Descriptive Studies on Stylized Facts of the German Business Cycle

W. Theis[1], K. Vogtländer, C. Weihs[2]

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

[2] Lehrstuhl für Computergestützte Statistik
Universität Dortmund, D–44221 Dortmund
e-mail: weihs@gigamain.statistik.uni-dortmund.de

Abstract

Thirteen Stylized Facts of the german economy are studied with different descriptive statistical methods. The results of this study are considered with respect to other results from Project B3 "Multivariate Bestimmung und Untersuchung von Konjunkturzyklen".

1 Introduction

This paper sums up different descriptive methods we applied to the 13 Stylized Facts for the German Economy, which have been selected by Heilemann and Münch for the consideration of the business cycle problem. They are listed in the Table 1:

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

Table 1: The 13 Stylized Facts

All variables were observed quarterly from 1955/4 until 1994/4, in total 157 observations (price index base=1991, y=yearly growth rates).

For the examination of the data with respect to business cycle phases we use the same four phase scheme as Heilemann and Münch (1999) with phases called (in their natural course of appearance) 'upswing' (1), 'upper turning points' (2), 'downswing' (3), and 'lower turning points' (4). For convenience this classification will be referred to as variable BC. The quarters are distributed among the four phases as follows:

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

For a detailed characterization in economic terms cp. Heilemann and Münch (1999).

The paper is organized as follows: In section 2 the statistical behaviour of the different stylized facts is discussed in more detail. Section 3 compares the behaviour in the four business cycle phases as proposed by Heilemann and Münch. Finally in section 4 a time series study is performed on the data set.

2 Empirical Distribution of Stylized Facts

2.1 Histograms

In the histograms of figure 1 the GNP (Y), the Private Consumption (C) and their respective price index / price deflator are shown. GNP (Y) has a steeper distribution on the right hand side. For Private Consumption (C) and the Price Deflator of the GNP (PY) the histograms show nearly symmetric behaviour. The Consumer Price Index (PC) shows two peaks, a high one around 3 and a smaller one around 5.5.

In figure 2 Wage and Salary Earners (L) shows two peaks a small one between -2 and -1 and a high peak between 1 and 2, whereas Unit Labour Cost (LC) is asymmetric, steep on the left hand side and flat on the right hand side. The two investment variables have nearly symmetric histograms. Investment in Construction (IC) has the main part of observations concentrated in the middle, in contrast IE is more equally distributed over its range.

Figure 3 contains all financial variables and the Export/Import-rate (X). Here we encounter the histograms differing most obviously from a symmetric distribution. The Short Term Interest Rate (RS) for example has a peak on the left hand side of the histogram and then decreases toward the right. The histogram of Government Deficit (GD) jumps up to its peak around -4 then decrases slowly again.

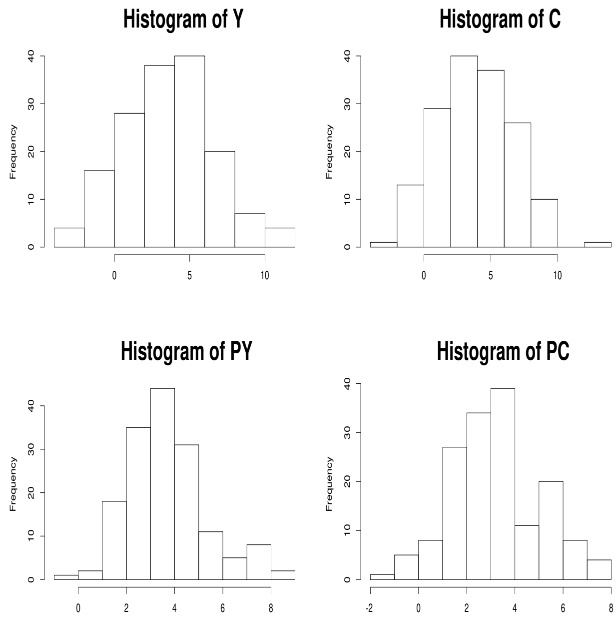

Figure 1: Histograms of Y, PY, C and PC

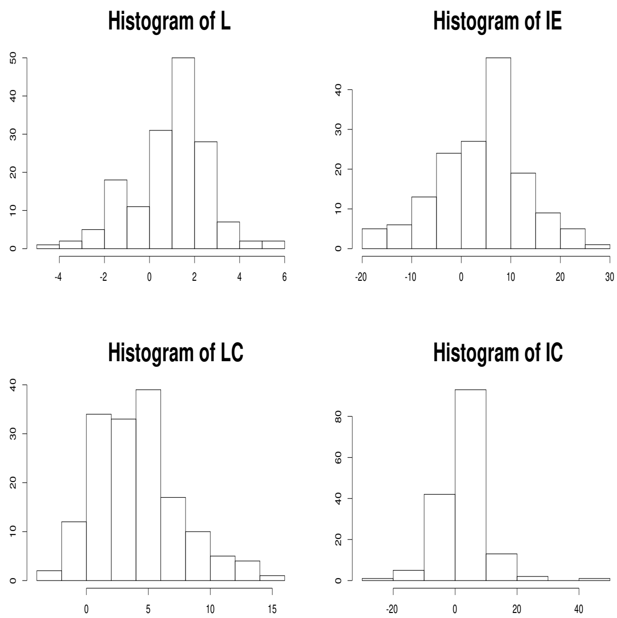

Figure 2: Histograms of L, LC, IE and IC

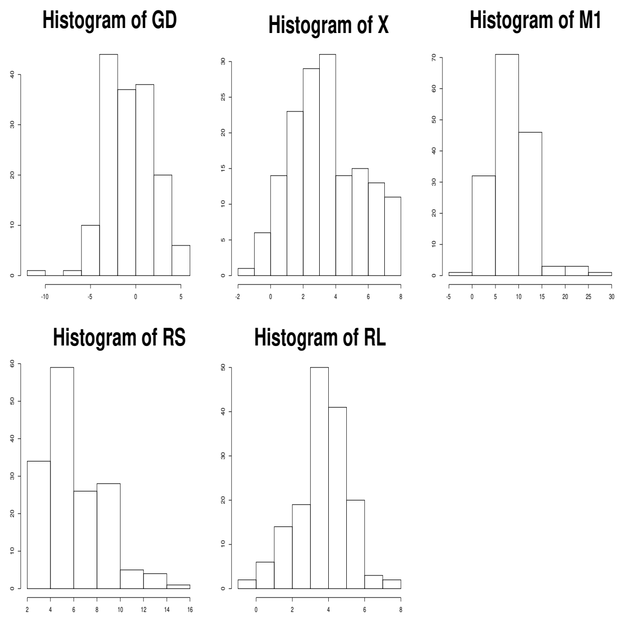

Figure 3: Histograms of GD, X, M1, RS and RL

2.2 Scatterplots

We now consider parts of the scatterplot-matrix of all 13 variables in which structure can be seen. In figure 4 nearly all variables show linear dependencies. This is not surprising since the variables are defined as linear combinations, or are known to be linearly dependent on each other (cp. Heilemann (1998)). The GNP (Y) is defined to be

$$Y = NetX + \text{state consumption} + C + I.$$

where $I = IE + IC$ and $NetX = Export - Import$. The other variables are considered by the experts at RWI to follow the following equations (the formulas contain only the variables of our dataset, dummy variables and unknown variables are collected in the term SV!) The indices indicate lag–variables, variables without index correspond to the current observation. α , β and γ represent unknown coefficients.

$$\begin{aligned} C &= -\alpha RS + SV \\ L &= \alpha(C + I + NetX) - \beta LC_{-2} + L_{-1} + SV \\ IE &= \alpha(C + I + NetX) - \beta(RL_{-2} - PY_{-2}) - \gamma LC_{-3} + SV \end{aligned}$$

Investment in Construction (IC) has a similar connection as IE to the other variables.

In figure 5 another part of the scatterplot matrix is shown. The linear connections between variables LC and PY as well as between PY and PC might be intuitively clear.

One can see high correlations for the above discussed pairs of variables in the correlation matrix, also (cp. Table 2).

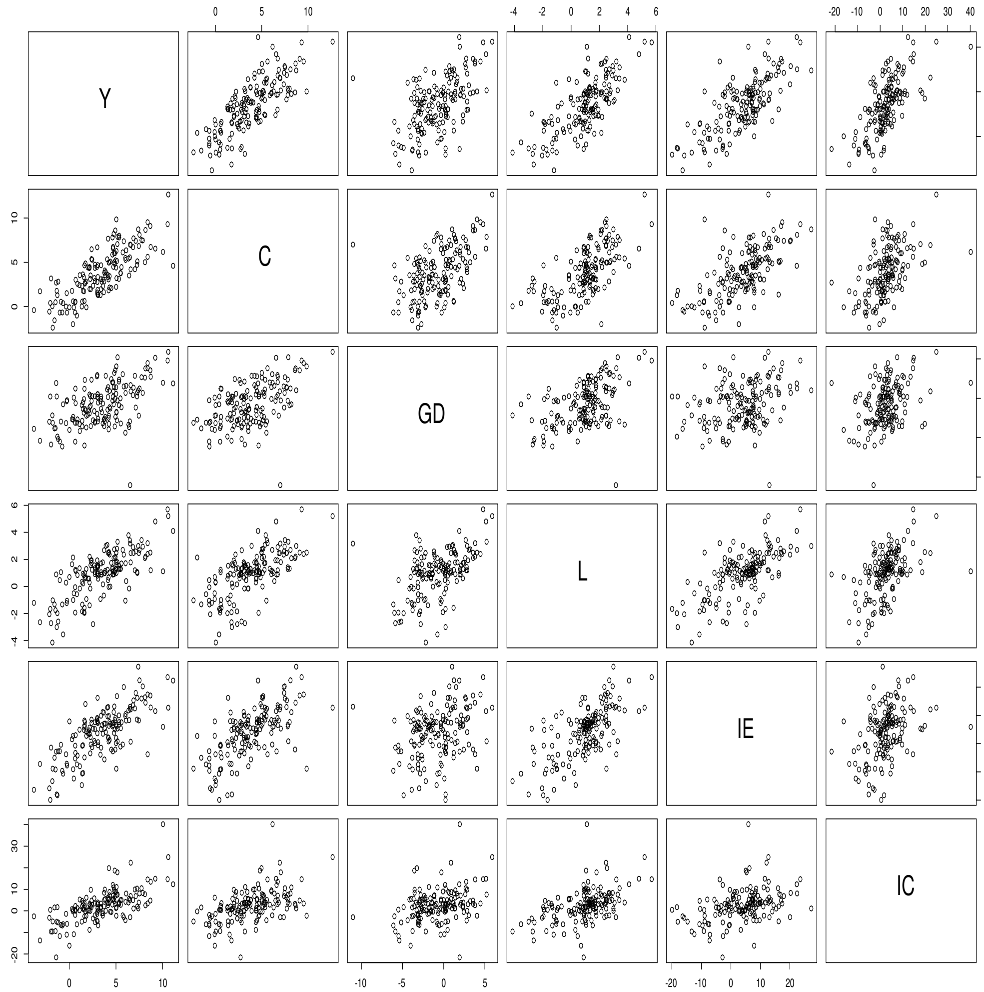

Figure 4: Scattermatrix of Y, C, GD, L, IE and IC

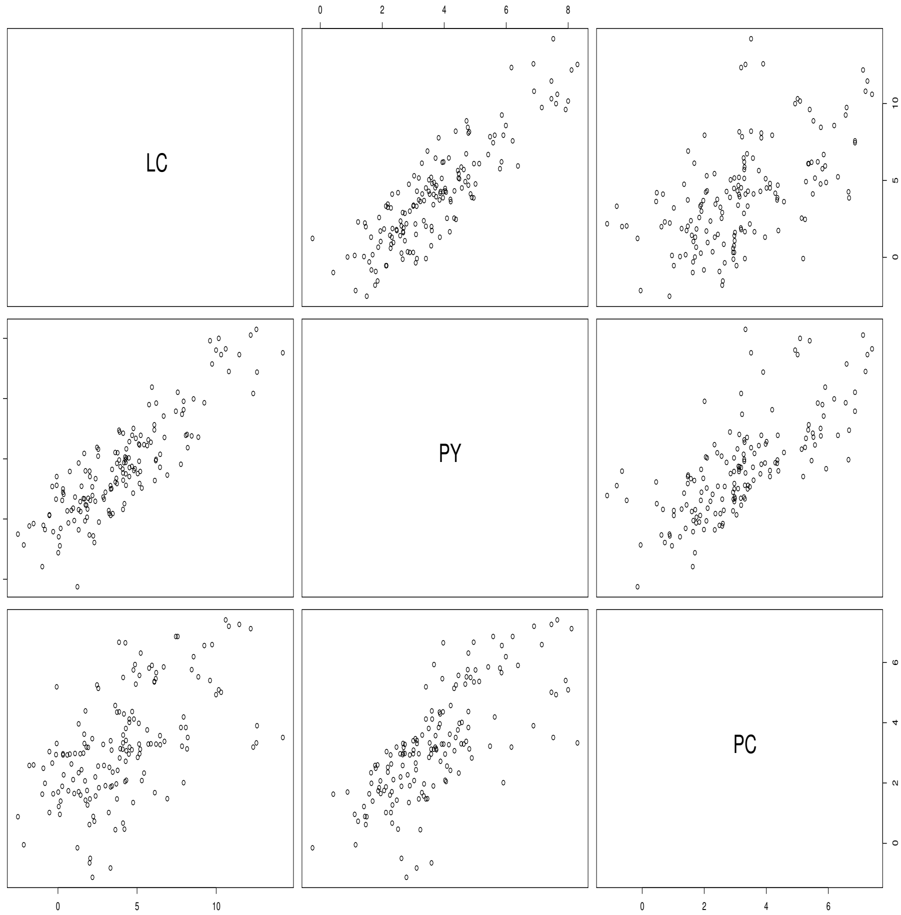

Figure 5: Scattermatrix of LC, PY and PC

	Y	C	GD	L	X	M1	IE	IC	LC	PY	PC	RS	RL
Y	1	**0.78**	**0.54**	**0.74**	-0.13	0.31	**0.74**	**0.68**	-0.16	-0.17	-0.35	-0.24	-0.10
C	0.78	1	**0.53**	**0.66**	-0.17	0.40	**0.64**	**0.51**	0.12	0.02	-0.34	-0.30	-0.37
GD	0.54	0.53	1	**0.48**	-0.16	-0.023	0.30	0.35	0.11	-0.06	-0.27	-0.21	-0.26
L	0.74	0.66	0.48	1	-0.09	0.18	**0.67**	0.49	0.16	0.055	-0.20	0.05	-0.23
X	-0.13	-0.17	-0.16	-0.09	1	0.18	-0.16	-0.04	-0.34	-0.27	-0.29	0.06	0.21
M1	0.31	0.40	-0.02	0.18	0.18	1	0.31	0.18	-0.15	-0.01	-0.14	-0.36	-0.18
IE	0.74	0.64	0.30	0.67	-0.16	0.31	1	0.39	-0.08	-0.09	-0.37	-0.27	-0.18
IC	0.68	0.51	0.35	0.49	-0.036	0.18	0.39	1	-0.18	-0.18	-0.27	-0.19	-0.09
LC	-0.16	0.12	0.11	0.16	-0.34	-0.15	-0.08	-0.18	1	**0.87**	**0.57**	0.42	**-0.66**
PY	-0.17	0.02	-0.06	0.05	-0.27	-0.01	-0.09	-0.18	0.87	1	**0.72**	**0.49**	**-0.66**
PC	-0.35	-0.34	-0.27	-0.20	-0.30	-0.14	-0.37	-0.27	0.57	0.72	1	**0.62**	-0.12
RS	-0.24	-0.30	-0.21	0.05	0.06	-0.36	-0.27	-0.19	0.42	0.49	0.62	1	0.15
RL	-0.10	-0.37	-0.26	-0.23	0.21	-0.18	-0.18	-0.09	-0.66	-0.66	-0.12	0.15	1

Table 2: Correlation Matrix of the stylized facts

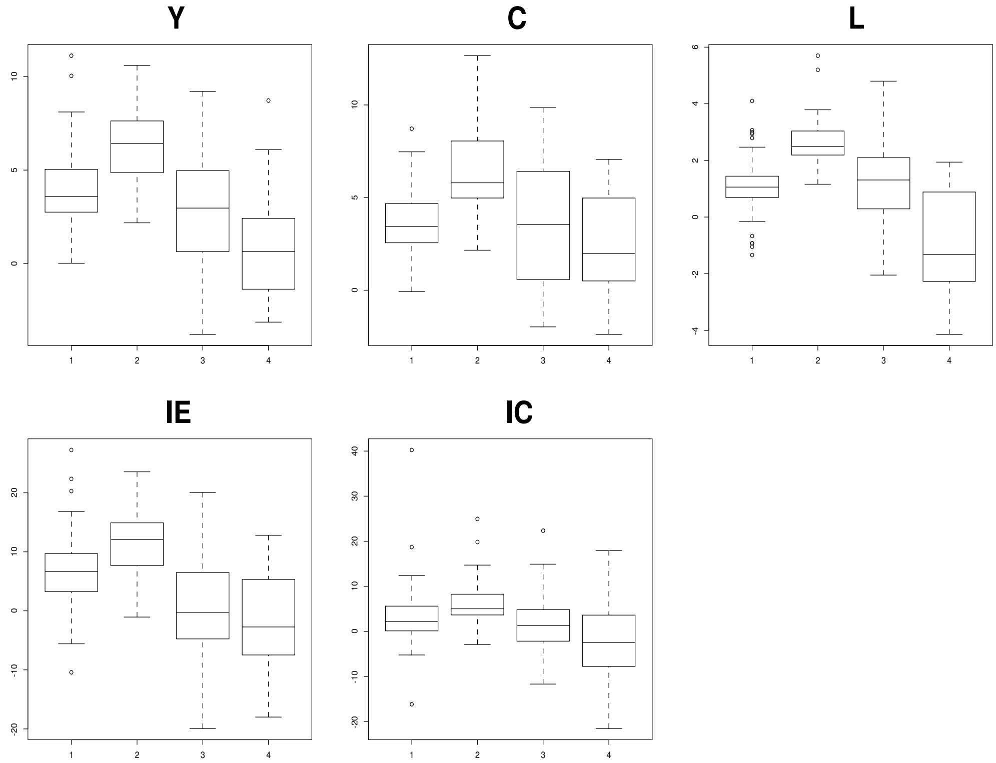

Figure 6: Variables Y, C, L, IE and IC against phases

3 Results with respect to RWI–Phases

In this section we consider the stylized facts with respect to the classification into the business cycle phases of Heilemann and Münch.

The figures 6, 7 and 8 show parallel boxplots of the variables separated into phases. The boxplots of the main phases, upswing (1), downswing (3) resp., have for all variables large overlaps with those of the corresponding turning point phases. This is not very astonishing since the turning point phases are defined as times of transition from one main phase to the other.

Figure 6 shows parallel boxplots of five of the variables from figure 4 separated into phases. The similarity of the boxplots of these variables is easily explained by the linear relations between them. Interesting is that all these variables have smaller ranges in upswing and upper turning point phase than in downswing and lower turning point phase. These five variables behave as one would expect for indicators for the economic behaviour — they are higher in median for the upper turning points than for the lower turning points.

Figure 7 shows the parallel boxplots for variables LC, PY and PC separated into phases. They show again smaller ranges for upswing and upper turning point, but here the median of the upper turning points is not higher than for the lower turning points.

The variables GD, X, RL and M1 — as shown in figure 8 — have similar medians and

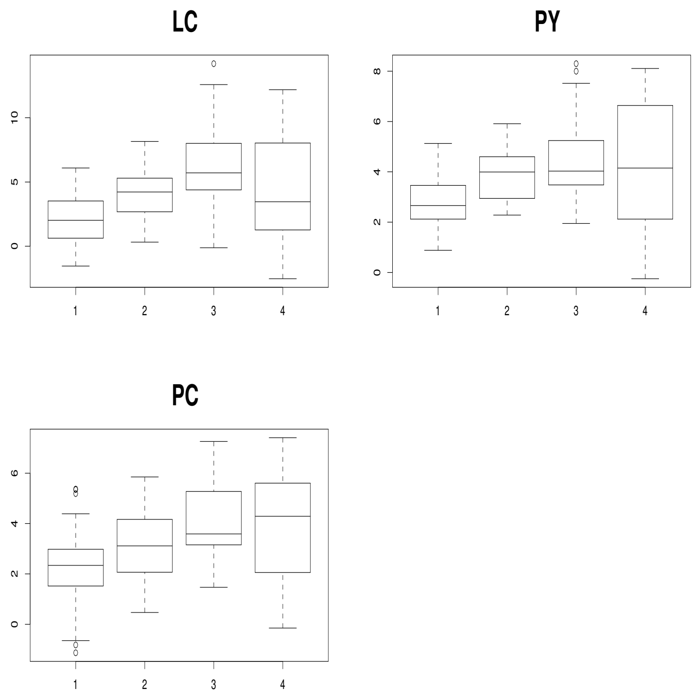

Figure 7: Variables LC, PY and PC against phases

ranges in nearly all phases. These variables are shown in Weihs, Röhl, Theis (1999) to be less important than the other variables with respect to minimal error rates in discriminant analysis. The big overlaps of the boxplots in the different phases for these variables may be a reason for this.

The short time interest rates (RS) shows an exceptional behaviour. The ranges in upswing and downswing differ immensely. But the turning point phases have similar ranges, though the box of the upper turning point phase (2) overlaps completely with the one of downswing. Even the medians of these phases lie near together so here upper turning point phase and downswing are closer together than upswing and upper turning point phase.

The big overlaps even of the main phases in all variables lead to great overlaps of the different phases even in 13–dimensional space. This explains the relatively high error rates of discriminant analysis using all 13 variables as can be seen in Weihs, Röhl, Theis (1999).

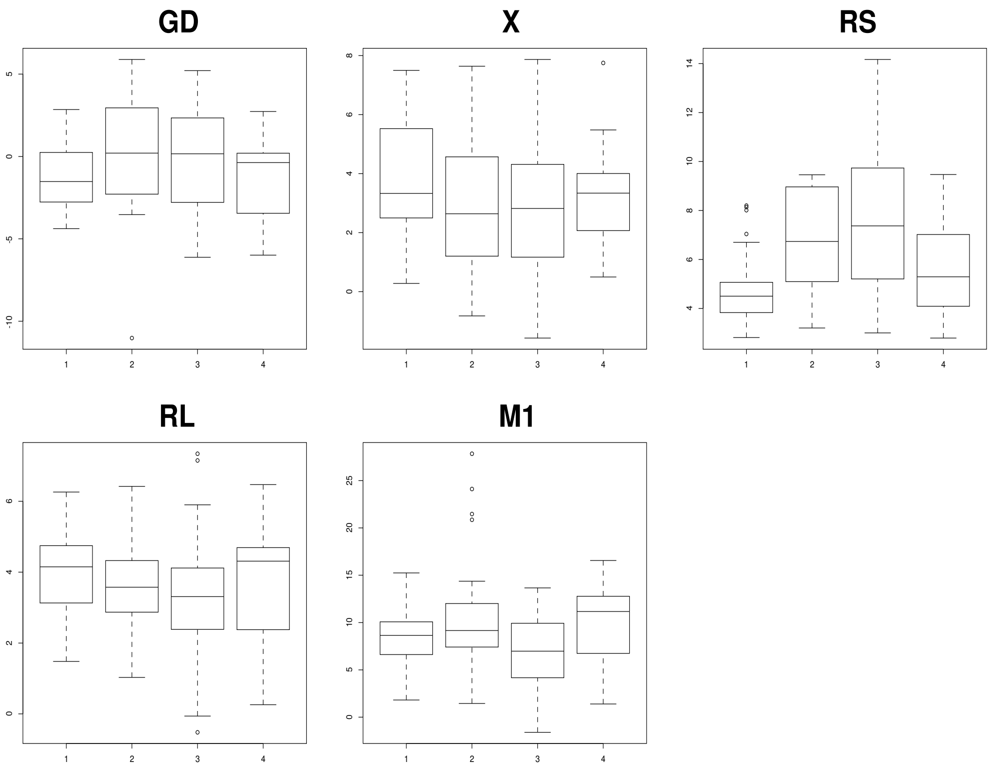

Figure 8: Badly separated variables (GD, X, RS, RL, M1)

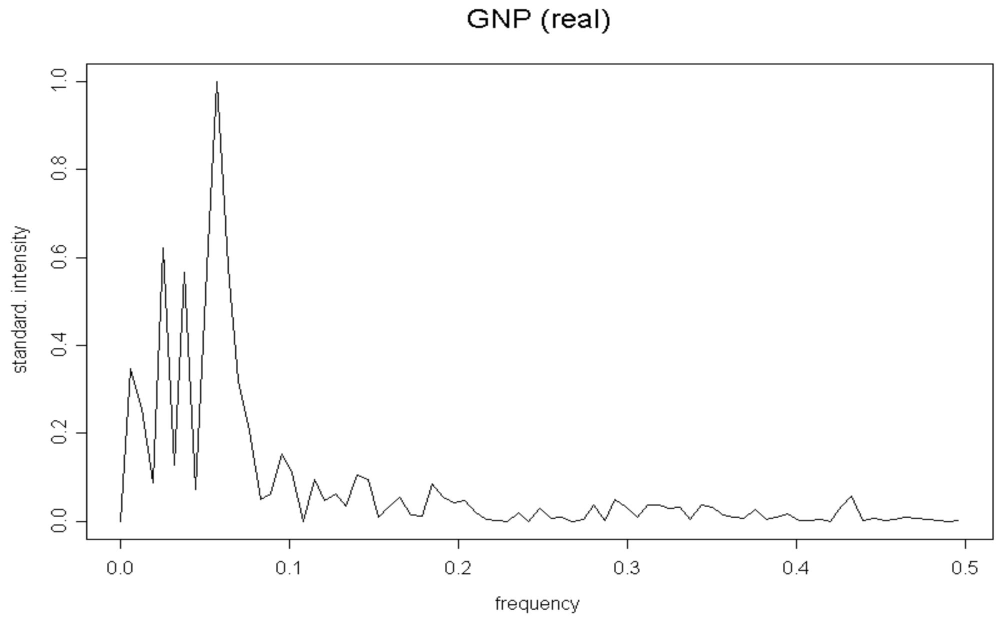

Figure 9: Periodogram for the GNP (Y)

4 Time Series study

A natural point of view is to consider the observations as time series. The aim of this section is to study the frequencies that cause the periodical behavior of the observations. The regular behavior may be interpreted as a summation of harmonics. Some of the time series are not stationary. They have at least a linear trend. Using a linear autoregression model of first order, the linear trend was eliminated. The relevant frequencies are found by using the periodogram which demands for stationarity (Schlittgen, Streitberg 1995).

Here we treat the grouping variable BC like the other variables. This means particularly that BC is used like a metric variable, although it is a nominal variable with a cyclical sequence, in order to make the results comparable to the results of the stylized facts.

The periodogram reports the intensity I of any Fourier-frequency λ_i between 0 and 0.5. The intensities of each variable are divided by their greatest value to make the peridograms comparable. So the values are between 0 and 1. We call a frequency λ_i relevant if the intensity is greater than 0.1 and the intensity of the neighbouring frequencies are distinctly lower.

Table 3 lists the first 23 of 78 Fourier-frequencies. The number of the relevant frequencies differs from variable to variable. E.g., the variables X, PY, PC, and RS have only 3, but the variable IC has 21 frequencies with $I_{stand}(\lambda) > 0.1$. IC has many high relevant frequencies, corresponding to periods of length 2.275 – 6.826 (the numbers of those Fourier-freqeuncies are between 23 and 78) so they are omitted in Table 3 (Figure 10). One reason for this might be that the building industry is dependent on season and weather. A smoother example is the periodogram of the variable Y (Figure 9).

No.	frequency	period	Y	C	GD	L	X	M1	IE	IC	LC	PY	PC	RS	RL	BC
1	0.0064	157.000	*		*	*	*			*	*	*	*		•	*
2	0.0127	78.500		*				*								
3	0.0191	52.333			*											*
4	0.0255	39.250	*	*		*	•		*		*	*	*	*		
5	0.0318	31.400			*					*					*	*
6	0.0382	26.167	*	*				*	*							
7	0.0446	22.429			*											
8	0.0510	19.625							*					•		
9	0.0573	17.440	*	*	•	*					*	*	•		*	*
10	0.0637	15.700								*						
11	0.0701	14.273		*				*	•					*		
12	0.0764	13.083			•						•					
13	0.0828	12.077								*						•
15	0.0955	10.467	•			•		*								•
16	0.1019	9.813								*	•					
18	0.1146	8.722								*						•
19	0.1210	8.263						*								
20	0.1274	7.850								•						
21	0.1338	7.476														•
22	0.1401	7.136	•					•								
23	0.1465	6.826		•						•	•				•	

Table 3: Relevant frequencies (No.: number of the Fourier-frequency (p=78);
*: $I_{stand} > 0.2$; •: $I_{stand} > 0.1$)

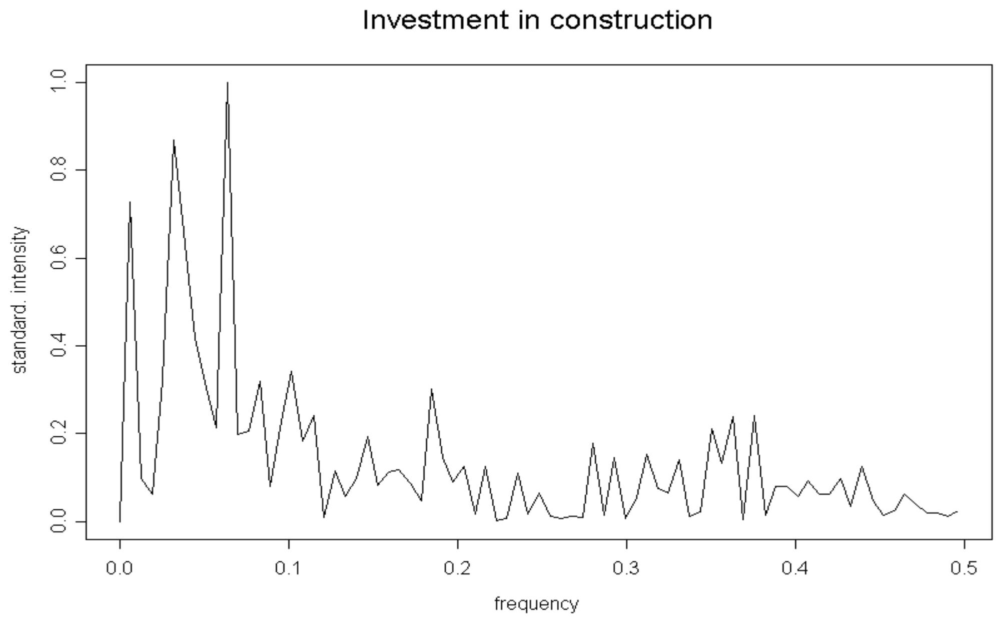

Figure 10: Periodogram for Investment in Constructions (IC)

The Fourier-frequencies 0.0064, 0.0255, and 0.0573 appear in 9 or 10 variables of 14 (Table 3). They correspond to the periods 157, 39.25, and 17.44. Other important frequencies are 0.0382 (period 26.167), 0.0701 (15.7), and 0.1465 (6.826). The variable BC has 8 relevant frequencies, particularly the frequencies 0.0064 and 0.0573. Overall, for many variables the same Fourier-frequencies are relevant.

The first relevant frequency of the variables Y, GD, L, X, IC, LC, PY, PC, and RL is 0.0064. This corresponds to the whole observed time period. Comparing the time plots of those variables a quadratic trend might be assumed (cp. Figure 11). After the elimation of this possible trend the frequency 0.0064 is not relevant anymore. But now the frequency 0.0127 is relevant for Y, X, LC, PY, PC, and RL. The other relevant frequencies do not change for most of the above listed variables except for GD and X. For these two variables some new short-wave frequencies appear. The robustness of the relevant frequencies gives a strong hint that the carried out transformation really eliminated a quadratic trend.

Table 3 shows that variables Y, L, LC, PY, and PC have the most relevant frequencies in common. Two of these frequencies are important for BC as well. This is interesting because Diebold and Rudebusch (1996) characterize the business cycle as "comovement of important economic variables". But it should be noted that there is a quite natural explanation by the high correlations for these variables. This group of five variables includes GNP, which is widely used as an indicator for the state of economy, and three of the variables found by Röhl (1998) in the search for the best combinations of variables with respect to the classification of business cycles with discriminant analysis.

The variables GD and IC are not important for classification (Röhl 1998), although the three most common frequencies are relevant for them as well. Indeed, they differ from the other variables because they have a number of relevant short wave frequencies (GD has another two short wave frequencies and IC has 13 short wave frequencies).

In order to judge how well the variables are represented by their most important frequencies, we fitted the harmonics of the relevant frequencies $\lambda_i, i = 1, \ldots, p$, for each variable $f(t)$, with a linear model of the form

$$f(t) = a_0 + \sum_{i=1}^{p}(a_{is} \cdot \sin(2 \cdot \pi \cdot \lambda_i \cdot t) + a_{ic} \cdot \cos(2 \cdot \pi \cdot \lambda_i \cdot t)).$$

The coefficients a_0, a_{is}, and a_{ic} are the unknown parameters of the regression.

	$I_{stand}(\lambda) > 0.1$		$I_{stand}(\lambda) > 0.2$	
	R^2	R^2_{adj}	R^2	R^2_{adj}
Y	0.4239	0.3759	0.385	0.3517
C	0.5094	0.461	0.4881	0.4455
GD	0.5615	0.5114	0.4976	0.4558
L	0.4503	0.4206	0.425	0.402
X	0.4964	0.4763	0.4407	0.426
M1	0.4529	0.4073	0.4398	0.401
IE	0.5249	0.4992	0.476	0.455
IC	0.542	0.3733	0.4031	0.3153
LC	0.5253	0.4858	0.4605	0.4389
PY	0.5903	0.5739	0.5903	0.5739
PC	0.5912	0.5749	0.5581	0.5465
RS	0.3735	0.3484	0.3278	0.3102
RL	0.5184	0.4924	0.4211	0.4059
BC	0.5138	0.4583	0.4199	0.3885

Table 4: Fit of the harmonics

Table 4 shows the R-square and the adjusted R-square of the models with reported relevant frequencies. The relevant frequencies fit the variables PC and PY best with $R^2_{adj} = 0.57$ (cp. Figure 11). The adjusted R-square's of the variables M1, IE, and GD are nearly 0.5, the values for the other variables lie between 0.34 and 0.48. The adjusted R-square regarding to BC is 0.45 (Figure 12).

Figure 12 shows that the fit for the last 30 observations of the time period is not as good as of the part before. One reason could be that around 1986 a so-called interrupted upswing occured (Tichy (1994)).

5 Conclusion

The descriptive methods shed a new light on the problems which were encountered in the discriminant analysis of Weihs, Röhl, Theis (1999). The marginal distributions, as

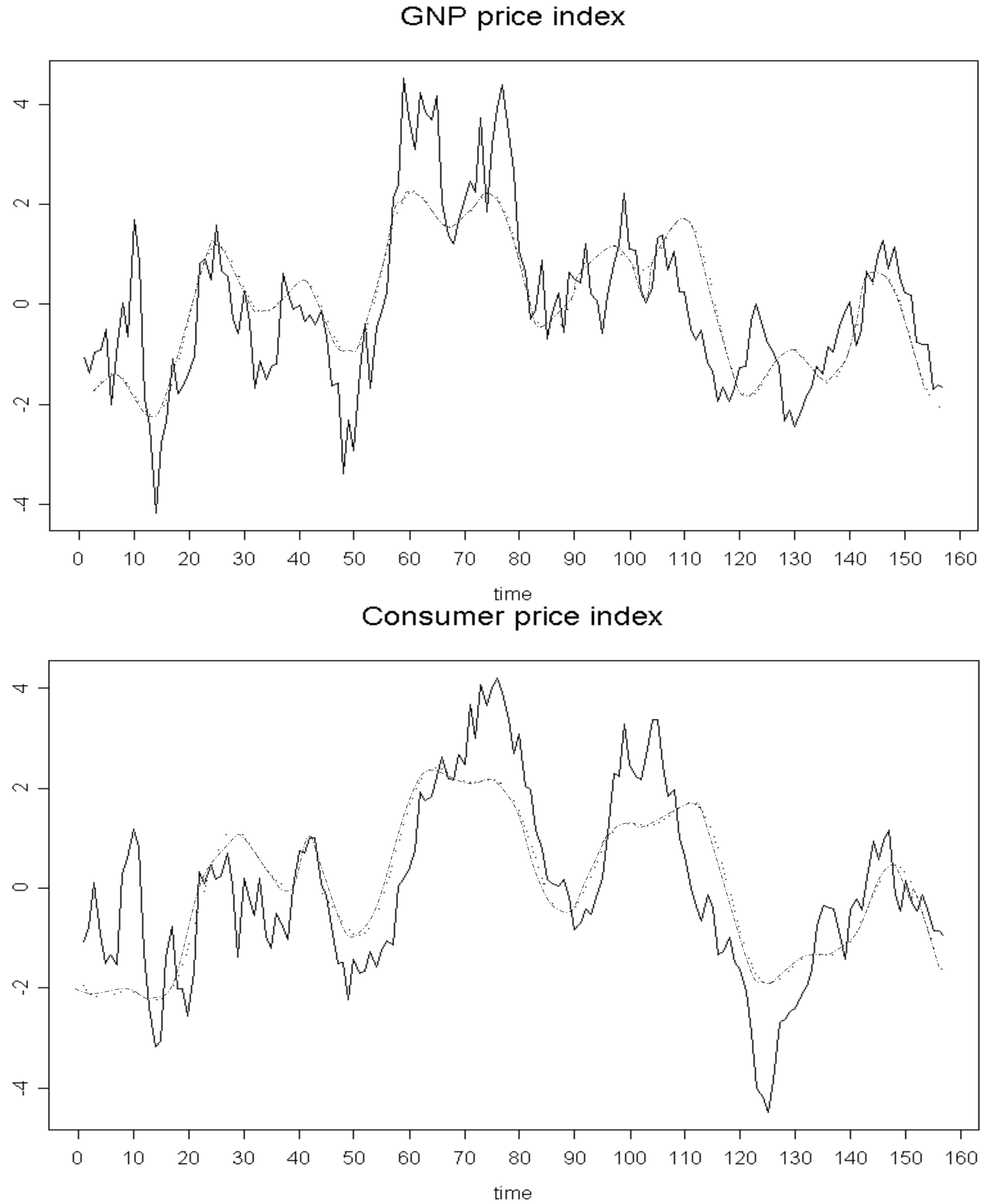

Figure 11: Price deflator of the GNP (PY) and Price index of private Consumption (PC) and fitted harmonics (dotted line)

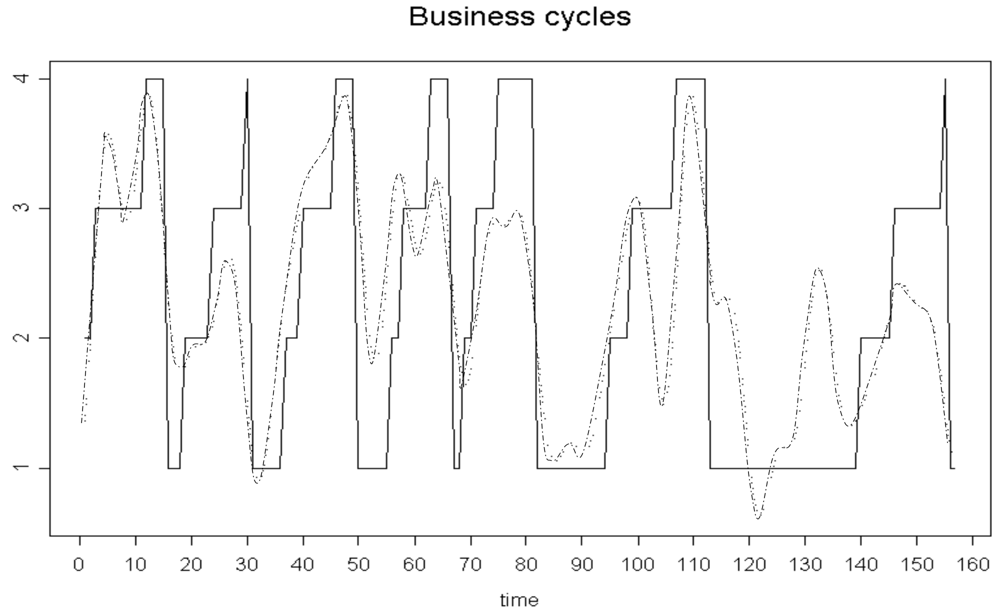

Figure 12: Business Cycle Classification as time series (BC) and fitted harmonics (dotted line)

seen in the histograms, make the normality assumption look not to far fetched. The scatterplots stress the point that a further study of the data with discriminant analysis should consider the highly linear dependencies in the variables by using reduced rank methods. The comparison of the variables in their four phases makes the problems with the classification into these phases obvious.

The time series study shows very nicely that the idea of "comovement of important economic variables" (Diebold, Rudebusch (1996)) applies very well to the used business cycle classification (BC) as can be seen from the common frequencies with the important variables for classification.

Acknowledgment

This work has been supported by the Collaborative Research Centre "Reduction of Complexity in Multivariate Data Structures" (SFB 475) of the German Research Foundation (DFG).

References

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

HEILEMANN, U. (1998) Erfahrungen mit dem RWI-Konjunkturmodell 1974 bis 1994. In: Ullrich Heilemann und Juergen Wolters (Hrsg.), Gesamtwirtschaftliche Modelle in der Bundesrepublik Deutschland: Erfahrungen und Perspektiven. Schriftenreihe des Rheinisch-Wesfälischen Instituts für Wirtschaftsforschung, Heft 61, Berlin 1998, S. 61-93

HEILEMANN, U. AND MÜNCH (1999) "West German Business Cycles 1963-1994: A Multivariate Discriminant Analysis", Paper presented at the 1995 CIRET-Conference in Singapore, CIRET-Studien 50, München

RÖHL, M.C. (1998) Computerintensive Dimensionsreduktion in der Klassifikation, Reihe: Qualitative Ökonomie, Bd. 95, Eul Verlag, Lohmar

SCHLITTGEN, STREITBERG (1995) Zeitreihenanalyse, *Oldenbourg Verlag München*

TICHY, G. (1976) Konjunkturschwankungen, *Heidelberger Taschenbücher, 174*, Berlin.

WEIHS, C., RÖHL, M.C., THEIS, W. (1999) Multivariate Classification of Business Phases, Technical Report **26**, SFB 475, Universität Dortmund