

Strunz, Franziska; Gödl, Maximilian

Conference Paper

An Evaluation of Professional Forecasts for the German Economy

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage"

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Strunz, Franziska; Gödl, Maximilian (2023) : An Evaluation of Professional Forecasts for the German Economy, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage", ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/277707>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

An Evaluation of Professional Forecasts for the German Economy*

Maximilian Gödl & Franziska Strunz
University of the Bundeswehr Munich

September 10, 2023
Preliminary draft - do not cite

Abstract

We evaluate the performance of professional forecasts of GDP growth and CPI inflation in Germany using a long time series from 1968 to 2022. We document that professional forecasts are biased, skewed and have excess kurtosis. The extent of the bias, skewness, and kurtosis depends on the the forecast horizon, the variable of interest, and on the time of the release of the forecast. We compare professional forecasts to standard time series models. While we find that professional forecasts with forecast horizons of less than four quarters outperform our benchmark time-series models in terms of root mean squared errors, the same is not true for forecasts at longer horizons. Additionally, we highlight the importance of accounting for the forecasters' information set when evaluating their performance. We find forecasts revisions for both variables to depend on past revisions, while recessions only affect real GDP growth forecast revision and lead to larger revisions and a faster updating during recession.

Keywords: Forecasting and Prediction Methods; Business Cycle Forecasting; Inflation Forecasting; Time-Series Models;

JEL codes: C22, C53, E37

*We would like to thank Jonas Dovern, Johannes Pfeifer and participants of the Bavarian Macro Day workshop for valuable feedback. Correspondence may be addressed to: Maximilian Gödl, Universität der Bundeswehr. E-mail: maximilian.goedl@unibw.de; Franziska Strunz, Universität der Bundeswehr. E-mail: franziska.strunz@unibw.de.

1 Introduction

Germany has a multitude of economic research institutes that regularly publish professional forecasts of key economic variables, such as output growth, inflation, and unemployment. In addition, international organizations, like the Organization for Economic Co-operation and Development (OECD) and the European Commission (EC), also produce macroeconomic forecasts for Germany. These forecasts often receive substantial coverage in the news and play a large role in the decision making process by firms and policy makers. Therefore the question how well these forecasts incorporate information and predict future developments is of high importance.

In this paper we evaluate the quality of professional economic forecasts for the German economy. We rely on a comprehensive dataset of forecasts for two economic variables assembled by the Halle Institute for Economic Research (IWH). This dataset comprises 15 professional forecasters and covers a period from 1969 until 2022. We focus on real GDP growth and CPI inflation, which tend to be the most closely watched forecasts.

We assess the performance of professional forecasters against forecasts obtained from simple time series models estimated on a rolling basis using real-time data based on root mean square error performance. We also analyze whether forecasts are efficiently incorporating information, i.e. whether forecast errors can be predicted from past forecast errors. Additionally, we assess how business cycle fluctuations affect real GDP growth and CPI inflation forecast errors.

We highlight four findings:

1. Overall, forecasts for real GDP growth and CPI inflation are biased. Professional forecasters tend to overestimate growth on average and underestimate inflation. The bias depends on the forecast horizon. Forecasts released within the last halfyear towards the end of the target year are not found to be significantly different from zero, while this changes for higher horizons. The distributions of forecast and nowcast errors relating to GDP growth and CPI inflation are skewed and have excess kurtosis. Forecasters seem to have been too optimistic about GDP growth more often than they were too pessimistic. In particular, forecasters failed to anticipate some severe recessions, which explains the skewness as well as the fat tails of the forecast error distributions.
2. Forecast accuracy critically depends on the forecast horizon. Root mean squared errors (RMSEs) increase with the forecast horizon, meaning that forecasts close to the target year tend to be more accurate. Meanwhile we cannot detect a tendency

for forecasts to have improved over the last decades in terms of RMSE. Instead, RMSEs are tightly connected to the standard deviations of GDP growth and CPI inflation.

3. A comparison of professional forecasts to univariate and bivariate time series models yields mixed results. While the forecasts of GDP growth and CPI inflation at a short horizon clearly outperform the time-series benchmarks, the same is not generally true for longer horizons. This is surprising since, at least in principle, professional forecasters have access to more information than our benchmark econometrician who only relies on real-time observations of one or two variables.
4. Consistent with the previous literature (e.g. Coibion and Gorodnichenko (2015) and Dovern, Fritsche, Loungani, and Tamirisa (2015)), we find evidence for information rigidities. Past forecast revisions are informative about future revisions and errors, which indicates weak inefficiency. Additionally, in line with findings by Loungani, Stekler, and Tamirisa (2013) recessions lead to larger forecast revisions with respect to GDP growth forecasts and a faster acquisition of information. For CPI inflation we do not find revisions to depend on the state of the business cycle.

These findings are generally in line with what has been previously found in the literature. There is a long-standing interest in evaluating the accuracy of economic forecasts. For instance, in an early contribution Cowles (1933) asked “Can Stock Market Forecasters Forecast?”, and came to a rather pessimistic conclusion. Since then, several studies have looked at differences across forecasters and their relative performance. For example, Döhrn and Schmidt (2011) find that forecasting accuracy highly depends on the information set available at the time of forecast, while differences across individual forecasters (i.e. individual views and institutional differences) appear to be insignificant. Similarly, Qu, Timmermann, and Zhu (2019) examine with respect to Bloomberg survey data on US economic developments whether any single forecaster is superior to others. They are unable to establish such superiority. Regarding changes over time, Heilemann and Stekler (2013) conclude that forecast performance has not improved consistently with time, but instead seems to reflect the degree of volatility of the variables of interest. Additionally, forecasters tend to overestimate growth during recessions and repeatedly fail to anticipate them correctly (Lewis and Pain (2014), An, Jalles, and Loungani (2018)). Other studies focus on the forecasting methodology (see Elliott and Timmermann (2008), for example) or evaluate the performance of just one forecasting institute. For instance, Rich and Tracey (2021) or Abel, Rich, Song, and Tracy (2016) look specifically at the

accuracy of the European Central Bank’s Survey of Professional Forecasters. While C. D. Romer and D. H. Romer (2000) or Clements, Joutz, and Stekler (2007) evaluate the forecast performance of the Federal Reserve Greenbook. These analyses have been conducted for a variety of countries. While most of the literature focuses on G7 and advanced countries (see above), there are also a few publications considering forecast performances in developing countries and Asia (for example, Capistrán and López-Moctezuma (2014) for Mexico or Deschamps and Bianchi (2012) for China). Dovern et al. (2015) jointly analyze forecasts from advanced and emerging economies in a large panel dataset.

Regarding the performance of economic forecasts for Germany, Köhler and Döpke (2023) use a variety of different criteria to rank forecasts for the German economy. However, they are unable to identify an institute that consistently outperforms the others. Significant differences can only be found across longer horizon, which the authors judge to arise mostly because of varying forecast horizons across institutes. With respect to the effect of recent crises, Döpke, Fritsche, and Müller (2019) find, based on various standard measures of forecasts accuracy (like RMSE, the mean absolute error or Theil’s Inequality Coefficient), that the accuracy of German forecasts remains mostly unchanged after the Financial Crisis of 2008/09, but there are signs for changes in forecasting behavior. In general, our analysis is similar to the approach by Döpke and Fritsche (2006) who consider time series for growth and inflation forecasts for the German economy from 1970 to 2004 and conduct a variety of statistical test regarding bias and efficiency of these forecasts. They find them to be unbiased but only weakly efficient. Pooling forecasts generates similar findings. Our approach differs from those above in that instead of ranking or evaluating the performance of individual professional forecasts, we put the focus on the question whether professional forecasters can be beaten by standard time-series models. With respect to business cycle fluctuations our methodology is closely related to Loungani et al. (2013) and Dovern and Jannsen (2017). The former use a larger data of real GDP growth forecasts for 46 countries and analyze differences in information rigidities across countries and business cycle states, while the latter look at differences in growth forecast errors over the business cycle for a sample of advanced economies. New about our approach is the use of a long time series. Thereby we are able to consider changes over time and identify trends with respect to business cycle fluctuations.

The structure of this paper is as follow. Section 2 presents general information on the data used. The main analysis is presented in Section 3. We assess the accuracy of real GDP growth and CPI inflation forecasts over time and consider effects by recessions and recoveries. We also evaluate the performance in comparison to simple time-series

models. Additionally, we test for informational rigidities, also with respect to business cycle states. Section 4 concludes.

2 Data Description

This section presents data sources and general characteristics of the forecast data. Additionally, information on the business cycle dating scheme we use is given.

2.1 Data sources

We use professional forecasts of real GDP growth and CPI inflation for Germany from 1969 to 2022, published by German economic research institutes, German government agencies as well as international organizations. The data are obtained through the Halle Institute for Economic Research (IWH), which collects forecasts from various sources and makes them available on the dashboard ForDas.¹ We use forecasts from 15 different forecasters (10 economic research institutes or groups, 2 national government agencies, and 3 international organizations). Table A.1 in the Appendix provides an overview of the institutes used in detail.

Most professional forecasters in Germany follow a semiannual or quarterly release calendar.² Each release typically contains forecasts for the ongoing calendar year as well as the next year. Therefore the forecast horizon h varies between 4 half-years (2 years) and one ahead ($h = 4, \dots, 1$). Let x_t denote the actual value of the variable x in year t and let $\hat{x}_{t,h,i}$ denote the h -quarter ahead forecast released by forecaster i for target year t . With a semi-annual frequency we therefore have 4 forecast for each target year, i.e. the data consists of fixed-event forecasts. Fixed event forecasts exhibit a special error structure (for more information see Davies and Lahiri (1995) and Doern and Weisser (2011)). We account for this when evaluating the forecasts. Forecast errors are constructed by subtracting forecasts from actual values. The forecast error at time t of horizon h by forecaster i is defined as $e_{t,h,i} = x_t - \hat{x}_{t,h,i}$. Note that a positive error means the actual value exceeds the forecast. We have an unbalanced panel as institutes do not follow identical release schedules and data availability improves over the sample period.³ We

¹The dashboard can be accessed here: <https://halle-institute-for-economic-research.shinyapps.io/economic-forecast/>. A detailed description of the dashboard can also be found in Heinisch, Behrens, Döpke, Foltas, Fritsche, Köhler, Müller, Puckelwald, and Reichmayr (2023).

²This is especially true for later years in our sample. For earlier years there are several institutes that have published only one forecast per year.

³Note that the number of observations varies significantly across individual institutes. For instance, for GDP growth the number of observations ranges for real GDP forecasts from 84 forecasts at horizon

are not interested in evaluating forecast on an individual forecaster basis. Instead we use aggregated forecasts throughout this paper.

In general, there are two types of actual values to calculate forecast errors, namely last-vintage values or real-time values. Last-vintage values refer to the latest release of official statistics while real-time values refer to the first release (typically published at the end of January of the year following the target year). Both types of data are provided by the IWH Dashboard.⁴ The main difference between real-time and last-vintage values is the information content they entail. While real-time values presumably reflect the information set available to the forecasters at the time of forecast better, last-vintage values reflect the target variable’s final and actual value, thus the forecasting target. We report errors based on real-time values, assuming it better captures the information set available to the forecasters.⁵ In general, the focus of our analysis is on identifying general characteristics of our dataset. Consequently we decide to rely on real-time values.

In the following, we evaluate the forecasting performance across institutes. We start with some descriptive statistics before using root mean squared errors (RMSEs) to evaluate the forecast accuracy.⁶ Furthermore, we evaluate the role the available information set plays for forecast performance. Therefore we analyze existing differences in forecast release quarters and across decades.

2.2 Business Cycle Dating

To analyze the role of business cycle fluctuations we start by identifying periods of recessions, recoveries and business cycle turning points within our sample period. We broadly follow the business cycle dating scheme used by Doornik and Jannsen (2017) and use both quarterly and yearly dating variables. This choice is motivated by the fact that

1 and 2 (by IFO and BMWK, respectively) to 7 forecasts (by SVR) at horizon 4. For CPI inflation forecasts most forecasts are available at horizon 1 (80 forecasts by HWWI) while the lowest number of forecasts is published at horizon 4 (by SVR). Table A.2 in the Appendix specifically lists the first target years for which data is available by institute.

⁴As stated in the help section for the IWH Dashboard, last-vintage values and first release data are taken from publications by the German Statistical Office (Destatis).

⁵We also conducted our analysis with errors based on last-vintage values. In general, the differences between both specifications are small, but forecasting precision appears to be slightly better for real-time errors, which indicates that the information and first releases available to the forecasters play an important role for the forecasting exercise. The results for our last-vintage value specification are available upon request.

⁶There is a large literature on accuracy measures and statistical tests to evaluate forecast performance, often with respect to forecasts derived from time-series models. For an overview of these measures see Gooijer and Hyndman (2006), for instance. Important contributions regarding suitable statistical test were made by Armstrong and Collopy (1992) or Diebold and Mariano (1995), among others. We focus on RMSE measures for now to facilitate comparison.

while the forecasts are released at a *quarterly* frequency (on an aggregated level) the forecasts themselves are made for variables at an *annual* level. Thus, we identify years of recession and recovery to match the annual forecasts and determine whether the forecast refers to a year that later turned out be a recession or recovery year. The definition here is simple. In line with the description by Dovern and Jannsen (2017) we define three dummy variables. First, a dummy variable indicating recession years which is 1 when the annual growth rate in real GDP is negative.⁷ Second, we identify a year as recovery year when the annual growth rate of real GDP is positive while the previous year has a negative growth rate. Third, to account for longer recession periods there is a dummy variable indicating the first year of a recession series.

Additionally, we define dummy variables for recession and recovery periods at a quarterly frequency to account for the *quarterly* release schedule by the professional forecasters. Thereby we capture effects relating to the forecasters' environment at the time of forecast release. The quarterly dating scheme similarly defines recession and recovery periods but with respect to the time of forecast release. Thus, the quarterly scheme can account for differences with respect to the forecasting environment, i.e. whether the forecasters are publishing their forecasts in a contractionary or expansionary phase of the business cycle. We again broadly follow Dovern and Jannsen (2017). We identify peaks and troughs of the business cycle at a quarterly frequency based on information on business cycle dates for the German economy, specifically peak and trough quarters, published by the German Council of Economic Experts (SVR) (published here: <https://www.sachverstaendigenrat-wirtschaft.de/themen/konjunktur-und-wachstum/konjunkturzyklus-datierung.html>) and then define a dummy variable for recessions and recoveries. With follow Dovern and Jannsen (2017): a recession starts in the quarter following a peak quarter while a recovery begins in the quarter following a trough quarter. Additionally, once quarter has been identified as a recession or recovery quarter, the dummy variable takes the value 1 until the end of the respective target year.

2.3 Descriptive Statistics

Tables 1 and 2 display basic summary statistics for real GDP growth and CPI inflation forecast errors summarized across all institutes by forecast horizon. As mentioned earlier, most institutes publish several forecasts per year. In general, the number of GDP growth

⁷The relevant data on the real GDP growth rate is obtained from the German Council of Economic Affairs and refers to the last release (as opposed to real-time data).

forecasts is larger than the number of inflation forecasts.⁸

We find evidence for a negative bias in the GDP growth forecasts as well as a for a positive bias in CPI inflation forecasts, with an exception of the forecasts released at horizon = 1. Across all institutes and horizons, the mean forecast error is approximately -0.35 percentage points for real GDP growth and 0.13 percentage points for CPI inflation; see Tables 1 and 2. This means that forecasters tend to overestimate growth and underestimate CPI inflation growth on average. This tendency is remarkably uniform across institutes with all 15 institutes showing negative mean growth forecast errors and positive mean CPI inflation forecast errors.

Next we find that the distribution of forecast errors is skewed. For GDP growth, forecasts are mostly negatively skewed, indicating that the tail of the distribution is longer on the left than on the right side. Again, this does not hold for the forecast horizon closest to the end of the target year. Overall, forecasters seem to have been too optimistic about GDP growth more often than they were too pessimistic. This effect could be due to recessions and we will elaborate on this point further below. Generally, 1991 proved to be a challenging year for forecasters since it was the first year they had to produce forecasts for the newly reunited Germany, including the former area of the German Democratic Republic.⁹ For inflation, the distribution of errors are positively skewed. The recent underestimation of inflation growth seems to be a driving force here.

In comparison to a normal distribution our GDP growth and CPI inflation nowcast and forecast errors have excess kurtosis, ranging from 31.96 in the case of GDP growth forecasts at horizon 1 to 4.2 for horizon 4. The source of excess kurtosis differs between target variables and forecast horizons. These findings are additionally illustrated by figure 1 displaying histograms of the distribution of forecast errors. For simplicity we show horizons 1 and 2 together as well as 3 and 4. While the former (blue bars) are centered around zero and have relatively thin tails, the distribution of the latter is more spread out and slightly shifted to the left.

Note that, the forecast errors are highly correlated across institutes.¹⁰ We see high and mostly positive correlations for both GDP growth and CPI inflation. This finding is to some extent reassuring as it would be surprising if institutes who in principle share the

⁸We also calculated summary statistics for each individual institute, but found statistics across institutes to be similar in size and direction to the overall summary statistics. The individual results are available upon request.

⁹Some uncertainty remains regarding the quality of our forecast errors for 1991. We assume that most forecasts refer to the reunited Germany. However, we cannot rule out that some of our forecasts still only consider West-Germany. The average errors for 1991 are larger than those for 1990 or 1992. We therefore decide to omit the errors for this year.

¹⁰Tables on pairwise correlations of errors across institutes are available upon request.

same (or very similar) information set produced divergent predictions for the same target variables. Interestingly, the cross-institute correlation seems to be higher for forecast errors at longer horizons than shorter ones.

Next, we will turn to evaluating the performance of forecasts.

Table 1: Descriptive statistics of real GDP growth forecast errors at different horizons

horizons h	Real GDP growth				
	1	2	3	4	Overall
Mean	0.02	-0.09*	-0.59***	-0.94***	-0.35***
Median	0.00	-0.09	-0.32	-0.50	-0.09
Standard Deviation	0.51	1.26	1.83	2.27	1.57
Minimum	-4.91	-6.11	-6.41	-7.01	-7.01
Maximum	5.30	5.15	4.60	4.75	5.30
Skewness	0.39	-0.82	-1.15	-0.92	-1.61
Kurtosis	31.96	8.66	5.10	4.20	8.35
Observations	775	654	728	473	2630
*** p<0.01, ** p<0.05, * p<0.1					

Notes: Descriptive statistics forecast errors of real GDP growth by forecast horizon h . Forecasts are measured relative to real-time data, in annual percentage rates of change (in percentage points).

Table 2: Descriptive statistics of CPI inflation forecast errors at different horizons

horizons h	CPI inflation				
	1	2	3	4	Overall
Mean	-0.01	0.17***	0.13**	0.32***	0.13***
Median	0.00	0.00	-0.10	0.00	0.00
Standard Deviation	0.51	0.87	1.37	1.72	1.16
Minimum	-2.00	-2.60	-3.30	-3.60	-3.60
Maximum	2.00	5.00	6.40	6.80	6.80
Skewness	0.24	1.48	2.11	2.20	2.64
Kurtosis	5.22	9.10	9.88	8.51	14.74
Observation	619	529	575	391	2114
*** p<0.01, ** p<0.05, * p<0.1					

Notes: Descriptive statistics forecast errors of CPI inflation by forecast horizon h . Forecasts are measured relative to real-time data, in annual percentage rates of change (in percentage points).

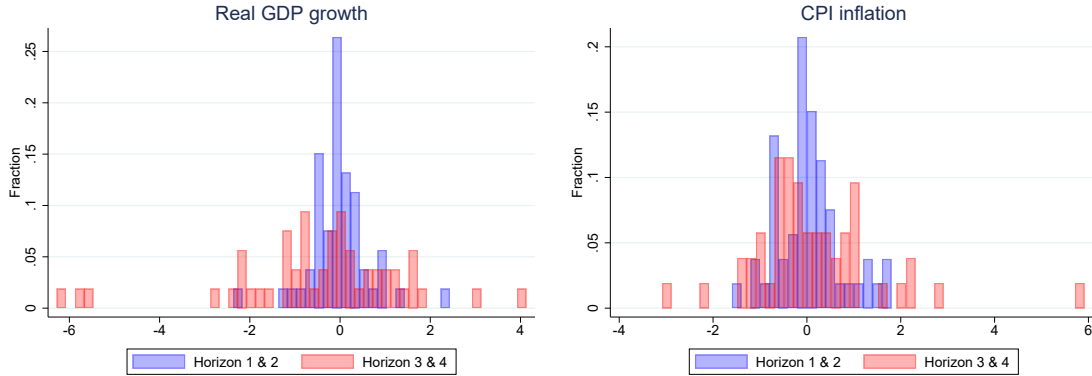


Figure 1: Histograms of nowcast and forecast errors for GDP growth (left panel) and CPI inflation (right panel) in percentage points.

3 Evaluation of Forecasts

We evaluate professional forecasts along three dimensions. First, we show how forecast accuracy developed over time and establish the dependence of forecast accuracy on the state of the business cycle. Second, we compare professional forecasts to simple time series models based on real-time data of GDP growth and inflation. Third, we test for information rigidities and consider differences across business cycle states.

3.1 Forecast performance over time and the role of the business cycle

Have professional forecasts improved over time? There are several reasons to expect such an improvement. First, technological progress in data collection, storage, sharing and processing during the past decades should have helped forecasters gain more accurate knowledge of economic data. Second, one would hope that research in forecasting methods had an impact on the quality of forecasts over and above the effect of increased data availability. Lastly, the reduction in output volatility, called the “Great Moderation” by Stock and Watson (2002) and documented by Buch, Pierdzioch, and Doepke (2004) for Germany, should have made it less difficult to forecast economic growth. Likewise, inflation volatility also decreased since the 1990s.

Evaluating forecast performance over time requires taking into account the forecasters’ information set. To account for the information set available to the forecasters at the time of forecast we evaluate whether the timing of when a forecast was released makes a difference regarding forecast performance. Moreover, we explore potential differences across decades to see whether forecasting might have become easier over time. This could

give an advantage to those forecasters with fewer and more recent forecasts. Additionally, we account for differences due to changes in business cycle states.

Figure 2 shows differences in RMSEs for GDP growth and CPI inflation as an average across institutes by horizon. We additionally distinguish between forecasts made before and after 1991, when forecasts started to be made for a reunified Germany. There is a clear trend for improvement in performance when forecasts are released closer to the end of the target year. Reunification did not affect forecast accuracy for either variable.

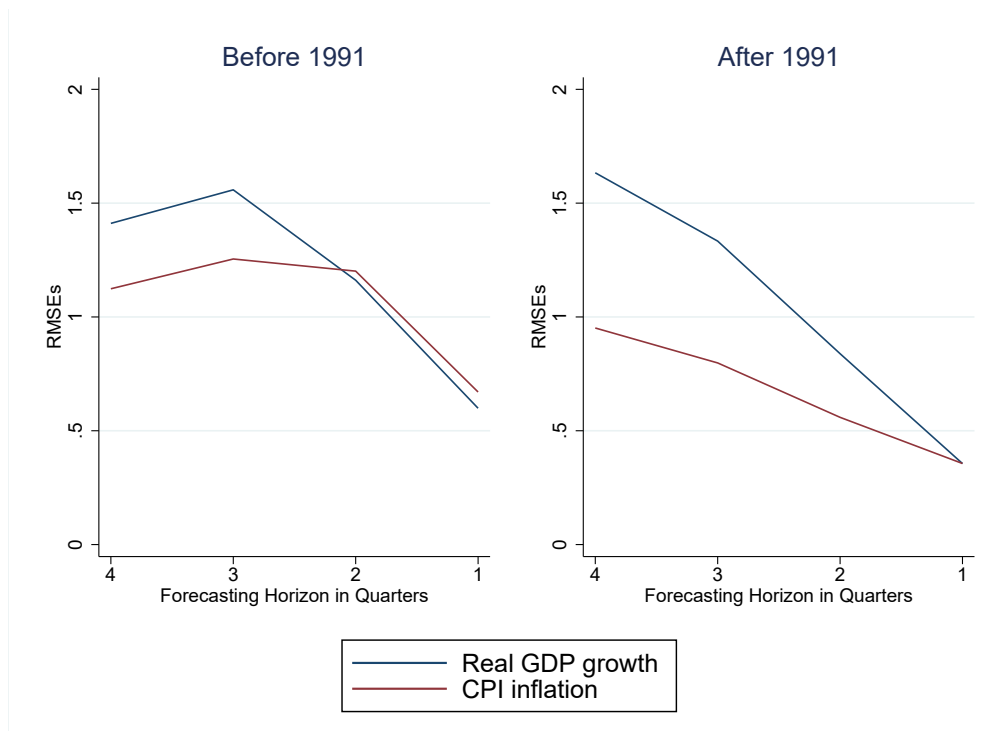


Figure 2: Root mean squared error of GDP growth forecasts by horizon in percentage points and for two subsamples: before and after 1991 (reunification of Germany).

To evaluate forecast performance over time, we also calculate RMSEs for each decade in our sample. Figures 3 and 4 show how the quality of forecasts evolved over time.

There is no clear time trend of improvement visible with respect to GDP growth. RMSEs for GDP growth forecasts are highest for the 2000s, mostly due to the Great Recession of 2008/09, which professional forecasters generally failed to predict. GDP growth forecasts seem to be most accurate in the 1980s, 90s, as well as in the 2010s. However, this can be explained by the fact that growth was less volatile during those decades. This can be seen by the green triangles in figures 3 and 4, which show the

standard deviation of growth rates for each decade (measured in real time).

For CPI inflation forecasts, there is a tendency for RMSEs to improve over time. Nonetheless, this tendency is again largely due to the decreasing volatility of inflation. This suggests that the main reasons why inflation forecasts have become more accurate is that inflation has become easier to predict in recent decades.

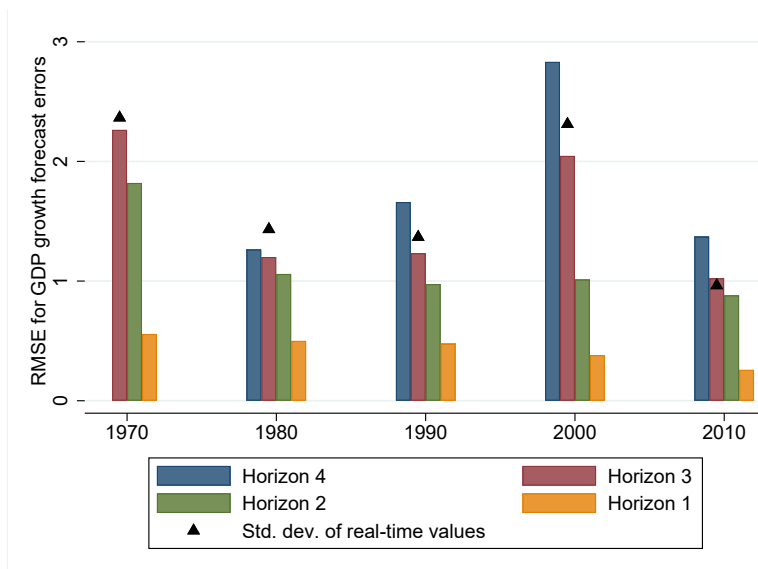


Figure 3: Root mean squared error of GDP growth forecasts and standard deviation of real GDP growth by decade based on real-time values in annual percentage rates of change (in percentage points).

3.2 Comparison to Simple Time Series Models

Professional forecasters have access to a wide range of economic data, both quantitative measurements such as first releases of monthly unemployment rates as well as qualitative information contained in news reports. One may therefore ask whether forecasts that (at least potentially) incorporate all these data significantly outperform simple time series models which rely on less information. To answer this question we assess the forecast errors of our 15 institutes and organizations against univariate autoregressive models as well as bivariate vector-autoregressions based on real-time data of GDP growth of CPI inflation. To facility the comparison in this section we summarize forecasts at horizon 1 and 2 as *nowcasts*, while we keep using the term *forecasts* for horizons 3 and 4.

We rely RMSEs as our basic measure of forecast quality.¹¹ The RMSE of forecaster i

¹¹We also used mean absolute errors (MAEs) to evaluate forecast performance. The conclusions were

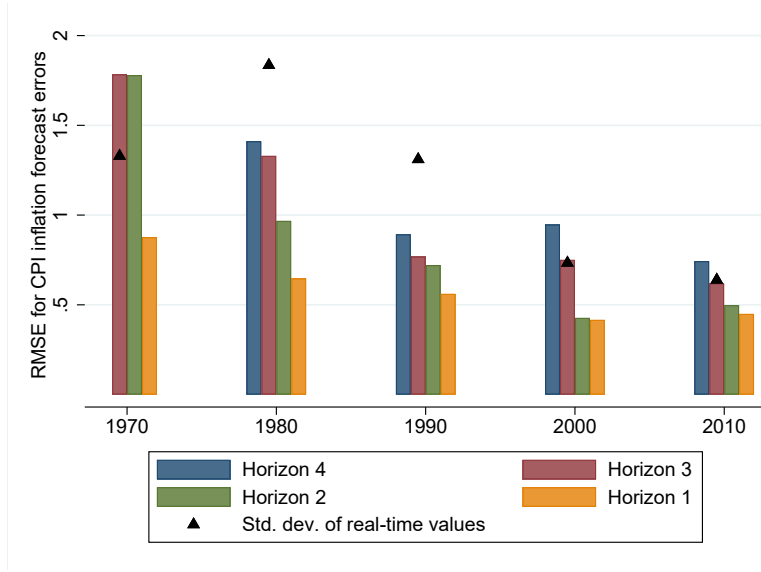


Figure 4: Root mean squared error of CPI inflation forecasts and standard deviation of CPI inflation by decade based on real-time values in annual percentage rates of change (in percentage points).

for horizon h is defined as follows:

$$RMSE_{i,h} = \sqrt{\frac{1}{N} \sum_{t=1}^N (e_{t,h,i})^2}. \quad (1)$$

We compare the RMSEs of professional forecasts to two benchmark models. The first benchmark is a univariate AR(1) model of the form

$$x_t = \rho_0 + \rho_1 x_{t-1} + \nu_t, \quad (2)$$

where ρ_0 and ρ_1 are parameters and ν_t are mean-zero innovations. The resulting forecast mimics an econometrician who makes predictions about x_t based only on past observations x_{t-1}, x_{t-2}, \dots of that same variable (GDP growth and CPI inflation, respectively). Our second benchmark is a two-variable VAR(1) model of the form

$$Y_t = A_0 + A_1 Y_{t-1} + \nu_t, \quad (3)$$

where Y_t collects the two individual variables x_t , namely GDP growth and CPI inflation.

When comparing the institutes' nowcast and forecast errors to these benchmarks, it equal to those from our RMSE analysis which is why they do not report the MAE results here.

is important to account for the information set available to the institutes at the time. The later the release date, the more information becomes available to the forecasters. Therefore, we estimate the benchmark AR and VAR models on a rolling basis using only real-time values available up to time t to generate one-step-ahead forecasts for the period $t + 1$.

Importantly, for this comparison we keep only the last nowcast and forecast of each institute in each year, i.e. $h = 5$ for forecasts and $h = 1$ for nowcasts. Most institutes release at least two forecasts and nowcasts per year, the first one typically in spring, the other one in the fall. We therefore only keep the nowcast and forecast with the largest information set for each institute. By keeping only the last nowcast and forecast for each target year, we try to improve comparability to our time-series models.

Figures 5 and 6 illustrate the RMSE comparison. First, we find that across all variables and institutes RMSEs for nowcasts are smaller than those for forecasts. This is to be expected as there is significantly more information available to the forecasters for nowcasts than for forecasts. This finding is slightly more pronounced for GDP growth than for CPI inflation nowcast and forecast errors.

Second, while there are differences in the forecast quality across institutes, we do not find a clear ranking of professional forecasters. This is consistent with the findings of Köhler and Döpke (2023).

Third, nowcasts perform significantly better than the AR and VAR model nowcasts. However, the same is not the case for the forecasts. The professional forecasters do not generally outperform our simple time series models in terms of RMSE. This result is surprising, since the benchmark AR and VAR models use significantly less data compared to the professional forecasters who presumably use a large variety of leading indicators when forecasting GDP growth and CPI inflation.

3.3 Testing for information rigidities

Next we turn to the question of informational efficiency of professions forecasts, i.e. whether forecasters incorporate available information into their forecast in an efficient way. A long literature starting with Nordhaus (1987, p.667) establishes empirical tests for forecasting efficiency, which requires that a forecast ‘[...] minimizes the loss function subject to available information’. Based on this definition a distinction between strong and weak efficiency can be made. For a forecast to be weakly efficient it is necessary that it includes all information from current and past forecasts.¹² Nordhaus (1987) shows

¹²For strong efficiency the relevant information set restriction consists of all information available at time t (including information about the structure of the economy and private knowledge)

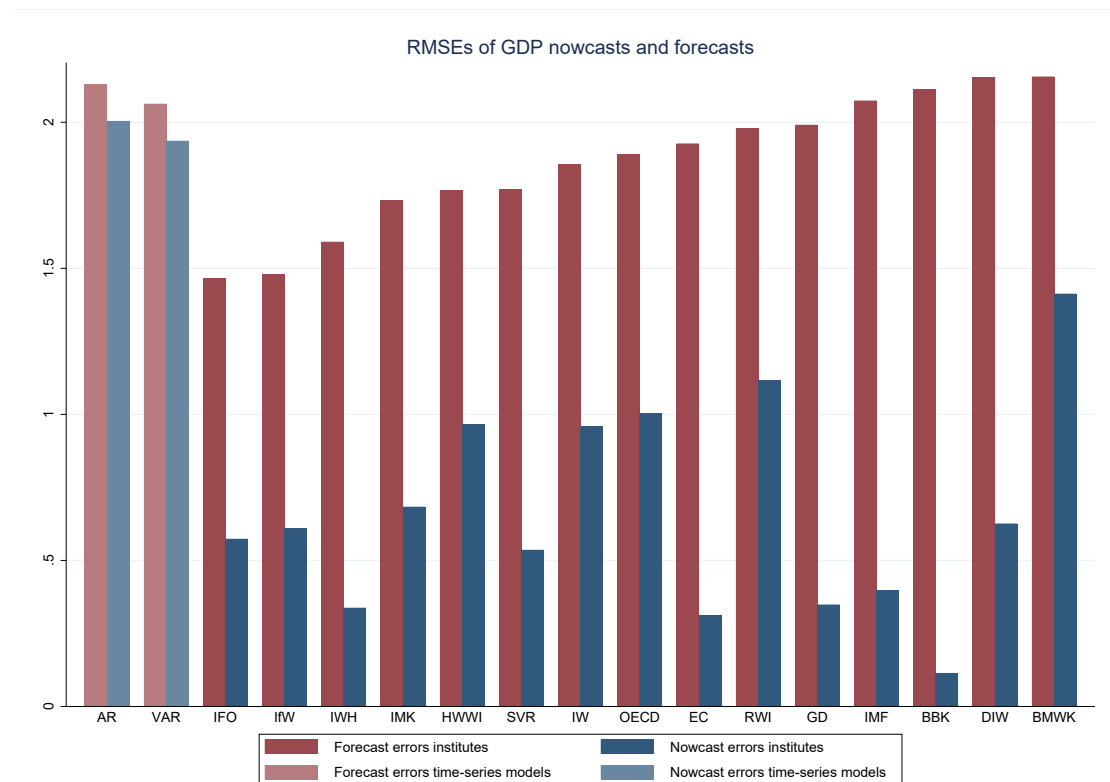


Figure 5: Root mean squared errors of GDP growth nowcasts and forecasts in percentage points.

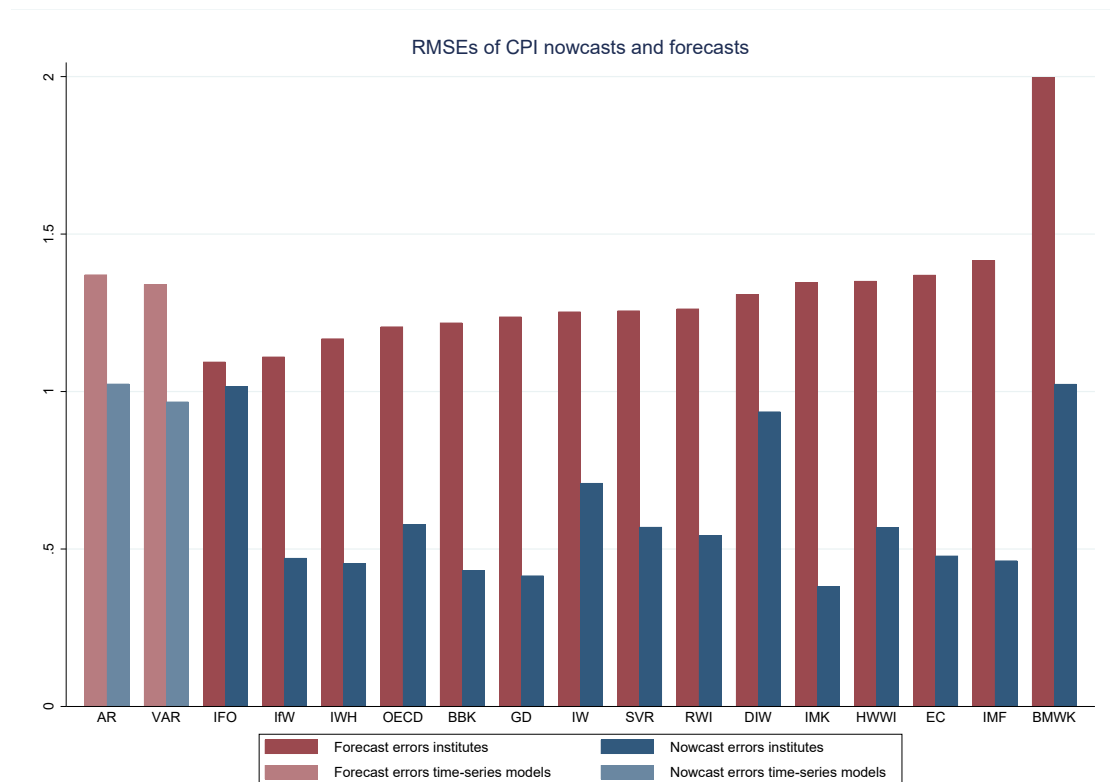


Figure 6: Root mean squared errors of CPI inflation nowcasts and forecasts in percentage points.

that forecast revisions can be used to evaluate forecast efficiency. More specifically, weak efficiency then requires that forecast errors and forecasts revisions at time t are independent of all past revisions until $t - 1$.¹³ Thus, we should not be able to predict future forecast revisions from past revisions, indicating positive or negative correlation. Thereby inefficient forecasts appear smoother than efficient forecasts.

Given the structure of our data we need to make some adjustments before applying the Nordhaus test. We calculate forecast revisions $r_{t,h}$ for horizons $h + 1$ to h , for each target year t : $r_{t,h} = \hat{x}_{t,h+1} - \hat{x}_{t,h}$ with $h = 1, \dots, 3$ and where $h = 1$ indicates the forecast horizon closest to the target year. Due to the semi-annual structure we are able to calculate three forecast revisions for each target year. Third, we then regress the revisions at time $h = 1$ on all lagged revisions:

$$r_{t,h} = \beta_0 + \beta_1 r_{t,h+1} + u_{t,h} \quad \text{for } h = 1, 2 \quad (4)$$

$$r_{t,h} = \beta_1 r_{t,h+1} + \beta_2 \lambda_{t,h} + \beta_3 \lambda_{t,h} r_{t,h+1} + u_{t,h} \quad \text{for } h = 1, 2 \quad (5)$$

where β_0 is a constant term and β_1 captures the effect of the lagged revisions. Coefficients β_3 and β_4 show effects with respect to recessions.¹⁴ Tables 3 and 4 display our results.

The results of this test are illustrated in Tables 3 and 4. We obtain positive and significant coefficients for both GDP growth and CPI inflation lagged forecast revisions and across all horizons. Thus, forecast revisions seem to be positively correlated, implying that information rigidities are present. This is in line with previous findings by Nordhaus (1987) and Dovern et al. (2015). Based on Loungani et al. (2013) we additionally test whether recessions have an effect on the extent on information rigidities (columns 4 and 5). With respect to real GDP growth we obtain negative and significant coefficients for both the recession indicator and the interaction term. This means that recessions imply larger forecast revision and information is incorporated quicker during that time (due to the negative sign on the interaction term). For CPI inflation our results are not significantly different from zero. We therefore do not find forecast revisions to depend on the state of the business cycle with respect to CPI inflation forecasts.

¹³Nordhaus (1987) argues that the series of forecasts for each target year should look like a random walk (more specifically, a martingale process). For forecasts to move up and down at random, it is necessary that past forecast revisions are not informative for the next revision.

¹⁴We focus on forecasts revisions in our analysis. The results for the test based on forecast errors are available on request.

Table 3: Informational rigidities for real GDP growth based on forecast revisions

horizons h	Lagged forecast revisions			Recession effects	
	$h = 2 - 3$	$h = 3 - 4$	$h = 2 - 4$	$h = 2 - 3$	$h = 2 - 3$
Forecast revision (lag 1)	0.406*** (0.116)		0.359*** (0.072)	0.183* (0.101)	0.411*** (0.110)
Forecast revision (lag 2)		0.407*** (0.122)	-0.195 (0.252)		
Recession indicator				-0.987*** (0.340)	-1.111*** (0.365)
Interaction term					-0.344** (0.154)
Constant	-0.106 (0.092)	-0.106 (0.092)	0.033 (0.122)		
Observations	82	81	41	82	82
R-squared	0.151	0.146	0.289	0.351	0.369

Notes: Testing for informational rigidities in real GDP growth forecasts based on forecast revisions. Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Informational rigidities for real CPI inflation based on forecast revisions

horizons h	Lagged forecast revisions			Recession effects	
	$h = 2 - 3$	$h = 3 - 4$	$h = 2 - 4$	$h = 2 - 3$	$h = 2 - 3$
Forecast revision (lag 1)	0.606*** (0.120)		0.455*** (0.120)	0.559*** (0.096)	0.357*** (0.129)
Forecast revision (lag 2)		0.532*** (0.201)	0.208 (0.151)		
Recession indicator				0.267 (0.184)	0.237 (0.184)
Interaction term					0.330* (0.189)
Constant	0.015 (0.058)	0.008 (0.063)	0.028 (0.050)		
Observations	80	79	39	80	80
R-squared	0.302	0.160	0.609	0.342	0.363

Notes: Testing for informational rigidities in CPI inflation forecasts based on forecast revisions. Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4 Conclusion

We use a large panel dataset on GDP growth and CPI inflation forecast errors of 15 professional forecasters for the years 1968 to 2022. Based on summary statistics we find that GDP growth forecast errors are negatively biased, which means that on average forecasters seem to overestimate the true values. We conjecture that this is mostly due to severe recessions that came unexpected for the forecasters. We find CPI errors to be positively biased. This bias seems to be strongly driven by the most recent underestimation of inflation in the aftermath of the Covid-19 pandemic.

We compare GDP growth and CPI inflation forecast errors based on professional forecasts to an AR(1) and a VAR(1)-model. The time-series models are on par with professional forecasters at forecast horizons of 5 quarters or more, but are outperformed by professional forecasters for shorter horizons. This is surprising as the information set of our time series models is purely based on past values while professional forecasters presumably account for more information than that. We additionally find that the forecasting horizon and the amount of information available to the forecasters matter. Judging by RMSE performance, forecasts published in the first quarter of the year, thus further away from the target variable, perform worse than those released in the fourth quarter. Information rigidities seem to be present and for GDP growth to depend on the state of the business cycle.

References

- Abel, Joshua, Robert Rich, Joseph Song, and Joseph Tracy (2016). “The measurement and behavior of uncertainty: evidence from the ECB survey of professional forecasters”. Journal of Applied Econometrics 31 (3), 533–550.
- An, Zidong, João Tovar Jalles, and Prakash Loungani (2018). “How well do economists forecast recessions?” International Finance 21 (2), 100–121.
- Armstrong, J.Scott and Fred Collopy (1992). “Error measures for generalizing about forecasting methods: empirical comparisons”. International Journal of Forecasting 8 (1), 69–80.
- Buch, Claudia M., Christian Pierdzioch, and Joerg Doepke (2004). “Business cycle volatility in germany”. German Economic Review 5 (4), 451–479.
- Capistrán, Carlos and Gabriel López-Moctezuma (2014). “Forecast revisions of mexican inflation and GDP growth”. International Journal of Forecasting 30 (2), 177–191.
- Clements, Michael P., Fred Joutz, and Herman O. Stekler (2007). “An evaluation of the forecasts of the Federal Reserve: a pooled approach”. Journal of Applied Econometrics 22 (1), 121–136.
- Coibion, Olivier and Yuriy Gorodnichenko (2015). “Information rigidity and the expectations formation process: a simple framework and new facts”. American Economic Review 105 (8), 2644–2678.
- Cowles, Alfred (1933). “Can stock market forecasters forecast?” Econometrica 1 (3), 309.
- Davies, Anthony and Kajal Lahiri (1995). “A new framework for analyzing survey forecasts using three-dimensional panel data”. Journal of Econometrics 68 (1), 205–227.
- Deschamps, Bruno and Paolo Bianchi (2012). “An evaluation of chinese macroeconomic forecasts”. Journal of Chinese Economic and Business Studies 10 (3), 229–246.
- Diebold, Francis X. and Roberto S. Mariano (1995). “Comparing predictive accuracy”. Journal of Business & Economic Statistics 13 (3), 253–263.
- Döhrn, Roland and Christoph M. Schmidt (2011). “Information or institution?” Jahrbücher für Nationalökonomie und Statistik 231 (1), 9–27.
- Döpke, Jörg and Ulrich Fritsche (2006). “Growth and inflation forecasts for Germany: a panel-based assessment of accuracy and efficiency”. Empirical Economics 31 (3), 777–798.
- Döpke, Jörg, Ulrich Fritsche, and Karsten Müller (2019). “Has macroeconomic forecasting changed after the Great Recession? Panel-based evidence on forecast

- accuracy and forecaster behavior from Germany”. Journal of Macroeconomics 62, 103–135.
- Dovern, Jonas, Ulrich Fritsche, Prakash Loungani, and Natalia Tamirisa (2015). “Information rigidities: comparing average and individual forecasts for a large international panel”. International Journal of Forecasting 31 (1), 144–154.
- Dovern, Jonas and Nils Jannsen (2017). “Systematic errors in growth expectations over the business cycle”. International Journal of Forecasting 33 (4), 760–769.
- Dovern, Jonas and Johannes Weisser (2011). “Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: an empirical comparison for the g7”. International Journal of Forecasting 27 (2), 452–465.
- Elliott, Graham and Allan Timmermann (2008). “Economic forecasting”. Journal of Economic Literature 46 (1), 3–56.
- Gooijer, Jan G. De and Rob J. Hyndman (2006). “25 years of time series forecasting”. International Journal of Forecasting 22 (3), 443–473.
- Heilemann, Ullrich and Herman O. Stekler (2013). “Has the accuracy of macroeconomic forecasts for Germany improved?” German Economic Review 14 (2), 235–253.
- Heinisch, Katja et al. (2023). “The IWH forecasting dashboard: from forecasts to evaluation and comparison”. Jahrbücher für Nationalökonomie und Statistik.
- Köhler, Tim and Jörg Döpke (2023). “Will the last be the first? ranking German macroeconomic forecasters based on different criteria”. Empirical Economics 64 (2), 797–832.
- Lewis, Christine and Nigel Pain (2014). “Lessons from OECD forecasts during and after the financial crisis”. OECD Journal: Economic Studies 2014 (1), 9–39.
- Loungani, Prakash, Herman Stekler, and Natalia Tamirisa (2013). “Information rigidity in growth forecasts: some cross-country evidence”. International Journal of Forecasting 29 (4), 605–621.
- Nordhaus, William D. (1987). “Forecasting efficiency: concepts and applications”. The Review of Economics and Statistics 69 (4), 667–674.
- Qu, Ritong, Allan Timmermann, and Yinchu Zhu (2019). “Do any economists have superior forecasting skills?” SSRN Electronic Journal.
- Rich, Robert and Joseph Tracey (2021). “All forecasters are not the same: time-varying predictive ability across forecast environments”. Mimeo.
- Romer, Christina D and David H Romer (2000). “Federal reserve information and the behavior of interest rates”. American Economic Review 90 (3), 429–457.
- Stock, James H. and Mark W. Watson (2002). “Has the business cycle changed and why?” NBER Macroeconomics Annual 17, 159–218.

Appendix

Table A.1: Research institutes and international organizations considered for forecast error evaluation.

Institute/ Organization ID	Description
BBK	Deutsche Bundesbank, German Central Bank
BMWK	German Ministry of Economic Affairs
DIW	Institute for Economic Research
GD	“Gemeinschaftsdiagnose”, Joint Economic Forecast
HWWI	Hamburg Institute of International Economics
IFO	ifo Institute Munich
IfW	Kiel Institute for the World Economy
IMK	Macroeconomic Policy Institute
IW	German Economic Institute
IWH	Halle Institute for Economic Research
RWI	Leibniz Institute for Economic Research
SVR	“Sachverständigenrat”, German Council of Economic Experts
EC	European Commission
IMF	International Monetary Fund
OECD	Organization for Economic Co-operation and Development

Table A.2: Data coverage by institute

horizons h	GDP growth		CPI inflation	
	1 & 2	3 & 4	1 & 2	3 & 4
BBK	2007	2008	2007	2008
BMWK	1969	2013	1969	2013
DIW	1970	1971	1970	1973
GD	1969	1969	1990	1993
HWWI	1969	1970	1969	1970
IFO	1969	1978	1971	1979
IfW	1970	1971	1992	1993
IMK	1970	1971	1997	1998
IW	1972	1985	1972	1985
IWH	1993	1995	1993	1995
RWI	1973	1987	1991	1992
SVR	1969	1970	1971	1973
EC	1998	1999	1999	2000
IMF	1992	1993	1993	1994
OECD	1969	1969	1976	1977

Notes: The table displays the first (target) year for which GDP and CPI forecast data is available. Forecasts for horizons $h = 1$ & 2 and $h = 3$ & 4 are combined. The table does not display the first year for which data at all horizons is available but for at least one of both horizons. Gaps in data coverage are not displayed.