Essay 4—The accuracy of European growth and inflation forecasts

Abstract. One-year-ahead forecasts by the OECD and by national institutes of GDP growth and inflation in 13 European countries are analysed. *RMSE* was large: 1.9 % for growth and 1.6 % for inflation. Six (11) OECD and 10 (7) institute growth forecast records were significantly better than an average growth forecast (the current year forecast). All full record-length inflation forecasts were significantly better than both naive alternatives. There was no significant difference in accuracy between the forecasts of the OECD and the institutes. Two forecasts were found to be biased and one had auto correlated errors. Directional forecasts were significantly better than a naive alternative in one-half of the cases. Overall, inflation forecasts were significantly more accurate than growth forecasts, and in contrast to growth forecasts, they generally improved over time. This has implications for economic policy. Positively biased revisions reveal large errors in data.

Keywords*:* Forecast accuracy, directional errors, forecast tests.

Acknowledgements

We want to thank the OECD and the national institutes mentioned in Appendix A for helping us with the data. Special thanks go to Michael Andersson, Karl-Gustav Hansson, Alfred Kanis, Svante Öberg, Erik Ruist and Steve Wentworth. We are also grateful to colleagues at NIER and to discussants at the ISF 99 for helpful comments and support, and to an associate editor and two referees who helped make this a better product. The responsibility for any remaining errors lies entirely with the authors.

Published in the *International Journal of Forecasting 16(2000) 293-315*.

1. Introduction[1]

Macroeconomic forecasts attract the interest of the general public, for obvious reasons, but only as long as they refer to the future. Few care about old forecasts of events in the past. But if we do not know the past record of a forecaster, how are we to judge his/her last statement about the future? It should be the responsibility of all serious forecasters to regularly publish reports with an analysis of their forecast records using adequate statistical methods. Granger (1996) suggests that point forecasts should be supplemented by confidence intervals, based on past performance.

The forecasts of large international organisations, such as the IMF and the OECD, and of some national forecasting institutes are occasionally scrutinised, cf. Artis (1996), Ash et al. (1998), Mills and Pepper (1999) and McNees (1992) for some of the most recent reports. This study compares the accuracy of real annual output growth and inflation forecasts made by the OECD for 13 European countries and forecasts made by an institute in the country studied[2].

The first hazard one encounters is becoming lost in dimensionality. There are many forecasters, and each forecasts many variables for several horizons. Different periods can be studied, there is a multitude of ways to assess accuracy, compare forecasters, etc. In order to keep the analysis manageable, we will study just two variables: *growth*, as measured by the annual percentage change in GDP, and *inflation*, measured by the annual percentage change in the consumption deflator. The horizon is one year ahead.

The next obstacle involves the very definition of accuracy. Assessing the accuracy of a forecast *ex post* may seem to be a simple problem: one just measures the distance between the forecast and the "known" outcome. But in forecasting GDP, the outcome is not known in the sense of aiming a weapon at an immobile target. All national statistical offices first publish a

---

[1] Two earlier reports from this study are Öller & Barot (1998, 1999).

[2] For a recent comparison of (mainly) US and UK forecasts, cf. Fildes and Stekler (1999).

preliminary figure, which can best be described as an informed guess, i.e. it is also a forecast. Successive revisions, some many years later, will reduce the share of approximation in the figure, but they never eliminate it completely. Hence, comparing forecasters according to how close they come to a published figure is also a comparison of the "outcome" data. McNees (1989) describes forecasting and revising as a continuous process that starts long before the period concerned, and continues long after.

In Ash et al. (1998)[3], forecasts made by the OECD of G7 countries and 20 macroeconomic variables are studied and tested, using forecast records from the period 1967-1987, and three forecast horizons: ½, 1 and 1 ½ years ahead. Taking ordinary differences of seasonally adjusted GDP, their main result is that what they call "quasi-forecasts", (i.e. forecasts for the current half-year or a forecast that is not a forecast in real time), generally are useful, in the directional sense. When the horizon is extended to one year ahead, there still is some indication that growth forecasts are valuable to users, but only in the case of France, the UK and the US. When the horizon is 1½ years, only the UK forecast is better than a naive alternative. Stekler (1994) analyses three organisations that have forecasted quarterly GNP for the United States for the period 1972-1983. Direction and rough size of change are studied. Again, the main conclusion is that current quarter quasi-forecasts are useful, while the results for one-quarter-ahead forecasts are ambiguous. Note that our comparison will use *annual* figures.

The data are presented in Section 2. In Section 3, we look at root mean squared errors *(RMSE)* and we test for improvement on two naive alternatives using the Wilcoxon signed rank test of Diebold and Mariano (1995). In choosing a naive alternative, we have endeavoured to reconstruct the situation in which the forecast was made. Hence, if the naive alternative is the average growth or the previous value of the series, we have only used data that were available to the forecaster. Consequently, the current year figure will be the forecast made in the autumn of the same year. We also test for accuracy improvement over time. A t-test is used to determine if inflation forecasts are significantly more accurate than growth forecasts. Weak form informational efficiency, i.e. bias and/or autocorrelation, is studied in Section 4. Non-parametric measures of accuracy based on direction allow for a different assessment. In Section 5 we look at acceleration/deceleration and test against a naive variant using a standard contingency table. In Section 6 we take a brief look at revisions and Section 7 summarises and discusses the results.

2. Data [4]

OECD annual growth and inflation forecasts have been collected from the December issue of the *OECD Economic Outlook,* 1971-1998. When these forecasts were made, preliminary data for the first half of the current year were the latest data available. Table A.1 in the Appendix A lists the European forecasting institutes that have kindly provided us with their forecast data. The dates of publication vary among institutes and even within the same institute, but they take place in the autumn[5], and treat the current and the following year. As can be seen in the last two columns, many institutes have forecasts only for more recent periods (an unbalanced panel). This has necessitated separate studies of the sub-periods. OECD inflation forecasts have been published for all involved countries only from 1975 onwards. Exhibit 1a contains graphs of the growth forecasts and Exhibit 1b shows the inflation forecasts. Generally speaking, the national institute and OECD forecasts are very close to each other; the coefficients of correlation for the 13 countries are in the interval 0.74 to 0.99 for growth and between 0.61 and 0.98 for inflation forecasts; so that the two are in most cases jointly

[3] This contains an excellent list of references.
[4] The data can be supplied upon request.
[5] ISCO/ISAE's inflation forecasts are issued in February and July. We used the February figures.

good or bad. Note also the general reluctance to forecast negative growth. Some forecasters report no negative figures, although all economies have experienced contractions; some as serious as –7 %, as in Finland and Switzerland, but in both cases forecasts were non-negative. These forecasters may have had an asymmetric loss function of the "linex" type, according to which negative growth forecast errors carry a much larger penalty than positive ones[6]. Inflation forecasts have smaller errors and the falling tendency has been captured quite well by the forecasters.

[EXHIBIT 1 HERE]

Outcome (for both the OECD and institutes) is defined as the figure published in the December issue of the *Economic Outlook* one year later. This is probably the figure most forecasters are aiming for, because it is sufficiently near the time when the forecast was made to be regarded as relevant to the debate[7].

3. Root mean squared errors and naive forecasts

The most common average error measure, *RMSE,* is shown in Exhibit 2a for growth and Exhibit 2b for inflation. Panel A contains the country forecasts of the OECD and Panel B shows those from the national institutes. We treat the growth forecasts first. For the entire period 1971-1997, the *RMSE* is in the range of 1.3 % for the OECD's forecast of growth in France, and 2.6 % for both forecasts for Finland. Even the minimum is above one percentage point, which must be considered unacceptably large. Still, they are of the same order as reported in other studies, see e.g. Zarnowitz (1992) for the US as well as Artis (1996) for the G7 countries; but the errors are smaller than for more volatile quarterly growth forecasts, cf. e.g. McNees (1986).

The standard deviation *(SD)* of the outcome is larger for Finland than for the other countries, which indicates that this variable is more difficult to forecast. In the second column, *RMSE* has been divided by *SD*. Now the most accurate growth forecaster is IFO.

[EXHIBIT 2 HERE]

Inflation forecasts in Exhibit 2b generally have lower *RMSE* (1.6) than growth forecasts[8]. MIFF has the lowest *RMSE/SD,* followed by the OECD forecast for inflation in France and in the UK and the NIESR[9] forecast. Note in Exhibit 1b the overall high volatility in the beginning of the period. *RMSE* appears to generally fall towards the end of the period, but not sufficiently to prevent a rise in *RMSE/SD* in the last period.

The ratio *RMSE/SD* can be interpreted as an approximation of the *Theil* U_2 *Index*[10]. A value above unity means that the forecast is no more accurate than a "naive" average change forecast and hence has no more value to the user than a simple naive projection. We have also compared the forecasts to a naive alternative of "same change as last year". Both measures are approximations. Instead of the mean growth over the entire period, one should use only the

---

[6] Cf. Varian (1975).

[7] Cf. OECD Economic Outlook (1995) and Ash et al. (1998).

[8] This is not the case for forecasts made by the EU, for which Keereman (1999) reports: RMSE (growth) = 1.3, wheras RMSE (inflation) = 1.5.

[9] Holden & Peel (1985) present a thorough analysis of NIESR forecasts.

[10] See Holden et al. (1994), p. 338.

information available at the time of the forecast. The notation * means significantly more accurate than in the first naive alternative. The average growth rate estimates were calculated using data known at the time of forecasting. The test against the naive alternative "latest growth/inflation" is based on the *forecast* for the current year, where † denotes significance. In both cases we have used the Wilcoxon signed-rank test (S_3) in Diebold and Mariano *(DM)* (1995).

The naive trend growth would be the best possible forecast for change in a variable y_t, if it would be close to a random walk with drift, and if nothing else would be known about y_t than its own history. A random walk with drift can be written:

$$\Delta y_t = \mu + \varepsilon_t \tag{1}$$

where Δ is percentage growth, $\mu > 0$ is average (percentage) change and the errors ε_t are i.i.d. $(0,\sigma^2)$. For testing if the forecasts are more accurate than what can be generated by (1), μ must be estimated recursively as a time varying mean.

The naive "latest change" would be optimal if y_t would be close to a random walk in percentage changes:

$$\Delta y_t = y_t + \varepsilon_t \tag{2}$$

When testing against (2), we project the autumn growth forecast for the current year one year ahead.

In Exhibit 2a both growth forecasts for Norway have *RMSE/SD* ratios above unity for the period 1971-97. A possible explanation is the unpredictability of the offshore economy of Norway[11]. In Exhibit 2b we find ratios above unity only for the last period for both forecasts for Ireland and for the one by FPB.

As in McNees (1978) we have tested published forecasts against naive alternatives, pooling all forecasts as if the same person had made them all. The null hypotheses of no better accuracy were rejected (last row), both using the binomial test suggested in ibid. and when applying the *DM* test. Note that we study one-year-ahead forecasts, calculated as *annual* growth rates, not as successive differences of semi-annual or quarterly figures as in Ash et al. (1998) and Stekler (1994), where forecasts quickly lost their accuracy. Using the *DM* test, we also checked if there is a difference in accuracy between the OECD and national institutes, but we found no significant difference.

The *DM* test shows that 6 OECD growth forecasts (Exhibit 2a) were significantly more accurate than projections of the average growth one year ahead. All but the OECD forecasts for growth in Norway and Switzerland were better than the naive alternative "current year forecast". For the national institutes, the corresponding numbers were 10 and 7. Note that some tests are based on fewer observations, because of lack of historical data from some institutes. All inflation forecasts with full length records were significantly better than both naive alternatives.

We did not expect that so many forecasters would prove better than naive projections, given the relatively large forecast errors. In Öller & Barot (1999) we used growth *outcomes* for the current year as naive alternatives. As correctly pointed out by the referees, this makes

[11] For a study on Norwegian macroeconomic forecast accuracy, see Bjønnes et al. (1998).

the comparison unfairly difficult for the forecasters, since they did not know the outcome for the current year when they made their forecasts for the next year. Indeed, by that comparison, only a few forecasters would have been significantly better than the naive alternative.

Could the two time series be regarded as random walks, generated by (1)? We tested to see if mean growth deviations could be regarded as white noise applying a Breusch-Godfrey LM-test. A triangle after the country name in Panel A of Exhibits 2a and 2b marks where this hypothesis could be rejected (we used longer data series, where available). In Exhibit 2a (growth) we see that the hypothesis could be rejected for all but five countries, and in Exhibit 2b (inflation), it is rejected for all countries. Hence, using (1) as a forecasting model, it should not have been difficult to dominate. For the shorter periods (1980-97 and 1985-97), tests were made only in cases where no data were available from the beginning of the period. The figures under the totals show the results when the forecasts are compared to the *final* outcome. In most cases forecasts are closer to the figures released in December the following year than to the final value[12]. However, the overall message of Exhibit 2 remains unchanged, so we decided not to report these results. Also, note the arbitrariness of the concept "final". The point in time when a figure is "final" may vary from country to country. Furthermore, all figures from the 1990s may still be revised.

Comparing Exhibits 2a and 2b, one immediately notices the much smaller *RMSE* and *RMSE/SD* for inflation than for growth; the inflation ratio is only one-half the size of that for growth forecasts. This is an interesting result from the policy viewpoint, and one would like to test to see if the difference in accuracy is statistically significant. The number of error observations being 647 and 568 for growth and inflation, respectively, we used an asymptotic t-test[13] for the null hypothesis that *RMSE/SD* is the same for growth and inflation. A t-value of 9.4 shows that the difference in accuracy hardly can be regarded as a result of pure chance.

The Mean Absolute Errors *(MAE)* were calculated, but are not reported here. The error distributions were found to be close to normal, in which case a rule of thumb is that $RMSE \cong 1.25 MAE$, see Granger (1996), and hence *MAE* does not contain much additional information.

Has the development of forecasting methods during a quarter of a century resulted in improved forecast accuracy? Looking at aggregate *RMSE/SD* at the bottom rows of Exhibits 2a and 2b, we find that there is little change (for the better). Treating the country/institute observed errors as a random sample the absolute errors made by each forecaster were regressed against a constant and a time trend[14]. Only the OECD growth forecasts for Italy and Sweden improved significantly over time. We may note in passing that Kennedy (1969) reports the following characteristics of the NIESR forecasts of growth 1959-1967: $RMSE = 1.4$, $SD = 1.8$ so that $RMSE/SD = 0.8$, which is the same figure as in Exhibit 2a.

Testing for improvement over time in inflation forecasts showed that all had improved significantly, except the OECD forecasts for the Netherlands and Norway and those by IFO and CPB. Not only are inflation forecasts more accurate than the growth forecasts, they also improve over time, but only in absolute terms, not in relation to *SD*.

4. Weak form informational efficiency

If one finds that a forecaster has a tendency to over-estimate or under-estimate the outcome, this is a systematic error (bias) that could easily be corrected, making the forecast more accurate. There is also another type of systematic error. If over-estimating is more likely to be

[12] Comparing errors from a simple econometric model using preliminary figures on the one hand and final ones on the other, Denton and Oksanen (1972) found no improvement in overall fit from using final instead of preliminary, figures.

[13] This test was suggested to us by Michael Andersson.

[14] Thus the test does not discriminate between the case of improved forecasting techniques and that of the task becoming easier because of smaller volatility in the data to be forecasted.

followed by another forecast above the target, and the same tendency applies to under-estimating, then the forecaster is rigid and this shows up as auto correlated one-step-ahead forecast errors. If autocorrelation is known to exist it is systematic and could be taken into account before releasing the final forecast. The absence of bias in one-step-ahead forecasts and no autocorrelation in forecast errors is called "weak form informational efficiency". This is regarded as rational forecasting in the limited sense of McNees (1978). The limitations are that:

(i) The forecaster's error loss function must be symmetric. As noted in Section 2, a higher penalty for negative errors makes it perfectly rational to make forecasts that are positively biased, cf. Clements and Hendry (1998) and Zellner (1986), and

(ii) In case of a symmetric loss function, unbiased ness and absence of autocorrelation are only necessary conditions for rationality. If information that could have reduced *RMSE* was available, but not utilised, the forecast is irrational.

In short, forecast errors should be innovations towards all available information. This will not be tested here, and we avoid the term "rational".

[EXHIBIT 3 HERE]

The conventional test method for bias in forecasts goes back at least to Theil (1966). The actual are linearly regressed on the forecasts and a constant. The null hypothesis is that the regression coefficient is 1 and the constant is 0. The problem with this test is that the residuals will often be positively auto correlated, thus inflating (biasing) the test of the null hypothesis. Autocorrelation is a measure of the rigidity of forecasters and should thus be included in the same test as pointed out by many authors. Instead of applying a Cochrane-Orcutt type transformation (McNees, 1978), or postulating an error-process as in Brown and Maital (1981)[15], we suggest testing both types of systematic errors simultaneously by regressing the forecast error on a constant and as many lagged errors as necessary to whiten the residuals:

$$e_t = \gamma_0 + \sum_{j=1}^{p} \gamma\, e_{t-1} + \varepsilon_t \qquad (3)$$

where e_t is the forecast error at t and the γ_i (i = 0, 1 ,..., p) are coefficients to be estimated. In macroeconomic applications one would expect p to be one or two. No bias or autocorrelation can be tested as a t-test for $\gamma_o = 0$ and an F test that any of the other γ 's are different from 0, respectively. Exhibits 1a and 1b suggest that both types of systematic errors may be present.

The results of the tests are shown in Exhibit 3. We have refrained from testing institute forecasts based on less than full-length records. There is only one significantly auto correlated forecast error record, that of MIFN for growth. Since autocorrelation does not seem to be a general problem here, we also performed a joint likelihood ratio *(LR)* test of actuals regressed on forecasts

[15] See also Mills & Pepper (1999).

INSEE's growth forecasts are biased according to the test based on (3), and inefficient according to the *LR* test. The test based on (3) signals bias in the OECD forecast of inflation for Sweden. OECD's forecasts of growth for Norway are inefficient according to the *LR* test.

Other researchers, e.g. Diebold et al. (1997) have found a tendency to under-estimate inflation during episodes of high inflation and to over-estimate it in periods of low inflation. This tendency should produce autocorrelation. If the tendency exists in European data, with one exception it is not sufficiently strong to trigger significance in tests of autocorrelation. We will return to this question in the next section.

Nothing has been said of heteroscedasticity in errors yet. Indeed, nearly all inflation forecast errors appear to decrease over time (see Exhibit 1b), which is a sign of heteroscedasticity. Regressing outcomes on forecasts, the positive correlation between forecasts and errors will *inflate* the t-values. We performed ARCH tests, which produced significance only in cases where the test for bias did not signal significance, so that the inference from Exhibit 3 does not change.

5. Directional forecasts

Leitch and Tanner (1995) suggest that the numerical accuracy measures *(RMSE, MAE*, etc.*)* have little relevance for users of forecasts in business enterprises, who are most concerned with the *direction* indicated by the forecast. One reason may be that businessmen examine professional growth forecasts in order to decide whether to invest in expanded production capacity. If the investor receives the wrong signal, the result will be either a loss of market share or over-capacity. A central bank wants to know if inflation will accelerate or decelerate to decide if the interest rate should be raised or lowered. Ibid. presents evidence that the US GNP forecasts of 42 professionals are useful in the directional sense.

[EXHIBIT 4 HERE]

Beginning with the growth forecasts, we see from Exhibit 1a that there have been three major recessions in the period studied: (1) the mid-1970s, (2) the early 1980s[16], and (3) the early 1990s. Did the forecasters issue correctly timed signals? The sad answer is, only in rare cases. IFO saw the first recession coming, but its severity was greatly under-estimated so the warning was of doubtful quality. IFO gave a perfect recession alert for the third recession. The OECD issued an almost correctly timed warning for recession (1) in Italy and a perfect one for recession (2) in the UK. In addition to these warnings, only the forecasts for recession 2 in Norway are worth mentioning. Note that there is a total of 29 episodes of substantial recession. Some negative growth forecasts can be seen, but they are poorly timed (see Section 2). Moreover, there are false recession and boom signals, the most remarkable ones: 1984 in Norway and 1977 in Finland and Sweden. MIFN issued a -2 % warning for 1984, when in fact the economy had accelerated from 3 % the year before to a healthy 4 %. The other case is a boom of 4 % growth forecasted by MIFF to occur in 1977, when in fact, essentially zero growth was recorded. NIER forecasted that growth would accelerate from 0 to 2.7 % for 1976-1977. The outcome was a deceleration from 1.4 % to -2.4 %.

Except for two observations, annual inflation has been positive in all countries considered and growth has almost always been positive. Thus, calculating the number of times that the sign has been correctly forecasted does not make sense for inflation, and for growth there are so few observations that we have chosen to comment on them verbally above. We found the following solution to this problem. At the end of the year, both the OECD and the institutes publish forecasts for the following as well as for the *current* year. This means that the

[16] The first two are discussed in Wallis (1989).

forecasts signal if there will be acceleration or a *deceleration* of production/inflation. We study this second order direction, which is done for prices in Ash et al. (1998).

Exhibit 4a shows the years when the forecasters missed the change in *growth*. The OECD's forecasts for growth in the UK have missed only four times (1975, 1977, 1981 and 1996) during the 27-year period studied. Among the institutes, only MIFF has equally few misses. IFO fares much worse in this comparison than when the criterion *RMSE/SD* is used. In order to compare with a naive alternative, the last column contains ratios of forecast misses to the number of misses that would have occurred with a simple projection of the last acceleration/deceleration one-year-ahead. Again, we used the *forecast* of the current year and the autumn statistics for the previous year's outcome, to place ourselves into the position of the forecaster. In an analogy with the Theil Index, it can be said that if the ratio is not below unity, the direction was not predicted better than by a naive forecast. There are eight ratios at or above unity, if sub-periods are also considered: the OECD forecast for Germany, Belgium, the Netherlands and Norway as well as those made by DEC, CPB[17], MIFN and KOF.

For ratios below unity, we have performed a simple contingency table test, see Appendix B. An asterisk in the last column of Exhibit 4a shows where growth forecasts are significantly correlated with the outcome, which is in only one-half of the cases. This contradicts the claim in Leitch and Tanner (1995) that macroeconomic forecasters may be better at direction than at numerical accuracy. When we tested the pooled data, the accuracy as a whole was significantly better than a naive projection.

Acceleration/deceleration forecasts of inflation are shown in Exhibit 4b. Twelve out of 26 of the forecasts are not significantly more correct than the naive alternative, which is almost the same number as for growth forecasts. According to this criterion directional growth and inflation forecasts appear to be equally accurate, but note that overall, the latter are wrong much less often than the former (24 % vs. 29 % for growth). Now both forecasts for the UK and the OECD forecast for Norwegian inflation miss only twice.

The OECD directional inflation forecasts for the Nordic countries, except for Norway, stand out as being particularly poor. The naive alternative (same acc./dec. as the year before) would have missed only three accelerations/decelerations of inflation in Finland, while the OECD emitted wrong signals in 10 cases! The naive variant would have proven better than DEC by 8 - 6 and NIER by 8 - 7. These forecasters may have had a cost function in mind, where wrongly predicting accelerating inflation carries a much higher penalty than wrongly predicting decelerating inflation. Considering that these countries had highly centralised labour markets with strong government involvement in this period, forecasters close to government would have an incentive to issue forecasts of decelerating inflation so as to ward off too high wage claims. Indeed, comparing over-predictions and under-predictions, one gets the following scores: OECD: Finland 10 - 0, DEC 6 - 2 and NIER 8 - 1.

Could acceleration/deceleration be particularly difficult to forecast for certain years? Exhibit 5 presents aggregated OECD and national institute directional forecast misses over time. The year 1988 stands out as the most difficult for forecasting growth. It is not difficult to guess why; in October 1987, too many forecasters still believed in positive and unavoidable correlation between the stock market and production. More surprising is the contrast between the years 1994 and 1995. Here it is difficult to find the reason. Again, inflation forecast misses behave quite differently; there is no exceptionally bad year.

[EXHIBIT 5 HERE]

---

[17] A report on the accuracy of CPB forecasts is given in Donders & Kranendonk (1999).

6. Statistical errors[18]

To understand why errors are so large, we refer to what was said in the introduction about measuring forecast errors as the difference between the forecast and a preliminary figure. Despite the importance of GDP and the consumption deflator as indicators, their measurement is uncertain. In Exhibit 6 we have calculated the *root mean squared revision (RMSR).* This is a measure of the errors detected in the official statistics, as of December, after one revision has already been made. The error measure is biased downwards also because the last observations will be further revised. Moreover, there are all kinds of errors, e.g. from sampling. Assume that two thirds are detected and removed. Then the average statistical error would be only slightly smaller than the average *RMSE* (1.9) of growth forecasts and 40 % of inflation *RMSE* (1.6). This simple calculation shows that forecasters may be able to share the blame for bad forecasts with the statistical authorities[19]. Exhibit 6 also reveals that some revisions are significantly biased. As already noted by Denton and Oksanen (1972), preliminary figures underestimate.

[EXHIBIT 6 HERE]

A rational forecaster dealing with uncertain data avoids big shifts in forecasts; he/she is rigid in the sense discussed in Section 4, inducing positive autocorrelation. There was very little evidence of autocorrelation in Exhibit 3, which means that forecasters were quite efficient when dealing with uncertain data, compensating for the negative autocorrelation that would appear in the forecasts if they would accept their model (real or intuitive) forecasts based on revised data, but using preliminary data as the starting point[20].

7. Conclusion and discussion

In summary, we have found that:

(1). Average errors in forecasts of both growth and inflation are large, in terms of both their variance and the importance of these variables. European macroeconomic forecasts for the following year, of both growth and inflation are as a whole (pooled) significantly more accurate than two naive alternatives. Significant superiority was also found for directional (acceleration/deceleration) forecasts. However, we found no significant difference in accuracy between the OECD and institute forecasts.

(2). Accuracy, as measured by *RMSE* is significantly higher for inflation than for growth forecasts. There are fewer directional misses in inflation forecasts than in growth forecasts.

(3). The following individual forecasts of growth (full-length records) are significantly better than all naive alternatives tested here: average growth, latest growth and random growth direction: the OECD forecasts for France, UK, Austria[21] and Sweden, and among the institutes, NIESR, MIFF and NIER. The following growth forecasts were found not to be useful according to any criterion: the OECD forecast for Norway, and the institute forecasts of MIFN and KOF.

---

[18] A classical reference on statistical errors is Morgenstern (1950).

[19] In Klein (1981) the statistical error in US growth figures is reported as 1.5 percentage points and this is regarded as the lower limit of the RMSE of forecasts, cf. also Granger (1996).

[20] See Clements & Hendry (1988), Ch. 8.3.

[21] For Austrian macroeconomic forecasts, see Thury (1986).

[21] We also looked at two-year ahead growth forecasts made by the OECD. They appear to contain (positive) bias, which is in accordance with the result in Milburn (1978) that optimism increases with the forecast horizon. Bias has also been found for shorter horizons, cf. Kirschgassner (1993) who examines German forecasts.

(4). According to *RMSE/SD,* all inflation forecasts with full-length record were better than both naive alternatives. The OECD forecasts for the UK and France again fare best in the competition, whereas some forecasters had their rankings reversed in comparison with growth forecasts (e.g. the forecasts for Norway). One-half of the directional inflation forecasts were significantly better than the naive alternative.
(5). The only growth forecasts that had improved over time were those of the OECD for Italy and Sweden. The only inflation forecasts that *did not* improve over time were the OECD forecasts for the Netherlands and Norway and those made by IFO and CPB. The improvement was in absolute terms, not in relation to the standard deviation.
(6). Weak form informational efficiency was rejected in very few cases[22].
(7). We found no support in annual data for the claim in Leitch and Tanner (1995) that macroeconomic *directional* forecasts are more accurate than forecasts measured on an interval scale, and turning points had generally been missed.
(8). Accuracy appears to be higher in annual forecasts than in shorter period change forecasts as reported by Ash et al. (1998) and Stekler (1994). This is in agreement with (British) evidence in Barker (1985) that forecasters with a longer perspective were more successful than those working with quarterly data.
(9). Although errors were too large, growth forecast accuracy could not have been substantially improved without improvement in the accuracy of the statistics. This also applies to a lesser extent to inflation forecasts.

As stated above, one of the most interesting results from this study is that inflation forecasts are better than growth forecasts. According to Exhibit 1b, inflation is falling and there has been a decline in the inflation variance over time (although not uniform). Inflation revisions are smaller than those for growth figures. Taken together, these results support stabilisation policies based on inflation targeting, instead of Keynesian fine-tuning of output, which was a policy still being pursued by many European countries in the 1970s. Economic policy requires accuracy in both statistics and in forecasts. The latter depends on the former, and it seems that we cannot achieve decisively better forecasts (and policy) without first improving the statistics, which today is possible through the use of modern data techniques to produce fast and highly accurate on-line statistics.

[22] We also looked at two-year ahead growth forecasts made by the OECD. They appear to contain (positive) bias, which is in accordance with the result in Milburn (1978) that optimism increases with the forecast horizon. Bias has also been found for shorter horizons, cf. Kirschgassner (1993) who examines German forecasts.

Appendix A. Table A1. National institutes

Initials	Name	Data period		Forecast date
		Growth	Inflation	
IFO	Institut für Wirtschaftsforschung, Germany (D)	1971-97	1975-97	December
INSEE/FM	Direction de la prévision, France / Ministry of Finance (F) [1)]	1971-97	1975-97 [2)]	October
(ISCO) ISAE	Instituto Nazionale per lo Studio della Congiuntura, Italy (I)	1981-97	1981- 97	December/ February [3)]
NIESR	The National Institute of Economic and Social Research, UK (GB)	1971-97	1975- 97 [4)]	November, 1971- 95, and October, 1996-97
WIFO	Austrian Institute of Economic and Social Research, Austria (A)	1971-97	1975- 97	December
FPB	Federal Planning Bureau, Belgium (B)	1983-97	1983- 97	Irregular
DEC	The Economic Council, Denmark (DK)	1974-97	1975- 97	December
MIFF	Ministry of Finance, Finland (FIN)	1971-97	1975- 97	September
ESRI	The Economic and Social Research Institute, Ireland (IR)	1971-76 [5)] 1978-97	1975- 97	October/November December
CPB	Bureau for Economic Policy Analysis, Netherlands (NL)	1971-97	1975-97	September
MIFN	Ministry of Finance, Norway (N)	1971-97	1975-97	October
NIER	The National Institute of Economic Research, Sweden (S)	1971-97	1975-97 [6)]	November/December
KOF	Swiss Institute for Business Cycle Research, Switzerland	1976-97	1976- 97	December

Note:

1). INSEE made the GDP forecasts, the Ministry of Finance the inflation forecasts.

2). The OECD forecast for France was used as a substitute for a missing value 1983.

3). The inflation forecasts are made in July and February. Since the latter is closer to December it was chosen.

4). For 1982 the inflation forecast is missing. As a substitute, we used a quarterly forecast made in the fourth quarter of year t for the same quarter in t+1.

5). The GDP forecast for 1977 is missing. Only figures for 1980 onwards were used.

6). In 1975-1980, two alternative inflation forecasts were published. For 1976 and 1978-80, the more likely one was indicated in the text and we chose that one. For the remaining years the forecast is the arithmetic average.

Appendix B . The contingency table test of directional accuracy

The acceleration/deceleration forecasts were tested for information content using the 2×2 contingency table:

		Forecast		
		Acc.	Dec.	Total
Outcome	Acc.	f_{11}	f_{12}	$f_{1.}$
	Dec.	f_{21}	f_{22}	$f_{2.}$
	Total	$f_{.1}$	$f_{.2}$	$f_{..}$

The test is:

$$\chi^2 = f_{..}(f_{11}f_{22} - f_{12}f_{21})^2 / f_{1.}f_{2.}f_{3.}f_{4.}\ ,\ df = 1,$$

where f_{ij}, i = 1,2 are the number of cases

Exhibit 1a

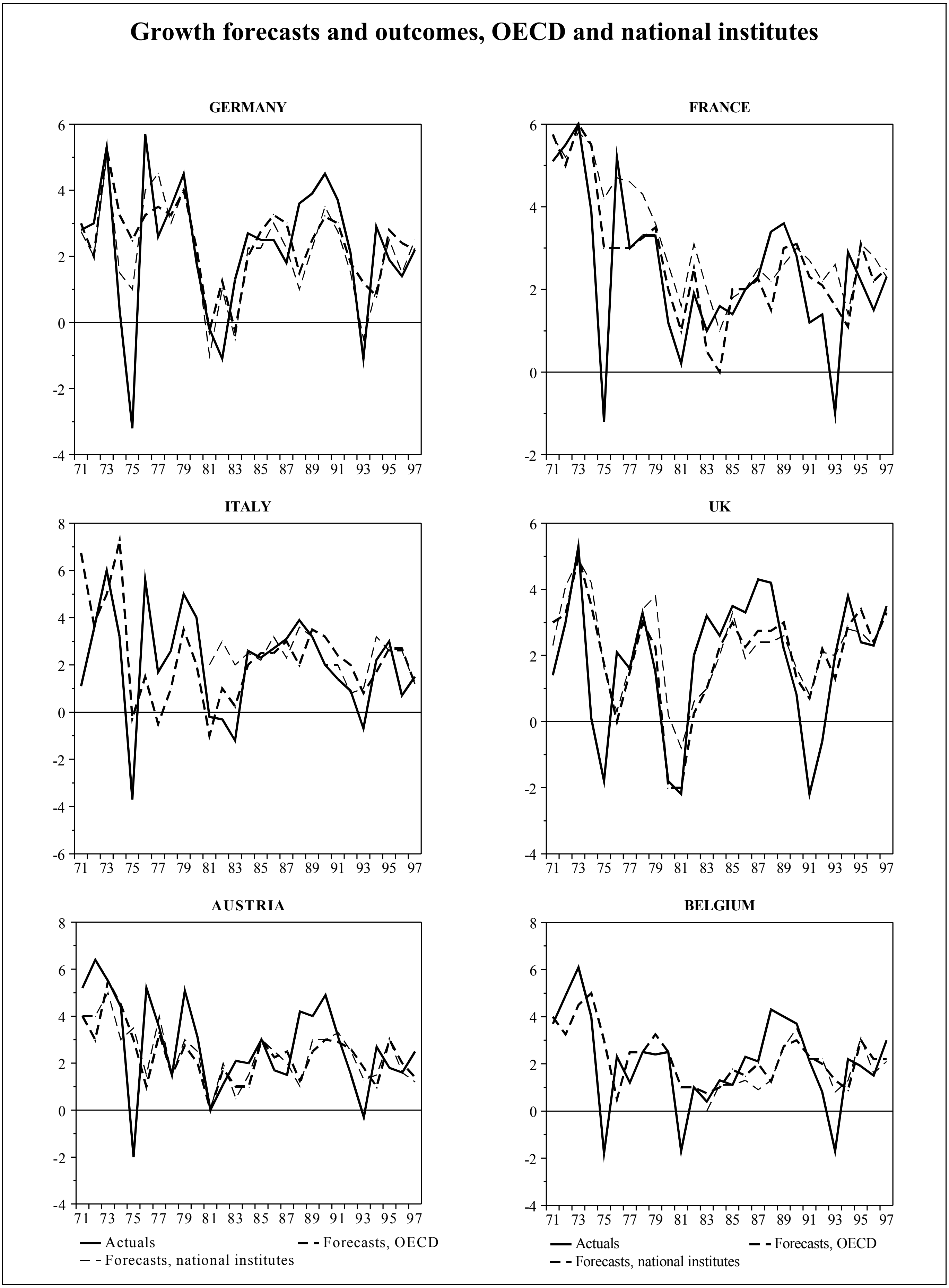
Growth forecasts and outcomes, OECD and national institutes
GERMANY
FRANCE
ITALY
UK
AUSTRIA
BELGIUM
71 73 75 77 79 81 83 85 87 89 91 93 95 97
Actuals
Forecasts, OECD
Forecasts, national institutes

Exhibit 1a continuing

Growth forecasts and outcomes, OECD and national institutes

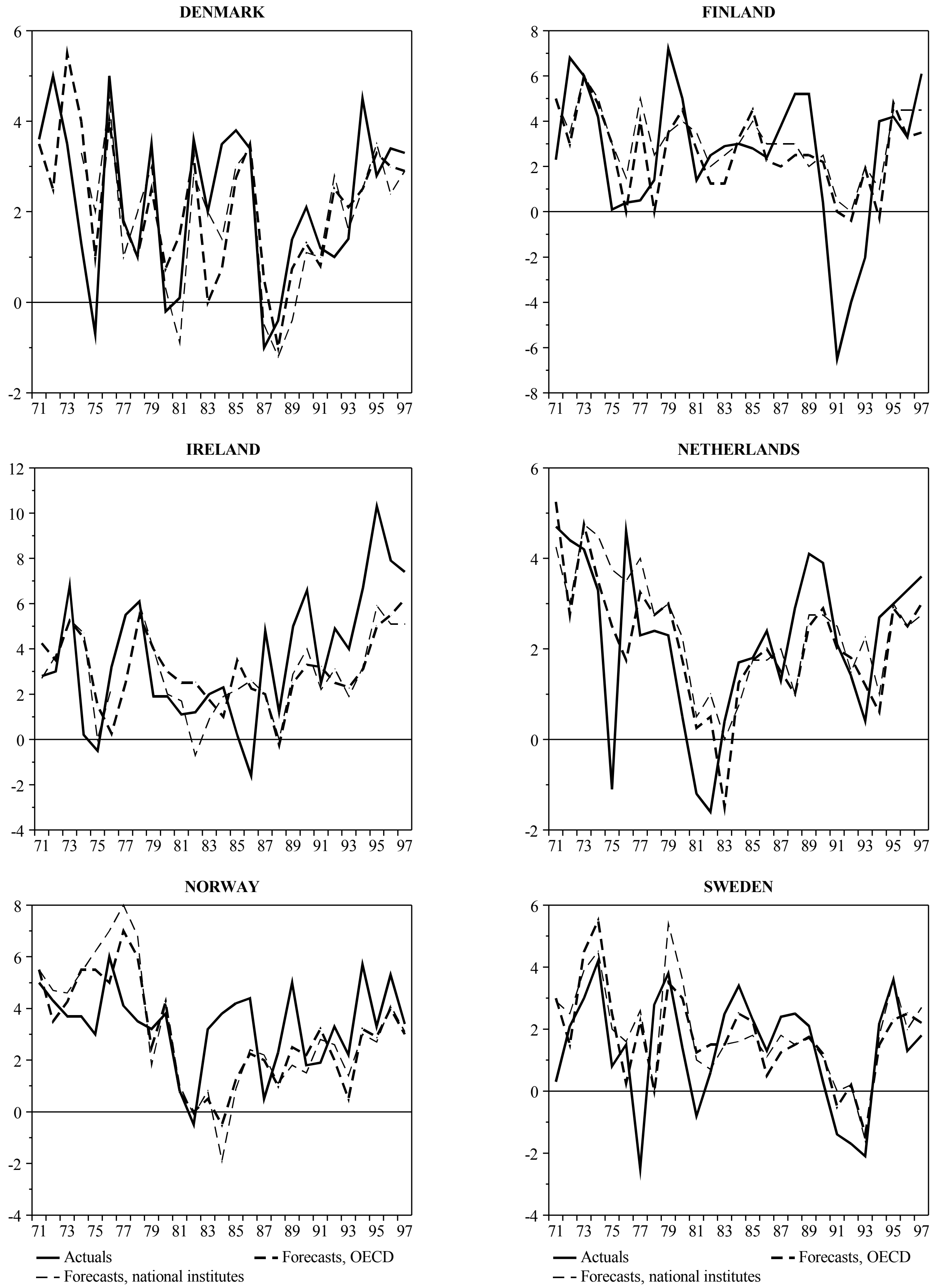

Exhibit 1a continuing

Growth forecasts and outcomes, OECD and national institutes

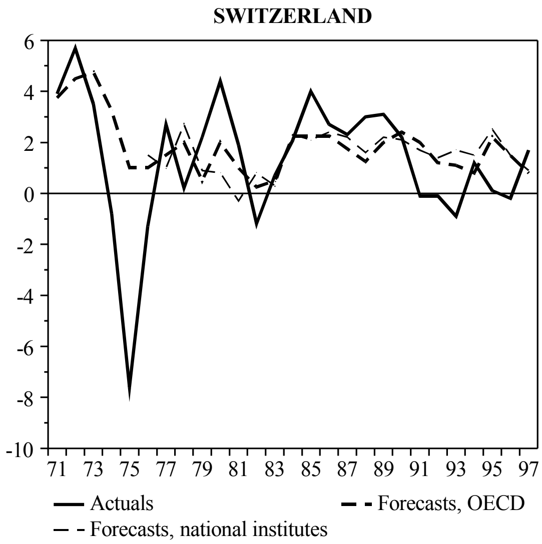

Exhibit 1b

Inflation forecasts and outcomes, OECD and national institutes

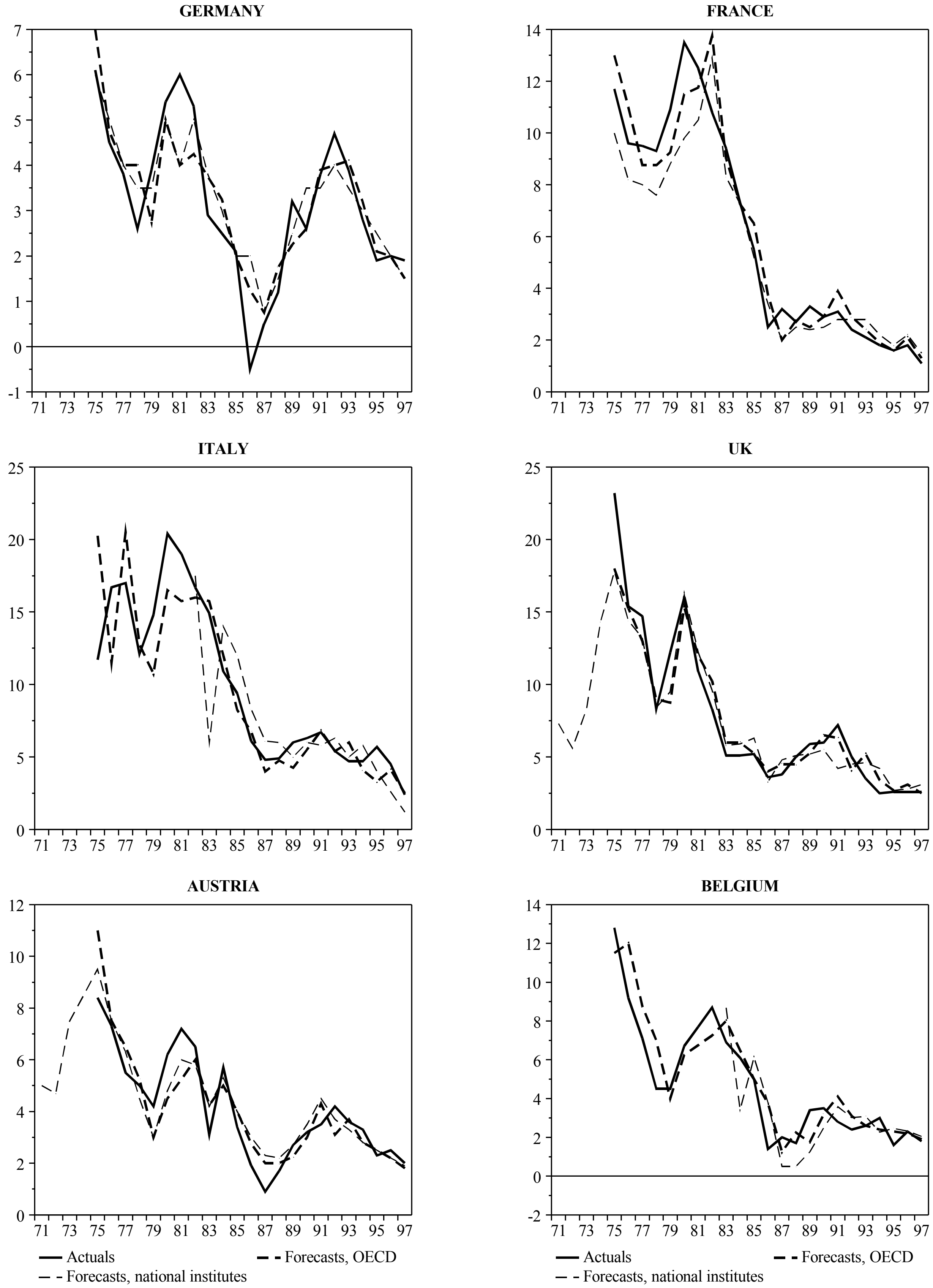

Exhibit 1b continuing

Inflation forecasts and outcomes, OECD and national institutes

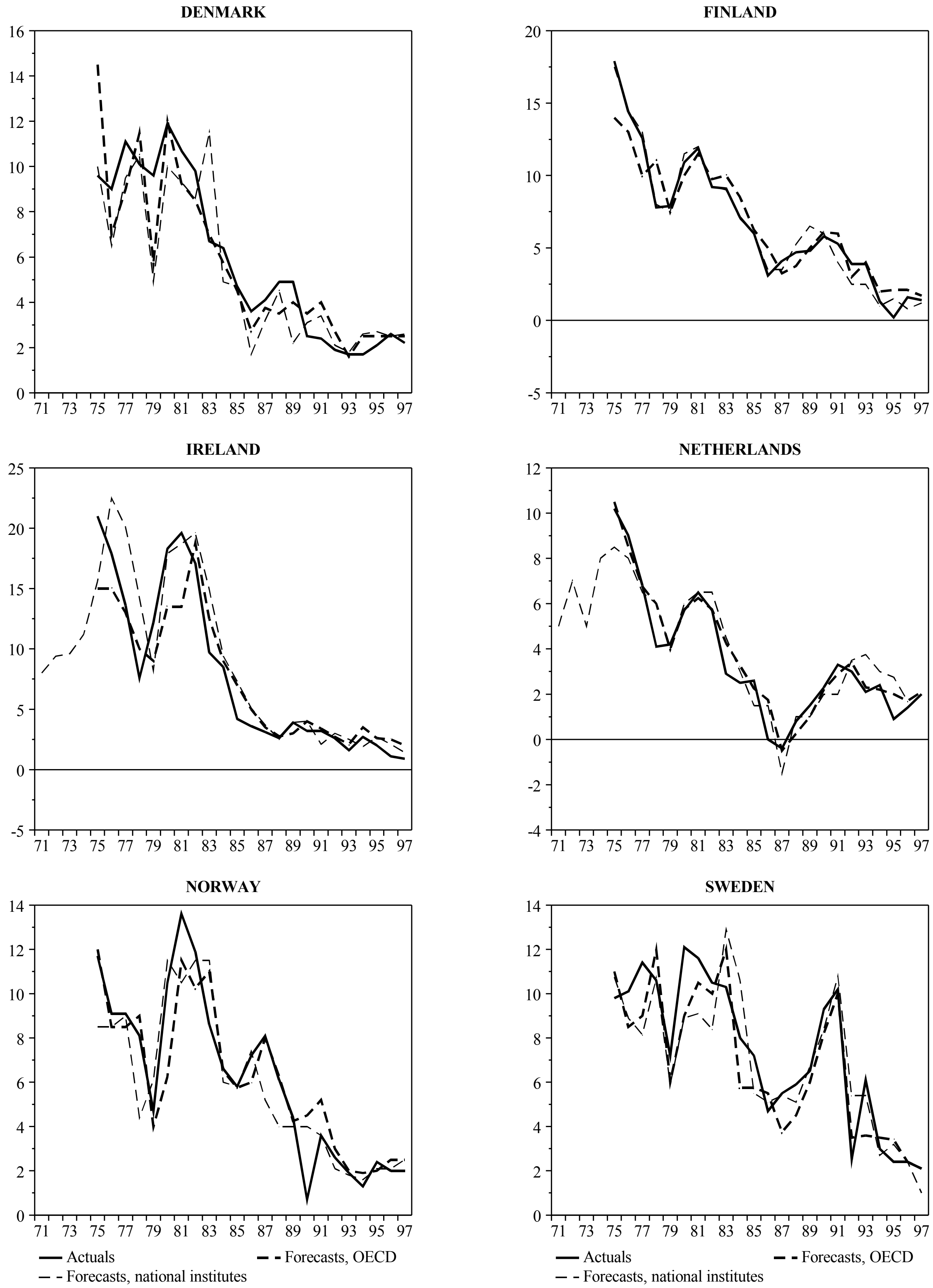

Exhibit 1b continuing

Inflation forecasts and outcomes, OECD and national institutes

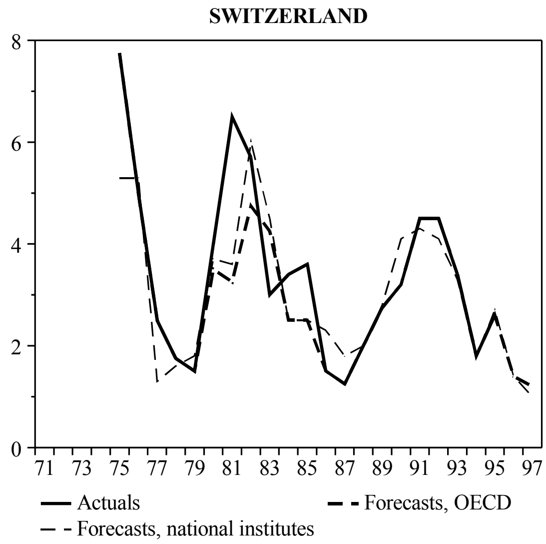

Exhibit 2a. Root Mean Squared Error (RMSE) of output growth forecasts, standardised and tested against naive projections

	1971 -1997		1980 –1997		1985 – 1997	
PANEL A	*RMSE*	*RMSE/SD*	*RMSE*	*RMSE/SD*	*RMSE*	*RMSE/SD*
GERMANY (D)**Δ**	1.72	0.86 †	1.31	0.83	1.31	0.93
FRANCE (F)**Δ**	1.32	0.74*†	1.09	0.97	1.16	0.98
ITALY (I)	2.02	0.94†	1.13	0.72	1.04	0.83
UK (GB)**Δ**	1.56	0.75*†	1.37	0.63	1.41	0.73
AUSTRIA (A)	1.85	0.93 †	1.31	0.97	1.45	1.04
BELGIUM (B)	1.60	0.84 †	1.38	0.84	1.44	0.92
DENMARK (DK)	1.35	0.78*†	1.22	0.74	0.97	0.59
FINLAND (FIN)**Δ**	2.62	0.83 †	2.65	0.79	3.04	0.79
IRELAND (IR) **Δ**	2.48	0.86 †	2.55	0.82	2.91	0.87
NETH (NL)**Δ**	1.51	0.79*†	1.40	0.75	1.34	0.91
NORWAY (N)	1.84	1.19	1.89	1.12	1.71	1.10
SWEDEN (S))**Δ**	1.59	0.87*†	1.10	0.64	0.98	0.54
SWITZ (CH)**Δ**	2.47	0.96*	1.80	1.09	1.95	1.24
Aver., OECD	1.89	0.88*†	1.63	0.85	1.72	0.90
	(2.09)	(0.96)	(1.91)	(0.95)	(1.97)	(1.00)
PANEL B						
IFO (D)	1.39	0.69*†	1.18	0.75	1.13	0.80
INSEE (F)	1.57	0.88*	1.30	1.16	1.34	1.14
ISCO (I)	NA	NA	NA	NA	0.87	0.70*
NIESR (GB)	1.71	0.83*†	1.49	0.69	1.44	0.75
WIFO (A)	1.76	0.89 †	1.19	0.88	1.29	0.93
FPB (B)	NA	NA	NA	NA	1.37	0.87*
DEC (DK)	NA	NA	1.08	0.65	1.06	0.64*†
MIFF (FIN)	2.64	0.83* †	2.63	0.79	3.02	0.78
ESRI (IR)	NA	NA	2.36	0.76	2.70	0.81*
CPB (NL)	1.58	0.83*†	1.39	0.74	1.27	0.86
MIFN (N)	2.13	1.37	2.10	1.24	1.77	1.14
NIER (S)	1.59	0.87*†	1.12	0.65	0.88	0.49
KOF (CH)	NA	NA	1.67	1.01	1.47	0.94
Aver., institutes	1.84	0.92*†	1.67	0.87	1.63	0.85
	(1.74)	(0.86)	(1.97)	(0.94)	(1.93)	(0.95)
Average, total	1.87	0.89*†	1.65	0.86	1.68	0.88
	(1.96)	(O.92)	(1.94)	(0.94)	(1.94)	(0.97)

Note: Denoting the forecast by P and the actual by A, RMSE and SD are calculated according to the formulas:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t} (P_t - A_t)^2}\,, \qquad SD = \sqrt{\frac{\sum_{t} (\bar{A} - A_t)^2}{n-1}}\;;$$

where Ã is the recursively calculated average growth and n is the number of observations. Forecasts that are significantly (5 %) better than naive are marked *, if compared to average growth, and by †, when naive is the current year forecast. Significant deviations in GDP growth from random walk with drift is denoted by Δ. NA means that data is not available. Figures in parentheses are comparisons to final outcomes.

Exhibit 2b. Root Mean Squared Errors *(RMSE)* of INFLATION forecasts, standardised and tested against naive projections

	1975 - 97		1980-97		1985 - 97	
PANEL A	*RMSE*	*RMSE/SD*	*RMSE*	*RMSE/SD*	*RMSE*	*RMSE/SD*
GERMANY (D) **Δ**	0.83	0.49*†	0.81	0.47	0.63	0.44
FRANCE (F) **Δ**	1.07	0.25*†	1.03	0.26	0.67	0.61
ITALY (I) **Δ**	2.73	0.49*†	1.53	0.28	1.04	0.65
UK (GB) **Δ**	1.57	0.29*†	0.89	0.26	0.74	0.48
AUSTRIA (A) **Δ**	1.00	0.49* †	0.87	0.49	0.61	0.66
BELGIUM (B) **Δ**	1.24	0.41* †	0.97	0.41	0.97	0.97
DENMARK (DK) **Δ**	1.65	0.46*†	0.85	0.26	0.82	0.66
FINLAND (FIN) **Δ**	1.47	0.32*†	0.92	0.28	0.93	0.49
IRELAND (IR) Δ	2.53	0.37*†	2.20	0.36	1.11	1.08
NETH (NL) **Δ**	0.72	0.26*†	0.66	0.35	0.65	0.58
NORWAY (N) **Δ**	1.54	0.40*†	1.70	0.44	1.28	0.52
SWEDEN (S) **Δ**	1.49	0.43*†	1.46	0.42	1.22	0.44
SWITZ (CH) **Δ**	1.38	0.72*†	1.19	0.65	0.94	0.54
Aver., OECD	1.58	0.43*†	1.24	0.40	0.92	0.65
	(1.64)	(0.45)	(1.26)	(0.40)	(0.90)	(0.61)
PANEL B						
IFO (D)	0.82	0.48*†	0.89	0.51	0.83	0.58
FM (F)	NA	NA	NA	NA	0.60	0.55*
ISCO (I)	NA	NA	NA	NA	0.89	0.56* †
NIESR (GB)	1.63	0.30*†	1.09	0.32	1.13	0.73
WIFO (A)	0.79	0.39 *†	0.78	0.44	0.66	0.71
FPB (B)	NA	NA	NA	NA	1.21	1.21
DEC (DK)	1.85	0.52*†	1.62	0.50	1.07	0.86
MIFF (FIN)	0.74	0.16*†	0.82	0.25	0.95	0.51
ESRI (IR)	2.98	0.44*†	1.71	0.28	1.16	1.13
CPB (NL)	1.06	0.39*†	1.00	0.53	1.05	0.94
MIFN (N)	1.75	0.46*†	1.57	0.40	1.45	0.59
NIER (S)	1.69	0.49*†	1.69	0.48	1.14	0.41
KOF (CH)	1.32	0.69*†	1.09	0.59	0.86	0.50
Aver., institutes	1.60	0.44*†	1.28	0.44	1.02	0.75
	(1.73)	(0.47)	(1.35)	(0.45)	(1.04)	(0.72)
Average, total	1.59	0.44*†	1.25	0.42	0.97	0.70
	(1.64)	(0.46)	(1.30)	(0.42)	(0.97)	(0.67)

Note: cf. Exhibit 2a

Exhibit 3. Testing for bias and autocorrelation in forecast errors

	BIAS		AUTOCORRELATION	
	1973 - 1997	1977 – 1997	1973 -1997	1977 -1997
PANEL A	Growth	Inflation	Growth	Inflation
GERMANY (D)	0.54	0.95	0.84	0.85
FRANCE (F)	0.27	0.93	0.73	0.62
ITALY (I)	0.87	0.35	0.19	0.71
UK (GB)	0.74	0.72	0.19	0.53
AUSTRIA (A)	0.27	0.64	0.49	0.38
BELGIUM (B)	0.48	0.57	0.96	0.79
DENMARK (DK)	0.95	0.30	0.11	0.98
FINLAND (FIN)	0.79	0.30	0.45	0.79
IRELAND (IR)	0.35	0.94	0.16	0.11
NETH (NL)	0.71	0.08	0.76	0.61
NORWAY (N) *	0.48	0.99	0.15	0.43
SWEDEN (S)	0.30	**0.03**	0.73	0.47
SWITZ (CH)	0.29	0.37	0.13	0.30
PANEL B				
IFO (D)	0.70	0.75	0.62	0.93
INSEE /FM (F) *	**0.05**	0.52	0.95	0.24
ISCO (I)	NA	NA	NA	NA
NIESR (GB)	0.47	0.79	0.07	0.40
WIFO (A)	0.32	0.93	0.26	0.12
FPB (B)	NA	NA	NA	NA
DEC (DK)	NA	0.41	NA	0.90
MIFF (FIN)	0.26	0.79	0.15	0.19
ESRI (IR)	NA	0.13	NA	0.18
CPB (NL)	0.59	0.13	0.56	0.87
MIFN (N)	0.66	0.17	**0.03**	0.62
NIER (S)	0.28	0.42	0.75	0.36
KOF (CH)	NA	0.52	NA	0.88

Note: Probabilities (bold = significant, 5%) in Student's t and Fisher's F test of the null hypothesis: H_{01}: $\beta_1 = 0$, H_{02}: $\beta_2 = \beta_3 = 0$, respectively in : $e_t = \beta_0 + \beta_1 e_{t-1} + \beta_2 e_{t-2} + u_t$. An asterisk signals rejection (5%) 0f the joint hypothesis: H_{01}: $\alpha = 0$, H_{02}: $\beta = 1$ in $A_t = \alpha + \beta P_t + \varepsilon_t$:

Exhibit 4A. Years when the acceleration / deceleration of GROWTH was missed and the ratio between these cases and naive forecast misses

PANEL A	Years missed 1971 – 1997	Ratio and 2 2 contingency table test results.
GERMANY (D)	72, 75, 82, 85, 86, 87, 88, 89, 90, 95, 96	11/11 (1.00)
FRANCE (F)	72, 79, 84, 85, 88, 89, 95	7/10 (0.70) *
ITALY (I)	74, 76, 78, 82, 88, 92, 96	7/15 (0.47) *
UK (GB)	75, 77, 81, 96	4/12 (0.33) *
AUSTRIA (A)	72, 73, 87, 90, 95	5/14 (0.36) *
BELGIUM (B)	72, 73, 78, 79, 80, 83, 86, 88 , 92, 93, 94, 95, 96	13/13 (1.00)
DENMARK (DK)	73, 78, 84, 85, 86, 88, 90, 92, 96, 97	10/13 (0.77)
FINLAND (FIN)	73, 82, 85, 88, 89	5/19 (0.26) *
IRELAND (IR)	77, 81, 85, 86, 90, 95	6/14 (0.43) *
NETH (NL)	73, 75, 79, 81, 82, 83, 84, 86, 87, 88, 89 95, 96	13/11 (1.18)
NORWAY (N)	73, 74, 75, 77, 78, 84, 85, 86, 88, 92, 96	11/11 (1.00)
SWEDEN (S)	77, 87, 88, 92, 95	5/13 (0.38) *
SWIT (CH)	72, 73, 79, 85, 88, 89, 92, 93 , 95 , 96	10/12 (0.83)
PANEL B		
IFO (D)	72, 75, 82, 86, 88, 89, 90, 95, 96	9/13 (0.69)
INSEE/ FM (F)	72, 83, 85, 89, 93, 95	6/12 (0.50) *
ISCO (I) (85 -97)	87	1/6 (0.16)
NIESR (GB)	75, 77, 79, 81, 87	5/15 (0.33) *
WIFO (A)	72, 83, 88, 90, 95	5/17 (0.29) *
FPB (B) (85 - 97)	86, 88, 92, 95	4/6 (0.66)
DEC (DK) (80 - 97)	81, 83, 84, 85, 86, 88, 89, 92, 93, 96, 97	11/10 (1.1)
MIFF (FIN)	73, 82, 88, 89	4/17 (0.24) *
ESRI (IR) (80 - 97)	80, 81, 82, 86, 90	5/16 (0.31)*
CPB (NL)	73, 75, 77, 79, 82, 86, 87, 88, 93, 95, 96	11/10 (1.01)
MIFN (N)	73, 74, 75, 77, 78, 84, 85, 86, 88, 92, 96	11/11 (1.10)
NIER (S)	77, 84, 88, 92, 93	5/17 (0.29) *
KOF (CH) (80 -97)	85, 88, 89, 92, 93, 95	6/6 (1.00)

Note: Border cases where the forecast expresses no change in growth are classified as correct if the change in the outcome is smaller than one half percentage point. Analogously, a forecast is considered correct if it predicts an acceleration/deceleration of less than one-half a percentage point in cases where the outcome is no change in growth. Remaining cases are registered as wrong. A * means significant on the 5 % level of the contingency test in Appendix B.

Exhibit 4B. Years when the acceleration / deceleration of INFLATION was missed and the ratio between these cases and naive forecast misses

PANEL A	Years missed 1975 – 1997	Ratio and 2 2 contingency Table test results.
GERMANY (D)	78, 79, 81 , 84	4/8 (0.50)*
FRANCE (F)	79, 82, 87, 89	4/10 (0.40) *
ITALY (I)	76, 79,88, 89, 90, 93, 95	7/7 (1.00)
UK (GB)	91, 94	2/7 (0.29) *
AUSTRIA (A)	81, 87, 92, 96	4/7 (0.57) *
BELGIUM (B)	79, 82, 87, 88 , 90, 91, 94	7/9 (0.78)*
DENMARK (DK)	77, 78, 88, 89, 91, 92, 94, 97	8/7 (1.14)
FINLAND (FIN)	81, 83, 84, 87, 88, 89, 90, 93, 95, 97	10/3 (3.33)
IRELAND (IR)	81, 90, 95, 96, 97	5/5 (1.00)
NETH (NL)	79, 81, 84, 85	4/9 (0.44)*
NORWAY (N)	77, 96	2/9 (0.22)*
SWEDEN (S)	76,77, 78, 83, 87,89, 91, 95	8/8 (1.00)
SWIT (CH)	78, 81, 84, 85, 90, 91, 97	7/7 (1.00)
PANEL B		
IFO (D)	81, 86, 90	3/8 (0.38)*
INSEE /FM (F)	79, 80, 87, 89, 91, 95	6/10 (0.60)*
ISCO (I) (83 - 97)	93, 94, 95	3/2 (1.50)
NIESR (GB)	91, 94	2/6 (0.33) *
WIFO (A)	81, 87, 96, 97	4/7 (0.57) *
FPB (B) (85 – 97)	87, 88, 93, 94	4/6 (0.67)
DEC (DK)	79, 80, 82, 88, 89 , 93, 94, 97	8/6 (1.33)
MIFF (FIN)	76, 77, 79, 84, 87, 96	6/8 (0.75) *
ESRI (IR)	76, 77, 82, 90, 91, 92	6/6 (1.00)
CPB (NL)	79, 81, 85, 89, 91, 92, 93, 95	8/8 (1.00)
MIFN (N)	77, 87, 91, 96	4/9 (0.44)*
NIER (S)	76, 77, 78, 83, 84, 87, 95, 97	8/7 (1.14)
KOF (CH)	81, 84, 85, 91, 97	5/8 (0.63)*

Note: c.f. Exhibit 4a.

Exhibit 5. Aggregated (OECD and institutes) acc/dec errror frequencies.

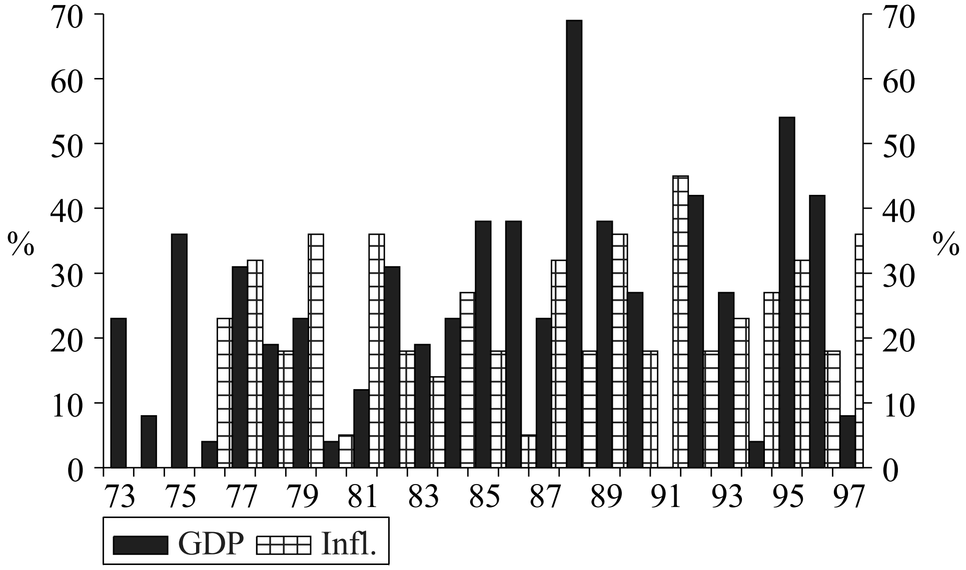

Exhibit 6. Root mean squared revisions (*RMSR*) and bias of revisions

	Output growth		Inflation	
	RMSR	*Bias* t – values	*RMSR*	*Bias* t – values
Germany (D)	0.65	0.62	0.21	0.09
France (F)	0.86	1.30	0.25	**2.48***
Italy (I)	1.22	**2.01***	1.16	**2.10***
UK (GB)	1.01	**2.66***	0.60	**3.28***
Austria (A)	0.72	0.48	0.42	**2.04***
Belgium (B)	0.87	1.53	0.49	0.63
Denmark (DK)	0.85	0.40	0.71	0.67
Finland (FIN)	0.52	**2.27***	0.45	0.87
Ireland (IR)	3.01	**2.66***	1.07	1.09
Ned (NL)	0.68	1.75	0.39	0.59
Norway (N)	1.28	**2.10***	0.91	0.20
Sweden (S)	0.71	**2.21***	0.62	**2.58***
Switz (CH)	0.88	0.22	0.39	0.08

Notes: Average *RMSR*: 1.18 for GDP and 0.60 for inflation. Revision = final – preliminary outcome. The column ”Bias” shows t-test values of the arithmetic average of revisions, where * denotes 5% significance.

References

Artis, M.J. (1996). How accurate are the IMF's short-term forecasts? Another Examination of the World Economic Outlook, *Staff Studies of the World Economic Outlook,* IMF.

Ash, J.C.K., Smyth, D.J. and Heravi, S.M. (1998). Are OECD forecasts rational and useful? A directional analysis, *International Journal of Forecasting, 14,* 381-391.

Bjønnes. G.H., Isachsen, A.J. and Stoknes, S.O. (1998). The big guessing competition. The accuracy of macroeconomic forecasts (in Norwegian), *Økonomiske Analyser*, 9, 34-41.

Brown, B.W. and Maital, C. (1981). What do economists know? An empirical study of experts' expectations, *Econometrica*, 49, no. 2, 491-504.

Clements, M.P. and Hendry, D.F. (1998). *Forecasting Economic Time Series*. Cambridge University Press, Cambridge.

Denton, F.T. and Oksanen, E.H. (1972). A multi-country analysis of the effects of data revisions on an econometric model, *Journal of the American Statistical Association*, (67), 286-291.

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

Diebold, F.X., Tsay A.S. and Wallis, K.F. (1997). Evaluating density forecasts of inflation, the success of professional forecasters, ESRC Macroeconomic Modelling Bureau Discussion Paper 48, University of Warwick, UK.

Donders, J. and Kranendonk, H. (1999). The accuracy of CPB forecasts, *Quarterly Review of CPB*, 2/99, 24-29.

Fildes, R. and Stekler, H.O. (1999). The state of macroeconomic forecasting, Discussion Paper No. 99-04, Lancaster University, UK.

Granger, C.W.J. (1996). Can we improve the perceived quality of economic forecasts? *Journal of Applied Econometrics,* 11, 455-473.

Holden, K. and Peel, D.A. (1985). An evaluation of quarterly National Institute forecasts, *Journal of Forecasting* (4), 227-234.

Holden, K., Peel, D.A. and Thompson, J.L. (1994). *Economic Forecasting: An Introduction,* Third Ed., Cambridge: Cambridge University Press.

Keereman, F. (1999). The track record of Commission forecasts, Economic Papers # 137, European Commission Directorate-General for Economic and Financial Affairs.

Kennedy, M.C. (1969). How well does the National Institute forecast? *National Institute Economic Review*, (3?), 40-52.

Kirschgassner, 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* , Free Press, New York.

Leitch, G. and Tanner, J.E. (1995). Professional economic forecasts: are they worth their costs? *Journal of Forecasting*, 14, 143-157.

McNees, S.K. (1986). Forecasting accuracy of alternative techniques: a comparison of U.S. macroeconomic forecasts, *Journal of Business & Economic Statistics*, (4), 5-15.

McNees, S.K. (1978). The "rationality" of economic forecasts, *American Economic Review*, 68 no. 2, 301-305.

McNees, S.K. (1989). Forecasts and actuals: The trade-off between timeliness and accuracy, *International Journal of Forecasting,* 5, 409-416.

McNees, S.K. (1992). How large are economic forecast errors? *New England Economic Review,* July/August, 25-42.

Milburn, M.A. (1978). Sources of bias in predicting future events. *Organisational Behavior and Human Performance*, 21, 17-26.

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 Institute, *International Journal of Forecasting,* 15, 247-257.

Morgenstern, O. (1950). *On the Accuracy of Economic Observations*, Second Ed., Princeton University Press, Princeton.

OECD Economic Outlook, 1971-1998, OECD, Paris.

Öller, L.-E. and Barot, B., (1998). How reliable are NIER's forecasts? A European comparison, (in Swedish), *Konjunkturläget,* March, 113-135.

- " - (1999). Comparing the accuracy of European GDP forecasts, Working Paper no. 64, National Institute of Economic Research, Sweden.

Stekler, H.O. (1994). Are economic forecasts valuable? *Journal of Forecasting* 13, 495-505.

Theil, H. (1966). *Applied Economic Forecasting*, Amsterdam: North-Holland Publishing Co.

Thury, G. (1986). Macroeconomic forecasting in Austria – an agnostic view, *Empirica* (20), 173-186.

Varian, H.R. (1975). A Bayesian approach to real estate assessment, in: S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, North-Holland, 195-208.

Wallis, K.F.(1989). Macroeconomic forecasting: A survey, *Economic Journal*, 99, 28-61.

Zarnowitz, V. (1992). Business cycles: theory, history, indicators and forecasting, *NBER Studies in Business Cycles* (27), University of Chicago Press, Chicago.

Zellner, A. (1986). Biased predictors, rationality and the evaluation of forecasts, *Economics Letters* (21), 45-48.