How accurate are GDP forecasts? An empirical study for Switzerland

by Eveline Ruoss, Economic Analysis, and Marcel Savioz, Research,
Swiss National Bank, Zurich

The authors would like to thank Reto Vögeli for his assistance in collating the data, and the forecasting institutes for helping to complete the data.
They are grateful for comments by Marlene Amstad, Franziska Bignasca, Aline Chabloz, Enzo Rossi and Mathias Zurlinden.

Despite their significance, economic forecasts do not have a very good reputation. The media frequently run reports on forecasts that have proved completely wrong. However, like weather forecasts, economic forecasts should not be condemned altogether simply because some have proved unreliable. It is surely preferable to form an opinion by looking at their accuracy over a relatively long period.

The first study of this type for Switzerland was carried about by Wasserfallen (1992). He showed that the forecasts for real gross domestic product (GDP) produced by nine selected forecasting institutes tended to slightly underestimate the real growth rates achieved between 1974 and 1991. He also showed that although the direction of changes in GDP was predicted correctly, the forecasts tended to overestimate the scale of such changes. The results indicated that there was very little difference in the accuracy of the forecasts made by the nine forecasting institutes. However, in view of the small sample, no significance test was carried out.

It is now ten years since this paper was published. Since then, there has been an increase in both the number of institutes producing regular forecasts and the frequency with which forecasts are issued. It therefore seems to make sense to examine this increased body of data on Swiss economic forecasting. The aim of this paper is to show what can be expected of economic forecasting. However, we do not intend to instigate a contest between the forecasting institutes. For this reason, the results are published without direct reference to the institutes.

In this paper we examine forecasts for annual growth in real GDP in Switzerland. We look at the forecasts made by 14 different organisations and institutes between 1981 and 2000. Unlike Wasserfallen (1992) and many surveys carried out for other countries (e.g. Öller and Barot, 2000), our study is not confined to the forecasts generated in the autumn for the following calendar year. As far as possible, we include all forecasts made by the institutes during the year for the current, following and next-but-one calendar year.

The paper is structured as follows. Section 1 outlines the data. In Section 2 charts and descriptive statistics are used to discuss the size of the forecast errors and whether the forecasts are unbiased and efficient. The third and fourth sections investigate whether the impressions gained from the charts and descriptive statistics are corroborated by econometric tests and whether the forecasts of the institutes are better than forecasts based on simple naive methods. In Section 5 we consider whether forecast errors would be lower if the estimates published by the Swiss Federal Statistical Office (SFSO) were taken as the outcome instead of the GDP growth data published by the State Secretariat for Economic Affairs (seco), which are released earlier. Section 6 contains a summary of our findings and conclusions.[1]

1 cf. Hendry and Ericsson (2001) for an introduction to interpreting and evaluating forecasts.

1 Data

1.1 Forecasts

This paper examines the forecasts for the percentage annual change in real GDP made by 14 different institutes, banks and organisations (referred to throughout as institutes). Alongside the Federal Commission for Economic Policy (KfK) and the Swiss National Bank (SNB), they include a number of commercial banks (CS, SBC, the old and the new UBS, and ZKB), international organisations (IMF, OECD) and various research institutes with organisational or personnel links to universities (BAK, CREA, KOF, MAT, SGZZ). Our sample thus includes all institutes which to our knowledge have published regular economic forecasts for Switzerland in recent years.

Table 1 gives the names and abbreviations of the 14 institutes in our sample. It also shows the start of the data series at each institute. The longest data series (KfK) goes back to 1971. Our evaluation is confined to the period 1981–2000 for two reasons: Firstly, the data set for the period before 1981 is relatively small because only a few institutes generated forecasts and some of these were for GNP rather than GDP. Secondly, the Swiss Confederation started to publish quarterly figures for real GDP in 1981. This greatly improved the information on real GDP during the year and probably influenced the quality of forecasting. Forecasts made since 1981 should therefore exhibit different statistical characteristics from those published before 1981.

The forecasts in our data set can be classified according to the month in which they were published and the period to which they refer. With reference to the forecasting horizon, we make a distinction between forecasts for (i) the current calendar year, (ii) the following calendar year and (iii) the next-but-one calendar year. The date when the forecasts were made is classified on the basis of the months (i) November, December and January, (ii) August, September and October, (iii) May, June and July and (iv) February, March and April. Aggregation over three-month periods mitigates the problem that with some institutes it is not always possible to determine exactly when a particular forecast was made.

Our classification gives 12 different forecasting horizons. h represents the number of months from the middle month in the three-month period in which the forecast was generated to the end of the calendar year to which it refers. The horizons $h = 0$ to $h = 9$ thus refer to forecasts for the current calendar year. $h = 0$ comprises forecasts made in November, December and (in some cases) January for the calendar year that is about to end/has just ended. $h = 12$ to $h = 21$ refer to forecasts for the next calendar year while

Forecasting institutes Table 1

Abbreviation	Name	Year when forecasting of real GDP growth commenced
BAK	BAK Basel Economics Ltd.	1983
CREA	Institute of Applied Macroeconomics, School of Higher Business Studies at the University of Lausanne	1977
CS[1]	Credit Suisse	1994
IMF	International Monetary Fund, Washington	1995
KfK	Commission for Economic Policy, Berne	1971[2]
KOF	Centre for the Research of Economic Activity at the Federal Institute of Technology in Zurich	1976
MAT	Aurelio Mattei, University of Lausanne	1977
OECD	Organisation for Economic Cooperation and Development, Paris	1981
UBS[3]	Union Bank of Switzerland, Zurich	1982
SBC[3]	Swiss Bank Corporation, Basel	1981
SGZZ	Center for Futures Research, St. Gallen	1976
SNB	Swiss National Bank	1977
UBS	UBS[4]	1998
ZKB	Zürcher Kantonalbank	1994

1 Vor 1997: Schweizerische Kreditanstalt (SKA).

2 Forecasts prior to 1980: Working Group for Economic Forecasts. From 1981: Economic Forecasting Subcommittee of the KfK.

3 Merged with SBC in December 1997 to form the new UBS.

4 Former Union Bank of Switzerland

$h = 24$ to $h = 33$ refer to forecasts for the next-but-one calendar year. $h = 33$ thus denotes the two-year forecasts produced in February, March and April, for instance forecasts made in March 1998 for the year 2000.

Table 2 shows the number of observations for each forecasting horizon and institute. The sample comprises a total of 766 observations. As the table shows, only a few institutes, i.e. KOF, KfK, the OECD and SBC (UBS "new" from 1998) generated 20 forecasts with the same forecasting horizon in the twenty-year period 1981–2000. This was principally because some of the institutes altered the timing of forecasts published during this time or did not make forecasts for some periods.

Almost all of the 766 forecasts were published. The sources are shown in the appendix. The exceptions are some of the forecasts by the SNB, which are taken from internal documents, specifically the monetary policy reports for the coming year elaborated each autumn between 1974 and 1999. Unlike the corresponding press releases, these documents in most cases contain precise forecasting data. Where a figure is qualified as "nearly" we deduct 0.25 percentage points. Similarly, we add 0.25 percentage points to those figures qualified with "at least". In four cases, only qualitative data were available on future economic trends, so that for those years a forecast figure for the SNB is missing.

Despite the gaps in our data set, our chart analysis and descriptive statistics provide an initial picture of the accuracy of the forecasts. By contrast, the gaps detract from the econometric tests outlined in section 3. Consequently, these are confined to institutes for which we have complete data series, thus enabling us to use standard econometric tests to assess the accuracy of the forecasts.

Overview of forecasts by time horizon and institutes 1981–2000 Table 2

Date of forecast	Time horizon	BAK	CREA	CS	IMF	KfK[1]	KOF	MAT	OECD	UBS[2]	SBC[2]	SGZZ	SNB	UBS	ZKB	Total of observations
Forecast for the current calendar year																
November to January	h=0	6	5			20			20	16	17			3	2	**89**
August to October	h=3	17	13	7	6		20					19		3	7	**92**
May to July	h=6	4	1			1			20		9	1		2	6	**44**
February to April	h=9	13	13	6	6	8	19			10	17			3	6	**101**
Forecast for the following calendar year																
November to January	h=12	5	7			20		20	19	16	16		14	3	6	**126**
August to October	h=15	16	12	6	5		20					16		2	6	**83**
May to July	h=18	3	1						19		5			2	5	**35**
February to April	h=21	13	12	4	5	8	18			10	5			2	4	**81**
Forecast for the next-but-one calendar year																
November to January	h=24	4	8						12	11	5			2	4	**46**
August to October	h=27	15	11				14							1	1	**42**
May to July	h=30	2	1								1					**4**
February to April	h=33	13	10													**23**
Total		**111**	**94**	**23**	**22**	**57**	**91**	**20**	**90**	**63**	**75**	**36**	**14**	**23**	**47**	**766**

1 From 1993 including the spring forecasts by the Swiss Confederation's Expert Group for Economic Forecasting

2 UBS and SBC merged in December 1997 to form the new UBS.

1.2 Outcomes

To assess the quality of the forecasts, they are compared with the actual GDP figures. Since GDP is normally revised several times, it is necessary to decide which figure to take as the outcome. Following the literature, we use the first available estimate for real GDP growth. In our case, this is the annual average calculated by the State Secretariat for Economic Affairs (seco) in March of each year on the basis of its quarterly estimates.[2]

The period 1981–2000 contains two business cycles. It thus includes both recessions and upswings, making it suitable for judging the accuracy of economic forecasts. On the basis of the seco data, annual average growth in real GDP was 1.4%. In six years, real GDP declined. The sharpest decline was 1.3% in 1982. The highest growth rate was 3.4% in 2000. The median is 1.95%, which is well above the average growth rate. This shows that the distribution of the rates of change in GDP in the period 1981–2000 is not symmetrical; downtrends (negative growth rates) have a greater impact than high growth rates. The standard deviation (SD), i.e. the square root of the squared deviations from the mean outcome, is 1.548. We use this figure as a benchmark.

2 How accurate are GDP forecasts?

2.1 Chart analysis

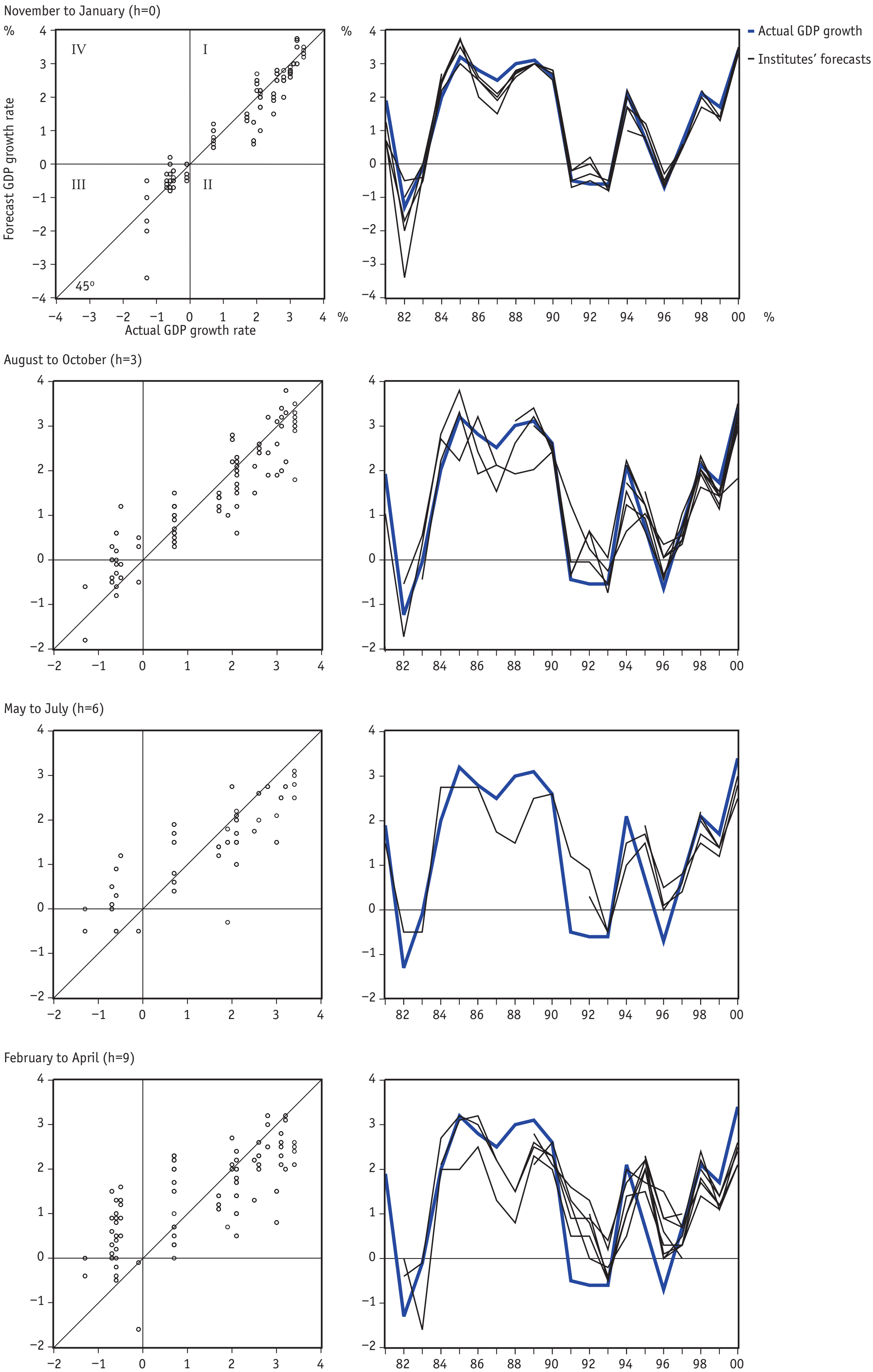
Figures 1 and 2 provide an initial insight into the quality of the forecasts. Fig. 1 shows the results for horizons $h = 0$ to $h = 9$, while the results for horizons $h = 12$ to $h = 21$ are shown in Fig. 2. The left-hand column in each chart contains four scatter diagrams while the right-hand column contains four time series plots. The curves show the GDP growth rates forecasts by the various institutes against the actual values (blue line). The deviations between the actual GDP growth rates and the forecasts indicate the development of forecast errors over time.

In the scatter diagrams, the x-axis represents actual GDP growth and the y-axis forecasts of GDP growth. Forecast errors, which are defined as the difference between the actual and forecast growth rate, correspond to the horizontal distance from a point to the 45° line. The closer the plots are to the 45° line, the more accurate the forecasts. The four squares indicated with Roman numerals show whether the direction of the GDP trend was forecast accurately. If the points are in squares I or III, a rise or decline in real GDP was anticipated correctly. By contrast, if the points are in squares II or IV, the institutes predicted the wrong direction, i.e. a rise rather than decline in GDP (square IV) or a fall rather than a rise (square II).

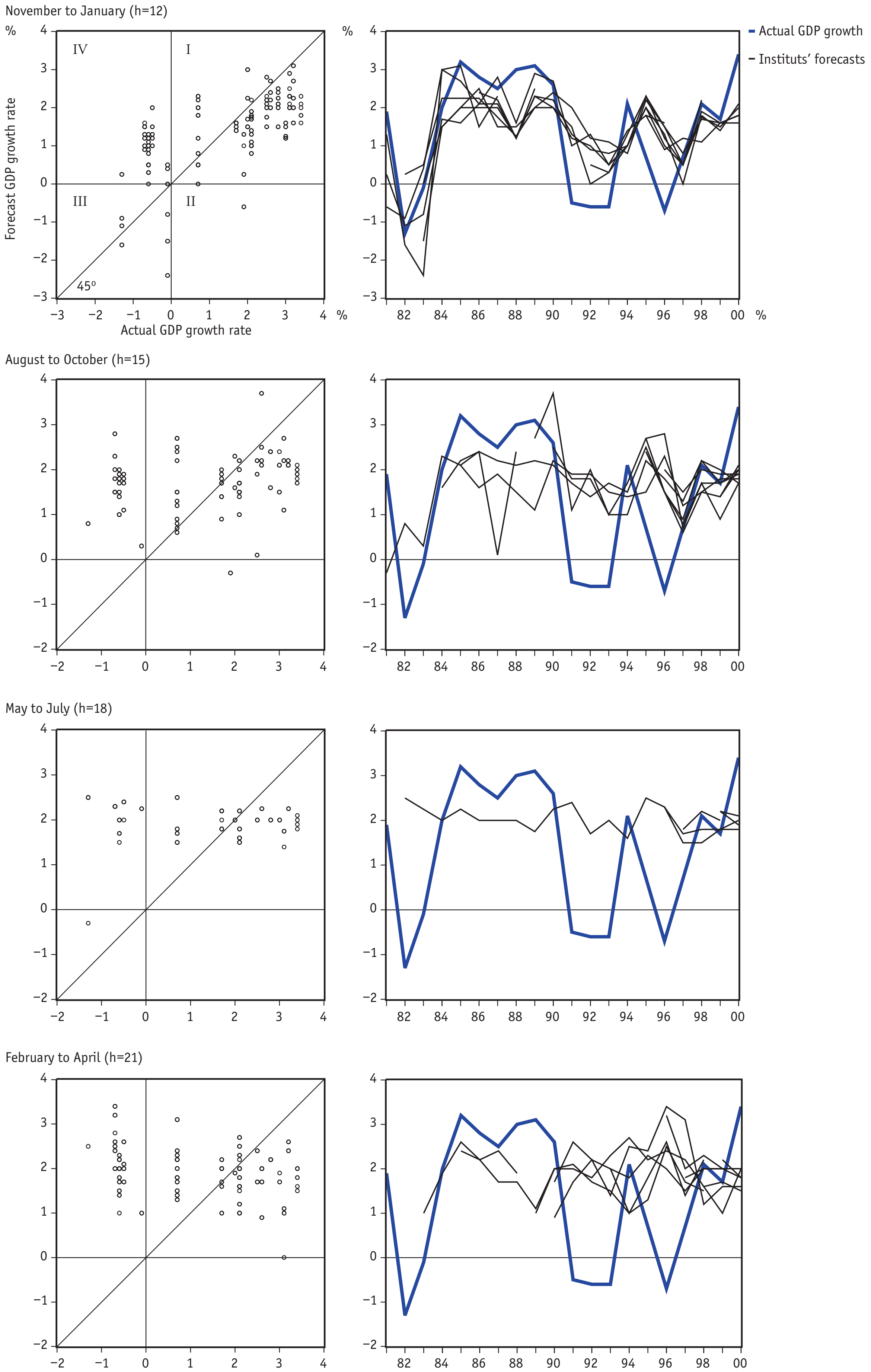
How accurate are the forecasts? Fig. 1 shows that the accuracy of the 89 $h0$-forecasts included in our evaluation is good. The projections generated in November to January for the current/last calendar year are close to the 45° line. The wrong direction was only forecast in one case, with a rise rather than a decline in GDP being anticipated. The forecasts for $h = 3$, i.e. forecasts made between August and October, are also close to the 45° line. By contrast, the forecasts for $h = 9$ (February to April) are a good deal less reliable. They scatter widely and in about a fifth of cases a rise in real GDP was indicated, when the actual trend was a decline.

2 From 1981–1986 the quarterly GDP estimates were carried out be the Swiss Federal Statistical Office (SFSO), which is part of the Federal Department of Home Affairs. In 1987 this task was entrusted to the Federal Office for Economic Questions within the Federal Department of Economic Affairs. Since 1999 the quarterly estimates for GDP have been calculated by the State Secretariat for Economic Affairs (seco).

Actual and forecast GDP growth rate for the current calendar year — Figure 1



Actual and forecast GDP growth rate for the following calendar year Figure 2



Two other observations can be made on the basis of the scatter diagrams in Fig. 1. Firstly, there is no indication that the forecasts are systematically distorted to the left or right of the 45° line. In other words, the forecasts are neither too high nor too low on average. They are thus unbiased. Secondly, the charts indicate that as from $h = 9$ the slope of the scatter plot is no longer around the 45° line but tends to be flatter. In other words, GDP forecasts tend to underestimate the trend in upward phases and overestimate it in periods of negative growth. From this we conclude that the higher the absolute forecast change, the higher the forecast error, i.e. the horizontal distance from the 45° line. They thus seem to correlate with the absolute forecast figures. This suggests that some information available to the institutes (including the correlation between the forecast and the forecast error) is not utilised in the forecasting process, resulting in informational inefficiency. Informational inefficiency is suspected for horizons $h = 9$ as well as for $h = 12$ and $h = 15$ in Fig. 2 because the slope of the scatter plot seems to be flatter than 45°.

The $h12$ forecasts in Fig. 2 contain 126 observations and represent the highest number of observations of all 12 time horizons. These are forecasts published between November and January for the coming year. Although they scatter substantially, there is still a positive correlation between the forecast and actual values. This recedes in the subsequent charts. For the $h15$, $h18$ and $h21$ forecasts there is hardly any positive correlation between forecast and actual GDP. The same applies to the predictions for $h24$ to $h33$, which are not shown here. They form a horizontal band in the upper two squares of the scatter diagrams. In other words, they only change slightly. They thus possibly provide information on the expected trend or potential growth, but no longer give any information on the future business cycle.

2.2 Descriptive statistics: ME, MAE, RMSE and Theil's U

Forecast errors can be quantified in various ways. Table 3 shows six common statistics for the time horizons $h = 0$ to $h = 33$. Note that for some horizons only a small number of observations are available (cf. Table 2).

The Mean Error (ME) shows whether the forecasts are too high or too low on average and thus indicates whether they are biased. In the forecasts for the current year ($h = 0,3,6,9$) the mean error is close to zero. It is also close to zero for $h = 12$, where Wasserfallen (1992) identifies a slightly positive forecast error. The mean errors increase substantially with the forecasting horizon. For the longest horizon ($h = 33$), it is 1.3 percentage points, which is only slightly lower than the average actual GDP growth rate. Moreover, from $h = 15$ the mean error is consistently negative, i.e. the actual growth rate is always overestimated. This is mainly because the extremely negative forecast errors are higher than the extremely positive errors from this horizon onwards (see MIN and MAX).

The Mean Absolute Error (MAE), i.e. the mean absolute difference between actual and forecast growth, is a yardstick of the accuracy of the forecasts. It rises significantly with the forecasting horizon.

The picture is similar for the Root Mean Squared Error (RMSE). This commonly used statistical measure gives major errors a higher weighting than minor ones, thus taking account of the fact that the main aim of forecasting is to avoid large errors. The RMSE also rises with the forecasting period. While it is below 1 for forecasts for the current year ($h = 0,3,6,9$), from $h = 18$ it exceeds the standard deviation for actual GDP (1.55).

Another descriptive statistic shown in Table 3 is the inequality coefficient of Theil. In analogy with Winker (2002, p. 257), we define it as RMSE divided by the standard deviation of the actual real GDP. If U is less than 1, this means the forecast error is smaller than the standard deviation, i.e. the average variation of the actual values. In this case the forecasters score better than an alternative forecasting method which uses the 1981–2000 average of the actual GDP growth rates as a forecast. The results show that for forecasts for the current year ($h = 0,3,6,9$) the value for U is well below 1. Similarly, forecasts for the next calendar year generated between August and January ($h = 12,15$) are more accurate than the average for the actual GDP growth rate. That changes from $h = 18$, where U is greater than 1. As the charts show, the long-term forecasts are virtually meaningless as regards future business cycles.[3]

Forecast error 1981–2000 Table 3

	ME	MIN	MAX	MAE	RMSE	Theil's U
Forecasts for the current calendar year						
h=0	0.110	−0.8	2.1	0.340	0.480	0.31
h=3	0.020	−1.7	1.6	0.446	0.575	0.37
h=6	0.039	−1.7	2.2	0.632	0.803	0.52
h=9	−0.112	−2.2	2.2	0.805	0.989	0.64
Forecasts for the following calendar year						
h=12	0.034	−2.5	2.5	0.988	1.175	0.76
h=15	−0.286	−3.5	2.4	1.175	1.462	0.94
h=18	−0.514	−3.8	1.7	1.306	1.635	1.06
h=21	−0.583	−4.1	3.1	1.439	1.790	1.16
Forecasts for the next-but-one calendar year						
h=24	−0.572	−4.0	2.1	1.393	1.730	1.12
h=27	−0.721	−4.1	2.0	1.607	1.919	1.24
h=30	−0.250	−4.3	1.6	1.900	2.393	1.55
h=33	−1.322	−3.7	1.5	1.670	2.069	1.34

Forecast error: Actual value minus predicted value. A plus sign indicates an underestimate while a minus sign indicates an overestimate.

Mean (ME), minimum (MIN) and maximum (MAX) forecast error

RMSE: Root Mean Squared Error

Theil's U: RMSE divided by the standard deviation (SD) of the actual values

3 This comparison is not entirely fair because it assumes a knowledge of the average GDP growth rate for 1981–2000. Later on we use a comparison based only on information that was actually available to the institutes when they made their forecasts.

3 Optimality conditions of the forecasts

In both the chart analysis and the descriptive statistics, there is a risk that random patterns may be interpreted as systematic patterns. Econometric tests can be used to minimise this risk.

This section looks at three necessary conditions required for optimal forecasting. Firstly, forecasts should be unbiased (Section 3.1). Secondly, there should be no autocorrelation between forecast errors (Section 3.2). Thirdly, forecasts should be efficient, i.e. forecast errors should not correlate with information that was generally available at the time when the forecast was made (Section 3.3).[4]

The tests are only carried out for institutes that made at least 18 forecasts for a given horizon. We calculate the results for each institute individually to avoid statistical problems arising from a possible heterogeneity of the forecasts.[5] Due to lack of data, only forecasts for the current year (h = 0,3,6,9) and the following year (h = 12,15,18,21) are tested.

Econometric tests generally give reliable results, even for small samples, if the forecast errors are normally distributed. Consequently, the normality of the distribution was tested first. These tests did not give any indication that the hypothesis of normal distribution of forecast errors should be rejected.[6] These findings also provide scope to calculate confidence intervals for various forecasting horizons. The method is outlined on page 62.

3.1 Are the forecasts biased?

We start by looking at whether the forecasts are biased, i.e. whether they are systematically too high or too low. A biased forecast is suboptimal because it can be corrected and thus improved on the basis of the bias recognised from past forecasts.

Unbiasedness is often tested using the Theil-Mincer-Zarnowitz equation. This is a regression of the actual values on a constant and the forecast values. In the following we test whether the intercept is equal to zero and the slope is equal to one. Holden and Peel (1990) pointed out that this null hypothesis is merely sufficient but not necessary for unbiasedness. Following their suggestion, we only run a regression of the forecast error e_t^h on the constant (cf. Clements and Hendry (1998), p.57.):

(1) $e_t^h = \alpha + \varepsilon_t$

If the constant α deviates significantly from zero, the hypothesis that the forecast is unbiased is rejected. It should be noted that the ordinary least squares method gives consistent results, but that the standard error is biased if there is autocorrelation in the residuals. For reasons outlined in the next section, autocorrelation of the first order has to be assumed for the forecasts for the following calendar year. In these cases, the standard errors were corrected using the method of Brown and Maital (1981).

Table 4a summarises the results of 15 regressions. The estimated values for the constant α are shown in column 2 with the standard error in brackets. In only one case is the null hypothesis rejected at the usual significance levels. In the other cases, there was no reason to assume that the constant is not zero. The results therefore show that the forecast errors can be considered unbiased, fulfilling the first condition for optimal forecasting.

4 Here we follow Granger and Newbold (1973), who recommend that the optimality of forecasts should be based on an analysis of forecast errors rather than a regression of the actual values on the predicted values. Cf. Clements and Hendry (1998, p. 56).

5 We will be utilising the panel structure of the data in a separate study.

6 The normal distribution of the forecast error was tested using the Jarque-Bera statistics. These tests show a normal distribution for most forecasting horizons. Moreover, the usual tests show that the forecast errors are stationary.

Optimality tests 1 – Unbiasedness of forecasts

Table 4a

	Bias α	P	Autocorrelation LM(3), LM(2)	p-value
1	2	3	4	5
h=0 (3 institutes)	0.180 (0.082)	0.040**	3.888	0.274
	0.103 (0.118[c])	0.397	8.554	0.036**
	0.045 (0.097)	0.649	1.450	0.694
h=3 (2 institutes)	0.080 (0.124)	0.527	6.018	0.111
	0.074[d] (0.209)	0.728	3.250	0.355
h=6 (1 institute)	−0.035 (0.191)	0.856	4.591	0.204
h=9 (2 institutes)	−0.179[d] (0.194)	0.370	2.163	0.539
	0.242[d] (0.230)	0.306	0.603	0.896
h=12 (4 institutes)	0.160 (0.327[a])	0.625	0.450[b]	0.778
	0.025 (0.289[a])	0.931	0.369[b]	0.831
	−0.184[d] (0.388[a])	0.635	0.598[b]	0.741
	0.083[e] (0.337[a])	0.805	0.014[b]	0.993
h=15 (1 institute)	−0.265 (0.343[a])	0.439	0.156[b]	0.925
h=18 (1 institute)	−0.658[d] (0.522[a])	0.207	0.754[b]	0.686
h=21 (1 institute)	−0.400[e] (0.540[a])	0.459	1.664[b]	0.435

α was estimated by the least squares method, whereby the standard error was calculated using the Brown and Maital method (1981). The LM-test for autocorrelation is the Breusch-Godfrey test. The start values for the test equation, i. e. the lagged right-hand variables outside the sample, were set at zero.

a) Standard error estimated using the Brown and Maital method (1981) with a lag of the first order
b) The residuals lagged by one period are omitted from the Breusch-Godfrey test equation.

c) Newey-West HAC standard error with truncation after the second lag
d) The data for 1981 are missing.
e) The data for 1981 and 1982 are missing.

***, **, *: Significance level 1%, 5%, 10%

3.2 Are the forecast errors autocorrelated?

The second condition for an optimal forecast requires that, for the forecasts of the current year, the forecast errors should not be autocorrelated. Autocorrelation would occur if, for example, a negative forecast error (overestimate) was always followed by a positive forecast error (underestimate) and vice versa. Such forecasts are sub-optimal because they can be corrected and thus improved on the basis of the observed pattern of forecast errors and the observed autocorrelation of the forecast errors.

Autocorrelation of forecast errors should be judged differently if the forecasting horizons overlap. Let us take as an example the forecasts generated in May, June and July for the following calendar year ($h = 18$): both the forecast errors of the forecast generated in June 2000 for 2001 and of the forecasts made in June 2001 for 2002 are likely to have been affected by the terrorist attacks in the USA on 11 September 2001. As a result, it is probable that both forecasts overestimate GDP growth. What has been illustrated with the example of 11 September is naturally repeated year for year. In other words, autocorrelation is to be expected for all forecasts for time horizons of a year or more ($h = 12, 15, 18, 21$) because the overlapping forecasting period contains common events (innovations). However, in such cases, the autocorrelation should only be related to the previous year's errors (autocorrelation of the first order), not to the forecast errors made two or more years before.

To test for autocorrelation of forecast errors, we can use the results shown in Table 4a. Since the residuals from equation (1) represent the forecast error adjusted for the mean value, testing for autocorrelation of the residuals is equivalent to testing for autocorrelation of forecast errors. Column 4 shows the results of the Lagrange Multiplier Test (LM) after Breusch-Godfrey. The p-value in column 5 shows the probability of the LM-statistics, assuming that there is no autocorrelation of the residuals ($h = 0,3,6,9$) or that only autocorrelation of the first order is identified ($h = 12,15,18,21$).[7] The results show that – with one exception – the p-values are over 5%. From this we conclude that there is no autocorrelation of the forecast errors that would contradict the optimality of the forecasts.

3.3 Informational efficiency of the forecasts

Efficiency is the third requirement for optimal forecasts. A forecast is efficient if it uses all information available at the time when it is made. Tests for informational efficiency therefore look at whether the forecast errors correlate with information known at the time when the forecast was made (orthogonality tests). They are based on a regression of the forecast error on variables that could be observed when the forecast was made:

$$(2) \quad e_t{}^h = \alpha + \beta x_{t-i} + \varepsilon_t \begin{cases} i = 1 \; h = 0,3,6,9 \\ i = 2 \; h = 12,15,18,21. \end{cases}$$

$e_t{}^h$ represents the forecast error in calendar year t and x_{t-i} an information variable (or a vector of information variables). The information variable has a lag of one period if the forecast error refers to forecasts for the current calendar year ($h = 0,3,6,9$), and a lag of two periods if the error refers to forecasts for the following calendar year ($h = 12, 15, 18, 21$). This guarantees that only information available when the forecast was made is used as right-hand variables. The coefficient (or a vector of coefficients) β is zero if the forecast error does not correlate with the variable x_{t-i}. By contrast, if β is not zero, x_{t-i} correlates with the forecast error and the forecast is not efficient relative to the information set to which the variable x_{t-i} belongs. Again, the error term of the regressions for the following year ($h = 12,15,18,21$) may be autocorrelated.

Was efficient use made of the information used for producing the forecast?

In the simplest efficiency test the information variable x_{t-i} is *the observed value* for the series being predicted. In this case, the variable x_{t-i} is the last actual real GDP growth rate. If there is no autocorrelation with the forecast errors and they do not correlate with the GDP growth rate, the forecasts are said to show *weak informational efficiency*.

We refrain from reporting the detailed results of this orthogonality test. The hypothesis $\beta = 0$ cannot be rejected in any of the regressions and, therefore, the forecast errors are not correlated with the GDP growth rate. Taking into account the results outlined in section 3.2 (no autocorrelation), it can be assumed that the forecasts are *weakly informational efficient*.

7 For ($h = 12,15,18,21$) the Breusch-Godfrey test is modified: the residuals for a one-period lag are omitted from the test regression. The results of the Q test (not shown) correspond to the results of the Breusch-Godfrey test.

Tests for stronger informational efficiency can be carried out by extending the set of information. By taking the forecast itself as the variable x_{t-i}, it is possible to check whether the forecasting institute made efficient use of the information available when the forecast was produced. If $\beta \neq 0$, the forecaster did not make efficient use of the information underlying the forecast. If $\beta > 0$, the change in GDP was systematically underestimated; if β is between −1 and 0 the change in GDP was systematically overestimated.[8]

This test can be illustrated with help of the charts. If $\beta \neq 0$, the plots in the scatter diagrams in Figs. 1 and 2 are not along the 45° line. If $\beta > 0$ (or $-1 < \beta < 0$), the line through the scatter plot is flatter (steeper) than the 45° line. By contrast, if $\beta < -1$, the slope in the scatter plot is negative, indicating that the institutes did not manage to forecast whether GDP would grow or contract in the calendar year to which the forecast refers. In this case, while positive (negative) GDP growth rates were predicted, the actual growth rates were negative (positive).

Optimality tests 2 – Informational efficiency regarding the institutes' own forecast Table 4b

	Orthogonality test β	p-value	Autocorrelation LM(3), LM(2)	p-value
1	2	3	4	5
h=0 (3 institutes)	−0.070 (0.051)	0.181	3.095	0.377
	−0.018 (0.073[c])	0.810	8.645	0.034**
	−0.072 (0.072)	0.331	2.694	0.441
h=3 (2 institutes)	−0.036 (0.087[c])	0.681	6.104	0.107
	−0.271[d] (0.201)	0.195	3.293	0.349
h=6 (1 institute)	0.201 (0.178)	0.274	4.793	0.188
h=9 (2 institutes)	0.228[d] (0.177)	0.214	0.963	0.810
	0.075[d] (0.211)	0.725	0.512	0.916
h=12 (4 institutes)	0.029 (0.165[a])	0.860	0.486[b]	0.784
	−0.218 (0.170[a])	0.200	0.635[b]	0.728
	0.375[d] (0.353[a])	0.288	0.553[b]	0.758
	0.575[e] (0.279[a])	0.039**	0.453[b]	0.798
h=15 (1 institute)	−0.090 (0.370[a])	0.807	0.154[b]	0.926
h=18 (1 institute)	−3.217[d] (0.792[a])	0.000***	0.912[b]	0.634
h=21 (1 institute)	−1.430[e] (0.850[a])	0.092*	0.184[b]	0.912

8 This only applies for unbiased forecasts.

β was estimated by the least squares method, whereby the standard error was calculated using the Brown and Maital method (1981). The LM-test for autocorrelation is the Breusch-Godfrey test. The start values for the test equation, i. e. the lagged right-hand variables outside the sample, were set at zero.

a) Standard error estimated using the Brown and Maital method (1981) with a lag of the first order
b) The residuals lagged by one period are omitted from the Breusch-Godfrey test equation.
c) Newey-West HAC standard error with truncation after the second lag
d) The data for 1981 are missing.
e) The data for 1981 and 1982 are missing.
***, **, *: Significance level 1%, 5%, 10%

The results of this orthogonality test are shown in Table 4b. The structure of this table is similar to Table 4a: the standard error for time horizons from $h = 12$ given in brackets is calculated according to Brown and Maital (1981). The third column gives the p-value while the fourth and fifth columns give the results of an autocorrelation test for the residuals, which can be used to test whether the standard errors were calculated correctly.

Up to $h = 18$, β does not normally deviate significantly from zero. It is true that from $h = 6$ there is a trend towards positive β coefficients, indicating a certain "inertia" in the forecasts. However, since the accuracy of the estimates declines (increasing standard error), in all cases except one the hypothesis $\beta = 0$ cannot be rejected.[9] To sum up, the forecasts up to $h = 15$ show *weak informational efficiency* and are also efficient in a stronger sense as they make optimal use of the information used to produce the forecasts. This is not true for $h = 18$ and $h = 21$. In the case of $h = 18$, β is even significantly below −1, so the institutes did not manage to forecast recessions over this forecast horizon.

Are the forecasts strongly informational efficient?

Informational efficiency is strongest when all publicly available information is utilised efficiently. To show that the forecasts are not efficient in this sense, it is sufficient to find information that was generally available to the institutes and that correlates with the forecast error. Such a result would be interesting as it would show how forecasting could be improved. However, a result indicating that no corresponding variables could be found is not very meaningful. Only if all available data had been taken into account – and that is an impossible task – would it be possible to conclude that the forecasts are *strongly informational efficient*.[10]

In this section, we show that information exists which could be used to improve forecasting – namely, forecasts from other institutes. Assuming informational efficiency, the forecast errors of one institute should not correlate with the forecasts of another. If there is a correlation, the institute has not sufficiently analysed the forecasts made by the other institute. There are either shortcomings in "forecasting technology" or the institute has overlooked the fact that the other institute utilises data which it ignores or does not pay sufficient attention to.

To test this, the forecasts made by other institutes are taken as the x_{t-i} variable in Equation 2. To make sure that this information was available when the forecast was made, we only use forecasts made in the previous three-month period. For instance, to test the informational efficiency of forecasts for $h = 0$ a regression of the corresponding forecast errors on the forecasts for $h = 3$ is performed.

Table 4c summarises the results. Columns 2-7 are headed by the six institutes whose forecasts are used as the variable. The results show that the forecasts made by one institute correlate in one case, while those of another correlate three times. The second institute is the OECD, which possibly has better information on the international economic situation than the other institutes. In any case, the results shown in this table indicate that the forecasts are not *strongly informational efficient*.

In the literature, forecasts that are both unbiased and show strong informational efficiency are described as *strongly rational*. As this section shows, there is information which is not utilised, so the forecasts are not *strongly rational*. As set out in Section 3.1, the forecasts for the current and following calendar year are nevertheless unbiased and thus *weakly rational*.[11]

9 This contradicts Wasserfallen (1992, p. 300). For $h = 12$ Wasserfallen assumed that the actual change in GDP was overestimated.

10 In the literature, strong informational efficiency is generally referred to simply as informational efficiency.

11 Alternative terminology refers to unbiasedness and weak informational efficiency as weak rationality (see e.g. Kirchgässner, 1993). The forecasts are weakly rational in this sense, too.

Optimality tests 3: Informational efficiency regarding the others' forecasts

Table 4c

Forecasting period	I1	I2	I3	I4	I5	I6	LM(3), LM(2)	p-value
1	2	3	4	5	6	7	8	9
h=0		−0.089[d]				0.020[d]	7.267	0.064*
(3 institutes)		(0.089[c])				(0.102[c])		
		−0.147[d]				0.240[d]	8.943	0.030**
		(0.171[c])				(0.244[c])		
		0.051[d]				0.086[d]	3.478	0.324
		(0.095)				(0.140)		
h=3				−0.003			6.032	0.110
(2 institutes)				(0.120)				
				0.370[d]**			4.600	0.204
				(0.175)				
h=6		0.610[d]			−0.354[d]		5.768	0.123
(1 institute)		(0.416)			(0.403)			
h=9	0.563[d]		0.228[d]	−1.181[d]*			4.041	0.257
(2 institutes)	(0.600)		(0.309)	(0.663)				
	0.215[d]		0.308[d]	−0.829[d]			2.161	0.540
	(0.758)		(0.390)	(0.837)				
h=12		−0.173					0.326[b]	0.850
(4 institutes)		(0.274[a])						
		−0.320					0.183[b]	0.912
		(0.344[a])						
		0.734[d]**					0.028[b]	0.986
		(0.336[a])						
		0.429[e]					0.179[b]	0.914
		(0.347[a])						
h=15				−2.094[d]***			0.837[b]	0.658
(1 institute)				(0.531[a])				
h=18		−0.515[e]					1.082[b]	0.582
(1 institute)		(0.920[a])						

β was estimated by the least squares method, whereby the standard error was calculated using the Brown and Maital method (1981). The LM-test for autocorrelation is the Breusch-Godfrey test. The start values for the test equation, i.e. the lagged right-hand variables outside the sample, were set at zero.

a) Standard error estimated using the Brown and Maital method (1981) with a lag of the first order
b) The residuals lagged by one period are omitted from the Breusch-Godfrey test equation.

c) c) Newey-West HAC standard error with truncation after the second lag
d) The data for 1981 are missing.
e) The data for 1981 and 1982 are missing.

***, **, *: Significance level 1%, 5%, 10%

4 Comparison with naive forecasts

In the previous sections, we saw that forecasts up to $h = 18$ satisfy the optimality criteria for forecasts. The question now is whether such forecasts are superior to naive forecasts. We define naive forecasts as simple forecasting methods requiring very little effort.

A first naive method, which we call naive forecast 1, takes the average GDP growth in the past 20 years as a forecast. For 2000 this naive forecast is thus the average GDP growth rate for 1980–1999.[12]

A second naive method called naive forecast 2 uses the last actual GDP growth rate as a forecast. Naive forecast 2 for 2000 thus corresponds to the actual growth rate in 1999.

Table 5a compares the RMSE of the two naive forecasting methods with the RMSE of the forecasts generated by the institutes. This comparison is based on four different time horizons. We assume that the two naive forecasts are published in March, as soon as the average GDP growth rate for the previous year is known. Hence, some of the $h = 9$ and $h = 21$ forecasts generated by the institutes in February to April were made before the GDP growth rate is published and are thus not based on the same information as the naive forecasts. By contrast, for the forecasts made between May and July for $h = 6$ and $h = 18$ the first GDP estimate was already available as these forecasts were made up to three months after the naive forecasts.

The results show that the forecasts made by the institutes for the current calendar year ($h = 6$ and $h = 9$) have far lower forecast errors than the two naive forecasts. Looking at the forecasts for the following calendar year, the quality of the forecasts for $h = 18$ is roughly the same as for naive forecast 1, while the forecasts for $h = 21$ are inferior to naive forecast 1. Moreover, both forecasts made by the institutes for the following year ($h = 18$ and $h = 21$) are better than the naive forecast 2.

Comparison with naive forecasting methods Table 5a

RMSE 1981–2000

Forecasting period	All institutes	Naive Forecast 1: Average growth rate	Naive Forecast 2: Last actual growth rate
h=6	0.803 [0.519]	1.534 [0.991]	1.618 [1.045]
h=9	0.989 [0.639]	1.534 [0.991]	1.618 [1.045]
h=18	1.635 [1.056]	1.564 [1.011]	2.263 [1.527]
h=21	1.790 [1.156]	1.564 [1.011]	2.263 [1.527]

12 For the effective values prior to 1981 we have to use the first data issued by the SFSO.

All institutes
Figures in square brackets:
RMSE/SD (Theil's U)

In a next step the Diebold and Mariano test (1995) is used to check the statistical significance of the results. The hypothesis tested is that the Mean Squared Error (MSE) is the same for the naive forecasts and the forecasts of the institutes. We still assume a squared loss function with positive and negative errors of the same magnitude leading to equally high losses and above-average weighting of large forecast errors. For the institutes' forecasts, we take the forecasting series of a forecaster with a large number of observations in the relevant period.

The results are shown in Table 5b. This table shows the difference between the MSE for the naive forecasts and the forecasts made by the institutes for $h = 0$ to $h = 21$. From this it is evident that the forecasts generated by an institute for $h = 0$ to $h = 12$ is significantly better than the two naive forecasting methods. The null hypothesis, which postulates that there is no difference, is consistently rejected at the 5% significance level and in most cases at the 1% significance level as well. Moreover, the forecasts generated by an institute for $h = 15$ and $h = 18$ are better than naive forecasts 2. In all other cases, the difference is not statistically significant. In other words, the forecasts made by the institutes are statistically not significantly worse than the two naive forecasts.[13]

Loss differentials between naive forecasts and institutes' forecasts 1981–2000 Table 5b

	Institutes' forecasts versus naive forecast 1	**Institutes' forecasts versus naive forecast 2**
Forecasts for the current calendar year		
h=0	2.192*** (0.299)	2.457*** (0.659)
h=3	2.053*** (0.263)	2.318*** (0.699)
h=6	1.660*** (0.272)	1.925** (0.740)
h=9[a]	1.742*** (0.337)	1.713** (0.631)
Forecasts for the following calendar year		
h=12	1.069** (0.411)	4.207** (1.734)
h=15	0.527 (0.518)	3.665** (1.536)
h=18[a]	−0.701 (0.849)	2.590* (1.273)
h=21[b]	−0.330 (0.767)	1.703 (1.179)

The forecasts made by the institutes are represented by the forecasting series of an institute with a large number of observations in the relevant period.

Standard error calculated using the Newey-West method
***, **, *: Significance level 1%, 5%, 10%
a) 1982–2000, 19 forecasts
b) 1983–2000, 18 forecasts

13 In a few cases, autocorrelation was observed so the Newey-West correction was used. Estimation with ARMA residuals gives similar results.

5 Forecast errors versus revision of GDP

The last question we wish to look at in this study is the impact of revisions of the actual values on forecast errors. Actual GDP figures are revised several times and occasionally completely reworked as a result of major changes in the calculation method of the national accounts. So far in this paper, we have used the first GDP estimate for the previous year published by the seco in March as the actual value.

Following publication of the first estimate by the seco, the SFSO publishes its own figures on real GDP and its components in the third quarter. These figures are based on a larger data set. The seco then adjusts its data in line with this figure. We start by looking at whether the forecast errors made by the institutes in our sample would be lower if we took the SFSO figure rather than the seco figure as our outcome.

Table 6 shows the forecast errors for the institutes for $h=0$ to $h=21$ measured by the RMSE. This shows that for the shorter time horizons ($h=0,3,6,9,12$) the RMSE is lower relative to the seco data than the SFSO data. For $h=15,18,21$ the opposite is the case, but the RMSE for these horizons are roughly of the same magnitude as the standard deviation of real GDP (1.55) regardless whether they are calculated with the seco or the SFSO data. For these forecasting horizons the forecasts are thus virtually meaningless in both cases.

How should these results be interpreted? Klein (1981) pointed out that revision represents the "limits of forecasting". He argues that if the data are revised by around 10% the mean error cannot be less than 10%[14] because genuine revision errors should by nature be unforecastable.

Institutes' forecasts versus the seco and SFSO actual values 1981–2000 Table 6

	RMSE relative to the actual value of the seco	**RMSE** relative to the actual value of the SFSO
Forecasts for the current calendar year		
h=0	0.480	0.642
h=3	0.575	0.662
h=6	0.803	0.843
h=9	0.989	1.051
Forecasts for the following year		
h=12	1.175	1.215
h=15	1.462	1.417
h=18	1.635	1.530
h=21	1.790	1.666

All institutes
The second column contains data from Table 3 for comparison purposes.

14 See Granger (1996), p. 463 and 464.

Knowledge of the size of the revision of GDP forecasts can therefore provide an indication of how far removed the forecasts are from the "limits of forecasting". We define the revision errors as the difference between the SFSO data and the annual GDP growth rate published by seco. In 1981–2000 the average revision error was 0.075 percentage points and was thus almost unbiased on average. However, individual revisions were between −0.9 and 1.0 percentage points. The average absolute revision error was 0.385, which was a quarter of the average real GDP growth rate. The RMSE is 0.489, which is about a third of the standard deviation in the real GDP growth rate.

In Table 6 the forecast error for $h=0$ relative to seco is 0.480 and the forecast error for $h=3$ is 0.575. That is close to the limit of forecasting. This result indicates that the forecasts can only be improved if the responsible statistical agencies succeed in reducing the revision error. As we have shown, from $h=18$ the forecasts are no longer informative. However, if revision errors were reduced by – for example – half, our rule of thumb for the confidence interval (see page 62) indicates that this limit could be extended by one quarter to $h=21$.

The final question is whether the seco estimates meet the optimality criteria. To test this we use the same method as for the forecasts made by the institutes.

Table 7 summarises the results. A regression of the revision error on a constant shows that the constant does not differ significantly from zero (column 2). In other words, the GDP data published by the seco are unbiased. Further, a LM-test does not indicate any autocorrelation of the revision errors. The third column shows that the revision errors cannot be predicted on the basis of the seco figure. The initial GDP estimate thus is *weakly informational efficient* compared with the annual SFSO estimate. A stronger test is to run a regression of the revision error on the institutes' forecasts that were available at the time when the first estimate was made. A link between the first GDP estimate and the autumn forecast of one institute can be identified, but only at the 10% significance level. This means that the first seco estimate of GDP growth ignores information that is contained in the forecasts made by this institute. The informational efficiency of the initial GDP figure is thus not strong.

Optimality criteria for the first GDP estimate (seco) 1981–2000 Table 7

	Unbiasedness; autocorrelation	**Weak informational efficiency**	**Informational efficiency[a]**
Constant	0.075 (0.111)	0.201 (0.149)	0.311 (0.266)
GDP growth		−0.090 (0.072)	
h=0			0.108 (0.570)
h=0			−0.288 (0.448)
h=0			−0.686 (0.425)
h=3			0.749* (0.395)
h=3			−0.032 (0.210)
LM(3)	5.356	3.233	2.256
p-value	0.147	0.357	0.521

The forecasts made by the institutes are represented by the forecasting series of an institute with a large number of observations in the relevant period.
a) The forecast for 1981 is missing.
***, **, *: significance level 1%, 5%, 10%

6 Summary and conclusions

This paper examines the accuracy of Swiss GDP forecasts on the basis of 766 observations by 14 different institutes.

The results show that the forecasts made during the year for the current year or in the autumn for the following year are both informative and clearly better than naive forecasting methods. However, even the forecasts made at year-end for the current year still have an average forecast error of about 0.5 percentage points, which is roughly equivalent to the revision error.

Our study also shows that the forecast error increases sharply as the time horizon increases. For forecasts made between May and July for the following year ($h = 18$), the forecast error is roughly equivalent to the standard deviation of the actual GDP growth rates. Forecasts for even longer horizons no longer provide any meaningful information about the future business cycle. At best they may provide an indication of the potential growth of the Swiss economy. Our findings do not change significantly if the SFSO data are taken as the outcome instead of the seco data.

These somewhat disappointing findings do not only apply to Swiss forecasts. Studies of forecasting in other countries come to similar conclusions. In an extensive survey of the accuracy of European growth forecasts generated in the autumn for the following year, Öller and Barot (2000) reported forecast errors of a similar order of magnitude. Like us, in a study of three leading British forecasting institutes Mills and Pepper (1999) came to the conclusion that as from $h = 18$ forecasts are not a useful tool for assessing business cycle.[15]

It is good to know that for the forecasting horizons where they are informative, the Swiss GDP forecasts meet the usual optimality criteria for forecasts. The forecasts are unbiased, so they can be described as *weakly rational*. Similarly, efficient use is made of the information included in the time series being forecast. The forecasts for the current year and the following year thus are *weakly informational efficient*. They are also efficient in terms of the set of information used to generate the forecast. However, Swiss GDP forecasts in our sample do not pass the most stringent test – i.e. the test of strong informational efficiency – because, in some cases, forecast errors correlate with the forecasts of other institutes.

All in all, our results highlight the considerable uncertainty associated with future GDP and the business cycle. This is important for decision-makers because it is a warning that they should not allow themselves to be lulled into false security.

Given that forecasts are informative at short forecasting horizons only, the forecasting institutes for their part should regularly review and rapidly adapt their forecasts in the light of new information. In recent years, there has been a clear trend to produce several forecasts a year, which is desirable from the users' viewpoint. For the forecasters, it would be helpful if statistical data on Swiss economic activity were published quickly and – if possible – monthly.

15 Mills and Pepper, 1999, p. 247: "It is found that forecasts are not of much use at horizons greater than 18 months (that is, 6 months before the year being forecast)."

A rule of thumb for confidence intervals

Assuming a normal distribution of the forecast errors, 50% of forecast errors are between +/−0.675 standard deviations and 80% are between +/−1.28 standard deviations. From this, a simple rule of thumb for conveying the uncertainty related to forecasts can be derived.

The starting point is an equation that approximates the root mean squared error (RMSE) as a linear function of the forecasting horizon h. Using the data from Table 3 we obtain the following:

1) RMSE = 0.45+0.06*h

where h is the number of months from the time when the forecast is made to the end of the year for which the forecast is made. The equation shows that the RMSE based on the average historical results for all the institutes in 1981–2000 increases by about 0.18 percentage points per quarter. In the forecasts for $h = 6$ the standard error is 0.81, while in the forecasts for $h = 12$ it is 1.17. In the forecasts for $h = 18$ it is 1.5 and thus roughly equivalent to the standard deviation of real GDP.[16]

In the next step, the standard error for the various horizons can be used to calculate confidence intervals. For this we follow the proposal made by Granger (1996) and give the 50% confidence interval in addition to the 80% confidence interval. For a forecast of 2% for $h = 6$ there is a 50% probability that the actual value will be between 1.5% and 2.5% and an 80% probability that it will be between 1% and 3%.

50% CI: $2 \pm 0.675 * 0.81 \approx 2 \pm 0.55$
80% CI: $2 \pm 1.28 * 0.81 \approx 2 \pm 1$

For a forecast of 2% for $h = 12$ there is a 50% probability that the actual value will be between 1.2% and 2.8% and an 80% probability that it will be between 0.5% and 3.5%.

50% CI: $2 \pm 0.675 * 1.17 \approx 2 \pm 0.79$
80% CI: $2 \pm 1.28 * 1.17 \approx 2 \pm 1.5$

This rule of thumb for the confidence interval of forecast error can be used to produce FAN-charts based on historical experience. These could help to improve interpretation of forecasts on the basis of the forecasting horizon and give an indication of the risks involved in the forecast.

16 For unbiased forecasts, RMSE corresponds to the standard error.

Sources

The *forecasts* made by the institutes in our sample were taken from the following publications:

BAK: CH-Plus, quarterly; CREA: Analyses et Prévisions, autumn and spring; CS: 1987–1996: "bulletin" der SKA, monthly, 1997–2000: "Bulletin" der CS, monthly; IMF: World Economic Outlook, autumn and spring; KfK: Mitteilungen der Kommission für Konjunkturfragen, supplement to December edition of "Die Volkswirtschaft", 1993–2000 (spring forecasts): Confederation's Expert Group for Economic Forecasting, seco; KOF: monthly as well as half-yearly reports, autumn and spring; MAT: Prévisions économétriques, annual; OECD: Economic Outlook, June and December; UBS "old": 1976–1986: "Wirtschafts-Notizen", monthly, 1987–1997: "Internationaler Konjunkturausblick", quarterly; SBC: "Der Monat"; SGZZ: "Lagebeurteilung der Bauwirtschaft", annually; SNB: Vorschläge für die Geldpolitik im Jahre 19xx, annually, unpublished; UBS "new": "Outlook Schweiz", quarterly; ZKB: "Konjunkturbarometer", monthly.

The *actual values* for the percentage annual change in real GDP were taken from the following publications:

Annual average on the basis of the quarterly data: 1981–1989: "Wirtschaftsspiegel", no. 3 (March), Swiss Federal Statistical Office , 1990–2000: Statistical Monthly Bulletin of the Swiss National Bank, no. 3 (March)

Annual estimate of the Swiss Federal Statistical Office: Swiss Statistical Year Books, volumes 1982–2001

Literature

Brown, W.B. and Maital, S.M. 1981. What Do Economists Know? An Empirical Study of Experts' Expectations. *Econometrica* 49: 491–504.

Clements, M.P. and Hendry, D.F. 1998. *Forecasting economic time series*. Cambridge: Cambridge University Press.

Diebold, F.X. and Mariano, R. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13: 253–265.

Granger, C.W.J. 1996. Can We Improve the Perceived Quality of Economic Forecasts? *Journal of Applied Econometrics* 11: 455–473.

Granger, C.W.J. and Newbold, P. 1973. Some Comments on the Evaluation of Economic Forecasts. *Applied Economics* 5: 35–47.

Hendry, D.F. and Ericsson, N.R., eds. 2001. *Understanding Economic Forecasts*. The MIT Press.

Holden, K. and Peel, D.A. 1990. On Testing for Unbiasedness and Efficiency of Forecasts. *Manchester School* 58: 120–127.

Kirchgässner, G. 1993. Testing Weak Rationality of Forecasts with Different Time Horizons. *Journal of Forecasting* 12: 541–558.

Klein, L.R. 1981. *Econometric Models and Guides for Decision Making*. New York: Free Press.

Mills, T.C. and Pepper, G.T. 1999. Assessing the Forecasters: An Analysis of the Forecasting Records of the Treasury, the London Business School and the National Institution. *International Journal of Forecasting* 15: 247–257.

Newey, W.K. and West, K.D. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55: 703–708.

Öller, L.-E. and Barot, B. 2000. The Accuracy of European Growth and Inflation Forecasts. *International Journal of Forecasting* 16: 293–315.

Steckler, H.O. 2002. The Rationality and Efficiency of Individuals' Forecasts. In *A Companion to Economic Forecasting*, M.C. Clements and D.F. Hendry (eds.), 222–240. Oxford: Blackwell.

Wasserfallen, W. 1992. Konjunkturprognosen in der Schweiz. *Finanzmarkt und Portfolio Management* 3: 296–302.

Winker, P. 2002. Vektor Autoregressive Modelle. In *Finanzmarkt-Ökonometrie*, M. Schröder (ed.), 213–262. Stuttgart: Schäffer-Poeschel Verlag.