JOSEF BAUMGARTNER

EVALUATION OF MACRO-ECONOMIC FORECASTS FOR AUSTRIA IN THE 1980S AND 1990S

The forecasts of the Austrian Institute of Economic Research (WIFO), the Institute for Advanced Studies (IHS) and the OECD for the Austrian economy have been evaluated for three key macro-economic variables (GDP growth, unemployment rate, and rate of inflation) for the period from 1983 to 1999. In terms of accuracy, no significant differences have emerged between the three institutions, as far as the projections of growth and inflation are concerned. However, the prospects for unemployment are more precisely assessed by the two Austrian institutes. Compared with previous studies, forecasting errors exhibit a slight downward trend. The forecasts by the three institutions are largely unbiased and efficient, and – with the exception of the unemployment rate forecast by the OECD – clearly superior to "naïve" forecasting strategies.

Josef Baumgartner is an economist at WIFO. The author is grateful to Helmut Kramer, Marcus Scheiblecker, Thomas Url and Michael Wüger for useful and constructive comments • The data were processed and analysed with the assistance of Christine Kaufmann and Martha Steiner •
E-mail addresses: Josef.Baumgartner@wifo.ac.at, Christine.Kaufmann@wifo.ac.at

Any forward-looking action under uncertainty by economic policy, firms, consumers, employers or employees is based on expectations and thereby, implicitly or explicitly, upon a forecast of economic developments. The reaction of economic agents to changes in future prospects can be rather different. Thus, the prediction of a business cycle downturn may induce private agents to a more cautious consumption and investment behaviour, and wage settlements may turn out more moderate. The reaction by the government will depend on the goals of economic policy. From the range of options, just a few shall be mentioned here: deliberate counter-cyclical action or – less pro-active – the operation of automatic stabilisers will drive up public expenditure and, with tax revenues falling at the same time, the budget deficit will increase. This will, by tendency, lean against the downward trend. If, on the other hand, fiscal policy wants to keep the government balance unchanged, it will try to compensate for the revenue losses by cutting expenditure. In this way, the downturn will be reinforced, at least in the short run[1]. Provided that the reactions by the different economic agents do not offset each other, the example illustrates the effect a (credible) forecast may have: depending on the reaction of economic agents, the forecast may be either self-fulfilling (or self-reinforcing) or self-destroying.

This renders the evaluation of the accuracy of a forecast difficult. The most straightforward way to proceed would be to compare the forecast with the actual economic outcome (the realisation) and to regard the forecast as good if the degree of congruence is high. In case that the forecast has no effect on the outcome – like the weather not being influenced by the weather forecast – such an approach is perfectly adequate. However, the situation is somewhat different for economic forecasts: in the extreme case, the forecast may have such a strong influence on economic agents that the outcome is based on entirely different conditions and the ex-post comparison would show little or no congruence between forecast and reality. Still, such a forecast should not be called bad or meaningless, as the alternative course of events could not have been anticipated[2].

[1] A general discussion of whether a particular policy measure of cyclical stabilisation is meaningful and adequate in a small open economy would go far beyond the scope of this paper. A few issues in this discussion shall nevertheless be recalled here: effectiveness of monetary and fiscal policy with fixed or flexible exchange rates and the appropriate policy mix, time inconsistency and policy ineffectiveness, decision, transmission and impact lags and the related question, whether a particular measure will have a counter- or rather pro-cyclical effect.

[2] If one were to anticipate the behaviour and reaction of economic agents to the forecast in the forecast itself, the result of the forecast would be different, leading again to reactions which themselves would have to be taken into account, etc. The result would be an infinite regression. The problem can only be solved under (highly) restrictive assumptions, using a forecasting model that adequately incorporates the formation of economic agents' expectations and explicitly contains policy reaction functions. In actual forecasting, such models

Moreover, in judging the quality of a forecast one has to bear in mind that we are usually dealing with conditional forecasts that are valid only on certain assumptions[3]. For a small open economy like Austria such assumptions usually refer to international economic developments (economic trends in the EU and the USA, exchange rates, prices of raw materials, etc.) as well as to policy conditions in the EU and in Austria (such as policy-controlled interest rates, tax rates, etc.). The very assessment of international business conditions is one of the major sources of uncertainty for forecasts in small countries.

The result of a forecast also depends on the (implicit or explicit) structure of the underlying macro-economic explanatory context (econometric model). If the forecast is validated by actual developments, although the underlying assumptions were false, the forecast ought to be regarded as incorrect even if the forecasting error, defined as realisation minus forecast, is zero. Conversely, an incorrect forecast derived from correct assumptions points to deficiencies in the underlying explanatory model.

Furthermore, the overall purpose of a forecast is not clearly defined, as different economic policymakers and forecasters may be guided by different interests. This problem area is reflected by the selection of a "loss function", i.e., the evaluation criterion for deviations of the forecast from realisations. For "rational" or "pure" forecasters it is important to produce a largely "correct" forecast coming as close as possible to the actual value of the projected variable. Upward and downward errors are given equal weight, implying a symmetric loss function.

However, for economic policymakers and their advisers upward or downward deviations may weigh differently. Thus, it may be that the non-recognition of an economic downturn or even more a recession carries a higher economic and political cost for a government than the incorrect perception of an upturn. If policy wants to react to a greater extent to a projected slower growth than to an announced higher rate of growth, then cautious economic forecasts will possibly be more appreciated than unbiased ones (*Aiginger,* 1979). This would point to an asymmetric loss function carrying a higher "sanction" for the over-estimation of the actual outcome (i.e., a negative forecasting error)[4].

For monetary policy mainly oriented towards price stability also an asymmetric loss function is plausible. However, with regard to real GDP (or potential output) it would be skewed rather to the opposite side of the one referred to above: the under-estimation of real GDP growth (and of inflation) would weigh more heavily in the loss function and carry a higher "sanction".

In the situations described, the asymmetric loss functions of the different users of the forecasts would have to be taken as the measure in assessing the quality of the forecasts. However, as the examples illustrate, policymakers may have (very) different loss functions for each of the key macro-economic variables; thus, the debate would shift from the accuracy of the forecasts towards the shape of the appropriate loss function.

Moreover, little is known about the statistical properties of forecast errors evaluated on the basis of asymmetric loss functions, and comparisons with previous research on the subject are not straightforward[5]. On the evaluation of forecasts on the basis of symmetric loss functions, however, there is an extensive literature that also allows comparisons over time and across countries.

For these reasons, the measures of accuracy and statistical tests used in the present analysis assume a symmetric loss function. The arguments mentioned above underline that there is no absolute measure of forecast accuracy. It is nevertheless meaningful to investigate the *relative* properties of forecasts over time, across different variables and between different authors of forecasts.

The following paper compares for the 1980s and 1990s the quality of the forecasts of the two leading Austrian economic research institutes – the Austrian Institute of Economic Research (WIFO) and the Institute for Advanced Studies (IHS) – with the OECD

---

are not available. Given that the users of forecasts will indeed react to the forecast, a perfect forecast is therefore logically impossible.

[3] An introductory discussion of unconditional forecasting methods is presented, e.g., by *Diebold* (1998, 2001).

[4] Giving preference to a cautious forecast over the most likely outcome is in line with some standard rules of business management for dealing with uncertainty (e.g., the minimax principle) as well as with accounting rules (*Aiginger*, 1979). This points to a similar asymmetric loss function also for firms with regard to GDP forecast.

[5] See *Christoffersen – Diebold* (1996, 1997).

forecasts for Austria. The present analysis covers three key macroeconomic indicators and tries to find out whether the forecast errors exhibit any systematic bias. In addition, the differences in forecasting accuracy between WIFO and IHS on the one hand and the OECD on the other, are tested for their statistical significance[6].

Several investigations of the forecasting performance of WIFO and IHS have been carried out in the past (see *Baumgartner,* 2002, for a review), with some studies also including the OECD forecasts for Austria (*Kramer,* 1980, *Wörz,* 1994, 1995, *Öller – Barot,* 2000). The quality of the OECD forecasts has been the subject of a great number of studies, particularly for the G-7 countries (e.g., *Ash – Smyth – Heravi,* 1997, 1998, *Koutsogeorgopoulou,* 2000, *Öller – Barot,* 2000, *Batchelor,* 2001, *Blix et al.,* 2001).

With regard to comparisons of its own projections with forecasts from the "private sector", the OECD has repeatedly pointed out that the OECD projections are serving other purposes, namely "to discuss risks and deliver policy messages that are deemed to be useful to policymakers in member countries" (*Lenain,* 2002). However, this statement is valid to at least an equal extent for the forecasts of the two Austrian economic research institutes: thus, the institutes' forecasts usually do not include projections of policy reactions, but rather formulate recommendations for policy on the basis of the given economic policy framework.

The data

The analysis covers the annual forecasts of WIFO, IHS and the OECD for the Austrian economy from 1983 to 1999 for three widely watched macro-economic variables:

- the growth rate of real GDP,
- the rate of unemployment (using the national definition for WIFO and IHS, and the OECD definition for the OECD forecast)[7],
- the rate of inflation based on the private consumption deflator.

The two Austrian economic research institutes present macro-economic forecasts four times a year (in March, June, September and December), each time for the current and the following year[8]. The OECD, in its semi-annual "Economic Outlook" issued in June and December, submits economic projections for the current and the next year[9]. The OECD holds the view that for the purpose of comparisons with other forecasts, the "preliminary version" of its "Economic Outlook" should be used that is finalised in May and November of each year (see *Lenain,* 2002). As these dates come to fall regularly between the forecasting dates of WIFO and IHS, the choice of the forecast issues from the national institutes for the present analysis has the following implications:

- When using the March and September issues (Tables 1 to 4), the OECD, knowing the institutes' forecasts, has an information advantage. If, in this case, the forecasts from WIFO and IHS prove more accurate than those from the OECD, they have to be regarded as clearly better, since they have been made on the basis of a comparatively smaller information set.
- When, alternatively, using the June and December issues (Tables 5 to 8), it is the Austrian institutes that have an information advantage, since their set of information includes the most recent OECD forecast. Should the OECD forecast turn out more accurate, it is then of better quality, as in this case the OECD is in the less favourable starting position, as far as the information set is concerned.

Thus, in judging the quality of the different forecasts, both alternatives have been used: from the autumn forecasts (September and December for WIFO and IHS) of year $t-1$

[6] The present analysis compares the forecasts of the Austrian economic research institutes (WIFO and IHS) with those of the OECD. The differences between WIFO and IHS have been examined in *Baumgartner* (2002) and are not discussed here.

[7] Due to a change in the definition of the OECD unemployment rate in 1996 (and a further one in 1999), the assessment period for the quality of the unemployment projections has been restricted (for all three forecasting institutions) to the years from 1983 to 1995.

[8] Since 1981, WIFO publishes forecasts for the next year as from the June projections, and since 1988 as from March. Since 1988, WIFO's December projections include main results and since 1994 a full-fledged two years ahead forecast. Since the beginning of its regular forecasting activity in 1974 the IHS has regularly submitted projections for the current and the next year (see *Fürst*, 1980).

[9] Three key dates have to be kept in mind: "a) the cut off rate for information used in the projections (normally April and October) with respect to new economic data and economic policy announcements; b) the release of a 'Preliminary Version' of the Economic Outlook (in May and November); and c) the publication of the printed version of the Economic Outlook (in June and December)" (see *Lenain,* 2002).

and from the respective spring forecasts (March and June for WIFO and IHS) of year t the projected values *for the current year t* have been taken[10].

For growth of GDP and the private consumption deflator, the realisation for year t, against which the forecasts are set, are the preliminary national accounts data calculated by WIFO, disposable in March of the following year[11]. This choice most closely corresponds to the state of information at the disposal of the forecasters at the time of their forecast. Later data revisions carried out by Statistics Austria often go back over several years and sometimes also include methodological changes, such as the changeover of the national accounts standards from SNA 68 to ESA 79 and ESA 95 (see in this regard also *Baumgartner,* 2002). For the rate of unemployment, the realisations taken for comparison with the WIFO and IHS projections (based on the national definition) are those given by the labour market service in March for the preceding year; for the OECD forecasts, the realisations were obtained from the "Economic Outlook" of the December issue of the following year.

Methods

The accuracy of the forecasts is assessed on the basis of different criteria which may be divided into five groups[12]:

1. measures of statistical accuracy with regard to the forecasting error,
2. comparison with "naïve" forecasts (Theil coefficient of inequality),
3. tests for statistical significance of the differences between the two institutes' forecasts,
4. sign tests in order to check the directional forecast accuracy,
5. statistical tests for the unbiasedness and efficiency of the forecasts.

Usually, the 95 percent level of significance is applied for the statistical tests used. This corresponds to a 5 percent probability of error for the underlying null hypothesis being rejected although being correct. In the tables, the error probability is given by the p value: a p value of 0.05 corresponds to an error probability of 5 percent, meaning that the null hypothesis is rejected (at least at the 95 percent level) if p is smaller than 0.05.

Measures of accuracy

The mean forecast error shows the average deviation of the projected from actual values, pointing to a forecast bias. If the forecast errors – defined as actual minus projected values – are distributed equally upwards and downwards, this figure is (close to) zero. A positive sign indicates that the forecast has a tendency of underestimating actual developments. Whether the forecast is actually biased is found out using the test procedures described below. An "optimal" predictor should, moreover, exhibit a lower variance than the realisations, since the forecast, unlike the actual value, includes no irregular component (see *Granger – Newbold,* 1977). In this regard, *Aiginger* (1979, p. 173) also points to a revealed "smoothing tendency": if information is scarce, it seems plausible to expect an average development. The more the future trend is uncertain, the more expectations will keep close to the arithmetic mean. The influence of a smoothing tendency is supported by the fact that it is stronger with forecasts of longer horizon than with short-term ones. The ratio between the standard deviations of the forecasts and the realisation should therefore be smaller than 1 and diminish with increasing forecast horizon.

In calculating the mean forecast error, the positive forecast errors are balanced against the negative ones. Statements going beyond the tendency of the forecast are therefore impossible to make. The mean absolute error (MAE; for all definitions see the methodological annex) and the mean squared error (MSE) can also measure the accuracy of a forecast. Hereby, the absolute values (MAE) – the squares, respectively (MSE) – of the forecast errors are added up and the average is taken. The two measures differ by the weight attributed to the forecast errors entering the calculation: for the MAE, they are weighted linearly, for the MSE by the power of 2; the latter measure therefore gives greater emphasis to large forecast errors. Frequently, instead of the MSE, the root mean square error ($RMSE$) is shown. This has the advantage that the calculated statistic is of the same dimension as the underlying variable. For all accuracy measures discussed holds: the smaller the value of the calculated statistics, the better the forecast.

[10] In June 1997, the two Austrian institutes did not submit forecasts. In this case, for the purpose of the present evaluation, the latest state of information available at that time, i.e., the forecast of March 1997, has been used.

[11] Alternatively, the realisations have been taken from Statistics Austria's first publication of the National Accounts in autumn of the following year. The findings hardly differ from the ones described here and are therefore not presented.

[12] For a definition of the assessment criteria see the methodological annex.

Furthermore the mean squared error can furthermore be split in two kinds of inequality proportions, in order to give further evidence on the accuracy of the forecast. The following relations hold:

$UM + US + UC = 1$,

$UM + UR + UD = 1$.

A good forecast is characterised by small bias (UM), variance (US), and regression (UR) proportions, while the co-variance (UC) or the distribution (UD) proportions should be close to 1 (see *Theil,* 1966, 1971, or the methodological annex).

Theil coefficients of inequality

No information on the difficulty of predicting a particular variable enters the calculation of the mean absolute or mean squared error. Yet, variables with a low degree of variation are easier to predict than those subject to wide variations. To this end, the $RMSE$ standardised with the standard deviation of the realisations SD_R is calculated $\left(\frac{RMSE}{SD_R}\right)$. This measure allows a better comparison of forecasting errors between different variables.

Theil (1966, 1971) developed several measures that also address the problem referred to above; these statistics are standardised to 1 and are therefore easier to interpret (see below). In general, they compare the $RMSE$s of different forecasts. In our case, the forecasts of the three research institutions are set against two "naïve" forecasting strategies suggested by Theil.

The statistic Theil W sets as "naïve" forecast the hypothesis of "no change in the rate of change": the last known rate of change of the realisation is used for projecting future changes. For example, the "naïve" forecast made in year t would thus assume that real economic growth in the current as well as in the subsequent year equals the rate of last year (known at the time of the forecast).

The "naïve" forecast according to the Theil U criterion hypothesises a "no change in level" situation. E.g., it would thereby assume a rate of real economic growth of zero for both the current and the following year.

For both measures holds that the mean squared errors of the institutions' forecasts are set against the mean squared errors of the "naïve" forecasts. If the Theil W or Theil U statistics are smaller than 1, the forecasts of the institutions are superior to the "naïve" forecasts. A good forecast therefore exhibits for the Theil inequality coefficients values *clearly* below unity.

Significance tests

The accuracy of two forecasts A and B is also examined using the modified Diebold-Mariano test (DM^*) and the non-parametric Wilcoxon signed rank test (W^*; see *Diebold – Mariano,* 1995, *Hartung,* 1991, *Harvey – Leybourne – Newbold,* 1997, *Mariano,* 2002, and the methodological annex). *Diebold – Mariano* (1995) have developed, for different loss functions, a test for the null hypothesis of "equal forecast accuracy" between two forecast providers. The present analysis uses both the absolute and the squared errors as loss functions to examine whether statistically significant differences between the forecasts (or the measures of forecast accuracy) of WIFO, IHS and the OECD exist.

The loss functions used here belong to the family of symmetric loss functions. Particularly the users of a forecast (e.g., economic policymakers) may rate positive or negative forecast errors differently than the forecasters themselves (see the discussion further above). The test developed by *Diebold – Mariano* (1995) is applicable also to non-symmetric loss functions. However, a generally accepted asymmetric loss function is not available, at least not at the present stage.

In the tables the values for DM^*-AE and W^*-AE, or DM^*-SE and W^*-SE, respectively, show the p values of these tests for the absolute (AE) or squared (SE) loss function, respectively. For values below 0.05, the null hypothesis of equal forecasting accuracy can be rejected at the usual levels of significance, i.e., the differences are statistically significant.

Tests for directional accuracy (sign tests)

For the users of forecasts it is sometimes not the accuracy of a forecast that matters most, but whether at least the direction of future developments has been correctly assessed. Using a non-parametric test (see methodological annex) it is examined whether the direction of change given in a forecast corresponds with the actual course of events. The null hypothesis of the test is stated as follows: the sign of a change in the forecast and the sign of a change of the same variable as actually observed are mutually independent[13]. The ratio of congruence (ER, in the interval [0, 1]) specifies the degree of congruence over the period of analysis. If this ratio is below 0.5, than a forecast by tossing a coin would better predict the direction of actual change. Good forecasts show ratios of congruence close to 1, and the p value should be smaller than 0.05.

Unbiasedness and efficiency

Unbiased forecasts have the same mean as the realisations and thus exhibit a mean forecast error of zero. A standard procedure in this regard is the estimation of the "realisation-forecast" regression equation by *Mincer – Zarnowitz* (1969),

$$R_t = \alpha_0 + \alpha_1 P_t + u_t ,$$

where R_t denominates the realisation, P_t the forecast and u_t the error term. Under the hypothesis of unbiasedness the parameter restrictions $\alpha_0 = 0$ und $\alpha_1 = 1$ must be fulfilled simultaneously, and the residual term u_t corresponds to the forecast error. This hypothesis can be tested by a joint F test. However, *Holden – Peel* (1990) have shown that this criterion is a sufficient, but not a necessary condition for an unbiased forecast: it is (theoretically) possible to establish unbiased forecasts that with certainty do not withstand that test. They therefore suggest to estimate the equation

$$R_t - P_t = \mu + u_{1t}$$

and verify the null hypothesis of unbiasedness through the ordinary t test $\mu = 0$. For a p value above 0.05, the null hypothesis cannot be rejected (at the usual level of significance).

The hypothesis of efficiency of a forecast assumes that the forecasters make optimal use of all the information at their disposal. This assumption is verified by testing the residuals u_{1t} from the "realisation-forecast" regression according to *Holden – Peel* (1990) for first-order autocorrelation through the Durbin-Watson statistic (DW). In addition, the orthogonality between forecast errors and available information is checked via the regression

$$R_t - P_t = \beta_0 + \beta_1 \left(R_{t-1} - P_{t-1}\right) + u_{2t} .$$

If the forecasters learn from past errors $\left(R_{t-1} - P_{t-1}\right)$ and use this information to be derived from the forecast errors, then the joint hypothesis $\beta_0 = 0$ and $\beta_1 = 0$ should not be rejected on the usual level of significance. The p value for the F test should thus be greater than 0.05. The residuals u_{2t} are tested for higher-order autocorrelation using the Box-Ljung Q test. The null hypothesis of "no autocorrelation" is rejected for a p value below 0.05.

Results

The results of the forecast comparison, as shown in Tables 1 to 8, refer to the forecasting rounds of September and December (projections for next year) and March and June (for the current year), as far as the WIFO and IHS forecasts are concerned; as for the OECD forecasts, they refer to the autumn (for next year) and the spring forecasts (for the current year), respectively. With the exception of the unemployment rate, the correlation among the forecasts is higher than that between the forecasts and realisations. This observation is consistent with earlier findings and those obtained for other countries. It is explained by the fact that the information set at the disposal of the research institutes and the OECD is highly similar. The advantage of the larger amount of information on international economic developments accessible to the OECD is apparently offset by the higher degree of institutional knowledge at the command of the domestic research institutes. However, in projecting the unemployment variable, the better institutional knowledge may be of decisive importance (see below).

The forecasts exhibit lower standard deviations than the actual outcomes, and the standard deviations diminish with the distance of the forecast horizon. For the variables examined, a systematic deviation towards caution is apparent: prospects generally deemed positive, such as growth of real GDP, tend to be projected too low, while those considered undesirable (e.g., unemployment and inflation) are rather over-estimated. This is an indication that forecasters may implicitly face an asymmetric loss function (e.g.,

[13] Since a meaningful forecast should match as closely as possible the actual outcome to be observed later, this null hypothesis has to be rejected for a good forecast.

Aiginger, 1979). Such a tendency is partly even statistically significant (see the findings on unbiasedness below).

The higher the volatility of a variable, the more difficult it is to predict; nevertheless, for all time series reviewed, the root mean square error ($RMSE$) is smaller (or only marginally higher) than one standard deviation of the realisations. An exception is the OECD projection of the unemployment rate for which the $RMSE$ is more than twice as high as the standard deviation of the actual outcome. For both forecast horizons, the lowest (standardised) $RMSE$ values have been found for the projections of inflation. Not surprisingly, the $RMSE$ is lower, the shorter the forecast horizon, given the lower degree of uncertainty and thus the risk of error.

A decomposition of the mean square error (MSE) shows that the larger part of the forecast errors is due to the covariance (UC) or regression (UR) proportion, i.e., the components beyond the influence of the forecaster. Exceptions in this regard are the unemployment projections by the OECD and the inflation projections of the IHS, each of which exhibit a (very) large bias proportion.

For some of the variables examined it is possible to compare $RMSE$ values over time, although not all studies refer to the same forecasting dates, forecast horizons and realisations[14]. On this basis it is difficult to assess the evolution of the forecasting errors. It seems, however, that for the $RMSE$ statistics the trend has been slightly declining. The results of *Öller – Barot* (2000) for the WIFO and OECD projections of growth and inflation for Austria point into the same direction[15]. One should nevertheless bear in mind that as from the mid-1980s cyclical variations were less strong than in the ten years before.

The projections of the three forecasting institutions are of better quality than simple "no change" alternatives, with the exception of the OECD unemployment forecast. The Theil-W measure of inequality for the OECD forecast of the unemployment rate is clearly above 1 for both forecast horizons. A similar result has been obtained by *Wörz* (1994, 1995). If the OECD were to take greater account of the figures projected by the domestic institutes in May and September that are already known at the time of its own forecast, it could have improved the quality of the latter. For the unemployment rate, the absolute and the squared forecasting errors are smaller for WIFO and IHS than for the OECD, a difference that is also statistically significant. This comes despite the fact that in the given situation the OECD operates from a better information base. The measures of forecast accuracy show, that the unemployment forecasts by WIFO and IHS are superior to those from the OECD.

The assessment is more ambiguous for the growth rate of real GDP and the rate of inflation: the OECD can benefit from a higher level of information and exhibits (slightly) smaller $RMSE$ values as compared with the forecasts of September and March of the two Austrian institutes. However, this result is turned around when the forecasts of December and June are used, thereby giving the information advantage to the Austrian institutes: in this setting, the forecasts by WIFO and IHS prove more accurate. Thus, from the present analysis no firm conclusion can be drawn as to whether the forecasts from the OECD or those from the domestic institutes are of better quality.

As far as the directional accuracy of forecasts for different economic variables is concerned, the forecasted changes for all variables are correct in more than 50 percent of the cases ($ER > 0.5$) by all three forecasting institutions; yet, the null hypothesis of independence can be rejected unambiguously on the 95 percent level for the shorter forecast horizon only.

As illustrated above, forecasts should on average not deviate from realisations (i.e., be unbiased) and make best use of the available information (i.e., be efficient). The hypothesis of unbiasedness and efficiency cannot be rejected at the 95 percent level of significance for the forecasts of GDP for all three institutions, of unemployment (of the Austrian institutes) and of the inflation forecast of the OECD.

---

[14] *Kramer* (1980), in his comparison of WIFO, IHS and OECD forecasts for the 1970s, uses the values projected in December of t-1 for year t, as does *Wörz* (1995) in her analysis comparing the forecasts of WIFO, IHS, IFO and OECD for the 1970s and 1980s. Likewise, the authors use different realisations: *Wörz* (1994, 1995) takes the official data for t available in March of t+2, *Öller – Barot* (2000) the data published in the OECD Economic Outlook of December t+1 for year t.

[15] In their comparative forecast analysis for growth and inflation for 13 European countries, *Öller – Barot* (2000) included for Austria only the forecasts by WIFO and OECD.

Table 1: Descriptive statistics for autumn forecasts for next year from WIFO, IHS and OECD

WIFO, IHS: September forecasts

	Real GDP growth					Unemployment					Inflation				
	Forecast			Realisation		Forecast			Realisation		Forecast			Realisation	
	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD
Mean	2.26	2.27	2.23	2.41	2.41	5.68	5.69	4.20	5.62	3.56	2.84	3.11	2.74	2.55	2.55
Variance	0.62	0.64	0.50	1.48	1.48	0.65	0.62	0.45	0.49	0.09	1.14	1.14	1.09	1.74	1.74
Standard deviation	0.79	0.80	0.71	1.22	1.22	0.81	0.79	0.67	0.70	0.30	1.07	1.07	1.04	1.32	1.32
Correlation															
With realisation	0.14	0.01	0.14			0.71	0.62	0.57			0.89	0.85	0.89		
With WIFO		0.88	0.90				0.98	0.46				0.97	0.99		
With IHS			0.84					0.55					0.96		

Realisations: first release by WIFO as of March $t+1$.

Table 2: Forecast comparisons for autumn forecasts for next year from WIFO, IHS and OECD

WIFO, IHS: September forecasts

	Real GDP growth			Unemployment			Inflation		
	WIFO	IHS	OECD	WIFO	IHS	OECD	WIFO	IHS	OECD
Accuracy measures									
ME	0.15	0.14	0.18	–0.07	–0.08	–0.64	–0.29	–0.56	–0.19
STDR	0.65	0.66	0.58	1.15	1.12	2.22	0.81	0.81	0.79
MAE	1.20	1.20	1.10	0.50	0.53	0.65	0.54	0.66	0.54
MSE	1.86	2.12	1.77	0.34	0.43	0.71	0.46	0.79	0.42
RMSE	1.36	1.45	1.33	0.58	0.66	0.84	0.68	0.89	0.65
UM	0.01	0.01	0.02	0.01	0.01	0.57	0.18	0.40	0.09
US	0.10	0.08	0.14	0.03	0.02	0.19	0.14	0.08	0.18
UC	0.89	0.91	0.84	0.95	0.97	0.24	0.68	0.52	0.73
UR	0.21	0.29	0.16	0.28	0.29	0.34	0.02	0.00	0.04
UD	0.78	0.70	0.82	0.71	0.70	0.09	0.80	0.60	0.87
$RMSE / SD_R$	1.12	1.20	1.10	0.83	0.94	2.80	0.51	0.68	0.49
Theil *W*	0.72	0.76	0.72	0.89	0.98	1.55	0.43	0.56	0.40
Theil *U*	0.52	0.55	0.50	0.11	0.12	0.25	0.27	0.36	0.26
Test of significance of forecast differentials									
With WIFO *DM*-AE*		1.000	0.259		0.395	0.217		0.102	1.000
W-AE*		0.959	0.420		0.081	0.001		0.161	0.224
DM-SE*		0.464	0.655		0.029	0.045		0.088	0.534
W-SE*		0.815	0.568		0.086	0.344		0.065	0.136
With IHS *DM*-AE*			0.398			0.312			0.236
W-AE*			0.917			0.002			0.153
DM-SE*			0.398			0.132			0.141
W-SE*			0.776			0.324			0.082
Directional forecast accuracy									
ER	0.69	0.63	0.81	0.50	0.58	0.50	0.81	0.69	0.69
p value	0.131	0.280	0.012	0.248	1.000	1.000	0.006	0.091	0.131
Unbiasedness									
μ	0.150	0.138	0.181	–0.067	–0.075	–0.638	–0.287	–0.563	–0.194
p value	0.674	0.718	0.602	0.710	0.710	0.003	0.089	0.007	0.246
DW	1.700	1.689	1.987	2.295	2.164	1.848	1.437	1.592	1.697
Efficiency									
β_0	0.026	–0.004	0.116	–0.006	0.006	–0.6	–0.254	–0.508	–0.228
p value	0.946	0.991	0.761	0.974	0.975	0.034	0.182	0.055	0.214
F test for all $\beta_i = 0$	0.919	0.926	0.953	0.653	0.747	0.021	0.124	0.02	0.343
BL	0.606	0.767	0.788	0.087	0.187	0.049	0.445	0.763	0.552

Realisations: first release by WIFO as of March $t+1$. – *ME* . . . mean error, *STDR* . . . standard deviation ratio (forecast/realisation), *MAE* . . . mean absolute error, *MSE* . . . mean squared error, *RMSE* . . . root of *MSE*, *UM* . . . bias proportion, *US* . . . variance proportion, *UC* . . . covariance proportion, *UR* . . . regression proportion, *UD* . . . distribution proportion, $UM+US+UC = 1$ and $UM+UR+UD = 1$, $RMSE / SD_R$ . . . ratio between *RSME* of forecast and standard deviation of realisation, Theil *W* . . . Theil inequality statistic *W*, Theil *U* . . . Theil inequality statistic *U*; *DM** . . . *p* value for the asymptotically *t*-distributed modified Diebold-Mariano test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*), *W** . . . *p* value for the asymptotically standard-normally distributed Wilcoxon signed rank test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*); *ER* . . . congruence between forecasted and actual sign changes of the variable observed, *p* value . . probability of error for the χ^2-distributed test for independence between forecasted and actual sign changes; μ . . . constant term of the regression line in the restricted estimation, *p* value for the *t* test $\mu=0$, *DW* . . . Durbin-Watson statistic; β_0 . . . constant term of the regression line, *p* value for the *t* test $\beta_0 = 0$, *F* Test . . . *p* value for the joint *F* test $\beta_0 = 0$ and $\beta_1 = 0$, *BL* . . . *p* value for the Box-Ljung *Q* test for the null hypothesis "no autocorrelation".

Table 3: Descriptive statistics forspring forecasts for the current year from WIFO, IHS and OECD

WIFO, IHS: March forecasts

	Real GDP growth					Unemployment					Inflation				
	Forecast			Realisation		Forecast			Realisation		Forecast			Realisation	
	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD
Mean	2.09	2.01	1.96	2.38	2.38	5.57	5.61	4.14	5.53	3.60	3.79	3.86	2.76	2.61	2.61
Variance	0.95	0.84	1.02	1.40	1.40	0.67	0.68	0.33	0.54	0.10	1.75	1.86	1.48	1.70	1.70
Standard deviation	0.98	0.92	1.01	1.18	1.18	0.82	0.83	0.57	0.74	0.32	1.32	1.36	1.22	1.30	1.30
Correlation															
With realisation	0.66	0.69	0.71			0.93	0.90	0.43			–0.34	–0.41	0.97		
With WIFO		0.96	0.98				0.99	0.20				0.99	–0.22		
With IHS			0.97					0.23					–0.28		

Realisations: first release by WIFO as of March $t+1$.

Table 4: Forecast comparisons for spring forecasts for the current year from WIFO, IHS and OECD

WIFO, IHS: March forecasts

		Real GDP growth			Unemployment			Inflation	
	WIFO	IHS	OECD	WIFO	IHS	OECD	WIFO	IHS	OECD
Accuracy measures									
ME	0.29	0.37	0.41	–0.04	–0.08	–0.54	–0.20	–0.26	–0.15
STDR	0.82	0.78	0.85	1.11	1.12	1.77	0.88	0.93	0.93
MAE	0.71	0.70	0.71	0.25	0.29	0.54	0.33	0.30	0.28
MSE	0.91	0.89	0.89	0.09	0.13	0.56	0.19	0.16	0.14
RMSE	0.96	0.95	0.94	0.31	0.36	0.75	0.44	0.40	0.37
UM	0.09	0.15	0.19	0.02	0.05	0.52	0.21	0.45	0.16
US	0.05	0.08	0.03	0.07	0.06	0.11	0.14	0.06	0.06
UC	0.86	0.77	0.77	0.91	0.89	0.38	0.65	0.50	0.78
UR	0.04	0.01	0.03	0.20	0.20	0.33	0.06	0.02	0.01
UD	0.87	0.83	0.78	0.79	0.76	0.15	0.73	0.53	0.83
$RMSE / SD_R$	0.81	0.80	0.80	0.42	0.49	2.32	0.33	0.30	0.28
Theil *W*	0.66	0.64	0.65	0.82	0.96	2.10	0.42	0.37	0.35
Theil *U*	0.36	0.36	0.36	0.06	0.06	0.21	0.15	0.14	0.13
Test of significance of forecast differentials									
With WIFO *DM*-AE*		0.550	0.922		0.184	0.109		0.176	0.227
W-AE*		0.715	0.714		0.037	0.000		0.331	0.875
DM-SE*		0.482	0.643		0.067	0.178		0.083	0.204
W-SE*		0.715	0.428		0.034	1.000		0.030	0.235
With IHS *DM*-AE*			0.394			0.189			0.785
W-AE*			0.794			0.000			0.937
DM-SE*			0.486			0.222			0.682
W-SE*			0.183			0.905			0.937
Directional forecast accuracy									
ER	0.81	0.94	0.88	0.67	0.67	0.58	0.88	0.94	0.94
p value	0.012	0.000	0.003	0.700	0.700	0.296	0.003	0.000	0.000
Unbiasedness									
μ	0.250	0.319	0.384	–0.042	–0.075	–0.550	–0.206	–0.262	–0.162
p value	0.312	0.177	0.105	0.672	0.513	0.006	0.063	0.004	0.086
DW	1.647	2.007	1.753	1.919	2.102	2.122	1.230	1.090	1.311
Efficiency									
β_0	0.243	0.331	0.349	–0.011	0.049	–0.574	–0.147	–0.143	–0.104
p value	0.381	0.238	0.220	0.919	0.698	0.059	0.236	0.190	0.331
F test for all $\beta_i = 0$	0.472	0.446	0.282	0.991	0.879	0.054	0.065	0.007	0.129
BL	0.439	0.422	0.290	0.251	0.255	0.708	0.158	0.390	0.129

Realisations: first release by WIFO as of March $t+1$. – *ME* . . . mean error, *STDR* . . . standard deviation ratio (forecast/realisation), *MAE* . . . mean absolute error, *MSE* . . . mean squared error, *RMSE* . . . root of *MSE*, *UM* . . . bias proportion, *US* . . . variance proportion, *UC* . . . covariance proportion, *UR* . . . regression proportion, *UD* . . . distribution proportion, *UM*+*US*+*UC* = 1 and *UM*+*UR*+*UD* = 1, $RMSE / SD_R$ . . . ratio between *RSME* of forecast and standard deviation of realisation, Theil *W* . . . Theil inequality statistic *W*, Theil *U* . . . Theil inequality statistic *U*; *DM** . . . *p* value for the asymptotically *t*-distributed modified Diebold-Mariano test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*), *W** . . . *p* value for the asymptotically standard-normally distributed Wilcoxon signed rank test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*); *ER* . . . congruence between forecasted and actual sign changes of the variable observed, *p* value . . probability of error for the χ^2 -distributed test for independence between forecasted and actual sign changes; μ . . . constant term of the regression line in the restricted estimation, *p* value for the *t* test μ=0, *DW* . . . Durbin-Watson statistic; β_0 . . . constant term of the regression line, *p* value for the *t* test $\beta_0 = 0$, *F* Test . . . *p* value for the joint *F* test $\beta_0 = 0$ and $\beta_1 = 0$, *BL* . . . *p* value for the Box-Ljung *Q* test for the null hypothesis "no autocorrelation".

Table 5: Descriptive statistics for autumn forecasts for next year from WIFO, IHS and OECD

WIFO, IHS: December forecasts

	Real GDP growth					Unemployment					Inflation				
	Forecast			Realisation		Forecast			Realisation		Forecast			Realisation	
	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD
Mean	2.22	2.16	2.23	2.41	2.41	5.69	5.68	4.20	5.62	3.56	2.90	2.92	2.74	2.54	2.55
Variance	0.55	0.51	0.50	1.48	1.48	0.64	0.64	0.45	0.49	0.09	1.18	1.21	1.09	1.60	1.74
Standard deviation	0.74	0.71	0.71	1.22	1.22	0.80	0.80	0.67	0.70	0.30	1.08	1.10	1.04	1.26	1.32
Correlation															
With realisation	0.32	0.37	0.14			0.79	0.78	0.57			0.92	0.94	0.89		
With WIFO		0.92	0.89				0.99	0.41				0.98	0.98		
With IHS			0.90					0.47					0.97		

Realisations: first release by WIFO as of March $t+1$.

Table 6: Forecast comparisons for autumn forecasts for next year from WIFO, IHS and OECD

WIFO, IHS: December forecasts

		Real GDP growth			Unemployment			Inflation		
		WIFO	IHS	OECD	WIFO	IHS	OECD	WIFO	IHS	OECD
Accuracy measures										
ME		0.19	0.24	0.18	–0.08	–0.06	–0.64	–0.36	–0.38	–0.19
STDR		0.61	0.59	0.58	1.14	1.14	2.22	0.86	0.87	0.79
MAE		0.99	0.99	1.10	0.41	0.39	0.65	0.51	0.46	0.54
MSE		1.49	1.40	1.77	0.25	0.25	0.71	0.38	0.35	0.42
RMSE		1.22	1.18	1.33	0.50	0.50	0.84	0.62	0.59	0.65
UM		0.02	0.04	0.02	0.02	0.01	0.57	0.34	0.42	0.09
US		0.15	0.18	0.14	0.04	0.04	0.19	0.08	0.08	0.18
UC		0.83	0.78	0.84	0.94	0.95	0.24	0.57	0.51	0.73
UR		0.09	0.05	0.16	0.24	0.25	0.34	0.02	0.02	0.01
UD		0.89	0.91	0.82	0.74	0.74	0.09	0.64	0.56	0.87
$RMSE / SD_R$		1.01	0.97	1.10	0.72	0.72	2.80	0.49	0.47	0.49
Theil *W*		0.64	0.64	0.72	0.81	0.79	1.55	0.39	0.43	0.40
Theil *U*		0.46	0.45	0.50	0.09	0.09	0.25	0.25	0.24	0.26
Test of significance of forecast differentials										
With WIFO	*DM*-AE*		0.888	0.058		0.536	0.063		0.244	0.425
	W-AE*		0.979	0.349		0.048	0.000		0.250	0.917
	DM-SE*		0.001	0.089		0.952	0.030		0.504	0.327
	W-SE*		0.856	0.204		0.316	0.782		0.600	0.716
With IHS	*DM*-AE*			0.211			0.049			0.232
	W-AE*			0.243			0.000			0.364
	DM-SE*			0.057			0.029			0.300
	W-SE*			0.055			0.843			0.533
Directional forecast accuracy										
ER		0.81	0.75	0.81	0.67	0.58	0.50	0.75	0.69	0.69
p value		0.012	0.049	0.012	0.700	1.000	1.000	0.016	0.037	0.131
Unbiasedness										
μ		0.188	0.244	0.181	–0.075	–0.058	–0.638	–0.363	–0.381	–0.194
p value		0.557	0.428	0.602	0.626	0.707	0.003	0.014	0.005	0.246
DW		1.961	1.873	1.987	2.258	2.040	1.848	1.503	1.433	1.697
Efficiency										
β_0		0.151	0.169	0.116	–0.039	–0.001	–0.600	–0.337	–0.321	–0.228
p value		0.672	0.619	0.761	0.812	0.996	0.034	0.057	0.057	0.214
F test for all $\beta_i = 0$		0.905	0.850	0.953	0.791	0.922	0.021	0.024	0.011	0.343
BL		0.659	0.583	0.788	0.066	0.064	0.049	0.391	0.181	0.552

Realisations: first release by WIFO as of March $t+1$. – *ME* . . . mean error, *STDR* . . . standard deviation ratio (forecast/realisation), *MAE* . . . mean absolute error, *MSE* . . . mean squared error, *RMSE* . . . root of *MSE*, *UM* . . . bias proportion, *US* . . . variance proportion, *UC* . . . covariance proportion, *UR* . . . regression proportion, *UD* . . . distribution proportion, $UM+US+UC = 1$ and $UM+UR+UD = 1$, $RMSE / SD_R$ . . . ratio between *RSME* of forecast and standard deviation of realisation, Theil *W* . . . Theil inequality statistic *W*, Theil *U* . . . Theil inequality statistic *U*; *DM** . . . *p* value for the asymptotically *t*-distributed modified Diebold-Mariano test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*), *W** . . . *p* value for the asymptotically standard-normally distributed Wilcoxon signed rank test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*); *ER* . . . congruence between forecasted and actual sign changes of the variable observed, *p* value . . probability of error for the χ^2 -distributed test for independence between forecasted and actual sign changes; μ . . . constant term of the regression line in the restricted estimation, *p* value for the *t* test $\mu=0$, *DW* . . . Durbin-Watson statistic; β_0 . . . constant term of the regression line, *p* value for the *t* test $\beta_0 = 0$, *F* Test . . . *p* value for the joint *F* test $\beta_0 = 0$ and $\beta_1 = 0$, *BL* . . . *p* value for the Box-Ljung *Q* test for the null hypothesis "no autocorrelation".

Table 7: Descriptive statistics for spring forecasts for the current year from WIFO, IHS and OECD

WIFO, IHS: June forecasts

	Real GDP growth					Unemployment					Inflation				
	Forecast			Realisation		Forecast			Realisation		Forecast			Realisation	
	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD	WIFO	IHS	OECD	WIFO, IHS	OECD
Mean	2.21	2.09	1.96	2.38	2.38	5.54	5.58	4.14	5.53	3.60	3.75	3.79	2.76	3.70	2.61
Variance	1.45	1.38	1.02	1.40	1.40	0.57	0.58	0.33	0.54	0.10	2.14	2.54	1.48	2.26	1.70
Standard deviation	1.20	1.17	1.01	1.18	1.18	0.76	0.76	0.57	0.74	0.32	1.46	1.59	1.22	1.50	1.30
Correlation															
With realisation	0.85	0.81	0.71			0.97	0.96	0.43			0.99	0.98	0.97		
With WIFO		0.97	0.95				0.99	0.17				0.99	–0.18		
With IHS			0.96					0.23					–0.25		

Realisations: first release by WIFO as of March $t+1$.

Table 8: Forecast comparisons for spring forecasts for the current year from WIFO, IHS and OECD

WIFO, IHS: June forecasts

		Real GDP growth			Unemployment			Inflation	
	WIFO	IHS	OECD	WIFO	IHS	OECD	WIFO	IHS	OECD
Accuracy measures									
ME	0.16	0.29	0.41	–0.01	–0.05	–0.54	–0.05	–0.09	–0.15
STDR	1.02	0.99	0.85	1.03	1.04	1.77	0.97	1.06	0.93
MAE	0.51	0.61	0.71	0.16	0.20	0.54	0.16	0.26	0.28
MSE	0.47	0.62	0.89	0.03	0.05	0.56	0.05	0.11	0.14
RMSE	0.68	0.79	0.94	0.18	0.23	0.75	0.22	0.33	0.37
UM	0.06	0.13	0.19	0.00	0.04	0.52	0.05	0.08	0.16
US	0.00	0.00	0.03	0.02	0.02	0.11	0.03	0.08	0.06
UC	0.94	0.87	0.77	0.98	0.94	0.38	0.92	0.84	0.78
UR	0.09	0.08	0.03	0.06	0.07	0.33	0.01	0.13	0.01
UD	0.86	0.79	0.78	0.94	0.89	0.15	0.94	0.78	0.83
$RMSE / SD_R$	0.58	0.67	0.80	0.24	0.31	2.32	0.15	0.22	0.28
Theil *W*	0.46	0.52	0.65	0.47	0.59	2.10	0.16	0.24	0.35
Theil *U*	0.26	0.30	0.36	0.03	0.04	0.21	0.05	0.08	0.13
Test of significance of forecast differentials									
With WIFO *DM*-AE*		0.172	0.039		0.265	0.029		0.152	0.218
W-AE*		0.231	0.139		0.421	0.000		0.449	0.275
DM-SE*		0.243	0.119		0.153	0.126		0.096	0.194
W-SE*		0.249	0.132		0.768	0.270		0.451	0.321
With IHS *DM*-AE*			0.114			0.046			0.652
W-AE*			0.133			0.000			0.756
DM-SE*			0.096			0.141			0.739
W-SE*			0.073			0.408			0.774
Directional forecast accuracy									
ER	0.94	1.00	0.88	0.92	1.00	0.58	1.00	0.94	0.94
p value	0.000	0.000	0.003	0.005	0.001	0.296	0.000	0.000	0.000
Unbiasedness									
μ	0.119	0.231	0.384	0.000	–0.033	–0.550	–0.106	–0.175	–0.162
p value	0.495	0.233	0.105	1.000	0.638	0.006	0.112	0.040	0.086
DW	1.778	2.231	1.753	2.250	1.995	2.122	2.602	2.076	1.311
Efficiency									
β_0	0.134	0.289	0.349	0.014	–0.005	–0.574	–0.158	–0.170	–0.104
p value	0.488	0.195	0.220	0.819	0.942	0.059	0.043	0.105	0.331
F test for all $\beta_i = 0$	0.693	0.418	0.282	0.728	0.927	0.054	0.111	0.202	0.129
BL	0.034	0.114	0.290	0.059	0.178	0.708	0.593	0.416	0.129

Realisations: first release by WIFO as of March $t+1$. – *ME* . . . mean error, *STDR* . . . standard deviation ratio (forecast/realisation), *MAE* . . . mean absolute error, *MSE* . . . mean squared error, *RMSE* . . . root of *MSE*, *UM* . . . bias proportion, *US* . . . variance proportion, *UC* . . . covariance proportion, *UR* . . . regression proportion, *UD* . . . distribution proportion, $UM+US+UC = 1$ and $UM+UR+UD = 1$, $RMSE / SD_R$ . . . ratio between *RSME* of forecast and standard deviation of realisation, Theil *W* . . . Theil inequality statistic *W*, Theil *U* . . . Theil inequality statistic *U*; *DM** . . . *p* value for the asymptotically *t*-distributed modified Diebold-Mariano test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*), *W** . . . *p* value for the asymptotically standard-normally distributed Wilcoxon signed rank test for the loss functions on the basis of the absolute errors (*AE*) and the squared errors (*SE*); *ER* . . . congruence between forecasted and actual sign changes of the variable observed, *p* value . . probability of error for the χ^2 -distributed test for independence between forecasted and actual sign changes; μ . . . constant term of the regression line in the restricted estimation, *p* value for the *t* test $\mu=0$, *DW* . . . Durbin-Watson statistic; β_0 . . . constant term of the regression line, *p* value for the *t* test $\beta_0 = 0$, *F* Test . . . *p* value for the joint *F* test $\beta_0 = 0$ and $\beta_1 = 0$, *BL* . . . *p* value for the Box-Ljung *Q* test for the null hypothesis "no autocorrelation".

A bias has occurred for the unemployment forecast of the OECD and the inflation forecasts of the Austrian institutes. For the latter also, the null hypothesis of efficiency is rejected for both variables. Since the choice of a loss function is of least influence for the results of the efficiency tests, this implies for practical purposes that for the variables mentioned the accuracy of the forecasts can still be improved by a more efficient use of the information available.

Summary and conclusions

The present analysis has examined the accuracy of the forecasts of WIFO, IHS and OECD for Austria for three key macro-economic variables, namely the growth rate of GDP, the unemployment rate and the rate of inflation. For GDP and inflation, the indicators of forecast quality show virtually no difference between the three institutions. The unemployment rate, however, is found to be foreseen clearly better by the domestic institutes (see also *Wörz,* 1994, 1995). The forecasts of the institutes are on the whole unbiased – i.e., on average, they do not deviate significantly from the realisations – and efficient – best use is made of the information available. With the exception of the unemployment forecast by the OECD, all forecasts have proved clearly superior to "naïve" forecasting strategies.

The method of evaluating the quality of forecasts selected here may make the forecasters look more unfavourable than they possibly deserve (see the discussion in the introduction). Nevertheless, the results of the analysis (with the exception of the unemployment rate projections by the OECD) warrant a clear affirmative answer to the question (raised by *Kramer,* 1980, p. 18) "whether business cycle forecasts are justified at all".

Methodological annex

Notation

d_t	difference of loss function,
e_t^i	$R_t - P_t^i$, error of forecast i for year t,
$g(e_t^i)$	loss function for forecast i,
h	forecast horizon, h = 1, 2 (years),
i	authors of forecast A, B,
MSE^i	mean square error of forecast i over the period of analysis,
P_t	forecast for year t,
P_t^i	forecast by author i for year t, $i = A, B$,
$P_t^{Autumn,\, t-1}$	autumn forecast of year t–1 for year t,
$P_t^{Spring,\, t}$	spring forecast of year t for year t,
$P_{t+1}^{Autumn,\, t}$	autumn forecast of year t for year t + 1,
R_t	realisation for year t,
T	number of observations.

Measures of accuracy

$$ME = \frac{1}{T}\sum_{t=1}^{T}\left(R_t - P_t\right)$$ mean error,

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left(\left|R_t - P_t\right|\right)$$ mean absolute error,

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\left(R_t - P_t\right)^2$$ mean squared error,

$$MSE = \left(\overline{R} - \overline{P}\right)^2 + \left(s_R - s_P\right)^2 + 2\,(1-r)\,s_R\, s_P$$ for decomposition 1,

$$MSE = \left(\overline{R} - \overline{P}\right)^2 + \left(s_P - r\, s_R\right) + \left(1 - r^2\right) s_R^2$$ for decomposition 2,

$$RMSE = \sqrt{MSE}$$ root mean square error,

$UM + US + UC = 1$ decomposition 1,

$UM + UR + UD = 1$ decomposition 2,

$s_P^2 = \frac{1}{T}\sum_{t=1}^{T}(P_t - \overline{P})^2$ variance of a forecast,

$s_R^2 = \frac{1}{T}\sum_{t=1}^{T}(R_t - \overline{R})^2$ variance of actual results,

$r = \frac{\frac{1}{T}\sum_{t=1}^{T}(R_t - \overline{R})(P_t - \overline{P})}{s_R\, s_P}$ correlation coefficient between forecast and result,

$\overline{P} = \frac{1}{T}\sum_{t=1}^{T} P_t$ mean of forecasts,

$\overline{R} = \frac{1}{T}\sum_{t=1}^{T} R_t$ mean of realisations,

$UM = \frac{(\overline{R} - \overline{P})^2}{MSE}$ bias proportion of mean squared error,

$US = \frac{(s_R - s_P)^2}{MSE}$ variance proportion of mean squared error,

$UC = \frac{2(1-r)\, s_R\, s_P}{MSE}$ covariance proportion of mean squared error,

$UR = \frac{(s_P - r\, s_R)^2}{MSE}$ regression proportion of mean squared error,

$UD = \frac{(1-r^2)\, s_R^2}{MSE}$ distribution proportion of mean squared error.

Theil measure of inequality

Theil $W = \sqrt{\frac{\sum_{t=h+1}^{T}(R_t - P_t)^2}{\sum_{t=h+1}^{T}(R_t - R_{t-h})^2}}$ Theil inequality measure W,

h . . . forecast horizon (h = 1 for the current year, h = 2 for the next year),

Theil $U = \sqrt{\frac{\sum_{t=1}^{T}(R_t - P_t)^2}{\sum_{t=1}^{T} R_t^2}}$ Theil inequality measure U.

Significance tests

Diebold-Mariano test

For a pair of forecasts (A, B) with forecast horizon h the forecast errors are defined as $\{e_t^i = R_t - P_t^i\}_1^T$ with i = A, B. The loss functions shall be given by $g(e_t^i)$, where g can be, e.g., the absolute value or the square of the forecast error. However, the functional form of the loss function may also be more general (see *Christoffersen – Diebold,* 1996, 1997). The loss differential between two forecasts is given by $d_t = g\left(e_t^A\right) - g\left(e_t^B\right)$.

Diebold – Mariano (1995) developed a general test for examining the difference between two forecasts. It is only assumed that the process of the forecast error loss differential $\{d_t\}_1^T$ is covariance-stationary and "short memory". They do *not* assume, as is usually in the case in other parametric tests for this issue, that many situations of practical relevance are excluded from the outset (see *Mariano,* 2002), such as non-quadratic and asymmetric loss functions, multi-period forecasts, forecast errors that are nongaussian, non-zero mean, serially correlated, and contemporaneously correlated. The mean $\bar{d}$ of $\{d_t\}_1^T$ is assumed to be asymptotically normally distributed, and

$$\sqrt{T}\left(\bar{d}-\mu\right)\xrightarrow{d} N\left(0, V\left(\bar{d}\right)\right)$$

holds, with

$$\bar{d}=\frac{1}{T}\sum_{t=1}^{T}\left(g\left(e_t^A\right)-g\left(e_t^B\right)\right)$$

and

$$V\left(\bar{d}\right)=\frac{1}{T}\left(\gamma_0+2\sum_{\tau=1}^{h-1}\gamma_\tau\right),$$

μ . . . expected value of d_t, $V\left(\bar{d}\right)$ . . . variance of $\bar{d}$, γ_τ . . . auto-covariance of d_t of order τ. When forming the derivation of $V\left(\bar{d}\right)$, a feature of the optimal error of forecasts with horizon h is applied, whereby all auto-correlations of d_t of the order $\geq h$ equal zero.

The auto-covariance function can be estimated by

$$\hat{\gamma}(\tau)=\frac{1}{T}\sum_{t=\tau+1}^{T}\left(d_t-\bar{d}\right)\left(d_{t-\tau}-\bar{d}\right).$$

Under the null hypothesis of equal accuracy of two forecasts $\mu = 0$ holds, and one obtains the asymptotically standard-normally distributed test statistic

$$DM=\frac{\bar{d}}{\sqrt{\frac{1}{T}V\left(\bar{d}\right)}}.$$

Harvey – Leybourne – Newbold (1997) show that a modification of DM in small samples exhibits better qualities than the original test variant. The present analysis therefore uses the student-t-distributed test statistic

$$DM^*=\frac{DM}{\sqrt{\frac{T+1-2h+\frac{h(h-1)}{T}}{T}}}.$$

Wilcoxon's signed rank test

The non-parametric Wilcoxon signed rank test is applicable for loss differentials $\{d_t\}_1^T$ of independent, identical and symmetric distribution around zero, and is asymptotically normally distributed under the null hypothesis of equal forecast accuracy.

This test is carried out for the loss functions of absolute and squared errors. The test strategy applied is illustrated using the MSE:

1. The differences of the squared forecasting errors are formed according to the following rule:

   If $MSE^A < MSE^B$ form $d_t=\left|e_t^B\right|-\left|e_t^A\right|$,

   otherwise $d_t=\left|e_t^A\right|-\left|e_t^B\right|$

   for all t, $t = 1, \ldots, T$.

2. Form $I_+\left(d_t\right) = 1$ if $d_t>0$,

   0 if $d_t<0$;

   if $d_t=0$ the observation concerned will not be used in the test statistic described below.

3. Determine, by disregarding the sign of d_t, the rank numbers $R\,g_t$ of $(d_1, \ldots, d_t)$, whereby the rank number 1 is attributed to the smallest and T to the highest value. If tied values occur (i.e., the same values for d_t and therefore $R\,g_t$ show up more than once), the respective rank numbers are averaged (see *Hartung,* 1991).

4. As test statistic the variable $W = \sum_{t=1}^{T} I_{+}(d_t) R\,g_t$ is computed, for which critical values are tabled. For more than 20 observations, W can be transformed by

   $$W^* = \frac{W - \frac{T(T+1)}{4}}{\sqrt{\frac{T(T+1)(2T+1)}{24}}}$$

   into an asymptotically standard-normally distributed test statistic.

   In the event of tied values a corrected form of the test statistic (W^*) should be applied (see *Hartung,* 1991):

   $$W^* = \frac{W - \frac{T(T+1)}{4}}{\sqrt{\frac{1}{24}\left(T(T+1)(2T+1) - \frac{1}{2}\sum_{j=1}^{n} t_j (t_j - 1)(t_j + 1)\right)}},$$

   n . . . number of *different* values in $\{d_t\}_1^T$, t_j . . . number of d_t in the j-th group (j = 1, . . ., n).

Directional accuracy (sign test)

For the examination of the directional accuracy of a forecast, four cases ought to be distinguished (see Table 9). In this context, the ratio of congruence (ER) is defined as the ratio of correctly projected changes of direction (congruent signs) to all changes of direction:

$$ER = \frac{a+b}{a+b+c+d}.$$

Table 9: Contingency table

Direction of		Forecast (1) $P_t^{Spring,t} - R_{t-1}$, (2) $P_{t+1}^{Autumn,t} - P_t^{Autumn,t}$		
		≥ 0	< 0	
Realisation (1) $R_t - R_{t-1}$, (2) $R_{t+1} - R_t$	≥ 0	a (Right)	b (False)	$a + b$
	< 0	c (False)	d (Right)	$c + d$
		$a + c$	$b + d$	$a + b + c + d$

(1) . . . current (year t), *(2)* . . . next year (t+1).

The ratio of congruence can be tested under standard assumptions (see *Bleymüller – Gehlert – Gülicher,* 1994) using a χ^2-distributed test of independence (in the present case with one degree of freedom). In that case, the null hypothesis is: the sign of the change of the forecast and the sign of the change of the realisation are statistically mutually independent. The test statistic χ^2 is defined as:

$$\chi^2 = \frac{(a+b+c+d)(a\,d - b\,c)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

The p value for this test statistic is shown in the Tables 2, 4, 6 and 8.

References

Aiginger, K., "Das Element 'Vorsicht' in Zukunftsdaten", WIFO-Monatsberichte, 1979, 52(4), pp. 167-184.

Ash, J.C., Smyth, D.J., Heravi, S. M., "Are OECD Forecasts Rational and Useful? A Directional Analysis", International Journal of Forecasting, 1997, (14), pp. 381-391.

Ash, J.C., Smyth, D.J., Heravi, S.M., "The Accuracy of OECD Forecasts of the International Economy: Balance of Payments", Journal of International Money and Finance, 1998, 16(6), pp. 969-987.

Austrian Institute of Economic Research (WIFO), Kurzfristige vierteljährliche Konjunkturprognose der österreichischen Wirtschaft, Vienna, 1983 to 2000.

Batchelor, R., "How Useful are the Forecasts of Intergovernmental Agencies? The IMF and OECD Versus the Consensus", Applied Economics, 2001, 33, pp. 225-235.

Baumgartner, J., "Die Wirtschaftsprognosen von WIFO und IHS. Eine Analyse für die achtziger und neunziger Jahre", WIFO-Monatsberichte, 2002, 75(11), pp. 701-716.

Bleymüller, J., Gehlert, G., Gülicher, H., Statistik für Wirtschaftswissenschaftler, 9th Edition, Vahlen Verlag, Munich, 1994.

Blix, M., Wadefjord, J., Wienecken, U., Adahl, M., "How Good is the Forecasting Performance of Major Institutions?", Sveriges Riksbanken Economic Review, 2001, 3, pp. 38-68.

Christoffersen, P.F., Diebold, F.X., "Further Results on Forecasting and Model Selection under Asymmetric Loss", Journal of Applied Econometrics, 1996, 11(5), pp. 561-571.

Christoffersen, P.F., Diebold, F.X., "Optimal Prediction under Asymmetric Loss", Econometric Theory, 1997, 13(6), pp. 808-817.

Diebold, F.X., "The Past, Present, and Future of Macroeconomic Forecasting", Journal of Economic Perspectives, 1998, 12(2), pp. 175-192.

Diebold, F.X., Elements of Forecasting, 2. Edition, Southwestern College Publishing, Cincinnati, 2001.

Diebold, F.X., Mariano, R.S., "Comparing Predictive Accuracy", Journal of Business and Economic Statistics, 1995, 13, pp. 253-265.

Fürst, E., "Wirtschaftsprognosen in Österreich", Wirtschaftspolitische Blätter, 1980, (5), pp. 28-36.

Granger, C.W.J., Newbold, P., Forecasting Economic Time Series, Academic Press, New York, 1977.

Hartung, J., Statistik. Lehr- und Handbuch der angewandten Statistik, 8th Edition, Oldenbourg Verlag, Munich, 1991.

Harvey, D., Leybourne, S., Newbold, P., "Testing the Equality of Prediction Mean Squared Errors", International Journal of Forecasting, 1997, 13, pp. 281-291.

Holden, K., Peel, D.A., "On Testing for Unbiasedness and Efficiency of Forecasts", The Manchester School, 1990, 58(2), pp. 120-127.

Institute for Advanced Studies (IHS), Vierteljährliche Prognose der österreichischen Wirtschaft, Vienna, 1983 to 1999.

Koutsogeorgopoulou, V., "A Post-Mortem of Economic Outlook Projections", OECD, Economics Department Working Papers, 2000, (274).

Kramer, H., "Wirtschaftsprognosen: Eine kritische Überprüfung aus Sicht eines Erstellers", Wirtschaftspolitische Blätter, 1980, (5), pp. 16-27.

Lenain, P., What is the Track Record of OECD Economic Projections?, OECD, Paris, 2002, http://www.oecd.org/pdf/M00024000/M00024465.pdf.

Mariano, R.S., "Testing Forecast Accuracy", in Hendry, D., Clements, M., A Companion to Economic Forecasting, Blackwell Publ., Oxford, 2002.

Mincer, J., Zarnowitz, V., "The Evaluation of Economic Forecasts", in Mincer, J. (Ed.), "Economic Forecasts and Expectations: Analysis of Forecasting Behavior", NBER Studies in Business Cycles, 1969, (19).

OECD, Economic Outlook, No. 33 to 66, 1983 to 1999.

Öller, L.-E., Barot, B., "The Accuracy of European Growth and Inflation Forecasts", International Journal of Forecasting, 2000, 16(3), pp. 293-315.

Theil, H., Applied Economic Forecasting, North Holland, Amsterdam, 1966.

Theil, H., Principles of Econometrics, New York, 1971.

Wörz, J., Evaluating Forecast Performance. An International Comparison Between Austria and Germany – An Empirical Approach, Master thesis at the University of Innsbruck, 1994.

Wörz, J., "Evaluierung der Leistungsfähigkeit von Prognosemethoden anhand ausgewählter Variablen", Wirtschaftspolitische Blätter, 1995, (2), pp. 157-162.