Incentivizing Accuracy: The Role of Public Rankings in Economic Forecasting

Marcos Ross Fernandes
FEA-USP*

Vinícius Sabbag
FGV-EESP†

February 2025

Abstract

Given the broad evidence of systematic forecast errors, we can say that forecasters respond to various incentives, making it entirely plausible that they are not solely concerned with the accuracy of their projections. With this in mind, we study how incentives created by the Central Bank of Brazil (BCB) — in the form of a competition among forecasters projecting GDP growth — drive them to be more accurate. By constructing a counterfactual scenario, we find that projection errors would have been significantly smaller during periods of crisis and instability: from 2013 to 2016 (Dilma's administration crisis) and from 2020 to 2021 (the pandemic crisis), if the competition had always existed. Considering that the reduction in projection errors could have occurred both due to the "copying" of the best forecasters' projections and the "incentive" created by the competition, which drives agents to seek publicity gains, we present a methodology that separates these two channels. Following this methodology and analyzing the counterfactual scenario in our sample, we observe that the "copying" channel is only active when the crisis results from an abrupt and unexpected shock (during the pandemic). However, when the crisis is structural (between 2013 and 2016), only the "incentive" channel appears to be active.
Keywords: *expectations, incentives, biases, forecasters, publicity*
JEL Classification: *D84, D90, E58, E70.*

1 Introduction

A forecaster is as agent who produces an estimate for a random variable, which will only be observed in the future. Initially, it is assumed that this agent faces the following problem:

$$\max_{p^e_{t+1}} \mathbb{E}_t U(p_{t+1}, p^e_{t+1}) = \mathbb{E}_t[-(p_{t+1} - p^e_{t+1})^2] \tag{1}$$

and must solve it in period t, where p_{t+1} is the random variable the agent aims to forecast, and p^e_{t+1} is their estimate. Muth (1961) states that a "rational" agent incorporates all available information at the time, so their forecast is as accurate as possible: ${p^e_{t+1}}^* = \mathbb{E}_t p_{t+1}$. Thus, they maximize their expected utility function, with $\mathbb{E}_t U(p_{t+1}, p^e_{t+1}) = 0$.

According to Muth, it is not plausible that the forecasts of these agents, in general, are worse than the forecast based on the "economic theory", considered the best one. If this was the case, there would be room for an opportunistic forecaster to achieve an extraordinary payoff by leveraging the theory to produce more accurate predictions than their peers.

*Email: marcosross@usp.br
†Email: vi.leosab@gmail.com

In equilibrium, therefore, all agents would have very similar estimates. Following this reasoning, an agent who systematically errs should naturally be excluded from the market for being of "lower quality". Thus, "Muth's hypothesis of 'rational expectations' implies unbiased and efficient forecasts using all available information" (Mincer and Zarnowitz (1969), p. 10).

Using the concept of "rational expectations", the literature in this area has evolved with the aim of empirically evaluating the quality of estimates, defining what constitutes a "good forecast" and a "bad forecast". The distinction between "good" and "bad forecasts" would then be based on objective metrics (see Mincer and Zarnowitz (1969) and Theil, Beerens, Tilanus, and De Leeuw (1966)).

Interestingly, these initial ideas conflict with the broad evidence of systematic biases. For instance, Batchelor (2007) reports that market consensus systematically overestimated GDP growth in G7 economies. Nasser and De-Losso (2021) found similar results in Brazil, also observing optimistic estimates for GDP and CPI. Thus, possibly, not every forecaster aims solely to maximize accuracy, as previously theorized. The utility function $U(p_{t+1}, p^e_{t+1})$ in (1), therefore, might not be correctly specified.

Several works have proposed hypotheses to explain the existence of these biases. These hypotheses can be divided into two groups. The first includes non-rational bias hypotheses, linked to heuristics studied by behavioral economists. For example, Ito (1990) finds that Japanese exporters systematically expect a more depreciated exchange rate than importers, which can be explained by the behavioral framework, violating "rational expectations".

The second group encompasses rational bias hypotheses, which are explained by models where agents do not consider accuracy as the sole input in their utility functions but also account for other factors. A rational bias could arise, for instance, when an agent seeks to maximize public exposure by producing deliberately exaggerated forecasts, thereby attracting media attention (Laster, Bennett, and Geoum (1999), Ashiya (2009)). This is the "publicity hypothesis" which implies an "anti-herding bias". Meanwhile, Ehrbeck and Waldmann (1996) theorize that rational forecasters imitate patterns of well-regarded forecasters (with less noise and smaller errors) in a game where they attempt to convince clients that they are of high quality. This leads to behavior where adjustments are made at a suboptimal frequency, leading to an "accommodation bias".

Incorporating the "publicity hypothesis" and the "accommodation bias" the extended forecaster problem, solved at t, is:

$$\max_{p^e_{t+1}} \mathbb{E}_t U = \mathbb{E}_t[-((p_{t+1} - p^e_{t+1})^2 + (p^{e,\text{previous}}_{t+1} - p^e_{t+1})^2) + F((p^{e,\text{consensus}}_{t+1} - p^e_{t+1})^2)] \quad (2)$$

where p_{t+1} is the random variable the agent aims to predict, p^e_{t+1} is their estimate for the variable, $p^{e,\text{previous}}_{t+1}$ is their previous estimate for the variable in $t+1$ (released in $t-1$), $p^{e,\text{consensus}}_{t+1}$ is the market consensus estimate for the variable in $t+1$ (released in $t-1$), and $F(x)$ is a function that measures publicity gains, with $F'(x) > 0$. This utility function

incorporates a penalty for forecast errors as in (1), adds a penalty for adjustments relative to the previous forecast, and includes a reward for deviating from the consensus, considering the publicity obtained. In this case, the optimal response would be ${p^e_{t+1}}^* \neq \mathbb{E}_t p_{t+1}$ (see Appendix A.1).

Therefore, given the wide range of explanations for systematic biases in macroeconomic forecasts, this study aims to deepen the discussion about the trade-offs forecasters face when deciding whether to "increase or not the accuracy of projections". From this, we seek to answer the following question: how does the public exposure of a forecaster's "quality" influence them to reduce bias? This will be investigated through a natural experiment, using initiatives created by the Central Bank of Brazil (BCB) in October 2021 for this purpose. We could hypothesize that these initiatives were a way to incorporate the agent's accuracy into the function $F(x)$[1], bringing ${p^e_{t+1}}^*$ closer to $\mathbb{E}_t p_{t+1}$ (see Appendix A.2).

The paper is structured as follows: Section 2 describes the Brazilian Central Bank's survey system and reflects on the scoring rules literature, relating it to the discussed problem; Section 3 describes the data used in the empirical investigation; Section 4 outlines the methodology employed; Section 5 presents the results, and Section 6 concludes.

2 Institutional Framework

2.1 The Inflation Targeting Regime and the Survey System (SEM)

Between 1998 and 1999, Brazil faced a turbulent period marked by a fragile fiscal policy and a loss of confidence in its fixed exchange rate regime, which ultimately led to the regime's flexibilization. The monetary authority was concerned about the unanchoring of inflation expectations, which threatened the accomplishments of the Real Plan, implemented in 1994 to stabilize the economy. To address this, the inflation targeting system was adopted in 1999 (Fraga (2009)).

For the system to work effectively, institutional improvements were required. These included the quick publication of the Monetary Policy Committee (COPOM) minutes, the quarterly release of the Inflation Report, greater investment in the BCB's technical staff enabling frontier research, and the implementation of a reliable system to aggregate forecasts from various market agents, based on surveys (F. A. Carvalho and Minella (2012)).

This system was created in May 1999 with 50 participating institutions[2], which shared their expectations of a few price indices and the GDP through various channels (fax, telephone, or email). Over time, the system improved. Expectations for more variables began to be collected across different horizons, and these forecasts started being shared through a dedicated website: the Market Expectations System (SEM), launched in November 2001. Additionally, the system gained popularity among forecasters, with participation increasing over the years. By 2012, there were around 100 participants (F. A. Carvalho and Minella

[1] Which changed from $F((p^{e,\text{consensus}}_{t+1} - p^e_{t+1})^2)$ to something like $F((p^{e,\text{consensus}}_{t+1} - p^e_{t+1})^2 - (p_{t+1} - p^e_{t+1})^2)$.
[2] Including banks, asset management firms, consulting companies, trade associations, among others.

(2012)). Currently, there are 171.

Today, SEM is the richest source of data on market agents' expectations in the country, and is internationally recognized for its quality, earning second place in the World Bank Regional Statistical Innovation Award in 2007, competing against 170 initiatives from other countries in the region (Marques (2012)). For each business day, the following statistics are available: (i) the median and (ii) the mean of the projections informed over the last 30 days by the institutions, (iii) the standard deviation of these projections, (iv) the maximum and (v) minimum, and (vi) the number of respondents. These same statistics are available for the best forecasters, selected based on the *Top 5 Ranking* (detailed below).

Currently, 27 variables are considered, divided into five groups: seven are price indices, two are rates (exchange rate and base interest rate, called Selic), four are external sector variables, ten are measures of activity, and four are related to fiscal conditions. Among the variables with the oldest data collection (since November 2001) are GDP, exchange rate, IPCA (the Brazilian CPI), and IGP-M (a Brazilian index which follows the prices to manufacturers), along with the Selic rate (since 2005).

2.2 *Scoring Rules*

The Central Bank's survey system employs a specific scoring rule whose primary purpose is to establish a criterion that ranks forecasters of a given variable based on their accuracy or, in other words, their "quality". The resultant ranking is called *Top 5*, and only the five best forecasters are disclosed.

There are many types of scoring rules in this system, depending on the periodicity of the variable[3] and the forecast horizon considered[4]. Consequently, there are also many different rankings. As an illustration, the so-called "annual long-term ranking – current year" for a variable p_t, conducted at the end of year t, is calculated as follows:

1. Projections of a forecaster j are collected twice a month in each of the 12 months of year t (24 times in total). At the end of year t, the absolute differences $e_{t,m}^{c,j}$ are calculated using the observed value and the predictions of j, where m is the month of year t ($m = 1, 2, ..., 12$) and c is the collection number within the month ($c = 1, 2$). The absolute difference is then:

$$e_{t,m}^{c,j} = |p_t - p_{t,m}^{c,j\ e}| \tag{3}$$

   where $p_{t,m}^{c,j\ e}$ is the estimate of forecaster j in collection c of month m of year t.

2. In each of the 24 collections, the agent with the smallest absolute difference receives a score of 10, and the agent with the largest absolute difference receives 0, interpolating the others to calculate their scores. The score of agent j in this specific collection, $n_{t,m}^{c,j}$, is therefore given by a decreasing function of its absolute difference: $n_{t,m}^{c,j} = f(e_{t,m}^{c,j})$.

[3] For instance, GDP is a quarterly variable, whereas IPCA is monthly.

[4] Forecast accuracy can be evaluated 6, 12, or 24 months before the variable is realized, for example.

3. Finally, the scores are weighted according to the month of the year. Under this criterion, a score in January is worth more than one in December, when the variable p_t is close to being known. The weight given to month m is $\beta_m = 13 - m$. Thus, January has a weight of 12 and December has a weight of 1.

4. The final score is:

$$r_t^j = \sum_{m=1}^{12} \beta_m (n_{t,m}^{1,j} + n_{t,m}^{2,j}) \tag{4}$$

As described in Section 1, the literature shows that agents might rationally be biased in favor of optimistic or pessimistic projections and disclose forecasts far from the expected value of the variable (which is equivalent to the agent's belief, according to the rational expectations hypothesis), acting strategically. Implementing a scoring rule is one way to mitigate this problem.

By definition, a "proper" scoring rule is a payoff function that is maximized when the agent reveals their true beliefs (Winkler and Murphy (1968), Gneiting and Raftery (2007)), thus reducing their incentives to lie[5]. Let's demonstrate whether the rule described above is a proper scoring rule (assuming there is only one collection per month and that p_t is a continuous random variable). The problem for j is:

$$\max_{p_{t,m}^{j,e}} \mathbb{E} r_t^j = \mathbb{E}[\beta_1 f(e_{t,1}^j) + \beta_2 f(e_{t,2}^j) + ... + \beta_{12} f(e_{t,12}^j)] \tag{5}$$

Therefore, the agent j has to minimize $e_{t,m}^j$ for each m, conditional on the information they have at m:

$$\min_{p_{t,m}^{j,e}} \mathbb{E}_m |p_t - p_{t,m}^{j,e}| \tag{6}$$

Since the absolute distance considers linear penalties for deviations from p_t, with no increasing penalties for larger deviations, the function is minimized when the agent chooses the median of the p_t distribution rather than its expected value $\mathbb{E}_m p_t$. Thus, ${p_{t,m}^{j,e}}^* = \text{Md}(p_t|\mathcal{I}_m)$, where $\mathcal{I}_m$ is the set of information available at m. Even if the agent observes the p_t distribution and predicts p_t to be equal to its expected value, they have incentives to share another forecast. Therefore, we conclude that the scoring rule is not proper for asymmetric distributions of p_t.

We should also note that further complications arise if agents are not risk-neutral, having nonlinear utility functions (A. Carvalho (2015)) when considering that forecasters make decisions under uncertainty. Interestingly, Hossain and Okui (2013) prove that another type of scoring rule (binarized scoring rule, BSR) effectively incentivizes agents to share their true beliefs, regardless of their preferences. In this scheme, a binary reward is given

[5]Formally, if the forecaster discloses the predictive distribution P and event x materializes, then their reward given by the scoring rule is $S(P, x)$. We define $S(P, Q)$ as the expected value of $S(P, \cdot)$ under distribution Q. Assuming the agent knows Q, we say that the scoring rule is strictly proper if $S(Q, Q) > S(P, Q)$ for any $P \neq Q$. (Gneiting and Raftery (2007))

to the agent. The agent receives A if their error is smaller than a random variable K[6] drawn by the principal, and B otherwise, with $A \succ B$. Under this arrangement, the forecaster's problem is to maximize the probability of obtaining the higher payoff. This probability is maximized if the agent reveals their true belief about the variable they are trying to predict and does not lie.

Note that the findings of Hossain and Okui (2013) are useful when analyzing the specific case of the BCB's *Top 5* ranking, as there is also an asymmetry in rewards. Similarly, the reward is binary, with forecasters obtaining "publicity returns" only if they are among the top five, and no such returns otherwise. Thus, even though the underlying scoring rule is not proper, *Top 5*, with its inherent asymmetry, might encourage forecasters to produce and disclose more accurate forecasts (see Appendix B)[7].

2.3 Competition and the natural experiment

In a *winner-takes-all* game, forecasters might, theoretically, avoid disclosing the expected value of the variable as their estimate in an attempt to "escape" the crowd of forecasters already disclosing that value. Yes, the probability of success decreases, but the probability of succeeding alone increases (Laster et al. (1999), Ottaviani and Sørensen (2006), Witkowski, Freeman, Vaughan, Pennock, and Krause (2023)). Following this reasoning, *winner-takes-all* competitions would distort incentives, contrary to the model of Hossain and Okui (2013).

Amid these conflicting ideas, we present the natural experiment to be analyzed in this work. On October 25, 2021, the BCB announced the creation of the "annual long-term ranking - current year" for GDP growth variable. Although GDP is one of the variables monitored since the beginning of the SEM, it did not, curiously, have this incentive for forecasters, unlike other variables tracked since the system's inception.

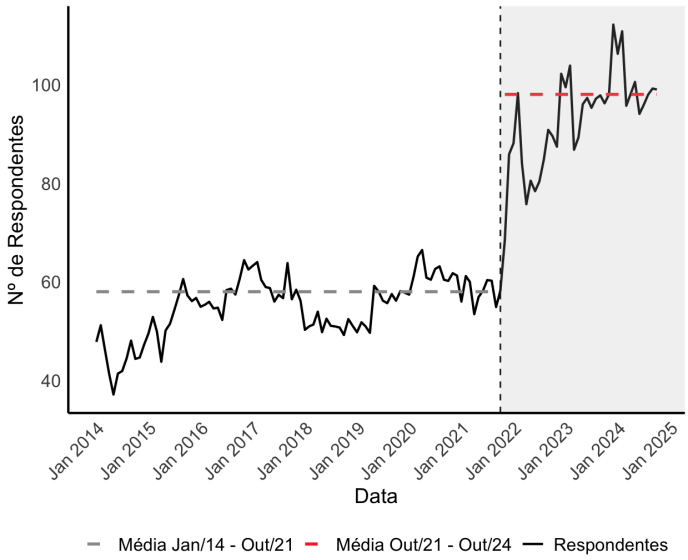
The SEM provides the number of forecasters sharing their projections since January 2014. Figure 1 shows that the number of respondents increased abruptly after the ranking's creation. Between January 2014 and October 2021, the number of participants was relatively stable, with an average of 58 respondents. From November 2021 onwards, the number of participants surged, reaching approximately 100 in January 2022. The average between October 2021 and October 2024 was approximately 98 respondents.

Few studies empirically analyze the impact of this competition among forecasters promoted by the BCB. F. A. Carvalho and Minella (2012) and Guillén (2008) observe that Brazilian forecasters react to the *Top 5*'s predictions, with experts being influential for other agents. Guillén (2008) even shows that this influence increases as the disclosure date of the predicted variable approaches. Agents in the Brazilian context would then trade the unlikely gains of being correct alone for a smaller expected forecast error, according to these

[6] $K \sim \text{Uniform}(0, \overline{K})$

[7] Nevertheless, we should recall that if other factors are still considered in the agent's utility function, such as in (2), even a proper scoring rule or a mechanism like the *Top 5* may not lead forecasters to maximum accuracy, even if it encourages them to improve it.

Figure 1: Evolution of the number of respondents for the GDP growth variable



studies, contrary to Ottaviani and Sørensen (2006).

That said, with this exogenous change caused by a BCB policy, we will analyze the relationship between the creation of the *Top 5* ranking and the GDP forecast errors through a natural experiment. Our objective is to understand whether implementing this mechanism, besides attracting more agents to the system, resulted in smaller forecast errors, leading forecasters to be more accurate, thus being an effective incentive mechanism. The incentive created by the ranking exists because of the reputational gains an agent obtains by appearing among the best, receiving positive publicity. But pecuniary returns might also be important, with the earnings of several agents being tied to their presence (or not) among the top five. Thus, with the present work, we empirically investigate the effect of this measure and whether it is significant.

3 Data

3.1 Projection Error Series

All time series used in this work were extracted from the SEM (BCB (2024c)). The first five series are the twelve-month-ahead projection errors for five macroeconomic variables: GDP (in p.p.), Selic rate (in p.p.), exchange rate (in R$/US$), IPCA (in p.p.), and IGP-M (in p.p.). That is, the projection error series e_t is equivalent to $y_t - {y_{t,t-12}}^e$, where y_t is the observed variable at t and ${y_{t,t-12}}^e$ is the forecast of the variable at t twelve months earlier (median of forecasts provided by agents to the BCB).

The series are monthly, so we are always working with rolling horizons. SEM directly provides the twelve-month-ahead expectations for IPCA and IGP-M in rolling horizons. It

is then sufficient to subtract this expectation from the observed value[8] to calculate e_t.

For the projection error series of the exchange rate and Selic rate, similarly, extensive changes are not necessary. SEM provides agents' expectations for the nominal exchange rate for each of the following twelve months[9] on a given date. We then calculate the average expected exchange rate over the next twelve months or the expected rate twelve months ahead (end-of-horizon). Thus, the average and end-of-horizon projection errors are obtained by subtracting this expectation from the observed value. For the Selic rate, the process is similar. We observe the Selic rate expectations for the next eight COPOM meetings, which corresponds to approximately 360 days, as the committee meets, on average, every 45 days. We can then calculate the average expected Selic rate for the next meetings or the rate set by the committee in the eighth meeting (end-of-horizon, one year ahead). The average and end-of-horizon projection errors for the Selic rate are obtained by subtracting this expectation from the observed value[10].

The GDP projection error series requires the most transformations to match the format of the other series. First, note that SEM provides GDP growth expectations by quarter compared to the same quarter of the previous year (*year-on-year, YoY*). The dataset extracted from SEM was supplemented with seasonally adjusted GDP growth rates[11] relative to the immediately preceding quarter (*quarter-on-quarter, QoQ*), enabling the calculation of the implicit GDP *QoQ* growth expectation.

Once *QoQ* rates are obtained, an interpolation is performed to calculate proxies for the monthly GDP growth rates (*month-on-month, MoM*) forecasts. In other words, we now aim to split the *QoQ* rate into three *MoM* rates. The next step is to estimate the weight that each of the three months has in the growth of each of the four quarters of the year.

The weights are estimated using the IBCbr time series, a monthly activity index compiled by the BCB, considered a good monthly proxy for GDP[12]. The series starts in January 2004 and ends in December 2023. The following problem is solved for a quarter Q, composed of months m_1, m_2, m_3:

$$\min_{\alpha,\beta,\gamma} \sum_{t=2004}^{2023} ((1+g_{Q,t})^{\alpha}-(1+g_{m_1,t}))^2+((1+g_{Q,t})^{\beta}-(1+g_{m_2,t}))^2+((1+g_{Q,t})^{\gamma}-(1+g_{m_3,t}))^2 \quad (7)$$

$$\text{s.t.} \quad \alpha+\beta+\gamma=1$$

where $g_{Q,t}$ is the growth rate of the IBCbr index in quarter Q of year t, $g_{m_i,t}$ is the growth rate of the index in the i-th month of quarter Q of year t, and α, β, γ correspond to the weights of the first, second, and third months, respectively. Note that the problem expressed in (7) is very similar to classical OLS estimation, as we are minimizing the sum of

[8] Retrieved from IBGE (2024a) and FGV-IBRE (2024).

[9] This horizon has increased in recent years, and we can currently know market expectations for each of the next 24 months.

[10] Retrieved from BCB (2024d) and BCB (2024a).

[11] Retrieved from IBGE (2024b).

[12] Retrieved from BCB (2024b).

three squared residuals for each year. Each residual corresponds to a month of the quarter. Also, the sum of the estimated parameters α, β, γ must be one, ensuring that the relative contribution of each month to quarterly growth totals 100%. This guarantees that the weighted interpolation between months m_1, m_2, m_3 is consistent with the total quarterly growth.

The estimated weights are shown in Table 1, with each row providing the results for one of the four quarters of the year. Higher weights indicate a greater contribution of the month to quarterly growth. If the weight is negative, the month's growth is generally negative, compensated by other months of the quarter.

Table 1: Estimated Weights for Each Month of Each Quarter

	α (m_1)	β (m_2)	γ (m_3)
Q_1	-0.422 (Jan)	0.120 (Feb)	1.302 (Mar)
Q_2	0.915 (Apr)	0.049 (May)	0.036 (Jun)
Q_3	1.218 (Jul)	-0.113 (Aug)	-0.105 (Sep)
Q_4	-0.090 (Oct)	0.601 (Nov)	0.489 (Dec)

With the expected *MoM* GDP growth rates, we can aggregate them into rolling twelve-month horizons. Using the IBCbr series, we can calculate the actual GDP growth observed in twelve-month rolling horizons. The difference between the two is then the GDP projection error.

Figures 2a to 2e visually present the series, spanning from March 2011 to August 2024. For all variables, the forecast error deviated significantly from zero during the 2014-2016 recession and the COVID-19 pandemic crisis, reflecting the high uncertainty in those periods.

First, it is observed in Figure 2a that the market generally projected lower consumer inflation than observed during the years 2014 to 2016. Due to this initial inflationary surprise, expectations deteriorated, and the market projected higher inflation than observed between 2017 and 2018, doubting the success of the monetary authority's disinflation plan. In 2021, with the supply shock in the global economy caused by the disruption of supply chains, we observe a large positive error, with the market predicting much lower inflation than what was observed between 2021 and 2022. A similar behavior is noted in Figure 2b, referring to the general price index (which incorporates producer and construction prices), with an even larger projection error between 2021 and 2022.

Figure 2c, which presents the forecast errors for the exchange rate, tells a similar story: positive errors for the exchange rate between 2015 and 2016, negative errors between 2017 and 2018, and positive errors again during the pandemic (between 2020 and 2021), reflecting the Dilma government crisis, the stabilization attempt during the Temer government, and the pandemic shocks, respectively. It is also noted that the behavior of the end-of-horizon projection error is more volatile than the average projection error, as expected. Figure 2d, depicting the forecast errors for the Selic rate, shows errors of smaller magnitude, generally not exceeding 1.5 p.p. The period between 2021 and 2023 is the major

exception, where the monetary authority was forced to raise interest rates beyond expectations in response to the intense inflationary shock.

Finally, Figure 2e shows weaker-than-expected GDP growth between 2014 and 2017. Moreover, a large negative surprise in 2020 is followed by a large positive surprise in 2021, reflecting the unexpected "V-shaped" recovery. The highlighted area after October 2021 indicates the period during which the *Top 5* ranking exists. During this period, the projection error was positive but smaller in magnitude compared to other times.

3.2 Projection Standard Deviation Series

We will also analyze the dispersion of agents' projections to deepen the understanding of the forecasters' distribution. To this end, we require the standard deviation series of twelve-month-ahead projections (in a rolling horizon) for the same five variables: GDP (in p.p.), Selic rate (in p.p.), exchange rate (in R$/US$), IPCA (in p.p.), and IGP-M (in p.p.). The series have monthly periodicity. In summary, these series aim to measure how much forecasters "disagree" when forecasting a variable that will be observed only a year later, also serving as a measure of uncertainty. The greater the "disagreement," the greater the estimates' dispersion and, consequently, the higher the standard deviation.

Again, the standard deviation series of the IPCA and IGP-M projections are directly provided by the SEM, requiring no transformation.

Regarding the Selic rate and exchange rate variables, we calculated the average standard deviation for the next year's forecasts. This is done by obtaining the average of the standard deviations of each of the following twelve months' projections for the exchange rate and the next eight Copom meetings for the Selic rate. We also collected the standard deviation of the end-of-horizon projections, showing the dispersion of estimates for the exchange rate twelve months ahead and the Selic rate eight meetings ahead.

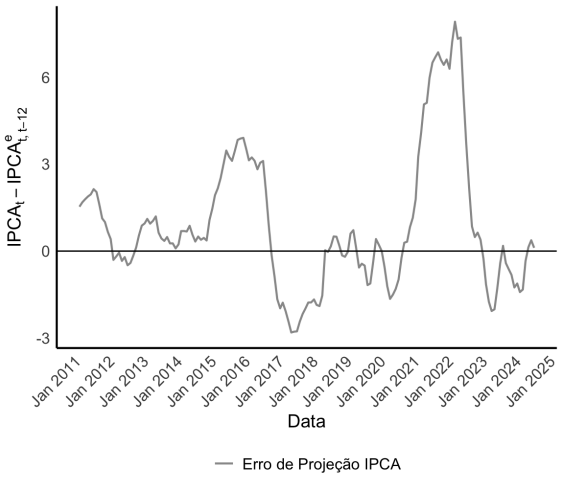
Finally, the case of GDP is, again, more complex. As we only have the standard deviation of quarterly *YoY* projections, some kind of interpolation is required to estimate the standard deviation of monthly *YoY* projections (or in a twelve-month rolling horizon). We cannot use the weights presented in Table 1 since they indicate each month's contribution to quarterly growth rather than its contribution to uncertainty in quarterly growth projections. Given the lack of information about the uncertainty associated with each month, we opted for a less sophisticated interpolation.

We start with the following observation: the *YoY* GDP growth projection for the first quarter of 2011, disclosed on the last day of March 2010, represents exactly a twelve-month horizon. The same applies to the *YoY* GDP growth projection for the second quarter of 2011, disclosed on the last day of June 2010[13]. Therefore, we must estimate the standard deviation of GDP growth estimates a year ahead for the days between the end of the first quarter (March 31) and the end of the second quarter (June 30). This standard deviation

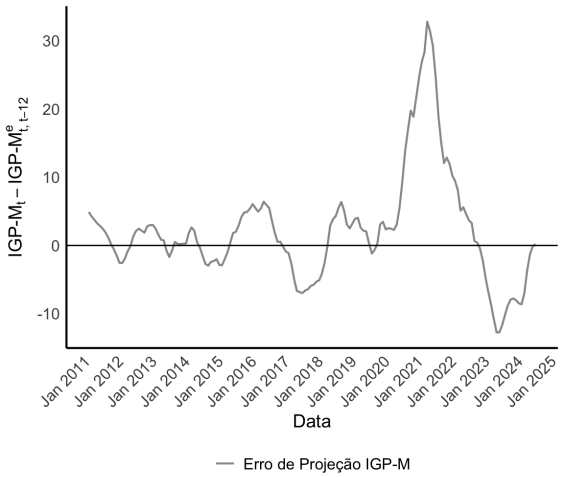
[13]If we used quarterly periodicity data, obtaining the standard deviation of estimates in twelve-month rolling horizons would be trivial.

Figure 2: Series Used - Projection Errors

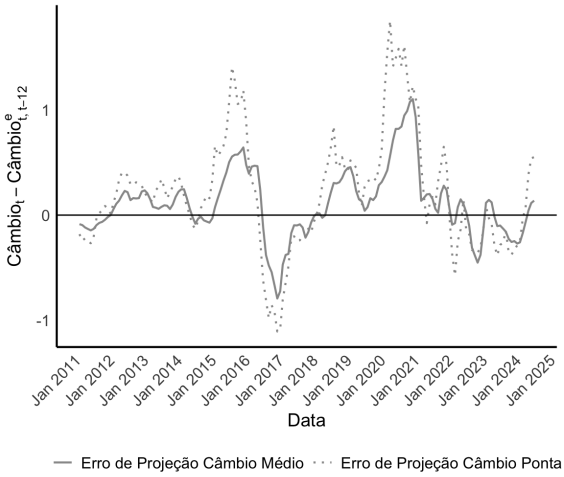
(a) Projection Error - IPCA



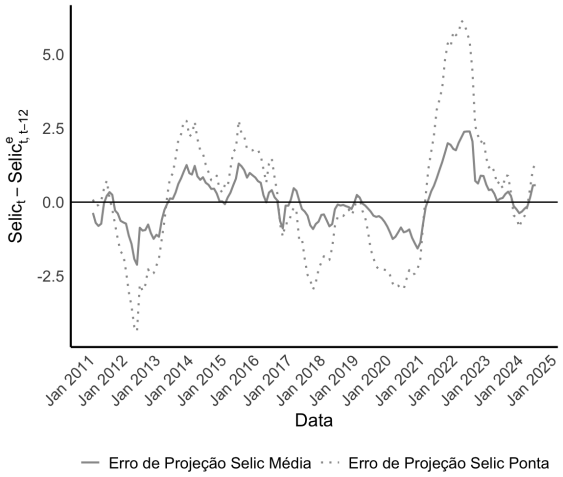
(b) Projection Error - IGP-M



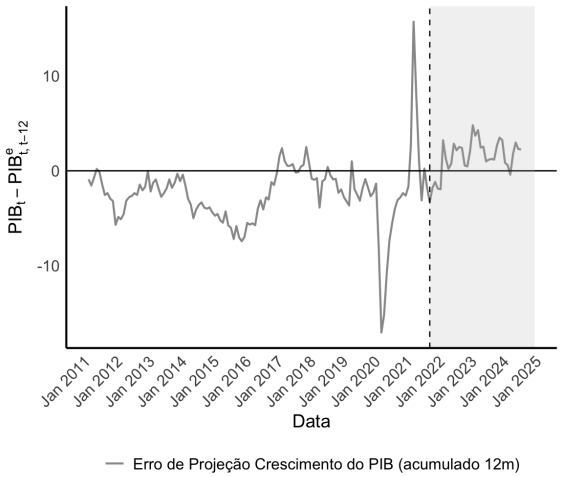
(c) Projection Error - Exchange Rate



(d) Projection Error - Selic



(e) Projection Error - GDP



should be closer to the March 31 estimate if the date of the estimate is closer to March 31 than June 30. Thus, for dates between March 31 and June 30, a weighted average is calculated:

$$\sigma_t = \frac{d_{t,\text{Jun 10}}\,\sigma_{\text{Mar 10}} + d_{t,\text{Mar 10}}\,\sigma_{\text{Jun 10}}}{d_{\text{Mar 10,Jun 10}}} \tag{8}$$

where t represents the day for which the standard deviation is to be estimated, and d_{t_1,t_2} represents the number of days separating t_1 and t_2. Note that $d_{t,\text{Jun 10}}$ serves as a weight for $\sigma_{\text{Mar 10}}$ because the further t is from June, the closer it is to March, with $\sigma_{\text{Mar 10}}$ deserving greater weight. These calculations, exemplified for the period between March 31, 2010, and June 30, 2010, apply to all other dates in the database, without loss of generality.

Afterward, the σ_t estimates are aggregated into months, resulting in a monthly periodic time series that begins in March 2010 and ends in October 2024, like the other four series.

Figures 3a to 3e depict the trajectories of the twelve-month-ahead projection standard deviations for the five variables considered. Generally, during periods of uncertainty, there are significant jumps in this dispersion measure. During the pandemic, projections for all variables, except the Selic rate, showed high standard deviations, reflecting the shocks experienced and the disagreement among forecasters.

Specifically, Figure 3a depicts high disagreement among IPCA variable forecasters during the years 2015 and 2016, the initial pandemic shock in 2020, and the subsequent supply shock (in 2021). Figure 3b shows that the period of greatest uncertainty and dispersion among IGP-M forecasters occurred precisely during the supply shock. Meanwhile, Figure 3c indicates that the exchange rate became increasingly unpredictable over time, with a rising standard deviation from 2014 onwards, reflecting the greater volatility of the Brazilian real following the Dilma government crisis. Figure 3d reveals that the standard deviation of Selic rate projections remained closer to the same average (approximately 0.4) with periods of peaks and troughs.

Finally, Figure 3e shows a GDP projection standard deviation around 0.5 p.p., initially peaking in 2011 and 2015. During the start of the pandemic, there is an intense increase in dispersion and forecasters disagreement. From the creation of the *Top 5* ranking in October 2021, there is a stabilization in the standard deviation and a slight decline in the final years of the series, possibly reflecting a convergence to the historical average.

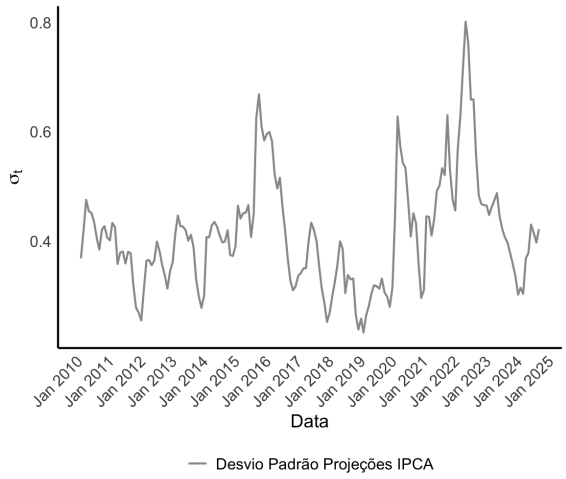
4 Methodology

4.1 Model for the synthetic forecast error of GDP

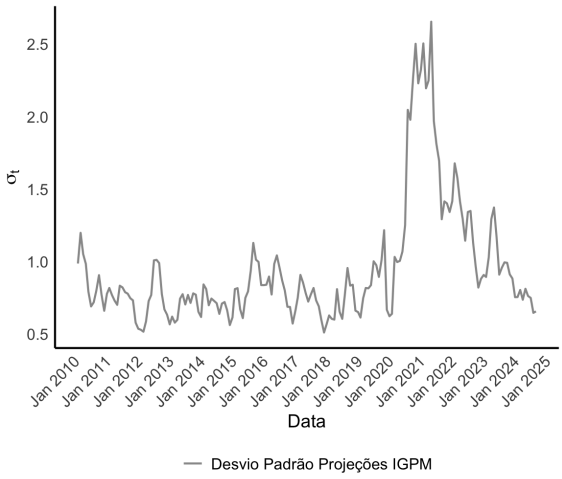
Four of the five series described in Section 3.1 are associated with variables that have always had a *Top 5* ranking. We can call them treated series. The other series (GDP forecast errors) only started being treated after the intervention in 2021. Given the small number of observed units, we base the empirical strategy of this study on an approach similar to the *synthetic control* method proposed by Abadie and Gardeazabal (2003).

Figure 3: Series Used - Standard Deviation

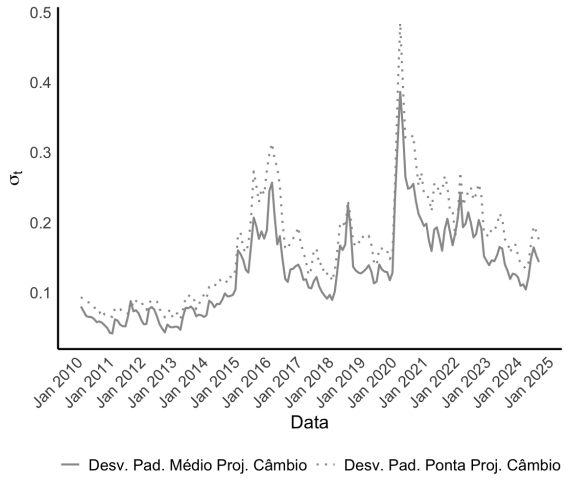
(a) Standard Deviation - IPCA Projections



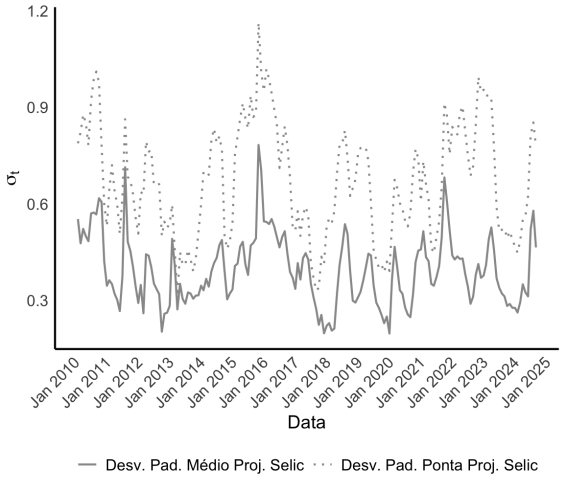
(b) Standard Deviation - IGP-M Projections



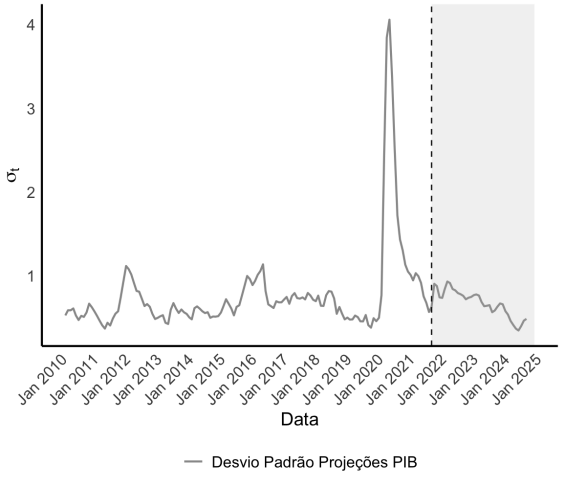
(c) Standard Deviation - Exchange Rate Projections



(d) Standard Deviation - Selic Projections



(e) Standard Deviation - GDP Projections



The synthetic control method is used when the researcher has n units in the sample, with the first $n-1$ units being untreated, and the nth unit receiving the treatment. The goal is to "create" a counterfactual for the treated unit, i.e., understand how it would have been without the intervention. Thus, the treatment effect T_t on a variable $y_t^{i=n}$ of the treated unit is identified as the difference between the observed value of that variable and its "synthetic" counterpart $y_t^{n,s}$: $T_t = y_t^n - y_t^{n,s}$. The variable of the "synthetic" unit is calculated as a linear combination of the same variable from the untreated units:

$$y_t^{n,s} = \sum_{i=1}^{n-1} w_i y_t^i \tag{9}$$

with w_i representing the weight of the ith unit in constructing the counterfactual for unit n. According to Abadie and Gardeazabal (2003), the $((n-1)$ x $1)$ weight vector W is traditionally obtained by minimizing $(X_1 - X_0W)'V(X_1 - X_0W)$, where X_1 is a $(K$ x $1)$ vector of K variables considered good predictors for y_t, X_0 is a $(K$ x $(n-1))$ matrix containing the same K variables for the $n-1$ untreated units, and V is a diagonal matrix whose entries reflect the relative weights of each K variable in predicting y_t. Two properties of the weight vector W are: (1) $w_i \geq 0$ for any i and (2) $\sum_{i=1}^{n-1} w_i = 1$, which serve as constraints in the proposed minimization.

There are some differences between the classical investigation using the synthetic control method and the investigation we intend to conduct in this study. First, our variable of interest (GDP forecast error) is the only untreated one, as all other variables always had the incentive of the *Top 5 ranking*. This is the opposite of the classical synthetic control case, where there is only one treated unit, and the remaining units are untreated (controls). Additionally, we do not have the matrix X_0 or vector X_1 of covariates, and thus cannot find the weights using the minimization proposed by Abadie and Gardeazabal (2003). Finally, the variables are measured in different units and may have negative correlation relationships; for example, a positive IPCA forecast error might be associated with a negative GDP forecast error.

Given these complications, W must be obtained by another way. First, we relax the two constraints, allowing the variable of interest, GDP forecast errors, to be a non-convex linear combination of the other variables (w_i can be negative or positive, and $\sum_{i=1}^{n-1} w_i \neq 1$ is allowed). Once the constraints are relaxed, the weights of the treated series in the untreated series can be obtained through a linear regression without intercept, estimated by OLS:

$$e_t^{\text{GDP}} = w_1 e_t^{\text{IPCA}} + w_2 e_t^{\text{IGP-M}} + w_3 e_t^{\text{Selic}} + w_4 e_t^{\text{Exchange Rate}} + u_t, t \geq \text{October 2022} \tag{10}$$

where e_t^X represents the forecast error series of variable X. The linear model is estimated for the subsample period where $t \geq$ October 2022, as we aim to estimate weights for a linear combination reflecting the existence of the *Top 5* ranking for the GDP growth variable as well. Since e_t^X reflects information from a year prior to t (incorporating a forecast made for t in $t-12$), this estimation can only be performed from $t =$ October 2022 (up to $t =$

August 2024).

This is how we construct a "synthetic treatment", representing what the GDP forecast error series would have been if the incentive to forecasters had always existed, as it does for the other series. After constructing this synthetic treatment, we can compare the observed and synthetic (or counterfactual) series and examine whether the creation of the ranking led to smaller forecast errors. The counterfactual series is therefore:

$$e_t^{\text{GDP},s} = \widehat{w_1} e_t^{\text{IPCA}} + \widehat{w_2} e_t^{\text{IGP-M}} + \widehat{w_3} e_t^{\text{Selic}} + \widehat{w_4} e_t^{\text{Exchange Rate}} \tag{11}$$

where $\widehat{w_i}$ represents a weight obtained by estimating Equation (10) using OLS. The treatment effect in this case is $T_t = |e_t^{\text{GDP},s}| - |e_t^{\text{GDP}}|$, and it is expected that $T_t < 0$, as the creation of the *ranking* should lead to smaller errors.

It is important to note that we are only interested first in predicting what the GDP forecast error series would have been, and not in identifying causal parameters that show, for example, how a higher exchange rate forecast error (or IPCA, Selic, IGPM...) affects GDP forecast error. That is, we are interested in how these other series predict $e_t^{\text{GDP}} (t \geq$ October 2022) in a narrow statistical sense. Thus, we do not need *identification hypotheses*, as the OLS method identifies the OLS "populational" parameters w_1, w_2, w_3, w_4[14].

Causality only appears when we estimate the treatment effect, when we subtract the observed value from the predicted one (or synthetic one). In that way, we need confidence intervals for the synthetic treatment to conclude if the difference between the observed and predicted value is significant, which implies that the treatment effect is different from zero. We define the confidence interval of a given t as:

$$CI_\alpha = [e_t^{\text{GDP},s} \pm t_{T-4,\alpha/2} se(e_t^{\text{GDP},s})] \tag{12}$$

where t_{T-4} is the Student's t statistic with $T-4$ degrees of freedom, α represents the significance level, and T is the number of periods from October 2022 to the end of the time series.

4.2 Model for the Synthetic Standard Deviation of GDP Forecasts

As previously mentioned, just as the SEM provides the median of the forecasts from responding agents, it also provides the standard deviation of the reported forecasts. Thus, it is possible to test whether the creation of the *Top 5* ranking resulted in a lower dispersion of these estimates in addition to reducing forecast errors. Here, we are not only interested in the accuracy of the projections themselves but also in determining whether the competition promoted by the BCB led to a more concentrated distribution of forecasts around their mean, thereby reducing disagreement among agents.

To this end, we follow the same methodology as in the previous section, estimating the

[14] Which, again not necessarily have a causal interpretation

following equation via OLS:

$$\sigma_t^{\text{GDP}} = \tilde{w}_1\sigma_t^{\text{IPCA}} + \tilde{w}_2\sigma_t^{\text{IGP-M}} + \tilde{w}_3\sigma_t^{\text{Selic}} + \tilde{w}_4\sigma_t^{\text{Exchange Rate}} + u_t, \quad t \geq \text{October 2021} \tag{13}$$

where σ_t^X represents the series of the standard deviation of one-year-ahead forecasts for variable X at time t. The linear model is estimated over the sample subperiod where $t \geq$ October 2021, as we aim to estimate the weights of a linear combination that reflects the presence of the *Top 5* ranking for the GDP growth variable as well.

From this, we construct a series of synthetic standard deviations for GDP forecasts, reflecting how dispersed they would have been had the *Top 5* ranking always existed. The counterfactual series is therefore given by:

$$\sigma_t^{\text{GDP},s} = \widehat{\tilde{w}_1}\sigma_t^{\text{IPCA}} + \widehat{\tilde{w}_2}\sigma_t^{\text{IGP-M}} + \widehat{\tilde{w}_3}\sigma_t^{\text{Selic}} + \widehat{\tilde{w}_4}\sigma_t^{\text{Exchange Ratr}} \tag{14}$$

where $\widehat{\tilde{w}_i}$ represents the weight obtained by estimating Equation (13) via OLS. The treatment effect is identified analogously, as $T_t = \sigma_t^{\text{GDP},s} - \sigma_t^{\text{GDP}}$. Similarly, we do not need *identification hypotheses* as we are concerned first with prediction.

The construction of confidence intervals follows the same logic:

$$CI_\alpha = [\sigma_t^{\text{GDP},s} \pm t_{T-4,\alpha/2} se(\sigma_t^{\text{GDP},s})] \tag{15}$$

where t_{T-4} is the Student's t statistic with $T-4$ degrees of freedom, α is the significance level, and T represents the number of periods from October 2021 to the end of the time series.

4.3 Two Channels: Is There an Incentive for Greater Accuracy or Just Imitation of the Top 5?

It is fair to question whether an improvement in projection errors might simply be the result of imitating the forecasts of the best forecasters[15], rather than stemming from the incentive generated by the competition in the *Top 5* ranking. To address this valid concern, we show that it is indeed possible to separate the effect of the BCB intervention into two channels: the "copy" channel and the "incentive through public exposure" channel. But how can we determine which of these channels is active? Or, in other words, how can we determine which one has a significant effect? Figure 4 helps answer these questions.

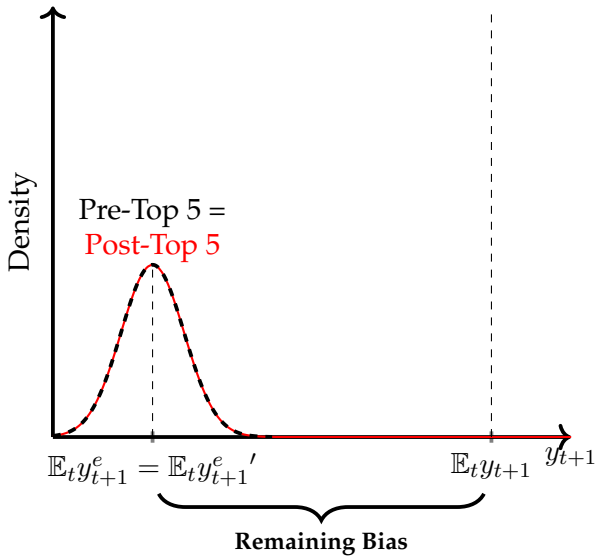
First, we can think that the "copy channel" should also influence the standard deviation of the forecasts, since if many forecasters copy the projections of the experts, in addition to the projection error decreasing, the density of forecasters becomes more concentrated around their mean. Figures 4a to 4d illustrate this dynamic.

In all cases, we assume that there is an initial non-zero systemic bias, since $\mathbb{E}_t y_{t+1}^e \neq \mathbb{E}_t y_{t+1}$, where $\mathbb{E}_t y_{t+1}^e$ is the mean of the projections of the forecasters made at t about y_{t+1}

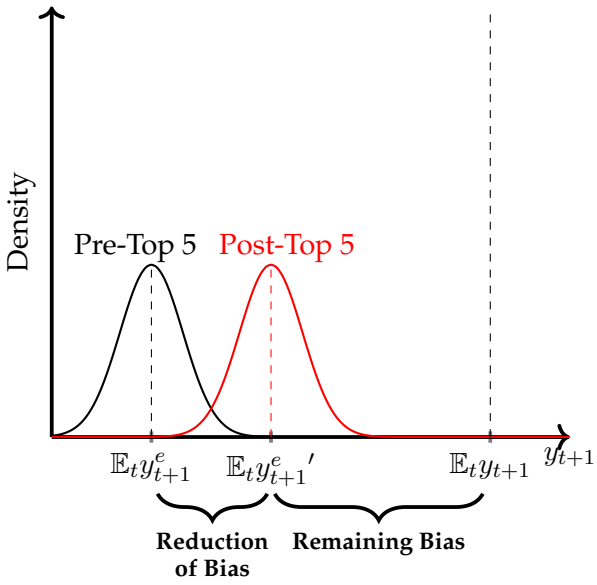
[15] As explained in section 2.1, the BCB also provides the median of the Top 5 projections.

Figure 4: Densities: Four Scenarios

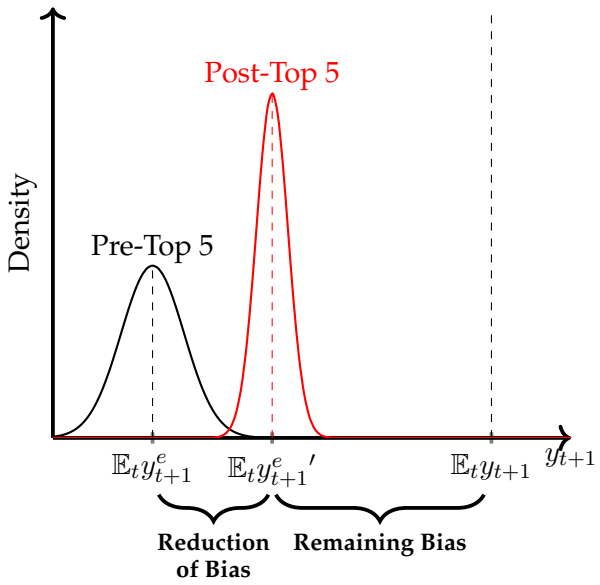
(a) Scenario 1: Weak Incentive, Weak Copy



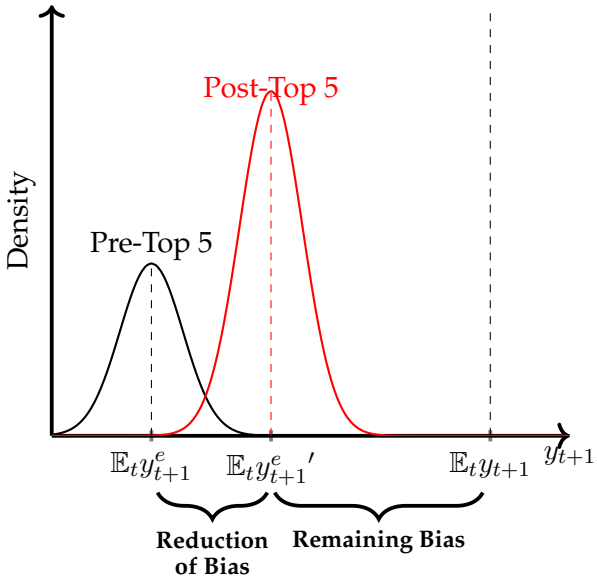
(b) Scenario 2: Strong Incentive, Weak Copy



(c) Scenario 3: Weak Incentive, Strong Copy



(d) Scenario 4: Strong Incentive, Strong Copy



and $\mathbb{E}_t y_{t+1}$ is the expectation of y_{t+1} conditional on the information available at t. We define $\mathbb{E}_t y^e_{t+1}{}'$ as the mean of the projections of the forecasters made at t about y_{t+1} in a counterfactual scenario, where the *Top 5* competition is introduced.

The first scenario (Figure 4a), where neither of the two channels is active, shows that the agents' density remains unchanged. In other words, the competition promoted by the *Top 5* would be completely ineffective in inducing improvements in the agents' projections. If only the "incentive channel" is active (Figure 4b), we should observe only a shift in the density, reducing the implicit bias. There is no reason to assume that the standard deviation of the forecasters' density is altered by the "incentive channel", although we cannot prove that this does not occur. Therefore, we will make the strong assumption that any reduction in the standard deviation is solely due to the "copy channel", which may also reduce the bias, as shown in Figure 4c. Finally, Figure 4d shows that a scenario with improved forecasting errors (reduction of bias) and a reduction in the standard deviation of the agents' density may have both channels active. Thus, if we observe both a reduction in projection errors and in the standard deviation in our counterfactual exercise, we cannot be certain whether we are in Scenario 3 or 4.

5 Results

5.1 Estimator for the Synthetic Forecast Error of GDP

Table 2 presents the results of estimating (10) by OLS. As mentioned, the model is estimated with 23 observations, corresponding to the period from October 2022 to August 2024. Column (1) shows the results for the specification with the forecast errors of the Selic rate and exchange rate calculated as the averages of the following twelve months. Column (2) shows the results for the specification with the forecast errors of these two variables measured at the end of the period. It is noteworthy that the estimation for both specifications was very similar, with only the estimator associated with the Selic rate forecast error (e_t^{Selic}) being statistically significant. The values of R^2 and the adjusted R^2 also remain close, indicating a similar level of fit for both regressions (relatively high, between 0.6 and 0.7).

The low number of observations is the main obstacle in estimating the synthetic treatment weights and, therefore, in identifying the treatment effect. The low degrees of freedom imply a wider confidence interval for the estimates of the counterfactual GDP projection error. Nevertheless, we observe in Figure 5 that between the end of 2013 and early 2016, there is a significant treatment effect on the counterfactual series (even with the wide confidence intervals[16]). This result is verified for both specifications (with the average forecast error of Selic and exchange rate and the end-of-period error), as between October 2013 and February 2016, the observed series of GDP forecast errors was either outside or very close to the 95% confidence interval of the synthetic forecast error estimate. The shaded

[16]The thinner red line in the figures represents the 95% confidence interval.

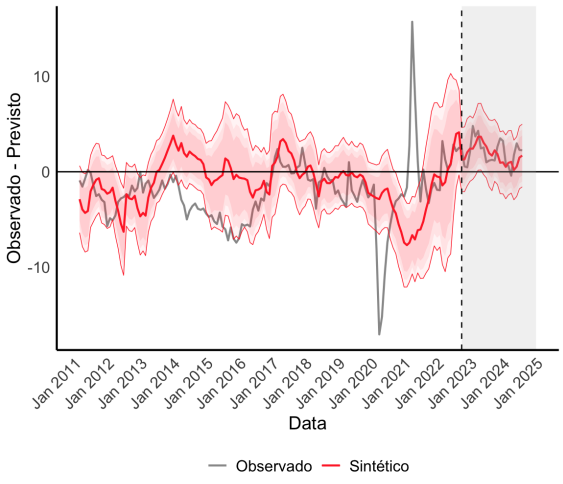
Table 2: Estimation Results of (10) by OLS

	Dependent Variable	
	e_t^{GDP}	
	(1)	(2)
e_t^{IPCA}	−0.797	−0.504
	(0.687)	(0.636)
$e_t^{IGP\text{-}M}$	−0.104	−0.160
	(0.100)	(0.099)
$e_t^{Exchange\ Rate}$ (average)	0.466	
	(1.765)	
e_t^{Selic} (average)	3.036***	
	(0.815)	
$e_t^{Exchange\ Rate}$ (end)		1.733
		(1.262)
e_t^{Selic} (end)		1.023***
		(0.298)
Observations	23	23
R^2	0.698	0.700
Adjusted R^2	0.634	0.637
Residual Std. Error (df = 19)	1.475	1.471
F-statistic (df = 4; 19)	10.977***	11.075***
Note:	*p<0.1; **p<0.05; ***p<0.01	

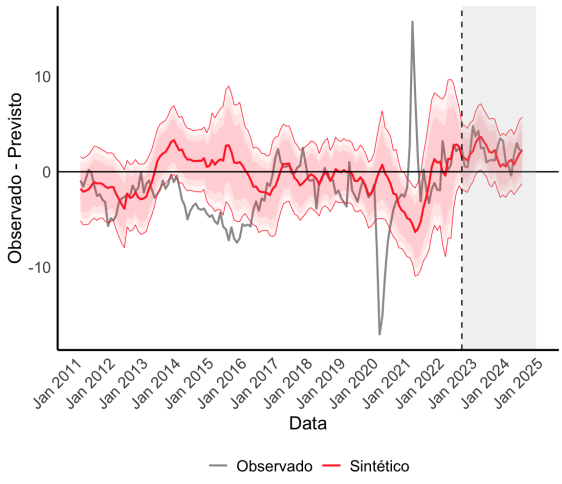
area (in gray) represents the subperiod in which the estimation was conducted.

Figure 5: Synthetic Forecast Error of GDP

(a) Specification 1 - Average Error



(b) Specification 2 - End-of-Period Error



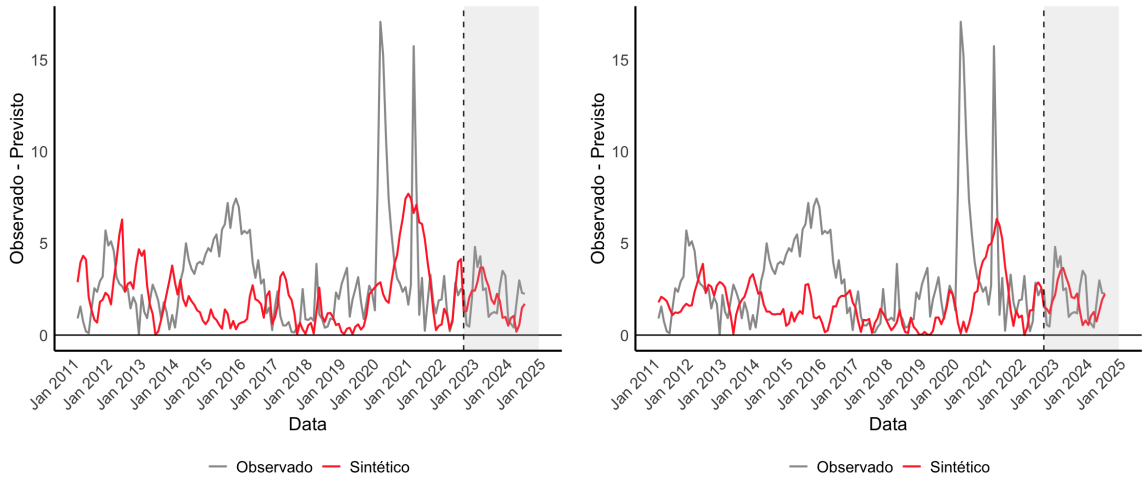
The synthetic error was also smaller during the pandemic period. Thus, we conclude that the existence of competition among forecasters induces smaller forecasting errors during times of crisis and uncertainty, and has little effect during periods of greater stability,

at least in the considered period. Figure 6 shows the absolute value series of the synthetic and observed errors, where we can observe the effect of the *Top 5* ranking more clearly. We see that the forecast errors would have indeed been closer to zero in general if the ranking had always existed (again, with emphasis on the period between 2013 and 2016).

Figure 6: Synthetic Forecast Error of GDP (Absolute Value)

(a) Specification 1 - Average Error

(b) Specification 2 - End-of-Period Error



Finally, Table 3 presents objective evaluation metrics for the forecasts. From these metrics, we observe that the synthetic errors (associated with both specifications) are smaller than the observed errors, with lower RMSE, MAE, and Theil U values for both synthetic series compared to the observed series. In fact, the synthetic model for specification 1 is the only one superior to the "naive" model, with a Theil U value below one, while the observed errors imply a Theil U value of approximately 2.1. The information in the table presents the same message as the the graphs, making it clear that the errors should have been smaller in the pre-ranking period ($t <$ October 2022).

Table 3: Forecast Evaluation Metrics

	Observed	Synthetic (1)	Synthetic (2)
RMSE (Root MSE)	3.907	2.685	**2.066**
MAE (Mean Abs. Error)	2.885	2.087	**1.671**
Theil U	2.127	**0.865**	1.212

5.2 Estimator for the Synthetic Standard Deviation of GDP Forecasts

Table 4 presents the results of estimating (13) by OLS, based on 37 observations covering the period from October 2021 to October 2024. Column (1) presents the results for the specification in which the standard deviations of the Selic rate and exchange rate projections were calculated as the averages of the standard deviations of the forecasts over the next twelve months (for the exchange rate) and the next eight meetings (for the Selic rate).

Column (2) presents the specification with the standard deviations of these two variables measured at the end of the period. The results of both specifications are similar. The standard deviation of the IGP-M projections ($\sigma_t^{\text{IGP-M}}$) is statistically significant in both models, while the estimator associated with σ_t^{Selic} is significant only in the second model. Additionally, the R^2 and adjusted R^2 values remained close and indicate a good fit for both regressions.

Table 4: OLS Estimation Results for (13)

	Dependent Variable	
	σ_t^{GDP}	
	(1)	(2)
σ_t^{IPCA}	0.337	0.153
	(0.249)	(0.220)
$\sigma_t^{\text{IGP-M}}$	0.263***	0.272***
	(0.088)	(0.076)
$\sigma_t^{\text{Exchange Rate}}$ (average)	1.298	
	(0.908)	
σ_t^{Selic} (average)	0.047	
	(0.173)	
$\sigma_t^{\text{Exchange Rate}}$ (end)		0.731
		(0.577)
σ_t^{Selic} (end)		0.228**
		(0.101)
Observations	37	37
R^2	0.981	0.984
Adjusted R^2	0.979	0.983
Residual Standard Error (df = 33)	0.101	0.091
F-statistic (df = 4; 33)	425.310***	522.582***
Note:	*p<0.1; **p<0.05; ***p<0.01	

In this case, since we now have a larger number of observations (since we can count the observations from October 2021 to September 2022) and given the high fit of the regression, the confidence intervals are narrower, making it easier to identify a possible treatment effect. Thus, Figure 7 shows the synthetic standard deviation of GDP forecasts.

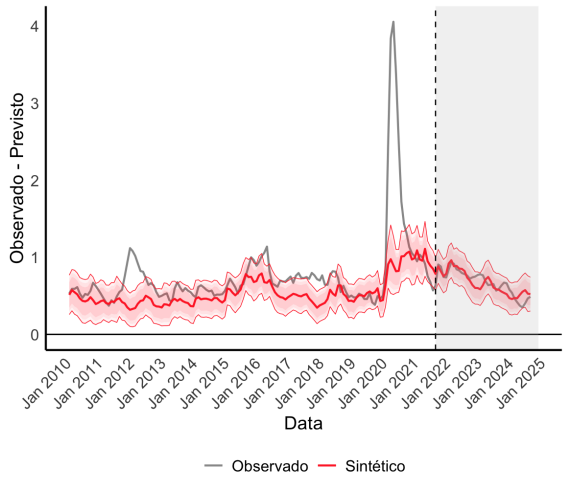
For both specifications, the synthetic standard deviation closely resembles the observed one. In some periods, it is lower: in 2012, when a small peak is observed; between early 2017 and mid-2018; and in the initial shock of the pandemic in 2020, when the observed standard deviation spikes and the synthetic one shows only a slight increase.

Therefore, by analyzing Figures 5 and 7 together, we can understand which of the scenarios depicted in Figure 4 occurred in the sample period (between 2011 and the creation of the ranking in 2021).

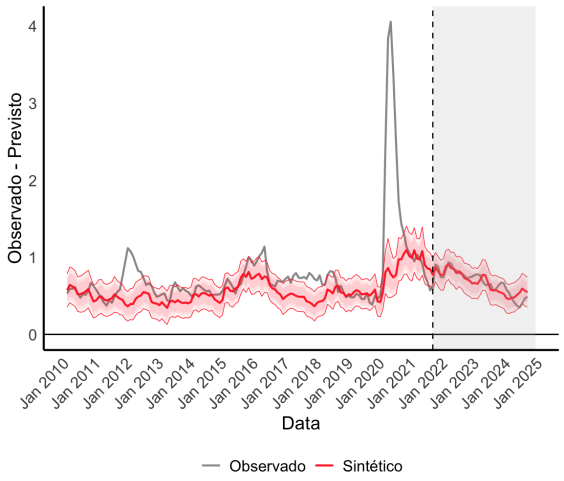
First, throughout most of the sample period, there was no effect in improving the accuracy of the forecasters' estimates. As mentioned, the two periods where a significant improvement in these estimates occurred were during the 2013-2016 crisis and during the

Figure 7: Synthetic Standard Deviation of GDP Forecasts

(a) Specification 1 - Average SD



(b) Specification 2 - End SD



pandemic in 2020-2021. In the first interval, there was no significant decrease in the standard deviation, showing that if the ranking had existed, the forecasters' density would likely not have been more concentrated. Even so, between 2013 and 2016, there was an improvement in projections, implying that only the "incentive channel" would have been active, and Scenario 2 would likely have been observed (Figure 4b, with only a shift in density). In the second period of improvement, between 2020 and 2021, the synthetic standard deviation is significantly lower, implying that the "copying channel" would likely have been used if the competition among agents had existed during that period. In this case, we would observe Scenario 3 or 4, though we cannot be sure whether the "incentive channel" would also have been active (Figure 4c and Figure 4d).

Thus, we can interpret that, possibly, during periods of structural crises, when agents have time to absorb and interpret information, gradually adapting their models, the "copying channel" is not as frequently used, and the "incentive channel" predominates. Meanwhile, in a crisis resulting from an intense and completely unexpected shock, when there is not enough time to absorb information and adapt models, agents prefer to copy the experts, leading to a lower standard deviation of the published projections and also a lower prediction error.

5.3 Possible Limitations

The first limitation of the exercise presented is related to the classical discussion of internal and external validity. It is not possible to assert that, with the creation of competition among forecasters, the "incentive" channel would be active in all structural crises, while the "copying" channel would be relevant only in unexpected crises. With the methodology used, we are unable to identify structural parameters. However, we offer a counterfactual illustration based on our sample, contributing significantly to the literature in this area.

Another limitation arises from the low degrees of freedom in the estimation of (10). However, this problem will disappear when we redo the exercise in the near future, as we gain one observation per month.

6 Conclusion

Forecasters may not only care about the accuracy of their projections. In their decision-making process, other factors are likely considered, as extensive empirical evidence illustrates. There are indications that forecasters avoid being labeled as "bad" by refraining from optimally revising their forecasts and seek publicity gains by distancing themselves from the consensus.

By creating a competition that rewards the "best" forecasters, the Central Bank of Brazil (BCB) links publicity to a measure of accuracy. We investigated whether this initiative, the creation of the *Top 5* ranking for GDP forecasts, is effective in reducing forecast errors. Furthermore, we sought to understand whether this potential reduction is caused precisely by the competition element between agents ("incentive channel") or by mere copying of the best forecasters, as the BCB also discloses the projections of the *Top 5* ("copy channel"). The key distinction between the two channels is that the "copying channel" increases the concentration of forecasters' density, reducing the standard deviation of estimates, while there is no reason to assume that the "incentive channel" affects this standard deviation.

For the empirical investigation of the issue, we used series of forecast errors twelve months ahead for the variables IPCA, IGP-M, Selic, Exchange Rate, and GDP, covering March 2011 to August 2024. In addition, series of standard deviations of forecasts twelve months ahead for these same variables were used, covering March 2010 to October 2024. We aimed to estimate what the time series of forecast errors and the standard deviation of GDP forecasts would have been had the ranking always existed. In other words, we sought a counterfactual series, or a "synthetic treatment", in an attempt to identify the effect of the ranking (or treatment).

Between late 2013 and early 2016, during the Dilma II administration's recession, forecast errors would have been lower, with the treatment effect being statistically significant. During this period, the standard deviation would not have changed significantly. This implies that between 2013 and 2016, only the "incentive channel" would have been active. During the pandemic, both forecast errors and the standard deviation of GDP forecasts would have decreased, implying that the "copy channel" was active during that period (with these results, we could not determine if the "incentive channel" was also active).

We interpret these results by considering that in structural crises, there is time for agents to incorporate new information and adapt their models, without the need for copying. In sudden, unexpected crises, agents rely on copying as a means of incorporating more information (which may be scarce or noisy). Thus, we conclude by stating that the Top 5 ranking has significant effects on the accuracy of forecasters, but only in times of

greater uncertainty and insecurity.

References

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, *93*(1), 113–132.

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

Batchelor, R. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, *23*(2), 189–203.

BCB. (2024a). *Dólar - compra (média mensal), série número 3697.* Disponível em: https://www3.bcb.gov.br/sgspub. (Acesso em: 20 out. 2024)

BCB. (2024b). *Ibc-br - com ajuste sazonal, série número 24364.* Disponível em: https://www3.bcb.gov.br/sgspub. (Acesso em: 20 out. 2024)

BCB. (2024c). *Sem (sistema de expectativas de mercado).* Disponível em: https://www3.bcb.gov.br/expectativas2/#/consultas. (Acesso em: 20 out. 2024)

BCB. (2024d). *Taxa de juros - selic, série número 11.* Disponível em: https://www3.bcb.gov.br/sgspub. (Acesso em: 20 out. 2024)

Carvalho, A. (2015). Tailored proper scoring rules elicit decision weights. *Judgment and Decision Making*, *10*(1), 86–96. doi: 10.1017/S193029750000320X

Carvalho, F. A., & Minella, A. (2012). Survey forecasts in brazil: a prismatic assessment of epidemiology, performance, and determinants. *Journal of International Money and Finance*, *31*(6), 1371–1391.

Ehrbeck, T., & Waldmann, R. (1996). Why are professional forecasters biased? agency versus behavioral explanations. *The Quarterly Journal of Economics*, *111*(1), 21–40.

FGV-IBRE. (2024). *Igp-m (Índice geral de preços - mercado).* Disponível em: https://extra-ibre.fgv.br/IBRE/sitefgvdados/default.aspx. (Acesso em: 20 out. 2024)

Fraga, A. (2009). Dez anos de metas para a inflação. *Dez Anos de Metas para A Inflação No Brasil*, 23.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359–378.

Guillén, D. (2008). *Ensaios sobre expectativas de inflação no brasil* (Tese de Mestrado). PUC-Rio.

Hossain, T., & Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, *80*(3), 984–1001.

IBGE. (2024a). *Ipca (Índice nacional de preços ao consumidor amplo).* Disponível em: https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html. (Acesso em: 20 out. 2024)

IBGE. (2024b). *Pib (produto interno bruto).* Disponível em: https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9300-contas-nacionais-trimestrais.html. (Acesso em: 20 out. 2024)

Ito, T. (1990). Foreign exchange rate expectations: Micro survey data. *The American Economic Review*, *80*(3), 434–449. Retrieved 2024-10-02, from http://www.jstor.org/stable/2006676

Laster, D., Bennett, P., & Geoum, I. S. (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics*, *114*(1), 293–318.

Marques, A. B. C. (2012). Central bank of brazil's market expectations system: a tool for monetary policy. *IFC Bulletin*, *36*, 304–324.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). NBER.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: journal of the Econometric Society*, 315–335.

Nasser, H., & De-Losso, R. (2021). Vieses comportamentais em projeções macroeconômicas. *Estudos Econômicos (São Paulo)*, *51*, 285–310.

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

Theil, H., Beerens, G., Tilanus, C., & De Leeuw, C. B. (1966). *Applied economic forecasting* (Vol. 4). North-Holland Publishing Company Amsterdam.

Winkler, R. L., & Murphy, A. H. (1968). "good" probability assessors. *Journal of Applied Meteorology (1962-1982)*, *7*(5), 751–758. Retrieved 2024-10-08, from http://www.jstor.org/stable/26174473

Witkowski, J., Freeman, R., Vaughan, J. W., Pennock, D. M., & Krause, A. (2023). Incentive-compatible forecasting competitions. *Management Science*, *69*(3), 1354–1374.

Appendices

A Extended Forecaster Problem

A.1 Advertising Not Linked to Accuracy

We address the problem of the extended *forecaster* described in (2), considering the agent's initial goal of being as accurate as possible, while also incorporating the "advertising hypothesis" ("anti-herding bias") and an "accommodation bias".

First, we establish weights $\delta > 0$ assigned to the agent's accuracy (δ_1), the aversion to changes from the previous forecast (δ_2), and advertising (δ_3). We propose a functional

form for $F(x)$ to obtain an analytical solution for p^e_{t+1}:

$$\max_{p^e_{t+1}} \mathbb{E}_t U = \mathbb{E}_t[-\delta_1(p_{t+1} - p^e_{t+1})^2 - \delta_2(p^{e,\text{previous}}_{t+1} - p^e_{t+1})^2 + \delta_3(p^{e,\text{consensus}}_{t+1} - p^e_{t+1})^2] \quad (16)$$

Which is equivalent to:

$$\max_{p^e_{t+1}} \mathbb{E}_t U = -\delta_1 \mathbb{E}_t[(p_{t+1} - p^e_{t+1})^2] - \delta_2(p^{e,\text{previous}}_{t+1} - p^e_{t+1})^2 + \delta_3(p^{e,\text{consensus}}_{t+1} - p^e_{t+1})^2 \quad (17)$$

Using the Leibniz rule for expected values, the FOC is as follows:

$$2\delta_1 \mathbb{E}_t[(p_{t+1} - p^e_{t+1})] + 2\delta_2(p^{e,\text{previous}}_{t+1} - p^e_{t+1}) - 2\delta_3(p^{e,\text{consensus}}_{t+1} - p^e_{t+1}) = 0 \quad (18)$$

$$\Leftrightarrow \delta_1 \mathbb{E}_t p_{t+1} - \delta_1 p^e_{t+1} + \delta_2 p^{e,\text{previous}}_{t+1} - \delta_2 p^e_{t+1} - \delta_3 p^{e,\text{consensus}}_{t+1} + \delta_3 p^e_{t+1} = 0$$

$$\Leftrightarrow (\delta_1 + \delta_2 - \delta_3) p^e_{t+1} = \delta_1 \mathbb{E}_t p_{t+1} + \delta_2 p^{e,\text{previous}}_{t+1} - \delta_3 p^{e,\text{consensus}}_{t+1}$$

Thus, the forecast chosen by the *forecaster* is a weighted average of the expected value of p_{t+1} with weight δ_1, the previous forecast $p^{e,\text{previous}}_{t+1}$ with weight δ_2, and the market consensus forecast $p^{e,\text{consensus}}_{t+1}$ with weight $-\delta_3$:

$$p^e_{t+1}{}^* = \frac{1}{\delta_1 + \delta_2 - \delta_3}(\delta_1 \mathbb{E}_t p_{t+1} + \delta_2 p^{e,\text{previous}}_{t+1} - \delta_3 p^{e,\text{consensus}}_{t+1}) \quad (19)$$

It follows that $p^e_{t+1}{}^* \neq \mathbb{E}_t p_{t+1}$ for $\delta_2, \delta_3 \neq 0$. The expected bias must be:

$$\mathbb{E}_t[p_{t+1} - p^e_{t+1}{}^*] = \frac{\delta_2}{\delta_1 + \delta_2 - \delta_3}(\mathbb{E}_t p_{t+1} - p^{e,\text{previous}}_{t+1}) - \frac{\delta_3}{\delta_1 + \delta_2 - \delta_3}(\mathbb{E}_t p_{t+1} - p^{e,\text{consensus}}_{t+1}) \quad (20)$$

A.2 Advertising Linked to Accuracy

Now, suppose that advertising gains are also associated with a measure of accuracy, where the agent becomes recognized for having small errors. The agent's utility function is:

$$\max_{p^e_{t+1}} \mathbb{E}_t U = \mathbb{E}_t[-\delta_1(p_{t+1} - p^e_{t+1})^2 - \delta_2(p^{e,\text{previous}}_{t+1} - p^e_{t+1})^2 + \delta_3((p^{e,\text{consensus}}_{t+1} - p^e_{t+1})^2 - (p_{t+1} - p^e_{t+1})^2)] \quad (21)$$

Which is equivalent to:

$$\max_{p^e_{t+1}} \mathbb{E}_t U = -(\delta_1 + \delta_3)\mathbb{E}_t[(p_{t+1} - p^e_{t+1})^2] - \delta_2(p^{e,\text{previous}}_{t+1} - p^e_{t+1})^2 + \delta_3(p^{e,\text{consensus}}_{t+1} - p^e_{t+1})^2 \quad (22)$$

The FOC now is:

$$(\delta_1 + \delta_3)\mathbb{E}_t[(p_{t+1} - p^e_{t+1})] + \delta_2(p^{e,\text{previous}}_{t+1} - p^e_{t+1}) - \delta_3(p^{e,\text{consensus}}_{t+1} - p^e_{t+1}) = 0 \quad (23)$$

$$\Leftrightarrow (\delta_1 + \delta_3)\mathbb{E}_t p_{t+1} - (\delta_1 + \delta_3) p^e_{t+1} + \delta_2 p^{e,\text{previous}}_{t+1} - \delta_2 p^e_{t+1} - \delta_3 p^{e,\text{consensus}}_{t+1} + \delta_3 p^e_{t+1} = 0$$

$$\Leftrightarrow (\delta_1 + \delta_2) p^e_{t+1} = (\delta_1 + \delta_3)\mathbb{E}_t p_{t+1} + \delta_2 p^{e,\text{previous}}_{t+1} - \delta_3 p^{e,\text{consensus}}_{t+1}$$

By isolating p^e_{t+1}, we have:

$$p^e_{t+1}{}^* = \frac{1}{\delta_1 + \delta_2}((\delta_1 + \delta_3)\mathbb{E}_t p_{t+1} + \delta_2 p^{e,\text{previous}}_{t+1} - \delta_3 p^{e,\text{consensus}}_{t+1}) \tag{24}$$

The expected bias is:

$$\mathbb{E}_t[p_{t+1} - p^e_{t+1}{}^*] = \frac{\delta_2}{\delta_1 + \delta_2}(\mathbb{E}_t p_{t+1} - p^{e,\text{previous}}_{t+1}) - \frac{\delta_3}{\delta_1 + \delta_2}(\mathbb{E}_t p_{t+1} - p^{e,\text{consensus}}_{t+1}) \tag{25}$$

Note that the bias in (25) is necessarily smaller in magnitude than in (20) (assuming $\delta_1 + \delta_2 > \delta_3$). This reflects the fact that, by making more extreme forecasts, there are both benefits and costs of advertising in the new model. Thus, the bias must be smaller.

B Representation of the *Top 5* as a Binarized *Scoring Rule*

The mechanism by Hossain and Okui (2013) creates a *scoring rule* in which agent j reports $p^{e,j}_{t+1}$, which they genuinely expect ($p^{e,j}_{t+1} = \mathbb{E}_t p_{t+1}$, assuming rational expectations), regardless of their exact preference structure. In other words, the value $p^{e,j}_{t+1} = \mathbb{E}_t p_{t+1}$ maximizes their expected utility. Thus, there is an alignment of interests between the agent (forecaster) and the principal (BCB). Adapting the mechanism of Hossain and Okui (2013) to the *Top 5* context, the timeline is as follows:

1. The agent reports $p^{e,j}_{t+1}$ to the principal (at t).
2. p_{t+1} is observed.
3. The random variable K is "drawn", independently of the reported $p^{e,j}_{t+1}$.
4. The agent receives reward A if $e^j_{t+1} = |p_{t+1} - p^{e,j}_{t+1}| \leq K$ and reward B otherwise, where they prefer A over B.

Here, K can be interpreted as the error of the fifth-best forecaster. If the error of forecaster j is less than this value, they enter the *Top 5*. Since K potentially depends on the strategic behavior of all forecasters in this game, our goal is to determine whether the strategy $p^{e,j}_{t+1} = \mathbb{E}_t p_{t+1}$ for all j constitutes a Nash equilibrium.

It is not plausible for K to have a uniform distribution as in Hossain and Okui (2013). We know:

$$K = |p_{t+1} - p^{e,j=5}_{t+1}| \tag{26}$$

where $p^{e,j=5}_{t+1}$ is the estimate of the fifth-best forecaster. Assuming $p_{t+1} \sim N(\mu, \sigma^2)$ and that all forecasters know the distribution of p_{t+1} (complete information) and have rational expectations, we can initially assume that the fifth-best forecaster, like all others, reports $p^{e,j=5}_{t+1} = \mathbb{E}_t p_{t+1} = \mu$. Therefore, $X = p_{t+1} - p^{e,j=5}_{t+1} \sim N(0, \sigma^2)$. Thus, $K = |X|$ follows a folded normal distribution.

Now, we can solve the problem for a forecaster j:

$$\max_{p_{t+1}^{e,j}} \mathbb{E}_t U = P(e_{t+1}^j \leq K)U(A) + (1 - P(e_{t+1}^j \leq K))U(B) \tag{27}$$

The problem reduces to maximizing $P(e_{t+1}^j \leq K)$:

$$\max_{p_{t+1}^{e,j}} P(e_{t+1}^j \leq K) = P(|p_{t+1} - p_{t+1}^{e,j}| \leq K) \tag{28}$$

This probability is maximized when the expected absolute error e_{t+1}^j is minimized, which occurs when $p_{t+1}^{e,j} = \mu$. This is because $K \perp p_{t+1}^{e,j}$ for each $j \neq 5$. Hence, we have proven that $p_{t+1}^{e,j} = \mu$ for all j is a Nash equilibrium. Forecaster j has no incentive to deviate.

However, as demonstrated in (6), this result does not hold for asymmetric distributions of p_{t+1}, as j has an incentive to deviate and report the median of p_{t+1} as their estimate, thereby minimizing the expected absolute error. Thus, depending on the distribution of p_{t+1}, $p_{t+1}^{e,j} = \mathbb{E}_t p_{t+1}$ for all j may not constitute a Nash equilibrium.