



Evaluation of economic forecasts for Austria

Ines Fortin¹ · Sebastian P. Koch¹ · Klaus Weyerstrass¹

Received: 5 July 2018 / Accepted: 4 December 2019 / Published online: 27 December 2019
© The Author(s) 2019

Abstract

In this paper, we evaluate macroeconomic forecasts for Austria and analyze the effects of external assumptions on forecast errors. We consider the growth rates of real GDP and the demand components as well as the inflation rate and the unemployment rate. The analyses are based on univariate measures like RMSE and Theil's inequality coefficient and also on the Mahalanobis distance, a multivariate measure that takes the variances of and the correlations between the variables into account. We compare forecasts generated by the two leading Austrian economic research institutes, the Institute for Advanced Studies (IHS) and the Austrian Institute of Economic Research (WIFO), and additionally consider the forecasts produced by the European Commission. The results indicate that there are no systematic differences between the forecasts of the two Austrian institutes, neither for the traditional measures nor for the Mahalanobis distance. Generally, forecasts become more accurate with a decreasing forecast horizon, as expected; they are unbiased for forecast horizons of less than a year considering traditional measures and for the shortest forecast horizon considering the Mahalanobis distance. Finally, we find that mistakes in external assumptions, in particular regarding EU GDP and the oil price, translate into forecast errors for GDP and inflation.

Keywords Forecast evaluation · MAE · RMSE · Theil's U2 · Mean directional accuracy · Mahalanobis distance · External assumptions · Small open economy · Austria

JEL Classification C53 · E37 · E66

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00181-019-01814-1>) contains supplementary material, which is available to authorized users.

✉ Klaus Weyerstrass
klaus.weyerstrass@ihs.ac.at

Extended author information available on the last page of the article

1 Introduction

Macroeconomic forecasts provide important information for economic policy makers, companies and private households. In Austria, two research institutes, the Institute for Advanced Studies (Institut für Höhere Studien, IHS) and the Austrian Institute of Economic Research (Wirtschaftsforschungsinstitut, WIFO), have a long tradition of providing economic forecasts. In addition, the Austrian National Bank (Oesterreichische Nationalbank, OeNB) as well as some private banks publishes macroeconomic forecasts for Austria. Further, international institutions like the European Commission (EC), the International Monetary Fund (IMF), and the Organisation for Economic Cooperation and Development (OECD) periodically produce forecasts for Austria.

This paper focuses on the evaluation of IHS and WIFO forecasts, because they share certain features that make them particularly suitable for comparison. First, both institutes publish their forecasts always and exactly at the same time in a joint press conference. This implies that the information sets for both institutes are very similar, as they use roughly the same cutoff day until which information is taken into account. The period between the cutoff day and the publication day is, at around one week, rather short, unlike for OeNB and international institutions (OECD, IMF, EC), which are part of a forecasting system that includes a large number of countries. The longer this interval, the more likely it is that recent events cannot be incorporated in the newest forecast publication. Both IHS and WIFO produce four short-term forecasts per year, which are published toward the end of each quarter. OeNB, on the other hand, publishes macroeconomic forecasts twice a year, in June and December.

Second, the target variables of IHS and WIFO are exactly the same. One might assume that this is true for all forecasts, but as practitioners know, there is a wide range of possible choices regarding the exact specification of the variable in question. On a quarterly basis the forecaster can, for example, choose whether to include a seasonal adjustment, a working-day adjustment, or the irregular component. Regarding the growth rates of GDP and its demand components, both IHS and WIFO forecast annual averages of the original series, i.e., not adjusted for working days. This might be true also for the private banks, but it is not for OeNB. The OeNB forecasts are part of the Eurosystem forecasts, which are based on series adjusted for working days to facilitate a better comparison across the different European countries. Different series are also available when it comes to inflation and unemployment. The national inflation rate usually differs only slightly from the harmonized European rate. However, this is not true for the unemployment rate due to substantial methodological differences.

Third, the objectives and therefore the evaluation and the underlying loss functions are the same for IHS and WIFO. Both institutes aim at producing “exact” point forecasts. Apart from definitional issues about what “exact” means, producing exact forecasts is not always the main goal of forecasters. It may well be, for example, that internal macroeconomic forecasts of a finance ministry should instead produce a conservative GDP forecast in order to reduce bad surprises when it comes to tax receipts. Overly pessimistic (and sometimes optimistic) forecasts, for their part, might increase

the forecaster's media coverage and thus enhance its public visibility.¹ In any case, we assume that the main objective of both IHS and WIFO is to generate "exact" point forecasts, i.e., forecasts with the smallest forecast error.²

Yet the precise definition of forecast error is far from obvious. On the one hand, there are different "realized" values with which the forecasts can be compared. We take as realizations the first release of the annual national accounts published by the Austrian statistical office (Statistik Austria) around nine months after the end of the year. Inflation and unemployment figures are usually not subject to revision. On the other hand, the choice of the loss function with respect to the forecasting error could alter the determination of the "best" forecast.

Traditionally, measures of accuracy like the mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean-squared error (MSE), the root-mean-squared error (RMSE), the mean directional accuracy (MDA), and Theil's inequality measure U2 have been used. Although these measures are widely accepted and regularly applied in forecast evaluations, they have the drawback of being one-dimensional; i.e., they examine each variable separately and do not take the relationship between different forecasts into account. Therefore, we additionally consider the Mahalanobis distance, which assesses jointly the forecasts of a group of variables.³ This multi-dimensional measure takes both the (potentially different) variances of the variables and the correlations between the different variables into account. In addition, we test for the unbiasedness of the forecasts and for differences between the forecasters.

Specifically, we evaluate the growth forecasts of GDP and its demand components: private consumption, gross fixed capital formation, exports, and imports. We also consider the forecasts of unemployment and inflation. The forecast assessment is based on the four forecasts published per year for the current year and the following year.

Furthermore, we consider forecasts of the European Commission, where we point out that a direct comparison with the national forecasts by IHS and WIFO may be flawed due to the different sets of information available at the time of forecasting. We examine the EC spring and autumn forecasts for the current year and the following year, published in May and November. The related analysis excludes inflation and unemployment, since national and international forecasters use different definitions.

Finally, we complement our forecast evaluation with an analysis of the effects of errors regarding external assumptions on forecast errors for GDP and inflation. For this purpose, we consider a number of assumptions related to the international environment, including the GDP growth of the European Union, the oil price, and the foreign exchange rate (euro versus US dollar) and examine whether and to what

¹ Obviously, both IHS and WIFO are also interested in public visibility, but not at the cost of producing "extreme" (false) forecasts.

² Alternative objectives might be to predict business-cycle turning points. See, for example, Giusto and Piger (2017) and Kovacs et al. (2017).

³ Sinclair and Stekler (2013) were probably the first to use the Mahalanobis distance in the context of macroeconomic forecasting, when they compared different vintages of US GDP and major component estimates. Other applications of this method in forecast evaluation include Sinclair et al. (2012, 2015), Döhrn (2015) and Sinclair et al. (2016).

extent mistakes in these assumptions translate into forecast errors for GDP growth and inflation.

Economic forecasts for Austria have been evaluated before. Baumgartner (2002a) compared IHS and WIFO forecasts; Baumgartner (2002b) examined the differences between national forecasts (IHS, WIFO) and OECD forecasts; Ragacs and Schneider (2007) investigated national (IHS, WIFO, OeNB) and international (EC, IMF, OECD) forecasts; and Schuster (2018) studied national (IHS, WIFO, OeNB) and international (EC, IMF, OECD) forecasts.

The paper is structured as follows. The next section sketches the dataset. We then describe the traditional evaluation measures and the Mahalanobis distance. In the following section we present and discuss the results, including the findings related to the effects of external assumptions on the forecast error. Finally, we summarize the main findings and conclude.

2 Data

The data include IHS and WIFO economic forecasts published in the period 1995 to 2017. In each year, the two institutes publish four forecasts, usually at the end of each quarter, i.e., in March, June, September, and December. At each forecast date and for each forecast variable, annual forecasts for the years t and $t + 1$ are published.⁴ Hence, the included forecast years cover the period 1995 to 2017, and for each year eight forecasts are available, four published in the year $t - 1$ (year-ahead forecasts), and four published in the year t (current-year forecasts). For the first year, 1995, only four forecasts are available, as the $t + 1$ forecasts of 1994 are missing.⁵ We consider the following variables: annual growth rates of real GDP, private consumption, real gross fixed capital formation (investment), exports, and imports. In addition to these demand components, forecasts of the inflation rate as measured by the national consumer price index and forecasts of the unemployment rate (according to national statistics, i.e., registered unemployment) are evaluated.

As actual data, we use the first (and usually final) publications of annual inflation and annual unemployment data; for the national accounts variables, we take the first release of the *annual* accounts, published by the Austrian statistical office approximately nine months after the end of a given year. While there are almost no revisions of unemployment and inflation rates, GDP and its components are frequently revised, and these revisions are sometimes large relative to the absolute values of the growth rates. This issue is particularly relevant for capital formation, due to its large variability. We use the first release of the *annual* accounts rather than the first “preliminary” release based on quarterly accounts, which are available three months after the end of the year, as the annual accounts are based on a larger information set. In a robustness check, we find that the differences between these data vintages are very small (see Sect. 4.3). Alternatively, it would be possible to take the latest data release.

⁴ However, the underlying models might run on a shorter periodicity.

⁵ Note that neither IHS nor WIFO published a forecast in June 1997 and hence the June t forecasts for 1997 and the June $t - 1$ forecasts for 1998 are missing.

The question of what values to take as realizations or actual values is a much debated issue. In the literature on forecast evaluations, both first and later data vintages (releases) have been used as benchmarks. Sinclair et al. (2016) evaluate German macroeconomic forecasts for the year 2013, published around December 2012. They compare these forecasts with the first release of actual data for 2013, published in January 2014, as well as the “final” (second) release of February 2014. The corresponding results are very similar. Sinclair and Stekler (2013) compare different vintages of US GDP and its ten major components: namely, the initial estimates available one month after the end of the quarter, and the estimates available three months after the end of the quarter. Despite the existence of some biases, overall the differences are rather small. Kirby et al. (2015) look at the accuracy of the NIELS’s forecasts of GDP growth in the UK, the USA, and the euro area, comparing these forecasts to the first release of the variables. Sheng (2015) evaluates real GDP growth, inflation, and unemployment forecasts of members of the FOMC of the US Federal Reserve System, using “final” estimates that are released roughly three months after the end of the quarter. Chen et al. (2016) evaluate forecasts of GDP growth and inflation for ten Asian countries, where forecasts are compared to initial releases. As a robustness check, they use revised data and find that the results are rather similar. The European Commission, in its regular forecast assessments,⁶ uses different realized values for current-year and year-ahead forecasts. The realizations are taken from the same publication as the forecasts, i.e., from the autumn publications for year-ahead forecasts and from the spring publications for current-year forecasts. Evaluations of macroeconomic forecasts for Austria usually take the first preliminary release of the national accounts provided by WIFO.⁷ Baumgartner (2002a, b) considers additionally the first “final” release produced by the Austrian statistical office, as do we, and concludes that the differences are very small.⁸

3 Evaluation measures

The selection of appropriate measures for the evaluation of forecasts is not straightforward. Obviously, the “best” economic forecast is a forecast that accurately predicts the realization of the forecast variable. Equally obviously, it is almost impossible to deliver “exact” predictions. Therefore, intervals, e.g., the 68% confidence interval, are sometimes published. In business-cycle research, the correct anticipation of so-called turning points is even more important than the exact forecast of a certain GDP growth rate. A standard evaluation criterion is that the forecast beats the naive no-change forecast. Furthermore, the forecasts should be unbiased, implying that the forecast should not be systematically too optimistic or too pessimistic. We evaluate the accuracy of the IHS and WIFO forecasts on the basis of traditional one-dimensional and more novel multi-dimensional measures.

⁶ The first study is Keereman (1999); the most recent evaluation is Fioramanti et al. (2016).

⁷ These studies include Baumgartner (2002a, b), Ragacs and Schneider (2007), and Schuster (2018)

⁸ We also find that the differences between using the first preliminary release and the first official release are very minor (see Sect. 4.3).

3.1 Traditional measures

We apply the following traditional measures: the mean absolute error (MAE)⁹, the root-mean-squared error (RMSE)¹⁰, the mean absolute percentage error (MAPE)¹¹, Theil's inequality coefficient (U2)¹², and the mean directional accuracy (MDA)¹³. The RMSE is defined as the square root of the average difference between the forecast and the actual realization. Since it is scale dependent, the RMSE is useful for a comparison between different forecasts, but its magnitude is not meaningful in itself. Theil's inequality coefficient (see Theil 1966) compares a given model-generated forecast with the naive forecast of no change. If the forecast is better than the naive no-change assumption, then U2 is smaller than one.¹⁴

The test for unbiasedness of the forecast is based on a procedure introduced by Mincer and Zarnowitz (1969). The test rests on regressing the realized values on a constant and the forecast. However, as Sinclair et al. (2010) point out, forecast errors might depend on the state of the economy, e.g., the stage of the business cycle. Therefore, these authors suggest to include a dummy for the state of the economy. This gives rise to estimating Eq. (1).

$$A_t = \alpha + \beta F_t + \gamma D_t + \epsilon_t \quad (1)$$

where F_t and A_t are the forecast for and the actual value at year t , respectively. D_t is the recession dummy, which is not present in the original version of the test. The recession dummy takes the value 1 if the economy is in a recession, and 0 otherwise. In order to identify the stage of the business cycle, first a simple Hodrick–Prescott (HP) filter is applied to the level of GDP over the entire data sample, and then the output gap is calculated as the deviation of actual GDP from the HP trend. We define a recession as a year with a negative output gap.¹⁵ For the forecast to be unbiased, α should not be significantly different from zero, β should not significantly deviate from one, and γ should not be significantly different from zero. We test this joint hypothesis with a Wald test. For the growth rates of GDP and the demand components, we employ both the original Mincer–Zarnowitz test, i.e., without the recession dummy, and the

⁹ $MAE = \frac{1}{n} \sum_{t=1}^n |F_t - A_t|$

¹⁰ $RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (F_t - A_t)^2}$

¹¹ $MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{F_t - A_t}{A_t} \right|$

¹² $U2 = \sqrt{\frac{\sum_{t=h+1}^n (F_t - A_t)^2}{\sum_{t=h+1}^n (A_t - A_{t-h})^2}}$

¹³ $MDA = \frac{1}{n-1} \sum_{t=2}^n 1(\text{sgn}(A_t - A_{t-1}) = \text{sgn}(F_t - F_{t-1}))$

where F_t and A_t are the forecast for and the actual value at t , respectively, and h is the forecast horizon ($h = 1$ for current-year forecasts and $h = 2$ for year-ahead forecasts). Note that in the definition of the mean directional accuracy we do not consider the values of zero separately but together with the values of ones. We thus conclude that the forecast direction of a given variable is assessed correctly if either the forecast goes down when the actual value goes down, or if the forecast goes up or stays the same when the actual value goes up or stays the same.

¹⁴ Note that for GDP and the demand components the assumption of no change refers to the growth rates of these variables.

¹⁵ Strictly speaking, this is a downturn rather than a recession.

modified version that includes this dummy. However, we believe that for the inflation rate and the unemployment rate, only the original test is meaningful: the labor market lags behind real economic activity. Hence, in general, unemployment does not rise at the same time as GDP growth falls (it may even be negative), but only later on, sometimes even only when real activity has already risen again. Furthermore, the labor market in Austria, as in many other European countries, is characterized by labor hoarding: companies react to an economic downturn only to an attenuated extent so as to avoid hiring costs in the following recovery. Inflation, likewise, does not follow the economic cycle very closely; hence, there are periods which according to our simple definition would be classified as a recession, but that also involve high and persistent inflation. At the same time, defining the recession dummy for each variable separately would also not be very meaningful, since a recession is usually characterized as a significant and widespread decline in activity across the economy lasting longer than a few months.

In addition, we apply the encompassing test introduced by Chong and Hendry (1986) in order to judge whether one institute’s forecast contains all the information inherent in the other institute’s forecast:

$$FE_t^{IHS} = \alpha_1 + \beta_1 FE_t^{WIFO} + \epsilon_t^{IHS}, \quad FE_t^{WIFO} = \alpha_2 + \beta_2 FE_t^{IHS} + \epsilon_t^{WIFO} \quad (2)$$

where $FE_t^i = F_t^i - A_t^i$, $i = IHS, WIFO$, is the forecast error for year t . The null hypothesis is that all information included in one forecast is already contained in the other forecast and hence $\beta_1 = 0$ or $\beta_2 = 0$. In the general version of the test, one given forecast is compared with a number of other forecasts, and the idea is that if a single forecast contains all the information contained in the other individual forecasts, that forecast will be just as good as a combination of all other forecasts.

3.2 The Mahalanobis distance

The aforementioned measures share the drawback of being one-dimensional. Therefore, we also apply the Mahalanobis distance, which is a multi-dimensional evaluation measure taking the variances of and the correlations between the variables into account. This measure allocates weights to the individual forecast errors, which are implied by the variance–covariance matrix of the variables. In utilizing this methodology, we follow Sinclair and Stekler (2013).

In order to formally define the Mahalanobis distance, let us assume that F_t is an m -dimensional vector of forecasts for time period t , and A_t is an m -dimensional vector of actual realizations of a variable at time period t . Let m be the number of variables to be predicted. If, for example, the growth rates of GDP, the inflation rates, and the unemployment rates are taken into account, m equals three.

$$F_t = \begin{pmatrix} F_{1t} \\ \vdots \\ F_{mt} \end{pmatrix}, \quad A_t = \begin{pmatrix} A_{1t} \\ \vdots \\ A_{mt} \end{pmatrix} \quad (3)$$

Let \bar{F} and \bar{A} be the mean column vectors of F_t and A_t , respectively, and let W be the pooled sample variance-covariance matrix of F_t and A_t . Then we define the Mahalanobis distance, M , as

$$M = \sqrt{(\bar{F} - \bar{A})' W^{-1} (\bar{F} - \bar{A})} \quad (4)$$

Under the assumptions of normality, one can construct an F -statistic based on the squared Mahalanobis distance, M^2 , to test the null hypothesis that two sets of forecasts have the same population means.¹⁶ We employ this test in order to examine the difference between the IHS and WIFO joint forecasting accuracies.

For assessing the unbiasedness of the joint forecasts, or rather the existence of any systematic errors, we follow the procedure used by Sinclair and Stekler (2013). This is a more general approach than investigating the variables separately (as done in Sect. 3.1). According to this approach, a given forecast error should not depend on past forecast errors of either the variable itself or other variables. We thus construct a first-order vector autoregression (VAR(1)) of the forecast errors of each variable, which is given by

$$FE_t = \beta_0 + \beta_1 FE_{t-1} + \epsilon_t \quad (5)$$

where FE_t is an m -dimensional vector of forecast errors at time t (with $FE_t = F_t - A_t$), β_0 is an m -dimensional vector of constants, and β_1 is an $m \times m$ matrix of coefficients on the lags of the forecast error. If the joint estimators are unbiased, then none of the coefficients in the VAR should be significant: the constant estimates should be zero, the coefficients on the own lags should be zero, and none of the past errors made in forecasting the other variables should Granger-cause any of the other errors.¹⁷ This implies a total number of $m \times 3$ single hypotheses that need to be examined. Instead of testing each null hypothesis separately at a given level α , we use the Bonferroni–Holm test (see Holm 1979). This is a multiple-level α test, whereby the probability of committing any type I error is always smaller than or equal to a given level α .¹⁸ In contrast to our procedure, Sinclair and Stekler (2013) examine the null hypotheses separately and do not employ multiple-level α tests.

4 Results

Both IHS and WIFO forecasts of GDP growth show a smoothing pattern over the business cycle. While in upturns both institutes tend to underestimate growth, in downturns they both overestimate growth. The pattern is shown in Fig. 1, which plots the forecast

¹⁶ $F = \frac{n_1 n_2 (n_1 + n_2 - m - 1)}{(n_1 + n_2) m (n_1 + n_2 - 2)} M^2$, with m and $n_1 + n_2 - m - 1$ the degrees of freedom, where n_1 and n_2 are the numbers of observations of the first and second group of variables (see McLachlan 1999).

¹⁷ This is a generalization of the Holden and Peel (1990) test for bias.

¹⁸ Testing each null hypothesis at the significance level α may lead to an inflated occurrence of multiple type I errors producing spurious significance. For more details on multiple-level α tests, see, for example, Hochberg and Tamhane (1987), Hsu (1996) and Alt et al. (2011).

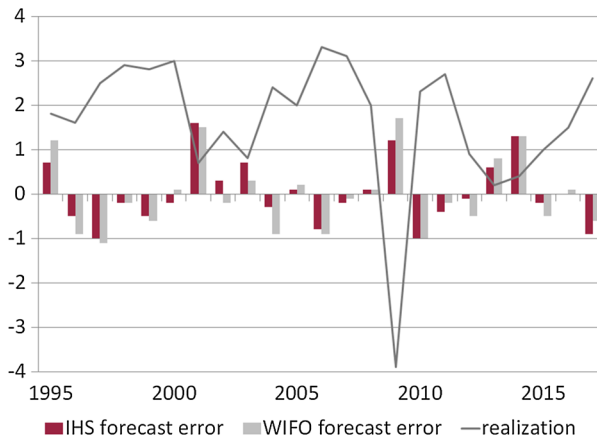


Fig. 1 Forecast biases and realizations of GDP growth. The figure shows the forecast errors (forecasts minus realized values) of IHS and WIFO March t forecasts and the realized values of GDP growth

errors of the current-year March forecasts against the realized values. This smoothing behavior seems to be characteristic of economic forecasters,¹⁹ as unexpected shocks (“extreme” events) are not highly predictable in general.

Another observation is that the forecast disagreement concerning GDP growth across the two institutes, measured in terms of the absolute difference between the respective forecasts, seems to be larger in downturns than in upswings. This is reflected by a negative correlation between the forecast differences and the realized values, for all current-year forecasts of GDP.²⁰ This pattern suggests that forecasters particularly disagree in the assessment of economic downturns or recessions.

4.1 Traditional measures

4.1.1 Accuracy

Tables 1 and 2 report the forecast evaluation results based on the traditional measures. Table 1 presents the values of the respective test statistics, while Table 2 assesses the improvement of the forecasts over time. For the latter, the evaluation measure for a given forecast horizon is divided by the “best” evaluation measure, no matter whether this best measure is provided by IHS or WIFO. Usually the best forecast is the one with the shortest forecast horizon (i.e., the December t forecast). Also, better forecasts usually go along with smaller values of the evaluation measures, except for the mean directional accuracy (MDA), where larger values imply better forecasts. The best forecast for a given variable thus shows a value of one. A value of two, for example,

¹⁹ See, for example, Ragacs and Schneider (2007), Sinclair et al. (2010), Sinclair and Stekler (2013), and An et al. (2018).

²⁰ The correlation is equal to -0.6 for the average of the four current-year forecasts, and it is significant at the 5% level ($p = 0.0027$).

Table 1 Evaluation results: MAE, RMSE, Theil's U2, MAPE, MDA

	GDP		Consump.		Invest.		Exports		Imports		Inflation		Unempl.	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
<i>MAE</i>														
Mar t-1	1.20	1.20	0.74	0.68	3.01	2.81	4.00	3.89	3.72	3.91	0.73	0.67	0.72	0.64
Jun t-1	1.21	1.23	0.76	0.73	2.86	2.71	4.07	4.12	3.77	3.93	0.65	0.65	0.62	0.62
Sep t-1	1.01	0.97	0.73	0.65	2.70	2.53	3.51	3.54	3.29	3.43	0.68	0.62	0.46	0.47
Dec t-1	0.79	0.77	0.66	0.60	2.36	2.05	3.12	3.06	2.95	2.95	0.54	0.53	0.39	0.38
Mar t	0.56	0.65	0.64	0.62	2.03	2.18	2.59	2.79	2.78	3.06	0.37	0.37	0.30	0.24
Jun t	0.43	0.47	0.60	0.57	1.80	2.02	2.12	2.20	2.41	2.52	0.17	0.18	0.14	0.15
Sep t	0.34	0.39	0.44	0.44	1.34	1.26	2.09	2.20	2.01	2.13	0.11	0.10	0.05	0.06
Dec t	0.32	0.32	0.43	0.39	1.28	1.24	1.91	2.00	1.87	1.86	0.02	0.04	0.03	0.03
<i>RMSE</i>														
Mar t-1	1.67	1.61	0.92	0.85	3.81	3.55	6.01	5.76	5.59	5.53	0.87	0.83	0.87	0.78
Jun t-1	1.66	1.62	0.94	0.89	3.69	3.46	5.89	5.92	5.50	5.54	0.83	0.83	0.78	0.73
Sep t-1	1.43	1.38	0.88	0.79	3.45	3.19	5.33	5.19	5.06	4.94	0.82	0.75	0.58	0.58
Dec t-1	1.11	1.06	0.80	0.73	2.93	2.51	4.29	4.40	4.15	4.33	0.62	0.63	0.47	0.45
Mar t	0.71	0.81	0.80	0.76	2.51	2.57	3.32	3.73	3.53	3.93	0.45	0.43	0.36	0.30
Jun t	0.55	0.56	0.75	0.71	2.19	2.38	2.77	2.84	3.08	3.09	0.22	0.21	0.18	0.18
Sep t	0.43	0.45	0.59	0.59	1.70	1.70	2.64	2.69	2.55	2.70	0.14	0.13	0.09	0.08
Dec t	0.37	0.38	0.57	0.52	1.64	1.58	2.36	2.45	2.32	2.45	0.05	0.06	0.05	0.06
<i>Theil's U2</i>														
Mar t-1	0.71	0.68	0.81	0.74	0.62	0.58	0.72	0.69	0.71	0.70	0.69	0.66	1.04	0.93
Jun t-1	0.70	0.68	0.81	0.77	0.59	0.55	0.69	0.69	0.68	0.69	0.64	0.64	0.91	0.85
Sep t-1	0.61	0.59	0.77	0.70	0.56	0.52	0.63	0.62	0.64	0.63	0.65	0.60	0.69	0.69
Dec t-1	0.47	0.45	0.70	0.64	0.48	0.41	0.51	0.52	0.53	0.55	0.50	0.50	0.56	0.54

Table 1 continued

	GDP		Consump.		Invest.		Exports		Imports		Inflation		Unempl.	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
Mar t	0.35	0.40	0.70	0.67	0.41	0.42	0.40	0.44	0.45	0.50	0.35	0.34	0.43	0.36
Jun t	0.27	0.27	0.64	0.61	0.35	0.38	0.32	0.33	0.38	0.38	0.17	0.17	0.22	0.21
Sep t	0.21	0.22	0.52	0.52	0.28	0.28	0.31	0.32	0.32	0.34	0.11	0.10	0.11	0.10
Dec t	0.19	0.19	0.50	0.46	0.27	0.26	0.28	0.29	0.29	0.31	0.04	0.05	0.06	0.07
<i>MAPE</i>														
Mar t-1	111.2	104.9	103.8	95.6	261.6	203.2	125.9	119.8	257.7	269.3	55.1	53.7	10.1	9.0
Jun t-1	116.3	106.7	108.5	100.6	263.8	236.3	129.2	127.6	269.2	285.5	51.7	54.6	8.7	8.8
Sep t-1	92.1	80.4	100.6	83.1	226.0	220.1	97.2	98.1	201.6	214.1	53.9	48.3	6.4	6.7
Dec t-1	67.3	72.2	85.4	80.2	181.5	208.4	84.0	84.6	159.6	188.8	39.8	36.6	5.4	5.4
Mar t	57.0	65.3	76.1	76.4	158.4	252.2	74.1	75.3	171.0	190.8	24.5	22.0	4.2	3.5
Jun t	44.0	41.2	66.3	60.9	189.8	304.9	64.5	58.7	154.2	133.8	10.9	10.8	2.0	2.1
Sep t	29.2	29.6	41.3	44.4	183.5	180.2	53.2	56.8	94.4	93.6	7.0	5.8	0.7	0.9
Dec t	22.7	19.5	42.1	38.0	185.0	227.6	46.1	45.8	76.9	73.7	0.9	2.6	0.4	0.5
<i>MDA</i>														
Mar t-1	0.41	0.41	0.55	0.55	0.50	0.59	0.23	0.36	0.36	0.55	0.45	0.55	0.55	0.55
Jun t-1	0.35	0.35	0.40	0.40	0.45	0.45	0.35	0.35	0.35	0.40	0.55	0.45	0.65	0.70
Sep t-1	0.32	0.50	0.45	0.55	0.55	0.64	0.41	0.50	0.41	0.41	0.50	0.59	0.77	0.73
Dec t-1	0.64	0.64	0.55	0.64	0.73	0.82	0.50	0.59	0.41	0.36	0.86	0.77	0.73	0.73
Mar t	0.77	0.77	0.82	0.73	0.77	0.82	0.64	0.73	0.55	0.73	0.91	0.82	0.73	0.77
Jun t	0.85	0.80	0.75	0.70	0.85	0.85	0.75	0.75	0.60	0.65	0.95	0.95	0.95	0.90
Sep t	0.82	0.86	0.73	0.73	0.86	0.86	0.77	0.77	0.64	0.82	0.95	0.95	1.00	0.95
Dec t	0.82	0.86	0.77	0.77	0.91	0.82	0.77	0.77	0.73	0.73	1.00	1.00	1.00	1.00

The table shows the univariate evaluation results for all forecast variables, for the eight forecasts published by IHS and WIFO

Table 2 Improvement of forecasts over time: univariate measures

	GDP		Consumpt.		Invest.		Exports		Imports		Inflation		Unempl.	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
<i>MAE</i>														
Mar t-1	3.73	3.72	1.92	1.76	2.43	2.27	2.10	2.04	2.00	2.10	33.40	31.00	27.67	24.67
Jun t-1	3.77	3.83	1.96	1.88	2.31	2.19	2.13	2.16	2.03	2.11	30.11	29.90	23.87	23.87
Sep t-1	3.14	3.00	1.88	1.67	2.18	2.04	1.84	1.85	1.77	1.85	31.20	28.60	17.67	18.00
Dec t-1	2.45	2.41	1.71	1.56	1.91	1.66	1.63	1.60	1.58	1.59	25.00	24.60	14.83	14.50
Mar t	1.74	2.03	1.65	1.61	1.64	1.76	1.36	1.46	1.50	1.64	17.20	16.80	11.50	9.33
Jun t	1.33	1.46	1.54	1.47	1.45	1.63	1.11	1.15	1.30	1.35	7.95	8.36	5.40	5.58
Sep t	1.07	1.22	1.13	1.15	1.08	1.02	1.09	1.15	1.08	1.14	5.20	4.60	2.00	2.33
Dec t	1.00	1.00	1.10	1.00	1.03	1.00	1.00	1.05	1.00	1.00	1.00	1.80	1.00	1.33
<i>RMSE</i>														
Mar t-1	4.46	4.29	1.77	1.62	2.42	2.25	2.55	2.44	2.41	2.39	18.64	17.84	16.98	15.24
Jun t-1	4.45	4.32	1.80	1.71	2.34	2.19	2.50	2.51	2.37	2.39	17.76	17.82	15.23	14.28
Sep t-1	3.81	3.68	1.69	1.52	2.18	2.02	2.26	2.20	2.19	2.13	17.52	16.16	11.33	11.31
Dec t-1	2.97	2.84	1.54	1.39	1.85	1.59	1.82	1.87	1.79	1.87	13.39	13.50	9.17	8.88
Mar t	1.89	2.16	1.53	1.45	1.59	1.63	1.41	1.58	1.52	1.70	9.57	9.25	7.01	5.94
Jun t	1.48	1.50	1.44	1.37	1.39	1.51	1.17	1.20	1.33	1.33	4.71	4.57	3.57	3.44
Sep t	1.15	1.21	1.13	1.14	1.08	1.07	1.12	1.14	1.10	1.16	3.03	2.79	1.73	1.63
Dec t	1.00	1.01	1.10	1.00	1.04	1.00	1.00	1.04	1.00	1.06	1.00	1.34	1.00	1.15
<i>Theil's U2</i>														
Mar t-1	3.83	3.68	1.77	1.62	2.42	2.25	2.55	2.44	2.41	2.39	18.64	17.84	17.83	16.00
Jun t-1	3.76	3.65	1.76	1.67	2.29	2.14	2.44	2.45	2.32	2.34	17.37	17.43	15.64	14.67
Sep t-1	3.27	3.16	1.69	1.52	2.18	2.02	2.26	2.20	2.19	2.13	17.52	16.16	11.89	11.88
Dec t-1	2.55	2.44	1.54	1.39	1.85	1.59	1.82	1.87	1.79	1.87	13.39	13.50	9.63	9.32

Table 2 continued

	GDP		Consumpt.		Invest.		Exports		Imports		Inflation		Unempl.	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
Mar t	1.89	2.16	1.53	1.45	1.59	1.63	1.41	1.58	1.52	1.70	9.57	9.25	7.36	6.24
Jun t	1.46	1.48	1.40	1.34	1.36	1.48	1.15	1.18	1.30	1.30	4.60	4.47	3.71	3.53
Sep t	1.15	1.21	1.13	1.14	1.08	1.07	1.12	1.14	1.10	1.16	3.03	2.79	1.85	1.71
Dec t	1.00	1.01	1.10	1.00	1.04	1.00	1.00	1.04	1.00	1.06	1.00	1.34	1.00	1.21
<i>MAPE</i>														
Mar t-1	5.70	5.38	2.73	2.52	1.65	1.28	2.75	2.61	3.50	3.66	58.19	56.75	25.72	23.07
Jun t-1	5.96	5.47	2.86	2.65	1.67	1.49	2.82	2.78	3.66	3.88	54.59	57.70	22.24	22.43
Sep t-1	4.72	4.12	2.65	2.19	1.43	1.39	2.12	2.14	2.74	2.91	56.92	51.06	16.39	17.04
Dec t-1	3.45	3.70	2.25	2.11	1.15	1.32	1.83	1.85	2.17	2.56	42.07	38.70	13.86	13.71
Mar t	2.92	3.35	2.00	2.01	1.00	1.59	1.62	1.64	2.32	2.59	25.93	23.27	10.81	8.91
Jun t	2.25	2.11	1.74	1.60	1.20	1.93	1.41	1.28	2.09	1.82	11.53	11.39	5.15	5.27
Sep t	1.50	1.52	1.09	1.17	1.16	1.14	1.16	1.24	1.28	1.27	7.36	6.10	1.84	2.26
Dec t	1.16	1.00	1.11	1.00	1.17	1.44	1.01	1.00	1.04	1.00	1.00	2.77	1.00	1.34
<i>MDA</i>														
Mar t-1	0.47	0.47	0.67	0.67	0.55	0.65	0.29	0.47	0.44	0.67	0.45	0.55	0.55	0.55
Jun t-1	0.41	0.41	0.49	0.49	0.50	0.50	0.45	0.45	0.43	0.49	0.55	0.45	0.65	0.70
Sep t-1	0.37	0.58	0.56	0.67	0.60	0.70	0.53	0.65	0.50	0.50	0.50	0.59	0.77	0.73
Dec t-1	0.74	0.74	0.67	0.78	0.80	0.90	0.65	0.76	0.50	0.44	0.86	0.77	0.73	0.73
Mar t	0.89	0.89	1.00	0.89	0.85	0.90	0.82	0.94	0.67	0.89	0.91	0.82	0.73	0.77
Jun t	0.98	0.93	0.92	0.86	0.94	0.94	0.97	0.97	0.73	0.79	0.95	0.95	0.95	0.90
Sep t	0.95	1.00	0.89	0.89	0.95	0.95	1.00	1.00	0.78	1.00	0.95	0.95	1.00	0.95
Dec t	0.95	1.00	0.94	0.94	1.00	0.90	1.00	1.00	0.89	0.89	1.00	1.00	1.00	1.00

The number for a given forecast variable and a given evaluation criterion is the ratio with respect to the minimum criterion (of either IHS or WIFO) for that variable

Table 3 Diebold–Mariano test for difference in accuracy

	GDP	Consumpt.	Invest.	Exports	Imports	Inflation	Unempl.
Mar t-1	0.46	0.11	0.08	0.40	0.76	0.50	0.09
Jun t-1	0.59	0.25	0.18	0.90	0.78	0.96	0.27
Sep t-1	0.43	0.10	0.18	0.59	0.65	0.22	0.98
Dec t-1	0.47	0.04*	0.11	0.51	0.31	0.92	0.64
Mar t	0.10	0.33	0.56	0.20	0.20	0.66	0.16
Jun t	0.87	0.59	0.25	0.80	0.97	0.82	0.84
Sep t	0.64	0.89	0.94	0.85	0.57	0.49	0.73
Dec t	0.89	0.01*	0.73	0.62	0.52	0.16	0.49

The table lists p values for testing the null hypothesis that IHS and WIFO forecasts show the same accuracy, where the loss function is the squared forecast error. Starred figures indicate that the null hypothesis is rejected at the 5% significance level and hence that IHS and WIFO forecasts show different levels of accuracy

means that the respective forecast error is twice the error of the best forecast for the respective variable.

In order to judge whether the forecasts across the two institutes differ significantly from each other, we employ a standard Diebold–Mariano test. For basically all variables and all forecast horizons, our results imply that the forecasts of IHS and WIFO do not differ significantly from each other. The only exception is consumption growth in the December year-ahead and current-year forecasts, where, as shown in Table 3, WIFO seems to provide a marginally better forecast (involving a smaller forecast error).²¹

Another way of comparing the forecasts is to perform an encompassing test, which tells us whether the forecast of one institute could be improved by using the other institute's forecast. The results, presented in Table 4, show that mostly the forecast of one institute indeed encompasses the forecast of the other institute. In particular, this is always true for GDP forecasts. In addition, all current-year IHS forecasts (with one exception) encompass the respective WIFO forecasts. In total, the IHS forecast does not encompass the WIFO forecast in four out of 56 cases, while the WIFO forecast does not encompass the IHS forecast in 17 out of 56 cases.

More important than the small differences between the forecasts is the common feature that all forecasts improve considerably over time (see Fig. 2 and Tables 1 and 2). This improvement is most distinct for the inflation rate and the unemployment rate, where the forecast errors are almost zero in September and December of the current year t . This result is to be expected, since inflation and unemployment data are published monthly, in a very timely fashion, and in the December forecast almost all of the realizations are known.

²¹ These conclusions rely on the Diebold–Mariano test using the squared forecast error as the loss function. If we take the absolute forecast error we do not find any difference at all between the two institutes when it comes to forecast accuracy.

Table 4 Encompassing test

		GDP	Consump.	Invest.	Exports	Imports	Inflation	Unempl.
IHS	Mar t-1	0.31	0.26	0.89	0.45	0.15	0.08	0.07
	Jun t-1	0.37	0.17	0.96	0.37	0.23	0.13	0.04*
	Sep t-1	0.71	0.24	0.41	0.33	0.79	0.48	0.09
	Dec t-1	0.11	0.37	0.01*	0.01*	0.08	0.47	0.06
	Mar t	0.27	0.08	0.29	0.02*	0.10	0.23	0.44
	Jun t	0.45	0.12	0.70	0.08	0.15	0.11	0.28
	Sep t	0.70	0.12	0.74	0.14	0.16	0.45	0.11
	Dec t	0.50	0.32	0.78	0.61	0.50	0.42	0.84
WIFO	Mar t-1	0.07	0.03*	0.20	0.03*	0.04*	0.04*	0.00*
	Jun t-1	0.08	0.05*	0.12	0.18	0.19	0.04*	0.00*
	Sep t-1	0.71	0.07	0.63	0.63	0.83	0.04*	0.01*
	Dec t-1	0.30	0.16	0.30	0.00*	0.03*	0.89	0.04*
	Mar t	0.18	0.12	0.43	0.00*	0.01*	0.73	0.04*
	Jun t	0.90	0.19	0.56	0.04*	0.19	0.89	0.14
	Sep t	0.95	0.37	0.61	0.08	0.12	0.43	0.22
	Dec t	0.81	0.27	0.80	0.25	0.17	0.50	0.95

The table lists p values for testing the null hypothesis that the IHS (and WIFO) forecast encompasses the forecast of the other institute. Starred figures indicate that the null hypothesis is rejected at the 5% significance level and hence the other institute contributes to explaining one's own forecast errors

Also with regard to Theil's U_2 , the improvement of the forecasts over time is clearly visible. As mentioned above, in contrast to the RMSE the absolute value of Theil's U_2 is meaningful. It should be below one, since only then is the forecast better than the naive no-change assumption. Theil's U_2 is above unity only in one case: namely the first IHS unemployment forecast. Especially for current-year forecasts, this measure is clearly below 1, particularly for the forecasts of the inflation rate and the unemployment rate, but also for the GDP growth forecast. The improvement is least distinct for the consumption growth forecast. This might be related to the fact that consumption is rather smooth over time. Hence, the growth rates do not fluctuate much over time, rendering the no-change assumption that is the benchmark of Theil's U_2 more difficult to beat.

Similarly to the other traditional measures, the difference between the two institutes with respect to getting the directional change right is rather small. Among all variables, inflation is assessed best: already at the beginning of the current year the directional change is projected correctly by IHS in more than 90% of all cases, and by WIFO in more than 80%. The directional change of GDP growth is forecast about equally well. The directions of import changes seem to be harder to predict: considering the March forecasts in year t , roughly 60% of all changes are anticipated correctly by IHS, and roughly 70% by WIFO.

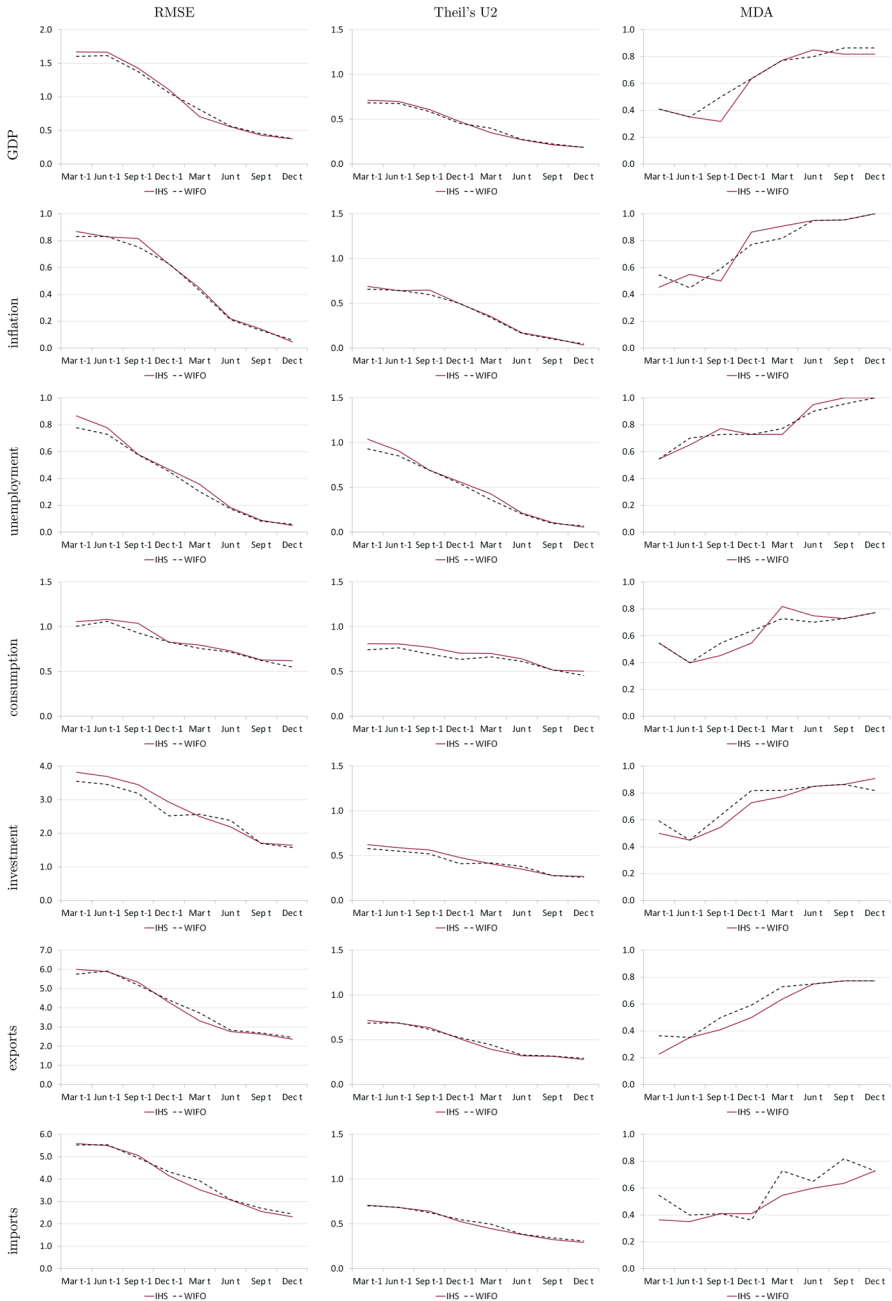


Fig. 2 RMSE, Theil's U2, and MDA for decreasing forecast horizon

4.1.2 Bias

The original and the modified Mincer–Zarnowitz tests show that in general the forecasts are neither too high nor too low (Tables 5, 6). Based on the original test, i.e., without taking the state of the business cycle into account, the growth forecasts of private consumption and exports, and the forecasts of the unemployment rate are the only ones in which the null hypothesis of no systematic forecast errors has to be rejected—and then only in a small number of cases. If the recession dummy is included, only some consumption growth forecasts show signs of being biased. Overall, the forecasters of IHS and WIFO do not seem to make any systematic errors, even when the state of the economy in the business cycle is taken into account.

Summarizing all results based on traditional, one-dimensional accuracy measures, we can draw the following conclusions: (i) all forecasts improve considerably over time, which is what one would expect; (ii) the forecasts published by the two institutes IHS and WIFO do not differ significantly from each other, except in two out of 56 cases; (iii) in most cases the forecast of one institute encompasses the forecast of the other institute; and (iv) there are basically no systematic forecast errors for the forecasts published in year t , including when the state of the economy is taken into account.

4.2 The Mahalanobis distance

We assess the joint forecasts of three different groups of variables. The first group includes all variables considered (hereafter called *All*), i.e., growth rates of real GDP, private consumption, real gross fixed capital formation, exports, and imports, as well as the inflation rate and the unemployment rate. The second group includes real GDP growth, the inflation rate, and the unemployment rate (hereafter called *Macro*), and the third group includes growth rates of real GDP, real gross fixed capital formation, private consumption, exports, and imports (hereafter called *Demand*). Very often forecasts of the *Macro* group are reported to provide a quick overview of the economic outlook.

4.2.1 Accuracy

Table 7 reports the joint evaluation results of IHS and WIFO forecasts using the Mahalanobis distance, considering all variables (*All*), the *Macro* group, and the *Demand* group. The differences between the two institutes are rather small. In each group the maximum difference between the two institutes is observed for the second earliest forecast, i.e., the June forecast published in year $t - 1$. This goes along with what one might expect: forecasts are usually more divergent the earlier they are produced. With a decreasing forecast horizon, forecasts normally move closer to the realized values (i.e., the forecast error decreases), and hence, the difference between the two institutes also shrinks, as more information becomes available. Overall, this observation is borne out in Fig. 3, which plots the evaluation results for the three groups of variables. IHS forecasts show a slightly larger joint forecast error at early forecast dates. By contrast, both institutes show nearly identical forecast errors for the *Macro*

Table 5 Mincer–Zarnowitz test for unbiased forecasts

	GDP		Consumpt.		Invest.		Exports		Imports		Inflation		Unempl.	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
Mar t-1	0.08	0.18	0.01*	0.02*	0.17	0.40	0.07	0.42	0.06	0.12	0.03*	0.09	0.02*	0.02*
Jun t-1	0.11	0.18	0.01*	0.02*	0.20	0.32	0.31	0.31	0.24	0.20	0.08	0.07	0.03*	0.02*
Sep t-1	0.59	0.73	0.02*	0.05	0.68	0.61	0.92	0.86	0.80	0.96	0.13	0.41	0.12	0.04*
Dec t-1	0.30	0.56	0.04*	0.09	0.48	0.29	0.04*	0.04*	0.13	0.35	0.81	0.70	0.09	0.08
Mar t	0.42	0.77	0.04*	0.10	0.48	0.90	0.03*	0.03*	0.25	0.21	0.43	0.96	0.30	0.11
Jun t	0.66	0.96	0.08	0.21	0.87	0.97	0.22	0.08	0.44	0.42	0.25	0.98	0.37	0.25
Sep t	0.80	0.91	0.22	0.47	0.78	0.92	0.19	0.10	0.23	0.25	0.85	0.74	0.08	0.23
Dec t	0.67	0.83	0.37	0.65	0.59	0.90	0.09	0.13	0.07	0.16	0.33	0.40	0.69	0.77

The table lists p values for testing the null hypothesis of no bias. Starred figures indicate that the null hypothesis of no bias is rejected at the 5% significance level

Table 6 Modified Mincer–Zarnowitz test for unbiased forecasts

	GDP		Consump.		Invest.		Exports		Imports	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
Mar t-1	0.06	0.14	0.03*	0.04*	0.17	0.42	0.09	0.40	0.11	0.18
Jun t-1	0.06	0.12	0.02*	0.04*	0.14	0.30	0.26	0.23	0.30	0.22
Sep t-1	0.57	0.69	0.05*	0.13	0.67	0.72	0.85	0.89	0.85	0.97
Dec t-1	0.49	0.75	0.10	0.19	0.67	0.44	0.09	0.09	0.25	0.49
Mar t	0.63	0.92	0.10	0.21	0.69	0.97	0.08	0.07	0.42	0.22
Jun t	0.84	0.92	0.18	0.37	0.90	0.99	0.35	0.13	0.56	0.33
Sep t	0.87	0.89	0.32	0.63	0.65	0.90	0.30	0.19	0.40	0.25
Dec t	0.76	0.85	0.52	0.75	0.73	0.97	0.19	0.25	0.14	0.29

The table lists *p* values for testing the null hypothesis of no bias, when the state in the business cycle is accounted for. Starred figures indicate that the null hypothesis is rejected at the 5% significance level

Table 7 Evaluation results: Mahalanobis distance

	Mahalanobis distance						Ratio with respect to minimum					
	<i>All</i>		<i>Macro</i>		<i>Demand</i>		<i>All</i>		<i>Macro</i>		<i>Demand</i>	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
Mar t-1	0.810	0.736	0.438	0.380	0.707	0.659	2.22	2.01	15.35	13.32	1.94	1.81
Jun t-1	0.901	0.710	0.493	0.356	0.795	0.631	2.47	1.94	17.28	12.48	2.18	1.73
Sep t-1	0.657	0.562	0.295	0.229	0.591	0.517	1.80	1.54	10.33	8.04	1.62	1.42
Dec t-1	0.586	0.504	0.191	0.221	0.532	0.482	1.60	1.38	6.71	7.74	1.46	1.32
Mar t	0.449	0.471	0.132	0.044	0.403	0.431	1.23	1.29	4.61	1.56	1.10	1.18
Jun t	0.476	0.455	0.058	0.040	0.450	0.389	1.30	1.24	2.03	1.41	1.23	1.07
Sep t	0.447	0.430	0.045	0.029	0.419	0.427	1.22	1.18	1.58	1.00	1.15	1.17
Dec t	0.480	0.365	0.039	0.038	0.479	0.365	1.31	1.00	1.36	1.32	1.31	1.00

The left panel of the table shows the Mahalanobis distance for the three different groups of variables (*All*, *Macro*, *Demand*) for the eight forecasts provided by IHS and WIFO. The right panel shows the ratio of that distance with respect to the minimum distance (of either IHS or WIFO) for a given group

group over the last three forecast dates. A bit surprisingly, the joint forecast errors of IHS for the group of all variables (*All*) and for the *Demand* group increase slightly from the penultimate to the last forecast, while the corresponding WIFO joint forecast errors decrease, as one would expect. Since this behavior is not visible in the univariate evaluation measures, it is probably inherent to the Mahalanobis distance, i.e., related to the covariances between the variables. For all forecast horizons, IHS and WIFO joint forecasts do not differ significantly from each other.²² Likewise, based on the univariate analysis presented in Sect. 4.1.1, we have concluded that the two institutes do not provide significantly different forecasts.

²² More precisely, the null hypothesis of equal means cannot be rejected, for any given forecast horizon. This is tested by considering the Mahalanobis distance between IHS and WIFO forecasts, for each forecast horizon (considering the implied *F*-distribution), see Sect. 3.2.

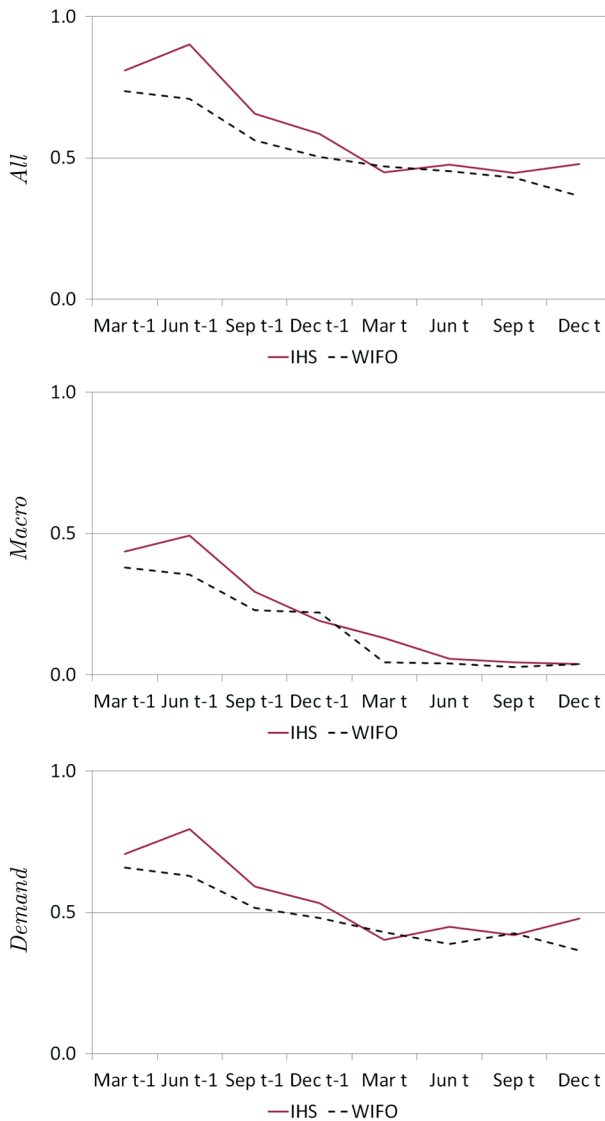


Fig. 3 Mahalanobis distance for decreasing forecast horizon. The figure shows the Mahalanobis distance for a decreasing forecast horizon for three different groups of variables (*All*, *Macro*, *Demand*)

The three graphs in Fig. 3 show that the joint forecasts clearly improve over time, and this is particularly true for the *Macro* group. The reason is probably the relatively high accuracy (small forecast errors) of the inflation and unemployment forecasts, already pointed out for the univariate results, which account for two of the three variables in that group. For the *Macro* group, the joint forecast errors implied by the latest forecasts are smaller than the errors implied by the first forecasts by a factor larger than ten. For the other two groups of variables (*All* and *Demand*), this factor is equal to or below two, and hence, the improvement over time is less pronounced (see Table 7).

Table 8 Unbiasedness of joint forecasts

	<i>All</i>		<i>Macro</i>		<i>Demand</i>	
	IHS	WIFO	IHS	WIFO	IHS	WIFO
Mar t-1	no (5)	no (4)	no (2)	no (2)	no (2)	no (1)
Jun t-1	no (5)	no (5)	no (3)	no (2)	no (3)	no (1)
Sep t-1	no (6)	no (5)	no (2)	no (1)	no (3)	no (2)
Dec t-1	no (5)	no (4)	no (1)	no (2)	yes	no (1)
Mar t	no (3)	no (4)	yes	no (1)	no (1)	no (2)
Jun t	no (1)	yes	no (2)	yes	yes	yes
Sep t	yes	no (1)	yes	no (1)	yes	yes
Dec t	yes	yes	yes	yes	yes	yes

The results are based on the Bonferroni–Holm procedure, a simple multiple-level α test, at a global significance level of 5%. Figures in brackets denote the number of rejections, where the total number of individual hypotheses is equal to 21 (*All*), 9 (*Macro*), and 15 (*Demand*)

4.2.2 Bias

In order to examine whether joint forecasts are unbiased, a fairly large number of null hypotheses need to be tested. The concrete number ($m \times 3$) depends on the number of variables considered in the joint forecast and is thus equal to 21, 9, and 15 in our case. Table 8 presents the results. Only for the most recent forecast, i.e., the December t forecast, are the joint forecasts unbiased for all the groups under consideration (*All*, *Macro*, *Demand*) for both Austrian research institutes. This means that in a VAR(1) system of the forecast errors implied by the variables under consideration, neither the constant terms nor the lagged errors are significant, and in addition none of the given errors is significantly Granger-caused by the other errors. Further, the September t forecasts are largely unbiased, and in the case of the IHS all forecasts are unbiased. If a joint forecast is not unbiased, we report the number of cases in which the null hypotheses are rejected. For all current-year forecasts, the maximum number of individual hypotheses rejected is three (IHS) or four (WIFO), out of a total of 21 cases (*All*).

We provide more detailed results in Tables 1 to 4 in the Online Appendix in order to be able to assess the precise source(s) of the bias. For example, the single rejection for the WIFO September t forecasts in the groups *All* and *Macro* originates from the unemployment forecast error being Granger-caused by the other forecast errors (see Tables 2 and 3 in the Online Appendix). In fact, it is nearly always true that the source of the bias can be found in a given forecast error being Granger-caused by the other forecast errors.

4.3 Robustness checks

We perform two sets of robustness checks. First, we exclude the year 2009 from our analyses since during this year of the “Great Recession” GDP in Austria dropped by 3.8%. The declines of fixed capital formation and exports were even more severe, at

7.2% and 14.4%, respectively. No forecasts predicted this extreme economic downturn during most of the year 2008. Second, we use another benchmark with which we compare our forecasts, namely the first estimate of annual national accounts published in March of year $t + 1$. We use this release only for the robustness check, and not as the benchmark, for two reasons. First, the annual accounts are based on a larger information set and are thus usually subject to smaller revisions than the first estimate of March. Second, the March release is produced by WIFO, i.e., one of the two forecasts institutes, on behalf of the Austrian statistical office, while the release published in late summer is produced by the statistical institute itself. Of course, inflation and unemployment figures are usually not subject to revision, unless methodological changes are implemented.

The results for both robustness checks are very similar to our main results. In particular, there are no systematic deviations, and thus, the conclusions remain unaffected. In contrast to our study, Baumgartner (2002a, b) uses the first estimate of the annual national accounts, provided by WIFO, as a benchmark. But he performs the analysis also with the realizations taken from the national accounts produced by Statistik Austria and finds, as we do, that there are hardly any differences in the results.

4.4 Comparison with European Commission forecasts

In this section, we compare the economic forecasts of IHS and WIFO with the forecasts for Austria made by the European Commission, which produces economic forecasts twice a year.²³ The EC publication dates, spring forecasts in May and autumn forecasts in November, differ from those of IHS and WIFO. As a consequence, the information available for national and international forecasters is not the same, which complicates a proper forecast comparison. If, for example, the national March issues are compared with the EC spring issues, then the EC has an information advantage and should, *ceteris paribus*, provide more accurate forecasts. If, on the other hand, the EC spring issues are compared with the national June issues, then IHS and WIFO have an information advantage.²⁴

If there is no systematic difference in the quality of forecasting across national and international institutions, then the forecasts should improve with a decreasing forecast horizon. Note that this section excludes inflation and unemployment from the analysis, since national and international forecasters use different definitions.²⁵

Figure 4 shows the different measures of accuracy for the national and international forecasts of GDP growth for different forecast horizons. We have twelve forecast dates in total, eight from the national institutes and four from the European Commission.

²³ A few years ago the EC started to publish an additional winter forecast. The winter forecasts are not considered in our analysis due to the short time series.

²⁴ Baumgartner (2002b) investigates the forecast accuracy of the two national institutes (IHS and WIFO) and OECD by grouping the national and international forecasts in two alternative ways (and performing the analysis twice). One grouping reflects an informational advantage for the national institutes; the other grouping an informational advantage for the international institute (OECD).

²⁵ IHS and WIFO use the national definitions of inflation (CPI) and the unemployment rate (Claimant Count based on administrative data from the benefits system), while the EC uses harmonized measures (HICP and LFS, respectively).

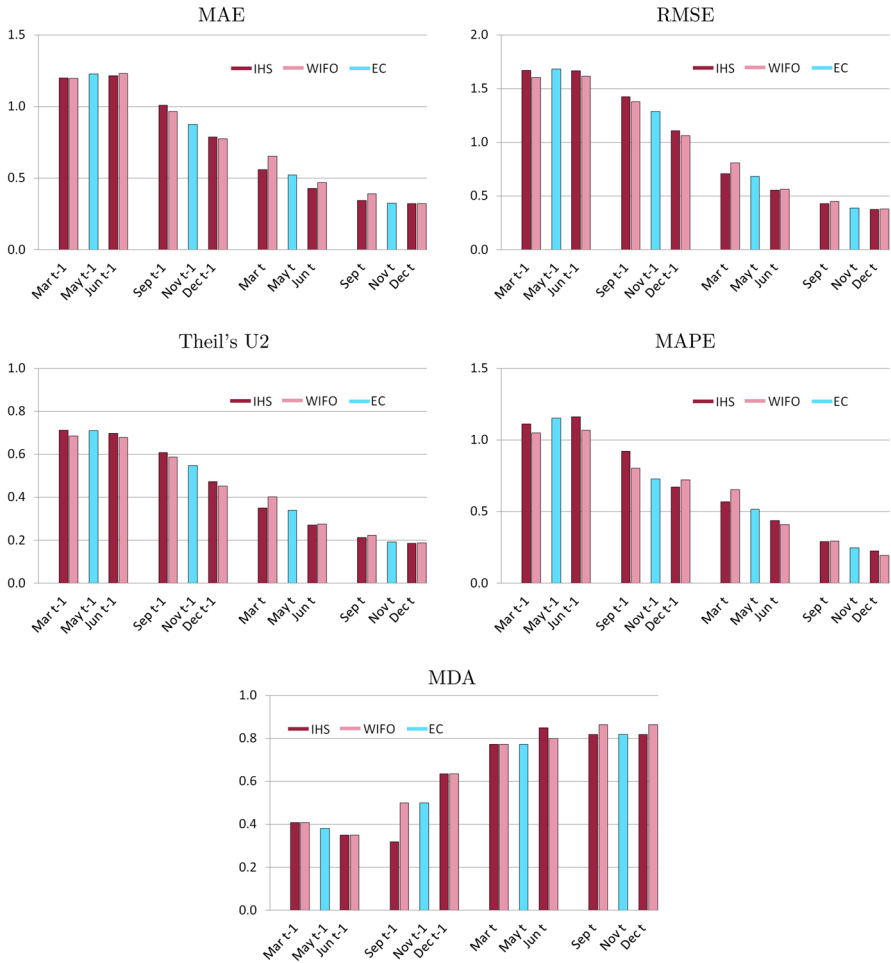


Fig. 4 Measures of accuracy for different dates of forecasts and different forecasting institutions for GDP growth

As we can see from the graphs, the first three GDP growth forecasts that we consider, i.e., the year-ahead forecasts in March, May, and June, are roughly the same and do not clearly improve over time. Similarly, the last three forecasts, i.e., the current-year forecasts of September, November, and December, do not seem to get better over time. By contrast, the GDP growth forecasts do seem to get more accurate from the year-ahead forecasts in September through the current-year forecasts in September. This general observation holds more or less across all the different measures of accuracy that we evaluated. More formally, as shown in Table 9, the Diebold–Mariano test for two successive forecasts²⁶ confirms our impression. The year-ahead forecasts in September, November, and December are significantly better

²⁶ The null hypothesis of the test states that the forecast accuracy is the same for both forecasts, while the alternative hypothesis says that the more recent forecast is more accurate. Note that this is slightly

Table 9 Diebold–Mariano test for successive national and international GDP forecasts

Compared forecasts	GDP		Consumption		Investment		Exports		Imports	
	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO	IHS	WIFO
Mar t-1, May t-1 (EC)	0.392	0.881	0.337	0.914	0.310	0.950	0.288	0.815	0.448	0.689
May t-1 (EC), Jun t-1	0.264	0.150	0.577	0.204	0.230	0.024*	0.309	0.374	0.187	0.235
Sep t-1, Nov t-1 (EC)	0.090*	0.057*	0.034*	0.100*	0.099*	0.215	0.101	0.035*	0.080*	0.008*
Nov t-1 (EC), Dec t-1	0.065*	0.093*	0.852	0.440	0.245	0.072*	0.207	0.201	0.223	0.273
Mar t, May t (EC)	0.356	0.101	0.074*	0.182	0.526	0.341	0.063*	0.063*	0.050*	0.034*
May t (EC), Jun t	0.050*	0.117	0.555	0.201	0.011*	0.160	0.532	0.672	0.814	0.872
Sep t, Nov t (EC)	0.147	0.059*	0.168	0.203	0.383	0.367	0.134	0.057*	0.355	0.077*
Nov t (EC), Dec t	0.347	0.400	0.692	0.230	0.443	0.267	0.604	0.732	0.309	0.512

The table lists p values for testing the null hypothesis that successive forecasts of GDP growth show the same accuracy against the alternative hypothesis that the more recent forecast is more accurate, where the loss function is the squared forecast error. Starred figures indicate that the null hypothesis is rejected at the 10% significance level and hence that the more recent forecast is more accurate

than their preceding counterparts (for both IHS and WIFO), and the same applies for the current-year September forecast (for both IHS and WIFO), the current-year June forecast (IHS), and the current-year November forecast (against the preceding WIFO forecast).

We observe a similar pattern in the quality of forecasts for the remaining variables, i.e., for the growth rates of consumption, investment, exports, and imports. For those variables as well, the forecasts do not considerably improve over the period of the first forecasts (March to June year-ahead forecasts) or over that of the last forecasts (September to December current-year forecasts), while they usually do improve over the remaining period (year-ahead September forecasts to current-year September forecasts), see Fig. 5 and Table 9.

Overall, this comparison between the IHS, WIFO, and EC forecasts corroborates our findings that the accuracy of the forecast depends much more on the time of publication, i.e., on the data available when preparing the forecast, than on the question of which institution publishes the forecast.

4.5 Effects of external assumptions

Macroeconomic forecasts are usually conditional on assumptions about the international economic environment, such as world trade, GDP growth for the main trade partners, the oil price, exchange rates, and monetary and fiscal policies. This is particularly true for small open economies like Austria's.²⁷ Forecast errors may then result from wrong external assumptions, a poor forecast model, or both. In addition, revi-

Footnote 26 continued

different from the Diebold–Mariano tests performed before, when the alternative hypothesis was that the two competing forecasts were different.

²⁷ Note, for example, that in 2017 the external trade to GDP ratio of Austria amounted to 104%, while it was 27% in the USA.

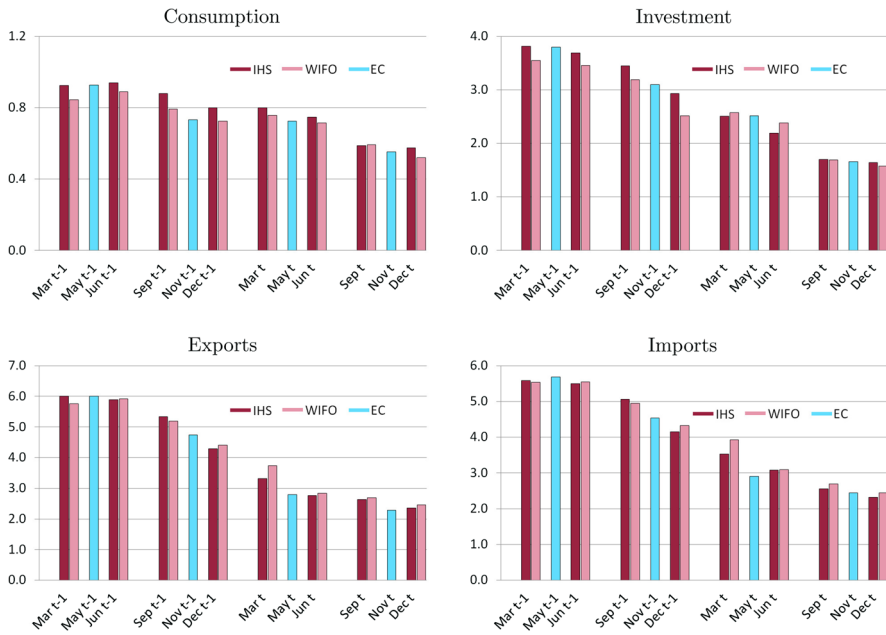


Fig. 5 RMSE for different dates of forecasts and different forecasting institutions for the growth of consumption, investment, exports, and imports

sions of national accounts data may be relevant. This section investigates the effects of errors in external assumptions of this sort. Since such errors usually propagate to the forecast errors, the following question arises: To what extent can forecast errors be explained by errors involving external assumptions?

In order to answer this question, we perform a regression analysis in the spirit of Keereman (2003), Fioramanti et al. (2016), and the European Commission (2016) to determine the influence of deviations of external assumptions on the forecast error of Austrian GDP growth. We consider the effect of an unexpected change in the growth rate of GDP of the European Union, the oil price, and the foreign exchange rate. A similar set of regressions is performed for the inflation forecast for Austria, with a view to answering the question whether and to what extent mistakes in external assumptions translate into inflation forecast errors. The analyses are based on 17 to 22 observations, depending on the forecasting horizon and the institute.

Table 10 summarizes the results with respect to prediction errors in the growth rate of GDP. We find a positive influence of EU GDP errors on the forecast error of the Austrian GDP, which is highly significant for the first five forecasts. With regard to the positive sign, this is what we would expect, since an overestimation of external growth should lead to a higher than realized national growth rate. With regard to oil prices our expectation about the sign is ambiguous. Overestimated oil prices might lead, on the one hand, to an underestimation of growth due to overestimated import prices. On the other hand, overestimated oil prices might just reflect the overestimation of the state of the global economy. The latter is what almost all our results show, in particular when the coefficients are significant. We find little evidence that mistakes in the oil price

Table 10 Effects of external assumptions on forecast errors for GDP

		Coefficients				P values				R^2	Obs.	F-test
		Const.	GDP _{EU}	Oil	FX	Const.	GDP _{EU}	Oil	FX			
IHS	Mar t-1	0.05	0.83*	0.02*	-0.04	0.753	0.000	0.006	0.14	0.92	17	51.7*
	Jun t-1	0.09	0.88*	0.02*	-0.01	0.601	0.000	0.035	0.74	0.90	17	38.9*
	Sep t-1	-0.03	0.89*	0.02	-0.03	0.866	0.000	0.052	0.40	0.86	18	27.6*
	Dec t-1	0.05	0.96*	0.01	0.00	0.735	0.000	0.213	0.92	0.79	20	20.1*
	Mar t	-0.03	0.87*	0.00	-0.02	0.801	0.001	0.976	0.50	0.50	21	5.6*
	Jun t	0.04	0.51	0.00	0.07	0.744	0.076	0.734	0.19	0.43	20	4.0*
	Sep t	0.00	0.53	-0.02	0.10	0.996	0.116	0.257	0.45	0.20	21	1.4
	Dec t	0.01	0.57	-0.03	0.05	0.874	0.140	0.295	0.80	0.26	21	2.0
WIFO	Mar t-1	-0.09	0.92*	0.01	0.01	0.503	0.000	0.142	0.46	0.90	22	54.3*
	Jun t-1	-0.05	0.87*	0.01	0.02	0.774	0.000	0.240	0.20	0.87	21	36.8*
	Sep t-1	-0.02	0.87*	0.01	0.01	0.915	0.000	0.147	0.63	0.82	22	27.3*
	Dec t-1	-0.04	0.95*	0.01	0.01	0.804	0.000	0.222	0.64	0.74	22	17.1*
	Mar t	-0.07	1.04*	0.00	0.00	0.621	0.000	0.988	0.86	0.55	22	7.4*
	Jun t	-0.04	0.38	0.00	0.08	0.747	0.256	0.874	0.14	0.39	21	3.7*
	Sep t	0.01	0.50	0.02	0.03	0.958	0.224	0.242	0.67	0.24	22	1.9
	Dec t	-0.04	0.56	-0.03	0.07	0.649	0.152	0.202	0.65	0.24	22	1.9

The table shows the results of regressing the forecast error of the Austrian GDP growth rate on unexpected changes in the growth rate of the EU GDP, the growth rate of the oil price, and the growth rate of the foreign exchange rate (euro versus US dollar). Starred figures indicate significance at the 5% level

prediction explain GDP forecast errors and no evidence at all for unexpected changes in the exchange rate. Note that about 90% of the variation of the GDP forecast error can be explained by the deviations in the external assumptions for the first (i.e., the March $t - 1$) forecast. This number drops to about 50% for the fifth (i.e., the March t) forecast. Note, in addition, that the results across the two institutes are very similar. A robustness check which excludes the crisis year 2009 shows no major changes in the results.

Table 11 shows the corresponding results for the effect of external assumption on the inflation forecasts. As anticipated, we find that unexpected changes in the oil price do indeed explain part of the inflation forecast error. If oil prices are thought to increase more, inflation is overestimated as well, and vice versa. However, for current-year forecasts the statistical significance either disappears (IHS) or becomes weaker (WIFO). In the case of the inflation forecast errors, roughly 50% to 60% of the variation can be explained by deviations in external assumptions for the first three forecasts. By contrast, we do not find any evidence for the propagation of mistakes in the external assumptions concerning GDP growth and exchange rates. Again, a robustness check excluding the crisis year 2009 yields only very minor changes in the analysis.

While other studies²⁸ consider the effects of external assumptions only for one current-year forecast and one year-ahead forecast, this analysis provides a more

²⁸ Keereman (2003), Fioramanti et al. (2016), and European Commission (2016).

Table 11 Effects of external assumptions on forecast errors for inflation

		Coefficients				P values				R^2	Obs.	F-test
		Const.	GDP _{EU}	Oil	FX	Const.	GDP _{EU}	Oil	FX			
IHS	Mar t-1	0.00	0.06	0.02*	-0.01	0.997	0.545	0.010	0.79	0.56	17	5.6*
	Jun t-1	-0.10	0.09	0.02*	-0.01	0.579	0.416	0.014	0.83	0.55	17	5.3*
	Sep t-1	-0.13	0.11	0.03*	-0.03	0.403	0.246	0.001	0.37	0.65	18	8.5*
	Dec t-1	-0.06	0.16	0.02*	0.01	0.700	0.250	0.035	0.73	0.29	20	2.1
	Mar t	-0.08	0.00	0.01	0.00	0.495	0.985	0.400	0.97	0.04	21	0.3
	Jun t	-0.04	-0.16	0.01	0.00	0.419	0.223	0.123	0.87	0.17	20	1.1
	Sep t	0.00	-0.04	0.01	0.02	0.978	0.727	0.156	0.56	0.32	21	2.7
	Dec t	-0.01	0.01	0.00	0.03	0.382	0.893	0.489	0.20	0.11	21	0.7
WIFO	Mar t-1	0.03	0.08	0.02*	0.00	0.844	0.414	0.002	0.79	0.51	22	6.3*
	Jun t-1	0.09	0.10	0.02*	0.00	0.568	0.276	0.001	0.76	0.59	21	8.1*
	Sep t-1	0.07	0.10	0.02*	-0.01	0.638	0.288	0.001	0.61	0.52	22	6.5*
	Dec t-1	0.06	0.06	0.02*	-0.01	0.713	0.673	0.025	0.51	0.27	22	2.2
	Mar t	0.00	-0.06	0.01	-0.02	0.964	0.716	0.282	0.26	0.14	22	0.9
	Jun t	0.03	-0.16	0.01*	0.00	0.489	0.148	0.001	0.98	0.52	21	6.1*
	Sep t	0.00	-0.11	0.01*	0.01	0.897	0.326	0.012	0.79	0.35	22	3.3*
	Dec t	-0.02	0.03	0.00	0.02	0.241	0.695	0.926	0.37	0.05	22	0.3

The table shows the results of regressing the forecast error of the Austrian inflation on unexpected changes in the growth rate of the EU GDP, the growth rate of the oil price, and the growth rate of the foreign exchange rate (euro versus US dollar). Starred figures indicate significance at the 5% level

detailed view of the influence of the external assumptions over different forecast horizons. This is due to the eight forecasts under analysis, which reflect an increasing information set and a shrinking forecast horizon. For both target variables, GDP growth and inflation, we find that the impact of mistakes in the external assumptions is rather strong for year-ahead forecasts but cannot be observed, or only weakly observed, for current-year forecasts. This is true for both institutes. These findings are in line with the literature previously mentioned. In Keereman (2003), for example, for 10 out of 12 countries the unexpected change in US GDP shows a higher significance level (lower p value) in explaining the year-ahead GDP forecast error than in explaining the current-year forecast. In Fioramanti et al. (2016), for the year-ahead forecasts the GDP forecast error in the base model and almost all additional models can be explained by errors in the external assumptions with respect to world GDP, while this is not the case for the current-year forecasts. With regard to unexpected changes in the oil price, the case is less clear. However, among the six models presented in Fioramanti et al. (2016) the year-ahead forecasts tend to result in lower p values and higher R^2 than those in the current year. Our findings are also in line with the results of Fenz et al. (2019), who use a different methodology to examine the forecast errors for Austrian GDP growth. They find that the variance of forecast errors can be explained largely by global and euro area shocks (91%), and only to a small degree by national shocks (9%) for year-ahead forecasts. This picture is nearly reversed for current-year fore-

casts, where 42% of the variance can be explained by national shocks and 58% by global and euro area shocks.

Our findings, as well as those in the related literature, suggest that the structure (decomposition) of the forecast error changes with the forecast horizon. In particular, assumptions with respect to the international environment seem to be much more important for longer (year-ahead) than for shorter (current-year) forecast horizons and additional factors not considered in our analysis are probably more important for current-year forecasts. One possible explanation for this change in structure could be that for the current-year forecasts data revisions play a more important role than for the year-ahead forecasts. This hypothesis is supported by the results in Fenz et al. (2019).

One implication of our analysis is that forecasters in small open economies like Austria's should focus more on making correct assumptions about the external environment, like GDP growth for important trade partners, oil prices, or US and EU monetary policy measures, if they want to reduce their forecast errors for GDP or inflation.

5 Conclusion

In this paper, we evaluate macroeconomic forecasts for Austria, published by the two leading Austrian economic research institutes, the Institute for Advanced Studies (IHS) and the Austrian Institute of Economic Research (WIFO). We evaluate the forecasts of growth rates of real GDP, private consumption, gross fixed capital formation, exports, and imports as well as the inflation rate and the unemployment rate. For each variable, we evaluate the year-ahead and the current-year forecasts published in March, June, September, and December. The analyses are based on traditional, univariate measures like the RMSE, Theil's inequality coefficient, and the mean directional accuracy, as well as on a more novel, multivariate measure, the Mahalanobis distance. The latter assesses jointly a group of variables and thereby takes the variances of and the correlations between these variables into account. Furthermore, we compare the forecasts of the two Austrian institutes with forecasts of the European Commission, taking into account the different publication dates. Finally, we examine how errors in the external assumptions affect forecast errors for GDP growth and inflation.

Our first finding is that the forecasts of the two Austrian economic research institutes are very similar. Considering both univariate and multivariate forecast evaluation measures, we basically do not find any significant differences between the two institutes (for any variable, any group of variables, or any forecast horizon). The only exceptions are the December year-ahead and the December current-year forecast of consumption growth, where WIFO seems to outperform IHS. If we examine the question of whether one institute's forecast can be improved by using the other institute's forecast, we conclude that mostly this is not the case; i.e., the forecast of one institute usually encompasses the forecast of the other institute. However, it happens a bit more often that WIFO forecasts do not encompass IHS forecasts than the other way around.

Our second finding is that the forecasts improve significantly over time, which is what one would expect; however, the improvement is usually less pronounced in

the first two and the last two forecasts. This is true both for the univariate measures and the Mahalanobis distance, and it particularly applies to the inflation rate and the unemployment rate, and to the *Macro* group (GDP, inflation, and unemployment). This pattern is probably due to the fact that the inflation and unemployment rates are published at a monthly frequency, in a very timely manner, and are usually not revised, so a lot of information is known to the forecaster toward the end of the year. The larger forecast errors at longer forecast horizons show that policy makers should be cautious when basing their budgetary planning on such forecasts. With respect to the Mahalanobis measure, a complementary analysis would be helpful, which allows to precisely disentangle the univariate forecast errors from consistency errors.

Our third finding relates to the unbiasedness of forecasts (univariate measures), where we also take explicit account of recessions, and the existence of systematic errors (Mahalanobis distance). We find that forecasts of GDP, investment, and import growth are always unbiased (for any given forecast horizon and for both institutes), and that all other forecasts are largely unbiased if they are published in the current year. In addition, we do not find any systematic errors in the forecasts published in December t (for any given group and for both institutes). Considering the September t forecasts, the IHS forecasts do not show any systematic errors (for any given group), while WIFO forecasts show errors in the group of all variables and in the *Macro* group.

With respect to the comparison of IHS and WIFO forecasts with those of the European Commission, we point out that any direct comparison may be flawed due to the different sets of information available to the national and international forecasters at the time of forecasting. We find, in general, an improvement in accuracy with a decreasing forecast horizon, except for the first three and the last three forecasts. This observation does not depend on whether a given forecast is produced by IHS, by WIFO, or by the EC. These results show that the forecast accuracy depends much more on the dataset available when preparing the forecast than on the question of which institute publishes the forecast. In this context, it would be useful to have methods that take explicit account of the different sets of information available to forecasters; this is left for future research.

Finally, we find that errors in external assumptions with respect to EU GDP growth translate into forecast errors for Austrian GDP growth, in particular for year-ahead forecasts. Similarly, mistakes in external assumptions with respect to the oil price are reflected in the forecast errors for inflation. All results regarding external assumptions are very similar across IHS and WIFO. This implies that forecasters in small open economies like Austria's should focus more on making correct assumptions about the external environment, like GDP growth for important trade partners, oil prices, or US and EU monetary policy measures, if they want to improve their forecasts, in particular for the year ahead. To what extent data revisions contribute to forecast errors is a question left for further research.

Acknowledgements Open access funding provided by Institute for Advanced Studies Vienna. Earlier versions of this paper were presented at the 36th Annual International Symposium on Forecasting in Santander, Spain, June 21, 2016; at the 1st Vienna Workshop on Economic Forecasting, February 15, 2018; and at the Jour Fixe Budget- und Steuerpolitik, WIFO, October 24, 2018. We thank participants in the conferences and seminars, Helmut Hofer from IHS, and Roland Döhrn from RWI - Leibniz Institute for Economic

Research, as well as two anonymous referees for valuable comments. In addition, we thank Marliese Wolf and Clara Zwettler, both IHS, for their help with data preparation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights The study was not externally funded, and it did not involve any experiments with animals or humans.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


References

- Alt R, Fortin I, Weinberger S (2011) The Monday effect revisited: an alternative testing approach. *J Empir Finance* 18:447–460
- An Z, Jalles JT, Loungani P (2018) How well do economists forecast recessions? IMF Working Paper, WP/18/39
- Baumgartner J (2002a) Die Wirtschaftsprognosen von WIFO und IHS. Eine Analyse für die achtziger und neunziger Jahre. In: WIFO (ed), WIFO-Monatsberichte 11/2002, pp 701–716
- Baumgartner J (2002b) Evaluation of macro-economic forecasts for Austria in the 1980s and 1990s. *Aust Econ Q* 4/2002:191–206
- Chen Q, Costantini M, Deschamps B (2016) How accurate are professional forecasts in Asia? Evidence from ten countries. *Int J Forecast* 32:154–167
- Chong YY, Hendry DF (1986) Econometric evaluation of linear macro-economic models. *Rev Econ Stud* 53(175):671–690
- Döhrn R (2015) Der Prognostiker des Jahres – ein Zufallsergebnis? Möglichkeiten einer mehrdimensionalen Evaluierung von Konjunkturprognosen. IBES Diskussionsbeitrag, No. 208
- European Commission (2016) European economic forecast. Winter 2016. EC Directorate-General of Economic and Financial Affairs European Economy Institutional Paper 20 (Box 1.2: Forecast accuracy revisited, pp 49f)
- Fenz G, Ragacs C, Schneider M, Vondra K (2019) A structural decomposition of forecast errors for Austrian GDP growth. OeNB, Mimeo
- Fioramanti M, González Cabanillas L, Roelstraete B, Ferrandis Vallterra S (2016) External assumptions, the international environment and the track record of the Commission forecasts, European Commission DG ECFIN Discussion Paper 27
- Giusto A, Piger J (2017) Identifying business cycle turning points in real time with vector quantization. *Int J Forecast* 33:174–184
- Hochberg Y, Tamhane AC (1987) Multiple comparison procedures. Wiley, New York
- Holden K, Peel DA (1990) On testing for unbiasedness and efficiency of forecasts. *Manch Sch* 48:120–127
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70
- Hsu JC (1996) Multiple comparisons. Chapman & Hall, London
- Keereman F (1999) The track record of the Commission forecasts, European Commission DG ECFIN Economic Papers 137
- Keereman F (2003) External assumptions, the international environment and the track record of the Commission forecasts, European Commission DG ECFIN Economic Papers 189
- Kirby S, Paramaguru K, Warren J (2015) The accuracy of NIESR'S GDP growth forecasts. *Natl Inst Econ Rev* 232:F41–F47

- Kovacs K, Boulrier B, Stekler H (2017) Nowcasting: identifying German cyclical turning points. *J Econ Stat* 237:329–341
- Mincer J, Zarnowitz V (1969) The evaluation of economic forecasts. In: Mincer J (ed) *Economic forecasts and expectations*. National Bureau of Economic Research, New York, pp 3–46
- OECD (2014) *OECD forecasts during and after the financial crisis: a post mortem*, OECD Economics Department Policy Notes 23
- Ragacs C, Schneider M (2007) Comparing the predictive accuracy of macroeconomic forecasts for Austria from 1998 to 2006. In: OeNB (ed) *Monetary Policy & the Economy*, Q4/2007, pp 29–49
- Schuster P (2018) *Evaluation of economic forecasts for Austria for the years 2005 to 2017*, Fiscal Advisory Council
- Sheng X (2015) Evaluating the economic forecasts of FOMC members. *Int J Forecast* 31:165–175
- Sinclair TM, Joutz F, Stekler HO (2010) Can the Fed predict the state of the economy? *Econ Lett* 108(1):28–32
- Sinclair TM, Stekler HO (2013) Examining the quality of early GDP component estimates. *Int J Forecast* 29:736–750
- Sinclair TM, Stekler HO, Carnow W (2012) A new approach for evaluating economic forecasts. *Econ Bull* 32(3):2332–2342
- Sinclair TM, Stekler HO, Carnow W (2015) Evaluating a vector of the Fed's forecasts. *Int J Forecast* 31:157–164
- Sinclair TM, Stekler HO, Muller-Droge HC (2016) Evaluating forecasts of a vector of variables: a German forecasting competition. *J Forecast* 35:493–503
- Theil H (1966) *Applied economic forecasting*. North-Holland, Amsterdam

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Ines Fortin¹  · Sebastian P. Koch¹  · Klaus Weyerstrass¹ 

Ines Fortin
fortin@ihs.ac.at

Sebastian P. Koch
koch@ihs.ac.at

¹ Macroeconomics and Economic Policy Research Group, Institute for Advanced Studies (Institut für Höhere Studien, IHS), Vienna, Austria