HAS THE ACCURACY OF GERMAN MACROECONOMIC FORECASTS IMPROVED?

by *Ullrich Heilemann*[1] and *H.O. Stekler*[2]

Abstract

The major focus of this paper is to determine whether the accuracy of German macroeconomic forecasts has improved over time. We examine 1-year-ahead forecasts of real GDP and inflation for 1967 to 2001 made by three major German forecasting groups and the OECD. We examine the accuracy of the forecasts over the entire period and in three sub-periods. We conclude that, with some exceptions, the errors of the German forecasters were similar to those of their US and UK counterparts. While the absolute size of the forecast errors has declined, this is not the case for relative accuracy. A benchmark comparison of these predictions with the ex post forecasts of a macroeconometric model indicates that the quality of the growth forecasts can be improved but that the expected increase in accuracy may not be substantial.

Keywords: Forecast evaluations, macroeconomic forecasting, accuracy limits

---

1 Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstraße 1-3, D-45128 Essen, FRG, Phone: +49/201/81 49-221, Fax: +49/201/81 49-284, email: heil@rwi-essen.de and University Duisburg-Essen, Location Duisburg, Lotharstraße 65, D-47048 Duisburg, FRG, Phone +49/203/3 7917 93.

2 H.O. Stekler, Department of Economics, George Washington University, Washington DC 20052 USA Phone + 202-994-6150; email: hstekler@gwu.edu.

1. Introduction

In a recent paper, Fildes and Stekler (2002) presented a survey of our current knowledge about the state of macroeconomic forecasting. While they mentioned some of the findings related to the forecasts of other countries, their survey primarily focused on the forecasts produced in the US and the UK. This paper presents an in depth examination of German macroeconomic forecasts to determine (1) whether the characteristics of these forecasts are similar to those of the US and UK and (2) whether the forecasts have improved over time. We also use an econometric model as a benchmark to determine the maximum increase in forecast accuracy that can be expected.

Quantitative forecasting in Germany began in earnest in the mid-1960s when the Joint Diagnosis[3] (JD) of the five (now six) large economic research institutes started to be published. This was followed by the forecasts of the newly established Council of Economic Experts (CEE) and the Annual Economic Report of the Federal Government (GAER). In the 1970s an increasing number of private forecasters, most of them from the banking sector, also started to issue macroeconomic forecasts. If the IMF, the OECD, the World Bank and the EU-Commission are included, there are now more than 30 institutions that regularly publish macroeconomic forecasts for Germany.

There have been a number of analyses of the accuracy of German macroeconomic forecasts (see e.g., Blix et al., 2001; Döpke, 2000; Öller & Barot, 2000; Pons, 2000; Kreinin, 2000). These studies report the usual statistics on absolute and relative accuracy or other forecast characteristics over a specific time span. Depending on the forecasters and the time period, the mean absolute errors (MAE) of the forecasts of the growth vary between 1.2 and 1.6 percentage points. The errors of the inflation forecasts vary between 0.6 and 0.8 percentage points. Many studies try to ascertain a ranking with respect to forecasters, methods, and variables. Most concluded that there is no forecaster

---

3 The names of these institutions in German are Joint Diagnosis, "*Gemeinschaftsdiagnose*", Council of Economic Experts, "*Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung*", Annual Economic Report of the Federal Government, "*Jahreswirtschaftsbericht der Bundesregierung*"

(or method) that is by all standards and for all variables always the best. This finding is similar to the results for the U.S. (e.g., Zarnowitz, 1992).

None of these studies has undertaken an explicit analysis of the way accuracy has changed over the past four decades. For shorter periods, deviations from what appears to be the standard are occasionally reported, but systematic studies over longer periods are missing. Implicit references can occasionally be found (e.g. Döpke and Langfeldt, 1995; Döpke, 2000; Heilemann, 1998).[4] Most studies that partition the sample period primarily examine the stability of the rankings of either the forecasters or the methods rather than analyze the time trend of forecast accuracy itself.

Although there has been no systematic analysis that has determined whether the accuracy of German forecasts has improved over time, this issue has been previously discussed in different contexts. In the 1950s and 1960s, with the development of large-scale econometric models, macroeconomists expected that the accuracy of their forecasts would improve over time. Since then things have changed. None of the contributions to the Centenary issue of the *Economic Journal* (1991) expected major improvements in the accuracy of forecasts. On the other hand, *Diebold* (1998) expressed a more optimistic view while *Hendry* (2001) doubted that this would occur. The major empirical studies, analyzing US forecasts, were undertaken by McNees (1986) and Zarnowitz (1992) , but they reached conflicting conclusions about the improvement in accuracy over time.

It is, therefore, appropriate to revisit the question of whether forecasts have improved over time, but this time with data that have not previously been used. This paper will examine four sets of German forecasts for the period, 1967-2001, primarily *focusing on whether the accuracy of the forecasts changed over time.* While this will be the primary focus, there will also be a discussion of forecast accuracy for the entire period and of the limits to the improvement in accuracy that can be expected. The next sections will

---

4 After the present study was finished, *Dicke and Glismann* (2002) analysed the forecast accuracy (over time) of one of the institutions studied here but they were rather brief on the subject.

discuss our sample of forecasters, the time periods that will be examined and the methods of analysis. We then present and explain the results. *We also use an econometric model as a benchmark in order to determine whether there are limits to the accuracy that can be expected from macroeconomic forecasts.*

2. Forecasters, samples, data, methods of analysis

2.1. Major macroeconomic forecasters

While a dozen major institutions produce macroeconomic forecasts for Germany, only four sets of forecasts are examined here. A number of criteria were used in selecting the organizations whose forecasts are analyzed. First, the organizations should play an important role in the public discussions of economic policies. The organizations should have produced a sufficient number of forecasts that would be available to determine whether accuracy has improved over time. Furthermore, the sample was selected to include forecasts from non-government as well as from government institutions and from one international organization. Finally, the forecasts had to be comparable as to the variables forecast, the forecast horizon, and the date of their publication. This led to the selection of the forecasts produced by (1) the Joint Diagnosis (JD)[5], (2) the Council of Economic Experts (CEE), (3) the Government Annual Economic Report (GAER), and (4) the OECD.

2.2. Data

Forecast accuracy and its evolution over time are analyzed here from the perspective of economic policy, or more specifically from fiscal policy. That is why we examine forecasts that are made infrequently and have a horizon of 6-18 months.[6] The study concentrates on two variables, the rates of change of real GDP and of the GDP deflator.

---

5 The composition of the JD has several times changed, currently members are: Deutsches Institut für Wirtschaftsforschung (DIW) Berlin, Ifo-Institut für Wirtschaftswirtschaftsforschung München, Institut für Weltwirtschaft (IfW) Kiel, Institut für Wirtschaftsforschung Halle (IWH), Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI) Essen.

"Growth" and "no inflation" are considered two of the most important macroeconomic goals. Given the strong dependencies of employment, the government deficit, etc., upon these two variables, they are also good indicators of the accuracy that might be expected if one evaluated the accuracy of the forecasts of these other variables.

In order to have a common base, the analysis begins in 1967, when the GAER published its first forecast. The sample ends with the year 2001. To examine the evolution of forecast accuracy, the sample is divided into three sub periods 1970-1979, 1980-1989 and 1990-2001.[7] While these sub periods are frequently used in analyses, their selection is still arbitrary.[8] Since each sub period is at least 10 years long, any cyclical bias should have been eliminated. Indeed, each decade experienced a recession. Other "events" affecting forecast accuracy such as the oil-shocks in the 1970s and 1980s, German unification, the Maastricht treaty and its fiscal consequences, and the Asia/Russia crisis 1997/8 are also included.

The forecasts are for the latter part of the current year and for the following year, but we only analyze the year-ahead predictions. The forecasts are published over a stretch of four months: {October (JD), November (CEE), December (OECD), and January (GAER)}, but the actual data on which they are based are not too different. The JD, CEE and also the OECD forecasts, given its three months of preparation, have to start from National Accounts (NA) data ending with the second quarter; the GAER, however, can start from data for the third quarter and can probably also use the *Federal Statistical Office*'s first estimate of GDP for the past year, which is issued in mid January of the following year. In the period studied here, there were only a few cases in which macroeconomic developments and events of essential importance happened between October and January. Although the GAER forecasts uses more information, notably

---

6 Monetary policy requires more frequent forecasts.

7 The inclusion of the 1967-69 period certainly would give a more optimistic impression of the evolution of forecast accuracy. At the same time it could be argued that the causes which led to these errors were so exceptional that there omission is well justified.

8 The splitting could have been based on a detailed break-point analysis but this seemed to be beyond the present question.

later data, and thus should be more accurate, it has been shown that this is hardly the case (Heilemann, 1998).

Many of the German forecasts have been presented with rates of change rounded to ½ percentage points. Consequently, in order for the forecasts and actual data to be comparable, all the forecasts and the actual data were rounded. (A preliminary analysis showed that in those cases where the original forecasts had not been rounded, the differences in the results were small.) In 1993 the German Federal Statistics Office changed its NA concepts and, as its measure of output, replaced GNP by GDP. Hence, until 1993 "growth" is associated with real GNP, thereafter with real GDP; the inflation indicator was changed correspondingly. The actual data were taken from the Federal Statistical Office's first release of NA data for the previous year. The data and sources are given in detail in Table 5 (Appendix).

2.4. Measures of forecast accuracy

Our measures of forecast accuracy include descriptive statistics, tests for directional accuracy and rationality tests.

2.4.1. Quantitative Measures

There are many statistics that may be used to measure forecast accuracy (Stekler, 1991; Diebold and Mariano, 1995; Döpke, 2000). Here, we focus on the bias, the mean absolute error (MAE), and the root-mean-square percentage error (RMSPE). As a benchmark, comparative accuracy is measured by Theil's U coefficient (based on extrapolating the previous rate of change $p_t = a_{t-1}$) and its decomposition is used to inform about the nature of forecast errors. Given that Germany has experienced a general decline in the rates of change of both growth and of inflation, the test is biased against an extrapolation of the previous year's rates of change. The forecast performance associated with the difficulty of the task is measured by the relationship of RMSE/σ (Ash, Smyth, and Heravi, 1993).

In determining whether forecast accuracy has changed over time, we adopt a method that is widely used in analyzing quality control and the stability of regression coefficients *but that has not been extensively applied in evaluating forecast accuracy*. The method consists of a CUSUM test.

2.4.2. Directional Accuracy

In analyzing directional accuracy we first describe the type of errors that were observed, namely the failure to predict turning points and the number of over and underestimates that occurred. Then we determine whether the accelerations and decelerations in the growth and inflation rates were correctly predicted. We use the concept of "Informational content" (IC) which compares the number of accelerations (decelerations) of changes that are forecast and realized (see e.g., Diebold & Lopez, 1996):

$$IC = \frac{AC}{AC + AW} + \frac{DC}{DC + DW}$$

with AC: increase forecast and realized; AW: increase forecast, decrease realized; DC: decrease forecast, and realized; and DW: decrease forecast, increased realized.

Following *Merton* (1981), we assume that for a forecast to have "informational content", IC has to be > 1. Under the null hypothesis that forecasts and realizations are independent and using past realizations, the probabilities for the four cases (cells) can be consistently estimated. They can be compared with the actual number and tested against a χ^2 distribution with one degree of freedom:

$$C = \sum_{i,j=1}^{2} \left(O_{ij} - \hat{E}_{ij}\right)^2 \Big/ \hat{E}_{ij} \qquad \rightarrow \chi_1^2$$

with: O_{ij}: observed cell counts and $\hat{E}_{ij}$: estimated cell counts.

2.4.3. Rationality

The rationality of forecasts, based on unbiasedness and efficiency, is tested in the "traditional way" (Kirchgässner, 1993). A sufficient condition that the forecasts are unbiased is that the joint null, $\alpha_1 = 0$ and $\beta_1 = 1$, in regression (1) cannot be rejected.

$$a_t = \alpha_1 + \beta_1 \cdot p_t + u_t \qquad (1)$$

The forecasts are efficient if $\beta_2 = 0$ in (2)

$$e_t = \alpha_2 + \beta_2 \cdot p_t + u_t, \qquad (2)$$

and $\rho = 0$ in (3).

$$e_t = \alpha_3 + \rho \cdot e_{t-1} + u_t. \qquad (3)$$

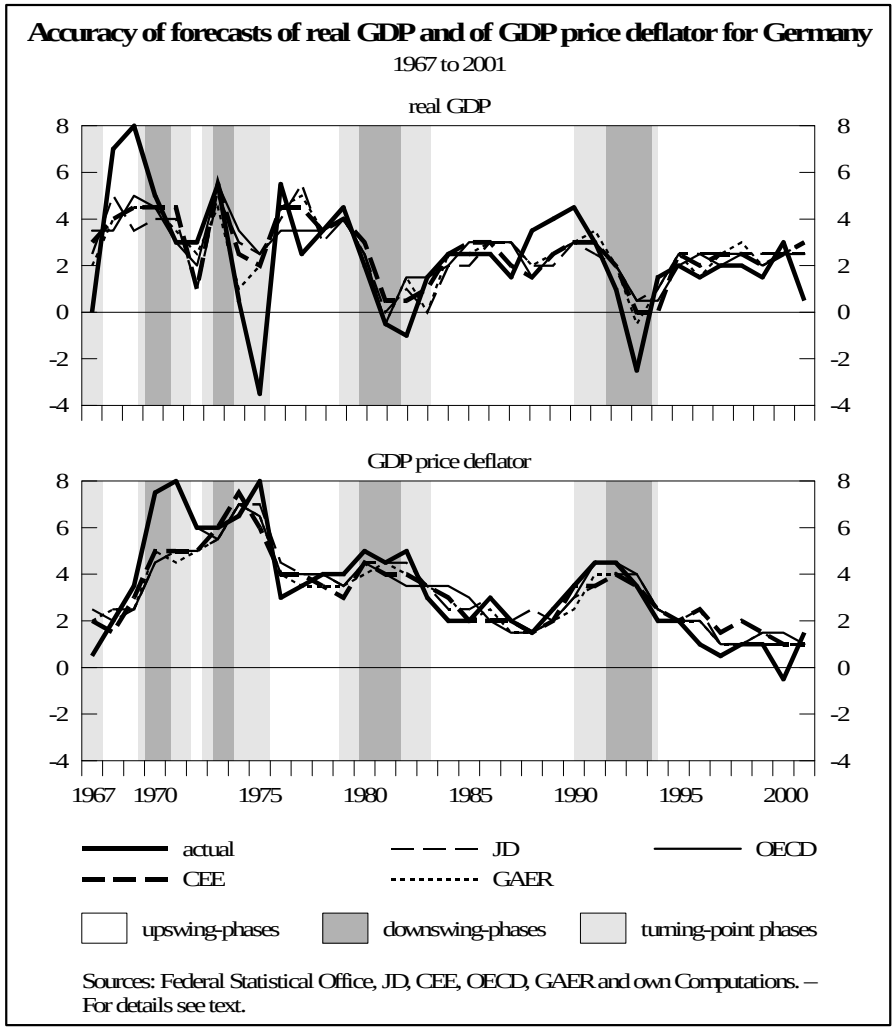
The test here is based – like Theil's inequality coefficient – on the assumption that the previous year's *actual* data are known.

3. Results: The complete sample – a summary

The main focus of our analysis is on the question of whether the German forecasts have improved over time. Nevertheless, we summarize the results for the entire period , 1967-2001. The forecasts and the actual data of growth and inflation are shown in Figure 1, and the results of the accuracy analysis are in Table 1. The MAE of the growth forecasts is about 1.2 percentage points. This was about 40% of the mean absolute change. Similarly, the MAE of the inflation forecasts was about 0.7 percentage points, but this was only 20% of the mean absolute change in the inflation rates.[9] The RMSPE is about

9 Although the forecast periods are not the same, it is possible to compare these results with those that Fildes and Stekler (2002, pp.443-44) reported for the US and UK. In the US the errors were about 25% of the mean absolute changes of both variables, while in the UK they averaged about 60%.

Figure 1



125 % and about 85 % for the growth and inflation forecasts, respectively.[10] The German inflation forecasts are more accurate than the growth predictions, contrary to the findings for the US and the UK.

A comparison of the forecasts with naïve forecasts using Theil's U coefficient indicates that all of the forecasts are very much superior to simple extrapolations of the

10 It should also be noted that between 1968/99 the MAE between the first and the final actual data had been 0.4 percentage points for growth and 0.3 percentage points for inflation.

Table 1

Annual forecasts of percentage changes of real GDP and of GDP price deflator for Germany: summary measures of error
1967 to 2001

	real GDP				GDP price deflator			
	JD	CEE	OECD	GAER	JD	CEE	OECD	GAER
				1967 to 2001				
MAE	1.5	1.3	1.3	1.2	0.7	0.8	0.7	0.7
RMSPE	128.6	126.6	138.7	108.3	92.4	87.6	78.1	78.7
Bias	0.2	0.3	0.3	0.2	-0.1	-0.1	0.0	-0.2
U	0.3	0.3	0.3	0.3	0.1	0.1	0.1	0.1
UM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
UV	0.3	0.3	0.4	0.4	0.3	0.3	0.2	0.4
UC	0.7	0.7	0.6	0.6	0.7	0.6	0.8	0.6
RMSE/σ	0.8	0.7	0.8	0.7	0.5	0.5	0.4	0.5
				1970 to 1979				
MAE	1.9	1.5	1.4	1.3	1.2	1.3	0.8	1.2
RMSPE	173.5	140.9	198.4	67.9	25.4	24.0	19.9	22.8
Bias	0.7	0.7	0.6	0.6	-0.7	-0.8	-0.4	-0.9
U	0.4	0.3	0.3	0.3	0.1	0.1	0.1	0.1
UM	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.3
UV	0.3	0.4	0.5	0.5	0.2	0.1	0.2	0.2
UC	0.6	0.5	0.4	0.5	0.6	0.6	0.7	0.5
RMSE/σ	0.9	0.8	0.9	0.8	0.8	0.9	0.7	0.9
				1980 to 1989				
MAE	1.1	0.9	1.0	1.0	0.4	0.5	0.7	0.5
RMSPE	87.3	85.3	88.9	93.1	25.6	22.2	33.1	22.1
Bias	-0.1	0.1	0.2	0.1	0.1	-0.2	-0.2	-0.2
U	0.3	0.2	0.3	0.3	0.0	0.0	0.1	0.0
UM	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1
UV	0.1	0.3	0.1	0.1	0.3	0.1	0.1	0.1
UC	0.9	0.7	0.8	0.9	0.7	0.8	0.9	0.8
RMSE/σ	0.8	0.7	0.8	0.8	0.4	0.5	0.7	0.5
				1990 to 2001				
MAE	1.0	1.0	1.0	0.9	0.6	0.7	0.5	0.5
RMSPE	127.9	154.1	128.2	150.8	97.4	120.9	124.0	84.6
Bias	0.5	0.4	0.3	0.4	0.2	0.3	0.3	0.1
U	0.3	0.3	0.3	0.2	0.1	0.1	0.1	0.1
UM	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.0
UV	0.6	0.3	0.5	0.3	0.3	0.4	0.1	0.4
UC	0.3	0.7	0.5	0.6	0.7	0.4	0.7	0.6
RMSE/σ	0.8	0.8	0.8	0.7	0.5	0.5	0.5	0.4

Author's computations. For sources, abbreviations and computation of the error measures see text.

Table 2

Correlations[1] of major institutions' forecasts for Germany
1967 to 2001

		JD	CEE	OECD	GAER
JD	1967 to 2001	-	0,951	0,948	0,971
	1970 to 1979	-	0,925	0,942	0,964
	1980 to 1989	-	0,896	0,794	0,925
	1990 to 2001	-	0,951	0,958	0,961
CEE	1967 to 2001	0,903	-	0,945	0,961
	1970 to 1979	0,919	-	0,925	0,963
	1980 to 1989	0,795	-	0,918	0,954
	1990 to 2001	0,943	-	0,915	0,921
OECD	1967 to 2001	0,831	0,895	-	0,971
	1970 to 1979	0,719	0,803	-	0,945
	1980 to 1989	0,828	0,813	-	0,897
	1990 to 2001	0,927	0,948	-	0,989
GAER	1967 to 2001	0,874	0,885	0,824	-
	1970 to 1979	0,736	0,802	0,546	-
	1980 to 1989	0,828	0,813	0,885	-
	1990 to 2001	0,873	0,925	0,869	-

Authors' computations. – 1) r between the real GDP forecasts (left of main diagonal) and forecasts of GDP price deflator (right of main diagonal).

previousactual rates of change.[11] Most of the errors are due to an incomplete capturing of the co-variance between forecasts and actual data (UC) which is considered as not disturbing.

The average errors of all four groups were similar for both variables, with perhaps the JD growth predictions being an exception. Although the forecasts were highly correlated (Table 2), we tested whether there was a statistically significant difference in the accuracy of the four groups. The forecasts for each year were, therefore, ranked on the

11 That is not all too surprising given the very long period with four major recessions. (It is hard to imagine that any mechanical use of any (naive) scheme will capture this).

basis of their accuracy and the average rankings test (also called analysis of variance by ranks) was used (Stekler, 1991). There was no significant difference among the four groups' predictions either of growth or of inflation.[12]

Based upon a classification[13] developed in Heilemann (2002), the forecasts are found to be more accurate during periods of recovery and growth than in periods of recession, with the failure to predict the recessions resulting in turning point errors. All of the institutions failed to predict at least some of the four recessions that occurred in this period. This result are similar to those observed in the US and UK forecasts.

The German forecasts also displayed some but not all of the systematic errors that had been observed in other predictions. Fildes and Stekler had noted that the US and UK forecasters underestimated GDP when it was growing and conversely when it was declining; similar errors were observed when inflation was accelerating and decelerating. On the other hand, the German forecasts contained an approximately equal number of underestimates and overestimates of the growth rate, but there was a tendency to underestimate the inflation rate when it was increasing and overestimating it when it was declining. The more refined analysis (IC) for the complete sample shows that the hypothesis of an independence of the accelerations and decelerations of the growth forecasts and actual values can be rejected at or close to the 5% level. (Table 3). In other words, the forecasters were able to determine whether the German economy would grow faster (slower) next year relative to this year. With the exception of the OECD forecasts, this was not the case for the inflation forecasts.

Finally, although the results are not presented here, the regression rationality test did not

---

12 The values of χ^2 were 4.43 and 1.57 for the growth and inflation forecasts, respectively. The critical 5% value of the statistic with three degrees of freedom is 7.82. The growth forecasts were based on all 35 observations, but we only used the last 30 observations for the inflation predictions because the OECD did not forecast inflation in either 1967 or 1971.

13 This classification, also shown in the figures, is based on a multivariate four-phase-scheme to classify business cycles consisting of upswing periods (Lower turning point phases and Upswings) and downswing periods (Upper turning point phase and Downswing).

Table 3

Accuracy of forecasts of directional change of real GDP growth and of GDP price deflator for Germany
1968 to 2001

	JD					CEE					OECD					GAER				
	IC (C)	AC	AW	DC	DW	IC (C)	AC	AW	DC	DW	IC (C)	AC	AW	DC	DW	IC (C)	AC	AW	DC	DW
										Real GDP										
1968 to 2001	1.35	8	4	15	7	1.71	12	2	17	3	1.41	9	4	15	6	1.43	12	7	12	3
	(3.826)					(16.703)					(5.384)					(6.333)				
1970 to 1979	1.40	3	2	4	1	1.86	3	0	6	1	1.86	3	0	6	1	1.40	3	2	4	1
											(6.429)					(1.667)				
1980 to 1989	0.90	1	2	4	3	1.58	3	1	5	1	1.17	2	2	4	2	1.17	2	2	4	2
1990 to 2001	1.78	3	0	7	2	1.66	4	1	6	1	1.31	3	2	5	2	1.63	5	3	4	0
										GDP price deflator										
1968 to 2001	1.30	8	4	14	8	1.33	7	3	15	9	1.52	6	1	16	8	1.30	8	4	14	8
	(2.862)					(2.993)					(6.004)					(2.862)				
1970 to 1979	1.25	3	1	3	3	1.10	2	1	3	4	1.50	2	0	3	3	0.83	2	2	2	4
1980 to 1989	1.38	2	1	5	2	1.75	2	0	6	2	1.75	2	0	6	2	1.58	3	1	5	1
1990 to 2001	1.25	2	2	6	2	1.25	2	2	6	2	1.44	2	1	7	2	1.44	2	1	7	2

Authors' computations, for computation see text and Appendix. – AC (AW): acceleration correctly (wrongly) forecast. DC (DW): deceleration correctly (wrongly) forecast. IC: information content, C : test on information content.

Table 4

Annual forecasts of percentage changes of real GDP and of GDP price deflator for Germany: summary measures of directional errors
1967 to 2001

	real GDP								GDP price deflator							
	JD		CEE		OECD		GAER		JD		CEE		OECD		GAER	
	1967 to 2001															
Number of																
Overestimates	4	(4)	3	(2)	1	(0)	3	(2)	3	(3)	5	(5)	4	(4)	4	(4)
Underestimates	17	(12)	17	(15)	13	(10)	16	(11)	16	(14)	14	(11)	14	(11)	17	(14)
Turning point errors	8	(8)	5	(5)	7	(7)	6	(6)	3	(3)	5	(5)	3	(3)	2	(2)
Coincidences	0	(5)	4	(7)	8	(12)	4	(10)	6	(8)	4	(7)	5	(8)	5	(8)
Other errors	5	(6)	5	(6)	5	(6)	5	(6)	6	(7)	6	(7)	6	(7)	6	(7)
	1970 to 1979															
Number of																
Overestimates	1	(1)	1	(0)	1	(0)	1	(0)	3	(3)	3	(3)	2	(2)	2	(2)
Underestimates	7	(5)	6	(4)	4	(3)	7	(3)	4	(3)	4	(2)	4	(2)	5	(3)
Turning point errors	2	(1)	1	(1)	2	(1)	1	(1)	1	(1)	2	(2)	0	(0)	1	(1)
Coincidences	0	(2)	2	(4)	3	(5)	1	(5)	1	(1)	0	(1)	2	(3)	1	(2)
Other errors	0	(1)	0	(1)	0	(1)	0	(1)	1	(2)	1	(2)	1	(2)	1	(2)
	1980 to 1989															
Number of																
Overestimates	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	1	(1)	2	(2)	1	(1)
Underestimates	6	(3)	6	(4)	4	(2)	4	(2)	6	(4)	5	(4)	4	(3)	5	(4)
Turning point errors	3	(3)	2	(2)	3	(3)	3	(3)	1	(1)	2	(1)	2	(2)	1	(1)
Coincidences	0	(2)	1	(2)	2	(3)	2	(3)	3	(4)	2	(3)	1	(2)	2	(3)
Other errors	1	(2)	1	(2)	1	(2)	1	(2)	0	(1)	0	(1)	1	(1)	1	(1)
	1990 to 2001															
Number of																
Overestimates	2	(2)	1	(1)	1	(0)	0	(1)	1	(0)	1	(1)	0	(0)	0	(0)
Underestimates	5	(4)	8	(8)	7	(7)	8	(6)	7	(6)	3	(2)	6	(5)	7	(6)
Turning point errors	4	(4)	1	(1)	2	(2)	2	(2)	1	(1)	3	(3)	1	(1)	0	(0)
Coincidences	1	(2)	2	(2)	2	(3)	2	(3)	0	(2)	2	(3)	2	(3)	2	(3)
Other errors	0	(0)	0	(0)	0	(0)	0	(0)	3	(3)	3	(3)	3	(3)	3	(3)

Authors' computations. For sources, abbreviations and computation of the measures of directional errors measures see text. In parentheses: coincidences: actual = ± 0.25 percent.

reject the null that the forecasts were unbiased. However, for the entire period, the hypothesis of the efficiency of both the growth and inflation forecasts is rejected (Table 4). The β-test indicates that the forecast errors are positively related to the forecasts and the ρ-test reveals that most forecast errors are autocorrelated. The exceptions are the inflation forecasts of the OECD and the GAER.

4. Results: Accuracy over time

We use four different approaches to determine whether forecast accuracy has improved over time. They involve (1) an examination of directional errors, (2) stability tests for forecast accuracy, (3) adjustments for the difficulty in forecasting in each time period, and (4) comparisons with benchmarks, including an econometric model.

4.1. Directional errors

The small number of observations in each sub-period precludes formal statistical tests, but descriptive results can be obtained from the information content statistics. If there had been an increase in accuracy over time, this statistic should be increasing monotonically from the 1970s to the 1990s. It can, however, be seen that the information content of the growth forecasts deteriorates in the 1980s but generally improves in the 1990s. A similar result can be observed in the inflation forecasts of the 1990s (Table 3). The biases are lower in the 1980s and 1990s than they were in the 1970s, but there is also no clear downward trend.[14] These results suggest that there is no tendency towards a monotonic improvement in accuracy.

4.2. Quantitative Errors

The time trend of the quantitative forecast errors for both variables also yields mixed results (Table 1). There were very large errors in the late 1960s. The MAEs in the 1970s ranged from 1.3 to 1.9 percentage points for growth and from 0.8 to 1.3 for inflation. These errors reflect the wage explosion in the early 1970s and the oil shock and its aftermath. The errors decline in the 1980s and 1990s to about 1.0 percentage point for growth and to 0.5 for inflation. While the MAEs show a decline from the 1970s through the 1990s, the RMSPEs rise between the 1980s and 1990s. These results require a further interpretation. We examine this issue by conducting a stability test and also by adjusting the errors for the difficulties involved in forecasting each period.

4.2.1. Stability Test

The stability tests of the forecast accuracy are analogous to the CUSUM tests of regression analysis[15]:

$$S_t = \sum_{k=1}^{t} \left(\frac{p_k - a_k}{\sigma_t} \right)^2 \Bigg/ \sum_{k=1}^{T} \left(\frac{p_k - a_k}{\sigma_T} \right)^2 \quad , \quad 2 \le t \le T$$

σ_t : standard deviation of actual values $x_1, \ldots, x_t$.

The CUSUM test here is based on a plot of the recursive errors. We restrict ourselves to the CUSUM-of-squares which plots the cumulative sum of squared residuals, expressed as a fraction of these squared residuals summed over all observations. If this sum goes outside a critical bound, this indicates that there was a structural break of the relationship of the average forecast accuracy (Brown et al., 1976).

The CUSUM of squares test is plotted in Figure 2. It shows that the forecasts of both variables made by the German forecasters display structural shifts from the early 1970s to the mid 1980s.[16] The performance of all the German forecasters is quite similar suggesting that there was forecasting improvement after the 1970s, but not subsequently.

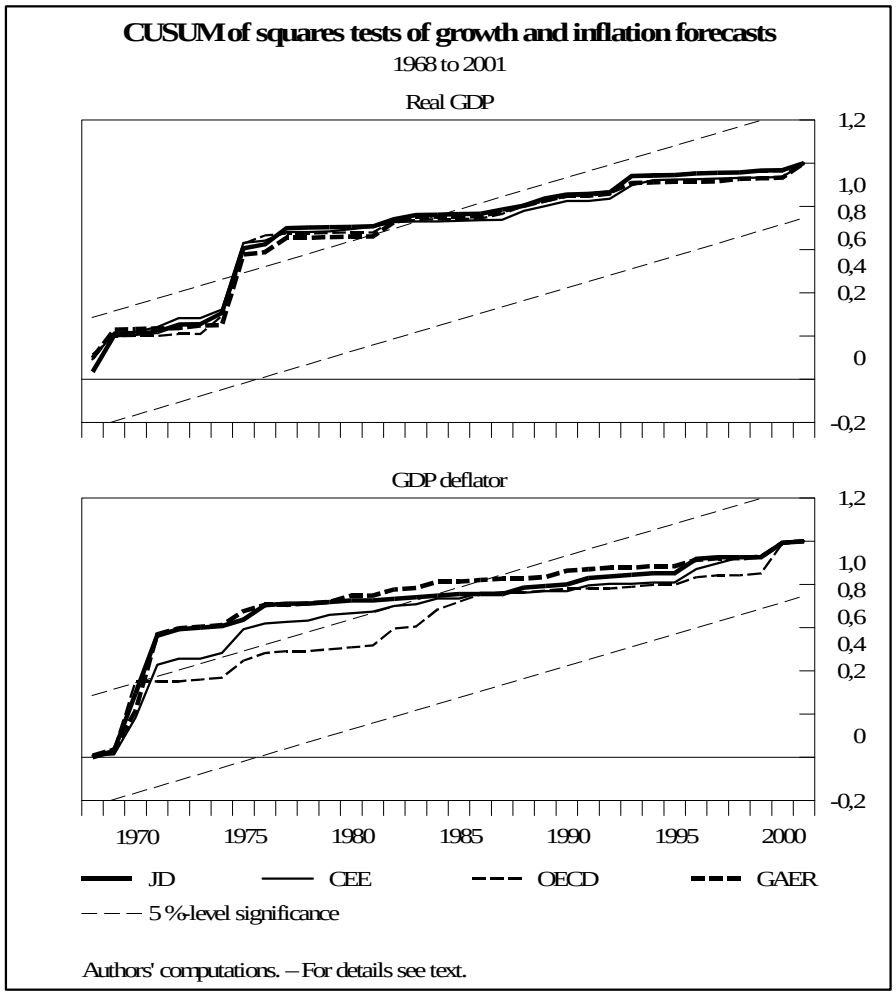
---

14 The bias of the growth forecasts is rather high in the 1970s and again in the 1990s, while the magnitude became negligible during the 1980s.

15 It should be remembered that CUSUM tests had long been used in quality control before they were transformed to be applied for stability analysis in regression analysis (see e.g. Brown et al., 1976).

16 It must be remembered that the results are very sensitive to the starting period and limited to forward recursive computations.

Figure 2



4.2.2. Adjusting for the difficulties of forecasting

The approach of the previous section, did not adjust for the difficulties involved in forecasting. One possible adjustment is to divide the RMSE by the standard deviation of the actual changes that occurred in each time period. The last entry in each panel of Table 1 presents this measure. This measure indicates that the forecast errors, adjusted for this variability, for both variables were similar in the 1980s and 1990s and slightly smaller than those of the 1970s. The stability tests using recursive RMSPEs (lower panels of Figures 2 and 3) yield similar results, with a slight increase in 2001 due to the recession. *All in all, there is some evidence of improvement in absolute forecasting accuracy, in particular if the oil and wage shocks in the 1970s are taken into account,*

but relative stability (based on the variance and rates of change) has been rather constant.

4.2.3. Explaining the results

The data reveal some of the factors that reduce accuracy and suggest areas where a forecaster should place his efforts in order to approach this limit. The effects of the errors made in predicting the recessions and downswings of 1974, 1980/81, and 2001 can be identified even in the recursive accuracy of growth forecasts.[17] While this finding suggests that greater efforts should be placed on predicting recessions in advance, it must be remembered that forecasters in other countries also have failed to predict the onset of recessions.

Similarly the impact that wage inflation and the oil-shocks had on inflation in the first half of the 1970s can be observed, but the statistics decline steadily towards a limit afterwards. The most plausible explanation is that exogenous inflation impulses and internal inflation behavior simply had normalized (see Figure 1) and forecasters have been able to forecast accurately in this environment.

However, a very important finding is that the recursive statistics show a declining trend that seems to be approaching a limit, i.e. a level beyond which accuracy cannot be improved, at least not with the current state of theory, forecasting methods, available data. While Fildes & Stekler (2002) did not discuss the limits of accuracy, their results are not in conflict with this view. We turn our attention to this issue in the next section.

4.3. Bench mark comparisons

In judging the quality of these forecasts, only the Theil U statistic has been used as a benchmark. This naïve model is rather simple because it mechanically extrapolates last period’s observed change. A more appropriate comparison would be with the

performance of macroeconometric models. While *ex ante* forecasts with these models show the usual inaccuracies of macroeconomic forecasts, their *ex post* performance is usually much better and may be used as a yardstick.

For this purpose we use the RWI-business cycle model, a medium sized (quarterly) macroeconometric model employed since the late 1970s for short term *ex ante* forecasting and simulations (see Heilemann, 2002, for details of the model). In our analysis this model was used to produce ex post *static* forecasts for each of the years 1980 to 1989. Each forecast was based on the actual values of all predetermined variables (exogenous variables and lagged endogenous variables). As an example, the data referring to the first half year of 1979 were used to forecast the second half of that year and all of 1980. This process was repeated to make forecasts for the other years. Hence the errors of these consecutive static simulations within the sample period are free from the errors that in ex ante forecasts are caused by (1) wrong assumptions about the predetermined variables, (2) the inability to capture the dynamics of multiperiod forecasts, and (3) the instability of the model outside the sample period. [18] The year ahead forecast was then based on consecutive simulated values for the current year's third and fourth quarters and for the complete next year. This procedure simulates the forecasting procedures that were actually but more importantly *it generates the highest forecast accuracy possible with a structural econometric model.*

The model's ex post growth MAE[19] was 0.6 percentage points, and the comparable RMSPE was 53.9 %. For inflation the respective errors were 0.4 percentage points and 17.9 %. The model's inflation errors for the period 1980-89 are very similar to those of the four forecasting groups. This suggests that the inflation forecasts for this period had

---

17 Surprisingly, the effects of German unification and the 1993 recession cannot be detected in this statistic.

18 The errors of static simulations can be further decomposed into stochastic equation errors and "model errors", that originate from the model's interaction within each solution period. It can be shown for the RWI-model (and probably for most models of this type), that for highly aggregated variables like GDP growth and the GDP deflator, the latter tends to be negligible. The main cause for this are the considerable aggregation gains, which, of course, do not show up on the single equation level.

achieved the highest accuracy level that was attainable. On the other hand, the model was substantially more accurate than the four organizations in predicting the rate of growth of the economy. It made no turning point errors, and the size of its errors was about 60% of those made by the four organizations. Since the model's errors represent the maximum accuracy attainable given the current state of macroeconomic forecasting, we provide the following interpretation. The quality of the growth forecasts can still be improved, but the expected increase in the accuracy of the ex ante predictions may not be that substantial.[20]

5. Summary, conclusions and recommendations

At the outset we posed a question: Has the accuracy of German macroeconomic forecasts improved over the last 40 years? The answer is that it depends, but certainly there is no clear cut trend towards improving accuracy. In terms of the *absolute* size of errors, the accuracy of both the growth and inflation forecasts have improved since the 1970s. The improvements, however, seem to be mainly due to the decline of the actual rates of change of growth and inflation and to the variability of these growth rates. The improvement is not so obvious if we are concerned with directional accuracy. The recessions in 1975, 1981/82 and 1993 were seen only after the fact, while the booms in the late 1960s and in the early 1990s were missed. These directional errors contributed substantially to the observed MAEs of the growth forecasts.

We believe that there is some room for improvement in these because the errors of these forecasts exceed the errors of the ex post forecasts obtained from the econometric model. The MAEs of the model's forecast were 0.6 percentage points for growth and of 0.4 for inflation. In general future forecast evaluations should determine the sources of forecast errors. Are they the result of faulty assumptions, misleading theories, empirical irregularities, insufficient data, etc.? Certainly, the errors cannot be blamed on the lack

---

19 Given that the model has been estimated by OLS, RMSE would have been a more adequate error measure but this would have caused problems of comparability with present results.

20 This is especially true since preliminary research indicates that some of the equations in the RWI model had larger errors in the 1986-2001 period than had been observed previously.

of resources because in the last few decades there has been substantial research activity on macroeconomic theory and forecasting methods in Germany and elsewhere that German forecasters could exploit. However, the quality of the German macro data may be a limiting factor in the ability to produce more accurate real time forecasts.[21]

We naturally recommend that theory, methods, and data be improved. While such efforts should be made, a more productive strategy in the short run may be to investigate *why* forecast accuracy differs over time, why forecasts for some countries are more accurate than for others (see e.g. Kreinin, 2002), whether some methods or forecasters are more "robust" than others, etc. In short: what determines forecast accuracy? Most forecast evaluations analyze "average" forecast accuracy, but we believe that it is equally necessary to undertake case studies to determine why the forecast errors occurred (see, for example, Fintzen & Stekler, 1999; Wallis (Ed.), 1987). We recommend as a first step that forecasters present an analysis of the accuracy of their last prediction at the same time that they are presenting their new forecast. Such an analysis should include a discussion of the role that assumptions, policy actions, random shocks, behavioral changes and interdependencies (offsetting errors) played in causing the observed errors. On the other hand, it may be that the one-percent-MAE for six-quarters-ahead GDP forecasts is a natural constant as this and other studies seem to suggest. If that is the limit to forecast accuracy, we will have to learn to accept it.

Acknowledgements

The authors gratefully acknowledge the Deutsche Forschungsgemeinschaft (SFB 475: Komplexitätsreduktion in multivariaten Datenstrukturen) for financial support.

---

21 The difference between the first and the final release of German real GDP data amounts to 1 percentage point (Heilemann, 2002).

References

Arbeitsgemeinschaft deutscher wirtschaftswissenschaftlicher Forschungsinstitute e.V., Essen a.o. (1967ff). Die Lage der Weltwirtschaft und der deutschen Wirtschaft im Herbst 1966 ff. Arbeitsgemeinschaft, Essen a.o.

Ash, J.C.K., Smyth, D.J. & Heravi, S.M. (1993). The accuracy of OECD forecasts for Canada and the United States. North American Journal of Economics & Finance 4, 179-210.

Blix, M., Wadefjord, K., Wienecke, U. and Adahl, M. (2001). How good is the forecasting performance of major institutions? Sveriges Riksbank Economic Review, no. 3, 38-68.

Brown, R.L., Durbin, J. & Evans, J.M (1975). Techniques for testing the constancy of of regression relationships over time, with comments. Journal of the Royal Statistical Society B37, 149-192.

Bundesregierung (1967 ff.). Jahreswirtschaftsbericht der Bundesregierung. Hegner, Bonn, after 1994: Bundesanzeiger-Verlag, Köln.

Dicke, H. & Glisman, H. G. (2002). Konjunkturprognosen und wissenschaftlich technischer Fortschritt. Wirtschaftsdienst 83, 167-169.

Diebold, F. X. (1998). The past, present, and future of macroeconomic forecasting. Journal of Economic Perspectives 12, 175-192.

Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253-263.

Diebold, F. X. & Lopez, J. A. (1996). Forecast evaluation and combination. NBER Technical Working Paper 192. NBER, Cambridge, MA.

Döpke, J. (2000). Haben Konjunkturprognosen in Deutschland einen politischen Bias? Schmollers Jahrbuch 120, 587-620.

Döpke, J. & Langfeldt, E. (1995). Zur Qualität von Konjunkturprognosen für Westdeutschland 1976-1994. (Kieler Diskussionsbeiträge, 247.) IfW, Kiel.

Economic Journal (1991). Centenary issue 101.

Fildes, R. & Stekler, H.O. (2002). The state of macroeconomic forecasting. Journal of macroeconomic forecasting 24, 435-468.

Fintzen, D. & Stekler, H. O. (1999). Why did forecasters fail to predict the 1990 recession? International Journal of Forecasting 15, 309-323.

Heilemann, U. (1998). Paradigm lost – Zu den Projektionen des Jahreswirtschaftsberichts der Bundesregierung. In: Heilemann, U., Kath, D. & Kloten, H. (eds.), Entgrenzung als Erkenntnis- und Gestaltungsaufgabe – Festschrift für Reimut Jochimsen zum 65. Geburtstag, Duncker & Humblot, Berlin, 79-100.

Heilemann, U. (2002). Increasing the transparency of macroeconomic forecasts: a report from the trenches. International Journal of Forecasting 18, 85-105.

Hendry, D. J. (2001). How economists forecast. In: D.J. Hendry & N.R. Ericsson (eds.), Understanding economic forecasts. MIT press, Cambridge, MA, 15-41.

Kirchgässner, G. (1993). Testing weak rationality of forecasts with different time horizons. Journal of Forecasting 12, 541-558.

Kreinin, M. (2000). Accuracy of OECD and IMF Projections. Journal of Policy Modeling 22, 161-79.

McNees, S. K. (1986). Forecasting accuracy of alternative techniques: a comparison of US macroeconomic forecasts. Journal of Business and Economic Statistics 4, 16-23.

Merton, R. C. (1981). On market timing and investment performance I: an equilibrium theory of value for market forecasts. Journal of Business 54, 363-406.

Öller, L.-E., & Barot, B. (2000), The accuracy of European growth and inflation forecasts. International Journal of Forecasting 16, 293-315.

OECD (1967 ff.). Economic outlook, 1ff. OECD, Paris.

Pons, J. (2000). The accuracy of IMF and OECD forecasts for G 7 countries. Journal of Forecasting 19, 53-63.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (1967ff.). Jahresgutachten 1967/68 ff. Kohlhammer, Stuttgart – Mainz, after 1989: Metzler-Poeschel, Stuttgart.

Stekler, H.O. (1991). Macroeconomic forecast evaluation techniques. International Journal of Forecasting 7, 375-384.

Theil, H. (1966). Applied economic forecasting. (Studies in mathematical and managerial economics, 4.). North Holland, New York, NY.

Wallis, K.F. et al. (ed.) (1987). Models of the UK economy. A second review by the ESRC Macroeconomic Modelling Bureau. Oxford, University Press.

Zarnowitz, V. (1992). Has macroeconomics failed? Cato Journal 12, 129-160.

Appendix

Table 5

Forecasts and actual data
1967 to 2001

	real GDP					GDP-Deflator				
	JD	CEE	OECD	GAER	actual	JD	CEE	OECD[1]	GAER	actual
1967	2.5	2.5	3.5	2.0	0.0	2.5	2.0	-	2.0	0.5
1968	5.0	4.0	3.5	4.0	7.5	2.0	1.5	2.5	2.0	1.5
1969	3.5	4.5	5.0	4.5	8.0	2.5	3.0	2.5	2.5	3.5
1970	4.0	4.5	4.5	4.5	5.5	4.5	5.0	4.5	5.0	7.5
1971	4.0	4.0	3.0	3.5	3.0	5.0	5.0	-	4.5	7.5
1972	1.0	1.0	2.0	2.5	3.0	5.0	5.0	5.0	5.0	6.0
1973	5.0	5.5	5.5	4.5	5.5	5.5	6.0	5.5	5.5	6.0
1974	3.0	2.5	3.5	1.0	0.5	7.0	7.5	7.0	7.0	7.0
1975	2.5	2.0	2.5	2.0	-3.5	7.0	6.0	6.5	6.5	8.0
1976	4.0	4.5	3.5	4.5	5.5	4.5	4.0	4.0	4.0	3.0
1977	5.5	4.5	3.5	5.0	3.0	4.0	4.0	4.0	3.5	3.5
1978	3.0	3.5	3.5	3.5	3.0	4.0	3.5	4.0	3.5	4.0
1979	4.0	4.0	4.0	4.0	4.5	3.5	3.0	3.5	3.5	4.0
1980	2.5	3.0	2.5	2.5	2.0	4.5	4.5	4.5	4.0	5.0
1981	0.0	0.5	-0.5	-0.5	0.0	4.5	4.0	4.0	4.5	4.0
1982	1.0	0.5	1.5	1.5	-1.0	4.5	4.0	3.5	4.0	5.0
1983	0.0	1.0	-0.5	0.0	1.0	3.5	3.5	3.5	3.5	3.0
1984	2.0	2.5	2.0	2.5	2.5	2.5	3.0	3.0	3.0	2.0
1985	2.0	3.0	3.0	2.5	2.5	2.5	2.0	2.5	2.0	2.0
1986	3.0	3.0	3.5	3.0	2.5	3.0	2.0	2.0	2.5	3.0
1987	3.0	2.0	3.0	2.5	2.0	2.0	2.0	1.5	1.5	2.0
1988	2.0	1.5	1.5	2.0	3.5	2.0	1.5	2.0	1.5	1.5
1989	2.0	2.5	2.5	2.5	5.0	2.0	2.0	2.0	2.0	2.5
1990	3.0	3.0	3.0	3.0	4.0	3.0	3.5	3.0	2.5	3.5
1991	3.0	3.0	3.0	3.0	4.0	3.5	3.5	4.5	4.0	4.0
1992	2.0	2.0	2.0	1.5	1.5	4.0	4.0	4.5	4.0	4.5
1993	0.5	0.0	1.0	-0.5	-2.0	4.0	3.5	4.5	3.5	3.0
1994	1.0	0.0	1.0	1.0	3.0	2.5	2.5	3.0	2.5	2.0
1995	2.5	3.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0	2.0
1996	2.5	2.0	2.5	1.5	1.5	2.5	2.5	2.0	2.0	1.0
1997	2.5	2.5	2.0	2.5	2.0	1.0	1.5	1.0	1.0	0.5
1998	3.0	3.0	3.0	3.0	2.0	1.0	2.0	1.0	1.0	1.0
1999	2.5	2.0	2.0	2.0	1.5	1.0	1.5	1.5	1.5	1.0
2000	2.5	2.5	2.5	2.5	3.0	1.0	1.0	1.5	1.0	-0.5
2001	2.5	3.0	2.5	3.0	0.5	1.0	1.0	1.0	1.0	1.5

Sources: Arbeitsgemeinschaft 1966ff., Sachverständigenrat 1966/67ff., OECD 1966ff., Bundesregierung 1967ff., rounded.