



Predicting the German Economy: Headline Survey Indices Under Test

Robert Lehmann¹ · Magnus Reif¹

Received: 3 December 2020 / Accepted: 25 March 2021 / Published online: 15 April 2021
© The Author(s) 2021

Abstract

This analysis investigates the predictive power of the headline indices of the four most important German survey providers. We conduct an out-of-sample, real-time forecast experiment for growth of total and private sector gross domestic product and growth of gross value added in both the manufacturing and the service sector. All providers publish valuable leading indicators for both GDP measures, with some advantages for the ifo indicators and the Economic Sentiment Indicator, respectively. For the manufacturing sector, indicators provided by the ifo Institute are clearly superior. For the service sector, all indicators prove to have a similar nowcasting performance, whereas the Economic Sentiment Services of the Centre for European Research is preferable for one quarter-ahead predictions.

Keywords Forecasting · ifo business climate · PMI · ESI · ZEW economic sentiment

JEL Classification E17 · E27 · E37

1 Introduction

Survey indicators are a well-established source to derive early predictions on current and future development of macroeconomic variables such as gross domestic product (GDP) (see, for example, Angelini et al. 2011). For Germany, two survey providers—namely the ifo Institute (ifo) and IHS Markit (IHS)—and their corresponding headline indices (the ifo Business Situation, the ifo Business Expectations,

We are grateful to two anonymous referees for very helpful comments and suggestions. Special thanks goes to IHS Markit for providing us the PMI data for Germany.

✉ Robert Lehmann
lehmann@ifo.de

Magnus Reif
reif@ifo.de

¹ ifo Institute Munich and CESifo, Munich, Germany

the Ifo Business Climate, and the PMI Composite Output Index) receive considerable media attention each month and are found to be important for tracking economic activity in both the Euro Area and Germany (see, for example, Basselier et al. 2018; de Bondt 2019; Fritsche and Stephan 2002; Lehmann 2020). Recently, the indicators of both survey providers are listed on Bloomberg's "12 Global Economic Indicators to Watch".¹ However, two further and very important survey providers that publish monthly headline indices are mostly neglected in the public debate: the Directorate-General for Economic and Financial Affairs of the European Commission (DG ECFIN) and the Centre for European Economic Research (ZEW). Whereas the indicators of the former provider are also based on either business or consumer surveys, the latter one uses a different source of information: the assessment of financial experts.

In the case of the Ifo Institute and IHS Markit, an ongoing debate across analysts takes place on which indicator is better suited to track the aggregate Germany economy (see, for example, the Wall Street Journal or tradingfloor.com)² or certain branches of the economy. An analysis by J.P. Morgan concludes that IHS Markit's service sector indicator is better in explaining movements in sectoral gross value added than the Ifo Business Expectations.³ This exemplary analysis, however, only investigates the in-sample fit of both indicators, that is, how much they can explain a variable's past fluctuations. From a forecaster's perspective such an analysis is of minor help as it does not take a stand on the indicators' forecasting properties, that is, on the reliability of the indicators' signals for a variable's current and future development.

We conduct an out-of-sample, real-time forecast analysis which compares the forecasting properties of the headline indices of the four very important survey providers in Germany for the current quarter and one quarter-ahead predictions. Instead of solely analyzing the properties for total GDP growth, we further study the forecasting performance of the providers' headline indices for private sector GDP growth (that is market-traded output excluding public activities) and gross value added (GVA) in the most important branches of the Germany economy, that is, manufacturing and services. Furthermore, we discuss the indicators' performance from an applied forecasting stance, investigate the impact of two indicator transformations on the forecast performance, and the accuracy of the real-time forecasts for revised data.

Our results can be summarized as follows. For total real GDP growth all providers publish meaningful and practical relevant leading indicators, with some advantages for the Ifo indicators in case of one quarter-ahead predictions. A similar picture emerges for market-traded GDP, but in this case, the PMI Composite Output Index

¹ Bloomberg, The 12 Global Economic Indicators to Watch, Update: December 1, 2020, <https://www.bloomberg.com/graphics/world-economic-indicators-dashboard>.

² The Wall Street Journal, Should You Believe German PMI or Ifo Data?, February 23, 2012. Trading-Floor, 3 numbers to watch: German & Eurozone PMI, French bond auctions, April 23, 2012.

³ J.P. Morgan, Germany: PMI tracks services activity better than the IFO, August 27, 2019. Unfortunately, there is no permanent link existing to this article.

is advantageous for nowcasts in one of the two models, whereas the Economic Sentiment Indicator of DG ECFIN is preferred for next quarter forecasts. For the manufacturing sector, the ifo indicators—and here especially the ifo Business Situation Manufacturing—are clearly superior with respect to the other providers' headline indices for both the nowcasting and the forecasting setup. For services, all providers publish headline indices with a similar nowcasting performance; the Economic Sentiment Services of ZEW is superior in case of one quarter-ahead predictions.

The paper is organized as follows. In Sect. 2 we introduce the headline indices of the four providers and the forecast experiment. Section 3 presents our baseline results, followed by a discussion in Sect. 4. The last section concludes and gives an outlook on future research activities.

2 Data and Forecast Experiment

2.1 Target Series

The four survey providers, which we evaluate, deliver headline indices for the aggregate German economy as well as for manufacturing and services. In the out-of-sample exercise, we therefore evaluate the forecasting performance for the following target series: GDP, GVA in manufacturing, and GVA in the service sector. Additionally, we consider GDP of the private sector economy, leaving activities of the government such as public administration or education aside. We do so because the providers do not survey the public sector at all which—given the public sector's large share of almost one-fifth in total GVA—might influence the results for total GDP. All four target series are price-, seasonally- and calendar-adjusted, and transformed into quarter-on-quarter growth rates in advance. Whereas total GDP and GVA in manufacturing are officially published by the German Federal Statistical Office, GVA in the aggregate service sector and private sector GDP are not readily available in official German statistics. We calculate both aggregates by applying growth contributions of the single sub-sectors, including all market-traded activities.⁴

To mimic the information set available to a forecaster at each period, we resort to the real-time database of Deutsche Bundesbank. Each data vintage includes quarterly observations for the target series. For each target series, our forecast exercise starts with the data vintage published in February 2012 and thus for the first release of the target series referring to the fourth quarter of 2011. We do so because of data

⁴ Our service sector aggregate comprises the following sub-sectors according to the German Classification of Economic Activities, Edition 2008 (WZ 2008): wholesale and retail trade; repair of motor vehicles and motorcycles (section G), transportation and storage (section H), accommodation and food service activities (section I), information and communication (section J), financial and insurance activities (section K), real estate activities (section L), professional, scientific and technical activities (section M), administrative, and support service activities (section N). GDP of the private sector economy is the aggregation of market-traded service activities plus agriculture, forestry and fishing (section A), mining and quarrying (section B), manufacturing (section C), energy and water supply (sections D+E), and construction (section F).

limitations for the survey indicators and stick to this issue in Sect. 2.3 that describes our forecasting exercise.

2.2 Headline Indices

The provider-specific headline indices for the target series are:⁵

- **ifo:** ifo Business Situation, ifo Business Climate, ifo Business Expectations;
- **IHS:** PMI Composite Output Index, Manufacturing PMI, PMI Services Business Activity Index;
- **DG ECFIN:** Economic Sentiment Indicator, Industrial Confidence Indicator, Services Confidence Indicator;
- **ZEW:** Current Economic Situation, Economic Sentiment Manufacturing, Economic Sentiment Services.

The two ifo headline indices—ifo Business Situation and ifo Business Expectations—are part of the monthly ifo Business Survey and are available for each single sector. Firms can choose from three qualitative answers reflecting a positive, neutral, or negative assessment. Both headline indices are calculated as the net balance of positive and negative answers; the ifo Business Climate is the geometric average of the two headline indices. The indices for the aggregate economy are calculated from the sector-specific results by applying weights based on gross value added from national accounts.

The headline index by IHS Markit for GDP—the PMI Composite Output Index—is based on the information from manufacturing (Output Index) and services (Business Activity Index). For services, construction, and retail, the headline indices of IHS Markit are based on a question aiming at economic activity. For manufacturing, the headline index is a composite indicator calculated as the weighted average of the following five survey questions: new orders, output, employment, suppliers' delivery times, and the stocks of materials purchased. All indicators are based on a formula given the positive answer full weight, the neutral answers half weight, and the negative answers zero weight. Thus, the indicator is centered around 50.

The Economic Sentiment Indicator is DG ECFIN's headline index for GDP and based on a weighted average of the (standardized) industry-specific Confidence Indicators plus Consumer Confidence. The single Confidence Indicators are the result of the Commission's monthly "Joint and Harmonized Business and Consumer

⁵ We use the term headline index also in case of ifo, DG ECFIN as well as the ZEW and therefore analog to IHS, that is, a headline index is the one closely watched by the markets (see IHS Markit 2017). The Online Appendix contains a comprehensive table with key conceptual differences across the survey providers' headline indices. Additional and more detailed information can be found in: Sauer and Wohlrabe (2018) for ifo, IHS Markit (2019) for IHS, European Union (2016) for DG ECFIN, and Köhler and Schmidt (2006) for ZEW. As we are restricted to show any indicator plots, we refer to the official representations of the survey providers as well as to the existing literature (see, for example Broughton and Lobo 2018; Christiansen et al. 2014; Hüfner and Schröder 2002; Lehmann 2020; Schröder and Hüfner 2002).

Survey". For the industrial sector, the Industrial Confidence Indicator is the mean of the questions on production expectations, order books, and stock of finished products. Each question can be answered by three qualitative answers and the results are represented in balances of positive and negative answers. The Services Confidence Indicator is calculated accordingly and based on questions on the business climate as well as past and future demand.

Compared to the previous three providers, the ZEW does not survey firms but financial experts instead. The monthly survey is based on answers from 350 financial experts who are, *inter alia*, asked on the current economic situation and on industry-specific expectations. As for the other surveys, the participants can choose from three qualitative answers and the headline indices are presented in balances. The ZEW Current Economic Situation can be seen as the headline index for GDP. The ZEW Economic Sentiment Services is directly based on an expectations question regarding the total service sector. Such an aggregate is, however, missing for manufacturing. Nevertheless, the ZEW asks its financial experts on the following industries, that account for more than 60% of GVA in manufacturing: automobile, chemicals/pharmaceuticals, electronics, mechanical engineering, and steel production. We apply official GVA weights from national accounts to weight these sub-sectors together and calculate the Economic Sentiment Manufacturing.

2.3 Forecasting Approach

To compare the indicators' predictive power for the current (nowcast, horizon: $h = 0$) and the next quarter (forecast, horizon: $h = 1$), we apply two indicator models, AR-X(p, q), where p and q denote the lag length of the target series and the indicator, respectively:

1. AR-X(0,0) model including a constant, and the contemporaneous value of an indicator,
2. AR-X(0, q) model including a constant and up to a maximum of two lags of the indicators to consider the indicators' dynamics. We select the optimal lag number q using the Akaike Information Criterion (AIC).⁶

Our sample for the forecasting experiment covers the period from the first quarter 2005 (the first period for which ifo service sector data is available) to the third quarter of 2019; the first 28 quarters are used to compute sensible coefficient estimates. Thus, the first now- and forecast is generated for the first quarter of 2012. After calculating the first predictions, the sample is enlarged by one quarter of observations which is equal to a new vintage of data, the models are re-estimated, and new

⁶ We further apply an AR-X(1,0), an AR-X(1, q), and a Random Walk specification. The lag length for GDP is fixed to unity because of the low persistence of German GDP growth (see Carstensen et al. 2020), which is also underpinned by the lag selection criteria. The three additional models delivered higher forecast errors than the two indicator models. We therefore refrain from discussing them in the main text, but put the results in the Online Appendix.

now- and forecasts are generated. This procedure is repeated until the end of our observation period. We assume that the now- and forecasts are generated at the end of each quarter, thus, the seasonally-adjusted monthly survey indicators are transformed to quarterly frequency by calculating quarterly averages.⁷

Moreover, we apply first differences of the indicators mainly because of the following two reasons. First, the literature focusing on Germany has found that difference specifications yield lower forecast errors than their level counterparts (see, for example, Henzel et al. 2015; Kholodilin and Siliverstovs 2006). Second, we follow the discussion in the literature survey by Lehmann (2020). If a business cycle is defined as a cyclical fluctuation around a trend, survey indicators that are defined by a balance statistic measure these fluctuations by construction. In this case, the “neutral” survey category represents the “normal” activity level of the economy, a positive balance can be interpreted as overutilization; the opposite holds for a negative balance statistic. Given this argumentation, the reference series of survey indicator in levels is the cyclical component of any macroeconomic aggregate (for example, GDP). As we, however, forecast quarterly growth rates, we apply an asymmetric filtering to the original target series. This filtering causes a phase shift of the series back in time and thus suppresses the leading properties of the survey indicators for the original series. Ultimately, the literature surveyed recommends to apply the same transformation to the indicators as to the target series. We follow this recommendation and also apply first differences in our baseline case. However, we will discuss the transformation issue again in Sect. 4 by comparing the forecast errors of the indicators in differences with those of its level counterparts.

The forecast comparison is based on root mean squared forecast errors (RMSFE). We evaluate the now- and forecasts with respect to the first release of data for a specific quarter, which usually receives the highest media attention. Nevertheless, we will also discuss the forecasting performance with respect to the final release of data in Sect. 4. To statistically identify the “best” indicators, we apply the model confidence set (MCS) procedure of Hansen et al. (2011). According to the MCS, “best” comprises all models that are statistically superior for a given confidence level, which we set to 90%. Thus, the MCS allows for making statements about statistical significance, which are not possible according to standard pairwise comparisons.

3 Baseline Results

Total gross domestic product. The RMSFEs (in percentage points) for total GDP growth, both forecast horizons, and each model are presented in Table 1. Figures in bold face highlight the indicator with the lowest RMSFE. Across the two models, all indicators deliver forecast errors within a quite similar range in the nowcasting case. For the AR-X(0,0), the span ranges from 0.34 p.p. (ifo Business Expectations)

⁷ We also tested a different timing assumption. Specifically, we calculate the predictions at the end of the second month of the quarter. The results are virtually identical and available upon request from the authors.

Table 1 Indicator performance for total GDP growth, in p.p

Horizon	Indicator	RMSFE	RMSFE
		AR-X(0,0)	AR-X(0,q)
Nowcast	ifo Business Situation Germany	0.38	0.36*
	ifo Business Climate Germany	0.37	0.37*
	ifo Business Expectations Germany	0.34*	0.40*
	PMI Composite Output Index	0.36	0.36*
	DG ECFIN Economic Sentiment Indicator	0.39	0.37*
	ZEW Current Economic Situation	0.44	0.49
Forecast	ifo Business Situation Germany	0.39*	0.39*
	ifo Business Climate Germany	0.45	0.47
	ifo Business Expectations Germany	0.45	0.55
	PMI Composite Output Index	0.44	0.43
	DG ECFIN Economic Sentiment Indicator	0.40*	0.46
	ZEW Current Economic Situation	0.52	0.51

The table shows root mean squared forecast errors (RMSFE) of the two indicator models for both forecast horizons (Nowcast horizon: $h = 0$, Forecast horizon: $h = 1$). Bold typeface indicates the best performing indicator for each model-horizon combination. Asterisks indicate that the indicator is part of the Model Confidence Set

to 0.44 p.p. (ZEW Current Economic Situation); the span for the AR-X(0,q) ranges from 0.36 p.p. (PMI Composite Output Index) to 0.49 p.p. (ZEW Current Economic Situation). However, the relative improvement of the best indicator over the worst is 23% (model 1) and 27% (model 2) or 0.09 p.p. and 0.13 p.p. in absolute terms. Setting these differences in relation to the standard deviation in quarterly growth rates of total GDP—0.91 p.p. in our sample—means that the worst indicator produces a higher RMSFE that corresponds to 15% of the series' volatility. By taking a closer look on the performance between IHS and DG ECFIN, our results for Germany are in line with those for the euro area: the PMI Composite Output Index performs slightly better than the DG ECFIN Economic Sentiment Indicator to nowcasting GDP growth (see European Union 2017).

The differences across the indicators are a bit larger in the forecasting case. For both models, the ifo Business Situation produces the lowest RMSFE (0.39 p.p.). The largest average forecast errors is provided by the ZEW Current Economic Situation with 0.52 p.p. (model 1) and 0.51 p.p. (model 2). This amounts to a relative improvement of 25% and 24% or to 0.14 p.p. and 0.12 p.p. in absolute terms. In total, we conclude that all providers produce leading indicators for total German GDP growth, with some advantages for the ifo Business Situation in the case of the forecast.

Especially the finding that the ifo Business Situation works best in the forecasting setup whereas the ifo Business Expectations work relatively better in the nowcasting case is very interesting and worthwhile to mention. We can offer no final conclusion to this result. However, it might be related to three issues. First, the wording of the question might play a role. In fact, it is not quite clear which horizon the individual

Table 2 Indicator performance for private sector GDP growth, in p.p

Horizon	Indicator	RMSFE	RMSFE
		AR-X(0,0)	AR-X(0,q)
Nowcast	ifo Business Situation Germany	0.47*	0.50
	ifo Business Climate Germany	0.48*	0.48*
	ifo Business Expectations Germany	0.47*	0.50
	PMI Composite Output Index	0.47*	0.44*
	DG ECFIN Economic Sentiment Indicator	0.50	0.49
	ZEW Current Economic Situation	0.57	0.59
Forecast	ifo Business Situation Germany	0.56	0.56
	ifo Business Climate Germany	0.60	0.60
	ifo Business Expectations Germany	0.57	0.65
	PMI Composite Output Index	0.58	0.55
	DG ECFIN Economic Sentiment Indicator	0.54*	0.52*
	ZEW Current Economic Situation	0.69	0.65

The table shows root mean squared forecast errors (RMSFE) of the two indicator models for both forecast horizons (Nowcast horizon: $h = 0$, Forecast horizon: $h = 1$). Bold typeface indicates the best performing indicator for each model-horizon combination. Asterisks indicate that the indicator is part of the Model Confidence Set

firm considers when forming its expectations as the question wording—“for the next six months”—leaves room for interpretation. If, on the one hand, the firm has a very short horizon in mind, the differences between both indicators might shrink. On the other hand, if the firm has a rather long horizon in mind, the expectations might be too blurry because the firm cannot anticipate all the relevant shocks. In this case, the expectations might include too much noise. Second, the aggregation of the monthly expectations series to quarterly values in order to match the target series’ frequency might impact the results. Obviously, reducing the indicators frequency absorbs both noise and signal from the indicators. Our results suggest that a significant fraction of the signal is lost by aggregation to the lower frequency. Third, the evaluation period might influence the average performance of the indicators. If our sample includes a period which is, for instance, driven by high uncertainty, the firms have difficulties in anticipating future shocks, leading to an inferior predictive power of the expectations indicators. This suggestion is underpinned by taking a closer look on the average forecasting performance in the period from 2014 to 2019. During this period, the expectations indicator clearly outperforms the business situation for one quarter-ahead forecasts.

Gross Domestic Product of the Private Sector Economy. Turning to total production of the private sector (see Table 2), we observe an interesting phenomenon: the RMSFEs for each headline index and forecast horizon are higher compared to those for total GDP growth. This phenomenon might be related to the smoothing properties of the public sector for GDP growth. Whereas the standard deviation in quarterly growth rates for total GDP in our observation period is 0.91 p.p., it increases to 1.15 p.p. for GDP of the private sector economy. As quarterly averaged

Table 3 Indicator performance for GVA growth in manufacturing, in p.p

Horizon	Indicator	RMSFE	RMSFE
		AR-X(0,0)	AR-X(0,q)
Nowcast	ifo Business Situation Manufacturing	1.19	1.18*
	ifo Business Climate Manufacturing	1.22	1.27
	ifo Business Expectations Manufacturing	1.13*	1.42
	Manufacturing PMI	1.21	1.27
	DG ECFIN Industrial Confidence Indicator	1.28	1.35
	ZEW Economic Sentiment Manufacturing	1.16	1.81
Forecast	ifo Business Situation Manufacturing	1.25*	1.38
	ifo Business Climate Manufacturing	1.33	1.32*
	ifo Business Expectations Manufacturing	1.38	1.56
	Manufacturing PMI	1.50	1.57
	DG ECFIN Industrial Confidence Indicator	1.39	1.41
	ZEW Economic Sentiment Manufacturing	1.80	2.01

The table shows root mean squared forecast errors (RMSFE) of the two indicator models for both forecast horizons (Nowcast horizon: $h = 0$, Forecast horizon: $h = 1$). Bold typeface indicates the best performing indicator for each model-horizon combination. Asterisks indicate that the indicator is part of the Model Confidence Set

survey indicators are quite smooth variables, this might explain why they exhibit a better average forecasting performance than for private sector GDP. We stick to this issue again when we discuss the practical relevance of our results.

Compared to total GDP growth, the span between the best and the worst indicator for private sector GDP is larger for both the two forecasting horizons and the two models. In the case of the nowcasts, the span for the AR-X(0,0) lies between 0.47 p.p. (ifo Business Expectations) and 0.57 p.p. (ZEW Current Economic Situation); for the AR-X(0,q), the span lies between 0.44 p.p. (PMI Composite Output Index) and 0.59 p.p. (ZEW Current Economic Situation). These spans correspond to the relative improvements of the best compared to the worst indicator of 19% (model 1) and 25% (model 2). Expressed in absolute terms the differences are 0.11 p.p. and 0.15 p.p., respectively.

The spans again become a bit larger in the case of the forecasting setup. For the AR-X(0,0), it ranges from 0.54 p.p. (DG ECFIN ESI) to 0.69 p.p. (ZEW Current Economic Situation). The span for the AR-X(0,q) instead lies between 0.52 p.p. and 0.65 p.p. (ZEW Current Economic Situation). The relative improvements amount to 22% (model 1) and 20% (model 2); these relative improvements correspond with 0.15 p.p. and 0.13 p.p. in absolute terms. Overall, all providers also publish leading indicators for private sector GDP growth, with some advantages for the PMI Composite Output Index in one out of the two nowcasting models and for the Economic Sentiment Indicator of DG ECFIN in case of the forecast.

Gross Value Added Manufacturing. Table 3 presents the RMSFEs for manufacturing. Overall, the average forecast errors are three times higher than those for total GDP, which is not surprising given the higher volatility of the manufacturing

series. In the nowcasting setup, the best performing indicators (model 1: ifo Business Expectations Manufacturing; model 2: ifo Business Situation Manufacturing) produce RMSFEs of 1.13 p.p. and 1.18 p.p., respectively. Compared to the worst performing indicators (model 1: DG ECFIN Industrial Confidence Indicator; model 2: ZEW Economic Sentiment Manufacturing), the relative improvements amount to 12% and 35%; these figures correspond to improvements of 0.15 p.p. and 0.63 p.p. in absolute terms. Given the volatility of 3.03 p.p. in quarterly GVA manufacturing growth, these differences can be expressed in RMSFEs that correspond to 5% and 21% of the series' volatility. Sticking again to the discussion of the performance between IHS and DG ECFIN for the euro area, our results for Germany are only partially comparable. The European Union (2017) finds a better nowcasting performance for the Industrial Confidence Indicator compared to the IHS indicator, whereas we find the opposite result. However, the evaluation for the euro area is based on monthly industrial production, but we apply quarterly GVA in manufacturing.

Turning to one quarter-ahead forecasts, the spans across indicators become even larger. For the AR-X(0,0), the span lies between 1.25 p.p. (ifo Business Situation Manufacturing) and 1.80 p.p. (ZEW Economic Sentiment Manufacturing). The span for model 2 ranges from 1.32 p.p. (ifo Business Climate Manufacturing) to 2.01 p.p. (ZEW Economic Sentiment Manufacturing). The relative improvement is 30% (model 1) and 34% (model 2) or 0.55 p.p. and 0.69 p.p. in absolute terms. Measured in terms of the standard deviation of the target series, these absolute improvements correspond to 18% and 23%, respectively. Taking both forecast horizons together, the ifo indicators—especially the ifo Business Situation Manufacturing—seem to be superior to the headline indices of the other three providers.⁸ However, we have to state that IHS Markit also publishes a PMI Manufacturing Output Index which is equivalent to their index for the service sector. Our application of their headline index, the Manufacturing PMI, might be a caveat of our analysis and can be investigated in follow-up studies. The same might hold true for other industrial indicators published by DG ECFIN.

Gross Value Added Services. The results for market-traded services are summarized in Table 4. In the case of the nowcasting setup and the AR-X(0,0), the RMSFEs are virtually identical and range between 0.32 p.p. (PMI Services Business Activity Index) and 0.35 p.p. (ifo Business Situation Services). Interestingly, all indicator RMSFEs worsen by applying the AR-X(0,q), with the exception of the ZEW Economic Sentiment Services. In the end, the in-sample AIC suggests for all indicators a higher lag order that comes with the price of a lower forecast accuracy. Again, the ZEW indicator is the exception as in the vast majority of cases the AIC recommends a model that only includes the contemporaneous indicator value. The application of the AR-X(0,q) leads to an increase of the span between the best and worst performing indicator, which runs from 0.33 p.p. (ZEW Economic Sentiment Services) to 0.43 p.p. (ifo Business Situation Services). The relative improvement amounts to 23% or 0.10 p.p. in

⁸ Similar results for the comparison between ifo and IHS are found by Pinkwart (2018) who documents a higher forecast accuracy of ifo manufacturing indicators.

Table 4 Indicator performance for GVA growth in services, in p.p

Horizon	Indicator	RMSFE	RMSFE
		AR-X(0,0)	AR-X(0,q)
Nowcast	ifo Business Situation Services	0.35	0.43
	ifo Business Climate Services	0.33	0.41
	ifo Business Expectations Services	0.32*	0.40
	PMI Services Business Activity Index	0.32*	0.42
	DG ECFIN Services Confidence Indicator	0.34	0.37
	ZEW Economic Sentiment Services	0.33	0.33*
Forecast	ifo Business Situation Services	0.51	0.47
	ifo Business Climate Services	0.46	0.43
	ifo Business Expectations Services	0.41	0.37
	PMI Services Business Activity Index	0.44	0.44
	DG ECFIN Services Confidence Indicator	0.38	0.34*
	ZEW Economic Sentiment Services	0.32*	0.34*

The table shows root mean squared forecast errors (RMSFE) of the two indicator models for both forecast horizons (Nowcast horizon: $h = 0$, Forecast horizon: $h = 1$). Bold typeface indicates the best performing indicator for each model-horizon combination. Asterisks indicate that the indicator is part of the Model Confidence Set

absolute values. In terms of the series' volatility (0.89 p.p.), the absolute improvement corresponds to an improvement of 11%. Our nowcasting results for Germany partially underpin the results for the euro area. Whereas the European Union (2017) documents a lower RSMFE of the IHS indicator compared to the one by DG ECFIN, this only holds for Germany in the case of the AR-X(0,0) model and reverses by investigating our second model. Thus, the performance seems to be a matter of modeling.

The forecasting setup reveals that the ZEW Economic Sentiment Services is the best performing indicator in both model specifications (RMSFE: 0.32 p.p. and 0.34 p.p.). For the two models the upper limits of the ranges are 0.51 p.p. and 0.47 p.p. (both ifo Business Situation Services), respectively. Nevertheless, the performance of all indicators—again with the exception of the ZEW indicator—improve by allowing for more dynamics in terms of the AR-X(0,q) model. However, the relative improvements of the ZEW indicator compared to the worst candidate are 21% and 16%, respectively. Overall, the providers publish quite similar indicators in terms of their nowcasting performance, but the ZEW Economic Sentiment Services seems to be superior in case of one quarter-ahead predictions. As for manufacturing, a possible caveat of our analysis might be the solely focus on the headline indices by DG ECFIN as it publishes other service indicators in addition to their Confidence Indicator.

4 Discussion

In the following, we discuss our baseline results in the light of the following three aspects. First, we assess their practical relevance for applied forecasting work. Second, we take up the discussion on the indicator transformation by comparing the

performance across first differences and levels. Third, we evaluate how well the indicators perform by comparing our real-time forecasting results with those based on revised data.

4.1 Relevance for Applied Forecasting Work

We base our discussion on the practical relevance on the Noise-to-Signal Ratio (NTS). The NTS compares the RMSFE of an indicator (enumerator) with the standard deviation of the target series (denominator). Based on this ratio, an indicator is practically relevant if its NTS is below unity, that is, the indicator produces forecast errors that are smaller compared to the volatility of a series. For our period under investigation (2005-Q1 to 2019-Q3), the standard deviations in quarterly growth rates of our target series are the following: 0.91 p.p. for total GDP, 1.15 p.p. for private sector economy GDP, 3.03 p.p. for GVA manufacturing, and 0.89 p.p. for GVA services.

The NTS for GDP growth nowcasts across both models range from 0.38 to 0.54. For one quarter-ahead predictions, the span of NTS lies between 0.43 and 0.60. Overall, each indicator is of practical relevance as all produce average forecast errors that are smaller than the underlying volatility of GDP growth. However, a practical gain exists by comparing the differences between the best and the worst performing indicators with the standard deviation of the target series. For both forecast horizons together, the improvement of the best indicator compared to the worst one is approximately 13% in terms of the series' volatility.

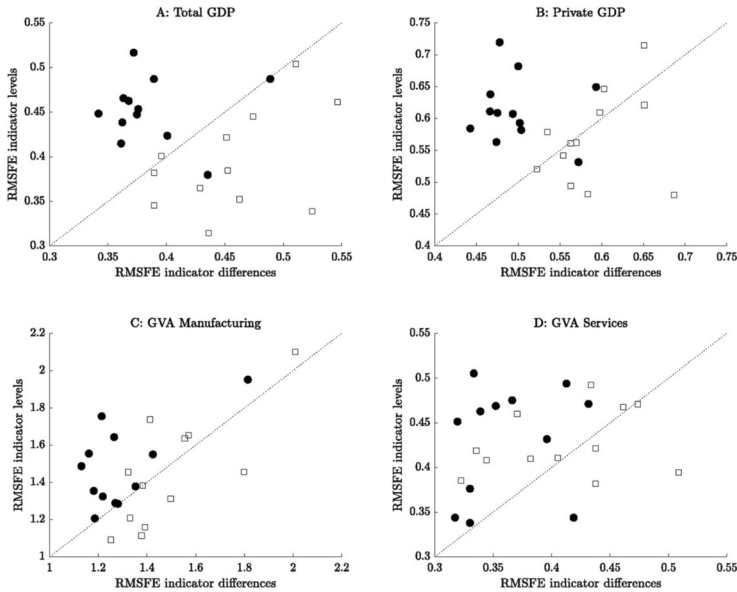
A similar picture emerges by looking at private sector GDP growth. All indicators seem to have practical relevance as the NTS range between 0.38 and 0.51 for the nowcasts and between 0.45 and 0.59 for the forecasts. Nevertheless also a practical gain exists for private GDP growth. On average over both forecasting horizons and models, the application of the best performing indicator can increase the RMSFE by 12% in terms of the series' volatility over the worst performing one.

The target series for manufacturing is the most volatile one—two to almost four times higher compared to the other series. However, all indicators do a good job as the NTS range from 0.37 to 0.60 in the nowcasting case and from 0.41 to 0.66 in the forecasting setup. The practical gain is substantial as for both GDP measures and amounts, on average, to 20% over the two forecasting horizons.

The results for the service sector are similar to those of the previous series. All NTS are below unity and range from 0.36 to 0.49 in the nowcasting setup and lie between 0.36 and 0.57 by looking at one quarter-ahead predictions. The relative improvement of the best over the worst performing indicator is, on average, 16% in terms of the series volatility. Overall, each indicator seems to have practical relevance as the NTS are below unity in each case. However, remarkable gains exist across indicators and it seems worthwhile to apply the best performing ones.

4.2 Levels vs. Differences

There is an ongoing discussion among researchers, applied forecasters, and policymakers on how the survey indicators should be transformed and used for



Notes: The figure compares the root mean squared forecast errors (RMSFE) from a forecasting experiment where the indicators enter in first differences with a forecasting experiment where the indicators enter in levels. The four target series are quarterly growth rates based on real-time data. Both forecast horizons are shown simultaneously. Dots refer to nowcasts, squares refer to forecasts.

Fig. 1 RMSFE comparisons across indicator transformations

economic forecasting. Based on our argumentation in Sect. 2.3 we decided to apply first differences as our baseline specification. However, studies such as Baselier et al. (2018) and de Bondt (2019) show that also level specifications provide sensible results, hence it might be an empirical matter what works best. We therefore compare the forecasting power of the indicators transformed into first differences with their accuracy based on levels.

For this purpose, Fig. 1 presents—for each target series—the comparison of the RMSFEs of the difference specification against those of the level specification. The abscissa depicts the RMSFEs from the baseline results (first differences transformation) for each of the two models and forecast horizons; the ordinate plots the RMSFEs from a comparable forecasting exercise based on the indicator levels.⁹ Dots refer to nowcasts, squares refer to forecasts. A marker above (below) the bisecting line indicates a case where the difference specification provides lower (higher) forecast errors than the level specification. A marker on the bisecting line indicates that both specifications provide an identical average forecast error.

⁹ Both forecasting experiments are comparable as we only exchanged the indicators and let the forecasting framework untouched (model specification, evaluation period etc.).

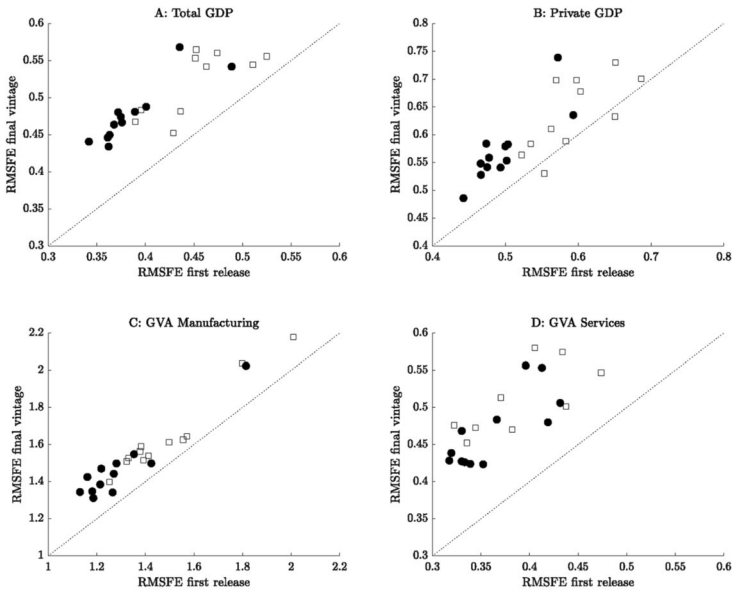
We analyze Fig. 1 along two dimensions. First, focusing on all indicators reveals that the difference specification delivers lower RMSFEs in most of the cases. This result is in line with the existing literature for Germany (see Kholodilin and Siliverstovs 2006; Henzel et al. 2015; Lehmann 2020). In particular, for services and manufacturing (bottom left and bottom right panel) the results are clear-cut—in 75% and 79% of all cases the difference specifications is superior. For private sector GDP growth (top right panel), the results are more blurred. However, also in this case the difference specification is superior in 62% of the cases. Only for total GDP (top left panel) the results are ambiguous—half of the combinations are in favor of first differences and the other half prefers the level specification. Interestingly, this result is driven by the forecast horizon ($h = 1$). Except for one case, all model-indicator-combinations for the forecast horizon (indicated by the squares) lie below the bisecting line. The opposite holds true for the nowcasting case ($h = 0$).

Second, focusing solely on the best performing indicator for each target series reveals that a difference specification is always superior for the nowcast case; for total GDP: AR-X(0,0) with ifo Business Expectation, for private sector GDP: AR-X(0,q) with PMI Composite Output Index, for GVA manufacturing: AR-X(0,0) with ifo Business Expectations Manufacturing, and for GVA services: AR-X(0,q) with PMI Services Business Activity Index. However, in the forecast case a level specification of the best performing indicator is superior to its difference specification for three out of four target series; for total GDP: AR-X(0,0) with PMI Composite Output Index, for private sector GDP: AR-X(0,0) with ZEW Current Economic Situation, and for GVA manufacturing: AR-X(0,0) with ifo Business Situation Manufacturing. Only for GVA services a difference specification (AR-X(0,0) with ZEW Economic Sentiment Services) provides superior RMSFEs. Overall, we cautiously take the results from this comparison as confirmation of our baseline specification. However, our results also suggest that both difference and level specifications provide useful signals. Thus, future research might concern the question of how to optimally combine forecasts from level and difference specifications.

4.3 Real-time vs. Final Data

Finally, we assess the indicators' forecast performance regarding the final data releases. We proceed in two ways. First, we conduct a similar forecast experiment as in Section 3, but in a *pseudo* out-of-sample setting, implying that we estimate the models based on the latest data vintage. This procedure hence abstracts from data revisions. Second, we use the forecasts from Sect. 3 and calculate RMSFEs based on the latest data vintage instead of first releases.

The *pseudo* out-of-sample forecast experiment provides two insights. First, the ranking across indicators remains virtually identical. Thus, our findings from the real-time forecasting experiment also hold for revised data. Second, almost all RMSFEs based on real-time data are smaller in magnitude than their counterparts based on revised data.



Notes: The figure compares the root mean squared forecast errors (RMSFE) from a real-time forecasting experiment that are either calculated by the first releases of the target series or their final vintages. The four target series are quarterly growth rates based on real-time data. Both forecast horizons are shown simultaneously. Dots refer to nowcasts, squares refer to forecasts.

Fig. 2 RMSFE comparisons across first release and final vintage

The previous finding ties up with our second part of this discussion which can be seen as an extension to the first part. Namely we ask whether the forecasts based on real-time data are also good predictors for the final release of the target series. This consideration is in line with the investigations by Strohsal and Wolf (2020), who assess whether the first release of a German macroeconomic aggregate is a good nowcast for the final release. Figure 2 plots the RMSFEs of each model-indicator-forecast horizon combination with regard to the first release (on the abscissa) against the corresponding RMSFEs regarding the final release (ordinate); dots refer to nowcasts, squares refer to forecasts.

For the majority of the cases, the RMSFEs of the first release are smaller than those based on final data. In addition, we ask whether these now- and forecasts are “good” predictions for the latest values. Therefore, we again resort to the NTS by comparing the RSMFE with respect to the final release to the standard deviation of the corresponding target series. For total GDP nowcasts, the NTS ranges from 0.48 to 0.62. Especially the NTS for the best indicators are of similar magnitude as in Strohsal and Wolf (2020). They find that the standard deviation of the revisions go hand in hand with a NTS of 0.44 compared to the series’ volatility. However, they analyze a different time period as we do. Nevertheless we interpret our results in favor of the good performance of the headline indices. The NTS for GDP forecasts lie between 0.51 to 0.68, which is also a quite “good”

performance. A similar picture emerges for private sector GDP: the NTS range for the nowcast is 0.42 to 0.55 and for the forecast 0.46 to 0.61.

Turning to the sectoral forecasts, the indicators also do perform well. For nowcasts in manufacturing, the NTS range from 0.43 to 0.67. Thus, the best performing indicator's forecast errors are less volatile than the underlying target series. The NTS span for the forecasts slightly increases to values between 0.46 to 0.72. Nevertheless, especially the best performing indicators are of high practical relevance. A similar conclusion can be drawn from the NTS of the service sector. The range in the case of nowcasts lies between 0.48 and 0.53; the span for the forecasts is 0.51 to 0.69. Overall, it seems worthwhile to use the indicators for final release predictions with pronounced gains from applying the best performing ones.

5 Conclusion

This analysis conducts a comprehensive forecasting experiment of the headline indices of the four main survey providers in Germany: the ifo Institute, IHS Markit, DG ECFIN, and the ZEW. We do not only focus on total GDP growth, but additionally examine the accuracy for private sector GDP and gross value added in both the manufacturing and the service sector. Our out-of-sample, real-time forecast exercise reveals the following results. For GDP growth, all providers publish useful leading indicators with advantages for the ifo indices in the case of one quarter-ahead forecasts. A similar picture emerges for private sector GDP, with an advantageous nowcast performance of the PMI Composite Output Index for one out of the two models and advantages of the Economic Sentiment Indicator in the case of the forecast setup. For the volatile manufacturing sector, the ifo indicators are clearly superior compared to the headline indices of the other providers. Regarding the service sector, the four institutions provide meaningful leading indicators to nowcast service sector gross value added growth. For one quarter-ahead predictions, the Economic Sentiment Services of the Centre for European Economic Research is superior.

From a researcher's perspective, it would be interesting to analyse the causes of the differences in forecast accuracy, for example in the service sector. However, this would require having detailed information on each provider's panel and its detailed sectoral coverage. Unfortunately, longer time series are not publicly available. Another interesting research idea deals with the "optimization" of the single indicators and the mixture of the different indicator concepts. Whereas, for example, the ifo Business Situation is only based on the outcome of one single question, the Economic Sentiment Indicator is designed as a composite index with several questions entering. In the case of the ZEW, one might also think of combining their Economic Situation Index with its expectations counterpart in style of ifo's Business Climate. For the Confidence Indicators of DG ECFIN one can also ask whether backward- and forward-looking questions have to be combined or only forward-looking questions are preferable for the forecasting performance. Furthermore—given detailed information on the samples—an optimization of the weighting scheme, for example for the ESI, might help to further increase the forecasting properties of each headline index. Finally, future research

could also combine the providers' headline indices and ask whether these combinations further increase the overall absolute forecasting performance for each target series at hand.

Supplementary Information The online version supplementary material available at <https://doi.org/10.1007/s41549-021-00055-5>.

Funding Open Access funding enabled and organized by Projekt DEAL.

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

- Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., & Rünstler, G. (2011). Short-term forecasts of euro area GDP growth. *Econometrics Journal*, *14*(1), 25–44.
- Basselier, R., de Antonio Liedo, D., & Langenus, D. (2018). Nowcasting real economic activity in the euro area: Assessing the impact of qualitative surveys. *Journal of Business Cycle Research*, *14*(1), 1–46.
- Broughton, J. B., & Lobo, B. J. (2018). Herding and anchoring in macroeconomic forecasts: The case of the PMI. *Empirical Economics*, *55*(3), 1337–1355.
- Carstensen, K., Heinrich, M., Reif, M., & Wolters, M. H. (2020). Predicting ordinary and severe recession with a three-state markov-switching dynamic factor model: An application to the German Business Cycle. *International Journal of Forecasting*, *36*(3), 829–850.
- Christiansen, C., Eriksen, J., & Møller, S. (2014). Forecasting US recessions: The role of sentiment. *Journal of Banking & Finance*, *49*, 459–468.
- de Bondt, G. J. (2019). A PMI-based real GDP tracker for the euro area. *Journal of Business Cycle Research*, *15*(2), 147–170.
- European Union (2016). The joint harmonised EU programme of business and consumer surveys, User Guide. https://ec.europa.eu/economy_finance/db_indicators/surveys/documents/bcs_user_guide_en.pdf.
- European Union (2017). European business cycle indicators 2nd quarter 2017. https://ec.europa.eu/info/sites/info/files/tp017_en.pdf.
- Fritsche, U., & Stephan, S. (2002). Leading indicators of German business cycles: An assessment of properties. *Journal of Economics and Statistics*, *222*(3), 289–315.
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.
- Henzel, S., Lehmann, R., & Wohlrabe, K. (2015). Nowcasting regional GDP: The case of the Free State of Saxony. *Review of Economics*, *66*(1), 71–98.
- Hüfner, F. P., & Schröder, M. (2002). Prognosegehalt von ifo-Geschäftserwartungen und ZEW-Konjunkturerwartungen: Ein ökonomischer Vergleich. *Journal of Economics and Statistics*, *222*(3), 316–336.
- IHS Markit (2017). An introduction to the PMI surveys. Retrieved July 18, 2017, from <https://cdn.ihs.com/www/pdf/1218/IHS-Markit-PMI-Introduction.pdf>.
- IHS Markit (2019). IHS Markit Flash Germany PMI: German economy continues to underperform in August. News Release on August 22, 2019.

- Kholodilin, K. A., & Siliverstovs, B. (2006). On the forecasting properties of the alternative leading indicators for German GDP: Recent evidence. *Journal of Economics and Statistics*, 226(3), 234–259.
- Köhler, M., & Schmidt, S. (2006). ZEW Financial Market Survey, short summary. http://ftp.zew.de/pub/zew-docs/div/Kurzinfo_English.pdf.
- Lehmann, R. (2020). The forecasting power of the ifo Business Survey. CESifo Working Paper No. 8291.
- Pinkwart, N. (2018). Short-term forecasting economic activity in Germany: A supply and demand side system of bridge equations. Deutsche Bundesbank Discussion Paper 36/2018.
- Sauer, S., & Wohlrabe, K. (2018). The new ifo Business Climate Index for Germany. *CESifo Forum*, 19(2), 59–64.
- Schröder, M., & Hüfner, F. P. (2002). Forecasting economic activity in Germany - How useful are sentiment indicators? ZEW Discussion Paper No. 02–56.
- Strohsal, T., & Wolf, E. (2020). Data revisions to German national accounts: Are initial releases good nowcasts? *International Journal of Forecasting*, 36(4), 1252–1259.

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