich deshalb häufig auf die Betrachtung der Entwicklung einer einzelnen Referenzgröße, wie z.B. die Zeitreihe des Bruttoinlandprodukts (BIP) oder der Industrieproduktion. Alternativ bzw. ergänzend zu einem solchen Messkonzept lassen sich auch aus mehreren Indikatoren zusammengesetzte Diffusionsindizes oder andere Sammelindikatoren zur Erfassung der Konjunktur konstruieren.

Konjunkturelle Wendepunkte sind dabei entweder Umkehrpunkte in der Hauptrichtung einer beobachteten (Referenz-)Zeitreihe oder durch das Erreichen „kritischer“ Werte von Diffusionsindizes bzw. von Sammelindikatoren definiert. Die Bestimmung solcher „kritischer“ Werte (Heuristiken) erfolgt in der Regel jedoch nicht ohne „Blick“ auf eine oder mehrere Referenzgrößen. Dabei basiert die Beschreibung der zu beobachtenden Konjunkturschwankungen mit Hilfe typischer (Konjunktur-)Phasen – wie z.B. Erholung, Boom, Abschwächung oder Rezession – auf unterschiedlichen Messkonzepten und Zyklusvorstellungen (z.B. Vosgerau 1978: 478– 507). Die Wende-

[1] Der Artikel basiert auf dem ersten Teil eines im Februar 2002 im RWI Essen gehaltenen Votrags. Der zweite Teil findet sich in Amstad (2002a). Neuere methodische Ansätze und aktuellere Daten sind nicht berücksichtigt.

punktdefinition und damit die Phasenunterteilung hängt somit wesentlich vom jeweils gewählten Mess- und Zykluskonzept ab[2].

Die unterschiedlichen Messkonzepte und Zyklusvorstellungen führen in der Regel auch zu abweichenden Erklärungsansätzen für die zu beobachtenden Konjunkturschwankungen (Tichy 1994; Maussner 1994). Empirische Analysen und darauf aufbauende wirtschaftspolitische Empfehlungen setzen deshalb eine Klärung der zugrundegelegten Konzepte voraus.

2. Die Erfassung des latenten Phänomens „Konjunktur"

Eine theoretisch befriedigende, zugleich leicht und eindeutig zu operationalisierende Definition des Begriffes „Konjunktur" existiert bis heute nicht. Wesentliches Element des Konjunkturphänomens ist und bleibt, dass sich dieses nicht im Verlauf einer Einzelreihe, sondern im Zusammenspiel verschiedener Reihen widerspiegelt. Um sich nicht auf die Beschreibung wesentlicher Charakteristika dieses Phänomens beschränken zu müssen, ist eine Operationalisierung durch geeignete – dem latenten Charakter dieses Phänomens gerecht werdende – Messkonzepte erforderlich. Naheliegend ist deshalb die Konstruktion von Diffusionsindizes und anderen Sammelindikatoren. Hierfür wiederum ist jedoch die Wahl entsprechend geeigneter einzelner Referenzgrößen von zentraler Bedeutung. Größen wie das BIP, die Industrieproduktion oder die Auslastung des Produktionspotenzials (*output gap*) stehen hier im Vordergrund. Doch auch bei der Verwendung dieser Einzelreihen als indirekte Repräsentanten der Konjunktur ergeben sich vielfältige Probleme.

Nachfolgend werden lediglich zwei, dafür aber sehr grundlegende Aspekte dieser Problematik näher erläutert: zum einen die Situation bezüglich der verzögerten Verfügbarkeit von Informationen und zum anderen die Saisonbereinigungsproblematik.

2.1 Bemerkungen zur Datenlage

Zur Vorbereitung einer Prognose ist es notwendig, möglichst früh über die aktuelle Verfassung einer Volkswirtschaft verlässliche Informationen zu erhalten. In der Regel stehen dabei mehrere Datenquellen zur Verfügung.

Zum einen Daten der offiziellen Statistik. Hier handelt es sich meist um quantitative Daten, die eine unterschiedliche – und oft eine für die Konjunkturbeobachtung unzulängliche – Periodizität aufweisen. Die Daten der Volkswirtschaftlichen Gesamtrechnung (VGR) – allen voran die BIP-Reihe – leisten unverzichtbare Dienste zur Messung der Konjunktur. Als beachtlicher Nach-

[2] Eine detaillierte vergleichende Darstellung zu unterschiedlichen Zyklusvorstellungen und Messkonzepten findet sich in Amstad (2000a, 2000b) und der dort angegebenen Literatur.

teil ist aber zu werten, dass sie sowohl einem Revisionsbedarf wie auch einem Publikationslag unterliegen. So lagen etwa im Februar 2002 für die Schweiz definitive Quartalswerte für das BIP erst bis Ende 1998 und vorläufige Schätzungen bis zum dritten Quartal 2001 vor.

Zum anderen stehen Daten von *business and consumer surveys* zur Verfügung. Die Resultate derartiger Umfragen sind überwiegend qualitativer Natur. Die monatlich, quartalsweise oder jährlich stattfindenden Befragungen bei „repräsentativ" ausgewählten Unternehmen oder Konsumenten über die gegenwärtige sowie über die in Zukunft von den Umfrageteilnehmern erwartete wirtschaftliche Lage geben eine Fülle von Informationen. Bei den Unternehmensumfragen ist die Stichprobe dabei meist ein rotierendes Panel, in dem vor allem die großen Unternehmen ständig enthalten sind und ein gewisser Anteil in meist unregelmäßiger Folge ausgetauscht wird. Ausgehend von Umfragen in Japan (*Tankan*) fand diese Methode inzwischen weltweit Anwendung. Einen internationalen Überblick der Institutionen, die solche Umfragen durchführen, gibt das CIRET Office (www.ciret.org).

In der Schweiz führte die KOF solche Umfragen ein. Heute werden die Industrie, Banken und Versicherungen, der Großhandel, das Gastgewerbe, die Bauwirtschaft und der Projektierungssektor (Architektur- und Ingenieurbüros), der Detailhandel sowie das Konsumkreditgewerbe regelmäßig befragt (Tabelle 1). Angesichts der erst ab 1980 verfügbaren Daten der VGR (nach ESVG 78) ist auch die Dauer einiger dieser Erhebungen ein wichtiges Argument, um diesen im Rahmen der Konjunkturanalyse mehr Beachtung zu schenken.

Die geschichteten Stichproben berücksichtigen die Verteilung nach Kantonen, Sprachregionen, Unternehmensgröße sowie Exportorientierung. Die Fragen umfassen vor allem die Veränderungen von Produktion, Umsatz, Lagerbeständen, Verkaufspreisen sowie die Beurteilung dieser Größen, Zukunftspläne und Erwartungen. Einen Überblick über die Vielfalt der Fragen geht aus dem im Anhang als Beispiel angefügten Fragebogen der Monatsumfrage in der Industrie hervor.

Tabelle 1

Branchen, Periodizität und Beginn der Umfragen der KOF

Branche	Periodizität	Umfragebeginn
Industrie	Monatlich und quartalsweise	1955
Konsumkredite	Quartalsweise	1971
Detailhandel	Monatlich und quartalsweise	1973
Großhandel	Quartalsweise	1977
Bastgewerbe	Quartalsweise	1988
Baugewerbe	Quartalsweise	1994
Projektierungssektor	Quartalsweise	1996
Banken/Versicherungen	Quartalsweise	2000

Konjunkturumfragen weisen viele spezifische Vorteile auf, die mittlerweile international zunehmend Beachtung finden. Zwei für die praktische Konjunkturanalyse und -prognose besonders gewichtige Vorzüge sind die Schnelligkeit, mit der Resultate vorliegen, und der Umstand, dass diese Daten keinem Revisionsbedarf unterliegen. So lagen z.B. zu Beginn des Februar 2002 bereits definitive Umfrageergebnisse bis Ende 2001 vor. Ein weiterer großer Vorteil ist der Umstand, dass Umfragedaten meist auch vorausschauende Angaben über erwartete Entwicklungen beinhalten.

2.2 Bemerkungen zur Saisonbereinigung

Bezüglich des Revisionsbedarfs bei Umfragedaten ist einschränkend anzufügen, dass eine Beantwortung der gestellten Fragen eine Ausblendung saisonaler Einflüsse durch die Antwortenden voraussetzt. Die Erfahrung zeigt aber, dass in den seltensten Fällen eine Abstraktion von saisonalen Effekten bei der Beurteilung durch die Befragten gelingt. Dies hat zur Konsequenz, dass auch diese Daten saisonbereinigt werden müssen, um Aufschluss über die eigentlich interessierende konjunkturelle Lage zu gewinnen.

Bei der Wahl des Saisonbereinigungsverfahrens ist jedoch der *trade-off* zwischen Randstabilität und Phasenverschiebung zu berücksichtigen. Saisonbereinigungsverfahren, die einen rekursiven Filter verwenden, sind zwar randstabil, haben aber mit dem Problem einer mehr oder weniger großen Phasenverschiebung zwischen der beobachteten und der gefilterten Reihe zu kämpfen. Die gebräuchlichen Saisonbereinigungsverfahren (wie etwa das X11- resp. das X12-Verfahren) handhaben das Problem der Phasenverschiebung am aktuellen Rand aber meist recht gut – wenn auch oft zu Lasten der Amplitude. Die aus konjunktureller Sicht besonders interessierenden Werte am aktuellen Rand unterliegen in der Regel deshalb deutlichen Revisionen, sobald weitere Beobachtungswerte vorliegen.

3. Alternative Zykluskonzepte zur Erfassung von Wachstumsfluktuationen

3.1 Das klassische Zykluskonzept

Das *klassische business cycle*-Konzept (*classical cycle*-Konzept) basiert auf den Schwankungen im absoluten Niveau der gesamtwirtschaftlichen Aktivitäten. Zur vereinfachten Operationalisierung wird als Messgröße meist das reale BIP verwendet. Dadurch wird das Phänomen Konjunktur lediglich auf eine Outputgröße reduziert. Für die Datierung von Wendepunkten und damit der Konjunkturphasen sind insbesondere die Vorzeichen der Zuwachsraten (Vorjahres- resp. Vorperiodenvergleich) der gewählten Referenzgröße bzw. die kritischen Werte von Diffusionsindizes oder anderen Sammelindikatoren maßgebend. Die oberen (*peaks*) und die unteren Wendepunkte (*troughs*) werden dabei durch die Nullstellen im Verlauf der Zuwachsraten bestimmt. Eine Boomphase entspricht dabei dem Zeitraum mit positiven, eine Rezessionsphase der Zeitspanne mit negativen Veränderungsraten.

Bei dieser sehr einfachen Regel zur Zykluseinteilung dürfen jedoch die spezifischen Einschränkungen je nach Berechnungsweise der Zuwachsraten nicht vergessen werden. So dient etwa die Berechnung von Vorjahreswachstumsraten als einfaches und – von revisionsbedingten Änderungen mal abgesehen – randstabiles Saisonbereinigungsverfahren. Dieses Verfahren eliminiert die Saisonschwankungen aber oft nur ungenügend, so dass ein Teil der Saisonkomponente der Trend-Zyklus-Komponente zugerechnet wird.

Die Berechnung der (meist auf Jahreswerte hochgerechneten) Wachstumsraten gegenüber der Vorperiode basiert dagegen auf bereits saisonbereinigten Daten. Hier wirkt sich einschränkend aus, dass die meisten Saisonbereinigungsverfahren keine Randstabilität aufweisen, was regelmäßig zu Revisionen der Vorperiodenwachstumsraten führt. Davon abgesehen ist – v.a. für eine bezüglich Saisonmuster so heterogen zusammengesetzten Größe wie das BIP – auch nicht garantiert, dass Saison- und Trend-Zykluskomponente sauber getrennt werden.

3.2 Das Wachstumszykluskonzept

Die an der Wachstumsrate orientierten klassischen Konjunkturzyklen vernachlässigen den Umstand, dass das gesamtwirtschaftliche Wachstum meistens mehr oder weniger trendbehaftet ist. Dadurch treten die für klassische Zyklen charakteristischen Phasen mit negativen Veränderungsraten sehr selten auf. Das *Wachstumszyklus*konzept (*growth cycle*-Konzept) entspricht einem den Trend berücksichtigenden Zykluskonzept. Wachstumszyklen orientieren sich nicht an Schwankungen in der gesamtwirtschaftlichen Produktion, sondern meist an der Auslastung des gesamtwirtschaftlichen Produktionspotenzials. Der obere Wendepunkt in einem Wachstumszyklus wird durch den Zeitpunkt mit einer maximalen positiven, der untere Wendepunkt durch den Zeitpunkt mit einer maximal negativen Trendabweichung bestimmt. Eine „Boomphase“ ist eine Zeitperiode mit zunehmenden und eine Rezessionsphase ein Zeitraum mit abnehmenden Abweichungen vom durch den „Trend“ bestimmten Wachstumspfad. Damit stimmen die Wendepunkte eines klassischen Konjunkturzyklus nur dann mit den Wendepunkten nach dem Wachstumszykluskonzept überein, wenn die als Referenzgrösse verwendete Zeitreihe keinen Trend aufweist.

Bei einem trendbehafteten gesamtwirtschaftlichem Wachstum dauern Boomphasen im klassischen Konjunkturzyklus länger als die entsprechenden Rezessionsphasen. Daraus ergibt sich, dass der obere Wendepunkt eines Wachstumszyklus häufig vor dem oberen Wendepunkt des entsprechenden

DOI https://doi.org/10.3790/978-3-428-52425-9 | Generated on 2025-02-22 16:18:59

klassischen Konjunkturzyklus liegt. Und im Gegensatz dazu beginnen Aufschwungsphasen nach dem Wachstumszykluskonzept oft zeitlich später als der untere Wendepunkt nach der klassischen Zyklusdefinition. Boomphasen sind in einem Wachstumszyklus daher in der Regel kürzer und Rezessionsphasen länger als in einem entsprechenden klassischen Konjunkturzyklus.

Wichtigstes Charakteristikum für Wachstumszyklen ist, dass sie durch die Abweichungen von einem mit der so genannten „Normalauslastung“ korrespondierenden Wachstumspfad bestimmt werden. Daraus ergibt sich die mit diesem Zykluskonzept verbundene Problematik der Ermittlung der Normalauslastung.

In der empirischen Wirtschaftsforschung stehen drei unterschiedliche Methoden zur Bestimmung der Normalauslastung im Vordergrund: durch die Schätzung eines Trends für die Zeitreihe der gesamtwirtschaftlichen Produktion, resp. des BIP, durch Schätzung des gesamtwirtschaftlichen Produktionspotenzials und dessen Auslastung oder durch einen bestimmten Wert eines Diffusionsindexes. Alle Methoden können unterschiedlich umgesetzt werden. Diese Vielfalt der Methoden und deren Umsetzung zur Bestimmung der „Normalauslastung“ sind der wesentliche Nachteil des Wachstumszykluskonzepts. Die Messung und die Analyse des Phänomens Konjunktur ist vom Entscheid bezüglich des jeweils zur Anwendung kommenden empirisch-methodischen Vorgehens abhängig.

Um dies zu veranschaulichen, betrachten wir die Probleme näher, die sich bei der Bestimmung der Normalauslastung mittels eines Trends ergeben. Hier stellt sich die „Gretchenfrage“ in Form der Entscheidung über den Charakter des Trends. Im Folgenden wird veranschaulicht, dass dieser Entscheid keine rein technische Frage darstellt. Vielmehr kommt der Entscheid für oder gegen ein bestimmtes Trendbereinigungsverfahren dem methodischen Entscheid über den Charakter und die Ursachen des Konjunkturzyklus gleich.

4. Aspekte verschiedener Trendbereinigungsverfahren

Die Zeitreihen vieler ökonomischer Variablen schwanken offensichtlich nicht um ein konstantes Mittel, sondern weisen einen mehr oder weniger ausgeprägten Trend auf und können deshalb nicht mehr als eine Realisation eines stationären stochastischen Prozesses mit einem zeitunabhängigen Erwartungswert aufgefasst werden. Eine ganze Reihe von statistischen Analyse- und Prognoseverfahren basiert aber auf der Annahme, dass die vorliegenden Daten eine Realisation eines stationären Prozesses sind. Modelle für stationäre Prozesse sind jedoch nicht in der Lage, eine trendmäßige Entwicklung, wie sie für viele ökonomische Zeitreihen charakteristisch ist, zu erfassen.

Eine Klasse bilden die Trend-stationären Prozesse, deren Nicht- Stationarität auf einen deterministischen Zeittrend zurückgeführt wird. Nach Elimination dieses Trends – z.B. durch die Anwendung eines Hodrick-Prescott-Filters (HP-Filter) – ergibt sich wiederum eine Zeitreihe, die als Realisation eines stationären Prozesses aufgefasst werden darf.

Eine andere Klasse bilden die Differenzen-stationären Prozesse. Liegt ein Differenzen-stationärer Prozess vor, dann ist eine entsprechende Differenzenbildung das adäquate Instrument zur Trendbereinigung. Die beiden Prozesstypen unterscheiden sich nun aber in Bezug auf gewisse Prozesseigenschaften ganz wesentlich. Wird etwa bei der Trendelimination irrtümlicherweise von einem deterministischen Zeittrend ausgegangen, dann führt dies zu einer Überbetonung der niederen Frequenzen im Spektrum der Trendabweichungen. Im umgekehrten Fall werden die hohen Frequenzen akzentuiert und die niederen Frequenzen abgeschwächt.

Weiters bleiben im Falle eines Trend-stationären Prozesses Störungen (Zufallsschocks oder Innovationen) auf den Zeitpunkt ihres Auftretens beschränkt, während bei Differenzen-stationären Prozessen solche Störungen noch „unendlich lange“ nachwirken. Es ist daher nicht unerheblich, ob für die langfristige Entwicklung einer für die Konjunkturanalyse herangezogenen Referenzgröße ein deterministischer oder ein stochastischer Trend unterstellt wird. Eine modellgestützte Antwort auf die Frage, ob Rezessionen einen dauerhaften Einfluss auf die zukünftige Entwicklung der gesamtwirtschaftlichen Aktivität haben oder ob temporäre Wachstumsverluste später wieder „wettgemacht“ werden können, hängt also entscheidend von der Wahl des „richtigen“ Prozesstyps bei der Modellierung ab.

Die für eine Diskriminierung der beiden Prozesstypen in der Regel verwendeten Ergebnisse von so genannten Einheitswurzel-Tests (Unit-root-Tests) sind jedoch nicht unproblematisch. Die Macht dieser Tests gegenüber lokalen Alternativhypothesen, d.h. gegenüber Einheitswurzeln, die relativ nahe bei Eins liegen, ist gering. Ein weiterer Einwand basiert darauf, dass die Macht dieser Tests auch gegenüber anderen als bei der Testkonstruktion unterstellten Trend-stationären Modellen ebenfalls meist ziemlich gering ist (Stier 2001: 286–313).

Die grundsätzlichen Anforderungen an die Trendbereinigungsverfahren sind durchaus mit den Kriterien für Saisonbereinigungsverfahren vergleichbar. Es geht auch hier darum, einen Mittelweg zwischen sich gegenseitig konkurrenzierenden Zielsetzungen – Erhaltung der Amplitude, Verhindern von Phasenverschiebungen und Erreichen von Randstabilität – zu finden. Nicht immer wird aber den spezifischen Eigenschaften der Verfahren zur Elimination von Trends gebührend Rechnung getragen. So hat z.B. der seit längerem besonders beliebte HP-Filter die Eigenschaft, dass die gefilterten Werte am aktuellen Rand konstruktionsbedingt nahe bei den Beobachtungswerten liegen. Die Trendabweichungen fallen dadurch am aktuellen Rand tendenziell kleiner aus als in einer ex post-Analyse, so dass die so bestimmten Konjunkturschwankungen in der Regel unterschätzt werden.

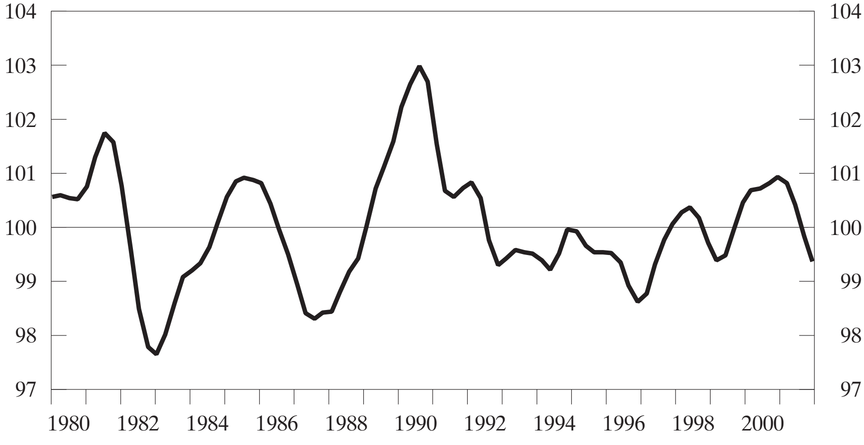
Die – hier lediglich angedeutete – Vielfalt der Methoden zur Trendbereinigung und damit zur Bestimmung der (Wachstums-)Fluktuationen findet ihren Niederschlag in unterschiedlichen Einschätzungen der konjunkturellen Entwicklung. Dieser Umstand wird nachfolgend anhand der aktuellen konjunkturellen Situation in der Schweiz dargestellt.

5. Beispiele zur unterschiedlichen Wendepunktdatierung

In der Einleitung wurde betont, dass der Begriff Konjunktur immer für ein Phänomen steht, das sich aus dem Zusammenspiel einer Vielzahl einzelwirtschaftlicher Aktivitäten ergibt. Für die folgende Darstellung kommen als Referenzgröße für die zu analysierende gesamtwirtschaftliche Entwicklung die Quartalswerte für das BIP der Schweiz gemäß den Schätzungen des *seco* zur Anwendung, die dann nachträglich, d.h. nach Vorliegen der Jahreswerte des Bundesamtes für Statistik (BFS), revidiert werden. Wie ebenfalls bereits betont wurde, erschwert der erfahrungsgemäß erhebliche Revisionsbedarf dieser Reihe die rechtzeitige Datierung von Wendepunkten wesentlich.

In Schaubild 1 wird die Abweichung des realen saisonbereinigten Schweizer BIP von seinem HP-Trend dargestellt. Aufgrund dieser Darstellung lässt sich

Schaubild 1
Abweichung des realen BIP vom Hodrick-Prescott-Trend
1980 bis 2001; Hodrick-Prescott-Trend = 100

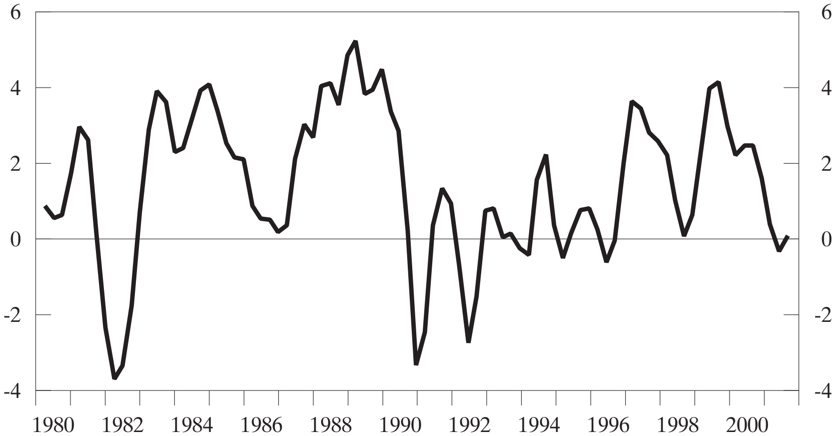


Hier und in allen folgenden Schaubildern: Ab 1999 provisorische Daten.

Schaubild 2

Vorperiodenveränderung des realen BIP, auf Jahresrate hochgerechnet

1980 bis 2001; in %

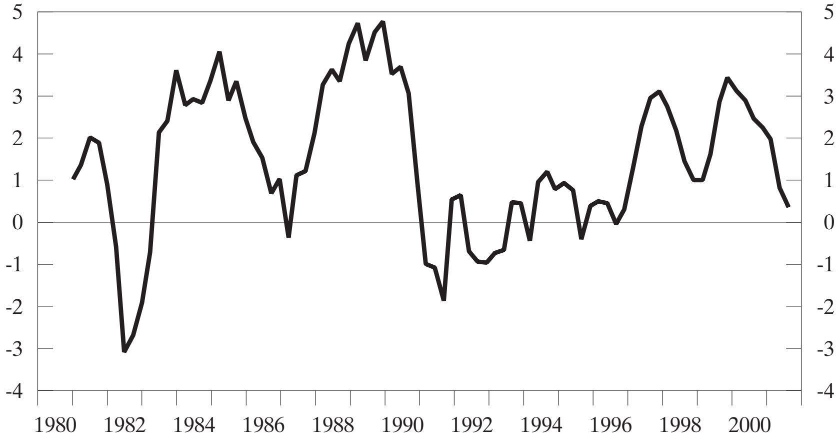


annehmen, dass sich die Schweizer Konjunktur seit Mitte 2001 unter dem Trendwachstum entwickelt und sich bis Ende 2001 keine Trendumkehr andeutet. Die relative HP-Trendabweichung liegt Ende 2001 auf dem vergleichsweise tiefen Niveau von Ende 1998. Diese Einschätzung ist aber in zweifacher Hinsicht mit Vorsicht zu interpretieren, da der HP-Filter nicht randstabil ist und die BIP-Werte erst provisorischer Natur sind.

Dies verdeutlicht sich, wenn anstelle des HP-Filters ein Differenzen-Filter zur Trendbestimmung verwendet wird. Schaubild 2 stellt die auf Jahreswerte hochgerechneten Veränderungsraten gegenüber der Vorperiode dar. Hier wird am aktuellen Rand ein ebenfalls Mitte 2001 unter den Trend fallendes Wachstum angezeigt. Im Gegensatz zur HP-Filterung verbleibt hier lediglich ein Quartal unter Trend und weist bereits im 4. Quartal 2001 wieder auf eine der Normalauslastung entsprechende Wirtschaftstätigkeit hin.

Die Gangart der Schweizer Konjunktur zu Jahresbeginn 2002 stellt sich nochmal leicht anders dar, wenn die Veränderungsraten gegenüber der jeweiligen Vorjahresperiode betrachtet werden (Schaubild 3). Der Vorjahresvergleich dient hier als Saisonbereinigungsverfahren. Hier zeigt sich der Eindruck eines seit Anfang 2000 sich verlangsamenden gesamtwirtschaftlichen Wachstums. Die Wachstumsraten nähern sich ab Mitte 2001 beschleunigt der Nulllinie, bleiben aber auch Ende 2001 – wenn auch vergleichsweise knapp – weiterhin im positiven Bereich.

Schaubild 3
Reales BIP: Veränderung gegenüber der Vorjahresperiode
1980 bis 2001; in %



6. Zykluskonzept und Erklärungsansätze

Nach der Darstellung der wichtigsten Aspekte in der Problematik der Konjunkturmessung und der Veranschaulichung an der aktuellen Konjunktursituation wird nun abschließend noch kurz auf das Zusammenspiel von Zykluskonzept und Erklärungsansätzen eingegangen. Es gilt zu beachten, dass unterschiedliche Zykluskonzepte in der Regel auch zu unterschiedlichen Erklärungsansätzen und Modellierungen führen.

Das *classical cycle*-Konzept korrespondiert mit traditionellen Modellen zur Erklärung von Wachstumsfluktuationen. Daher steht dieser Zyklustyp bei strukturellen ökonometrischen Modellen zur Analyse und Prognose der kurz- bis mittelfristigen Entwicklung im Vordergrund. Das am Trend orientierte *growth cycle*-Konzept findet dagegen v.a. in Real business cycle (RBC)- oder vektorautoregressiven Modellen (VAR) Anwendung. Dabei setzen RBC-Modelle einen deterministischen Trend voraus, während in VAR-Modellen in der Regel von stochastischen Trends ausgegangen wird, die sich durch Differenzenbildung eliminieren lassen. Abweichungen vom deterministischen resp. stochastischen Trend haben dabei die bereits skizzierten unterschiedlichen Auswirkungen auf die weitere Entwicklung der betrachteten Größen.

Im Sprachgebrauch – auch unter Ökonomen – wird häufig nicht sorgfältig genug zwischen den alternativen Konzepten und zugehörigen Erklärungsansätzen unterschieden. So wird beispielsweise von einer Rezession im Sinne eines *classical cycle*-Konzepts gesprochen und gleichzeitig für die Erklärung der Wachstumsfluktuationen ein RBC-Modell verwendet.

7. Fazit

Der Begriff Konjunktur bezeichnet ein latentes Phänomen, das sich aus der Beobachtung verschiedener Reihen ergibt. Die Operationalisierung dieses Phänomens in der empirischen Wirtschaftsforschung bedarf einer konkreten Vorstellung über den Zyklustyp, die Wendepunkte und den sich daraus ergebenden Konjunkturphasen. Leider wird in der theoretischen Diskussion wie auch in der empirischen Analyse oft nur von einer ganz bestimmten, oft nicht einmal explizit gemachten Vorstellung vom Zyklus ausgegangen. Über Vor- und Nachteile verschiedener Zykluskonzepte sollte man sich aber im Klaren sein und unter Umständen auch alternative Konzepte in die Analyse mit einbeziehen. So gilt es wie gezeigt zu beachten, dass vor allem die trendbasierten Ansätze zur Bestimmung der Konjunkturphasen immer nur vorläufigen Charakter haben. Im Weiteren sollten Konjunkturanalysen möglichst auf definitiven und mit der gewählten Referenzgröße hoch korrelierten Daten basieren. Hier empfiehlt sich die Verwendung von Umfrageergebnissen, die v.a. bezüglich Revisionsbedarf und rascher Verfügbarkeit klare Vorteile gegenüber den quantitativen Reihen aufweisen.

Literatur

Amstad, M. (2000a), *Chronologie unterschiedlicher Wendepunkttypen und Entwicklung eines Frühwarnsystems mittels Markov-Switching-Modellierung von Schweizer Unternehmensumfragedaten*. Dissertation, Universität St. Gallen.

Amstad, M. (2000b), Chronologie konjunktureller Wendepunkte der Schweiz- Methoden und Resultate. Arbeitspapiere der Konjunkturforschungsstelle der ETH Zürich 52. KOF, Zürich.

Maussner, A. (1994), *Konjunkturtheorie*. Berlin et al.: Springer.

Stier, W. (2001), *Methoden der Zeitreihenanalyse*. Berlin et al.: Springer.

Tichy, G. (1994), *Konjunktur, stilisierte Fakten, Theorie, Prognose*. Berlin et al.: Springer.

Vosgerau, H.J. (1978), Konjunkturtheorie. In W. Albers et al. (Hrsg), *Handwörterbuch der Wirtschaftswissenschaft*. Stuttgart et al.: Fischer et al., 478–507.

Anhang: Fragebogen der Monatsumfrage in der Industrie

Konjunkturumfrage

Industrie: Monatsfragen

Konjunkturforschungsstelle der ETH
ETH-Zentrum, 8092 Zürich
E-Mail: kof@kof.gess.ethz.ch
Telefon 01 632 42 39
Telefax 01 632 12 18
Internet: http://www.kof.ethz.ch

Bitte beachten:
- Nur für die auf dem Fragebogen angegebene Produktegruppe antworten.
- Zutreffendes Feld ankreuzen. Bei Unklarheiten bitten wir Sie, die "Erläuterungen" auf der Rückseite beizuziehen.
- Nur mengenmässige Veränderungen berücksichtigen (rein preisbedingte Veränderungen sind auszuschalten).
- Fragebogen bis zum **10. des Monats** zurücksenden.

Rückblick und Beurteilung

1. Bestellungseingang

a) Er war im Vergleich zum Vormonat *

☐ höher ☐ gleich ☐ niedriger

b) Er war im Vergleich zum Vorjahresmonat

☐ höher ☐ gleich ☐ niedriger

2. Auftragsbestand ○ kein Auftragsbestand

a) Er war im Vergleich zum Vormonat *

☐ höher ☐ gleich ☐ niedriger

b) Wir beurteilen den Auftragsbestand **insgesamt** als

☐ gross ☐ normal ☐ zu klein

c) Wir beurteilen den Bestand an **Auslandsaufträgen** als ○ kein Export

☐ gross ☐ normal ☐ zu klein

3. Produktion

a) Sie war im Vergleich zum Vormonat *

☐ höher ☐ gleich ☐ niedriger

b) Sie war im Vergleich zum Vorjahresmonat

☐ höher ☐ gleich ☐ niedriger

4. Lager an Vorprodukten ○ keine Lager

a) Sie waren im Vergleich zum Vormonat *

☐ höher ☐ gleich ☐ niedriger

b) Wir beurteilen die Lager an Vorprodukten als

☐ zu gross ☐ normal ☐ zu klein

5. Lager an Fertigprodukten ○ keine Lager

a) Sie waren im Vergleich zum Vormonat *

☐ höher ☐ gleich ☐ niedriger

b) Wir beurteilen die Lager an Fertigprodukten als

☐ zu gross ☐ normal ☐ zu klein

Erwartungen

6. In den kommenden 3 Monaten wird im Vergleich zu den vergangenen 3 Monaten *...

a) der Bestellungseingang

☐ zunehmen ☐ gleich bleiben ☐ abnehmen

b) die Produktion

☐ zunehmen ☐ gleich bleiben ☐ abnehmen

c) der Vorprodukteeinkauf

☐ zunehmen ☐ gleich bleiben ☐ abnehmen

7. Die für die kommenden 3 Monate erwartete Geschäftsentwicklung wird sich **in den nachfolgenden Monaten ***

☐ verbessern ☐ fortsetzen ☐ verschlechtern

* Unter Ausschluss von saisonalen Schwankungen: Saisonale Schwankungen sind jahreszeitlich bedingte Nachfrageschwankungen und/oder Veränderungen infolge der üblichen Betriebsferien und Festtage, regelmässig wiederkehrender Reparaturen u.ä.

Erläuterungen zum Konjunkturtest

Industrie: Monatsumfrage

Allgemeine Hinweise

Der Konjunkturtest eilt den öffentlichen und teilweise auch den innerbetrieblichen Statistiken voraus und spiegelt daher nur die Tendenzen wider, ermöglicht aber ein frühzeitiges Erkennen von konjunkturellen Veränderungen. Die Antworten sollen auf den allgemeinen Lagekenntnissen leitender Personen basieren, eine Konsultation der "Bücher" ist nicht notwendig. Jeder Fragebogen bezieht sich auf eine bestimmte Produktegruppe. Die Antworten sind daher ausschliesslich auf diese Erzeugnisse zu beziehen. Bei Unternehmen mit nur einem Fragebogen entspricht dies allerdings der Geschäftstendenz der Firma. Alle Fragen beziehen sich auf Ihre Aktivitäten in der Schweiz.

Die Rubrik "Bemerkungen" ist reserviert für allgemeine Lagebeurteilungen, Hinweise auf besondere Verhältnisse in der entsprechenden Produktegruppe, der Unternehmung oder der Branche allgemein. Ebenfalls sollten vorgenommene Änderungen in der Beantwortungsmethode hier vermerkt werden. In Zeitabständen werden in diesem Feld von der KOF/ETH zusätzliche Fragen platziert.

Die folgenden Hinweise zu den einzelnen Fragen sind nur unverbindliche Ratschläge, da die Verhältnisse von Produkt zu Produkt und von Betrieb zu Betrieb sehr unterschiedlich sind. Wir bitten Sie jedoch, an der einmal gewählten Beantwortungsmethode festzuhalten.

Vergleich zum Vormonat

Insbesondere beim Bestellungseingang, aber auch bei anderen Fragen mit starken Schwankungen, gibt der Vormonatsvergleich in einzelnen Fällen die konjunkturelle Tendenz nicht richtig wieder. Es kann daher der Vergleich mit der Entwicklung der letzten Monate vorgenommen werden.

Fragen:

1. Bestellungseingang

Beim Bestellungseingang handelt es sich um Kundenaufträge; interne Aufträge sollen nicht berücksichtigt werden. Grundsätzlich ist von der bestellten Menge auszugehen (vor allem bei standardisierten Produkten). Wo dies nicht möglich ist, kann der Wert der Bestellungen als Grundlage dienen (rein preisbedingte Änderungen dabei ausschliessen).

2. Auftragsbestand

Er umfasst die Menge oder den (preisbereinigten) Wert der noch nicht in Arbeit genommenen Kundenaufträge.

Der Auftragsbestand ist zu niedrig, wenn er die normale Kapazitätsauslastung nicht ermöglicht, oder in Zukunft gefährdet. Er gilt als gross, wenn die Ausführung nicht in der gewünschten (normalen) Frist ausgeführt werden kann.

Wenn Sie regelmässig ins Ausland liefern, so beantworten Sie bitte auch Frage 2c. Es sind dabei auch Bestellungen zu berücksichtigen, die nicht direkt, sondern über Exportfirmen ins Ausland gelangen.

3. Produktion

Darunter ist die Menge oder der (preisbereinigte) Wert der erzeugten Zwischen- und Endprodukte, allenfalls die Summe der aufgewendeten Arbeits- und Maschinenstunden zu verstehen.

4. Lager an Vorprodukten

Es sind dies Bestände an Rohstoffen und unfertigen Erzeugnissen, die ausschliesslich von Dritten bezogen wurden. Es interessieren nur die mengenmässigen Veränderungen.

Sie sind zu hoch, bzw. zu klein, wenn ihr übliches - vielleicht saisonal unterschiedliches - Verhältnis zur geplanten Produktion in der einen oder anderen Richtung gestört ist.

5. Lager an Fertigprodukten

Es sind nur jene Bestände gemeint, die nicht auftrags- oder bestellungsgebunden sind. Kundenlager oder Endprodukte, die aus terminlichen oder technischen Gründen noch bei Ihnen lagern, zählen nicht dazu.

Die Lager sind zu hoch, wenn das gegenwärtige Bestandsniveau Ausdruck einer Absatzstockung ist und zu klein, wenn die Bestellungen nicht in der gewünschten Zeit ab Lager ausgeführt werden können.

6.+7. Erwartungen und Pläne

Abgrenzungen bezüglich Bestellungseingang und Produktion siehe Erläuterungen zur Frage 1 und 4 des Monatstests.

Wenn eine Vielzahl von Vorprodukten (Rohstoffen und Halbfabrikaten) eingekauft werden muss, sollen nur die wichtigsten bei der Beantwortung der Frage berücksichtigt werden.

Bei der künftigen Geschäftsentwicklung ist die Entwicklung ab dem 4. Monat (vom Berichtsmonat aus betrachtet) im Verhältnis zur erwarteten Entwicklung in der Zeit vom 1. - 3. Monat zu beurteilen.

Konjunkturforschungsstelle
Umfrage Industrie
ETH Zentrum WEH
8092 Zürich

Konjunkturforschungsstelle
Umfrage Industrie
ETH Zentrum WEH
8092 Zürich

Bernd Lucke und Malte Knüppel

Unternehmensgrößenklassen im ifo-Konjunkturtest: eine Burns-Mitchell-Analyse

1. Einleitung

Diese Arbeit wendet sich einem speziellen Aspekt des ifo-Konjunkturtests zu, dem normalerweise in der öffentlichen Rezeption nur geringe Aufmerksamkeit zukommt: der Analyse disaggregierter, in diesem Fall nach Größenklassen kategorisierter Unternehmensdaten. Diese Daten werden regelmäßig im Rahmen des ifo-Konjunkturtests durch Befragungen gewonnen und geben Aufschluss über die Geschäftslage, die Geschäftsbeurteilung und die Nachfragesituation im Vergleich zum Vormonat.

Wir wenden uns deshalb der Frage zu, wie diese mikroökonomischen, produktionsseitigen Einschätzungen in Zusammenhang mit den makroökonomischen konjunkturellen Gegebenheiten stehen. Dabei messen wir die konjunkturelle Situation anhand der Entwicklung des realen Bruttoinlandsprodukts (BIP) und klassifizieren einzelne Stadien des Konjunkturzyklus mit Hilfe der von Burns/Mitchell (1946) am *National Bureau of Economic Research* entwickelten Systematik. Diese Systematik impliziert eine – im Hauptteil der Arbeit eingehender beschriebene – Transformation der beobachteten Daten aus Echtzeit in „Zykluszeit". Diese Transformation entspricht ungefähr einer Stauchung der Kalenderzeit in der Aufschwungphase bzw. einer Streckung der Kalenderzeit in der Abschwungphase, so dass den bekannten Asymmetrien des klassischen Konjunkturzyklus besser Rechnung getragen werden kann, als dies in linearen Zeitreihenmodellen der Fall ist. Aber auch für die in dieser Arbeit hauptsächlich analysierten Wachstumszyklen ermöglicht die Burns-Mitchell-Methodik die Erkennung von Asymmetrien, die in Standardansätzen der Zeitreihenanalyse nicht unbedingt gewährleistet ist.

Da die Burns-Mitchell-Methodologie nicht als allgemein bekannt unterstellt werden kann, beginnt dieser Beitrag mit einem kurzen Abriss dieses Verfahrens. Es schließt sich eine Datierung deutscher Konjunktur- und Wachstumszyklen an, bevor die disaggregierten Unternehmensdaten detailliert untersucht werden. Insbesondere fokussieren wir auf das Potenzial dieser Daten für die Früherkennung von Wendepunkten sowie auf konjunkturelle Beschäftigungsimpulse, die von Unternehmen unterschiedlicher Größenklasse ausgehen.

2. Burns-Mitchell-Methodologie

Traditionell wird die Konjunkturanalyse des NBER mit saisonbereinigten Daten (Census X11-Methode) durchgeführt. Obwohl Forschungen der jüngeren Zeit deutlich gezeigt haben, dass das Arbeiten mit saisonbereinigten Daten zu nicht unerheblichen statistischen Fehlschlüssen Anlass geben kann (Maravall, Gomez 2001), soll diese Konvention hier nicht problematisiert werden.

Nach der Saisonbereinigung besteht der erste eigentliche Schritt zur Implementierung der NBER-Methodologie in der Datierung des Konjunkturzyklus für einen gegebenen Beobachtungszeitraum. Unter Datierung ist dabei die Erstellung einer Wendepunktchronologie zu verstehen, die die Abgrenzung zwischen expansiven und rezessiven Phasen der wirtschaftlichen Entwicklung ermöglicht. Die zeitliche Festsetzung oberer und unterer Wendepunkte ist freilich nicht trivial, da Konjunkturzyklen definitionsgemäß als gemeinsame Fluktuationen vieler ökonomischer Variablen angesehen werden, die typischerweise nicht exakt phasengleich sind. Oft jedoch häufen sich die Wendepunkte vieler Reihen mehr oder weniger konzentriert um einen bestimmten Zeitpunkt, so dass dieser als Wendepunkt identifiziert werden kann[1]. Die Gesamtheit aller identifizierten Wendepunkte definiert dann die Folge der sog. Referenzzyklen[2].

Eine wichtige Idee des Burns-Mitchell-Ansatzes besteht darin, Merkmale des Konjunkturverlaufs nicht nur mit Bezug auf Kalenderzeit, sondern auch mit Bezug auf „Zykluszeit" herauszuarbeiten. Ein voller Zyklus (hier als Hoch-

[1] Dieses Vorgehen ist offensichtlich sehr stark von subjektiven Einschätzungen der beteiligten Konjunkturforscher abhängig und seine Resultate daher u.U. nicht unkontrovers. Die naheliegende Alternative, den Konjunkturzyklus durch die Betrachtung nur eines Maßes für die aggregierte ökonomische Leistung (wie z.B. das Bruttoinlandsprodukt) zu definieren, wird daher auch von anderen Staaten und z.B. der OECD vorgezogen (Klein, Moore 1985: 54). Auch wir folgen in unserer Arbeit dieser Vorgehensweise und bestimmen den Referenzzyklus allein durch Rückgriff auf das reale Bruttoinlandsprodukt. Man muss freilich berücksichtigen, dass zuverlässige Messungen der aggregierten ökonomischen Aktivität zur Zeit des Entstehens von „Measuring Business Cycles" noch nicht verfügbar waren (Simkins 1994).

[2] Die folgende Beschreibung des methodischen Instrumentariums von Burns und Mitchell führt einige dem Leser vielleicht unvertraute Konzepte ein. Um die Lesbarkeit des Textes nicht zu stark zu erschweren, ist die Erläuterung der entsprechenden Begrifflichkeit jedoch relativ knapp gehalten. Zur besseren Übersicht konsultiere der interessierte Leser daher die präzisen Definitionen, die im Anhang bereitgestellt werden.

punkt-Tiefpunkt-Hochpunkt-Zyklus betrachtet) wird zu diesem Zweck in neun Stadien, die Zykluszeiteinheiten, unterteilt. Stadium I besteht aus den drei Monaten, deren mittlerer den ersten, oberen Wendepunkt enthält, Stadium V aus den drei Monaten, deren mittlerer den unteren Wendepunkt enthält, und Stadium IX ist identisch mit Stadium I des folgenden Zyklus. Die Zeit zwischen Stadium I und Stadium V (beginnend mit dem Monat nach dem oberen Wendepunkt und endend mit dem Monat vor dem unteren Wendepunkt) wird dann in die drei möglichst gleich großen Stadien II, III und IV disjunkt zerlegt[3]. Analog definiert man die Stadien VI, VII und VIII des Aufschwungs.

Der Mittelwert einer zu untersuchenden Zeitreihe während eines vollständigen Konjunkturzyklus, die sog. Zyklusbasis, besteht nun aus dem Mittelwert der Zeitreihe vom Beginn des Stadiums I bis zum Ende des Stadiums IX. Um diesen Mittelwert nicht nach oben zu verzerren, werden die sechs Monate der Stadien I und IX nur mit dem halben Gewicht versehen. Im zweiten Schritt der Burns-Mitchell-Methodologie werden alle Beobachtungen dann als Prozentsätze ihres jeweiligen Zyklusmittelwertes ausgedrückt. Die durch diese einfache Normierung entstehenden Variablen (die sog. *cycle relatives*) sind dimensionslos und ermöglichen daher eine weitgehende Vergleichbarkeit von Werten unterschiedlichen Ursprungs. Sie erlauben aber auch den Vergleich zwischen Beobachtungen in unterschiedlichen Konjunkturzyklen[4].

Wenn allerdings Wachstumsraten (und nicht Niveauwerte) analysiert werden sollen, werden die Beobachtungen nicht als Prozentsätze ihres jeweiligen Zyklusmittelwertes, sondern als die absoluten Abweichungen von diesem Wert ausgedrückt. Andernfalls würden die *cycle relatives* von Zyklen, deren Zyklusbasis nahe an null liegt, im Vergleich zu den *cycle relatives* anderer Zyklen zu extreme Werte annehmen, so dass die folgende Berechnung von Durchschnitten der *cycle relatives* kaum aussagekräftig wäre. (In Ermangelung eines handlicheren Ausdrucks für „absolute Abweichungen vom Zyklusmittelwert“ behalten wir den Begriff *cycle relative* bei, obwohl gerade „relative“ dann nicht mehr besonders treffend ist.)

Offenbar in der Annahme, allen Konjunkturzyklen liege eine ungefähr gleiche, nur durch stochastische Einflüsse gestörte Struktur zugrunde, berechnen Burns/Mitchell für die meisten ihrer Statistiken auch Mittelwerte über alle beobachteten Referenzzyklen. Für jede Reihe bestimmen sie z.B. zunächst

---

3 Man beachte aber, dass Phase II um einen Monat mit Phase I und Phase IV um einen Monat mit Phase V überlappt. King/Plosser (1994: 414) behaupten, dass es keine Überschneidungen gebe, haben aber offenbar Burns/Mitchell (1946: 144ff., 161) nicht sorgfältig gelesen.

4 Tatsächlich kann die Transformation in *cycle relatives* als eine (unvollständige) Form der Trendbereinigung aufgefasst werden: Sie eliminiert den Interzyklustrend, nicht aber den Intrazyklustrend. Entfernt wird also eine Trendfunktion, die konstant während des Zyklus ist, aber Sprungstellen zwischen den Zyklen aufweist.

den durchschnittlichen Wert der *cycle relatives* eines jeden Stadiums und mitteln diese durchschnittlichen *cycle relatives* dann über alle Referenzzyklen. Die resultierenden neun Kenngrößen bilden als sog. *business cycle plots* (oder Neun-Punkt-Diagramme) eine wichtige graphische Charakterisierung des Konjunkturverlaufs makroökonomischer Variablen.

Der dritte Schritt des Burns-Mitchell-Ansatzes besteht in der Berechnung statistischer Kenngrößen auf der Basis der *cycle relatives*. Dies beginnt bei der durchschnittlichen Dauer (über alle Referenzzyklen) von Aufschwung, Abschwung und vollem Zyklus. Auch Amplituden werden ermittelt: So wird die Amplitude in der Aufschwungphase als *cycle relative* des oberen Wendepunkts abzüglich des *cycle relatives* am vorhergehenden unteren Wendepunkt definiert. Die Amplitude des Abschwungs ist analog definiert und die Zyklusamplitude einer Variablen ergibt sich als die Summe der Absolutbeträge von Aufschwung- und Abschwungamplitude[5].

Durchschnittliche monatliche Zuwächse (in Prozentpunkten) werden von Burns/Mitchell erneut für Aufschwung, Abschwung und den vollen Zyklus betrachtet. Auch die durchschnittlichen Monatszuwächse zwischen den Konjunkturstadien (also von der Mitte des Stadiums I zur Mitte des Stadiums II z.B.) finden ihr Interesse. Besondere Bedeutung aber erlangt die sog. „Konformität“, die zum Ausdruck bringen soll, wie stark die Fluktuationen bestimmter Variablen mit der Entwicklung des Referenzzyklus übereinstimmen.

Die Konformität des Aufschwungs wird errechnet, indem jedem vollen Referenzzyklus, bei dem die betrachtete Variable in der Aufschwungphase im Durchschnitt steigt, ein Wert von 100 zugewiesen wird (ein Wert von –100, falls sie im Durchschnitt sinkt). Das Mittel dieser Werte über alle Referenzzyklen ist die Aufschwungkonformität. Entsprechend wird die Abschwungkonformität errechnet, nur erhält hier jede Abschwungphase, in der die Variable im Durchschnitt sinkt, den Wert von 100 zugewiesen und jede Abschwungphase, in der die Variable steigt, den Wert von –100.

Die Konformität über den gesamten Zyklus definieren Burns/Mitchell nicht als Durchschnitt aus Aufschwung- und Abschwungkonformität, sondern als ein Maß dafür, ob der durchschnittliche Monatszuwachs im Abschwung kleiner ist als in der vorhergehenden und der folgenden Aufschwungphase. Diese Definition ist vergleichsweise kompliziert und anders als alle anderen bisher

---

5 Anders King/Plosser (1994), die die Zyklusamplitude als Aufschwungamplitude minus Abschwungamplitude definieren. Dies kann zu anderen Resultaten führen, z.B. bei Konsumdaten, deren Abschwungamplitude oft positiv ist. Unsere hier gegebene Definition ist konsistent mit Burns/Mitchell, die ihrige nicht. Allerdings betrachten Burns/Mitchell auch für die Aufschwung- und Abschwungamplitude nur Absolutbeträge, worauf wir hier verzichten, um nicht sinnvolle Information zu vernichten (Burns, Mitchell 1946: 131ff.).

dargestellten Maße von der *Folge* der Konjunkturzyklen abhängig. Die sog. Zykluskonformität dynamisiert also die bislang auf isolierte Zyklen konzentrierte Betrachtungsweise. Für die Zwecke dieses Beitrags ziehen wir es allerdings vor, King/Plosser (1994) zu folgen, die die Zykluskonformität errechnen, indem sie für jeden Referenzzyklus, bei dem der mittlere monatliche Zuwachs im Aufschwung größer ist als der mittlere monatliche Zuwachs im Abschwung, einen Wert von 100 zuweisen (–100 im umgekehrten Fall) und den Durchschnitt dieser Werte über alle Referenzzyklen bilden. In der Terminologie von Burns/Mitchell heißt dieser Index *cycle conformity, peak to peak* und geht als eine von zwei Komponenten in die Berechnung der Zykluskonformität ein (Burns, Mitchell 1946: 176ff.).

Burns/Mitchell weisen Konformitäten lediglich für Aufschwung, Abschwung und den ganzen Zyklus aus. Es ist aber vielleicht sinnvoll, den Konformitätsbegriff etwas weiter aufzufächern und Konformitäten auch für die von ihnen vorgeschlagenen Abschwungstadien II, III und IV bzw. Aufschwungstadien VI, VII und VIII zu berechnen. Wir werden von diesen als „Stadienkonformitäten" sprechen. Die Stadienkonformität im Abschwung ist das gewichtete Mittel der drei Stadienkonformitäten II, III und IV, die Stadienkonformitäten im Aufschwung bzw. im gesamten Zyklus sind analog die gewichteten Mittel der Stadienkonformitäten VI, VII und VIII bzw. II-IV und VI-VIII[6]. Diese Stadienkonformitäten sind vermutlich imstande, ein etwas differenzierteres Bild von Kobewegungen einzelner Reihen zu vermitteln, als die von Burns/Mitchell verwendeten Maße.

3. Bestimmung der Referenzzyklen

Die oben dargestellten Konzepte sollen im Folgenden auf Daten des ifo-Konjunkturtests angewendet werden. Dafür ist zunächst die Ermittlung der Referenzzyklen erforderlich, d.h. die Datierung der konjunkturellen Wendepunkte. Zu diesem Zweck soll auf ein am NBER entwickeltes Computerprogramm von Bry/Boschan (1971) zurückgegriffen werden[7]. Das Bry-Boschan-Programm wurde vom NBER explizit zur Wendepunktbestimmung in der Konjunkturanalyse eingesetzt. Allerdings wurden die vom Programm identifizierten Wendepunkte in der Regel nicht mechanisch übernommen, sondern einer auf Erfahrungswissen basierenden Kritik unterzogen, die u.U. zu abweichenden Datierungen führen konnte.

Bry/Boschan haben die rein mechanische Fähigkeit ihres Programms zur Wendepunktidentifikation getestet, indem sie es auf mehr als 50 vom NBER

[6] Jede Stadienkonformität wird mit ihrer Länge (in Monaten) gewichtet.

[7] Wir haben eine in GAUSS geschriebene Version des Bry-Boschan-Programms benutzt, die uns freundlicherweise von Mark Watson zur Verfügung gestellt wurde.

Übersicht 1

Bry-Boschan-Prozedur zur computergestützten Wendepunktbestimmung

I. Ausreißerbestimmung und Substitution extremer Beobachtungen

II. Zyklusbestimmung in einem zwölfmonatigen gleitenden Durchschnitt; extreme Beobachtungen ersetzt
   A. Bestimmung von Werten, die höher (oder niedriger) sind als alle Werte im Umkreis von 5 Monaten
   B. Erzwingung alternierender Wendepunkte: Höchster von multiplen oberen Wendepunkten; niedrigster von multiplen unteren Wendepunkten[1]

III. Zyklusbestimmung in einem weniger stark geglätteten gleitenden Durchschnitt (sog. Spencer-Kurve); extreme Beobachtungen ersetzt
   A. Bestimmung der höchsten (oder niedrigsten) Werte im Umkreis von 6 Monaten um die Wendepunkte des zwölfmonatigen gleitenden Durchschnitts
   B. Erzwingung der Mindestzykluslänge von 15 Monaten durch Elimination niedrigerer oberer oder höherer unterer Wendepunkte bei Zyklen kürzerer Dauer

IV. Bestimmung der Wendepunkte in einem kurzfristigen gleitenden Durchschnitt von zwischen 3 und 6 Monaten (nach Vorgabe des Nutzers oder datengestützt)
   A. Bestimmung der höchsten (oder niedrigsten) Werte im Umkreis von 6 Monaten um die Wendepunkte der Spencer-Kurve

V. Bestimmung der Wendepunkte in der ungeglätteten Zeitreihe
   A. Bestimmung der höchsten (oder niedrigsten) Werte im Umkreis von 6 Monaten um die Wendepunkte des kurzfristigen gleitenden Durchschnitts
   B. Elimination von Wendepunkten, die nicht weiter als 6 Monate vom Anfang oder Ende der Zeitreihe entfernt liegen
   C. Elimination von oberen (oder unteren) Wendepunkten an beiden Enden der Zeitreihe, die niedriger (oder höher) sind als noch näher an den Enden platzierte Werte
   D. Elimination von Zyklen mit einer Dauer von weniger als 15 Monaten
   E. Elimination von Phasen mit einer Dauer von weniger als 5 Monaten

VI. Ausgabe der endgültigen Wendepunkte

[1]Illustration zur Möglichkeit der Identifizierung von aufeinanderfolgenden Hochpunkten ohne dazwischenliegenden Tiefpunkt: Oktober: 90; November: 100; Dezember: 90; Januar: 80; Februar: 100; März: 120; April: 110; Mai: 100; Juni: 90; August: 110; September: 130; Oktober: 120; November: 100; Dezember: 110; Januar: 120; Februar: 110. Bei dieser Zeitreihe würde man zunächst zwei aufeinanderfolgende Hochpunkte im März (120) und im September (130) finden, denn zwischen diesen beiden Hochpunkten erfüllt der Juni (90) nicht die Anforderung an einen Tiefpunkt, da fünf Monate zuvor im Januar (80) ein niedrigerer Wert vorliegt. Von diesen beiden Hochpunkten würde dann der Septemberwert als Hochpunkt ausgewählt werden.

analysierte Reihen anwendeten und seine Resultate mit denen des NBER verglichen. Von insgesamt 435 Wendepunktfestsetzungen des NBER wurden 94% vom Programm repliziert, davon rund 90% mit identischen Datierungen. Allerdings schlug das Programm ungefähr 15% mehr Wendepunkte vor, als den NBER-Einschätzungen entsprach.

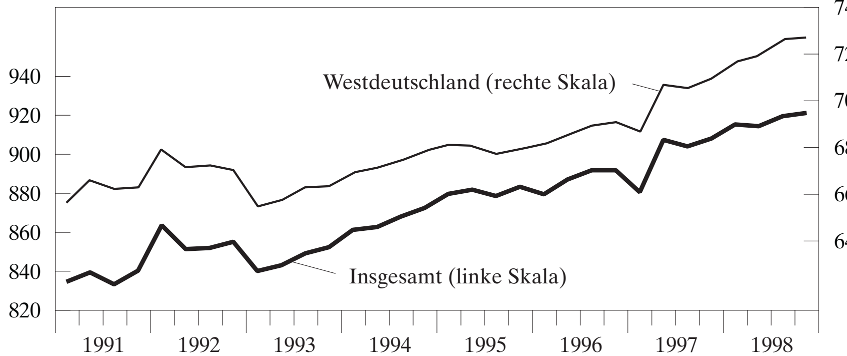
Das Bry-Boschan-Programm ist für Monatsdaten[8] konstruiert und arbeitet im Prinzip wie folgt: Zunächst werden mögliche Ausreißer ermittelt und durch weniger extreme Beobachtungen ersetzt. Dann wird die Reihe mit einem *moving average*-Filter geglättet, und es werden vorläufige Wendepunkte ermit-

[8] Bry/Boschan bearbeiteten Quartalsdaten, indem sie (für Stromgrößen) den Quartalswert gleichmäßig auf die drei Monatswerte aufteilten. Wir folgen dieser Praxis hier. Eine Anpassung des Programms an Quartalsdaten ist zwar im Prinzip möglich, würde aber eine Reihe von willkürlichen Festsetzungen erfordern, z.B. bei der Konstruktion von Gewichten für eine „Spencer-Kurve" für Quartalsdaten. Ein solches Vorgehen scheint darüber hinaus nicht ratsam zu sein, da die Vergleichbarkeit mit dem NBER-Ansatz beeinträchtigt wäre, ohne dass überzeugende Anhaltspunkte für dessen offensichtliche Nachteiligkeit vorlägen.

Schaubild 1

Bruttoinlandsprodukt in Deutschland

1991 bis 1998



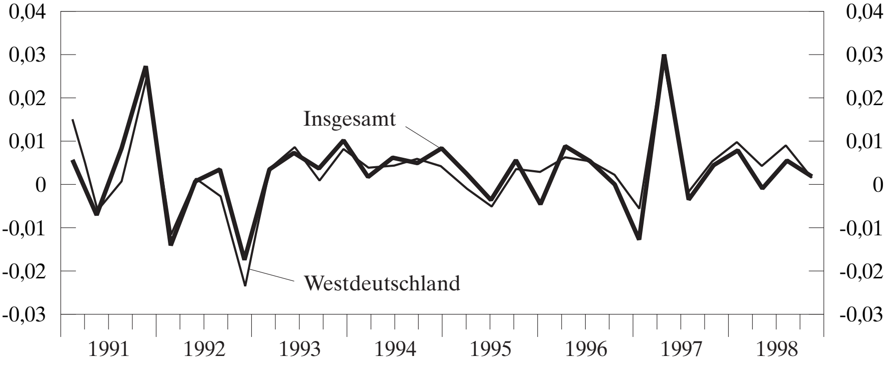
telt, die bestimmten Restriktionen genügen. Dieser Schritt wiederholt sich noch zweimal mit anderen *moving average*-Filtern, wobei jeder neue *moving average*-Filter geringere Glättungswirkung hat als sein Vorgänger. Schließlich werden die Wendepunkte der Ausgangsreihe (ohne Ausreißerbereinigung) ermittelt, die wiederum einer Reihe von Restriktionen genügen müssen. Eine genauere Darstellung des Vorgehens findet sich in Übersicht 1, die an Bry/Boschan (1971: 21) angelehnt ist[9].

Eine computergestützte Datierung des Konjunkturzyklus hat den unbestreitbaren Nachteil, dass sie auf vorhandenes Expertenwissen und -erfahrung nur in relativ eingeschränktem Maße zurückgreifen kann, nämlich nur insoweit diese bei der Programmierung des Quellcodes berücksichtigt wurden. Sie besitzt allerdings den entscheidenden Vorteil, dass sie für jedermann nachvollziehbar und von subjektiven Urteilen weitgehend unabhängig ist.

Bei der Konstruktion der Referenzzyklen mit Hilfe der Bry-Boschan-Prozedur ergibt sich zunächst das Problem, dass Daten für das reale BIP Westdeutschlands nur bis 1998 existieren. Wir haben uns damit beholfen, die westdeutschen Daten ab 1999 aus dem gesamtdeutschen BIP zu schätzen. Die dazu erforderlichen Parameter werden durch eine Regression des westdeutschen auf das gesamtdeutsche BIP im Zeitraum 1991 bis 1998 gewonnen. Schaubild 1 macht deutlich, dass insbesondere für die hier interessierenden zyklischen Eigenschaften des BIP in den neunziger Jahren kein maßgeblicher Unterschied zwischen gesamtdeutschen und westdeutschen Daten zu bestehen scheint.

[9] Es gibt kleine Abweichungen zwischen Bry/Boschans Beschreibung ihres Vorgehens und ihrem Computerprogramm. Übersicht 1 bezieht sich auf den tatsächlich programmierten Code.

Schaubild 2
Veränderungsrate des Bruttoinlandsprodukts in Deutschland
1991 bis 1998



Ein weiteres Problem ergibt sich aus der Tatsache, dass die im Mittelpunkt der Analyse stehenden disaggregierten Unternehmensdaten lediglich seit 1984 erhoben werden, d.h. die Stichprobe umfasst lediglich ca. 17 Jahre. In einem Zeitraum dieser Länge findet man im Normalfall nur sehr wenige Zyklen in den Niveaus des BIP, was kaum eine Typisierung konjunktureller Muster zulässt.

Alternativ bietet sich die Bestimmung von Referenzzyklen für die *Wachstumsraten* des BIP an, die üblicherweise eine größere Anzahl von Zyklen aufweisen. Auch hier liegt eine recht enge Korrelation zwischen westdeutschen und gesamtdeutschen Wachstumsraten vor, so dass die fehlenden Beobachtungen für den alten Gebietsstand aus gesamtdeutschen Wachstumsraten geschätzt werden können (Schaubild 2). Der Vollständigkeit halber (und aus illustrativen Gründen) beginnen wir mit der Bestimmung eines Referenzzyklus für das Niveau des deutschen BIP, allerdings für den über unseren Untersuchungshorizont hinausreichenden Zeitraum 1968 bis 2001.

Betrachtet man das bundesdeutsche reale BIP im Zeitraum von 1968 bis 2001, so ergibt die Anwendung der Bry-Boschan-Prozedur eine Datierung von Konjunkturzyklen, deren Hoch- und Tiefpunkte in Schaubild 3 durch vertikale Linien markiert werden. Demnach wurde in Westdeutschland im Zeitraum von 1984 bis 2001 kein einziger Konjunkturzyklus vollständig durchlaufen, so dass das BIP für eine Burns-Mitchell-Analyse in dem Zeitraum, in dem Geschäftsklimadaten vorliegen, ungeeignet ist.

Bezogen auf die Wachstumsrate des realen BIP finden sich im Analysezeitraum von 1984 bis 2001 immerhin vier Hochpunkt-Tiefpunkt-Hochpunkt-Zyklen (Schaubild 5 oben). Da sich lediglich drei Tiefpunkt-Hochpunkt-Tief-

Schaubild 3

Konjunkturelle Wendepunkte des Bruttoinlandsprodukts

1965 bis 2005

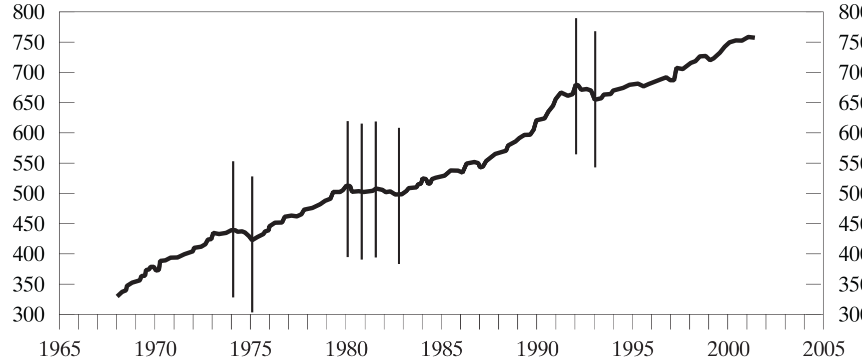
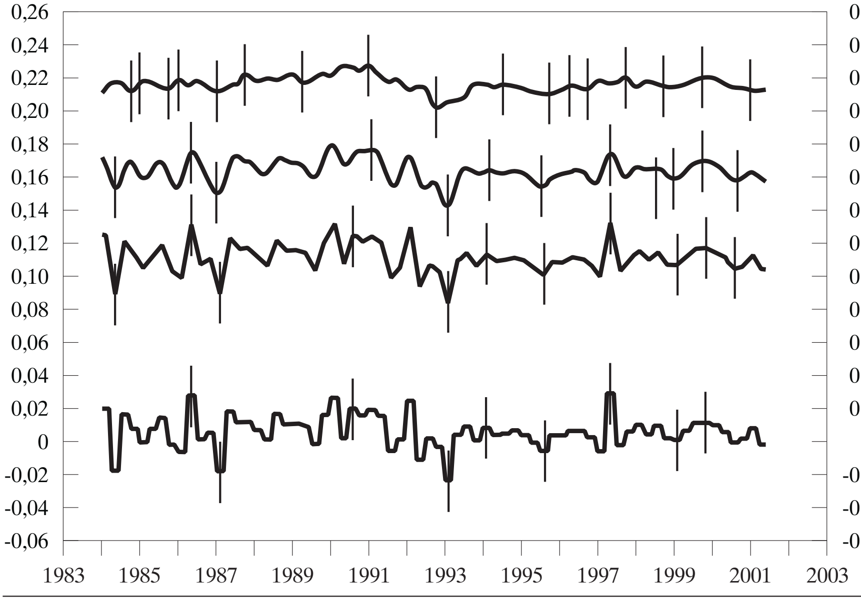


Schaubild 4

Bry-Boschan-Prozedur: Sukzessive Wendepunktbestimmung für Wachstumsraten des BIP

1983 bis 2003

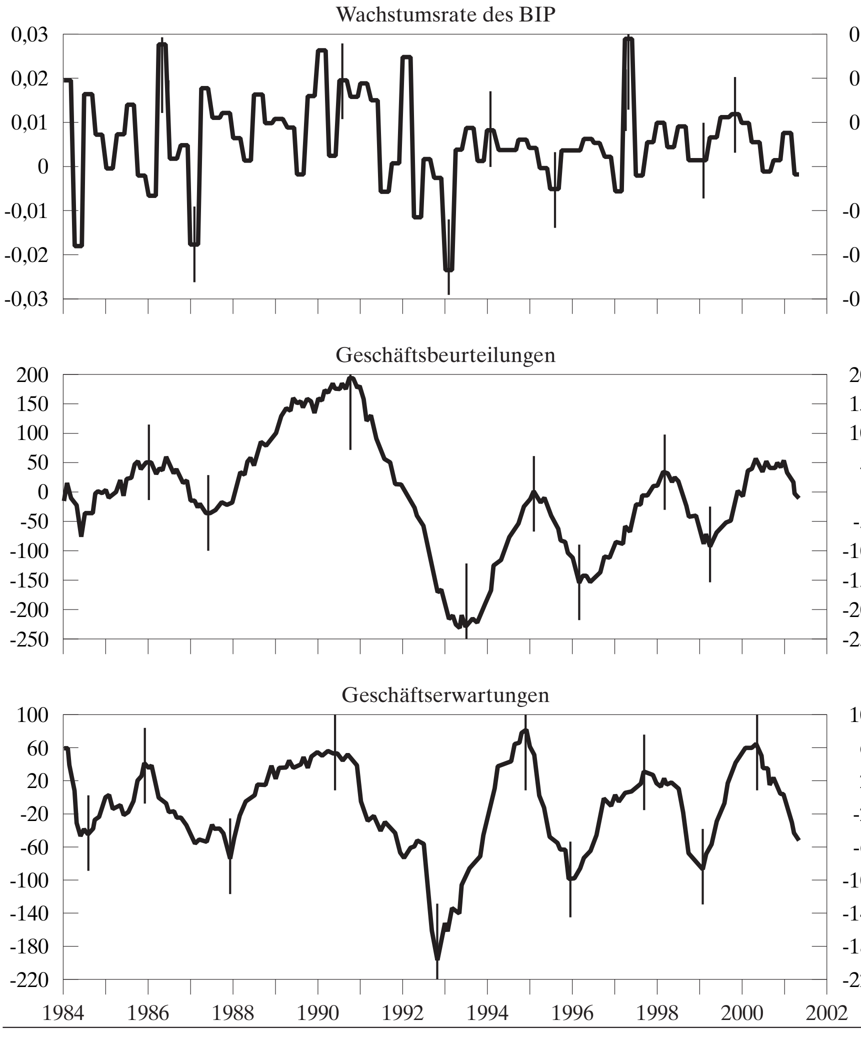


punkt-Zyklen finden, wird der ersteren Zyklusabgrenzung der Vorzug gegeben. Die Hoch- und Tiefpunkte sind in Schaubild 5 ebenfalls durch vertikale Linien markiert. Schaubild 4 illustriert darüber hinaus, wie die Bry-Boschan-Prozedur von der Analyse der am stärksten geglätteten Reihe bis zu der der

Schaubild 5

Bry-Boschan-Prozedur: Sukzessive Wendepunktbestimmung ausgewählter Zeitreihen

1984 bis 2002



Originalreihe voranschreitet. Die Werte der Ordinate in Schaubild 4 beziehen sich auf die unten stehende Originalreihe.

Zur besseren Vergleichbarkeit werden in Schaubild 5 auch die (noch aggregierten) ifo-Reihen „Geschäftsbeurteilungen“ und „Geschäftserwartungen“ mit Hilfe der Bry-Boschan-Prozedur datiert, d.h. deren Wendepunkte bestimmt. Dies ist für die Bestimmung eines Referenzzyklus zwar nicht relevant (der Referenzzyklus wird in unserer Arbeit allein aus den BIP-Daten ermittelt), jedoch ist es für die folgende Analyse zumindest instruktiv zu sehen, dass Wendepunkte im Referenzzyklus weitgehende Entsprechung in den Wendepunkten der Geschäftsbeurteilungen und der Geschäftserwartungen finden.

Das Bry-Boschan-Programm identifiziert also bei zwei der drei untersuchten Zeitreihen jeweils vier Hochpunkt-Tiefpunkt-Hochpunkt-Zyklen. Der fehlende *peak* der Geschäftsbeurteilungen im Zeitraum von 2000 bis 2001 erscheint allerdings nicht problematisch, denn die Bry-Boschan-Prozedur identifiziert für diese Reihe zunächst einen Hochpunkt im Januar 2001, eliminiert diesen dann jedoch, da er zu nahe am Ende der Reihe liegt. Bei einem minimal veränderten Verlauf der Zeitreihe hätte dieser Hochpunkt z.B. im Frühjahr 2000 identifiziert werden können, was zu seiner Beibehaltung geführt hätte. Bemerkenswert ist auch, dass die Bry-Boschan-Prozedur in der etwas volatileren Reihe der Geschäftserwartungen nicht mehr als vier Zyklen identifiziert. So wäre z.B. die Feststellung je eines weiteren Hoch- und Tiefpunkts in den Jahren 1985, 1987 und 1992 vom optischen Eindruck her vorstellbar gewesen. Die exakte Datierung der ermittelten Hoch- und Tiefpunkte findet sich in Tabelle 1.

Wie man Tabelle 1 entnimmt, sind die Hoch- und Tiefpunkte der Geschäftsbeurteilungen und der Geschäftserwartungen keineswegs koinzident, noch haben sie eine klare *lead*- oder *lag*-Eigenschaft gegenüber der Wachstumsrate des BIP. Jedoch findet sich zumindest das grobe Muster der vier Zyklen in allen drei Reihen wieder, während, wie bereits erwähnt, das Niveau des BIP gar keinen vollen Zyklus im Untersuchungszeitraum 1984 bis 2001 aufweist. Dies ist insofern bemerkenswert, als die vom ifo gestellten Fragen nach der Beurteilung oder Erwartung der Geschäftslage Fragen nach dem *Niveau* der wirtschaftlichen Aktivität darstellen. Dennoch scheinen die ifo-Reihen dynami-

Tabelle 1

Hoch- und Tiefpunkte gemäß Bry-Boschan-Prozedur

Veränderungsrate BIP[1]		Geschäftsbeurteilungen		Geschäftserwartungen	
Hochpunkt	Tiefpunkt	Hochpunkt	Tiefpunkt	Hochpunkt	Tiefpunkt
1986, Mai	1987, Feb	1986, Jan	1987, Juni	1985, Dez	1987, Dez
1990, Aug	1993, Feb	1990, Okt	1993, Juli	1990, Juni	1992, Nov
1994, Feb	1995, Aug	1995, Feb	1996, März	1994, Dez	1995, Dez
1997, Mai	1999, Feb	1998, März	1999, Apr	1997, Sep	1999, Feb
1999, Nov				2000, Mai	

[1] Die Bry-Boschan-Prozedur wurde zur Auswertung der Wachstumsrate des BIP minimal modifiziert, da die ursprüngliche Prozedur Hochpunkte immer am Quartalsbeginn, Tiefpunkte jedoch am Quartalsende identifiziert. Dies hängt mit der einfachen Konvertierung von Quartals- zu Monatsdaten zusammen, die hier verwendet wird. Die Modifikation bewirkt nun, dass sowohl Hochpunkte als auch Tiefpunkte in der Quartalsmitte identifiziert werden. Die so gewonnenen Ergebnisse stimmen gut mit denen überein, die man erhält, wenn man aufwendigere, glättende Verfahren zur Konvertierung der Quartals- zu Monatsdaten verwendet.

sche Eigenschaften zu haben, die eher mit jenen der *Wachstumsrate* des BIP einhergehen.

4. Geschäftsklimadaten nach Unternehmensgrößenklassen

Die soeben in aggregierter Form bereits benutzten Daten entstammen einer monatlich unter dem Titel „ifo-Konjunkturtest“ in deutschen Unternehmen durchgeführten Umfrage. Sie gibt Aufschluss über die konjunkturelle Lage und über die kurzfristige Planung in Industrie, Bauwirtschaft, Groß- und Einzelhandel. Der ifo-Konjunkturtest ist in mehr als 300 Produktgruppen bzw. Märkte untergliedert. Die stark zusammengefassten Resultate des Konjunkturtests werden am Ende des Folgemonats veröffentlicht.

Die hier durchgeführte Analyse bezieht sich auf drei Kategorien des ifo-Konjunkturtests, die uns von Januar 1984 bis Mai 2001 für Firmen unterschiedlicher Größenklassen aus dem verarbeitenden Gewerbe in Westdeutschland vorliegen. Es handelt sich dabei um die Salden *Geschäftsbeurteilung*, *Geschäftserwartungen* und *Nachfragesituation gegenüber dem Vormonat* von Unternehmen, die auf Basis ihrer Beschäftigtenzahlen in fünf Größenklassen eingeteilt werden:

Größenklasse 1:	1 bis 49 Beschäftigte
Größenklasse 2:	50 bis 199 Beschäftigte
Größenklasse 3:	200 bis 499 Beschäftigte
Größenklasse 4:	500 bis 999 Beschäftigte
Größenklasse 5:	1000 und mehr Beschäftigte

Für die Fragen nach Geschäftsbeurteilung, Geschäftserwartungen und der Nachfragesituation gegenüber dem Vormonat gibt es jeweils drei Antwortkategorien (+, =, –), die so über alle Unternehmen aggregiert werden, dass die Salden einen Maximalwert von +100 und einen Minimalwert von –100 annehmen können[10]. Die Salden weisen im Zeitverlauf ein ausgeprägt zyklisches Verhalten auf, wie in Schaubild 6 zu sehen ist, das noch einmal die über die Größenklassen aggregierten Zeitreihen darstellt.

Dabei fällt auf, dass die Zeitreihe der „Nachfragesituation gegenüber dem Vormonat“ sehr stark den Geschäftserwartungen ähnelt. Die Nachfragesituation scheint lediglich einen etwas volatileren Charakter zu besitzen. Da diese Eigenschaft auch in den disaggregierten Zeitreihen besteht, soll, um eine bessere Übersichtlichkeit zu gewährleisten, die Nachfragesituation nicht weiter betrachtet werden. Es sollte jedoch der Beobachtung Aufmerksamkeit geschenkt werden, dass die Erwartungen der Unternehmen möglicherweise in rückblickender Form gebildet werden.

[10] Also: Saldo = 100 · (Anzahl „+“ – Anzahl „–“) ÷ Anzahl der Unternehmen.

Schaubild 6

Geschäftsbeurteilung, Geschäftserwartung und Nachfrage

1984 bis 2001; Salden

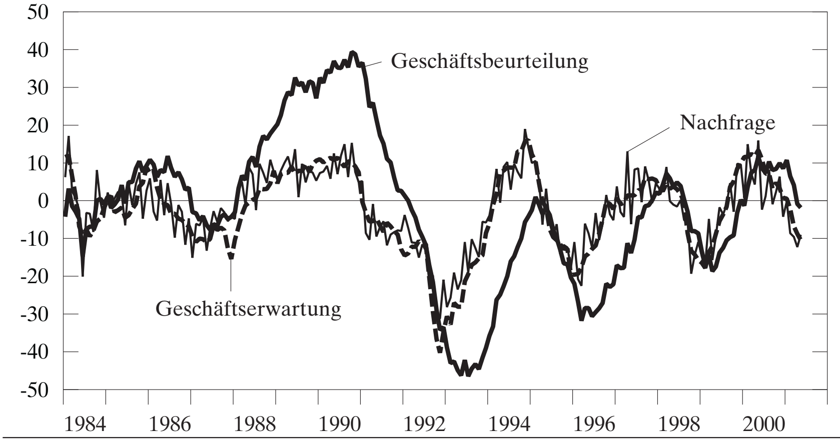
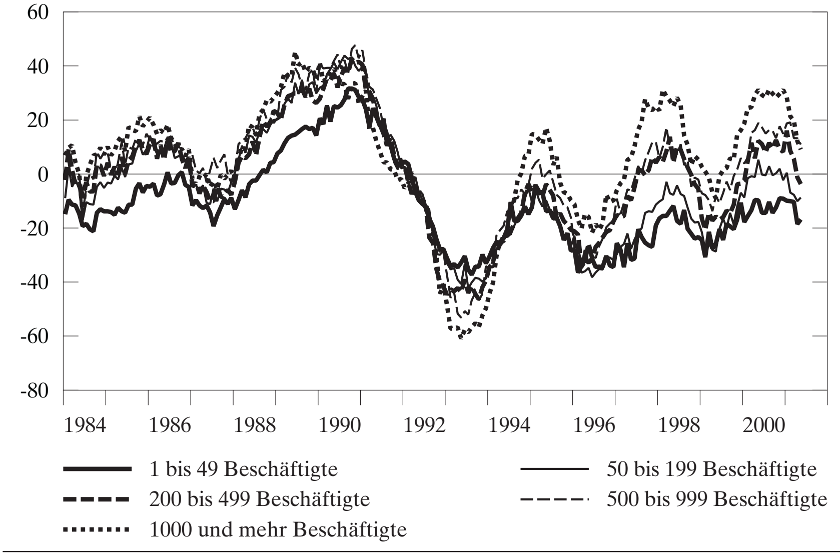


Schaubild 7

Geschäftsbeurteilung nach Größenklassen

1984 bis 2001; Salden

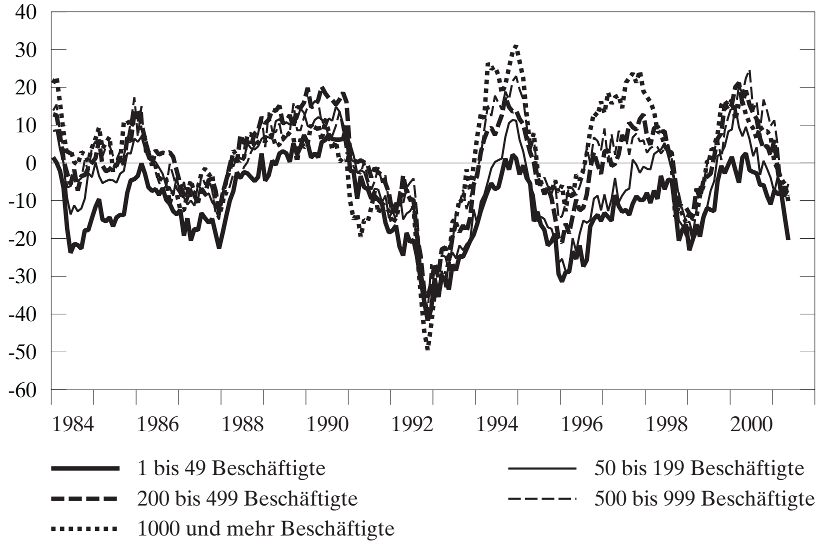


Der erste visuelle Eindruck legt dennoch – nicht überraschend – einen Vorlauf der Geschäftserwartungen gegenüber den Geschäftsbeurteilungen nahe. Allerdings folgen den Veränderungen der Geschäftserwartungen nicht immer Veränderungen der Geschäftsbeurteilungen in dieselbe Richtung. So schlie-

Schaubild 8

Geschäftserwartung nach Größenklassen

1984 bis 2001; Salden



ßen sich z.B. an die Verschlechterungen der Geschäftserwartungen Anfang 1985 oder auch Ende 1987 keine Verschlechterungen der Geschäftsbeurteilung an, und auch die Anfang 1992 beobachtete leichte Verbesserung der Geschäftserwartungen zieht keine Verbesserung der Geschäftsbeurteilung nach sich.

Die nach Größenklassen unterteilten Geschäftsbeurteilungen verlaufen, ebenso wie die Geschäftserwartungen, auf den ersten Blick recht gleichförmig (Schaubild 7 und 8). Der einzige offensichtliche Unterschied zwischen Unternehmen verschiedener Größenklassen scheint die oft positivere Einschätzung von Geschäftsbeurteilungen und -erwartungen bei den größeren Unternehmen zu sein. Das dynamische Verhalten der verschiedenen Größenklassen scheint sich dagegen nur wenig zu unterscheiden. So weist auch ein Neun-Punkt-Diagramm aus, dass sich die durchschnittlichen *cycle relatives* der Geschäftsbeurteilungen verschieden großer Firmen kaum unterscheiden (Schaubild 9).

Für alle Größenklassen gilt, dass die Geschäfts*beurteilungen* der Wachstumsrate des BIP nachlaufen. Allenfalls scheint ein konjunktureller Tiefpunkt auf größere Unternehmen stärkere negative Auswirkungen zu besitzen in dem Sinne, dass die Beurteilung der Geschäftslage relativ zur durchschnittlichen Geschäftsbeurteilung des Zyklus im Stadium nach dem Tiefpunkt umso schlechter ausfällt, je größer das Unternehmen ist.

Schaubild 9

Neun-Punkt-Diagramm der Geschäftsbeurteilungen

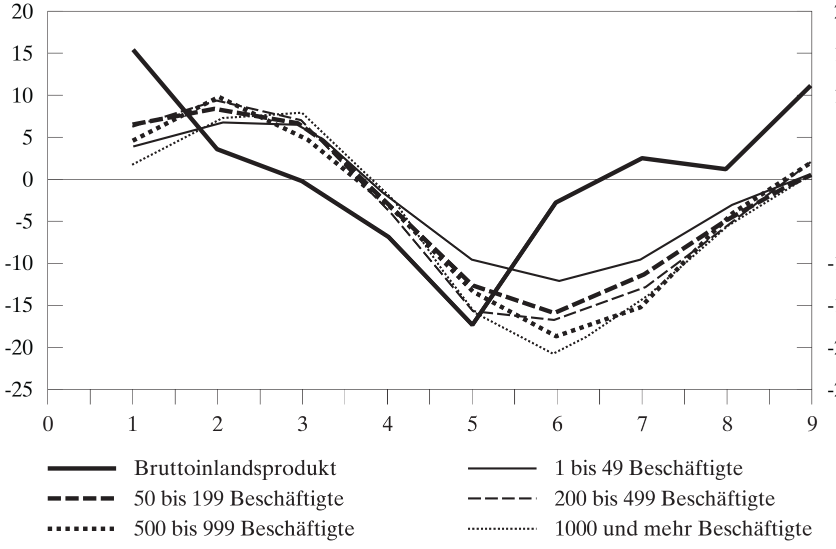


Schaubild 10

Neun-Punkt-Diagramm der Geschäftserwartungen

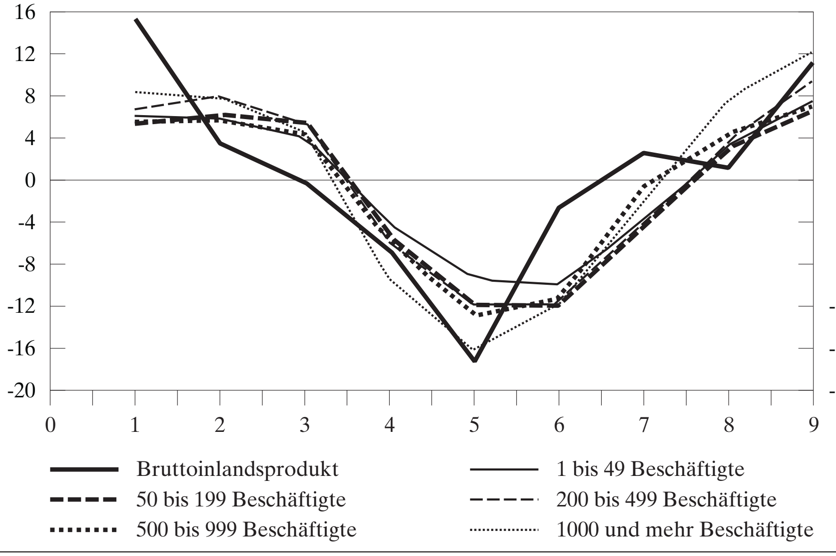
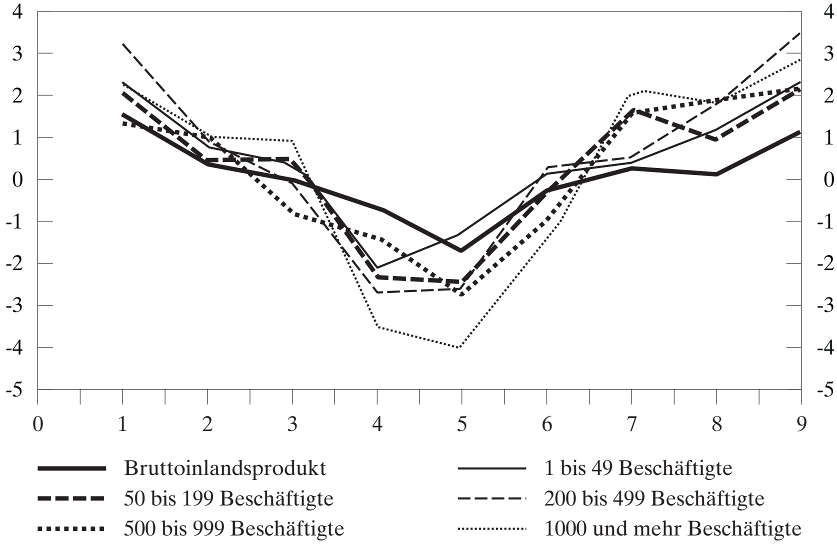


Schaubild 11

Neun-Punkt-Diagramm der 1. Differenz der Geschäftsbeurteilungen



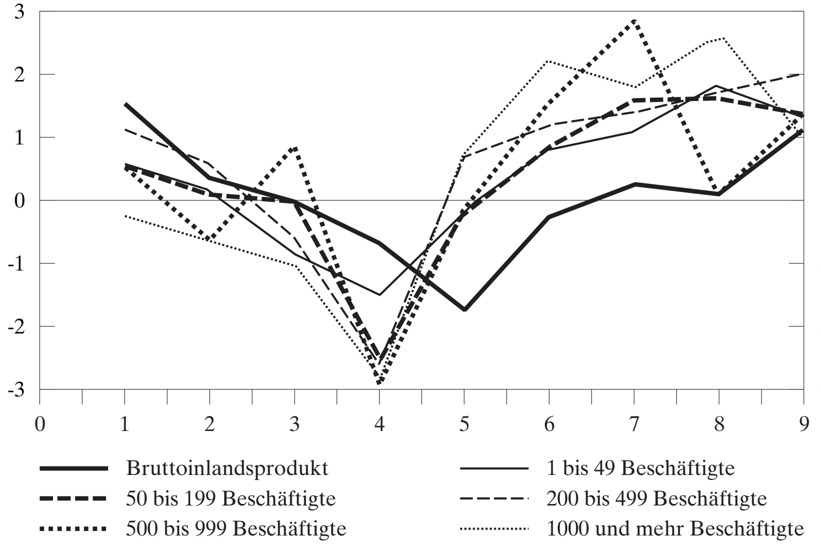
Die *cycle relatives* der unterschiedlichen Größenklassen besitzen auch bezüglich der Geschäfts*erwartungen* eine ausgeprägte Homogenität (Schaubild 10): Hier kann man sehen, dass die Geschäftserwartungen ungefähr zeitgleich mit der Wachstumsrate des BIP verlaufen. Allerdings hinken die Erwartungen der kleineren Betriebe den Tiefpunkten der Wachstumsrate des BIP leicht nach. Dass die Geschäftserwartungen der Wachstumsrate des BIP nicht vorauseilen, ist nicht weiter verwunderlich, wenn die Erwartungen, wie eingangs vermutet, im Wesentlichen die Veränderung der Nachfrage gegenüber dem Vormonat reflektieren.

Es wäre freilich vorschnell, daraus folgern zu wollen, dass die Geschäftserwartungen wenig Information über die zukünftige Entwicklung des BIP zur Verfügung zu stellen vermögen. Denn einerseits sind die ifo-Reihen schneller verfügbar (sie werden am Ende des Folgemonats veröffentlicht, BIP-Daten jedoch erst knapp zwei Monate nach Quartalsende), andererseits können wertvolle Informationen aus der *Veränderung* von Geschäftsbeurteilungen oder Geschäftserwartungen entnommen werden. Wir betrachten deshalb jetzt die ersten Differenzen dieser Variablen.

Beginnend mit den Neun-Punkt-Diagrammen der Geschäfts*beurteilungen* (Schaubild 11), ist festzuhalten, dass für Mittelstandsunternehmen (hier definiert als die Größenklassen 1–3) die Geschäftsbeurteilungen der Wachstumsrate des BIP im Mittel um ein Stadium vorzulaufen scheinen. Beim Hoch-

Schaubild 12

Neun-Punkt-Diagramm der 1. Differenz der Geschäftserwartungen



punkt freilich ist kein Vorlauf ersichtlich, jedoch signalisieren bei Unternehmen der Größenklassen 2 und 5 die Geschäftsbeurteilungen die Spätphase des Aufschwungs durch ein temporär gemindertes Wachstum in Stadium 8.

Für die Geschäfts*erwartungen* ergibt sich bei allen Größenklassen deutlich ein Vorauseilen um ein Stadium vor dem Tiefpunkt der Wachstumsrate des BIP (Schaubild 12). Bezüglich des Hochpunkts ist die Dynamik der unterschiedlichen Größenklassen weniger homogen. Nur bei den Größenklassen 1, 2 und 5 ist ein Vorauseilen um ein Stadium feststellbar. Der Hochpunkt der Größenklasse 3 ist zeitgleich mit dem der Wachstumsrate des BIP, Größenklasse 4 ist über den gesamten Zyklus sehr volatil, so dass sie sich als Prädiktor aufgrund ihrer hohen Varianz wohl kaum eignet. Im Allgemeinen scheinen die Geschäftserwartungen unzuverlässiger für das Erkennen des Hochpunkts als für das Erkennen des Tiefpunkts zu sein.

Die bisher erzielten Erkenntnisse lassen folgende kurze Zusammenfassung zu:

- Das Prognosepotenzial insbesondere der Geschäftserwartungen ist am Tiefpunkt deutlich anders als am Hochpunkt, d.h. es liegt eine klare Asymmetrie gegenüber den beiden Wendepunkten vor. Diese kann in traditionellen linearen Zeitreihenmodellen nicht wiedergegeben werden, tritt aus der deskriptiv-statistischen Darstellung eines Burns-Mitchell'schen Neun-Punkt-Diagramms aber klar hervor.

– Unterschiedliche Größenklassen haben unterschiedliche Prognoseeigenschaften. Beispielsweise wird die Spätphase des Aufschwungs durch eine Betrachtung von Geschäftsbeurteilungen von Unternehmen der Größenklassen 2 und 5 vielleicht als solche erkannt; umgekehrt sind die Angaben kleiner und mittelständischer Unternehmen vielleicht tauglicher zur Diagnose eines bevorstehenden Tiefpunkts. Dies impliziert, dass die bloße Betrachtung aggregierter Größen u.U. wertvolle disaggregierte Information „verschmiert" und damit tendenziell zu schlechteren Einschätzungen Anlass gibt.

Freilich ist zu berücksichtigen, dass *cycle relatives* lediglich Durchschnittswerte (hier über vier Zyklen) darstellen. Neben dem Mittelwert einer Stichprobe ist jedoch auch deren Varianz relevant. Obwohl es sicherlich wenig sinnvoll ist, die Varianz einer Stichprobe von Größenordnung vier zu berechnen, kann weitergehendes Burns-Mitchell-Instrumentarium schnell deutlich machen, dass manche Ergebnisse der Neun-Punkt-Diagramme zu suggestiv sind, da sie die Unterschiedlichkeit der einzelnen Zyklen nicht berücksichtigen. Das Konformitätsmaß (Definition siehe Anhang) ermittelt beispielsweise, wie synchron zwei Reihen über mehrere Zyklen hinweg verlaufen, wobei ein Konformitätswert von 100 völlige Synchronität und ein Konformitätswert von –100 völlige Antikonformität bezeichnen.

In Tabelle 2 wird die Konformität zwischen der Wachstumsrate des BIP und den um ein Stadium verzögerten Veränderungen der Geschäftserwartungen am Hochpunkt und am Tiefpunkt berechnet, um zu überprüfen, wie robust die aus den Neun-Punkt-Diagrammen entnommenen Vorlauf-Eigenschaften sind. Es stellt sich z.B. heraus, dass diese sich in völliger Synchronität nur in den Größenklassen 3 und 4 finden, nicht aber z. B. in Größenklasse 1 (diese ist offenbar nur in zwei Zyklen synchron, die aber bei der Durchschnittsbildung im Neun-Punkt-Diagramm dominiert). Die Konformität im Hochpunkt ist insgesamt unbefriedigend, evtl. mit Ausnahme von Größenklasse 5.

Tabelle 2

Burns-Mitchell-Analyse der Wachstumsrate des BIP und der Veränderung der verzögerten Geschäftserwartungen

Mittelwert über vier Referenzzyklen	BIP-Wachstum	Größenklasse				
		1	2	3	4	5
Amplitude im						
Aufschwung	2,9	3,4	4,1	4,3	3,0	5,4
Abschwung	–3,3	–2,8	–3,8	–4,1	–3,7	–4,8
Zyklus	6,1	6,1	7,9	8,4	6,8	10,3
Konformität im						
Tiefpunkt	0	0	50	100	100	50
Hochpunkt	–50	0	0	0	–100	50

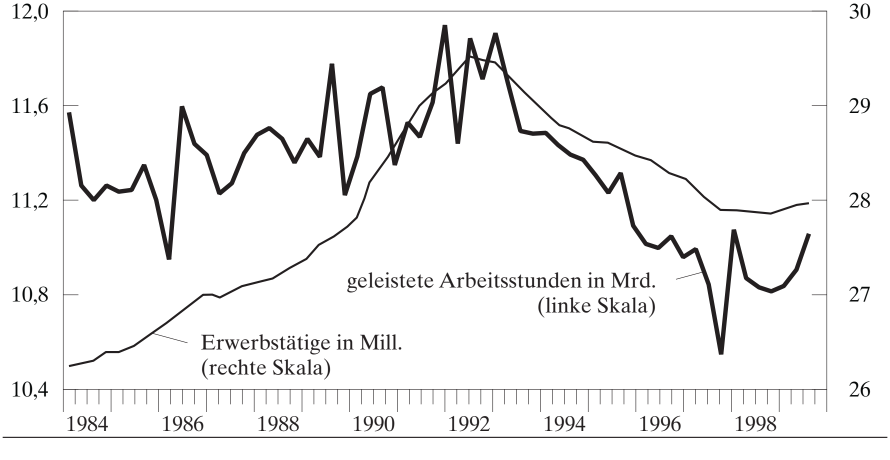
Tabelle 2 beschreibt auch die Amplitude der betrachteten *cycle relatives* im Aufschwung, Abschwung und im Zyklus. Tendenziell nimmt hier die Amplitude mit wachsender Unternehmensgröße zu, d.h. kleinere Unternehmen haben tendenziell stabilere Geschäftsbeurteilungen und Erwartungen.

5. Beschäftigungseffekte

Abschließend soll noch kurz die Frage untersucht werden, welchen Aufschluss die disaggregierten Unternehmensdaten über konjunkturelle Beschäftigungswirkungen erlauben. Insbesondere soll der Frage nachgegangen werden, ob mittelständische Unternehmen besondere Beschäftigungsakzente setzen. Zur Abgrenzung verwenden wir auch hier das beispielsweise vom Institut für Mittelstandsforschung verwendete Kriterium einer Beschäftigtenzahl von unter 500. (Die Europäische Union dagegen bezeichnet Unternehmen mit weniger als 250 Beschäftigten als kleine und mittelständische Unternehmen). Als Beschäftigungsindikatoren verwenden wir die Zahl der Erwerbstätigen und die der geleisteten Arbeitsstunden. Schaubild 13 zeigt zunächst die Niveaus dieser beiden Reihen.

Wie leicht ersichtlich ist, ist die Zahl der Erwerbstätigen von 1984 bis 1998 für eine konjunkturelle Analyse nicht geeignet, so dass auch hier die Wachstumsraten Analysegrundlage sein sollen. Dazu werden wieder die Werte der Wachstumsraten für Westdeutschland ab 1999 aus den gesamtdeutschen Werten geschätzt. Es ist natürlich nicht überraschend, dass die Wachstumsrate der Arbeitsstunden höher mit der Wachstumsrate des BIP korreliert ist, als dies

Schaubild 13
Arbeitsstunden und Erwerbstätige
1984 bis 1999

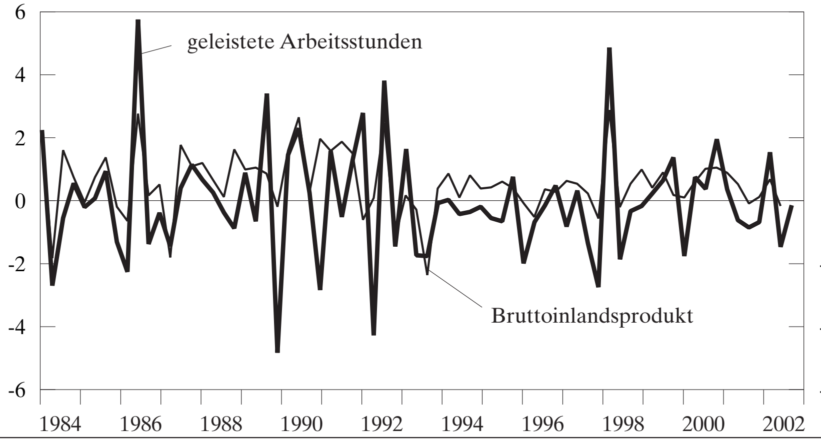


DOI https://doi.org/10.3790/978-3-428-52425-9 | Generated on 2025-02-22 16:18:59

Schaubild 14

Arbeitsstunden und BIP

1984 bis 2002; Veränderung gegenüber dem Vorjahr in %



für die Wachstumsrate der eher glatten Beschäftigtenzahlen gilt (Schaubild 14).

Das Neun-Punkt-Diagramm (Schaubild 15) zeigt klar einen engen Zusammenhang zwischen der Wachstumsrate der Stunden und der Wachstumsrate des BIP auf. Außerdem ist ersichtlich, dass sich die Wachstumsrate der Erwerbstätigen leicht nachlaufend verhält, denn sowohl nach einem Hoch- als auch nach einem Tiefpunkt kommt es erst mit leichten Verzögerungen zu quantitativ bedeutenden Reaktionen. Dieses Verhalten erscheint wegen des erheblich größeren Aufwands, den ein Arbeitgeber bei Neueinstellungen und Entlassungen im Vergleich zu dem Aufwand bei der Anweisung von anderen Arbeitszeiten betreiben muss, sehr plausibel.

Um einen genaueren quantitativen Einblick der Beschäftigungseffekte zu gewinnen, wird nunmehr für jede Größenklasse die Wachstumsrate der Erwerbstätigen auf eine Konstante und auf das *Niveau* der Geschäftsbeurteilung der jeweiligen Größenklasse regressiert. Es handelt sich dabei um eine Regression in „Zykluszeit", d.h. um eine Regression, die spezifisch auf das Burns-Mitchell-Instrumentarium abstellt. Die Regressionsgleichung lautet formal:

WR Erwerbstätige= c0 + c1* *cycle relative* Geschäftsbeurteilung (Niveau).

Für alle Größenklassen ist die Geschäftsbeurteilung hochsignifikant für die abhängige Variable „Wachstumsrate der Erwerbstätigen" (Tabelle 3). Wichtiger ist jedoch vielleicht das Ergebnis, dass das Erklärungspotenzial des Re-

Schaubild 15

Neun-Punkt-Diagramm der Veränderungsraten von Arbeitsstunden und Zahl der Erwerbstätigen

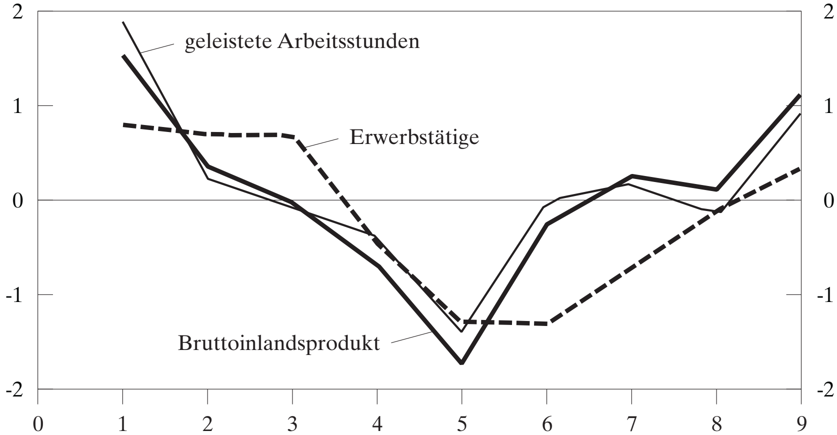


Tabelle 3

Regression der Wachstumsrate der Erwerbstätigen auf die Geschäftsbeurteilung in Zykluszeit

Regressor	t-Statistik	R^2	DW
Größenklasse 1	14,5	0.87	2,1
Größenklasse 2	14,2	0.87	2,0
Größenklasse 3	12,2	0.82	1,7
Größenklasse 4	12,3	0.83	1,9
Größenklasse 5	10,2	0.77	1,6

gressors abzunehmen scheint mit wachsender Unternehmensgröße (und parallel dazu das Bestimmtheitsmaß sinkt). Dies kann als eine gewisse Evidenz dafür angesehen werden, dass in der Tat der Mittelstand einen stärkeren Einfluss auf die konjunkturelle Beschäftigungsentwicklung ausübt als Großunternehmen.

6. Schlussbemerkungen

Drei wesentliche Resultate lassen sich den vorstehenden Analysen entnehmen, zwei davon methodischer, das andere eher inhaltlicher Natur:

- Zum Ersten scheint die Burns-Mitchell-Methodologie ein aussagekräftiges und mithin vielleicht zu Unrecht vernachlässigtes deskriptiv-statistisches Verfahren zu sein, das in der angewandten Konjunkturforschung komplementär zu den Standardinstrumentarien eingesetzt werden könnte. Insbesondere die Transformation der Kalenderzeit in Zykluszeit ermöglicht neue Einsichten in konjunkturelle Zusammenhänge. Aber auch das graphische Hilfsmittel der Neun-Punkt-Diagramme erweist sich als wertvoll, da es die Aufdeckung bestimmter Asymmetrien (z.B. unterschiedliches Prognosepotential an Hoch- und an Tiefpunkten) möglich macht.

– Zum Zweiten betont der Beitrag die Bedeutung disaggregierter Information. Je nach Fragestellung ist der Informationsgehalt der vom ifo erfragten Reihen zwischen den einzelnen Unternehmensgrößenklassen unterschiedlich. Dies ist fast eine Trivialität, aber die zwingende Konsequenz dieser Trivialität ist es, dass die Betrachtung ausschließlich der aggregierten Information informationsökonomisch ineffizient ist. Es liegt also nahe zu vermuten, dass die Abschätzung konjunktureller Entwicklungen qualitativ fortentwickelt werden kann, wenn verfügbare, aber öffentlich kaum beachtete disaggregierte Information stärker als bislang berücksichtigt wird. Dass dies nicht nur mit Burns-Mitchell-Methodik, sondern mit der Gesamtheit des zur Verfügung stehenden Instrumentariums der empirischen Wirtschaftsforschung geschehen sollte, braucht an dieser Stelle nicht ausdrücklich betont zu werden.
– Zum Dritten – und dies ist der eher inhaltliche Aspekt – legt die Auswertung der disaggregierten Information eine stärkere Forschung über kleine und mittelständische Unternehmen auch im konjunkturellen Umfeld nahe. Wenngleich die empirische Evidenz, die hier präsentiert wurde, keine allzu starken Schlussfolgerungen erlaubt, so ist die Mehrzahl der erzielten Resultate doch kompatibel mit der Behauptung, dass die wirtschaftliche Lage von kleinen und mittleren Unternehmen bessere Rückschlüsse auf die konjunkturelle Situation gestattet, als dies bei Großunternehmen der Fall ist.

Literatur

Bry, G. and C. Boschan (1971), Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. NBER Technical Paper 20. NBER, New York.

Burns, A.F. and W.C. Mitchell (1946), *Measuring Business Cycles.* NBER, New York.

King, R.G. and C.I. Plosser (1994), Real Business Cycles and the Test of the Adelmans. *Journal of Monetary Economics* 33: 405–438.

Klein, P.A. and G.H. Moore (1985), *Monitoring Growth Cycles in Market-Oriented Countries.* NBER, Cambridge, MA.

Maravall, A. and V. Gómez (2001), Seasonal Adjustment and Signal Extraction in Economic Times Series. In D. Peña, G.C. Tiao and R.S. Tsay (eds.), *A Course in Time Series Analysis.* New York: J. Wiley & Sons, 202–246.

Simkins, S.P. (1994), Do Real Business Cycle Models Really Exhibit Business Cycle Behavior? *Journal of Monetary Economics* 33: 381–404.

Anhang

In diesem Anhang sollen die wesentlichen Konzepte des Burns-Mitchell-Ansatzes möglichst genau definiert werden. Es handelt sich freilich nicht um mathematisch exakte Definitionen, sondern lediglich um Begriffsbestimmungen, die das Prinzip erläutern sollen, im Anwendungsfall aber einer weiteren Präzisierung bedürfen. Auch werden den Definitionen in der Anwendung weitere qualifizierende Anforderungen (z.B. über Mindestdauern des Referenzzyklus, Nähe benachbarter Wendepunkte etc.) beigefügt, die im Hauptteil der Arbeit größtenteils erwähnt, hier aber nicht mit aufgenommen werden, um den Text der Definitionen nicht über Gebühr aufzublähen.

Konjunktureller Wendepunkt
Für eine gegebene Menge ökonomischer Zeitreihen ist ein Zeitpunkt ein konjunktureller Wendepunkt, falls die Werte dieser Zeitreihen in einer geeignet bestimmten Umgebung dieses Zeitpunktes ein lokales Extremum annehmen. Untere Wendepunkte werden als *troughs*, obere als *peaks* bezeichnet.

Referenzzyklus
Der Referenzzyklus ist eine Folge von drei Zeitpunkten, von denen der erste und dritte obere konjunkturelle Wendepunkte (*troughs*) bezeichnen, der zweite aber den unteren konjunkturellen Wendepunkt (*peak*) darstellt. Der letzte untere Wendepunkt eines Referenzzyklus fällt mit dem ersten oberen Wendepunkt des nächsten Referenzzyklus zusammen. Die Folge aller Referenzzyklen während eines gegebenen Zeitraumes wird als Abbild der unbeobachtbaren Konjunktur angesehen, die dem Verlauf der beobachteten Zeitreihen zugrunde liegt.

Phase
Als Phasen werden Aufschwung und Abschwung unterschieden. Die Abschwungphase eines Referenzzyklus besteht aus der Zeit vom ersten oberen Wendepunkt bis zum unteren Wendepunkt, die Aufschwungphase aus der verbleibenden Zeit des Referenzzyklus.

Stadium
Ein Stadium ist eine Untereinheit eines Referenzzyklus. Es werden neun Stadien unterschieden, die eine vollständige (aber nicht disjunkte) Zerlegung eines gegebenen Referenzzyklus darstellen. Stadium I besteht aus den drei Monaten, deren mittlerer den oberen Wendepunkt enthält, Stadium V aus den drei Monaten, deren mittlerer den unteren Wendepunkt enthält und Stadium IX ist identisch mit Stadium I des folgenden Referenzzyklus. Die Zeit zwischen Stadium I und Stadium V (beginnend mit dem Monat nach dem oberen Wendepunkt und endend mit dem Monat vor dem unteren Wendepunkt) wird dann in die drei möglichst gleich großen Stadien II, III und IV disjunkt zerlegt. Analoges gilt für die Stadien VI, VII und VIII des Aufschwungs.

Zyklusbasis
Die Zyklusbasis einer gegebenen Zeitreihe für einen gegebenen Referenzzyklus ist der Mittelwert der Zeitreihe während des Referenzzyklus.

Cycle Relative
Der *cycle relative* einer gegebenen Zeitreihe in einem gegebenen Stadium eines gegebenen Referenzzyklus ist im Allgemeinen der Quotient aus dem Mittelwert der Zeitreihe während dieses Stadiums und der entsprechenden Zyklusbasis. In dieser Arbeit werden statt des Quotienten die absoluten Abweichungen des Mittelwerts der Zeitreihe während des gegebenen Stadiums von der entsprechenden Zyklusbasis als *cycle relative* bezeichnet. Der Grund für die Betrachtung von absoluten Abweichungen statt Quotienten liegt in der Verwendung von Zeitreihen, die als Wachstumsraten oder Salden vorliegen und um den Wert null schwanken.

Monatszuwachs
Der Monatszuwachs in einer gegebenen Phase ist die durchschnittliche Veränderung des *cycle relatives* einer gegebenen Zeitreihe während dieser Phase. Der Monatszuwachs von einem Stadium zum nächsten ist die durchschnittliche monatliche Veränderung des *cycle relatives* der Zeitreihe im Zeitraum von der Mitte des ersten Stadiums zur Mitte des zweiten Stadiums.

Amplitude
Die Abschwungamplitude ist die Differenz zwischen den *cycle relatives* einer gegebenen Zeitreihe im Stadium V und im vorgelagerten Stadium I. Die Aufschwungamplitude ist die analoge Differenz zwischen den *cycle relatives* in Stadien IX und V. Die Zyklusamplitude ist die Summe der Absolutbeträge von Aufschwung- und Abschwungamplitude.

Durchschnittliche Amplitude
Für eine gegebene Zeitreihe ist die durchschnittliche Amplitude der Mittelwert aller derartigen Amplituden über alle Referenzzyklen.

Konformität
Die Konformität ist ein Index, der ein Maß der Prozyklizität einer gegebenen Zeitreihe darstellt. Für einen gegebenen Zeitraum (Abschwung, Aufschwung, Stadium II, III, IV oder VI, VII, VIII) wird ein Wert von +100 angesetzt, wenn die Zeitreihe sich in diesem Zeitraum prozyklisch verhält, und –100, wenn sie sich antizyklisch verhält. Dabei bedeutet prozyklisch hier, dass der Zuwachs in dem Zeitraum positiv in der Aufschwungphase oder negativ in der Abschwungphase ist. Der Durchschnitt dieser Zahlen über alle Referenzzyklen ist die (Phasen- oder Stadien-)Konformität der Zeitreihe.

Zykluskonformität
Für eine gegebene Zeitreihe erhält ein gegebener Referenzzyklus den Indexwert +100, wenn der mittlere monatliche Zuwachs im Aufschwung größer ist als der mittlere monatliche Zuwachs im Abschwung, und –100 im umgekehrten Fall. Der Durchschnitt dieser Werte über alle Referenzzyklen ist die Zykluskonformität der Zeitreihe.

Katharina Morik and Stefan Rüping[1]

An Inductive Logic Programming Approach to the Classification of Phases in Business Cycles

1. Introduction

The ups and downs of business activities have been observed since a long time[2]. It is, however, hard to capture the phenomenon by a clear definition. The National Bureau of Economic Research (NBER) defines business cycles as "*recurrent sequences of altering phases of expansion and contraction in the levels of a large number of economic and financial time series*". This definition points at the multi-variate nature of business cycles. It does not specify many of the modeling decisions to be made. There is still room for a variety of concepts.

- What are the indices that form a phase of the cycle? Production, employment, sales, personal income, and transfer payments are valuable indicators for cyclic economic behavior. Are there others that should be included?
- Which measurements of indices are to be taken? Where the classical business cycle is expressed according to the level of indicators, the growth cycle is measured with respect to the deviation from the trend of indicators.
- What is the appropriate number of phases in a cycle? The number of phases in a cycle varies in economic models from two to nine. The NBER model indicates two alternating phases. The transition from one phase to the next is given by the turning points *trough* and *peak*. In the RWI model, a cycle

[1] This work has been partially sponsored by the Deutsche Forschungsgemeinschaft (DFG) collaborative research center 475 "Reduction of Complexity for Multivariate Data Structures". The authors thank Ullrich Heilemann, vice president of the Rheinisch-Westfälische Institut für Wirtschaftsforschung, for data of high quality and just as valuable suggestions as well as Claus Weihs and Ursula Sondhauss for raising our interest in the task and providing insight in its statistical nature.

[2] Amstad (2000) reports the first definition from Clement Juglar in 1860. She investigates several models of the business cycle and discusses their distinctions with respect to dating turning points of the business cycle.

consists of a lower turning point, an upswing, an upper turning point, and a downswing. Here, the turning points are phases that cover several months.

– Are all cycles following the same underlying rules or has there been a drift of the rules? What is the appropriate sample for classifying current business data?

All modeling decisions are to be (comparatively) validated with respect to economic theory and to business data. One approach to validation is the formalization by macro-economic equations. A model of business activities is calculated *ex post* and the deviation of the results of the equations from the observed values assesses the model. For instance, the business cycle model of the Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen) only deviated 1.2 per cent for the spring 2000 state of affairs in Germany (RWI Essen 2000). The main focus here lies on the prediction of level or growth of business activities. We do not contribute to this approach.

The other approach is an empirical one, in which statistical methods are adjusted to business data and used for prognoses. Again, the statistical models are validated on past data. We are concerned with the development and comparison of methods for the empirical modeling of business cycles. Empirical methods are particularly demanded for the task of dating turning points or phases of the business cycle. This task is less clearly defined than the task of predicting business activities, because business cycles themselves are basically a theoretical model to explain the variation in business data. So the quality of a classification of business cycles cannot be judged by simple oberservations, but has to be evaluated in the context of the greater model.

Our question is: Which methods can support modeling and validating models of the business cycle? More precisely: Can inductive logic programming support economists in dating and predicting phases of the business cycle? We may re-formulate the questions into two general problems:

– *Dating*: Given current (and past) business measurements, in which phase is the economy currently? In other words, the current measurements are to be classified into the phases of a business cycle.

– *Prediction*: Given current (and past) business measurements, what do we expect next?

Linear discriminant analysis has been proposed as the baseline of empirical models (personal communication with C. Weihs at the workshop). Univariate rules were learned that used threshold values for separating phases. The accuracy of the 18 learned rules was 54% in cross validation. Using this result as the baseline means that the success of any other method has to be shown in comparison to this accuracy. It has been investigated how the classification can be enhanced by the use of monthly data (Heilemann, Münch 2001). More sophisticated statistical models have been developed and achieved 63% accuracy (Weihs, Sondhauß 2001). However, even this substantial enhancement still reflects how hard it is to classify business phases correctly.

The difficulty of the problem lies in its multi-variate nature, which follows from the definition of business cycles. Moreover, the business cycle cannot be observed directly and main factors of influence may well be hidden. Hence, we may want to incorporate economic knowledge (theory) into business cycle data analysis. In fact, the advanced Markov Switching model as it was used by Sondhauss/Weihs expresses knowledge about the past phase and the transition probability to the next phase (Weihs, Sondhauß 2001). Also the approach which we describe in this paper, exploits domain knowledge. Here, economic knowledge is used to restrict the space of possible rules in order to exclude those that would not make sense or are trivially true.

In this paper, we investigate the applicability of inductive logic programming to the problem of dating phases of a business cycle. We were given quarterly data for 13 indicators concerning the German business cycle from 1955 to 1994, where each quarter had been classified as being a member of one of four phases (Heilemann, Münch 1996). The indicators are:

IE	real investment in equipment (growth rate)
C	real private consumption (growth rate)
Y	real gross national product (growth rate)
PC	consumer price index (growth rate)
PYD	real gross national product deflator (growth rate)
IC	real investment in construction (growth rate)
LC	unit labour cost (growth rate)
L	wage and salary earners (growth rate)
Mon1	money supply M1
RLD	real long term interest rate
RS	nominal short term interest rate
GD	government deficit
X	net exports

We experimented with different discretizations of the indicator values. The discretization into ranges (levels) of values was also used in order to form time intervals. A sequence of measurements within the same range is summarized into a time interval. For instance, the money supply being *high* for quarters 8 to 18 is summarized by the fact mon1(i1,high), where i1 corresponds to the time interval from 8 to 18. Hence, the time intervals differ from indicator to indicator. Relations between the different time intervals express precedence or domination of one indicator's level to another ones level. We also compared the two phase with the four phase business cycle. In summary, the following three models were inspected:

– business cycle with four phases, without time intervals,

– business cycle with four phases, time intervals,
– business cycle with two phases, without time intervals.

Particular attention was directed towards the appropriate sample size for the dating problem. The homogeneity of the data set of business cycles with two phases was investigated. The hypothesis being that at the end of cycle 3 (i.e., third quarter of 1971) the rules for dating phases could change.

2. Inductive Logic Programming

Inductive Logic Programming (ILP) establishes the intersection of logic programming and machine learning. A logic program is learned from observations by inductive reasoning. The logic program expresses a theory in the form of facts and rules in a restricted first-order logic. The theory describes diverse concepts together with the relations among them. This contrasts with propositional logic, where only one concepts and its sub-concepts can be modeled. A simple example illustrates this.

Given the observations

mother	*(ann,*	*brigid).*	*mother*	*(alice,*	*bonnie)*
mother	*(brigid,*	*cecilie).*	*mother*	*(bonnie,*	*christie)*
old	*(ann).*		*old*	*(alice).*	
grandmother	*(ann,*	*cecilie).*	*grandmother*	*(alice,*	*christie)*

ILP may learn the rule

$$mother\ (X, Z), mother\ (Z, Y) \rightarrow grandmother\ (X, Y).$$

Whereas a propositional learner cannot exploit the relations but can only learn the heuristic:

$$old\ (X) \rightarrow grandmother\ (X, Y)$$

A logic program is directly executable. The learned rules derive the conclusion from new facts. For instance, the learned grandmother rule derives

grandmother(agatha,carol).

as soon as the facts are stated:

mother(agatha,beth). mother(beth,carol).

The expressive power of first-order logic proves especially successful in relations between intervals. Explicitly the starting and end point of an interval can be stated together with the relations between intervals. For instance, direct precedence can easily be expressed between time intervals, here between the time intervals from $T1$ to $T2$, from $T2$ to $T3$, and from $T3$ to $T4$:

cooking($C,T1,T2$),*serving*($S,T2,T3$),*eating*($E,T3,T4$) (direct precedence)
Similarly, inclusion and overlap of intervals is written.
cooking($C,T1,T4$),*serving*($S,T2,T3$),$T1 \leq T2, T3 \leq T4$ (inclusion)
cooking($C,T1,T3$),*serving*($S,T2,T4$),$T1 \leq T2, T3 \leq T4$ (overlap).

It has been shown that the time relations of Allen's calculus (Allen 1984) can be expressed in the form of a logic program (Rieger 1996; 1998).

Note, that the first-order logic rules are inherently multi-variate. Distinct events can be expressed with their properties. For instance, the activities cooking (C), serving (S), and eating (E) can be described independently from another, naming the involved places ($Y1$,$Y2$), persons $X1$ to $X5$, and their properties (*salary*) and relations (*mother, father*).

cook	($X1$,	C).	*guest*	($X3$,	E).
salary	($X1$,	$W1$).	*guest*	($X4$,	E).
kitchen	($Y1$,	C).	*guest*	($X5$,	E).
recipe	(Z,	C).	*mother*	($X3$,	$X5$).
waiter	($X2$,	S).	*father*	($X4$,	$X5$).
salary	($X2$,	$W2$).			
diningRoom	($Y2$,	S).			

Of course, this representation can be compiled down to propositional logic, if the number of objects is finite (Lavrac, Dzeroski 1994). However, the ease of understanding is lost in the compilation. The understandability of first-order logic eases the formulation of hypotheses that the learning algorithm should test on the data. We shall see, how user-specified sets of hypotheses are represented and tested by the Rule Discovery Tool.

A last advantage of ILP to be mentioned is the explicit statement of background knowledge, commonly in terms of facts. In the grandmother example, the facts stating the grandmother role are the examples and the facts stating the mother role are the background knowledge.

We may now state the task of concept learning within ILP formally.

Concept Learning or Learning Classifications

Given positive and negative examples $E = E^+ \cup E^-$ in a representation language L_E and background knowledge B in a representation language L_B, **find** a hypothesis H in a representation language L_H, which is a (restricted) first-order logic, such that

(1) $$B, H, E^+ \not\models \Box \text{ (consistency)}$$

(2) $$B, H \models E^+ \text{ (completeness of } H\text{)}$$

(3) $$\forall e^- \in E^- : B, H \not\models e^- \text{ (accuracy of } H\text{).}$$

2.1 MOBAL

MOBAL is a workbench which allows users to easily enter facts and rules, detects inconsistencies in the knowledge base, and proposes minimal changes to facts and rules in order to make it consistent (Morik et al. 1993). In addition to the support of users in building up a knowledge base, the rule discovery tool automatically learns rules from facts and adds the learned rules to the knowledge base.

2.1.1 The Rule Discovery Tool RDT

For learning rules from facts, the Rule Discovery Tool RDT forms all possible rules according to a user given hypothesis space (Kietz, Wrobel 1991). The user specifies rule schemata. A rule schema has predicate variables that can be instantiated by predicates of the domain. An instantiated rule schema is a rule. Rule schemata are partially ordered according to their generality. For our learning task of dating business data according to the business cycle, we first used the following rule schemata:

m1 (Index1, Value, Phase):
$Index1(T, V), Value(V) \rightarrow Phase(T)$

m2 (Index1,Value,Index2, Phase):
$Index1(T, V), Value(V), Index2(T, V) \rightarrow Phase(T)$

m3 (Index1, Value1,Index2,Value2,Phase):
$Index1(T, V1), Value1(V1), Index2(T, V2), Value2(V2), opposite(V1, V2) \rightarrow Phase(T)$

Here, $m1$ is more general than $m1$ and $m2$. The predicates that fit to instantiate the predicate variable *Index* are the 13 indicators of the economy (see above), e.g., *lc*(*Time,Value*) for unit labour cost. The predicates that fit to instantiate the predicate variable *Value* are *low,medium,high* and express the discretization of the real values of the indicators. The phase variable can be instantiated by *down*,*ltp*,*up*,*utp* for four phases or by *down*,*up* for two phases of the business cycle. The *opposite* predicates is used to write which qualitative value intervals are excluding each other. The background knowledge consists of such facts:

opposite(*low,medium*).
opposite(*high,medium*).
opposite(*high,low*).

Hence, the hypothesis space consists of all indicators or combinations of two indicators that allow to predict the phase of the business cycle. Uni-variate rules with just one indicator are excluded, because they are not considered to be sensible. Where $m2$ states that two indicators have to have values within the same range (e.g., both are low), $m2$ states that two indicators must have opposite value ranges. The three rule schemata here enforce the selection of the most informative indicators. By giving more complex rule schemata, the user enables RDT to learn more complex rules. The rule schemata are the means by which the set of interesting rules is specified.

RDT's learning procedure consists of two steps: hypothesis generation and testing. In a top-down, breadth-first manner, all possible instantiations of the rule schemata are generated and tested according to an acceptance criterion on the basis of all facts. For instance, the following rules which instantiate $m1$ and were learned in one of our experiments:

$$mon1(T,V), medium(V) \rightarrow up(T)$$
$$lc(T,V), low(V) \rightarrow up(T)\,.$$

The following instantiation of $m2$ has been learned in another experiment:

$$ic(T,V), medium(V), pc(T,V) \rightarrow down(T)\,.$$

A illustration for $m3$ is the following learned rule:

$$rs\,(T,V1), medium(V1), x(T,V2), low(V2) \rightarrow down(T).$$

If a rule has enough support but too many non supporting examples, it is considered too general. Later on, it becomes a partial instantiation of a more specific rule schema if this exists. If a rule does not have enough support, it is considered too specific. In this case, the rule need not be specialized further, since this cannot increase the number of supporting examples. RDT safely prunes the search in this case. RDT learns all valid rules that fit the rule schemata. Hence, a rule which is not learned, definitely does not hold given the facts.

2.1.2 Rule Inspection

For both, user given and learned rules, we might be interested in their coverage of examples and in their redundancy. The coverage can be measured by the percentage of the positive examples for a concept, that is covered by the rule. For instance, if there are 58 time points classified as *down*, we are interested in how many of them are covered by a certain rule predicting a downswing. We are also interested in the number of examples that are covered by all the rules together. The rule inspection of MOBAL indicates for each rule, how many of the input facts (in our experiments, time points that are classified into a certain phase of the business cycle) are correctly classified by the rule. For a set of rules, MOBAL indicates how many examples are covered by all the rules, or, in other words, how many examples cannot be explained by the rules. Redundancy of rule sets can be determined intensionally or extensionally. The intensional redundancy refers to the logical models that form the semantics of the rule set. The extensional redundancy of a rule set refers to the examples that are covered by the rules: if the same examples are covered by several rules, these rules are extensionally redundant (Sommer 1996). Rules that are 100% extensionally redundant are not necessarily intensionally redundant. They might cover new examples not yet seen which would not be explained by another rule. It is up to the domain expert to assess which of the extensionally redundant rules should be kept within the knowledge base.

2.1.3 Knowledge Revision

A set of rules can easily become contradictive. Most user-given rule sets first show contradictions because the user is not aware of all implications of all rules. Also learned rule sets can become contradictive. The most frequent contradiction occurs when applying the rules learned from a set of data (i.e. the training set) to another set of data (i.e. the test set): the predicted phase differs from the one given by the expert. In general, the detection and revision of inconsistencies is a hard problem. Due to the well-defined semantics of MOBAL and its restrictions of first-order logic, the problem could be solved (Wrobel 1994a, b). The system determines the facts and rules that are involved in the contradiction. It calculates all minimal changes to the knowledge base that would make it consistent again. The user chooses among the proposed changes and the system revises the knowledge base accordingly. Hence, the user is supported in building up a knowledge base about a domain by integrating rule sets, either learned or input. In particular, the user may input domain (causal) knowledge and the system watches that no learned rule contradicts the theoretical insight.

3. Experiments on German Business Cycle Data

Our leading question was whether ILP can support economists in developing models for dating phases of the business cycle. Given the quarterly data for 13 indicators concerning the German business cycles from 1955 to 1994 where each quarter is classified as member of one of four phases, we used all but one cycle for learning rules and tested the rules on the left-out cycle. The leave-one-cycle-out test assesses the *accuracy* (how many of the predicted classifications of quarters corresponded to the given classification) and the *coverage* (how many of the quarters received a classification by the learned rules).

We now come back to the questions raised in the introduction. The learned rules automatically select pairs of relevant indicators. Hence, all learning experiments contribute to the question, which indicators actually influence the classification into one phase of the business cycle. Since the data we have are measuring the growth, the learning results refer to the growth cycle. However, we experimented with automatically finding ranges or levels of values of the indicators in order to base the learning results on a view of the business cycle which favours the level of indicator values (Section 3.1). In order to tackle the question about the number of phases in business cycles, we have modeled four phases (Sections 3.2 and 3.3) and two phases (Section 3.4). An additional modeling decision needs to be made according to the handling of the time aspect in the data. In two experiments (Section 3.2 and Section3.4), we just used the quarters as time points. No time intervals were formed. The rule schemata are the ones shown in Section 2.1.1. Hence, the rules only classify a quarter based on the measurements of this quarter. In a third experiment, we formed time intervals for the indicators and learned rules between them (Section 3.3).

3.1 Discretization

Before ILP can be applied, the originally real-valued time series of indicator values have to be transferred into discrete-valued temporal facts about this indicators. The goal of discretization is to provide the learning algorithm with data from which it can generalize maximally. This means, the discretization must be general enough such that rules learned from one situation can be transferred to another situation but specific enough such that non-trivial rules can be found. An example for a too specific discretization is to assign different values to every observation, an example for a too general discretization is to assign the same value to every observation. We use the number of generated facts to judge the quality of a discretization.

Actually, the task of discretization consists of two different subtasks:

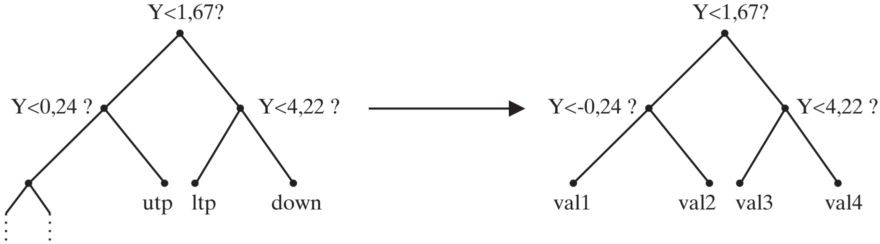
– *Discretization of Values*: split the continuous range of possible values into finitely many discrete values, e.g. by using equidistant thresholds or calculating suitable quantiles. For example, a gross national product of 4.93 in the fifth quarter could be expressed as the fact $y(5,\mathit{high})$.

– *Interval segmentation*: for a given time series, find a segmentation of the time points into maximal sub-intervals, such that the values of the series in this interval share a common pattern, e.g. by approximating the time series by piecewise constant or piecewise linear functions. For example, the time series of gross national products Y=(10.53,10.10,9.21,5.17,4.93) could be described as the temporal facts $y(1,3,\mathit{high})$,$y(4,5,\mathit{medium})$, but can also be described as $y(1,5,\mathit{decreasing})$.

Interval segmentation can be viewed as discretization of the temporal values, therefore in this chapter we will use the name discretization as a generic term for both discretization of values and interval segmentation.

Figure 1

Decision Tree and its Induced Discretization into val1...val4



The two subtasks are closely intertwined: Discretized data can be very easily segmented by joining consecutive time points with identical discretization. Also, segmented data can be discretized by building a discretization based on the patterns that lead to the segmentation. In this work, we chose the first approach to discretize the data, first because it is simpler and secondly because the indicators are already given free of trends (growth rates etc.), so it can assumed the relevant information lies in the value of the indicator alone.

To improve the quality of the discretization, we can also use the information that is given by the class of the examples (Zighed et al. 1999). In this case, we used C4.5 (Quinlan 1993), a decision tree learner, to induce decision trees about the cycle phase based on only one indicator. The resulting trees were cut off at a given level and the decisions in this resulting tree were used as discretization thresholds (Figure 1). Decision trees of depth 2, i.e. using 4 discrete values, proved to build a suitable number of facts.

A closer look at the resulting discretization showed that in certain cases, the indicators had a very high variation, which lead to many intervals that contained only one time point. In this case, the relevant observation may not be the value of the indicator, but the fact that this indicator was highly variating, i.e. that no definite value can be assigned to it. This can be expressed by a new fact $indicator(T1, T2, unsteady)$, which replaces the facts $indicator(T1, T1+1, value_1)$, $indicator(T1+1, T1+2, value_2), \ldots, indicator(T2-1, T2, value_n)$.

3.2 Modeling Four Phases Without Time Intervals

The data correspond to six complete business cycles, made of four phases each. For the upper and lower turning point phases, no rule could be learned. Only for the upswing, each learning run delivered rules. Here are some examples of learned rules:

$$gd(T,V),\ pc(T,V), low(V) \rightarrow up(T)$$

stating that a low government deficit and a low consumer price index determine the phase as an upswing.

$$c(T,V), l(T,V), medium(V) \rightarrow up(T)$$

stating that a medium private consumption and a medium number of wage and salary earners classify a quarter as belonging to an upswing.

$$rld(T,V), mon1(T,V), low(V) \rightarrow down(T)$$

stating that a low long term interest rate and a low money supply can be used to date a downswing.

$$rs\,(T,V), ic(T,V), medium(V) \rightarrow down(T)$$

stating that a medium nominal short term interest rate together with a medium investment in construction point at being in the downswing.

Illustrating the rule inspection, we show the result for the first rule, called r171 in the leave-fifth-cycle-out learning run (Table 1). The difference between the total number of facts about $up(t)$ and the input occurrences of $up(t)$ is explained by the forward inferences of the (learned) rules. They derive further facts not given in the input.

Table 1

Statistics on Rule r171

Number of rules with same conclusion – up(t):	10
Coverage	
Total:	
Number of occurences of up(t):	77
Number of occurences covered by r171:	43
r171s coverage of all occurences:	55.8442 %
Number of occurences covered by all rules:	59
Total coverage of all occurences:	76.6234 %
Of inputs:	
Number of input occurences of up(t):	59
Number of inputs covered by r171:	35
r171s coverage of inputs:	59.322 %
Number of inputs covered by all rules:	41
Total coverage of inputs:	69.4915 %
Redundancy	
Total:	
Number of occurences also covered by other rules:	28
r171s internal redundancy (redundant/covered):	65.1163
On inputs:	
Number of inputs also covered by other rules:	27
r171s internal redundancy (redundant/covered) on inputs:	77.1429

Table 2

Results in the Four Phase Model Using Time Points

Cycle	Accuracy	Coverage	No.of learned rules
LOO1	0.125	0.25	13 upswing
LOO2	0.5	1.0	12 upswing
LOO3	0.462	0.462	10 upswing, 2 downswing
LOO4	0.375	1.0	11 upswing
LOO5	0.696	0.696	10 uspwsing, 1 downswing
LOO6	1.0	0.36	1 upswing
Average	0.526	0.628	total: 60

If no rule states that the classification is exclusive, then no contradiction will be detected between the input fact *utp*(31) and *up* or *utp*(30) and *down*(30). Hence, for testing, we entered rules of the form:

$utp(t) \rightarrow not(up(t))$

Then, we also find a misclassification.

Contradictory instances covered by rule r171: auf(31) – [1000,1000].

In fact, time point 31 (corresponding to the second quarter of 1963) starts the upswing and time point 30 (first quarter of 1963) finalizes the downswing. The first two quarters of 1963 are classified as the lower turning point. Misclassifications at the turning points are strikingly more frequent than in other phases. For the downswing, only two learning runs, namely leaving out cycle 3 and leaving out cycle 5, delivered rules. Table 2 shows the results.

The results miss even the baseline of 54% in the average. Leaving out the fifth cycle (from 1974 until 1982) delivers the best result where both, accuracy and coverage, happen to approach 70%. This might be due to its length (32 quarters), since also in the other experiment dealing with four phases the prediction of upper turning point and upswing is best, when leaving out the fifth cycle. Since the sixth cycle is even longer (45 quarters), we would expect best results in LOO6 which is true for the accuracy in this experiment. In the other experiment with four phases, the accuracy is best for upswing in LOO6 and second best for it in LOO5.

3.3 Modeling Four Phases with Time Intervals

Let us now see, whether time intervals can improve the learning results. We have used the discretization of the indicator values for the construction of time intervals. As long as the indicator value stays within the predefined level, the time interval is continued. As soon as the indicator value exhibits a level change, the current time interval is closed and the next one is started. We end up with facts of the form *Index*(*I,Range*), and for each time point within the

Figure 2

The Temporal Relations *Contains* and *Overlaps*

time interval *I* a fact stating that this time point *T* (i.e. quarter) lies in the time interval *I: covers*(*I, T*).

We then described the relations between different time intervals by means of Allen's temporal logic (Allen 1984). From the 13 possible relationships between time intervals, we chose *contains* and *overlaps*. The relation *contains*($I1$,$I2$) denotes a larger interval $I1$ in which somewhere the interval $I2$ starts and ends. *contains*($I1$, $I2$) is true for each time point within the larger interval $I1$. *overlaps*($I1$, $I2$) is true for each time point of the interval $I1$ which starts before $I2$ is starting (Figure 2). We left out the other possible relations, because they were either too general or too specific to be used in a classification rule or would violate the constraint, that only information about past events can be used in the classification. For example, a relation that would require that the end point of one interval was identical to the starting point of another interval would be too specific and a relation that would only require that an interval would happen before another interval, regardless of the amount of time in between, would be too general. The time intervals were calculation before the training started. The rule schemata were defined such that they link two indicators with there corresponding time intervals.

m1 (Index1, Value1, Value2, Phase):
$Index1(I1, V1), Value1(V1), covers\,(I1, T),$
$Index2(I2, V2), Value2(V2), covers\,(I2, T \rightarrow Phase(T)$

m2 (Index1, Value1,Index2,Value2,Rel, Phase):
$Index1(T, V1), Value1(V1), covers\,(I1, T),$
$Index2(T, V2), Value2(V2), Rel(I2, I1) \rightarrow Phase(T)$

$m1$ substantially not differs from $m3$ in the preceeding model (time point model). It finds two indicators which determine the phase on the basis of a quarter which is shared by both time intervals.

$m2$ is more special in that it requires the time intervals of the two indicators to either overlap or include each other. Instantiations of $m2$ express rules where the behavior of one indicator must preceed or embrace the other indicator's behavior. These more specific rule schemata were intended to find rules for the turning phases, where no rules were learned in the previous experiment. In

Table 3

Results in the Four Phase Model Using Time Intervals

Cycle	Phase	Accuracy	Coverage	No. learned rules
LOO1	upswing	0.167	1	73
	downswing	–	0	1
	utp	–	0	0
	ltp	–	0	2
LOO2	upswing	–	0	103
	downswing	–	0	3
	utp	–	0	2
	ltp	–	0	0
LOO3	upswing	0.461	1	87
	downswing	1	0.200	2
	utp	0	0	2
	ltp	–	0	2
LOO4	upswing	0.167	1	59
	downswing	0.333	1	7
	utp	–	0	0
	ltp	–	0	4
LOO5	upswing	0.481	1	88
	downswing	0	0	3
	utp	–	0	0
	ltp	0.75	0.857	4
LOO6	upswing	0.667	0.296	6
	downswing	0.243	1	2
	utp	–	0	0
	ltp	–	0	0
Average	upswing	0.388	0,716	69.3
	downswing	0.194	0.500	3
	utp	0	0	0.667
	ltp	0.75	0.143	2

fact, rules for the upper turning point, upswing, and downswing were learned, but no rules could be learned for the upper turning point.

This rule states, that a period with high consumer price index growth, that is overlapped by a period of high growth rate in the private consumption, is indicative of an upswing:

$$pc(I1, V1), high(V1), covers\,(I1, T),$$
$$c(I2, V2), high(V2), overlaps\,(I2, I1) \rightarrow up(T)$$

The next rule states, that a downswing happens, if during a period with medium growth in the number of wage and salary earners, the short term interest rate is high:

$$l(I1, V1), medium(V1), covers\,(I1, T),$$
$$rs\,(I2, V2), high(V2), contains\,(I2, I1) \rightarrow down(T).$$

Another intention behind the time interval modeling was to increase the accuracy of the learned rules. Indeed, rules for the upper turning point could be learned with the average accuracy of 75% in the leave-one-cycle-out runs. However, the accuracy for upswing decreased to 34% in the average. Hence, overall the time interval model did not enhance the results of the time point model in as much as we expected (Table 3).

3.4 Modeling Two Phases

Theis/Weihs (1999) have shown, that in clustering analyses of German macroeconomic data at most three clusters can be identified. The first two cluster roughly correspond to the cycle phases of upswing and downswing and the eventual third cluster corresponds to the time period of the oil-crisis around 1971. This suggests, that two phases instead of four may be more suited for the description of business data. Therefore, in our third experiment we mapped all time points classified as upper turning point to upswing and all quarters of a year classified as lower turning point to downswing.

We then applied the rule schemata of the first experiment. An example of the learned rules is:

$$ie(T,V1), low(V1), c(T,V2), high(V2) \rightarrow down(T)$$

stating that a low investment into equipment together with high private consumption indicates a downswing.

Again, leaving out the fifth or the sixth cycle gives the best results in the leave-one-cycle-out test. Accuracy and coverage are quite well balanced (Table 4).

These learning results are promising. They support the hypothesis that a two phase model is of advantage for the dating task. Concerning the selection of indicators, the learning results show that all indicators contribute to the dating of the phase. However, the short term interest rate does not occur in three of the rule sets. Consumption (both the real value and the index), net exports,

Table 4

Results in the Two Phase Model Using Time Points

Cycle	Accuracy	Coverage	No. learned rules
LOO1	0,8125	0,795	9 up, 69 down
LOO2	0,588	1,0	17 up, 35 down
LOO3	0,823	0,571	2 up, 15 down
LOO4	0,8	0,35	6 up, 8 down
LOO5	0,869	0,8	10 up, 39 down
LOO6	1,0	0,701	6 up, 41 down
Average	0,815	0,703	total 50 up, 207 down

money supply, government deficit, and long term interest rate are missing in at least one of the learned rule sets. For the last four cycles, i.e. leaving out cycle 1 or cycle 2, some indicators predict the upswing without further conditions: high or medium number of salary earners (*l*), high or medium investment in equipment (*ie*), high or medium investment in construction (*ic*), medium consumption (*c*), and the real gross national product (*y*). It is interesting to note, that a medium or high real gross national product alone classifies data into the upswing phase only when leaving out cycle 1,2, or 4. Since RDT performs a complete search, we can conclude, that in the data of cycle 1 to cycle 4, the gross national product alone does not determine the upswing phase. Further indicators are necessary there, namely money supply (*mon*1) or consumer price index (*pc*).

3.5 Concept Shift

Starting from the two-phase model, we analyzed the homogeneity of the business cycle data. We want to know whether there are rules that are learned in all training sets, or, at least, whether there are rules that are more frequently learned than others. There is no rule which was learned in all training sets. Eight rules were learned from three training sets. There is one rule, which was learned in four training sets, namely leaving out cycle 1, cycle 4, cycle 5, or cycle 6:

$$rld(T,V), l(T,V), low(V) \rightarrow down(T).$$

We now turn around the question and ask: which training sets share rules? Eighteen rules were shared in the training sets leaving out cycle 5 and leaving out cycle 6. Four of the rules predict an upswing, fourteen rules predict a

Table 5

Two Distinct Samples (Concept Shift)

Cycles 1 – 3			Cycles 4 – 6		
Cycle	Accuracy	Coverage	Cycle	Accuracy	Coverage
LOO1	0,765	0,555	LOO4	0,501	0,937
LOO2	0,634	0,715	LOO5	0,795	1
LOO3	0,625	0,889	LOO6	0,509	0,97
Average	0,635	0,720	Average	0,602	0,846
Cycles 1 – 4			**Cycles 5 – 6**		
Cycle	Accuracy	Coverage	Cycle	Accuracy	Coverage
LOO1	0,592	0,727	LOO5	0,555	0,97
LOO2	0,62	0,944	LOO6	0,55	1
LOO3	0,62	1			
LOO4	0,596	0,75			
Average	0,61	0,855	Average	0,552	0,985

downswing. This means, that cycles 1 to 4 have the most rules in common. The data from the last quarter of 1958 until the third quarter of 1974 are more homogeneous than all the data from 1958 until 1994. When leaving out cycle 1 or cycle 2, eleven rules occur in both learning results. This means, that cycles 3 to 6 have second most rules in common. The data from the second quarter of 1967 until the end of 1994 are more homogeneous than all data together. This raises the question of the sample size for the dating problem (RWI Essen 2000):page 16. Klinkenberg has investigated methods for handling *concept drift* by adaptively selecting the sample size for prediction and classification (Klinkenberg 2001; Klinkenberg, Joachims 2000). Concept drift means that a concept changes over time. Concept shift is more specific and means that a concept changes at a certain point in time. Here, we investigate whether a concept shift has occurred in business cycles.

We perform the same learning task on two disjoint data sets. We split the overall data set into two parts, cycles 1 to 3 and cycles 4 to 6. We apply training and leave-one-cycle-out testing to each part. As you can see in Figure 6, accuracy as well as coverage decreased in the two distinct sets. Where the average accuracy has been 81% using all cycles, it is now decreased to 64% in the first three cycles and to 60% in the last three cycles.

If we split the overall data set into cycles 1 to 4 and cycles 5 and 6, the accuracy decreases for the first cycles to 61 % and for the last two cycles even to 55,2%. Then, we check whether the decreased accuracy is due to the smaller size by putting together cycles from early and recent years and see whether this also decreases accuracy. We selected cycles 2, 4, and 6 and performed the leave-one-cycle out training and testing. Surprisingly, the accuracy was 78% in the average and the coverage was 76,5%. Hence, we conclude that we cannot detect a concept shift.

4. Conclusion and Further Work

Coming back to the questions asked in the introduction, our research has delivered some answers and some new questions. Let us start with the answers.

- ILP offers opportunities for the analysis of business cycle data. It is easy to interpret the results and the learned rules can be inspected with respect to redundancy and contradictions. The multi-variate nature of ILP and the automatic selection of most relevant indicators fits the needs of dating problem. However, numerical processes are not captured by ILP but a discretization must preceed ILP processing.
- Although some indicators are more dominant than others, no subset of the given indicators could be formed. All the thirteen indicators contribute to the dating of the phase.

- The two-phase model of the business cycle clearly outperformed the four-phase model. Where the best average accuracy in the four-phase model was 53%, the average accuracy of the two-phase model was 82%.
- There is no clear concept shift between cycle 3 and cycle 4 (around mid of 1971).

It still needs to be investigated, what indices form a phase of a cycle. In this paper, we restricted ourselves to analyze the relationship between two indices, but there might be easily three or more economic indices that describe a cycle phase. Also, there might be other indices that are more informative than the 13 indicators in our data set.

We used discretization in a straight-forward manner by creating the interval segmentation based on the discretization of values. This can be extended by using piecewise constant or piecewise linear regression to get the interval segmentation directly. However, in this approach it is still unclear, how the slope of an approximating linear function can be interpreted. Because understandability is a main goal of ILP, a discretization needs to be found that gives both accurate results and is meaningful in the experts theory. This discretization might also consist of more complex patterns like peaks or valleys or patterns with outliers. Algorithms that find these patterns (Berndt, Clifford 1996) can be used as to preprocess the time series.

Other enhancements include the use of experts' background knowledge, a smoother partition into two phases, and a closer look at the appropriate sample size for dating phases.

References

Allen, J.F. (1984), Towards a general theory of action and time. *Artificial Intelligence* 23: 123–154.

Amstad, M. (2000), *Konjunkturelle Wendepunkte*: Datierung und Prognose. Chronologie unterschiedlicher Wendepunkttypen und Entwicklung eines Frühwarnsystems mittels Markov-Switching-Modellierung von Schweizer Unternehmensumfargedaten. Dissertation, Universität St.Gallen.

Berndt, D. J. and J. Clifford (1996), Finding patterns in time series: A dynamic programming approach. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (eds.), *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press and MIT Press, chapter 3, 229–248.

Heilemann, U. and H.J. Münch (1996), West German Business Cycles 1963–1994: A Multivariate Discriminant Analysis. In Selected Papers Submitted to the CIRET-Conference in Singapore. CIRET-Studien 50. Ifo, München, 219–250.

Heilemann, U. and H.J. Münch (2001), Classification of German Business Cycles Using Monthly Data. SFB 475 Technical Report 8/2001. University of Dortmund.

Kietz, J.-U. and St. Wrobel (1991), Controlling the complexity of learning in logic through syntactic and task-oriented models. Arbeitspapiere der GMD 503. GMD, St. Augustin.

Klinkenberg, R. (2001), Using labeled and unlabeled data to learn drifting concepts. In M. Kubat and K. Morik (eds.), *Workshop notes of the IJCAI-01 Workshop on Learning from Temporal and Spatial Data*. Menlo Park, CA: IJCAI, AAAI Press, 16–24.

Klinkenberg, R. and T. Joachims (2000), Detecting concept drift with support vector machines. In P. Langley et al. (ed.), *Proceedings of the Seventeenth International Conference on Machine Learning* (ICML-2000). San Francisco, CA: Kaufmann, 487–494.

Lavrac, N. and S. Dzeroski (1994), Inductive Logic Programming – Techniques and Applications. Artificial Intelligence 148. Hertfortshire: Ellis Horwood.

Morik, K. . S. Wrobel, J.-U. Kietz and W. Emde (1993), *Knowledge Acquisition and Machine Learning* – Theory, Methods, and Applications. London: Academic Press.

Quinlan, J.R. (1993), C4.5: Programs for Machine Learning. Machine Learning. San Mateo, CA: Morgan Kaufmann.

Rieger, A. (1996), MP: An efficient method for calculating the minimum Herbrand model of chain Datalog programs. In W. Wahlster (ed.), *Proceedings of the Twelveth European Conference on Artificial Intelligence*. Chichester et al.: John Wiley, 385–389.

Rieger, A.D. (1998), *Program Optimization for Temporal Reasoning within a Logic Programming Framework*. PhD thesis, University of Dortmund.

RWI Essen – Rheinisch-Westfälisches Institut für Wirtschaftsforschung (ed.) (2000), Arbeitsbericht 2000. Essen.

Sommer, E. (1996), *Theory Restructering*: A Perspective on Design & Maintenance of Knowledge Based Systems. PhD thesis, University of Dortmund.

Theis, W. and C. Weihs (1999), Clustering techniques for the detection of business cycles. SFB475 Technical Report 40. University of Dortmund.

Weihs, C. and U. Sondhauß (2001), Using labeled and unlabeled data to learn drifting concepts. In M. Kubat and K. Morik (eds.), Workshop notes of the IJCAI-01 Workshop on Learning from Temporal and Spatial Data. Menlo Park, CA: IJCAI, AAAI Press, 38–44.

Wrobel, St. (1994a), *Concept Formation and Knowledge Revision*. Dordrecht: Kluwer Academic.

Wrobel, St. (1994b), Concept formation during interactive theory revision. *Machine Learning Journal* 14(2): 169–192.

Zighed, D.A., S. Rabaseda, R. Rakotomalala and F. Feschet (1999), Discretization methods in supervised learning. In *Encyclopedia of Computer Science and Technology*. New York: Marcel Dekker Inc, vol. 40: 35–50.

Ursula Garczarek and Claus Weihs[1]

Univariate Characterization of the German Business Cycle 1955–1994

1. Introduction

In order to find simple univariate characterizations of business cycle phases, in this paper we apply simple statistical methods to quarterly after-war data of the German economy classified into four business classes called upswing (Up), upper turning points (UTP), downswing (Down), and lower turning points (LTP). The aim was to find simple univariate rules based on so-called "stylized facts" (Lucas 1987) with acceptable predictive power, i.e. with acceptable ability predicting the correct business cycle phase from the state of the economy.

In order to adapt the notion of predictive power to our problem, the cross-validation methods standard in statistical analysis like leave-one (-observation)-out- or 10-fold-cross-validation were replaced by the so-called double-leave-one-cycle-out analysis (Weihs, Garczarek 2002). By considering rules which contain one stylized fact only, we looked for those "stylized facts" being best able to characterize the business cycle over the whole time period available.

Our data set consists of 13 stylized facts 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-

[1] This work has been supported by the Deutsche Forschungsgemeinschaft (DFG), SFB 475. – We also thank Prof. Dr. U. Heilemann and H.-J. Münch from the RWI Essen (Rheinisch-Westfälisches Institut für Wirtschaftsforschung) for valuable discussions in our joint project in the SFB 475.

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 (Heilemann, Münch 1986) 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 the above mentioned four business cycle phases to each quarter from 1963/1 to 1994/4. This classification (and its extension to 1955/4) was supposed to be the "correct" classification for the purpose of our study. In this paper the following questions will be examined.

– Is the development of the stylized facts proceeding "in parallel" to the business cycle?
– Are the chosen stylized facts "independent" factors of the economic development?
– Are there ranges of the values of the stylized facts indicating a certain business cycle phase?

The solution of the first two questions should be seen as a preparation of the solution of the third question.

2. Analysis of the Course of the Economic Variables in the Business Cycle

As a first step we will compare the course of the considered economic variables over time with the development of the business cycle. We will illustrate this comparison by plotting the time series of the variable versus the business cycle indicated by means of an increasing line for upswing, a horizontal high level line for upper turning points, a decreasing line for downswing, and a horizontal low level line for lower turning points. We give some examples first.

Figure 1
Growth Rate of Gross National Product
1955 to 1994

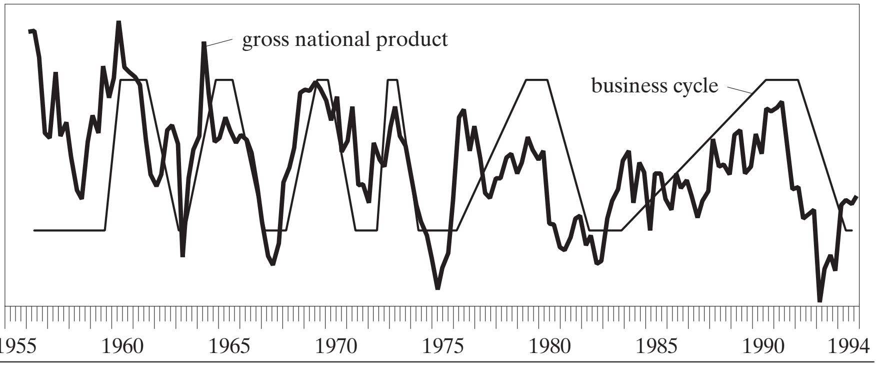


Figure 2
Growth Rate of Unit Labor Cost
1955 to 1994

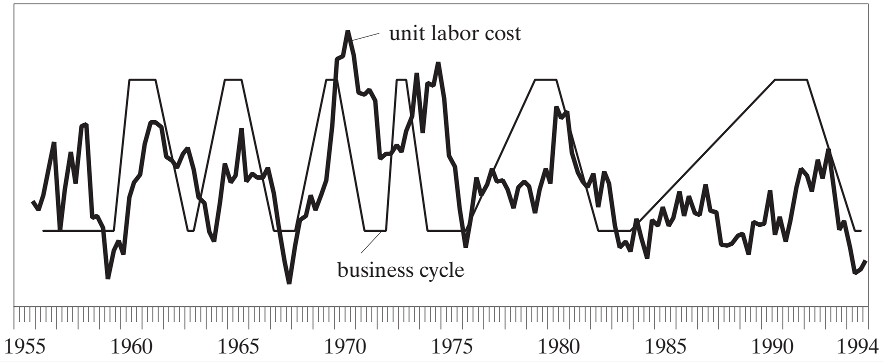
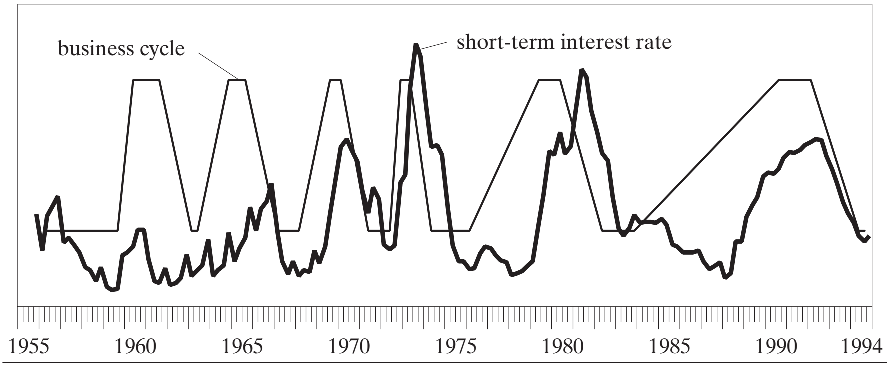


Figure 3
Nominal Short-Term Interest Rate
1955 to 1994



– The growth rate of gross-national-product Y is, as expected, a leading indicator of the cycle and is more or less stable in its level in the cycle phases over time (Figure 1).

– The growth rate of unit-labor-cost (LC) is a lagging indicator, and more or less level stable (Figure 2).

– The nominal short-term-interest-rate RS is lagging. The first cycles, however, cannot be identified (Figure 3).

– The growth rate of wage-and-salary-earners L is leading, and not all cycles can be identified (Figure 4).

Figure 4
Growth Rate of Wage and Salary Earners
1955 to 1994

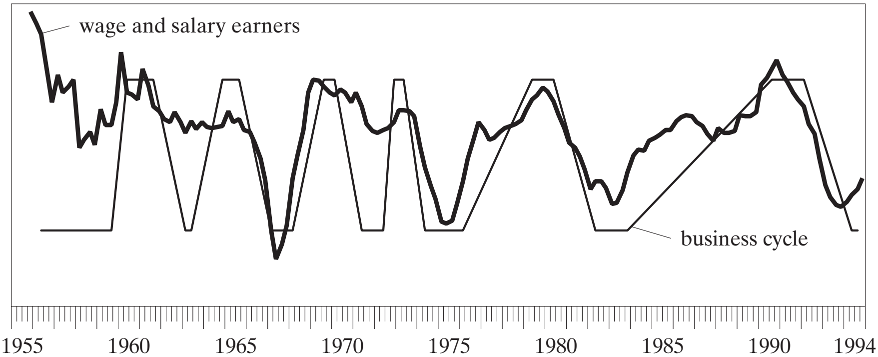
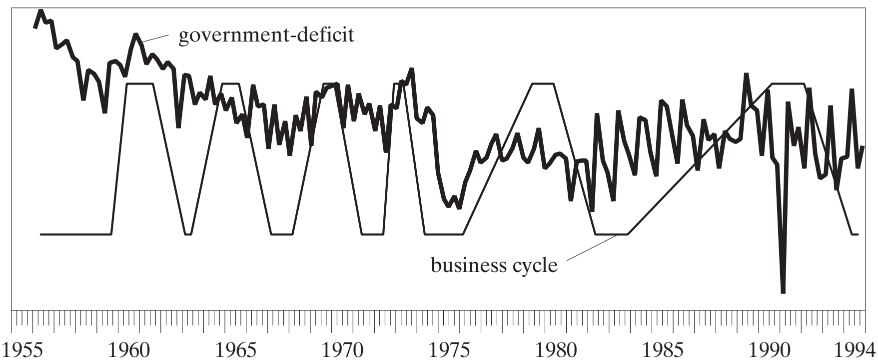


Figure 5
Government-deficit
1955 to 1994



– In the time series of the government-deficit GD the business cycle cannot be identified at all (Figure 5).

We can summarize the results of this analysis as follows:

– GD, RL, and X do not seem to have any relationship to the business cycle.

– IC and M1 are questionable as stylized facts.

– C, IE, L, LC, PC, PY, RS, and Y remain as possible stylized facts.

Table 1

Correlation Matrix

	IE	C	Y	PC	PY	IC	LC	L	M1	RL	RS	GD	X
IE	1.00	0.64	0.74	–0.37	–0.09	0.39	–0.08	0.67	0.31	–0.19	–0.27	0.30	–0.16
C	0.64	1.00	0.78	–0.34	0.02	0.51	0.12	0.66	0.41	–0.37	–0.30	0.53	–0.17
Y	0.74	0.78	1.00	–0.35	–0.17	0.68	–0.16	0.74	0.31	–0.10	–0.24	0.54	–0.13
PC	–0.37	–0.34	–0.35	1.00	0.72	–0.27	0.57	–0.20	–0.15	–0.12	0.62	–0.27	–0.30
PY	–0.09	0.02	–0.17	0.72	1.00	–0.18	0.87	0.06	–0.01	–0.66	0.49	–0.06	–0.27
IC	0.39	0.51	0.68	–0.27	–0.18	1.00	–0.18	0.50	0.18	–0.09	–0.19	0.35	–0.04
LC	–0.08	0.12	–0.16	0.57	0.87	–0.18	1.00	0.16	–0.15	–0.66	0.43	0.11	–0.34
L	0.67	0.66	0.74	–0.20	0.06	0.50	0.16	1.00	0.18	–0.23	0.05	0.48	–0.09
M1	0.31	0.41	0.31	–0.15	–0.01	0.18	–0.15	0.18	1.00	–0.18	–0.36	–0.02	0.19
RL	–0.19	–0.37	–0.10	–0.12	–0.66	–0.09	–0.66	–0.23	–0.18	1.00	0.15	–0.27	0.21
RS	–0.27	–0.30	–0.24	0.62	0.49	–0.19	0.43	0.05	–0.36	0.15	1.00	–0.21	0.07
GD	0.30	0.53	0.54	–0.27	–0.06	0.35	0.11	0.48	–0.02	–0.27	–0.21	1.00	–0.17
X	–0.16	–0.17	–0.13	–0.30	–0.27	–0.04	–0.34	–0.09	0.19	0.21	0.07	–0.17	1.00

3. Independence of Economic Variables

The above stylized facts were chosen in order to represent the different spheres of economics. In order to find more or less independent stylized facts for a low dimensional characterization of the business cycle we will study the correlation structure of the chosen variables by means of the correlation matrix (Table 1) and the corresponding graphical analogue, the scatterplot matrix (Figure 6).

From these representations we at least learn:

– The correlation matrix is nearly singular (spectral condition number = 121), thus we have to face multicollinearity.

– According to the scatterplot matrix there is a high linear relationship between PY and LC, and also between IE, C, Y, and L. According to the correlation matrix, PC is highly correlated to PY and RS, too.

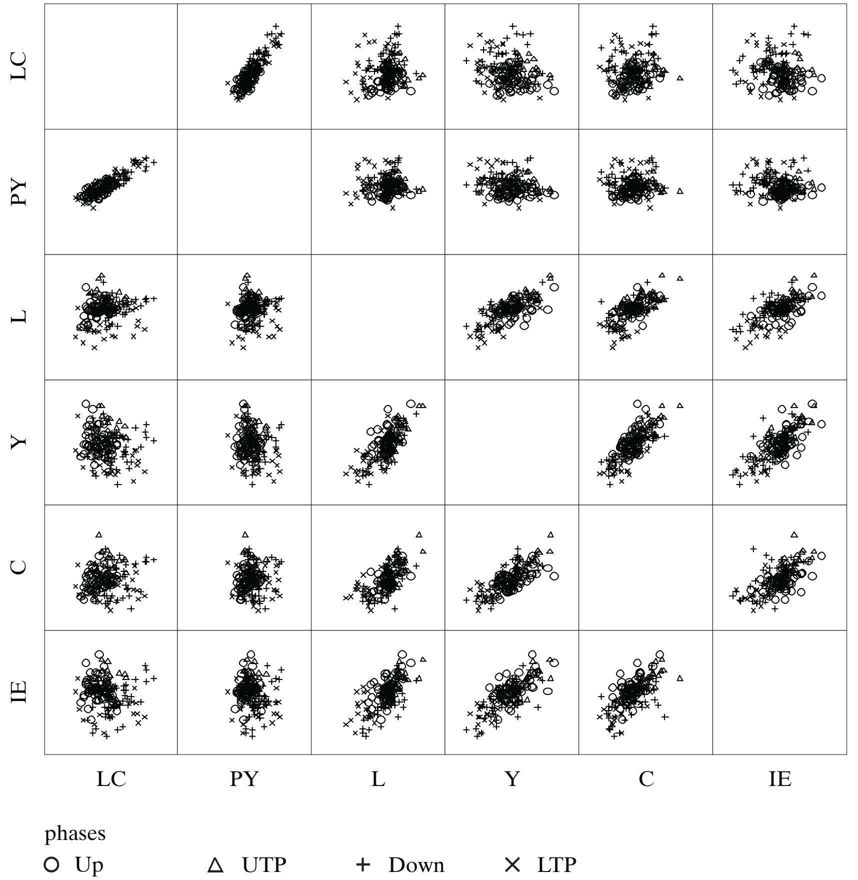
Overall, this correlation analysis might lead to the suspicion that the relevant factor dimension is even lower than eight as found in Section 2.

4. Univariate Rule Finding

In the following, ad-hoc rules for the classification of observations of economic variables into the 4 business cycle phases are derived by means of parallel box plots. The instructions chosen for rule construction look as follows:

> *Draw a parallel box plot for the observed values of some stylized fact in each of the four business cycle phases. The four boxes have horizontal lines at the corresponding lower quartile, median, and upper quartile values.*

Figure 6
Scatterplot Matrix – German Business Cycles



If the highest box is above two other boxes, that is if the greatest lower quartile is greater than the 2nd smallest upper quartile, then choose the greatest lower quartile as a separation limit;

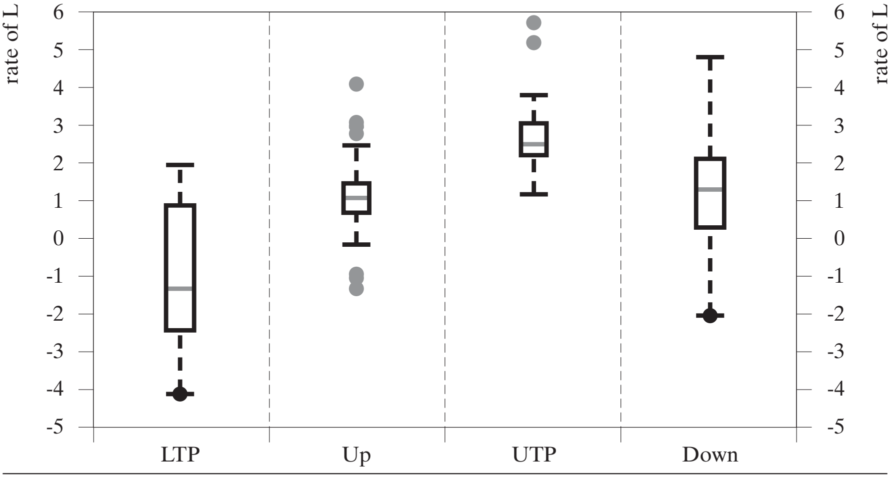
else if the inner line in the highest box is above two other boxes that is, if the greatest median is greater than the 2nd smallest upper quartile, then choose the median.

If the lowest box is below two other boxes, that is if the smallest upper quartile is smaller than the 3rd smallest lower quartile, then choose the upper quartile;

else if the inner line in the lowest box is below two other boxes that is, if the smallest median is smaller than the 3rd smallest lower quartile, then choose the median.

For each of the separation limits found in the above way, a rule is constructed by classifying into the phase with maximum frequency above (first pair of rules), and below (second pair of rules) the limit, respectively.

Figure 7
Parallel Boxplots of the Growth Rates of Wage and Salary Earners



These instructions are illustrated by means of examples.

– The rule '$L \geq 2.19 \Rightarrow UTP$' separates directly below the highest box (Figure 7). In the different phases the condition of the rule is fulfilled with the following frequencies (N: absolute frequency, P: relative frequency):

Up	UTP	Down	LTP
N: 6	18	10	0
P:.18	.53	.29	0

The rule '$L \leq -1.32 \Rightarrow LTP$' separates in the median of the lowest box (Figure 7).

– The rule '$Y \geq 6.34 \Rightarrow UTP$' separates in the greatest median (corresponding to UTP).
The rule '$Y \leq 2.43 \Rightarrow LTP$' separates directly above the lowest box (Figure 8).

– The rule '$LC \geq 5.71 \Rightarrow Down$' separates in the greatest median with the following frequencies:

	Up	UTP	Down	LTP
N	2	5	24	10
P	0.05	0.12	0.58	0.24

The rule '$LC \leq 2.03 \Rightarrow Up$' separates in the smallest median (Figure 9).

– For the long-term-interest-rate RL no rule can be constructed (Figure 10).

Figure 8

Parallel Boxplots of the Growth Rates of Gross National Product

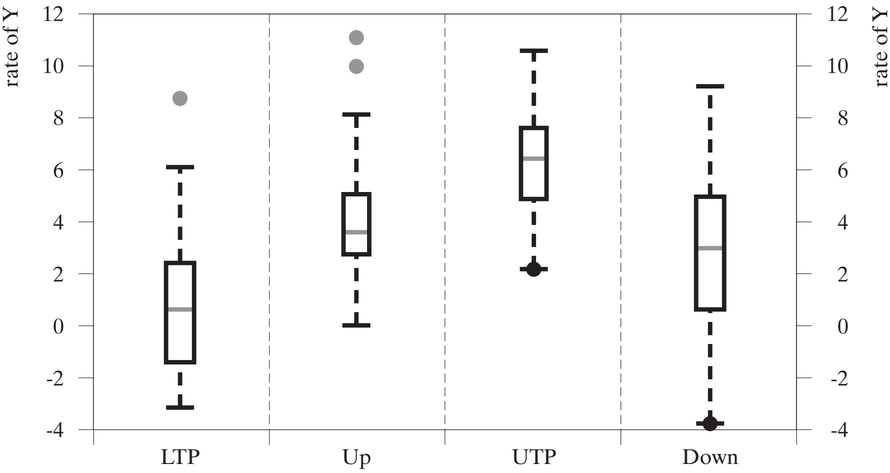
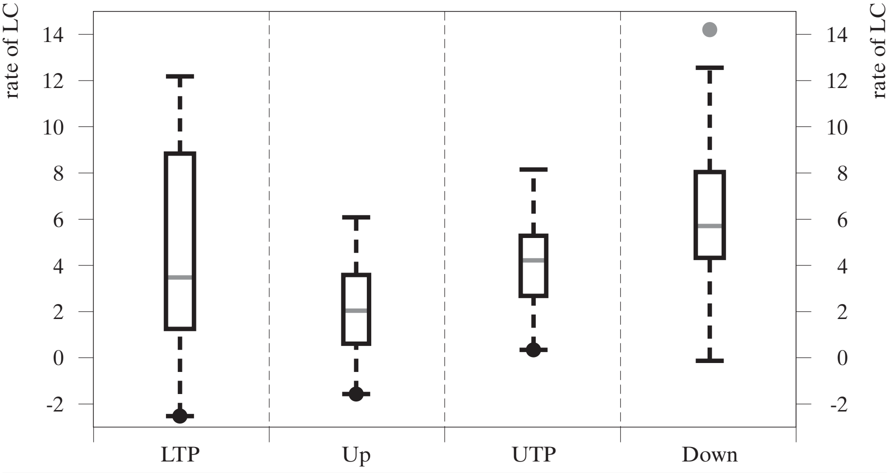


Figure 9

Parallel Boxplots of the Growth Rates of Unit Labor Costs

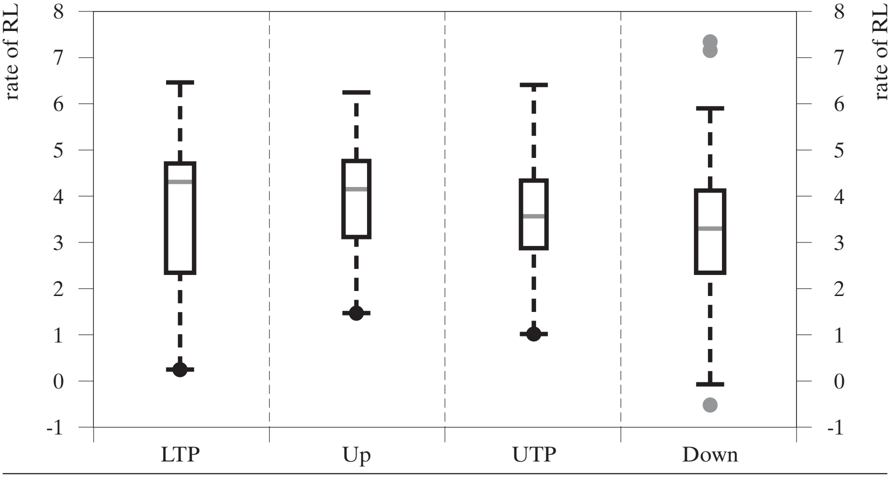


Since it may well be that more than one rule is applicable to one observation, a *higher order rule* is necessary for the decision which rule should be used. We use the following decision process:

For each observation of stylized facts, determine the valid rules.

If there is no valid rule, then randomly decide on the phase corresponding to the frequency of the phases in the learning sample.

Figure 10
Parallel Boxplots of the Growth Rates of Long-Term Interest Rates



If there is exactly one valid rule, then use this rule.

If there is more than one valid rule, then use the rule with minimum error, i.e. with the maximum of the maxima of frequencies P of validity of phases.

Example: Consider the following observation:
IE = 0.34, C = 6.13, Y = 5.73, PC = 3.29, PY = 3.98, IC = 2.08, LC = 6.20, L = 2.70 , M1 =11.00, RL = 3.09, RS = 4.83, GD = 3.06, X = 4.57.
Thus, C ≥ 5.56, LC ≥ 5.71, and L ≥ 2.19.

Therefore, the following three rules are applicable with relative frequencies P:

	Up	UTP	Down	LTP
If C ≥ 5.56 then Down:	P = 0.10256	0.30769	0.43590	0.15385
If LC ≥ 5.71 then Down:	P = 0.04878	0.12195	0.58537	0.24390
If L ≥ 2.19 then UTP:	P = 0.17647	0.52941	0.29412	0

This leads to the application of the 2nd rule, i.e. to the choice of the phase Down.

5. Ranges Corresponding to Phases

Altogether the above instructions lead to 18 rules, whereof only 11 are chosen using the above higher order rule (Table 2). Note that three rules lead to a correct decision in only one case, including the rules based on the growth rate of real-investment-in-construction IC, and the growth rate of money-supply-M1, the variables identified to be questionable stylized facts in Section 2.

Table 2

Rules Chosen by Higher Order Rule

Phase	Rule			Chosen	correct
LTP	L	≤	–1.32	20	14
	IE	≤	–2.73	1	1
Up	PY	≤	2.67	36	28
	RS	≤	4.5	26	13
	LC	≤	2.03	12	8
UTP	L	≥	2.19	8	4
	C	≥	5.56	4	2
	IC	≥	4.98	1	1
Down	LC	≥	5.71	29	18
	RS	≥	7.37	14	7
	M1	≤	7	4	1
				155	97

The stylized facts involved in the chosen rules are C, IC, IE, L, LC, M1, PY, and RS. Note that the real-gross-national-product growth rate Y is not used in the rules. The coverage of the rules, when learnt and applied on the whole data set, is 100%, the correctness 63%. Note, however, that the mean prediction error rate found by our standard method leave-one-cycle-out-cross-validation is as high as 54%, which is actually better than the corresponding rate for the standard Quadratic Discriminant Analysis (58%), but clearly worse than the corresponding rate for the standard Linear Discrimination Analysis (47%) (Weihs, Garczarek 2002).

6. Conclusion

Using ad-hoc instructions for identifying univariate rules characterizing the German business cycle 1955-1994, leads to an error rate comparable to standard multivariate methods. Nevertheless, all these error rates are that high that further analysis is urgently necessary (cp. Weihs, Garczarek 2002).

References

Heilemann, U. and H.J. Münch (1996), West German Business Cycles 1963-1994: A Multivariate Discriminant Analysis. In Selected Papers Submitted to the CIRET-Conference in Singapore. CIRET-Studien 50. Ifo, München, 219–250.

Lucas, R.E. (1987), *Models of Business Cycles.* Oxford: Basil Blackwell.

Weihs, C. and U. Garczarek (2002), Stability of multivariate representation of business cycles over time. SFB 475 Technical Report 20/2002. Universität Dortmund.

Claudia Becker and Winfried Theis[1]

Combining Dimension Reduction and Fuzzy-clustering: An Application to Business Cycles

1. Introduction

A question which can often be heard is whether economics is actually in a state of upswing or recession. Experts try to answer this question, based on information from actual economic data. These answers are to some extent subjective. A method by which we can predict the current business cycle phase by means of an objective statistical analysis of the data would be rather helpful. Since there are many different economic variables giving information about the current economic state, from a statistical point of view, dimension reduction methods seem to be a helpful tool in the analysis of such data. Before we can develop a method predicting the current phase, we have to make sure that the data used really contain the information needed. The aim of this article is to show to what extent we are able to re-discover a given experts' classification of business cycle phases. We use a data set of quarterly observations of 13 selected economic variables from about 40 years. An experts' judgement about the business cycle phase corresponding to each quarter is given. The data set and the four phases are described in more detail in Section 3.

The data have already been investigated earlier (Theis, Weihs 1999; Weihs et al. 1999). Here we apply a third approach, where in a first step the data set of observed economic variables is projected into a lower dimensional space, before a clustering of the dimension-reduced data is performed. The experts' information is used in the dimension reduction procedure, leading to a "supervised learning" type of projection of the data. To the reduced data, we then apply the "unsupervised" fuzzy-clustering method. If information about the business cycle phases can be extracted from the data, and if the experts'

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

classification is congruent with this information, the given business cycles should reappear after the clustering.

The paper is organized as follows. The used dimension reduction method SIR and the clustering algorithm are briefly described in Section 2. In Section 3 we introduce the data and discuss the results. We finish with some concluding remarks.

2. Dimension Reduction and Clustering

2.1 Sliced Inverse Regression

As discussed in Becker/Theis (2002), the classification of business cycles can be interpreted as a regression problem with a discrete univariate response variable. Hence, the idea of Li (1991) for dimension reduction in regression can be applied here. Assume that

$$(1) \qquad Y = f\left(\beta_1^T \boldsymbol{X}, \dots, \beta_K^T \boldsymbol{X}, \varepsilon\right).$$

The discrete response Y (business cycles) depends on the economic variables $X_1, \dots, X_p$ only via the linear combinations $\beta_1^T \boldsymbol{X}, \dots, \beta_K^T \boldsymbol{X}$. It surely can be assumed that there are interdependencies between the economic variables, hence this assumption is well motivated in our problem. A dimension reduction from p to K is achieved, if $K << p$. Here, $\boldsymbol{X} = (X_1, \dots, X_p)^T$, $\boldsymbol{X}, \varepsilon$ are stochastically independent, and $\beta_i \in R^p, i = 1, \dots, K$.

Sliced inverse regression (SIR; Li 1991) is a method to estimate the space B spanned by the so-called effective dimension reduction (edr) directions $\beta_1, \dots, \beta_K$. We do not discuss the SIR method in detail here. The main idea is to use the information contained in the inverse regression curve $E(\boldsymbol{X}|Y)$ instead of estimating $E(Y|\boldsymbol{X})$ directly. Under certain conditions, $E(\boldsymbol{X}|Y)$ almost surely lies in the space B. The inverse regression curve is roughly approximated, and the directions of its main variability are determined, yielding estimates for the edr directions (see Li 1991 for an extensive presentation). The $\boldsymbol{X}$ observations are then projected into the reduced space B. After the dimension reduction step, the functional relationship f can be estimated in the reduced space (Becker 2001), i.e. based on the projected data. In our investigation of the economic data, we do not try to estimate f in the second step, but to re-reveal a certain structure. 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.

In the original work of Li (1991), it is assumed that $Y \in R$. But when approximating the inverse regression curve, only categories of Y values and corresponding $\boldsymbol{X}$ observations are needed. Thus, the SIR procedure can easily be applied to the case of a discrete response Y (Cook, Lee 1999; Chen, Li 2001). For the economic application we take four categories, because the business classification used is composed of four different cycle phases (Section 3).

To get a first impression of how our combination of dimension reduction and clustering may be helpful in the context of the classification of business cycles, we ignore the time-series structure of the data and treat the $\boldsymbol{X}$ values as if they were i.i.d. observations. Further work will be concerned also with the time-series aspect (also Becker, Fried 2003: 3–11; Becker et al. 2001: 201–214).

2.2 Fuzzy-clustering

In earlier investigations of the same data (Theis, Weihs 1999), fuzzy-clustering based on a certain distance measure turned out to be the best version of clustering for this problem. Generally, contrary to classical clustering procedures fuzzy-clustering does not divide a data set into well-separated groups. Instead, different groups are identified, and every observation gets a set of so-called membership values, one corresponding to each group. The membership values range from 0 to 1 and can be interpreted as the degree of an observation belonging to a certain group. Hence, an observation may belong to several of the groups with different degrees. Points lying in overlapping regions get membership values smaller than 1 to belong to a specific group, whereas points lying in only one group get a membership of 1 to belong to this group and 0 for all other groups. The overall goal is to assign the observations to the groups in a way such that a maximum number of high membership values is achieved. The number k of groups or fuzzy-clusters has to be chosen beforehand, motivated by knowledge of the problem at hand or by hints from descriptive analysis. It is obvious that fuzzy-clustering leads to the same result as the classical k-means approach if there are well separated groups of data points.

For our analysis, we apply fuzzy-k-means clustering as implemented in the R/S-function FANNY (Kaufman, Rousseeuw 1992: 164–197). To judge the goodness of "separation" into the groups, we use the normalized Dunn-coefficient (Kaufman, Rousseeuw 1992: 187), which takes values between 0 for no partition (total fuzziness) and 1 for a hard partition.

To perform fuzzy-clustering, distances between observations have to be measured. As found out by Theis/Weihs (1999), 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. Theis/Weihs (1999) adjust for this by normalizing the observations with their euclidean norm and hence reducing the information in the observations to their direction in p-dimensional space. The euclidean distance is then taken for the normalized observations, leading to a distance $d(\boldsymbol{x},\boldsymbol{y}) := \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} - \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \right\|$ between two observations $\boldsymbol{x}$ and $\boldsymbol{y}$.

3. Results

The data set consists of $p = 13$ economic variables, the so-called stylized facts for the German business cycle. Quarterly observations are available from 1955/4 to 1994/4 (price index base: 1991, *y*: yearly growth rates), yielding $n = 157$ observations. The variables and their abbreviations used here are displayed in Table 1.

The stylized facts are selected in Heilemann/Münch (1996) from a total of 120 variables with respect to the objective that these 13 variables represent all information necessary to classify the observations into four business cycle phases. The phases are called "upswing", "upper turning point phase", "downswing" and "lower turning point phase". The experts' classification for the data can be seen in Figure 1. The phases are arranged in a special order on the ordinate axis. The dominating phases "upswing" and "downswing" form the outer ordinates whereas the turning point phases are arranged in between.

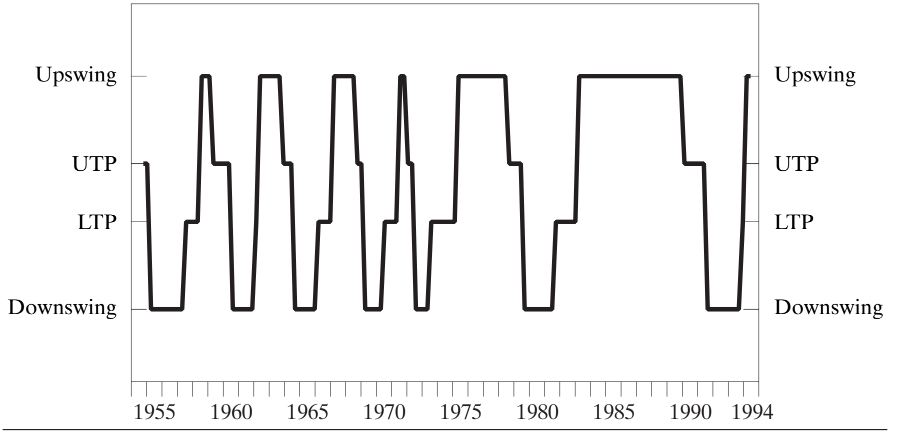
Applying SIR to the data set yields a reduced space of dimension three. The variables contributing most to this space are the GNP price deflator (PY), the long and short term interest rates (RL, RS) and the private consumption (C). This set of variables is similar to the sets chosen when determining the variables optimal for classification of the original data. For example, the variables L, LC, and RL, which are all contained in the projections found by

Table 1

The 13 Stylized Facts

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

Figure 1
Experts' Classification of Business Cycles
1955/IV to 1994/II



SIR, also form the best set of three variables for quadratic discriminant analysis (Weihs et al. 1999).

In the second step of our analysis, we now apply the fuzzy-clustering procedure as described in Section 2 to the dimension-reduced data. We compare the results with the outcome of applying the fuzzy-clustering procedure to a set of three "optimal" variables (L, IE, PC). These variables are determined to yield the best possible fuzzy-clustering based on any three of the 13 stylized facts (Becker, Theis 2002).

To determine the number *k* of classes, we calculate the Dunn-coefficient for the possible choices k=4,3,2, starting with k=4 classes like in the experts' classification. Only for k=2 we get clusterings with a Dunn-coefficient clearly larger than zero. Hence, we choose k=2.

With the optimal variables, the value of the Dunn-coefficient exceeds the value reached for the projected data. This means that we can generally get better separated clusters using the optimal variables than using the projection from SIR. But these "optimal" clusters do not reflect the experts' classification as good as the clusters based on the projected data. This can be seen in Figures 2 and 3 (see the discussion below).

The two classes found can be best interpreted as the two dominating phases "upswing" and "downswing". The turning-point phases are inherently found by looking at the membership value of each observation. Distinguishing only two groups, high membership values with respect to the first group (upswing,

Figure 2

Fuzzy-clustering of Optimal Variables: Membership to Group "Upswing" and Experts' Classification

1955/IV to 1994/II; membership values

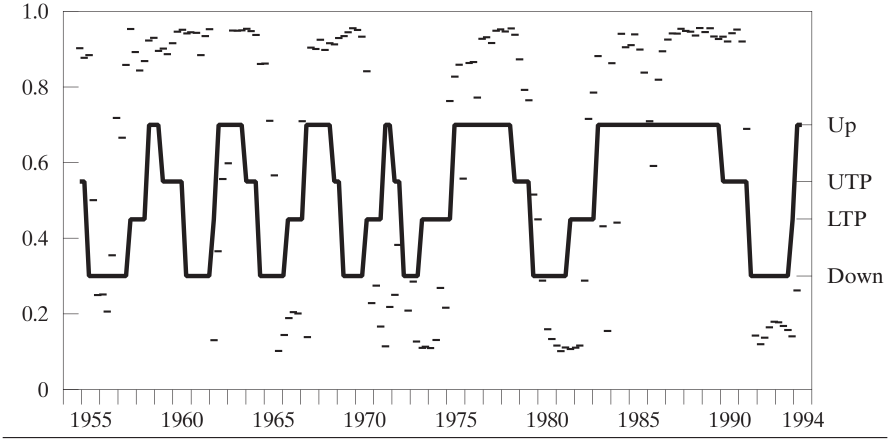
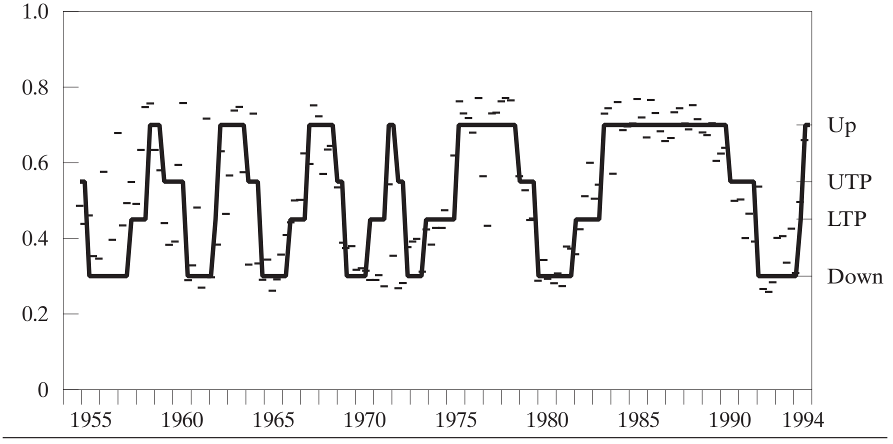


Figure 3

Fuzzy-clustering of SIR Projected Data: Membership to Group "Upswing" and Experts' Classification

1955/IV to 1994/II; membership values



say) tell us that the corresponding observations belong (with a high degree of reliability) to an upswing phase. Low membership values with respect to this group speak for the observations belonging to a downswing phase. Membership values around 0.5 can then be interpreted as observations in between the two phases. Plotting the membership values with respect to one of the groups over time we expect to find times of high values alternated with low ones and in between some points with membership values near 0.5, belonging to the respective turning-point phase. Such plots can be found in Figures 2 and 3, for the clustering based on the optimal variables and the SIR projection, respectively, with respect to the upswing phase. The experts' classification as displayed in Figure 1 is depicted as an additional line in the plots. The heights chosen for the line are chosen in a way that the "middle" values (standing for the turning point phases) lie around 0.5, corresponding to our interpretation of the membership values. Obviously, the clustering is less clear when using the projected data (Figure 3) than for the optimal variables (Figure 2). But in Figure 3, the membership values of the observations arrange themselves in a pattern very similar to the experts' classification. Due to the greater fuzziness coming with the projection by SIR we gain an implicit detection of the turning-point phases.

4. Conclusion

In the considered problem the proposed combination of the supervised construction of a dimension reduction with the unsupervised fuzzy-clustering has revealed that the necessary information about the business cycle phases is contained in the used data set. Even the drawback that only two groups could be found was overcome by interpretation of the results in a problem-oriented way. These results recommend the combination for a further use in this field, since if it is possible to re-discover the business cycle phases, the method should also be able to help in constructing business cycles from new data.

References

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

Becker, C. and R. Fried (2002), Sliced Inverse Regression for High-dimensional Time Series. In O. Opitz and M. Schwaiger (eds.), *Exploratiory Data Analysis in Empirical Research*. Berlin et al.: Springer.

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

Becker, C. and W. Theis (2002), Clustering of Business Cycles in Optimal Directions found by SIR and DAME. SFB 475 Technical Report 13/2002. Universität Dortmund.

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

Cook, R.D. and H. Lee (1999), Dimension Reduction in Binary Response Regression. *Journal of the American Statistical Association* 94: 1187–1200.

Heilemann, U. and H.J. Münch (1996), West German Business Cycles 1963–1994: A Multivariate Discriminant Analysis. In Selected Papers Submitted to the CIRET-Conference in Singapore. CIRET-Studien 50. Ifo, München, 219–250.

Kaufman, L. and P.J. Rousseeuw (1992), *Finding Groups in Data*. New York: Wiley.

Li, K.-C. (1991), Sliced Inverse Regression for Dimension Reduction (with discussion). *Journal of the American Statistical Association* 86: 316–342.

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

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

Gabriela Guimarães

Self-organizing Maps for Time Series Analysis

1. Introduction

In recent years a large amount of data was gathered from quite different processes such as industrial processes, medical applications, meteorological phenomena, and also from financial processes. Artificial neural networks (ANN) and methods from Statistics are particularly well suited for handling such noisy and inconsistent data. The application of ANN and statistics often refers to problems of discrimination (supervised learning) or to clustering problems (unsupervised learning).

In this paper we present a special ANN proposed by Kohonen (1982), called Self-organizing Map (SOM). SOM have been successfully used in a large range of applications (Kohonen 2001). Since they are neural networks based on unsupervised learning, they have been widely used for exploratory tasks (Deboeck, Kohonen 1998; Kaski 1997). The main aim in exploratory data analysis is to discover patterns (clusters) in high dimensional data usually performing a projection from a high dimensional space onto a lower dimensionality. For SOM this means that an adequate visualization of the network structures is required. For example, component maps (Simula, Alhoniemi 1999), U-matrices (Ultsch, Siemon 1990), and hierarchical and colored visualization techniques (Murtagh 1995; Vesanto, Alhoniemi 2000; Himberg 2000) have been used for this propose.

In addition, SOM have also been successful in applications, where temporal or sequential data are processed, for instance, in speech recognition, process control and time series analysis in medicine (Behme et al. 1993; Walter, Schulten 1993; Guimarães 2000). However, the basic SOM algorithm like all the other clustering algorithms does not take time into account. Many different approaches have been developed to adapt the basic SOM algorithm for temporal sequence processing. All these approaches have particular advantages and drawbacks, and some are reviewed in (Kohonen, 2001; Guimarães, Moura-Pires 2001; Chappelier et al. 2001). The use of each of these approaches depends highly on the application, and none is universally better than the other. The best results are usually obtained by carefully combining these methods. SOMs for temporal processing usually are used for one of the following purposes: prediction (Ultsch et al. 1996; Koskela et al. 1998), control (Ritter et al. 1989; Ritter 1994), monitoring (Joutsiniemi et al. 1995), and data mining (Guimaraes 2000; Guimaraes et al. 2001a; Deboeck, Kohonen 1998).

In this paper an overview of the different types of adaptation will be given. The main ideas that have been proposed will be described. We will also discuss their strengths and drawbacks as well as provide some examples, if needed.

First the basic SOM algorithm is explained in Section 2. Since three main "families" of SOMs have been identified, they will be explained in the following sections. This taxonomy was first proposed in Guimaraes/Moura-Pires (2001). In Section 3 the first approach is presented, where time is introduced only during the pre- or post-processing stage and the basic SOM algorithm is preserved. Section 4 presents the approach that modifies the learning rule (and/or the classification rule) in order to take temporal dependencies into account. In Section 5 the approach with the modification of the topology of the basic SOM is presented, either by introducing feedback or a hierarchical structure. Conclusions are presented in Section 6.

2. The Basic SOM-Algorithm

SOMs is strongly biology-motivated, where biological principles, the generation of topographical maps in the brain through self-organization, play an important role. In the following, the learning process will be described from a more algorithmic point of view. During learning SOMs adapt their weights such that a n-dimensional input space is projected onto a m-dimensional map with $m < n$, preserving the neighborhood of the input data on the map. Usually, a two-dimensional map is chosen. The map is formed by the properties inherent to the data itself. Consequently, no previous classification of the data is needed. The input layer has n units representing the n components of an input vector $x_k = (x_{k1}, \ldots, x_{kn}), k = 1, \ldots, N$. The output layer is a two dimensional array of units arranged on a map. Each unit in the input layer is connected to each unit in the output layer with a weight $w_i = (w_{i1}, \ldots, w_{in}), i = 1, \ldots, p \cdot q$ associated. All weights are initialized randomly. They are adjusted according to Kohonen's learning rule

$$\Delta w_i = \eta(t) \cdot (x_k - w_i) \tag{1}$$

that uses a distance measure $\|w_r - x_k\| = \min_i \|w_i - x_k\|$ to determine the best-match r and a neighborhood function

(2) $$h_{ir}(t) = e^{-\frac{(i-r)^2}{2\,\sigma(t)^2}}$$

that realizes the lateral inhibition. The learning rate determines the strength of learning with $0 < \eta(t) < 1$. The radius $\sigma(t)$ determines the set of neurons in a neighborhood of the bestmatch that are included into the learning process. Both functions usually decrease monotonously during learning.

On the map neighboring units form regions that correspond to similar input vectors. These neighborhoods form disjoint regions, thus enabling a classification of the input vectors. However, in order to perform a classification, a visualization of the network structures is needed, since the Kohonen algorithm converges to an equal distribution of the units on the output layer (Kohonen 2001). Therefore, a visualization of the network structure may give an enhanced insight into the network sturctures. A possible, and often used visualization tool is the U-Matrix (Ultsch, Siemon 1990), a three dimensional landscape, which represents structural properties of the high dimensional input space on the map. At each point of the grid the distance between the weights of two neighboring units is calculated and then is displayed as height into the third dimension. A U-Matrix has valleys where the vectors on the map are close to each other and represent data that are in the same class. Hills or walls represent larger distances indicating dissimilarities of the input data. Alternative visualization tools for SOMs have also been proposed, such as the agglomerative clustering where the SOM neighborhood relation can be used to construct a dendrogram on the map (Vesanto, Alhoniemi 2000), and a hierarchical clustering of the units on the map with a simple contraction model (Himberg 2000).

3. Temporal Sequence Processing without Modifying the Basic SOM-Algorithm

There are mainly two distinct approaches of SOMs for temporal processing that do not require an adaptation of the original algorithm or network topology. On the one hand a pre-processing of a temporal sequence can be performed before it is presented to the SOM. In this case time is embedded into the pattern vector (which we will call SOM with embedded time). On the other hand the network outputs are submitted to some kind of post-processing, whereby a time-related visualization or processing of the data on the map with trajectories enables the identification of temporal dependencies between the output patterns or vectors (Trajectory-based SOM).

In *embedded time approaches* time series are pre-processed before the vector is presented to the SOM. Here the input vector is treated in the standard manner without taking time into account. There are several ways to bring in time into the pattern vector. The simplest way is to use a tapped-delay of the input as pattern vector, i.e it is a vector of time-shifted samples of the temporal sequence (Kangas et al. 1990; Príncipe, Wang 1995). This approach is very simple and intuitive, and in many cases it provides good results. However, one of the main drawbacks is the determination of the length of the tapped delay, which has to be large enough in order to capture the dynamics of the system.

The domain transformation is another way of introducing time into the pattern vector. The most used is a short-time Fourier transform (Kohonen 1988), but other transformations as cepstral features (Kangas 1992), wavelet transforms (Moshou, Ramon 2000), and time-frequency transformations (Jossa et al. 2001) are also very popular.

The main advantage of SOM with embedded time lies in the preservation of the well-known characteristics of the SOM algorithm, although enabling a time processing. This also means that a simple integration of standard SOM software packages with appropriate pre-processing software enables the handling of time series with SOM.

This technique has been first used in speech recognition (Kohonen et al. 1984). Leinonen et al. (1992), for example, calculated a 256 point FFT of the short time power spectra of each 9.83s chunk of signal (spoken Finnish) in order to detect dysphonia. Another application is the identification of different classes of ships, using the sound they make in the ocean and transforming the signal into the frequency domain.

Trajectory-based SOM either don't adapt the original algorithm, but perform a post-processing of the results obtained during the classification phase considering temporal relations among succeeding best-match units and connecting them with a path, often named trajectory. This approach can be regarded as a new way of dealing with time, since temporal information is recovered during the classification phase, though revealing inherent structures of the time series. In fact the success of trajectory-based approaches is based on the topological information provided by SOM extrapolating it into the time domain. Additional information can be obtained from the direction of the path in order to predict a given state of the underlying process before it actually occurs (Tryba, Goser 1991; Ultsch 1993).

Trajectories are often combined with other visualization techniques for the graphical representation of the weights of a learned SOM. These are, for instance, component maps where one of the components of the weights is projected onto a third dimension, as well as U-Matrices (Ultsch, Siemon 1990), where the distances between neighboring units calculated in the original space, i.e. the weights, are projected onto a third dimension. Often these additional visualization techniques lead to an enhanced interpretation of the trajectory. Other interesting visualization techniques for SOMs consider hierarchical agglomerations of the map where the SOM neighborhood relation can be used to construct a dendrogram on the map (Murtagh 1995; Vesanto, Alhoniemi 2000; Himberg 2000).

The visualization of trajectories on the map itself was first applied to speech recognition problems (Kohonen 1988). At the early 90's this approach was then applied to several medical applications, such as the identification of co-articulation variation and voice disorder (Utela et al. 1992), and the recognition of topographic patterns in EEG spectra from several patients having different sleep/awake stages (Joutiniemi et al. 1995). Trajectory-based SOMs have also been proposed to model low dimensional non-linear processes, such as non-linear time series obtained from a Markey-Glass system (Principe, Wang 1995). All these approaches perform a visualization of the trajectories on the map itself without additional visualization of the network structures. The following examples describe how such an visualization may enhance the post-processing process.

Kasslin et al. (1992), for instance, uses components maps together with trajectories for process state monitoring. Here the values for one parameter are visualized as gray values on a map. An interpretation of the map is straightforward. The lighter the unit on the map, the higher the parameter value is. Their main aim was the classification of system states and the detection of faulty states for devices based on several device state parameters, such as temperature. Faults in the system could be detected with trajectories, if a transition to a forbidden area on the map marked with a very dark color occurred. This approach was also applied to process control in chemistry for monitoring a destillation process (Tryba, Goser 1991).

Visualization of trajectories on U-matrices have been used for monitoring chemical processes (Ultsch 1993), and have been applied to complex processes, such as the dynamic behavior of a computer systems with regard to utilization rates and traffic volume (Simula et al. 1996), to industrial processes, such as a continuous pulp digester, steel production and pulp and paper mills (Alhoniemi et al. 1999), and to different subjects with distinct sleep apnea diseases (Guimaraes et al. 2001).

4. Modifying the Activation/Learning Rule

The adaptation of the original Kohonen activation and/or learning rule represents another possibility for processing temporal data with SOMs, where a distinction between two approaches can be made. First the input vector is decomposed into two distinct parts, a past or context vector and a future or pattern vector, and both parts are treated in different ways during learning.

This approach was introduced by (Kohonen 1991) and named Hypermap. Second, there exists the possibility to choose the best match in a neighborhood of the last best match. Since this approach was first proposed by (Kangas 1992), we will call it Kangas map.

The *Hypermap* decomposes the input vector into two distinct parts, a "past" or "context" vector and a "future" or "pattern" vector and both parts are treated in different ways. Usually the context part is used to select the best match or a region around the best-match region, and then both parts are used to adapt the weights, separately or together.

For time series a Hypermap means that the future (prediction) is learned in the context of it's past. During the classification phase the prediction is made using only the "past" vector for the best-match search. The "future" part of the vector is then retrieved from the weights of the map associated with the best match obtained by the "past" part of the vector.

We however have to be aware that this approach transforms the original algorithm towards a supervised learning algorithm, since an output vector (the future) is added to the input (the past). This often makes sense when an extrapolation of the data into the future is required, for instance, in time series prediction (Ultsch et al. 1996), or in robot control (Walter, Schulten 1993).

This architecture was generalized by (Brückner et al. 1992). Here hierarchical relationships with $n-1$ levels define the context for the classification of EEG signals. This type of model was also studied for phoneme recognition (Midenet, Grumbach 1994), for simulating a sensory-motor task (Ritter et al. 1989), as well as for robot control (Walter, Schulten 1993; Ritter 1994). In the latter application, the output is the target position of the robot arm (for instance, given by the angles), while the input is given as a four-dimensional vector describing the spatial position of the robot arm obtained by the images of two cameras.

Hypermaps have also been used for prediction tasks, for instance, using SOMs for local models in the prediction of chaotic time series (Koskela et al. 1998). In Ultsch et al. (1996) a two step implementation of this approach was used for the prediction of hailstorm, first performing a classification of different types of hailstorm, in order to perform an enhanced prediction during a second prediction step.

The *Kangas Map* considers the previous best match as context and uses only the neighboring units when choosing the next best match, though enabling the learning in a given context on the map. In this approach, the original learning rule is preserved.

The area for the search of next best matches can be regarded as a "focus region", where the changes in the input are reported. This means that we may have various distinct areas with almost the same information (relating to the same type of input signal), but with different neighboring areas. Consequently, the activation of the units depends on its past history, i.e. on how the signal reached that region.

Approaches based on Kangas Maps have been used for speech recognition tasks (Kangas 1992), texture identification, and 3-D object identification (Chandrasekaran, Liu 1998).

5. Modification of the Network Topology

Another possibility in handling temporal data with SOM lies in an adequate modification of the network topology, either introducing feedback connections or introducing several hierarchical layers, each with one or more SOMs. Feedback SOM are closely related to digital filters and ARMA models. Hierarchical SOM are used when a segmentation of complex and structured problems is needed, for instance, in image recognition, speech recognition, time series analysis, process control, and protein sequence determination.

SOM with Feedback can be ranked among all of the other classical methods in control theory that handle temporal sequences feeding some kind of output back into the inputs. Usually an internal memory stores past outputs, and uses them in order to generate new outputs. This also means that they overcome some main drawback of tapped-delay approaches, since the length of the delay has not to be specified. There are a few different ways to introduce feedback into the system.

First, SOM with feedback appeared in 1993 (Chappell, Taylor 1993), and was named Temporal Kohonen Map. The main idea here lies in using the output values of each unit in the calculation of the next output value of that unit. This is done introducing a leaky integrator in the output of each SOM unit. However, this approach only keeps the magnitude of the output of the units, but not the direction of the error. To overcome this limitation, a recurrent SOM was proposed (Critchley 1994).

Both approaches only use local feedback. This means that only the previous output of that unit is used. An alternative is to feedback the outputs of all the units of the map to each of them (Von Harmelen 1993). This approach was later on analysed in detail in (Voegtlin, Dominey 2001). Another approach, similar to recursive SOM, was proposed and analysed in (Euliano, Príncipe 1999) and, in the latter paper, named SOM with Temporal Activity Diffusion. In this approach only the activations of neighboring units are fed back. This leads to a sort of shock wave that is generated in the best match unit, and propagates throughout output space of the map.

Recurrent SOMs have been successfully applied to the Mackey-Glass chaotic series, infrared laser activity, and electricity consumption (Koskela et al. 1997), as well as for clustering epileptic activity based on EEG (Koskela et al. 1998). The SOMTAD model, also related to these kind of models, has been applied to digit recognition, and to other small illustrative problems (Euliano, Príncipe 1999).

Hierarchical SOMs perform a structured decomposition of a very complex problem into smaller and layered problems, using one or more than one SOM at different layers, usually operating at different time scales. Hierarchical SOM in temporal sequence processing have been successfully applied to speech recognition (Kemke, Wichert 1993), electricity consumption (Carpinteiro, Silva 2000), vibration monitoring (Jossa et al. 2001), motion planning (Barreto, Araújo 1999) and temporal data mining in medical applications (Guimarães 2000).

The main difference between the different hierarchical SOM lies in the type of codification of the results of one level SOM to the next upper level. They also differ in the number of levels used, which strongly depends on the type of application. Finally, they differ in the number of SOM at each level, and the interconnections between the levels.

Hierarchical SOM principle is that of "divide and conquer". A relatively low complexity model is used, such as a SOM to handle each of the small groups of inputs, and then this partial information is merged in order to extract higher-level results. Good results can thus be obtained with hierarchical SOM in complex problems that cannot be modeled by a single SOM (Guimarães et al. 2001).

There are mainly three different ways to calculate the input vector for the next level SOM. First, the weights of the lower-level SOM are used as input without any further processing, either taking into account the information of the previous known classes (Kemke, Wichert 1993) or without considering any information on the classes (Walter, Ritter 1996). Second, a transformation of the network results is possible, for instance: 1) calculating the distances between the units (Carpinteiro 2000); 2) concatenating subsequent vectors into a single vector, thus representing the history of state transitions (Simula et al. 1996); or 3) taking into account the information about clusters formed at this level, and adjusting the weights towards the cluster center (Guimarães 2000). The third possibility lies in interposing other algorithms or methods, such as segment classifiers (Behme et al. 1993).

6. Conclusions

In this paper, several approaches of adaptations of the Self-organizing Maps (SOM), as proposed by Kohonen (1982) for dealing with time series were presented. Three main approaches for temporal sequence processing with SOM were identified. These are: 1) methods requiring no modification of the basic SOM algorithm, such as embedded time and trajectory-based approaches; 2) methods that adapt the activation and/or learning algorithm, such as Hypermaps or Kangas Maps; and 3) methods that modify the network structure, introducing feedback connections, or hierarchical levels. The use of each of these approaches, which are not mutually exclusive, depends highly on the application, and none is universally better than any other. The best results are usually obtained by using a carefully tailored combination of these methods.

References

Alhoniemi, E., J. Hollmén, O. Simula and J. Vesanto (1999), Process Monitoring and Modeling using the Self-Organizing Map. *Integrated Computer Aided Engineering* 6 (1): 3–14.

Barreto, G. and A. Araújo (1999), Unsupervised Context-based Learning of Multiple Temporal Sequences. In *Proceedings IEEE-INNS International. Joint Conference on Neural Networks* (IJCNN'99), Washington, DC, 1102–1106.

Behme, H., W.D. Brandt and H.W. Strube (1993), Speech Recognition by Hierarchical Segment Classification. In S. Gielen and B. Kappen (eds.), *Proceedings of the International Conference on Aritificial Neural Networks* (ICANN 93). London: Springer, 416–419.

Brückner, B., M. Franz and A. Richter (1992), A Modified Hypermap Architecture for Classification of Biological Signals. In I. Aleksander and J. Taylor (eds.), *Artificial Neural Networks*, 2. Amsterdam: Elsevier Science, 1167–1170.

Carpinteiro, O. and A. Silva (2000), A hierarchical neural model in short-term forecasting. In A. Shun-Ichi et al. (eds.), Proceedings of the IEEE-INNS-ENNS Interntional Joint Conference on Neural Networks. Los Alamitos et al., CA: IEEE Computer Society, 6: 241–248.

Chandrasekaran, V. and Z. Liu (1998), Topology Constraint Free Fuzzy Gated Neural Networks for Pattern Recognition. *IEEE Transactions on Neural Networks* 9 (3): 483–502.

Chappelier, J., M. Gori and A. Grumbach (2001), Time in Connectionist Models. In R. Sun and C.L. Giles (eds.), Sequence Learning: Paradigms, Algorithms, and Applications. Lecture Notes in Artificial Intelligence Serie 1828. Berlin et al.: Springer, 105–134.

Chappell, G.J. and J.G.Taylor (1993), The Temporal Kohonen Map. *Neural Networks* 6: 441–445.

Crichley, D.A. (1994), *Extending the Kohonen Self-Organizing Map by use of adaptive parameters and temporal neurons.* Ph.D. Thesis, University College London, Department of Computer Science.

Deboeck, G. and T. Kohonen (1998), *Visual Explorations in Finance with Self-organizing Maps.* London et al.: Springer.

Euliano, N. and J. Principe (1999), A Spatio-Temporal Memory Based on SOMs with Activity Diffusion. In E. Oja and S. Kaski (eds.), *Kohonen Maps.* Amsterdam: Elsevier, 253–266.

Guimarães, G. (2000), Temporal knowledge discovery with self-organizing neural networks. *International Journal of Computers, Systems and Signals* – Special issue Knowledge Discovery from Structured and Unstructured Data, Part I, 5–16.

Guimarães, G. and F. Moura-Pires (2001), An Essay in Classifying Self-Organizing Maps for Temporal Sequence Processing. In N. Allison, H. Yin, L. Allinson and J. Slack (eds.), *Advances in self-Organizing Maps.* London et al.: Springer, 259–266.

Guimarães, G., J.-H. Peter, T. Penzel and A. Ultsch (2001), A method for automated temporal knowledge acquisition applied to sleep-related breathing disorders. *Artificial Intelligence in Medicine* 624: 1–27.

Himberg, J. (2000), A SOM based cluster visualization and its application for false coloring. In In A. Shun-Ichi (ed.), *Neural computing – New Challenges and perspectives for the new millennium.* Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'2000-Como). Los Alamitos et al., CA: IEEE Computer Society, 3: 587–592.

Jossa, I., U. Marschner and W.J. Fischer (2001), Signal Based Feature Extraction and SOM based Dimension Reduction in Vibration Monitoring Microsystem. In N. Allinson, H. Yin, L. Allison and J. Slack (eds.), *Advances in self-Organizing Maps.* London et al.: Springer, 283–288.

Joutsiniemi, S.L., S. Kaski and T.A. Larsen (1995), Self-Organizing Map in Recognition of Topographic Patterns of EEG Spectra. *IEEE Transactions on Biomedical Engineering* 42 (11): 1062–1068.

Kangas, J. (1992), Temporal Knowledge in Locations of Activations in a Self-Organizing Map. In I. Aleksander and J. Taylor (eds.), *Artificial Neural Networks*, 2. Amsterdam: Elsevier Science, 117–120.

Kangas, J., T. Kohonem and J. Laaksonem (1990), Variants of Self-Organizing Maps. *IEEE Transactions on Neural Networks* 1(1): 93–99.

Kaski, S. (1997), *Data Exploration Using Self-organizing Maps.* PhD Thesis, Helsinki University of Tecnology.

Kasslin, M., J. Kangas and O. Simula (1992), Process State Monitoring using Self-Organizing Maps. In I. Aleksander and J. Taylor (eds.), *Artificial Neural Networks*, 2. Amsterdam: Elsevier Science, 1531–1534.

Kemke, C. and A. Wichert (1993), Hierarchical Self-Organizing Feature Maps for Speech Recognition. In Proceedings of the World Congress on Neural Networks (WCNN'93-Portland). Hillsdale, NJ: L. Erlbaum, 3: 45–47.

Kohonen, T. (1982), Self-organized formation of topologically correct feature maps. - *Biological Cybernetics* 43: 141–152.

Kohonen, T. (1988), The "Neural" Phonetic Typewriter. *Computer* 1988: 11–22.

Kohonen, T. (1991), The Hypermap Architecture. In T. Kohonen, K. Mäkisara, O. Simula and J. Kangas (eds.), *Artificial Neural Networks.* Amsterdam: Elsevier Science, 1357–1360.

Kohonen, T. (2001), *Self-Organizing Maps.* New York: Springer.

Kohonen, T., K. Makisara and T. Saramaki (1984), Phonotopic Maps – Insightful representation of phonological features for speech recognition. In *7th International Conference on Pattern Recognition*, Montreal, Canada, 182–185.

Koskela, T., M. Varsta, J. Heikkonen and K. Kaski (1997), Time Series Prediction using Recurrent SOM with local linear models. Research Report B 15. Lab. of Computational Engineering, Helsinki University of Technology, Finland.

Koskela, T., M. Varsta, J. Heikkonen and K. Kaski (1998), Temporal Sequence Processing using Recurrent SOM. In *Proceedings of the 2nd International Conference on Knowledge-Based Intelligent Engineering Systems*, Adelaide, Australia, 290–297.

Leinonen, L., J. Kangas and A. Juvas (1992), Dysphonia Detected by Pattern Recognition of Spectral Composition. *Journal of Speech and Hearing Research* 35: 287–295.

Midenet, S. and A. Grumbach (1994), Learning Associations by Self-Organization: The LASSO Model. *Neurocomputing* 6: 343–361.

Moshou, D. and H. Ramon (2000), Wavelets and self-organizing maps in financial time series analysis. *Neural-Network-World* 10: 231–238.

Murtagh, F. (1995), Interpreting the Kohonen self-organising map using contiguity-constraint clustering. *Pattern Recognition Letters* 16: 399–408.

Principe, J.C. and L. Wang (1995), Non-Linear Time Series Modeling with Self-Organizing Feature Maps. In F. Girosi et al. (eds.), *Neurals Networks for signal processing*. Proceedings of the 5th IEEE Workshop on Neural Networks for Signal Processing. New York: Institute for Electrical and Electronics Engineers, 11–20.

Ritter, H. (1994), Parametrized Self-Organizing Maps for Vision Learning Tasks. In M. Marinaro and P.G. Morasso (eds.), *Proceedings of the International Conference on Artificial Neural Networks* (ICANN 94-Sorento). London et al.: Springer, 803–810.

Ritter, H., T. Martinetz and K. Schulten (1989), Topology-conserving maps for learning visuo-motor coordination. *Neural Networks* 2: 159–168.

Simula, O. and E. Alhoniemi (1999), SOM Based Analysis of Pulping Process Data. In J. Mira et al. (eds.), Engineering applications of bio-inspired neural networks. Proceedings of International Work-Conference on Artificial and Natural Neural Networks (IWANN '99-Alicante). Lecture Notes in Computer Science 1607. New York: Springer, 2: 567–577.

Simula, O., E. Alhoniemi, J. Hollmén and J. Vesanto (1996), Monitoring and modeling of complex processes using hierarchical self-organizing maps. In *Proceedings of the IEEE International Symposium on Circuits and Systems* (ISCAS'96-Atlanta, GA). Piscataway, NJ: IEEE Operations Center, 73–76.

Tryba, V. and K. Goser (1991), Self-Organizing Feature Maps for Process Control in Chemistry. In T. Kohonen, K. Mäkisara, O. Simula and J. Kangas (eds.), *Artificial Neural Networks*. Amsterdam: Elsevier Science, 847–852.

Ultsch, A. (1993), Self-Organized Feature Maps for Monitoring and Knowledge Acquisition of a Chemical Process. In S. Gielen and B. Kappen (eds.), Proceedings of the International Conference on Artificial Neural Networks (ICANN'93-Amsterdam). London et al.: Springer, 864–867.

Ultsch, A., G. Guimarães and W. Schmidt (1996), Classification and Prediction of Hail using Self-Organizing Neural Networks. In Proceedings of the International Conference on Neural Networks (ICNN'96-Washington, DC). Piscataway, NJ: IEEE Service Center, 1622–1627.

Ultsch, A. and H.P. Siemon (1990), Kohonen's Self-Organizing Neural Networks for Exploratory Data Analysis. In *Proceedings of the International Neural Network Conference* (INNC'90-Paris). Dordrecht: Kluwer Academic, 305–308.

Utela, P., J. Kangas and L. Leinonen (1992), Self-Organizing Map in Acoustic Analysis and On-Line Visual Imaging of Voice and Articulation. In I. Aleksander and J. Taylor (eds.), *Artificial Neural Networks*, 2. Amsterdam: Elsevier Science, 791–794.

Vesanto, J. and E. Alhoniemi (2000), Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks* – Special Issue on Data Mining 11 (3): 586–600.

Voegtlin, T. and P. Dominey (2001), Recursive Self-Organizing Maps. In N. Allinson, H. Yin, L. Allison and J.Slack (eds.), *Advances in self-Organizing Maps*. London et al.: Springer, 210–215.

Von Harmelen (1993), *Time-dependent self-organizing feature map for speech recognition*. Master's thesis, University of Twente, Netherlands.

Walter, J. and H. Ritter (1996), Investment learning with hierarchical PSOM. In D. Touretzky, M. Mozer and M. Hasselmo (eds.), Advances in Neural Information Processing Systems 8 (NIPS'95), 570–576.

Walter, J.A. and K.J. Schulten (1993), Implementation of Self-Organizing Neural Networks for Visual-Motor Control of an Industrial Robot. *IEEE Transactions on Neural Networks* 4 (1): 86–95.

Victor Zarnowitz

Modern Trends and Their Effects on International Business Cycles

1. Globalization, New Information Age, and Monetary Regimes

Recent years have seen a remarkably strong and pervasive integration of markets for goods and services and for capital. Will this new globalization cause a significant increase in the synchronization of economic and financial fluctuations between the many different countries involved? According to a seemingly plausible hypothesis (Zarnowitz 1985) the process could speed up the international transmission of business cycles substantially. A conversion of more dispersed national fluctuations into roughly synchronized international cycles is apt to prove increasingly destabilizing globally.

To be sure, globalization conveys many major advantages on the economies that open up to the international trade and flows of capital and labor. Multinational firms find less expensive sources of the production factors they use, consumers benefit from cheaper imports, poor countries from more opportunities for exports to rich countries, etc. But the process creates some difficult problems, too, especially when it is reinforced by the concurrent technical progress and changes in the economic and financial institutions and structures. Open borders invite all that want to or can be gainfully moved, legally or illegally. People move to where they earn more, jobs to where they cost less. When flows of useful goods, services and investments are accommodated readily with less costly controls, then so are often the flows of bad and even criminal products (e.g., viruses infecting computers and people) or projects (e.g., money to help terrorists).

However, it must be noted that the same period of the past thirty years or so witnessed other major events and developments that had their own effects on national and international economic instability. The ongoing "computer revolution" generated rapid advances of powerful new information technologies and their business applications, which likewise spread worldwide. Many expect the new information age to bring on improvements in data processing and data quality, in business conditions analysis and forecasting. If so, then this should result in better stabilization policies and greater moderation of business cycles. Thus, the national cycles may become milder as well as more synchronized. The net effect on global instability would be uncertain; moreover, all of this may add up to a long and bumpy process.

There is also the important matter of changes in the prevailing monetary and exchange-rate regimes. The international gold exchange standard set up in 1946 following the Bretton Woods conference was eliminated when the U.S. closed the gold window in August 1971. This removed the last link to the ability of the public to convert, at fixed price, paper money into gold, a valuable real commodity, and vice versa. Full convertibility means that the supply of money is determined by the demand for money so that it cannot be controlled by the government. Therefore, with the end of the last global convertible regime and the move to a managed float, the national monetary and fiscal policies grew more independent and potentially more effective as tools of countercyclical, stabilizing action.

Economic history and the chronology of business cycles can help sort out the mixed effects of modem trends. I shall use new data from The Conference Board (TCB) for the recent era but find it useful to provide first an overview of fluctuations under different monetary regimes (for more on this subject, see Bergman et al. 1998).

2. Business Cycles under the Gold Standard before World War I

The United States switched from the gold-silver bimetallism to gold de facto in 1834, de jure in 1900. From 1834 until 1933, the mint price of gold was fixed at $ 20.67 per ounce. Great Britain adopted the gold standard as early as 1717, other major countries in the 1870s. The last quarter century before World War I, the classical gold standard period, was a time of great prosperity, free trade and free international movement of labor and capital: the first era of globalization and probably the widest so far.

Fixing the price of gold ensured fixed exchange rates among the participating countries. The "needs of trade" determined the demand for, and hence the supply of, money. Large gold discoveries resulted in temporary global surges in money and prices, as well as lower central bank discount rates, all of which discouraged future gold production. Shortages of gold would have the opposite temporary effects. So short-term procyclical price level movements did occur, but persistent inflation or deflation did not: there was, in fact, a long-term price stability. Contemporary economists for the most part appreciated the latter fact highly but were critical of the system allowing too much short-term variability.

DOI https://doi.org/10.3790/978-3-428-52425-9 | Generated on 2025-02-22 16:18:59

Table 1

Business Cycles in Great Britain and the United States: Monthly Dates and Relative Timing of Peaks and Troughs
1854–1918 and 1919–1938

Peaks		GB/US	Troughs		GB/US
US	GB		US	GB	
06/1857	09/1857	+3	12/1854	12/1854	0
10/1860	09/1860	–1	12/1858	03/1858	–9
04/1865	03/1866	+11	06/1861	12/1862	+6
06/1869			12/1867	03/1868	+3
10/1873	09/1872	-13	12/1870		
03/1882	12/1882	+9	03/1879	06/1879	+3
03/1887			05/1885	06/1886	+13
07/1890	09/1890	+2	04/1888		
01/1893			05/1891		
12/1895			06/1894	02/1895	+8
06/1899	06/1900	+12	06/1897		
09/1902	06/1903	+9	12.1900	09/1901	+9
05/1907	06/1907	+1	08/1904	11/1904	+3
01/1910			06/1908	11/1908	+5
01/1913	12/1912	–1	01/1912		
08/1918	10/1918	+2	12/1914	09/1914	–3
01/1920	03/1920	+2	03/1919	04/1919	+1
05/1923	11/1924	+18	07/1921	06/1921	–1
10/1926	03/1927	+5	07/1924	07/1926	n.m.
08/1929	07/1929	–1	11/1927	09/1928	+10
05/1937	09/1937	+4	03/1933	08/1932	–7
			06/1938	09/1938	+3
Pre–1918	Mean	6,13	Pre–1918	Mean	3,45
	Median	6		Median	3
	St. Dev	4,55		St. Dev	6,01
Post–1918	Mean	7,25	Post–1918	Mean	5.00
	Median	5		Median	2
	St. Dev	7,27		St. Dev	10,83

Source: Burns/Mitchell 1946: 78–79, Table 16. Matching and calculations of leads and lags are my own. GB/US denotes leads (–) or lags (+), in months, of the GB dates relative to the US date. All means and medians are average lags (+ omitted). – n.m.: not matched (lags of two years or more).

In fact, the data suggest that cyclical instability was high in the gold standard era. Table 1 compares the timing of business cycle peaks and troughs in the United States and Great Britain using the monthly reference cycle chronology of the National Bureau of Economic Research (NBER), which goes back to 1854. This list includes 15 trough-to-trough business cycles in the U.S., 1854–1914, and ten such cycles for Great Britain in the same period. All the G.B. turning points can be matched with like U.S. turning points. All but three of the 11 peaks and all but three of the 11 troughs occurred earlier in the United States than in Great Britain (note that lags marked + prevail over leads marked – in the columns labeled GB/US). Interestingly, then, it appears that the U.S. economy was more cyclical and subject to earlier cycles even in the era when it was Great Britain that dominated the global economy.

As a necessary caveat, let us remember that these comparisons are based on data for early cycles, which are quantitatively and qualitatively limited, but also on the best source there is. The charge that the reference dates of Burns/Mitchell (1946) are basically flawed by reflecting fluctuations in growth rates rather than levels of economic activity, for example, does not stand a serious scrutiny. Still, some of the episodes are more dubious than the rest. In the U.S., 1869–70 may have witnessed a retardation rather than recession; it is also possible that 1887–88 and 1892–1900 were periods of very weak growth rather than actual declines (for more detail and countercritique, see Zarnowitz 1981; Romer 1986). Notably, no recessions occurred in the major foreign countries in the same periods. So omitting these episodes from Table 1 would strengthen considerably the correspondence of business cycles in US and GB in the gold standard era. The leads and lags varied from short to intermediate but averaged only 3–6 months.

Table 2 shows that the cycles in Europe were very well synchronized in 1879–1918; for the earlier times the data are too scanty to tell. It is clear that overall the classical gold standard era was marked by a high degree of synchronization of business cycles in the principal industrialized nations. The variability of cyclical amplitudes and durations was probably much greater over time in each of the major countries covered than across these countries. Some long and severe contractions and also financial crises occurred in each of the last three decades of the 19th century and early in the 20th century, but most of the declines had moderate durations and amplitudes. Relatively high and stable growth rates of economic activity prevailed in the U.S. in 1882–92 and 1903–13 (for evidence, see Glasner 1997; Zarnowitz 1981).

The sixteen U.S. business cycles in 1854–1919 (which include also one cycle that occurred during World War I) had average durations of 27 and 22 months for expansions and contractions, respectively; but when the three rather dubious episodes are omitted then the numbers would change to the more plausible 37 and 23 (Zarnowitz 1981: 230). For Great Britain, the corresponding mean duration estimates are similar, 42 and 21; for France, 32 and 26; and for Germany, 40 and 29.

3. The Interwar Period: 1919–1939

In the 1920's and 1930's, the U.S. economy suffered three major depressions, in 1920–21, 1929–33, and 1937–38. The first post-WWI cycle was short but virulent, with a highly inflationary and rapid boom followed by a uniquely sharp

Table 2

Business Cycles in Great Britain, France, and Germany:
Monthly Dates and Relative Timing of Peaks and Troughs
1867–1918 and 1919–1938

Peaks			F/GB	G/GB	Troughs			F/GB	G/GB
France	Germany	GB			France	Germany	GB		
11/1867		3/1866	+20		12/1865				
08/1870					10/1868		03/1868	+7	
09/1873		9/1872	+12		02/1872				
04/1878					08/1876				
12/1881	1/1882	12/1882	–12	–11	09/1879	02/1879	06/1879	+3	–4
01/1991	1/1890	9/1890	–8	–8	08/1887	08/1886	06/1886	+14	+2
03/1900	3/1900	6/1900	–3	–3	01/1895	02/1895	02/1895	–1	0
05/1903	8/1903	6/1903	–1	+2	09/1902	03/1902	09/1901	+12	+6
07/1907	7/1907	6/1907	+1	+1	10/1904	02/1905	11/1904	–1	+3
06/1913	4/1913	12/1912	+6	+4	02/1909	12/1908	11/1908	+3	+1
06/1918	6/1918	10/1918	–4	–4	08/1914	08/1914	09/1914	–1	–1
09/1920	5/1922	3/1920	+6	n.m.	04/1919	06/1919	04/1919	0	+2
10/1924	3/1925	11/1924	–1	+4	07/1921	11/1923	06/1921	+1	n.m.
10/1926		3/1927	–5		06/1925	03/1926	07/1926	–13	–4
03/1930	4/1929	7/1929	+8	–3	06/1927		09/1928	–15	
07/1933					07/1932	08/1932	08/1932	–1	0
06/1937		9/1937	–3		04/1935				
					08/1938		09/1938	–1	
	Pre–1918	Mean	1,22	-2,71		Pre–1918	Mean	4.50	1.00
		Median	–1	-3			Median	3	1
		St. Dev	10,01	5,47			St. Dev	5,95	3,16
	Post–1918	Mean	1.00	9.00		Post–1918	Mean	4.83	0,75
		Median	–1	4			Median	–1	1
		St. Dev	5,7	15,13			St. Dev	7,17	3,77

Source: Burns/Mitchell 1946: 78–79, Table 16. Matching and calculations of leads and lags are my own. F/GB and G/GB are leads (–) or lags (+), in months, of French and German dates relative to GB dates, respectively. – n.m.: not matched (lags of two years or more).

deflation, slump and recovery. The Great Depression of the 1930's consisted of a long and severe contraction, a sluggish and incomplete recovery, and a painful relapse late in the decade. Unquestionably, the interwar period was the worst, most unstable and depressed interval in modem economic history.

Britain experienced a cycle in 1919–21 similar to that in the U.S., but France had a much milder cycle, and in Germany a recession occurred over two years later and for different reasons. The two mild U.S. recessions in 1923–24 and 1926–27 had no clear counterparts in continental Europe.

All four countries participated in the depression of the late 1920's and early 1930's, but the original and worst experience by far was that of the United States. The decline in Britain was mainly imported, shorter, and followed by a forceful recovery. In France, the depression developed more slowly and was less severe but lasted longer (almost a decade, until 1938), probably due largely to a long overvaluation of the franc. Germany suffered an early and severe but relatively short depression, in part because as a major debtor country it pursued a very deflationary monetary policy to accumulate gold and avoid default (all of which changed drastically with Hitler's rise to power).

The Smoot-Hawley tariff of June 1930 raised U.S. import duties on great many products and let to a pernicious outbreak of protectionist "beggar-thy-neighbor" trade war policy moves. The U.S. export surplus initially increased, but the international trade in dollars declined by more than two thirds between 1929 and 1933. The tariff made exporting to the U.S. more difficult, raised the world demand for gold, and increased the deflationary pressures, especially on the debtor countries (Batchelder, Glasner 1995).

The 1937–38 depression in the United States had domestic sources: fall in profits and investment due to rises in unionization, wages, and other costs as well as shifts t more restrictive fiscal and monetary policies. Other economies did not suffer similar declines: Britain and France experienced mild recessions, Germany none. Overall, as shown by the lower panels in Tables 1 and 2, there was much less synchronization between business cycles in 1919–39 than in 1879–1914.

4. The Post-World War II Era

The 1944 Bretton Woods conference established a system of pegged exchange rates but allowed the member countries to alter their parities when facing "fundamental disequilibrium". Temporary disturbances were to be offset by domestic policy measures.

The plan was to avoid the turbulences of the interwar period by combining the flexibility of a floating-rate system with the stability of the gold standard rule in nominal terms. The system was in force until 1971.

After World War I, the harsh treatment of Germany and Austria by the Allies, through reparations, etc., contributed to the economic and financial troubles of the interwar period. This was avoided after World War II when the victorious Allies were much more magnanimous versus the vanquished Axis powers, partly because of the developing tensions of the Cold War. As a result of this, plus the better monetary and exchange-rate regime, there was far less global economic instability after the second than after the first world war.

Indeed, I find the Bretton Woods period, that is, mainly the 1950's and the 1960's, to have witnessed the least synchronization of the national business cycles. The United States experienced five recessions during this period: (1) 11/1948–10/1949; (2) 7/1953–5/1954; (3) 8/1957–4/1958; (4) 4/1960–2/1961; and

Table 3

Business Cycles in the United States, United Kingdom, France and Germany: Relative Timing of Peaks and Troughs
1973–2006

Peaks					Leads (–) or lags (+) in months		
US	UK	France	Germany		UK/US	France/US	Germany/ US
11/73	09/73	08/74	03/73		–2	+9	–8
01/80	11/79	03/80	03/80		–2	+2	+2
07/81							
07/90	05/90	02/92	02/92		–2	+19	+19
03/06			1/01				+1
				Mean	–2	10	7,5
				St. Dev	0	8,54	8,27

Troughs					Leads (–) or Lags (+) in Months		
US	UK	France	Germany		UK/US	France/US	Germany/ US
03/75	08/75	06/75	07/77		+5	+3	n.m.
07/80							
11/82	02/82	05/81	12/82		–9	–18	+1
03/91	03/92	07/93	09/93		+12	n.m.	n.m.
11/01				Mean	2.67	–7.5	1
				St. Dev	10,69	14,85	0

Source: US dates: NBER; all other dates: TCB. – n.m.: not matched (lags of two years or more).

(5) 12/1969–11/1970. Each of them was due to domestic causes, including inventory liquidation, drop in profits and business capital investment (in 57/58), and restrictive monetary policy (in 69/70). None of these U.S. recessions had a clear counterpart in either Western Europe or the Far East.

The advent of peace in 1945 in Western Europe and Japan was followed by a restoration of free markets and sound currencies, great reconstruction to undo the huge damages of the war, and rapid growth from initially low levels of output and other measures of total economic activity. Growth slowdowns largely replaced recessions, i.e., level declines in the next two decades; for example, the first post-WW II recessions in West Germany and Japan occurred in 1966–67 and 1973–75, respectively. Real GDP declines, were few and far between (Britain in 1952, Belgium in 1958, Switzerland in 1949 and 1958). So, while leading in economic power, the U.S. also was for some time standing out for cyclical instability.

The lack of international diffusion of recessions during the early post-1945 period is remarkable. But the growth cycles, i.e., sequences of rises and declines in the detrended indicators of macroeconomic activity, show considerable synchronization in the Bretton Woods era, as they do generally at other times as well (The adjustment for inter-country differences in growth tends to increase synchronization; that is, growth cycles are generally more diffused internationally than business cycles.).

In the 1970s, growth slowed in many countries as world oil prices rose sharply twice. Soon business cycles reappeared everywhere. Table 3 compares the chronologies for U.S., U.K., France and Germany based on The Conference Board composite indexes of monthly coincident indicators and on quarterly real GDP series for these countries (cf. e.g. TCB 2006). Recessions occurred in all four countries in 1973–75 and 1980–82 (but U.S. had two back-to-back recessions in the early 1980's, the other countries each only one). The U.S. had a recession in 1990–91, U.K. in 1990–92, France and Germany in 1992–93. The 2001 recession in the U.S. was accompanied by only slowdowns in U.K. and France but by an about concurrent recession in Germany.

References

Batchelder, R. and D. Glasner (1995), Debt, Deflation, the Gold Standard, and the Great Depression. In George Edward Durell Foundation (ed.), *Money and Banking*: The American Experience. Fairfax, VA: George Mason University Press, 277–310.

Bergman, U.M., M.D. Bondo, and L. Jonung (1998), Historical Evidence on Business-Cycles: The International Experience. In J.C. Fuhrer and S. Schuh (eds.), Beyond Shocks: What Causes Business Cycles? Ferderal Reserve Bank of Boston Conference Series 42. Boston, 65–113.

Burns, A.F. and W.C. Mitchell (1946), Measuring business cycles. NBER Studies in Business Cycles 2. NBER, New York.

Glasner, D. (ed.) (1997), Business Cycles and Depressions – An Encyclopedia. New York.

Romer, C. (1986), Is the Stabilization of the Postwar Economy a Figment of the Data? *American Economic Review* 76 (2): 314–334.

TCB – The Conference Board (ed.) (2006), *Business Cycle Indicators* 11 (4).

Zarnowitz, V. (1981), Business cycles and growth. In W.J. Mücke und A.E. Ott (Hrsg.), *Wirtschaftstheorie und Wirtschaftspolitik*. Gedenkschrift für Erich Preiter. Passau: 475–508.

Zarnowitz, V. (1985), Recent Work on Business Cycles in Historical Perspective: A Review of Theories and Evidence. *Journal of Economic Literature* 23 (June): 523–580.

Zarnowitz, V. (1992), Business Cycles. Theory, History, Indicators, and Forecasting. NBER Studies in Business Cycle Research 27. Chicago and London: University of Chicago Press.

List of Authors

Marlene Amstad, Swiss National Bank, Fraumünsterstraße 8, 8001 Zürich / Switzerland
marlene.amstad@snb.ch

Claudia Becker, Lehrstuhl für Statistik, Wirtschaftswissenschaftliche Fakultät, Martin-Luther-Universität Halle-Wittenberg, 06099 Halle
Claudia.Becker@wiwi.uni-halle.de

Gabriela Dick-Guimarães, Johann-Clanze-Straße 53, 81369 München, gabriela.dick.guimaraes@gmail.com

Ursula Garczarek, Roche Diagnostics GmbH, B. 241 R. 401, Nonnenwald 2, 82377 Penzberg
ursula.garczarek@roche.com

Ullrich Heilemann, Institut für Empirische Wirtschaftsforschung, Wirtschaftswissenschaftliche Fakultät der Universität Leipzig, Marschnerstraße 31, 04109 Leipzig
heilemann@wifa.uni-leipzig.de

Malte Knüppel, Deutsche Bundesbank, Economic Research Centre, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main
Malte.Knueppel@bundesbank.de
Der Autor vertritt seine persönliche Auffassung, die nicht notwendigerweise die Meinung der Deutschen Bundesbank widerspiegelt.

Bernd Lucke, Department of Economics, von Melle-Park 5, Universität Hamburg, 20146 Hamburg
lucke@uni-hamburg.de

Katharina Morik, Lehrstuhl Informatik 8, Universität Dortmund, 44221 Dortmund
morik@Is8.cs.uni-dortmund.de

Heinz Josef Münch, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstraße 1-3, 45128 Essen
Muench@rwi-essen.de

Stefan Rüping, Fraunhofer Institut AIS, Schloss Birlinghoven, 53754 Stankt Augustin
stefan.rueping@ais.fraunhofer.de

Bernd Schips, Tanneichenstraße. 5, 9010 St. Gallen / Switzerland
bschips@ethz.ch

Winfried Theis, MSR Consulting, Bonner Straße 211, 50968 Köln
Winfried.Theis@msr.de

Claus Weihs, Lehrstuhl „Computergestützte Statistik“, Fachbereich Statistik, Universität Dortmund, 44221 Dortmund
weihs@statistik.uni-dortmund.de

Victor Zarnowitz, The Conference Board, 845 Third Avenue, New York, NY 10022-6679, USA
victor.zarnowitz@conference-board.org