Bank i Kredyt 52(2), 2021, 123-142

BANK I KREDYT
BANK & CREDIT

Polish GDP forecast errors: a tale of inefficiency

Jakub Rybacki*

Submitted: 28 January 2020. Accepted: 2 September 2020.

Abstract

The aim of this paper is to evaluate gross domestic product (GDP) forecast errors of Polish professional forecasters based on the individual data from the *Rzeczpospolita* daily newspaper. This dataset contains predictions obtained from forecasting competitions during the years 2013–2019 in Poland. Our analysis shows a lack of statistical efficiency of these predictions. First, there is a systemic negative bias, which is especially strong during the years 2016–2019. Second, the forecasters failed to correctly predict the effects of major changes in fiscal policy. Third, there is evidence of strategic behaviour; for example, the forecasters tended to revise their estimates too frequently and too excessively. We also document herding behaviour, i.e. an alignment of the most extreme forecasts towards market consensus with time, and an overly strong reliance on forecasts from NBP inflation projections in cases of estimates for longer horizons.

Keywords: GDP forecasting

JEL: E32, E37

* Warsaw School of Economics, Collegium of Economic Analysis and Polish Economic Institute; e-mail: jr70663@sgh.waw.pl.

1 Introduction

The aim of this paper is to evaluate gross domestic product (GDP) forecast errors of Polish professional forecasters based on individual data from the *Rzeczpospolita* daily newspaper. The dataset contains predictions from forecasting competitions and covers the years 2013 to 2019. Based on statistical tests, we analyse the unbiasedness and Nordhaus' efficiency (1987) of the forecasts, as well as potential irregularities in the forecast revisions and consensus formation. We propose three hypotheses:

1. There is no systematic bias in the GDP growth forecast.
2. Forecasts are efficient by the Nordhaus definition, i.e. prior revision does not help to predict next forecast modifications.
3. Professional forecasters do not alter their forecasts based on the information regarding market consensus.

All the three hypotheses were rejected in our study. We identified the following problems:

1. The systematic underestimation of growth dynamics, particularly visible during the years 2016 to 2019.

2. An inability to correctly forecast the effects of changes related to fiscal policy or structural reforms. Professional forecasters significantly underestimated the consequences of the transition to the new EU budget perspective, which resulted in a contraction of investment growth. They probably also overestimated the effects of the introduction of child benefit. These errors resulted in forecast revisions of the greatest magnitude and the biggest surprises in the analysed sample.

3. Excessive and overly frequent revisions of activity forecasts. A strong revision in quarter t tends to be reversed in the next quarter ($t + 1$). Similar to the forecast errors, revisions are more likely to be positive rather than negative, especially after 2016.

4. Evidence of strategic behaviour. Two forecasters tended to more strictly follow market consensus rather than produce controversial estimates. There was also a tendency to align the most extreme forecasts toward a market consensus; this is known as herding behaviour. Finally, disagreement tends to be lower for the forecast with a longer time horizon. Probably, professional forecasters are putting less effort in developing in-house models for long-term forecasting and they are anchoring their expectations closely on the official NBP GDP estimates presented in the inflation projection of the central bank.

Problems like herding behaviour of forecasters or inability to predict downturns or structural changes are not unique for Poland. They are also reported in the G10 economies. However, we highlight two solutions which may limit the inefficiency of forecasts. First, in Poland, the market for economic forecasts is dominated by the commercial banks – 90% of the forecasts are produced by representatives of those entities. Greater participation of public sector entities, i.e. the Ministry of Finance, the Ministry of Economy, Narodowy Bank Polski (NBP), and the Polish Economic Institute, may be beneficial. Second, there are no systemic incentives for the academic sector to shape the public debate and regularly present economic forecasts. Modelling competitions organized by public sector entities may help to activate this group. During such contests participants should provide models capable of forecasting macroeconomic variables in the subsequent quarters (a good example is G-research's NBA Data Challenge). Publicly available information on the models' consensus should also enforce greater discipline amongst the professional forecasters, as this group would definitely need to explain deviations of their opinions from the indications of algorithms.

This manuscript is structured as follows. Section 2 provides a literature review on GDP forecasting, describing irregularities visible especially in the G7. Section 3 delivers information about the dataset. Section 4 summarizes the methodology. Section 5 discusses the results of the estimation. Finally, section 6 concludes the paper.

2 Literature review

The aim of this section is to present problems related to GDP growth forecasting reported previously in the academic literature. The prediction of a business cycle is probably one of the most sophisticated exercises done by economists and forecast errors are usually greater compared to other economic figures such as inflation (Lahiri, Sheng 2010; Loungani, Stekler, Tamirisa 2013). Furthermore, researchers report persistent systematic biases which are especially visible in the forecasts with longer horizons (Ager, Kappler, Osterloh 2009). There is also strong evidence of failure to predict severe downturns (Loungani 2001) or effects of structural changes, e.g. those related to fiscal policy (Blanchard, Leigh 2013).

The Polish economic debate is particularly driven by commercial economists representing the banking sector. The literature on the subject highlights a few significant problems related to such situations.

First, the primary aim of such forecasters is not necessarily to minimize forecast errors, but rather to realize some other strategic objectives, for example, a greater presence in the media or triggering some policy actions (Pons-Novell 2003; Dovern, Weisser 2011). This may result in two opposite phenomena: either strong and systematic deviations from the market consensus or self-censorship to avoid such discrepancies (i.e. herding behaviour). Some authors (e.g. Ashiya 2009) claim that forecasting may reflect the interests of the forecaster employer – for example representatives of banking sectors in some periods may be more pessimistic than academics and the difference between estimates is statistically significant.

Second, there is a widespread debate about irregularities visible in the revisions of the forecasts made by professionals. In a perfect world, the pattern of the revisions would be totally unpredictable and follow a random walk process (Nordhaus 1987); however, this is not always the case. Several studies operating on monthly data show that forecasts are too rigid and too sluggish in incorporating incoming information (Lahiri, Sheng 2010; Loungani, Stekler, Tamirisa 2013; Capistrán, López-Moctezuma 2014). Such inefficiency can be explained by two phenomena: 1) smoothing estimates for reputational reasons or 2) informational rigidities of forecasters, i.e. analysts are not willing to send multiple notes regarding small amendments to forecasts and instead wait for an opportunity for a greater correction (Jain 2018).

There are also reports providing examples of strategic behaviour such as presenting overly optimistic or pessimistic estimates in order to acquire publicity (Ashiya 2003). Finally, forecast revisions sometimes may be used to trigger significant policy actions or affect valuation in a financial instrument. However, these effects are rather more frequently seen in the case of publicly listed companies' earnings announcements (e.g. Gleason, Lee 2003; Kasznik, McNichols 2002), rather than in relation to macroeconomic variables.

Finally, several papers suggest there is an interaction between public and private sector forecasts. Behavioural economists suggest the effect of an anchoring bias (Campbell, Sharpe 2009); there is also evidence that the establishment of public forecasts may crowd out some efforts from the private sector (Tong 2007).

We propose three different statistical tests to identify if forecasts are unbiased and free of strategic behaviour. The detailed information will be presented in the methodology section.

3 The *Rzeczpospolita* forecasting competition

This section describes the dataset used in this study. The *Rzeczpospolita* competition was established in 2008 by NBP President Sławomir Skrzypek to promote better macroeconomic forecasting. The competition initially contained five categories of forecasts: gross domestic product (GDP), gross fixed capital formation (GFCF), consumer price index (CPI), unemployment rate, and the current account of the balance of payments. NBP abandoned supporting the competition in 2015. After this event, *Rzeczpospolita* modified the forecast variables: private consumption and exchange rate forecasts were added; there was no further interest in forecasting the current account of the balance of payments or unemployment rate.

The *Rzeczpospolita* survey is conducted quarterly. Polled analysts provide their estimates for the four quarters ahead. For example, at the end of September, analysts provide their estimates for the last quarter of the survey year and the first, second, and third quarters of the following year. At the time of the survey, information regarding GDP growth in the current quarter is still unavailable, and the analysts must base their estimates on monthly data (e.g. industrial production and construction output). In December, the window moves by one quarter, and at that time, the surveyed analysts are unaware of the GDP reading for the current quarter, and so on.

The dataset used for this study consist of the individual forecasts covering the period from 2013 to 2019. We excluded the participants who posted their estimates irregularly or belonged to student associations (due to frequent rotations of the forecasters). Therefore, we were left with forecasts from 20 permanent contributors: 90% of them representing financial institutions and 10% representing academic institutions or think tanks.

4 Methodology

This section presents the methodology of our research. Our aim was to analyse the efficiency of the GDP forecasts for the Polish economy based on two independent tests. We also verify whether there is evidence of strategic behaviour by the forecasters, following the approach of Pons-Novell (2003).

Below, we define the key variables used in the analysis:

GDP_t denotes the annual growth of gross domestic product in the quarter t;

$GDP^{f_i}_{t,h}$ represents the i-th professional forecaster's prognosis of gross domestic product in the quarter t, formulated h quarters prior to the reading;

i takes values from 1 to n, where n denotes the number of forecasters; we will use the superscript f_i every time a variable is related to the forecasts and not to a realized macroeconomic reading;

μ_i stands for the individual error of the *i*-th professional forecaster, estimated using fixed effects model;
θ_t denotes a time period effect;
ε_t represents a random disturbance;
β_x are estimated parameters.

4.1 Efficiency of forecasts: the first statistical test

The first test of forecast efficiency assumes that forecast errors should have no systematic bias. Therefore, the test should validate our first hypothesis. Unbiasedness is an initial requirement to meet criteria of the rational expectations hypothesis (Lucas, Sargent 1981).

We follow an approach used previously by Ashiya (2003, 2009), Loungani (2001), and Lahiri and Sheng (2010). This test is also widely adopted in different contexts, for example, with fiscal forecasts (Artis, Marcellino 2001; Brück, Stephan 2006; Pina, Venes 2011).

We formulate the following equation using an ordinary linear regression with cross-section and period fixed effects:

$$GDP_t = \beta_0 + \beta_1 \cdot GDP_{t,h}^{f_i} + \mu_i + \theta_t + \varepsilon_t \tag{1}$$

An efficient forecast should meet the following criteria:

1. There are no systematic biases. Therefore, parameters β_0 and μ_i should both be statistically insignificant for each forecaster.

2. Forecasts should correctly describe the final realization of GDP, except for some random disturbances related to θ_t and ε_t. This implies that $\beta_1 = 1$.

θ_t is a white noise series.

We assume there is no multiplicative error in the forecast. Therefore, we simplify the first equation and provide the following model:

$$GDP_{t,h}^{f_i} - GDP_t = \beta_0 + \mu_i + \theta_t + \varepsilon_t \tag{2}$$

Our aim is to test and verify the hypotheses presented in criteria (1) and (3). If those criterions are met, we would be capable of validating our first main hypothesis.

4.2 Efficiency of forecasts: the second statistical test

To verify the second hypothesis, we follow the approach proposed by Nordhaus (1987). We attempt to verify whether the forecast revisions indeed follow a white noise process. Let us denote $REV_{t,h}^{f_i}$ as the magnitude of a forecast revision of GDP in the quarter *t*, prepared *h* quarters prior to the reading. The indicators denote the difference between the most recent forecast at the time $(t - h)$ and the previous one, made in the period $(t - h - 1)$. The computation is given by the following formula:

$$REV_{t,h}^{f_i} = GDP_{t,h}^{f_i} - GDP_{t,h-1}^{f_i} \tag{3}$$

We attempt to estimate the following autoregressive model with fixed and period effects:

$$REV_{t,h}^{f_i} = \beta_0 + \beta_1 \cdot REV_{t,h-1}^{f_i} + \mu_i + \theta_t + \varepsilon_t \tag{4}$$

Our aim is to verify the following hypotheses:

1. A past revision should not give information regarding the forecaster's next decision. If the forecasts are effective, then parameter β_1 is required to be statistically insignificant.

2. The cross-section fixed effect μ_i should be statistically insignificant.

3. θ_t should be a white noise series.

We will also verify whether the magnitude of forecast revisions differ substantially between professionals. Some analysts may have a greater propensity to perform stronger revisions and to do so more frequently just to attract greater attention from the media. Therefore, we will compute the absolute values of the magnitudes of the forecast revisions and average them.

4.3 Strategic behaviour of forecasters

Finally, based on the approach of Pons-Novell (2003), we attempt to verify whether there exists any evidence of strategic behaviour on the part of the forecasters (hypothesis 3). This methodology focuses on deviation of individual forecasts from the market consensus. We define the market consensus as the median of available forecasts.

$$Consensus_{t,h}^{f_i} = Median\left(GDP_{t,h}^{f_1}, GDP_{t,h}^{f_2}, \ldots\ldots, GDP_{t,h}^{f_n}\right) \tag{5}$$

Our aim is to analyse deviations from the market consensus, calculated by a simple subtraction.

$$Deviation_{t,h}^{f_i} = GDP_{t,h}^{f_i} - Consensus_{t,h}^{f_i} \tag{6}$$

First, we analyse whether the magnitude of such deviations differs significantly between the forecasters. Forecasters can estimate the value of a market consensus prior to its publication as real--time information is available on the Bloomberg terminal; monthly estimates are also aggregated by a Consensus Economics poll. Therefore, some groups of analysts may have the temptation to self-censor their estimates and not deviate strongly from the median.

To perform this exercise, we calculate the deviations of each forecaster's projections from the consensus using equation 6. Then, we compute the absolute values of those deviations and average them separately for each forecaster. We perform a single *t*-test to verify whether the average deviation produced by a single forecaster is substantially different from those of other professionals. Forecasters strictly following the consensus should have substantially lower deviations, whereas economists lobbying for some policy action would produce greater deviations.

Second, we attempt to identify evidence of herding behaviour using the following model:

$$Deviation_{t,h}^{f_i} = \beta_0 + \beta_1 \cdot Deviation_{t,h-1}^{f_i} + \mu_i + \theta_t + \varepsilon_t \tag{7}$$

Values of β_1 lower than 1 suggest that forecasters are prone to correct their deviations and move closer to the consensus values as the forecast horizon shortens and estimates start to gain greater publicity. In such case conditions for validating the 3rd hypothesis are not met.

5 Estimation results

In this section, we present and discuss the results of our estimation. We find evidence of strategic behaviour, and of both group and individual biases in the forecasts.

5.1 First test: analysis of forecast errors

The first test confirms the inefficiency of professional market forecasts. First, the constant parameter β_0 is negative and statistically significant for all forecast horizons. Forecasts published in the examined window (2013–2019) tend to underestimate GDP growth dynamics by 0.4 to 0.6 percentage points (further pp). Detailed results are presented in Table 1.

Second, based on the statistical tests for redundant cross-section effects, we reject the null hypothesis that parameters corresponding to such effects are statistically insignificant (equal to zero). The problem of individual biases is quite visible in the parameter estimates. One survey respondent tended to systematically present much more pessimistic forecasts as compared to the average derived for other respondents; the discrepancy amounts to another 0.5–0.6 pp for the longer horizon (Q3–Q4). The respondent with the second largest negative bias provided forecasts that were 0.2–0.3 pp lower than the average. The estimated cross-section of effects is presented in Table 2.

Finally, the estimated time period effects do not represent a white noise process. There are two episodes confirming problems in forecasting structural changes and downturns. First, analysts overestimated the potential effect of the introduction of the child benefit programme in 2015 and early 2016. Second, they were incapable of predicting the duration of the slowdown related to the contraction of investment during the transition between the EU budget perspectives. There was also a systematic shift in GDP forecasting errors for the years 2016 to 2019. During this time, forecasts were overly negative. This phenomenon may be related to a negative assessment of the economic policies proposed by the PiS government. The period effects are presented in Figure 1. In any case, based on this test we rejected the first hypothesis.

5.2 Second test: analysis of forecast revisions

The second test also confirms that forecasts are statistically ineffective. Before analysing the model output, we should note that the magnitude of revision is different, depending on the time horizon and market participant. Detailed data is presented in Table 3.

The strongest revisions occur in the quarters directly preceding the publication of data. The magnitude of revisions becomes lower, as the forecast's horizon increases. There is a group of respondents (i.e. 8, 9, 12, and 16) who tend to revise forecasts more sharply compared to others. There is also a group that make significantly smaller revisions (respondents 14 and 20).

The model confirms the existence of autoregressive patterns visible in the data. There is statistically significant evidence that forecasters are prone to making excessively strong changes in their prognosis. A negative parameter of β_1 (-0.3) indicates that a revision made in the previous quarter is usually corrected in the next round of polls. There is also evidence of systematic upward revisions: parameter β_0 is positive in the case of both forecast horizons. This evidence confirms the problem of systematic bias observed with the previous test. A detailed description of the model is available in Table 4.

Estimates of cross-fixed effects and period effects are presented in Figures 2 and 3. Similar to the findings with the first test, there is evidence of persistent one-sided revisions during the years 2017–2019. Therefore, this test also rejected the second hypothesis about forecast efficiency.

5.3 Third test: herding behaviour

Finally, we also studied whether there is visible strategic behaviour regarding an approach to market consensus. As with the previous test, we start from an analysis of descriptive data shown in Table 5.

First, absolute values of deviation from market consensus and forecast disagreement are greater in the short term. Contrary to intuition and statistics, there is evidence of decaying disagreement with a longer forecast horizon. Commercial research teams are usually small (3–5 people) and have a limited capacity. From the perspective of their employer, greater importance is related to commenting the most recent macroeconomic figures and their nowcasting, rather than building in-house long-term models. Therefore, chief economists in these institutions are strongly tempted to follow central bank inflation projections with minor subjective adjustments. Such a phenomenon was confirmed earlier by the subject literature (see e.g. Kotlowski 2015). However, NBP does not provide public access to quarterly forecasts; therefore, we are not capable of replicating such research.

Second, the first two participants tended to more frequently formulate forecasts that do not deviate from the current market consensus (indexed as 1 and 2). Simultaneously, one of the pessimistic participants identified in the first test was also much more likely to deviate more strongly than the others from the market consensus in the Q3–Q4 horizon.

The third model, specified in equation 7, confirms the existence of herding behaviour. Parameter β_1 is lower than 1 in each time horizon. Alignment toward consensus is most visible during a period of one to three quarters prior to publication. The model's specification is presented in Table 6. All in all, we rejected the last third hypothesis.

6 Policy conclusions

The analysis of the Polish macroeconomic forecasts shows all the major imperfections identified in the subject literature, i.e. the existence of systematic biases and problems with correct forecasting of structural and fiscal changes. There is also strong evidence of strategic behaviour, seen in excessively strong forecast revisions and a willingness to align with the market consensus. Test 3 and the statistically different magnitudes of revision presented in Table 2 suggest that this behaviour is partially intentional, i.e. forecasters who make the most excessive revisions are likely to attract greater media attention. Therefore, total elimination of the identified problems is probably impossible. However, it is worth considering ways to minimize the influence of dishonest behaviour. The probably viable option for public institutions is to intervene in order to alter the structure of the forecasters' market.

There is a discrepancy between the share of professional forecasters from the banking sector in Poland (90%) and G7 (around 50% in the eurozone according to Bowles et al. 2007). A greater diversification of forecasters' backgrounds may be beneficial, i.e. the commercial analysts frequently revise their estimates instantly after publication of new data for the sake of appearance and in order to provide guidance to bank dealers. Test 2 showed that in the case of a strong surprise, these revisions are often excessive. The existence of forecasters which have different priorities should balance market consensus.

A short-term solution may be provided by a more active engagement of public institutions in the debate in Poland. Presently, the government's forecasts are provided twice per year (around April and October), and the central bank's forecasts are provided three times per year (March, July, and November); more frequent projections and auditing of errors should foster the debate. As mentioned earlier, such decisions also have adverse effects, such as eliminating some private participants (Tong 2007).

The long-term problem is related to the low activity of academic and non-governmental institutions in these debates. Again, the public sector should provide incentives for greater participation in the public debates, for example, by using grant schemes. Some competition in developing forecasting models may also help to improve the forecasting market (see the example of G-research's NBA Data Challenge).

References

Ager P., Kappler M., Osterloh S. (2009), The accuracy and efficiency of the Consensus Forecasts: a further application and extension of the pooled approach, *International Journal of Forecasting,* 25(1), 67–181.

Artis M., Marcellino M. (2001), Fiscal forecasting: the track record of the IMF, OECD and EC, *Econometrics Journal,* 4, 20–36.

Ashiya M. (2003), Testing the rationality of Japanese GDP forecasts: the sign of forecast revision matters, *Journal of Economic Behavior & Organization,* 50(2), 263–269.

Ashiya M. (2009), Strategic bias and professional affiliations of macroeconomic forecasters, *Journal of Forecasting,* 28(2), 120–130.

Blanchard O.J., Leigh D. (2013), Growth forecast errors and fiscal multipliers, *American Economic Review,* 103(3), 117–120.

Bowles C., Friz R., Genre V., Kenny G., Meyler A., Rautanen T. (2007), *The ECB survey of professional forecasters (SPF): a review after eight years' experience*, ECB Occasional Paper, 59.

Brück T., Stephan A. (2006), Do Eurozone countries cheat with their budget deficit forecasts? *Kyklos,* 59(1), 3–15.

Campbell S.D., Sharpe S.A. (2009), Anchoring bias in consensus forecasts and its effect on market prices, *Journal of Financial and Quantitative Analysis,* 44(2), 369–390.

Capistrán C., López-Moctezuma G. (2014), Forecast revisions of Mexican inflation and GDP growth, *International Journal of Forecasting,* 30(2), 177–191.

Dovern J., Weisser J. (2011), Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: an empirical comparison for the G7, *International Journal of Forecasting,* 27(2), 452–465.

Jain M. (2018), *Sluggish forecasts*, Bank of Canada Staff Working Paper, 2018-39.

Gleason C.A., Lee C.M. (2003), Analyst forecast revisions and market price discovery, *The Accounting Review,* 78(1), 193–225.

Kasznik R., McNichols M.F. (2002), Does meeting earnings expectations matter? Evidence from analyst forecast revisions and share prices, *Journal of Accounting Research,* 40(3), 727–759.

Kotlowski J. (2015), Do central bank forecasts matter for professional forecasters?, *Czech Journal of Economics and Finance (Finance a uver),* 65(6), 432–454.

Lahiri K., Sheng X. (2010), Learning and heterogeneity in GDP and inflation forecasts, *International Journal of Forecasting,* 26(2), 265–292.

Loungani P. (2001), How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth, *International Journal of Forecasting,* 17(3), 419–432.

Loungani P., Stekler H., Tamirisa N. (2013), Information rigidity in growth forecasts: some cross--country evidence, *International Journal of Forecasting,* 29(4), 605–621.

Lucas R.E., Sargent T.J. (1981), *Rational Expectations and Econometric Practice*, vol. 2, University of Minnesota Press.

Nordhaus W.D. (1987), Forecasting efficiency: concepts and applications, *The Review of Economics and Statistics,* 69(4), 667–674.

Pina A., Venes N. (2011), The political economy of EDP fiscal forecasts: an empirical assessment, *European Journal of Political Economy,* 27(3), 534–546.

Pons-Novell J. (2003), Strategic bias, herding behaviour and economic forecasts, *Journal of Forecasting,* 22(1), 67–77.

Tong H. (2007), Disclosure standards and market efficiency: evidence from analysts' forecasts, *Journal of International Economics,* 72(1), 222–241.

Appendix

Table 1

Test 1 – bias in estimated forecasts

	Forecast horizon			
	1Q ahead forecast	**2Q ahead forecast**	**3Q ahead forecast**	**4Q ahead forecast**
Model Constant	-0.46 (0.01; 0.00***)	-0.52 (0.01; 0.00***)	-0.59 (0.01; 0.00***)	-0.60 (0.01; 0.00***)
R-squared	0.88	0.89	0.90	0.90
Observations	427	408	425	424
Periods	22	21	22	22
Cross-sections	20	20	20	20

Notes:
This table presents the parameter estimates of β_0 for different forecast horizons. The model specification is presented in equation 2. Negative parameters for the model constant (the second row) denote that GDP forecasts were overly pessimistic in the analysed period (2013–2019).

Table 2
Test 1 – estimated cross-section effects

Respondent	Estimated fixed effects for different forecast horizons				Standardized values (number of standard deviations from the mean)			
	1Q ahead forecast	2Q ahead forecast	3Q ahead forecast	4Q ahead forecast	1Q ahead forecast	2Q ahead forecast	3Q ahead forecast	4Q ahead forecast
1	2	3	4	5	6	7	8	9
1	-0.01	0.08	0.07	0.07	-0.15	0.62	0.41	0.33
2	0.02	0.02	0.12	0.16	0.17	0.18	0.66	0.81
3	-0.12	-0.10	-0.05	-0.06	-1.27	-0.77	-0.29	-0.30
4	-0.08	-0.06	-0.02	-0.09	-0.92	-0.41	-0.10	-0.45
5	0.06	0.11	0.03	0.04	0.62	0.82	0.14	0.19
6	0.05	0.02	0.00	0.04	0.57	0.14	-0.03	0.21
7	-0.12	-0.14	-0.10	0.00	-1.28	-1.06	-0.56	0.00
8	-0.11	-0.22	-0.25	-0.26	-1.25	-1.63	-1.39	-1.29
9	0.02	0.09	0.25	0.26	0.21	0.69	1.35	1.30
10	0.18	0.18	0.23	0.27	2.01	1.35	1.26	1.35
11	0.03	0.05	0.03	-0.06	0.32	0.36	0.17	-0.29
12	0.04	0.07	-0.04	-0.01	0.42	0.49	-0.21	-0.03
13	-0.01	0.02	0.02	-0.07	-0.13	0.15	0.12	-0.34
14	-0.03	-0.03	-0.01	-0.01	-0.32	-0.24	-0.06	-0.06
15	0.02	0.06	0.06	0.04	0.17	0.47	0.32	0.19
16	0.02	-0.06	-0.12	-0.04	0.22	-0.45	-0.65	-0.19
17	0.09	0.06	0.13	0.19	1.01	0.43	0.71	0.96
18	-0.04	-0.02	0.04	-0.05	-0.42	-0.17	0.19	-0.24
19	-0.17	-0.36	-0.58	-0.64	-1.84	-2.67	-3.19	-3.21
20	0.17	0.23	0.21	0.21	1.86	1.70	1.14	1.06
Average	0.00	0.00	0.00	0.00				
Std. dev.	0.09	0.14	0.18	0.20				

Notes:
This table presents the parameter estimates of μ_i for different forecast horizons (columns 2–5). Columns 6–9 show standardized values. The model specification is presented in equation 2. Respondents 8 and 19 systematically present more negative forecasts compared to other professionals; given that the β_0 values are negative (Table 1), they are more biased than their competitors. Forecasters 10 and 20 are less biased.

Table 3
Test 2 – Magnitude of revision (pp)

Respondent	Magnitude of revisions (absolute value)			Standardized values (number of standard deviations from the mean)		
	3Q before publication	2Q before publication	1Q before publication	3Q before publication	2Q before publication	1Q before publication
1	2	3	4	5	6	7
1	0.21	0.33	0.36	-1.11	0.06	0.17
2	0.30	0.30	0.26	0.66	-0.32	-1.18
3	0.23	0.26	0.45	-0.67	-0.78	1.30
4	0.26	0.38	0.34	-0.10	0.59	-0.12
5	0.26	0.34	0.35	-0.18	0.17	-0.03
6	0.27	0.17	0.30	0.17	-1.71	-0.69
7	0.25	0.25	0.23	-0.32	-0.91	-1.63
8	0.28	0.48	0.47	0.40	1.68	1.62
9	0.31	0.41	0.48	0.97	0.94	1.70
10	0.28	0.30	0.33	0.38	-0.32	-0.26
11	0.24	0.26	0.31	-0.55	-0.80	-0.55
12	0.28	0.51	0.41	0.36	2.02	0.74
13	0.22	0.30	0.34	-0.98	-0.32	-0.16
14	0.18	0.23	0.26	-1.85	-1.04	-1.21
15	0.31	0.36	0.38	1.03	0.31	0.44
16	0.35	0.47	0.38	1.86	1.57	0.44
17	0.32	0.28	0.37	1.21	-0.56	0.32
18	0.30	0.31	0.34	0.84	-0.22	-0.08
19	0.27	0.41	0.44	0.08	0.93	1.17
20	0.16	0.21	0.20	-2.22	-1.28	-1.99
Average	0.26	0.33	0.35			
Std. dev.	0.05	0.09	0.08			

Notes:
This table presents the absolute values of forecast revisions (columns 2–4). Columns 5–7 show standardized values. The magnitude of the revisions was calculated with the formula presented in equation 3. Respondents 8, 9, 12, and 16 tended to make bigger revisions compared to the others; respondents 14 and 20 made smaller revisions.

Table 4

Test 2 – autoregressive models of forecast revisions

	Horizon	
	1Q ahead revision	**2Q ahead revision**
Model constant	0.11 (0.01; 0.00***)	0.05 (0.01; 0.00***)
Previous revision	-0.32 (0.05; 0.00***)	-0.30 (0.06; 0.00***)
R-squared	0.63	0.58
Observations	438	438
Periods	23	23
Cross-sections	20	20

Notes:
This table presents the parameter estimates of β_0 and β_1 for different forecast horizons. The model specification is presented in equation 4. Positive values of the model constant β_0 (the second row) confirm that the forecasters were overly pessimistic – the number of positive revisions is greater than the number of negative ones. Negative values of the model constant β_0 after the previous revision (the third row) imply that the forecasters are prone to making overly strong revisions; the changes are often reversed in the next quarter.

Table 5

Test 3 – average of absolute deviations from the market consensus

Respondent	Average deviation from market consensus (absolute value)				Standardized values (number of standard deviations from the mean)			
	1Q ahead forecast	**2Q ahead forecast**	**3Q ahead forecast**	**4Q ahead forecast**	**1Q ahead forecast**	**2Q ahead forecast**	**3Q ahead forecast**	**4Q ahead forecast**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	0.16	0.15	0.17	0.16	-2.53	-2.72	-2.36	-2.12
2	0.37	0.41	0.29	0.28	-1.58	-1.12	-1.36	-1.06
3	0.56	0.43	0.33	0.34	-0.72	-1.01	-1.04	-0.57
4	0.51	0.46	0.46	0.39	-0.92	-0.79	0.03	-0.12
5	0.52	0.40	0.40	0.34	-0.90	-1.18	-0.46	-0.53
6	0.58	0.51	0.45	0.42	-0.61	-0.50	-0.10	0.15
7	0.74	0.71	0.49	0.34	0.10	0.75	0.28	-0.53
8	0.85	0.89	0.64	0.46	0.62	1.86	1.43	0.52
9	0.79	0.68	0.45	0.44	0.33	0.59	-0.07	0.36
10	0.75	0.60	0.47	0.43	0.15	0.08	0.09	0.29
11	0.86	0.74	0.58	0.52	0.66	0.91	0.95	1.03
12	0.73	0.62	0.37	0.34	0.05	0.17	-0.71	-0.54
13	0.84	0.61	0.44	0.53	0.57	0.13	-0.15	1.10
14	0.87	0.59	0.43	0.37	0.71	0.03	-0.26	-0.28
15	1.07	0.66	0.46	0.35	1.61	0.47	-0.04	-0.40
16	0.98	0.67	0.62	0.42	1.22	0.51	1.28	0.19
17	0.94	0.69	0.56	0.49	1.04	0.63	0.79	0.77
18	0.86	0.74	0.43	0.35	0.67	0.93	-0.21	-0.43
19	0.64	0.59	0.72	0.72	-0.34	0.00	2.07	2.81
20	0.69	0.63	0.44	0.33	-0.12	0.26	-0.17	-0.66
Average	0.71	0.59	0.46	0.40				
Std. dev.	0.22	0.16	0.12	0.11				

Notes:
This table presents the absolute values of forecast deviations from the market consensus (columns 2–5). Columns 6–9 show standardized values. Deviations were calculated with the formulae presented in equations 5 and 6. Respondents 1 and 2 tended to strictly follow the consensus. This may be a result of strategic behaviour rather than use of independent models.

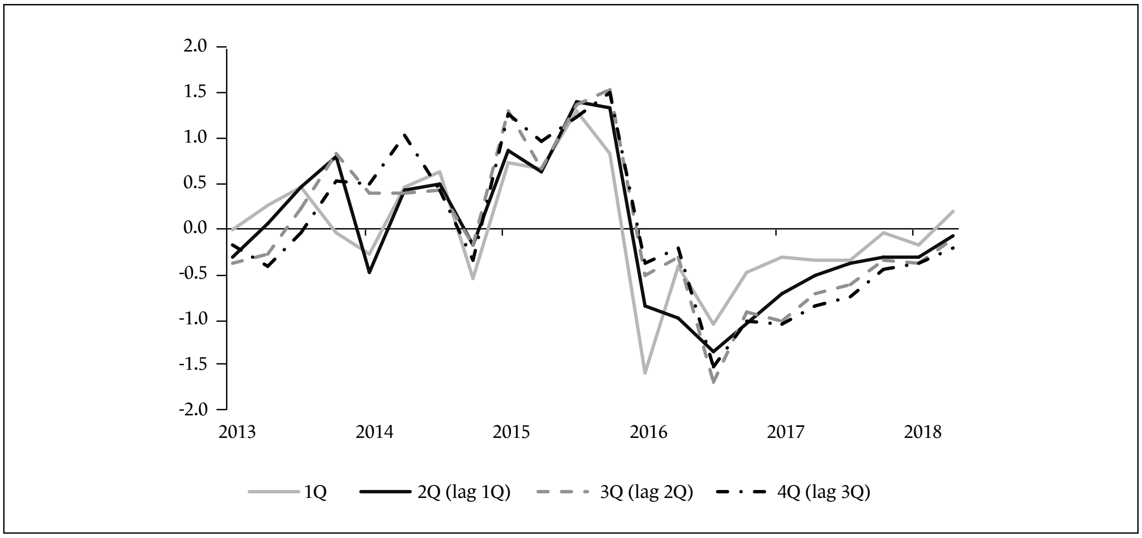
Table 6
Test 3 – deviation from market consensus

	Forecast horizon			
	1Q ahead forecast	**2Q ahead forecast**	**3Q ahead forecast**	**4Q ahead forecast**
Model Constant	0.48 (0.05, 0.00***)	0.31 (0.04, 0.00***)	0.21 (0.03, 0.00***)	0.19 (0.03, 0.00***)
Deviation in the previous quarter	0.49 (0.07, 0.00***)	0.70 (0.11, 0.00***)	0.55 (0.10, 0.00***)	
Deviation of the previous forecast				0.19 (0.07, 0.01***)
R-squared	0.40	0.44	0.66	0.62
Observations	286	317	381	381
Periods	22	23	25	25
Cross-sections	20	20	20	20

Notes:
This table presents the parameter estimates of β_0 and β_1 for different forecast horizons. The model specification is presented in equation 7. Values of β_1 lower than 1 (rows 3 and 4) imply that the forecasters are self-censoring to avoid large deviations from the market consensus (herding behaviour).

Figure 1

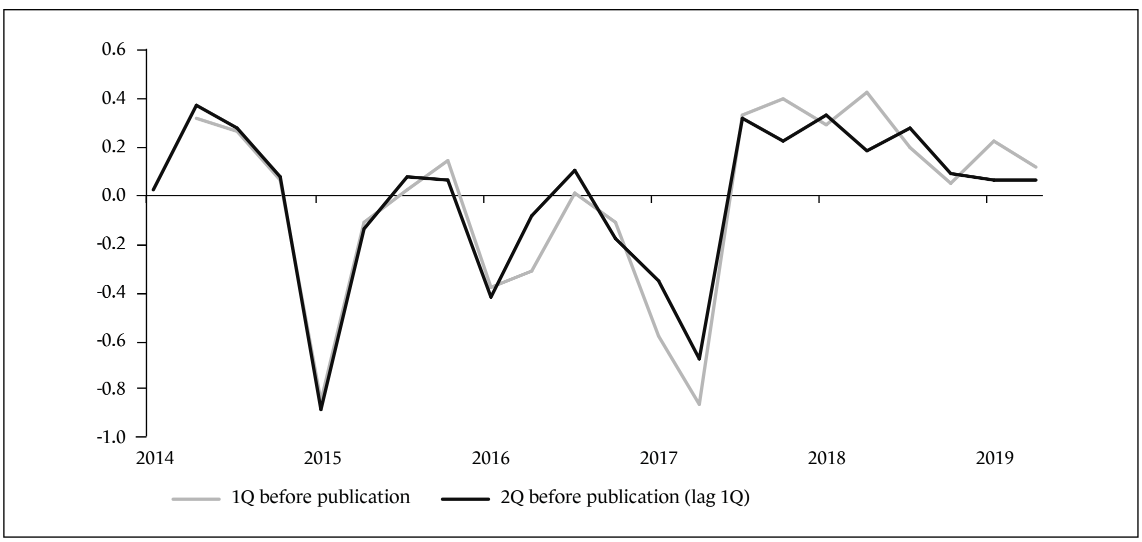
Test 1 – visualization of period effects



Notes:
This figure presents the estimated values of θ_t for different forecast horizons. The model specification is presented in equation 2. The lag relationships were used to group the forecasts corresponding to the same poll release. A negative systematic bias persists in the years 2016–2019.

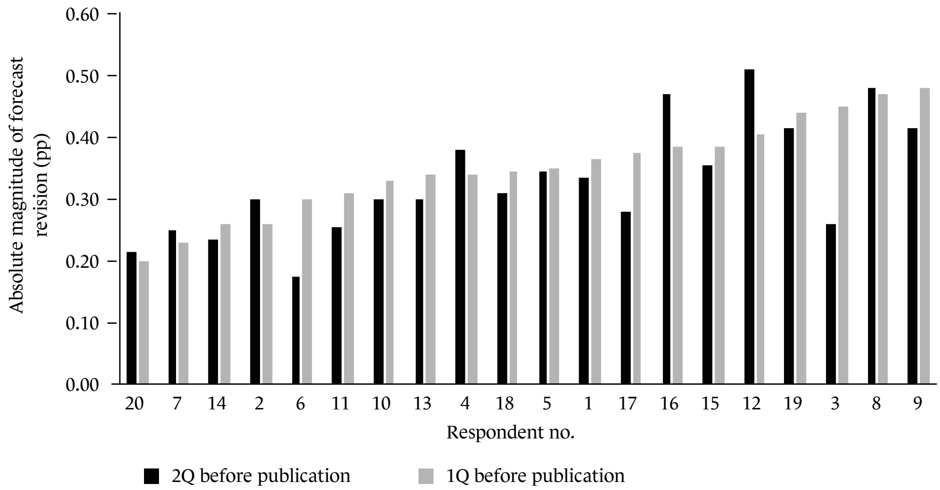
Figure 2

Test 2 – visualization of period effects



Notes:
This figure presents the estimated values of θ_t for different forecast horizons. The model specification is presented in equation 4. The lag relationships were used to group the forecasts corresponding to the same poll release. This series is not an example of a white noise process – we see some persistence of one-sided revision (e.g. during 2017–2019).

Figure 3
Test 2 – magnitude of revisions



Notes:
This figure presents the estimated values of μ_i for different forecast horizons. The model specification is presented in equation 4. The estimates confirm that some forecasters tend to make overly strong revisions. For greater details, see also Table 3.

Prognozowanie PKB w Polsce: historia ciągłych niepowodzeń

Streszczenie

Praca zwiera analizę błędów prognoz wzrostu PKB wysyłanych do dziennika *Rzeczpospolita* przez instytucje komercyjne w ramach konkursu na najlepszego analityka makroekonomicznego w latach 2013–2019. W ramach badania wykonany został model panelowy efektów stałych, za pomocą którego zweryfikowano 3 hipotezy:

1. Prognozy nie zawierają systematycznych obciążeń.
2. Prognozy są efektywne według definicji Nordhausa, tj. przeszłe rewizje prognoz nie pozwalają wnioskować na temat kolejnych kierunków zmian.
3. Profesjonalni analitycy kształtują swoje szacunki niezależnie od tego, jaka jest aktualna mediana oczekiwań.

Jak wskazuje tytuł pracy, badanie pokazało brak statystycznej efektywności prognoz – każda z powyższych hipotez została odrzucona.

1. Analitycy komercyjni popełniali systematyczne i jednostronne błędy, które były szczególnie widoczne w latach 2016–2019. Prognozowane tempo wzrostu było niższe od późniejszych danych.
2. Analitycy nie byli w stanie poprawnie przewidzieć konsekwencji działań fiskalnych rządu oraz zmian strukturalnych związanych ze zmianą unijnej perspektywy budżetowej.
3. Ekonomiści mieli tendencję do nadmiernego rewidowania prognoz. W przypadku każdej większej zmiany można było oczekiwać, że kolejna rewizja odwróci część dokonanej wcześniej korekty.
4. Zidentyfikowane zostały przykłady działań strategicznych. Analitycy, których prognozy znacznie odchylały się od rynkowego konsensusu, mieli tendencję do rewidowania swoich szacunków w kierunku mediany (ang. *herding behaviour*). Co więcej, obserwujemy spadek niezgodności między prognostami wraz z oddalającym się horyzontem prognozy. Takie zjawisko sugeruje, że profesjonalni analitycy komercyjni w znacznym stopniu bazują na wynikach projekcji makroekonomicznej NBP – może być to efekt dysponowania mniejszymi możliwościami w zakresie budowy i utrzymania modeli do prognoz długookresowych.

W pracy rozważono możliwość korekty zaistniałego stanu. Autor wskazuje, że jedną z przyczyn problemów może być zdominowanie rynku prognoz przez instytucje finansowe – główni ekonomiści w tych organizacjach cechują się zbliżonymi życiorysami, środowiskiem pracy etc. Dlatego też autor rekomenduje większe zaangażowanie instytucji sektora publicznego w komentowanie bieżącej rzeczywistości gospodarczej.

Systematycznym problemem w Polsce jest także niskie zaangażowanie sektora akademickiego. W opinii autora NBP stosunkowo niedużymi nakładami finansowymi jest w stanie poprawić tę sytuację. Warto przyjrzeć się działaniom promocyjnym prowadzonym przez londyńskie firmy finansowe. Organizują one konkursy modeli prognostycznych w wybranych dziedzinach (np. różnica punktów w meczu NBA w NBA data challenge firmy G-research). Stworzenie podobnego przedsięwzięcia w Polsce może zachęcić np. zdolnych technicznie doktorantów do generowania własnych predykcji. Zapewniłoby to większy pluralizm w debacie ekonomicznej oraz wymusiło znacznie większą dyscyplinę przygotowywania prognoz przez profesjonalnych prognostów.

Słowa kluczowe: prognozy PKB