Central European Economic Journal

sciendo

ISSN: 2543-6821 (online)

Journal homepage: http://ceej.wne.uw.edu.pl

Jakub Rybacki

Macroeconomic forecasting in Poland: The role of forecasting competitions

To cite this article

Rybacki, J. (2020). Macroeconomic forecasting in Poland: The role of forecasting competitions . Central European Economic Journal, 7(54), 1-11.

DOI: 10.2478/ceej-2020-0001

To link to this article: https://doi.org/10.2478/ceej-2020-0001

Central European
Economic Journal

Jakub Rybacki

SGH Warsaw School of Economics, Collegium of Economic Analysis, Warsaw, Poland,
corresponding author: jr70663@sgh.waw.pl

Macroeconomic forecasting in Poland: The role of forecasting competitions

Abstract

Macroeconomic forecasters are often believed to idealistically work on improving the accuracy of their estimates based on for example the Root Mean Squared Error (RMSE). Unfortunately, reality is far more complex. Forecasters are not awarded equally for each of their estimates. They have their targets of acquiring publicity or to earn prestige. This article aims to study the results of Parkiet's competitions of macroeconomic forecasting during 2015–2019. Based on a logit model, we analyse whether more accurate forecasting of some selected macroeconomic variables (e.g. inflation) increases the chances of winning the competition by a greater degree comparing to the others. Our research shows that among macroeconomic variables three groups have a significant impact on the final score: inflation (CPI and core inflation), the labour market (employment in the enterprise sector and unemployment rate) and financial market indicators (EUR/PLN and 10-year government bond yields). Each group is characterised by a low disagreement between forecasters. In the case of inflation, we found evidence that some forecasters put a greater effort to score the top place. There is no evidence that forecasters are trying to somehow exploit the contest.

Keywords

forecasting | strategic behaviour | incentives | Parkiet

JEL Codes

E17, E37

1 Introduction

Macroeconomic forecasters are often believed to idealistically work on improving the accuracy of their estimates based on for example the Root Mean Squared Error (RMSE). Unfortunately, reality is far more complex. First of all, forecasters are not awarded equally for each of their estimates. They have their targets of acquiring publicity or to earn prestige. Therefore, they may put greater attention to improve only some predictions, while sacrificing accuracy in the case of others. To illustrate the problem, Hann, Ogneva, and Sapriza (2012) compared the prediction of aggregate earnings, which was made by equity analysts and macroeconomists. While the first group makes frequent revisions based on companies' news, the latter is much more sluggish in incorporating information as macroeconomics models rely mainly on GDP data, which is released quarterly.

Furthermore, preferences of public opinion regarding optimal forecasting and preferences of forecasters may significantly differ. For example, professional forecasters tend to make more frequent revisions during a period of recession comparing to normal times (Loungani, Stekler, & Tamirisa, 2013), affectively react to economic surprises (Eroglu & Croxton 2010). Multiple articles are trying to solve whether analysts are fully rational or if they are using the most up-to-date information. But research on forecasters' incentives is relatively scarce.

This article aims to study the results of Parkiet's competitions of macroeconomic forecasting during 2015–2019. The announcement of winners has a strong influence on economic debate in the Polish media—winners give many interviews about their economic views, advertising expertise of their institutions. Furthermore, the influence of such media comments or research reports on financial instruments' valuations and actions of the public institutions is frequently greater compared to academic research papers. Therefore, such competitions affect the work of commercial economic research time, for example, set priorities in forecasting.

Studies regarding analysts' performance and competition between them are not publicly available. From this perspective, Poland is an interesting example—the majority of commercial research is published to a wide audience without additional charges.

Based on a logit model, we analyse whether more accurate forecasting of some selected macroeconomic variables (e.g. inflation) increases the chance of winning the competition by a greater degree comparing to the others. Second, we verify if there exists some persistence of scoring one of the best five places among the forecasters in the respective categories.

Our research shows that among macroeconomic variables, three groups have a significant impact on the final score: inflation (CPI and core inflation), labour market (employment in the enterprise sector and unemployment rate) and financial market indicators (EUR/PLN and 10-year government bond yields). Each group is characterised by a low disagreement between forecasters. Logit models suggest that top forecasters are capable to persistently be in the lead in case of the inflation block. The competition gives little benefit for correctly forecasting activity and leading sentiment surveys.

2. Literature review

This section presents a review of the literature regarding the motivations of professional economists to regularly present their macroeconomic forecasts to the public. The frequent illusion is that financial analysts or commercial economists tend to optimise their forecast based on some widely renowned metrics, for example, RMSE. Subject literature shows that this is not the truth.

First of all, forecasts are not produced in a vacuum. There is strong evidence that forecasters are influenced by other decisions (e.g. Scharfstein & Stein 1990, Pons-Novell 2004, Ottaviani & Sørensen 2006), which results in two opposite phenomena. Individual analysts may strategically approach their jobs. One solution is to self-censor his or her forecasts and closely follow the market consensus to avoid major mistakes. This phenomenon is known as herding behaviour. The opposite strategy suggests exaggerating the model finding to crowd-out from the others and attract attention. Both strategies have acquired the mathematical and game theory foundations (e.g. Elliott, Komunjer, & Timmermann, 2008; Marinovic, Ottaviani, & Sorensen, 2013; Pierdzioch, Rülke, & Stadtmann, 2012).

A second strain of literature reports that forecasters are prone to behavioural biases—we will focus on this directly related to motivation and credibility. Lamont (2002) showed that older and more recognised forecasters tend to more frequently stand out of the crowd and produce non-consensus ideas (not exactly more accurate). Ashiya (2009) reported the existence of affiliation biases—there is a tendency that forecasters working in the same sector, for example, commercial banks are likely to generate similar mistakes. The author suggests that this problem may be related to the wishes of the forecasters' employers.

A common feature of the analysts is publishing their estimates in the competition, for example, from Consensus economics, Focus economics, Bloomberg or Reuters ranking. The subject literature is relatively silent regarding the impact of the rules of such competition on economic forecasting. The results of such contests frequently have a significant impact on the assessment of commercial economists' work and their financial payoff. Furthermore, the mechanisms evaluating forecasts are prepared by non-professionals and may strongly deviate from academic state-of-the-art practices. To fill this gap, we evaluate the impact of the most prestigious competition in Poland from Parkiet daily.

3 Parkiet daily's competition

This section aims to present the Parkiet competition for macroeconomic forecasting and review the statistical properties of the formula. The competition is the most prestigious contest in Poland—it gathers the greatest number of forecasters participating in the panel, and results are published in the media. The influence of the Parkiet's ranking is visible, for example, in the analysts' interactions on Twitter. According to the author's knowledge, some economists in the research teams have directly linked the financial bonuses to performance in the competition.

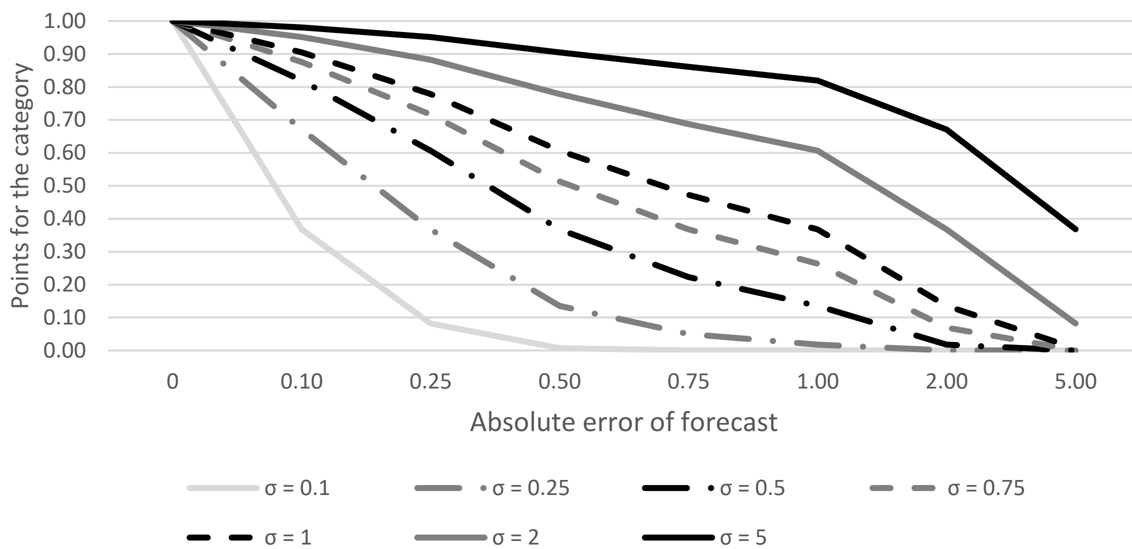

Fig. 1. Final score in Parkiet's competition depending on a standard deviation of forecast's disagreement (σ) and absolute forecast's error. *Source*: The author.

The competition requires monthly submission of Nowcasts for 14 macroeconomic variables and 5 financial indicators. Macroeconomic variables consist of: PMI index, CA balance, exports and imports value, CPI, core inflation, employment and wage growth in the enterprise section, PPI, growth of industrial production, construction output, retail sales and M3, and unemployment rate. Financial variables consist of two exchange rates (EUR/PLN and USD/PLN), two interest rates (on 2-year and 10-year government bonds) and NBP policy rate. Once per quarter there is also a submission of three additional indicators: growth of gross domestic product, private consumption expenditures and gross fixed capital formation.

Participants represent mainly the financial sector. In the most recent edition, there is also a public sector institution (Polish Institute of Economics) and an independent think-tank (prognozy-gospodarki.pl). The panel composition is unbalanced due to mergers in the banking sector or new entries. The dataset spans over the years 2015–2019.

Forecasts sent by analysts are evaluated based on a function described in Eq. (1).

$$\text{points}_{cat,t,i} = \mathrm{e}^{-\frac{|\varepsilon_{cat,t,i}|}{\sigma_{cat,t}}} \tag{1}$$

where: $\varepsilon_{cat,t,i}$ denotes forecast error of i-th analysts for the economic variable cat in the period t. $\sigma_{cat,t,t}$ describes the standard deviation between forecasts submitted by all competition's participants for the variable cat. $\text{points}_{cat,t,i}$ is the final amount of scores earned by i-th analyst. The final value lies between zero and one.

The final score is calculated as a simple sum of points in all categories, without weighting. The calculation is presented in Eq. (2). n denotes the number of variables in the competition.

$$\begin{aligned} \text{score}_{t,i} &= \sum_{cat}^{n} \text{points}_{cat,t,i} \\ \text{position}_{t,i} &= \text{rank}(\text{score}_{t,i}) \end{aligned} \tag{2}$$

From a perspective of the competition's participants, an important feature of this function is strong penalisation of relatively small forecast errors, when disagreement is low. This property is important especially in the case of categories with low forecast disagreement (e.g. employment, unemployment rate) where differences related to the rounding of numbers can halve the final score. The distribution of points depending on disagreement and the forecast error is presented in Figure 1.

Second, the function is strongly penalising big forecast errors in a situation when final realisation missed every individual estimate. The exact amount of points related to standardised values of error is presented in Figure 2.

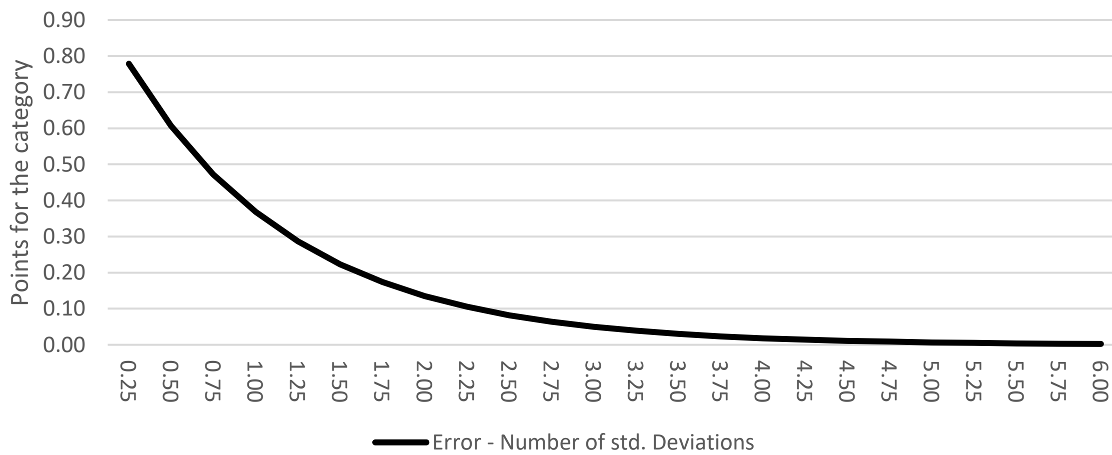

Fig. 2. Final score depending on standardised error. *Source*: The author.

In case of a big surprise, when the error exceeds 2.5 standard deviations of disagreement between forecasts, the number of earned points is not greater than 0.1 in any case. Therefore, even if there is a statistically significant difference in the accuracy (e.g. best forecaster has an error of $2.5^*\sigma_t$ and market consensus equal $5^*\sigma_t$), the difference between his/her reward and other participant scores is not meaningful. Therefore, this contest may discourage greater scrutiny of more volatile categories or during the period of greater uncertainty.

The mentioned problems have a strong impact on the final scores. First of all, there are strong discrepancies in the expected number of scores between the macroeconomic variables. The median score in the unemployment rate is frequently three times higher than the PMI index. The numbers are presented in Table 1.

Second, the dispersion between scores of top performers and median forecasters also differs strongly. Standard deviations for categories are presented in Table 2. The differences between scores for PMI or Industrial production can be relatively small. Furthermore, they can vary over time. In case of variables where strong surprises are more often, no one is capable to ex-ante predicting the total amount of scores but can expect the discrepancies between forecasters should be small (see Table 2).

Those facts suggest that the competition gives greater motivation towards accurate forecasting of categories with low disagreement. We attempt to verify whether successful projections of these variables result in a greater chance of winning the contest.

Tab. 1. Median score in selected categories of Parkiet's competition

Year	PMI index	Industrial production	CPI	Unemployment rate
2019	3.65	4.71	4.82	7.76
2018	3.13	5.65	4.95	7.15
2017	3.79	5.54	4.45	5.85
2016	2.90	4.72	3.79	7.34
2015	3.10	4.91	3.16	6.28

Source: The author. Full table is presented in Appendix.

Tab. 2. Standard deviation of scores in selected categories of Parkiet's competition

Year	PMI index	Industrial production	CPI	Unemployment rate
2019	1.44	1.68	1.70	3.02
2018	0.73	1.18	1.40	1.52
2017	0.80	1.19	1.19	1.14
2016	0.89	1.22	1.18	1.99

Source: The author. Full table is presented in Appendix.

4 Methodology

This section presents the methodology of our research. First of all, we will present a logit model that aims to answer whether successful forecasting in some categories has greater importance for final success in the competition than the others. Success is defined as a place in the top 5—only these participants are announced in the newspaper. Second, we will look at whether there exists evidence that winners in the previous years in some categories are more likely to triumph also in the next edition. If so, such teams are more likely to utilise greater resources compared to others either by developing more complex models, by analysing a greater amount of information or by purchasing non-public data.

Let's define a binary variable that describes whether a competitor achieved a top five performance score. The representation is described in Eq. (3).

$$\text{success}_{i,t} = \begin{cases} 1 & \text{position}_{t,i} \leq 5 \\ 0 & \text{position}_{t,i} > 5 \end{cases} \tag{3}$$

where index i identifies forecaster and index t denotes a time period.

The probability of finishing the contest among five top performers is described by a logit regression. Such model classification is described by an unobservable latent variable $y_{i,t}$. This latent variable takes a positive value when participants are among top performers, and negative or equal to zero otherwise (see Eq. 4).

$$\text{success}_{i,t} = \begin{cases} 1 & y_{i,t} > 0 \\ 0 & y_{i,t} \leq 0 \end{cases} \tag{4}$$

To describe the value of those latent variables, we use information about the top five performers in specific macroeconomic or financial variables. The formula is presented in Eq. (5).

$$y_{i,t} = \beta_0 + \beta_1 * \text{success}_{cat1,i,t} + \cdots + \beta_n * \text{success}_{catn,i,t} + \varepsilon_{i,t} \tag{5}$$

where $cat1$ denotes the first macroeconomic variable (category) i.e. PMI index, $cat2$ the 2nd one—CPI, and so on. $Catn$ describes the last n-th category—Polish government bond 10-year yield. The following enumeration was used only in methodological sections to shorten formulas—each table presenting the outcome of the model has a full description of variable instead.

Tab. 3. Which variables increase the odds of scoring top five in Parkiet's competition

Model parameters – Eq. (5)				
Variable	**Coefficient**	**Std. Error**	**z-Statistic**	**Prob.**
CPI	0.94	0.44	2.15	0.03
Core inflation	1.25	0.45	2.75	0.01
Unemployment rate	1.18	0.45	2.63	0.01
Employment	1.53	0.48	3.21	0.00
CA	1.67	0.47	3.58	0.00
EUR/PLN	0.91	0.53	1.71	0.09
POLGB10	1.46	0.41	3.58	0.00
Constant	−3.18	0.54	−5.87	0.00
Model diagnostics				
McFadden R-squared	0.51	Obs with Dep = 0	96	
LR statistic	62.57	Obs with Dep = 1	25	
Prob (LR statistic)	0.00	Total obs	121	

Source: The author. Model diagnostics are presented in the Section A2 in Appendix.

We hypothesize that some of the β parameters have positive and statistically significant values. In such circumstances, some variables have a greater influence on the final result than others.

Second, we will analyse if forecasters are exploiting these opportunities. We aim to verify if some analysts are capable to persistently achieve the top five places in case of variables, which brings a higher score. We propose a simple logit model where the binary variable will be positive in case when forecasters finished the contest with positions 1–5 in the selected category and zero otherwise. The latent variables will be explained by last year's performance.

$$y_{cat_j,i,t} = \beta_0 + \beta_1 * \text{success}_{cat_j,i,t-1} + \varepsilon_{i,t} \tag{6}$$

Positive values of β_1 suggest that the best forecaster can maintain pole positions and earn a higher score (on average). This implies that they dedicate greater effort to score in the category.

5 Model outcomes

This section aims to present the results of the models presented in the previous chapter. The model describing the probability of victory in the final competition is presented in Table 3. Statistically non-significant variables were removed.

First of all, our models suggest that good performance in the current account (CA) category to the greatest extent increases odds of the final score, but the corresponding components (exports/imports) do not guarantee any advantage. This phenomenon is puzzling—differences between analysts' scores in this category typically are not strong. One of explanations is that analysts who have a greater motivation to win the competition makes an effort to correctly predict the annual flow of the EU funds typically occurring in January (reading is published in March). The information on what amount of money was located on a CA is provided by the ministry of finance, but less-motivated analysts may not use this report.

The problems described in Section 3 are visible in case of the employment component—this is the second category where a good score increases the odds of winning the overall competition in a greater manner. In this case, the score in January is dependent on GUS statistical procedure. The statistical office annually rebalances its panel of enterprises, which creates unpredictable distortions. Furthermore, in other months, small errors may result in significant losses of scores. This problem applies also to forecasting the unemployment rate where a good score also increases the odds of victory.

The labour market variables are rather not the most forward-looking predictors of the business cycle. At the same time, there is no incentive to improve accuracy in case of monthly activity indicators (e.g. construction output, industrial production) or sentiment survey.

Finally, there is consistent evidence that focuses on inflation and financial markets forecasting pays off. The importance of inflation forecasting is obvious—these figures influence central banks' decisions regarding interest rate policy. But there are several obstacles that make financial variables at least controversial. First of all, there is a rich literature confirming that short-term foreign exchange movements are an example of a random walk process (e.g. Kilian & Taylor 2003, Bacchetta & Van Wincoop 2007). Second, in case of stronger volatility on FX/FI market and bigger shifts in valuations, the function is unlikely to grant scores to analysts.

Tab. 4. Are some analysts persistently better than others in forecasting? Models' estimations output of Eq. (6)

Variable	β_1	β_0	Prob (x1 = 0)	Prob (x1 = 1)
CPI	0.70 0.37 (0.06)	−1.02 0.18 (0.00)	15%	38%
Core inflation	0.83 0.38 (0.03)	−1.09 0.18 (0.00)	17%	38%
CA	0.66 0.35 (0.06)	−0.94 0.18 (0.00)	17%	39%
Export	0.23 0.38 (0.54)	−1.00 0.18 (0.00)		
Import	0.64 0.36 (0.08)	−0.96 0.18 (0.00)		
Construction output	0.19 0.36 (0.59)	−0.78 0.17 (0.00)		
Production	0.07 0.41 (0.86)	−0.91 0.17 (0.00)		
Retail sales	0.07 0.37 (0.86)	−0.83 0.17 (0.00)		
Employment	−0.74 0.41 (0.07)	−0.59 0.17 (0.00)		
Unemployment rate	0.02 0.37 (0.96)	−0.78 0.17 (0.00)		
M3	0.65 0.35 (0.06)	−0.99 0.18 (0.00)	16%	37%
PMI	0.35 0.36 (0.33)	−0.89 0.17 (0.00)		
EUR/PLN	−0.44 0.43 (0.3)	−0.74 0.17 (0.00)		
USD/PLN	0.02 0.37 (0.95)	−0.74 0.17 (0.00)		
POLGBs 10Y	−0.13 0.36 (0.72)	−0.71 0.17 (0.00)		

Source: The author.

After identifying the shortcomings and advantages in the scoring of Parkiet's competition, we analysed whether research teams are utilising these flaws to perform better (Table 4).

Three values in columns β_1 and β_0 represent respectively parameter estimate, its standard deviation and *p*-value for Z-statistics. Positive values of β_1 denote that the forecaster who achieved a top five position in the previous edition of the contest is more likely to repeat this achievement next year, than others.

We found some evidence that top forecasters are capable to more persistently achieve a top five position in the categories of CA and inflation. On the other hand, we do not see a bigger persistence in case of employment and the unemployment rate. These categories usually have the biggest number of scores and limited relevance—the number of instant comments after these releases is much lower compared to CPI or industrial production. Therefore, there is no evidence that participants are trying to somehow exploit the contest by achieving greater scores only in this group.

6 Conclusions

In this article, we have also analysed the statistical properties of the function that evaluates forecaster estimates in Parkiet's competition. Our analysis showed that the scoring mechanism is not rewarding analysts, when surprises are exceeding two standard deviations of the disagreement. Therefore, this function discourages analysts to make a greater effort in case of periods of elevated uncertainty. Ultimately, it is not supporting more accurate forecasting when the potential error is greater during a depression or structural changes. At the same time, accurate forecasting is usually most valuable during such periods. Therefore, instead of promoting accurate forecasting generally, this function promotes forecasting in stable conditions only.

Our model highlighted that Parkiet's forecasting competition generates incentives to focus on predictions of inflation and labour market data rather than economic activity. Each of the mentioned categories is characterised by low forecast disagreement and usually small forecast errors. The motivation to accurately forecast these variables may result in utilising a greater amount of time for forecast where potential errors are not that important (e.g. employment and unemployment rates). But, accurate nowcasting in these two categories does not lead to greater scrutiny of long-term forecasts—commercial economists rarely publish their estimates of labour market for the horizons greater than 1 month.

Finally, we analysed whether some persistence in victories of top-performing analysts occurs. Our models showed that last year's winners are statistically more likely to triumph in the next year. Still the advantage is not that great—proposed models do not generate a greater gain in predictability comparing to a constant probability model. Therefore, we found no evidence that participants are trying to somehow exploit the contest by achieving greater scores only in most scoring groups.

References

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

Bacchetta, P., & Van Wincoop, E. (2007). Random walk expectations and the forward discount puzzle. *American Economic Review, 97*(2), 346–350.

Elliott, G., Komunjer, I., & Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association, 6*(1), 122–157.

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting, 26*(1), 116–133.

Hann, R. N., Ogneva, M., & Sapriza, H. (2012). *Forecasting the macroeconomy: Analysts versus economists* 2012. Available at SSRN.

Hess, D., & Orbe, S. (2013). Irrationality or efficiency of macroeconomic survey forecasts? Implications from the Anchoring Bias Test. *Review of Finance, 17*(6), 2097–2131.

Kilian, L., & Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics, 60*(1), 85–107.

Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior & Organization, 48*(3), 265–280.

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

Marinovic, I., Ottaviani, M., & Sorensen, P. (2013). Forecasters' objectives and strategies. *Handbook of Economic Forecasting, 2*, 690–720.

Ottaviani, M., & Sørensen, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics, 81*(2), 441–466.

Pierdzioch, C., Rülke, J. C., & Stadtmann, G. (2012). A note on forecasting emerging market exchange rates: Evidence of anti-herding. *Review of International Economics, 20*(5), 974–984.

Pons-Novell, J. (2004). Behavioural biases among interest rate forecasters? *Applied Economics Letters, 11*(5), 319–321.

Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *The American Economic Review, 80*, 465–479.

Appendix

A1 Median scores in the Parkiet ranking

Standard deviations of final scores in competition categories						Median scores in the competition categories					
Year	**2019**	**2018**	**2017**	**2016**	**2015**	**Year**	**2019**	**2018**	**2017**	**2016**	**2015**
PMI	1.44	0.73	0.80	0.89	0.80	**PMI**	3.65	3.13	3.79	2.90	3.10
CPI	1.70	1.40	1.19	1.18	1.11	**CPI**	4.82	4.95	4.45	3.79	3.16
CA balance	1.60	0.85	1.23	1.42	0.95	**CA balance**	3.66	4.91	4.74	5.28	4.49
Export	2.18	1.55	0.97	0.85	0.84	**Export**	5.08	5.78	4.59	4.35	3.18
Import	1.76	1.35	1.50	1.55	1.57	**Import**	4.50	5.24	6.34	5.35	5.74
Core inflation	1.52	1.39	0.91	1.40	1.13	**Core inflation**	3.50	4.91	5.18	5.37	4.33
Employment	1.55	1.68	1.26	1.36	1.11	**Employment**	4.89	6.56	5.19	5.02	4.38
Wage growth	1.28	1.04	0.97	0.94	1.35	**Wage growth**	4.05	4.73	5.01	4.03	5.16
PPI	1.89	1.36	1.07	1.18	1.33	**PPI**	4.63	5.61	3.90	4.94	4.40
Industrial Output	1.68	1.18	1.19	1.22	1.68	**Industrial Output**	4.71	5.65	5.54	4.72	4.91
Construction Output	1.39	1.13	0.84	1.43	1.38	**Construction Output**	4.21	5.25	4.19	3.94	4.79
Retail Sales	1.60	0.91	0.54	0.95	0.97	**Retail Sales**	4.58	4.58	3.63	3.01	4.14
M3	1.49	1.40	0.91	1.01	0.98	**M3**	4.50	5.65	5.26	4.16	4.42
Unemployment rate	3.02	1.52	1.14	1.99	1.79	**Unemployment rate**	7.76	7.15	5.85	7.34	6.28

A2 Diagnostic tests

The aim of this chapter is to present diagnostic tests for Eq. (5), which was presented in Table 3. The distribution of dependent variable $success_{i,t}$, which describe if analysts reached a top five position in the contest is presented in the table below. Approximately 20% of occurrences were concluded with success, while nearly 80% describes failures.

Variable's value	Count	Percent
$success_{i,t} = 0$	96	79.34%
$success_{i,t} = 1$	25	20.66%

Firstly, we present Expectation-Prediction evaluation, which compares performance of the estimated model with Constant Probability Model (CPM). Specification with single constant parameter should classify each occurrence of $success_{i,t}$ at zero, as this value is dominant in the sample. For our model, we set a cut-off level of 0.8 i.e. similar to frequency of failures in the sample. Each time the model will state that probability of success is greater than 80% it will classify the occurrence as a success. Otherwise failure will be recorded. The output of this analysis is presented below:

Expectation-Prediction Evaluation – Model response vs. observed variables

	Estimated equation			Constant probability model		
	$success_{i,t} = 0$	$success_{i,t} = 1$	Total	$success_{i,t} = 0$	$success_{i,t} = 1$	Total
Correct	95	12	107	96	0	96
%	99	48	88	100	0	79

Estimated model does not sacrifice much accuracy of predicting failures—only one observation was classified incorrectly (column 2). At the same time the model is correctly predicting 48% of successful occurrences (column 3). Therefore, the share of correctly classified observations is greater by 9.09pp comparing to CPM.

Even greater accuracy is visible, while comparing expected number of success/failure observations in the split sample. The table with results is presented below:

Expected number of observations in the sample based on the model specification

	Estimated equation			Constant probability model		
	$success_{i,t}$ = 0	$success_{i,t}$ = 1	Total	$success_{i,t}$ = 0	$success_{i,t}$ = 1	Total
Total	96	25	121	96	25	121
Correct	87	16	102	76	5	81
%	90	62	84	79	21	67

The CPM indicates that the percentage share of successes or failures in every subsample should be equal to the one derived from the original sample. For example, if we select only observations when $success_{i,t}$ =0, CPM would still indicate that share of failures is equal to 79.34%. Given 96 observation in the subsample expected number of occurrences is equal to 76 (column 5).

Our model presented a more accurate number of successes/failures in both subsamples. Based on the observable characteristics model stated that in subsample of 96 $success_{i,t}$=0 occurrences, 87 observation are expected to be a failure vs. 79 in CPM. Similarly, for subset of observations where $success_{i,t}$ =1, expected number of successes is 16 of 25 (vs. 5 in CPM).

Finally, we performed formal Andrews and Hosmer-Lemeshow Tests. Both procedures evaluate Goodness-of-Fit of the model with a null hypothesis that specification is correct. Both tests assess whether occurrences of success in observed data match expected successes implied by the model in subsets divided by selected number of percentiles. The output with the results of the two tests is presented below:

Goodness-of-Fit test statistics

	$success_{i,t}$ = 0		$success_{i,t}$ = 1		
Percentile	Actual	Expected	Actual	Expected	H-L Value
1	12	11.99	0	0.01	0.01
2	12	11.99	0	0.01	0.01
3	12	11.90	0	0.10	0.10
4	12	11.80	0	0.20	0.20
5	11	11.49	1	0.51	0.50
6	11	11.29	1	0.71	0.12
8	8	8.24	4	3.76	0.02
9	6	6.05	6	5.95	0.00
10	1	1.27	12	11.73	0.06
Total	96	96.18	25	24.82	1.47

H-L Statistic	1.47	Prob. Chi-Sq (8)	0.99
Andrews Statistic	55.82	Prob. Chi-Sq (10)	0.00

The listing shows that expected number of occurrences stays close to the actual number. Therefore, the Hosmer-Lemeshow test is not rejecting the null and states specification is correct. On the other hand, the contradictory signal was present by the 2nd Andrew test. Based on available data, we were incapable to create specification, which could successfully pass two tests.