Macroeconomic Forecasting in Austria An Analysis of Accuracy

Gerhard Thury

Juni 1984

6

Macroeconomic Forecasting in Austria
An Analysis of Accuracy

Gerhard Thury

WIFO Working Paper 6, Juni 1984

Österreichisches Institut für Wirtschaftsforschung
Austrian Institute of Economic Research

Macroeconomic Forecasting in Austria
An Analysis of Accuracy

Gerhard Thury

WIFO Working Paper 6, Juni 1984

Österreichisches Institut für Wirtschaftsforschung
Austrian Institute of Economic Research

MACROECONOMIC FORECASTING IN AUSTRIA
An Analysis of Accuracy

Gerhard Thury

Austrian Institute of Economic Research
A-1103 Vienna P.O. Box 91 Austria

In the present paper, we attempt a critical evaluation of macroeconomic forecasting in Austria. For this purpose, we calculate conventional magnitude measures of accuracy as well as probabilities of correctly predicting directional changes for the forecasts made by two Austrian institutions and by the OECD. ARIMA model extrapolations serve as benchmarks for comparison.

1 Introduction

It was demonstrated in Thury (1970) and Schebeck and Thury (1980) that, during the sixties and in the beginning of the seventies, the quality of forecasts of Austrian gross domestic product and its components was quite satisfactory. It should be remembered, however, that the period 1964 to 1974 was characterized by fairly constant economic growth and, therefore, was relatively easy to predict. Beginning with 1975, the economic climate changed drastically. Years of still substantial economic growth were interrupted by a year of stagnating economic activity and, finally, economic growth came to an halt almost completely with the beginning of the eighties. So, it might now be interesting to investigate how forecasters really came to grip with this new economic situation. Especially, since first attempts to judge the Austrian forecasting performance in recent years by Fleissner (1980) and Kramer (1980) produced rather contradictory results.

2 A Short History of Macroeconomic Forecasting in Austria

First attempts to predict important macroeconomic indicators for Austria, such as gross domestic product and its components, the inflation and unemployment rates, etc., were made by the Austrian Institute of Economic Research (WIFO) in the early sixties. In the beginning, these forecasts were of a purely judgemental nature. The economists of the Institute, each disposing of a large amount of inside information in his special field of research, developed in several meetings certain ideas about the development of the Austrian economy in the next future. This exchange of ideas and information then formed the basis for the Institute's forecasts. Since the early seventies, econometric models have been used to an increasing extent as a supplement to judgemental forecasting. In some rare cases, methods of time series analysis have been applied also. So, at a first glance, the forecasters at WIFO seem to follow a procedure which is very similar to that suggested in Newbold and Granger (1974). A closer investigation, however, immediately reveals that this is not the case. On the contrary, the way how forecasters at WIFO try to combine the outcome of different forecasting methods is extremely open to criticism. Instead of independently applying different forecasting methods and then forming some kind of average from the resulting outcomes, a completely contrary procedure is adopted. Slightly exaggerating, one could say that first the final results are derived by judgemental forecasting, and then it is checked whether these results can be produced by more formalized forecasting techniques also. This is the main reason why methods of time series analysis are hardly used at WIFO, because it is rather difficult to trim predictions, obtained by these methods, into a certain direction. The exact consequences of the WIFO forecasting procedure on the accuracy of the resulting forecasts are difficult to judge. Without doubt, however, it prohibits any form of learning from past experience since, for that to be possible, a careful analysis of the forecast errors of each specific method would be a prerequisite. But exactly this type of analysis is not feasible with the procedure adopted at WIFO.

The Institute of Advanced Studies (IHS) started its forecasting activities in the early seventies. Right from the beginning, the IHS generated its predictions with the help of an econometric model. Unfortunately, no information is given whether adjustments - and if yes, what type of adjustments - have been made to the published forecasts. As a consequence, it is unclear to what extent a judgemental factor (apart from the assumptions about the future values for the exogenous variables) is of importance here or not.

A third source of forecasts for macroeconomic variables is the OECD Secretariat. However, since Austria is one of the smaller member countries, only predictions for a very limited number of variables are published regularly. Forecasts for these variables are available since the mid seventies. Methodological details on the forecasting methods can be found in a special OECD report (Techniques of Economic Forecasting, Paris 1965). Special care is taken here to ensure the international consistency of foreign trade forecasts for individual countries.

3 Comparative Evaluation of the Forecast Accuracy

In this section, the accuracy of forecasts for gross domestic product, consumption, investment, exports and imports of goods and services (all in real terms), consumer prices and the unemployment rate of Austria, made by OECD, WIFO and IHS, is analysed. ARIMA model extrapolations are also included into this investigation as benchmarks for comparison. The analysis covers the period 1974 to 1983. Details on the different measures of forecast accuracy, which are used in this study, can be found in the Appendix. Besides usual magnitude measures, indices of directional accuracy are given. All these measures are calculated using predicted and realized log--changes of the relevant variables; this fact should be kept in mind when judging the forecast performance. The relevant results are summarized in exhibits 1 to 7.

Before discussing these results in greater detail, some general remarks about the usefulness of particular accuracy measures seem to be appropriate. Absolute magnitude measures, which are used to a large extent in applied work as can be seen from Carbone and Armstrong (1982), are not too informative. Relative magnitude measures and indices of directional accuracy all seem offer much better information on the quality of the analysed predictions. So, it would be highly recommendable to urge the use of this latter type measures in applied analysis. Certain problems may arise however with the relative root mean square error (RRMSE) and the relative mean absolute deviation (RMAD) in cases where the observed log-changes are close to zero. Even very small forecast errors may give rise to extremely high values for these measures. When working with log-changes, such a situation may be encountered quite often. Therefore, these two measures are not given for gross domestic product and its components, for which zero or close to zero log--changes are observed for some years of the sample period.

Exhibit 1 contains the measures of forecast accuracy for real gross domestic product. The first column under each institution gives the various measures calculated for the years 1974 to 1983, while the measures in the second column are obtained after deleting the years 1975 and 1976 from the computation.

Measures of Forecast Accuracy
Gross Domestic Product

Exhibit 1

A first glance at Exhibit 1 immediately reveals (considering first the full period 1974 to 1983) that the GDP forecasts of all three institutions (OECD, WIFO and IHS) are of rather poor quality. They are slightly better than naive no-change extrapolations, but no really significant improvement in quality over ARIMA model

	OECD		WIFO		IHS		ARIMA	
Absolute Magnitude Measures								
RMSE	.022	.010	.023	.010	.025	.014	.021	.014
MAD	.015	.007	.016	.008	.018	.011	.019	.012
Relative Magnitude Measures								
U	.703	.351	.711	.349	.711	.462	.710	.490
U^M	.016	.196	.008	.225	.001	.084	.043	.108
U^R	.096	.001	.109	.006	.228	.127	.202	.052
U^D	.888	.803	.883	.769	.771	.789	.755	.840
Regression-Based Measures								
α	.014 (.991)	.004 (.579)	.015 (1.000)	.004 (.499)	.019 (1.340)	.011 (1.266)	.014 (.988)	.002 (.144)
β	.432 (.707)	1.020 (3.378)	.378 (.604)	1.067 (3.529)	.199 (.382)	.676 (2.057)	.330 (.721)	.774 (2.080)
R^2	.059	.685	.044	.675	.018	.414	.061	.419
Measures of Directional Accuracy								
AMDA	.667	.714	.556	.714	.667	.714	.444	.429
RMDA (Last Change Model)	1.502	1.666	1.252	1.666	1.502	1.666	1.000	1.000
RMDA (ARIMA Model)	1.502	1.666	1.252	1.666	1.502	1.666	1.000	1.000

Measures of Forecast Accuracy
Gross Domestic Product

Exhibit 1

projections is observed. The most devastating result, however, follows from a regression of the realizations on the predictions. As realizations we use throughout this paper those values of a variable prevailing at that date, when the corresponding forecast was made. In the meantime, these values might have been revised several times. For gross domestic product, this regression shows that predicted and realized log-changes are hardly related. The slope coefficient, $\hat{\beta}$, is far from 1, and the coefficient of determination, R^2, practically zero. In the light of these findings, it is only cold comfort that the forecasts made by OECD, WIFO and IHS are marginally better than ARIMA model projections with respect to directional accuracy. The often made claim that judgemental forecasting, as it is applied at WIFO for example, is superior in predicting turning points of business cycles, is not supported by our empirical evidence.

Looking at the forecast errors for individual years, we found that the observed poor forecasting performance is primarily due to large errors in the years 1975 and 1976. For 1975, the predicted percentage changes lie between +3.0 and +3.9, and observed was a decline of -2.0 percent. The situation in 1976 was not much better. The forecasts range from +1.0 to +1.8 percent, and the actual growth rate turned out as +5.2 percent. Excluding these two years computations improves the forecast accuracy substantially. We have difficulties, however, in finding invulnerable arguments for this way of proceeding. One might argue that this was an unique situation then, with a long lasting period of constant economic expansion coming to an abrupt end, and that a similar situation is unlikely to occur again in the future. But, be that as it may, what we are really demonstrating here is the fact that all the institutions under study can only produce satisfactory forecasts of gross domestic product for years, in which no extraordinary events are observed. Similar conclusions are reached by Macfarlane and Hawkins (1983) in a study on economic forecasts in general.

Exhibit 2 presents measures of forecast accuracy for private consumption. The predictions for this variable are somewhat better than those for gross domestic product. Really surprising, however, is the relatively good performance of the ARIMA model in this context. It outperforms all other forecasting methods. Even with respect to directional accuracy, this model is not very inferior compared with other methods. This finding contradicts the often made claim that ARIMA models mechanically extrapolate the present

Measures of Forecast Accuracy
Consumption
Exhibit 2

state of a variable into the future. Details about this model can be found in Thury (1982). Perhaps, a general remark on the ARIMA models used in this paper should be added here. These models are of relatively uneven quality. While a considerable amount of time

	OECD	WIFO		IHS		ARIMA	
Absolute Magnitude Measures							
RMSE	.025	.024	.024	.025	.025	.013	.014
MAD	.016	.020	.019	.019	.018	.011	.013
Relative Magnitude Measures							
U	.675	.674	.649	.679	.675	.398	.464
U^M	.009	.002	.022	.017	.010	.036	.049
U^R	.220	.073	.000	.075	.023	.002	.000
U^D	.771	.925	.978	.908	.967	.962	.951
Regression-Based Measures							
α	-.020	.009	.004	.006	.002	-.003	-.003
	(-.941)	(.690)	(.244)	(.446)	(.138)	(-.372)	(-.296)
β	2.195	.656	1.002	.653	.809	1.036	.998
	(2.197)	(1.521)	(1.764)	(1.536)	(1.588)	(3.545)	(2.844)
R^2	.491	.224	.342	.229	.296	.611	.574
Measures of Directional Accuracy							
AMDA	1.000	.667	.857	.778	.857	.667	.714
RMDA (Last Change Model)	2.252	1.502	2.000	1.752	2.000	1.502	1.666
RMDA (ARIMA Model)	1.499	1.000	1.200	1.166	1.200	1.000	1.000

Measures of Forecast Accuracy
Consumption

Exhibit 2

and effort was spent on the identification and estimation of the above consumption model, for lack of time, the ARIMA models for the remaining variables were identified and estimated in a rather cursorily fashion. The relatively good forecast performance of the consumption model now seems to indicate that there might be some scope for improvement of the remaining ARIMA models, provided that more care is spent on their identification and estimation.

Exhibit 3 gives the forecast accuracy measures for investment. Investment expenditure on plant and equipment is a highly volatile variable. The naive no-change model is a serious competitor here. Taking this fact into account, the relatively good performance of the WIFO forecasts is a great surprise. This might be due to the

Measures of Forecast Accuracy
Investment
Exhibit 3

fact that an investment survey is conducted by this organisation, which apparently produces quite reliable results. It should be noticed that the investment forecasts of WIFO (and of the other forecasters also) are somewhat biased, and that the size of the fluctuations is substantially underrated. Thus, there would be scope for further improving these predictions by attributing special attention to these problems and, possibly, applying linear corrections.

Exhibits 4 and 5 contain the forecast accuary measures for

	OECD	WIFO		IHS		ARIMA	
Absolute Magnitude Measures							
RMSE	.033	.037	.028	.052	.039	.050	.042
MAD	.023	.032	.023	.040	.024	.044	.036
Relative Magnitude Measures							
U	.733	.807	.656	.880	.932	1.302	1.120
U^M	.196	.148	.303	.188	.376	.082	.168
U^R	.191	.110	.352	.099	.001	.336	.173
U^D	.613	.742	.345	.713	.623	.582	.659
Regression-Based Measures							
α	-.026 (-1.767)	-.024 (-1.668)	-.030 (-3.357)	-.008 (-.409)	-.023 (-1.301)	-.002 (-.142)	-.007 (-.464)
β	1.648 (3.176)	1.599 (2.911)	1.738 (5.826)	.445 (.844)	.958 (2.231)	.132 (.326)	.471 (1.117)
R^2	.669	.514	.850	.082	.453	.013	.172
Measures of Directional Accuracy							
AMDA	.857	.778	.857	.444	.428	.556	.572
RMDA (Last Change Model)	1.930	1.752	2.000	1.000	.999	1.252	1.170
RMDA (ARIMA Model)	1.541	1.399	1.498	.799	.748	1.000	1.000

Measures of Forecast Accuracy
Investment

Exhibit 3

exports and imports of goods and services at constant prices. For the OECD, only predictions for exports and imports of goods are available for a larger number of years. It is extremely unlikely, however, that their predictions for total exports and imports will be of better quality. On the contrary, just the opposite is to be expected. Only measures for the full period 1974 to 1983 are given here, since an exclusion of the years 1975 and 1976 did not improve the results significantly. It can be seen immediately that, for foreign trade, the performance of all forecasters is extremely poor. For total exports, these forecasts are only marginally better

Measures of Forecast Accuracy
Exports of Goods and Services
Exhibit 4

Measures of Forecast Accuracy
Imports of Goods and Services
Exhibit 5

than naive no-change extrapolations and, for total imports, the situation is even worse. Forecasts with RMSE's of almost 7 percent in the case of exports, and of more than 8 percent for imports are practically useless. Not only huge magnitude errors are committed, but the directional accuracy of foreign trade forecasts is also very low. Simple last-change models would have resulted in more accurate forecasts. A regression of realizations on forecasts reveals that, for total exports, these two magnitudes are inversely related. In the case of total imports, this absurd result is only

	OECD[1]	WIFO	IHS	ARIMA
Absolute Magnitude Measures				
RMSE	.065	.061	.068	.061
MAD	.052	.055	.057	.040
Relative Magnitude Measures				
U	.794	.827	.919	.880
U^M	.030	.101	.042	.001
U^R	.262	.220	.537	.695
U^D	.708	.679	.421	.304
Regression-Based Measures				
α	.088 (1.812)	.059 (1.943)	.093 (3.182)	.097 (3.840)
β	-.576 (-.629)	-.131 (-.186)	-.976 (-1.579)	-.736 (-1.812)
R^2	.047	.004	.238	.291
Measures of Directional Accuracy				
AMDA	.333	.333	.222	.222
RMDA (Last Change Model)	2.664	.750	.500	.500
RMDA (ARIMA Model)	-	1.500	1.000	1.000

[1] Exports of Goods

Measures of Forecast Accuracy
Exports of Goods and Services

Exhibit 4

	OECD[1]	WIFO	IHS	ARIMA
Absolute Magnitude Measures				
RMSE	.083	.074	.085	.068
MAD	.062	.058	.063	.054
Relative Magnitude Measures				
U	.957	.903	1.159	1.000
U^M	.002	.008	.002	.098
U^R	.138	.081	.348	.027
U^D	.860	.911	.650	.875
Regression-Based Measures				
α	.044	.033	.057	.016
	(.882)	(.828)	(1.618)	(.204)
β	-.108	.238	-.450	.435
	(-.110)	(.264)	(-.641)	(.385)
R^2	.002	.009	.049	.018
Measures of Directional Accuracy				
AMDA	.556	.445	.445	.445
RMDA (Last Change Model)	1.483	1.000	1.000	1.000
RMDA (ARIMA Model)	-	1.000	1.000	1.000

1 Imports of Goods

Measures of Forecast Accuracy
Imports of Goods and Services

Exhibit 5

observed for the IHS forecasts.

As already mentioned above, an exclusion of the years 1975 and 1976 does not improve the quality of forecasts for the foreign trade variables, which are among the cyclically most sensitive components of gross domestic product. Being already very disappointing in itself, in addition, this finding must be considered as strong evidence of existing consistency problems, since we observed above an evident improvement of the forecast performance for gross domestic product in the shorter period. This fact implies that the errors in foreign trade forecasts either cancelled directly within this sector or were offset by errors in predictions for other variables. Inventory investment, for example, could be one of these variables. Be that as it may, this result is suitable to raise additional doubts about the quality of the forecasts for gross domestic product.

Total exports and imports contain as important component services, and here especially tourism. It is clear that accurate predictions for these variables are very difficult to make. So, we concentrated on exports and imports of goods. Unfortunately, the forecast performance for these subcategories is only marginally better than that for total exports and imports.

Summarizing one can only state that, at present, the forecast accuracy as far as foreign trade is concerned, is extremely discouraging. For a small open economy like the Austrian, where foreign trade constitutes such an important component of gross domestic product, this fact has serious consequences. The poor forecasts of foreign trade destroy all advances which are made in predicting the domestic components of gross domestic product. Thus, an improvement of the forecast performance in general will be only attainable by raising the accuracy of the foreign trade forecasts.

Exhibits 6 and 7 show forecast accuracy measures for consumer prices and the unemployment rate. For both variables, the forecasts from all four sources are of very satisfactory quality here. It is definitely not our intention to belittle the achievements of the various forecasters, but it must be stated that the credit for the observed good forecast performance is not exclusively due to their expertness alone. Consumer prices and unemployment rate are relatively easy to predict for Austria. These two variables show much less cyclical variation than one would expect to observe according to mainstream economic theory. This relatively stable behavior is the consequence of unique institutional factors prevailing in Austria. The price formation is not completely left to the free play of market forces, but is controlled by the

Measures of Forecast Accuracy
Consumer Prices
Exhibit 6

Measures of Forecast Accuracy
Unemployment Rate
Exhibit 7

	OECD	WIFO	IHS	ARIMA
Absolute Magnitude Measures				
RMSE	.015	.009	.010	.013
MAD	.013	.008	.009	.010
Relative Magnitude Measures				
RRMSE	.075	.029	.045	.058
RMAD	.238	.154	.180	.184
U	.241	.144	.161	.206
U^M	.045	.000	.010	.002
U^R	.341	.015	.007	.001
U^D	.614	.985	.983	.997
Regression-Based Measures				
α	.023 (2.240)	.004 (.342)	.002 (.158)	.002 (.136)
β	.638 (3.715)	.940 (5.454)	.951 (4.756)	.968 (3.222)
R^2	.633	.788	.739	.565
Measures of Directional Accuracy				
AMDA	.778	.778	.750	.778
RMDA (Last Change Model)	1.167	1.167	1.125	1.167
RDMA (ARIMA Model)	1.000	1.000	.964	1.000

Measures of Forecast Accuracy
Consumer Prices

Exhibit 6

	WIFO	IHS	ARIMA
Absolute Magnitude Measures			
RMSE	.004	.006	.004
MAD	.004	.005	.004
Relative Magnitude Measures			
RRMSE	.042	.086	.035
RMAD	.177	.250	.155
U	.167	.250	.179
U^M	.162	.098	.031
U^R	.001	.108	.200
U^D	.837	.794	.769
Regression-Based Measures			
α	-.002 (-.483)	.004 (.648)	.005 (1.550)
β	1.019 (5.986)	.773 (3.555)	.794 (5.565)
R^2	.818	.612	.795
Measures of Directional Accuracy			
AMDA	.875	.556	.572
RMDA (Last Change Model)	2.333	1.483	1.525
RMDA (ARIMA Model)	1.530	.972	1.000

Measures of Forecast Accuracy
Unemployment Rate

Exhibit 7

government via a special commission for prices and wages. Firms have to propose price increases for sensitive consumer goods before this commission. We do not intend to enter a discussion of the merits or demerits of this commission. Suffice it to say, that it considerably facilitates the task of the forecaster by offering a priori information on impending price increases. The unemployment rate exhibits even less cyclical variation. Apart from recent years, where it showed a strongly rising tendency, it was almost constant during our sample period. The prevailing economic climate is reflected in variations of participation rates (especially that for women), the amount of early retirement and the number of workers from abroad, but hardly in variations of the number of registered unemployed persons. The institution with the best knowledge of these factors, and that is, beyond doubt, WIFO, should make the best forecasts for these two variables. Moreover, the predictions for consumer prices and the unemployment rate are probably self-fulfilling to a certain extent. In a way, they are considered as target variables by policy makers, and these persons do their best to prevent them from exceeding the predicted figures, sometimes even by applying rather odd methods.

4 An Attempt to Combine Forecasts

In this section, we briefly turn to the question whether a combination of the OECD, WIFO, and IHS forecasts and the ARIMA extrapolations can provide more accurate predictions than a single method. To test this question, we apply a simple average and a weighted average, where the latter is of the form

$$F_t = \sum_{i=1}^{p} w_i \, F_t^{(i)}.$$

Here, $F_t^{(i)}$ is the predicted log-change for period t from source i and p is the number of sources (4 in our example). The weights w_i are calculated using Procedure 1 of Newbold and Granger (1974), i.e.

$$w_i = \left(\sum_{s=t-\gamma}^{t-1} e_s^{(i)2} \right)^{-1} \Bigg/ \sum_{j=1}^{p} \left(\sum_{s=t-\gamma}^{t-1} e_s^{(j)2} \right)^{-1},$$

where $e_t^{(i)} = F_t^{(i)} - R_t^{(i)}$ is the forecast error of method i, and the value of γ reflects the weight given to past history. Smaller values of γ restrict the calculation of the weights to the most recent observations. The choice of this procedure was suggested by the findings of Winkler and Makridakis (1983). The weights are positive and sum to one.

Our attempt to combine forecasts was not successful. The use of a simple average brought no improvement of forecast accuracy. With the weighted average, we observe in two cases, namely for the exports and imports of goods and services, a significant improvement in directional accuracy. After taking a closer look on the structure of the forecasts, which were combined here, these findings are not very surprising any longer. Exhibit 8 contains the outcome of a correlation analysis between forecasts for gross

Correlation Coefficients for Predicted Log-Changes ($F_t^{(i)}$) and for Forecast Errors ($e_t^{(i)}$)

Exhibit 8

domestic product stemming from different sources. The results for the other variables are very similar, so that we can confine the discussion to gross domestic product. We observe not only high correlation coefficients between the individual forecasts but also between the corresponding forecast errors. Combining now forecasts, which do not contain any independent information, cannot lead to an improvement of accuracy.

Another question which emerges immediately, however, is that about possible causes for these high correlation coefficients. A rather trivial, but also very discouraging explanation would be that it is simply impossible to make more accurate forecasts for gross domestic product, irrespective which method is applied. We have certain doubts about the validity of this conclusion. We believe that some of the causes must be sought in the process of forecast formation. There is exchange of information between OECD, WIFO and IHS about the exogenous variables, on which the forecasts are based. Additionally, the OECD forecasts are already available

Gross Domestic Product

Forecasts ($F_t^{(i)}$)

	OECD	WIFO	IHS	ARIMA
OECD	1.000			
WIFO	.877	1.000		
IHS	.891	.939	1.000	
ARIMA	.739	.597	.690	1.000

Forecast Errors ($e_t^{(i)}$)

	OECD	WIFO	IHS	ARIMA
OECD	1.000			
WIFO	.964	1.000		
IHS	.978	.964	1.000	
ARIMA	.906	.942	.908	1.000

Correlation Coefficients for Predicted Log-Changes ($F_t^{(i)}$)
and for Forecast Errors ($e_t^{(i)}$)

Exhibit 8

at the time when WIFO and IHS forecasts are formed. And, finally, there exist contacts, especially between WIFO and IHS, during the process of forecast formation. The extremely high correlation coefficients between WIFO and IHS forecasts clearly show that it is an obvious strategy of the two institutes to minimize the differences between their forecasts.

In addition to the problems just discussed, there seem to exist also certain problems with our weighting scheme. Our small sample size did not allow us to pick larger values for γ, so we had to work with a γ equal to 3. The consequence were relatively large variations in weights of different methods from year to year. Moreover, we have the impression that weighting schemes allowing negative weights could be of advantage.

5 Conclusions

Unfortunately, our findings are likely to raise serious doubt about the analysed forecasters' ability to predict business cycle movements in Austria correctly. In the light of this evidence, it is difficult to see how Kramer (1980) arrived at his relatively positive judgement about the WIFO forecast performance. Our findings are more in line with the conclusions of Fleissner (1980), although we would not fully support his somewhat hypercritical judgement.

Above all, cyclical variations of gross domestic product are not predicted very accurately. This failure is primarily due to the fact that forecasts of foreign trade are of very poor quality. Possibly, this might be the consequence of a too strong reliance on the information provided by the OECD. It often turned out in the past that the economic performance of Austria diverged substantially from that of other OECD countries. The domestic components of gross domestic product, such as consumption and investment, could be predicted with reasonable accuracy. The predictions for consumer prices and unemployment were of even excellent quality.

When judging these results, two facts must be taken into account. First, by working with log-changes, we applied a very stringent test. Second, the forecasts made by WIFO and IHS in December of the preceding year, are revised every three month of the subsequent year. In general, these revisions tend to greatly improve the quality of the respective forecasts. So, by March or, at the latest, June of the subsequent year, quite an accurate assessment of the prevailing trend will be available.

Before concluding, we would like to emphasize an important point. Although our findings are likely to nourish doubts about the reliability of economic forecasts, they are not aimed at banning forecasting activities. As a matter of fact, there is no choice: forecasts have to be made. For a government, it would be impossible to devise a budget, the means of its financing etc. other than on the basis of economic forecasts. It does not seem to be really indispensable that these forecasts are highly accurate. The doubtlessly successful economic policy of the Austrian government during the past decades could serve as a good example here.

Appendix

A1. Magnitude Measures of Forecast Accuracy

For a detailed discussion of magnitude measures of forecast accuracy refer to Cicarelli and Narayan (1980), Granger and Newbold (1973), Makridakis et al. (1982), Pack (1982) and Theil (1966). We present here only a short summary of the various formulas.

But, before doing this, some more general comments seem to be appropriate. In the computation of our forecast accuracy measures we apply percentage changes of forecasts and realizations, and not levels of these magnitudes. This is done in order to secure stationarity of the predicted and realized values implicit in the assumed existence of means and standard deviations. The asymmetry between percentage increases and decreases is avoided by converting them to log-changes using the tables in Theil (1966). Finally, a short comment on the ARIMA model extrpolations must be added. Annual predictions from ARIMA models are obtained as sum of four quarterly forecasts with a forecast horizon ranging from two to five quarters, while all remaining forecasts are annual predictions with a horizon of one year. Thus, these two sets of forecasts might not be strictly comparable.

Following Pack (1982), we distinguish absolute and relative magnitude measures of forecast accuracy. For absolue measures, the size is dependent on the scale of the variable which is predicted. Relative measures are unitless and, additionally, some are based on an explicit comparison of forecasts produced by different methods.

A1.1 Absolute Measures

Root Mean Square Error

$$RMSE = \left(\frac{1}{n} \sum_{t=1}^{n} e_t^2 \right)^{1/2}$$

where $e_t = F_t - R_t$,

F_t = forecasted log-change,

R_t = realized log-change (at that time when the forecast was made).

RMSE is a standard deviation measure and, as such, preferable to the MSE variance measure.

Mean Absolute Deviation

$$MAD = \frac{1}{n} \sum_{t=1}^{n} |e_t|$$

The difference to RMSE lies in the form of the underlying loss function.

A1.2 Relative Measures

Relative Root Mean Square Error

$$\mathrm{RRMSE} = \left(\frac{1}{n} \sum_{t=1}^{n} e_t^2 / R_t^2 \right)^{1/2}$$

Relative Mean Absolute Deviation

$$\mathrm{RMAD} = \frac{1}{n} \sum_{t=1}^{n} |e_t| / |R_t|$$

Both measures are the relative logical cousins of the absolute measures given above. In both cases, problems arise when the realizations are near zero. When working with percentage changes, such a situation may be encountered quite frequently.

Inequality Coefficient

$$U = \left(\sum_{t=1}^{n} e_t^2 \Big/ \sum_{t=1}^{n} R_t^2 \right)^{1/2}$$

This measure compares forecasts derived by a certain technique to the predictions resulting from the naive "no change" model. Values

$U < 1$, $U = 1$, $U > 1$ imply that the forecasts at hand are better than, equivalent to or worse than the naive "no change" predictions, respectively.

Inequality Proportions

$$U^M = \frac{(\bar{F} - \bar{R})^2}{MSE}$$

$$U^R = \frac{(s_F - rs_R)^2}{MSE}$$

$$U^D = \frac{(1 - r^2)\, s_R^2}{MSE}$$

$$1 = U^M + U^R + U^D$$

where

$$\bar{F} = \frac{1}{n}\sum_{t=1}^{n} F_t \; ; \quad \bar{R} = \frac{1}{n}\sum_{t=1}^{n} R_t \; ;$$

$$s_F = \left(\frac{1}{n}\sum_{t=1}^{n} (F_t - \bar{F})^2\right)^{1/2} ; \quad s_R = \left(\frac{1}{n}\sum_{t=1}^{n} (R_t - \bar{R})^2\right)^{1/2} ;$$

$$r = \frac{\frac{1}{n}\sum_{t=1}^{n} (F_t - \bar{F})(R_t - \bar{R})}{s_F\, s_R} .$$

The measures U^M and U^R will tend to zero for optimal predictions, and so U^D will tend to one. Theil (1966) presents also another set of inequality proportions. But, Granger and Newbold (1973) have shown that the logic behind this decomposition is not very clear.

Regression-Based Measures

$$R_t = \alpha + \beta F_t + u_t$$

Here the estimated regression coefficients with t-values in parenthesis and the coefficient of determination (R^2) are given in the exhibits of the text. Ideally, when no discrepancies between forecasts and realizations exist, the slope of this regression line should be one and the constant term zero, with the coefficient of determination equal to one.

A2. Measures of Directional Accuracy

A detailed discussion of measures of directional accuracy can be found in Cicarelli (1982). In the following, we draw heavily from that paper.

A2.1 An Absolute Measure

Here too, we use log-changes as starting point of the computations. In terms of direction, we speak of an increase if the present log-change is greater than the previous one, and of a decrease if the opposite is true (we disregard the small number of cases where the present and the previous log-change are identical). Now, for a particular variable, we partition the predictions into increases, (F+), and decreases (F-). Comparing these with actual changes (R+ and R-), we can determine the number of times the series increased or decreased as predicted, $(R+|F+)$ and $(R-|F-)$ respectively.

But, sheer enumeration alone is not very informative. Therefore, we try to convert the above partition process into a probability framework. First dividing the number of predicted increases or decreases by the total number of forecasts, (F+) + (F-), gives the prior probability that a specific forecast method predicted an increase, P(F+), or a decrease, P(F-). Next, dividing conditional outcomes, $(R+|F+)$ and $(R-|F-)$, by the appropriate number of forecasts, (F+) and (F-), will yield the probability that predicted increases or decreases actually materialize, $P(R+|F+)$ and $P(R-|F-)$ in symbolic notation. Finally, multiplying these conditional probabilities by the relevant prior probabilities and summing, yields the probability that a forecast method correctly predicts directional change, i.e.

$$AMDA = P(F+)\ P(R+|F+) + P(F-)\ P(R-|F-).$$

A2.2 A Relative Measure

A relative measure of directional change can be derived by comparing a certain forecast technique to some competing benchmark. Thus, we have

$$\text{RMDA} = \frac{\text{AMDA}_m}{\text{AMDA}_b},$$

where RMDA stands for relative measure of directional accuracy, and AMDA_m and AMDA_b denote the respective absolute measures of a particular forecast method and a competing benchmark. In the present paper, we utilize the naive "same change" model and ARIMA models as benchmarks.

References

Carbone, R. and Armstrong, J.S., 'Note - Evaluation of Extrapolative Forecasting Methods: Results of a Survey of Academicians and Practioners', Journal of Forecasting, 1 (1982), 215 - 217.

Cicarelli, J., 'A New Method of Evaluating the Accuracy of Economic Forecasts', Journal of Macroeconomics, Fall 1982, 469 - 475.

Cicarelli, J. and Narayan, J., 'The Performance of Eleven Economic Forecasting Models in the 1970s', Business Economics, 15 (September 1980), 12 - 16.

Fleissner, P., 'Wirtschaftsprognosen zwischen Orakel, Politik und Wissenschaft', Das Prognoseproblem in der Nationalökonomie, Wirtschaftspolitische Blätter, 5 (1980), 37 - 49.

Granger, C.W.J. and Newbold, P., 'Some Comments on the Evaluation of Economic Forecasts', Applied Economics, 5 (1973), 35 - 47.

Kramer, H., 'Wirtschaftsprognosen: Eine kritische Überprüfung aus der Sicht eines Erstellers', Das Prognoseproblem in der Nationalökonomie, Wirtschaftspolitische Blätter, 5 (1980), 16 - 27.

Macfarlane, I.J. and Hawkins, J.R., 'Economic Forecasts and their Assessment', Reserve Bank of Australia, Research Discussion Paper 8302 (August 1983).

Makridakis, S. et al., 'The Accuracy of Extrapolation (Time Series) Methods', Journal of Forecasting, 1 (1982), 111 - 153.

Newbold, P. and Granger, C.W.J., 'Experience with Forecasting Univariate Time Series and the Combination of Forecasts', Journal of the Royal Statistical Society, (A), 137 (1974), 131 - 165.

Pack, D.J., 'Measures of Forecast Accuracy', presented to the ORSA/TIMS 1982 Joint National Meeting, San Diego, October 25 - 27, 1982.

Schebeck, F. and Thury, G., 'WIFO-Prognosen, Analyse der Treffsicherheit und Verbesserungsvorschläge an Hand von Zeitreihenmodellen', Österreichisches Institut für Wirtschaftsforschung, Monatsberichte 8/1980, 417 - 434.

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

Thury, G., 'Treffsicherheit und Qualität der Institutsprognosen', Österreichisches Institut für Wirtschaftsforschung, Beilage 88 zu den Monatsberichten (Oktober 1970).

Thury, G., 'Modelling Private Consumer Expenditure in Austria by Intervention Analysis', in: O.D. Anderson (ed.), Time Series Analysis: Theory and Practice 1, Amsterdam: North-Holland, 1982, 307 - 322.

Winkler, R.L. and Makridakis, S., 'The Combination of Forecasts', Journal of the Royal Statistical Society, (A), 146 (1983) 150 - 157.