Inefficient Use of Competitors' Forecasts?*

André Reslow†

December 3, 2021

Abstract

Do forecasters use competitors' forecasts efficiently? Empirical results using a large panel of forecasters suggest that forecasters *underuse* information from their competitors in their estimates for current and next year's annual GDP growth and inflation. The results also show that forecasters pay more attention to their competitors when releasing short-term forecasts compared to medium-term forecasts. A baseline model with noisy and private information supports the underuse interpretation and predicts that it is optimal to pay considerable attention to competitors' work. The theoretical framework also rules out overconfidence as the primary explanation of the observed behavior. When the model is extended to include a revision cost, it can only explain the observed inefficiency and horizon dependency if asymmetric horizon discounting between revisions and forecast errors is assumed.

Keywords: Forecasting, Forecast Efficiency, Consensus Forecasts, Imperfect Information

JEL Classification: C53, D82, E17, E37

*I would like to thank Mikael Carlsson, Jesper Lindé and Erik Öberg for the extensive feedback, and Daniel Bougt, Tobias Broer, Fabio Canova, David Cesarini, Davide Cipullo, Matz Dahlberg, Nils Gottfries, Sofia Hernnäs, Isaiah Hull, Sebastian Jävervall, Gisle Natvik, Oskar Norström Skans, Sven Oskarsson, Luca Repetto, Anna Seim, Janne Tukiainen, Karl Walentin, and the participants in the seminars held at Uppsala University and Sveriges Riksbank for their dialogue and comments. Financial support from *Handelsbankens Forskningsstiftelser* is gratefully acknowledged. The forecast data used in this paper is copyright protected against redistribution but can be purchased from the provider. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of Sveriges Riksbank.

†Payments Department, Sveriges Riksbank and Uppsala Centre for Fiscal Studies (UCFS), Uppsala University. Email: andre.reslow@riksbank.se

1 Introduction

Uncertainty poses a problem for economic agents, who all need to make important decisions based on expectations about the future. Forming expectations is not trivial. While the average consumer, investor, voter, and even policymaker is likely to lack the competence to derive optimal predictions, they can obtain helpful information from those who have made it their livelihood to predict the future: *professional forecasters*. Firms update their expectations when exposed to forecasts from professionals (see, e.g., Coibion et al., 2018), and firms' forecasts are positively and significantly associated with their input choices (see, e.g., Tanaka et al., 2019). When economic agents use forecasts in their decision process, they would like those forecasts to be as accurate and efficient as possible.

Even so, professional forecasters also operate in a world with incomplete information. Even though they are highly skilled, they have different skillsets as well as different information sets. Therefore, an essential source of information for a macroeconomic forecaster is the forecasts from competitors. The starting point of this paper is to assess the extent to which macroeconomic forecasters make efficient use of forecasts from their competitors. If forecasts are inefficient, the resulting outcome can be inefficient firm and household decisions. Nordhaus (1987) presents a useful concept of weak and strong forecast efficiency when evaluating efficiency in macroeconomic forecasts. This paper augments the weak efficiency concept in Nordhaus (1987) to include competitor information in addition to the own lagged information. The paper also deploys a theoretical model with noisy and private information to better understand the forecasters' optimal and actual behavior.

The empirical analysis makes use of high-frequency forecaster firm-level panel data from Consensus Economics. The data used cover ten countries—Canada, France, Germany, Italy, Japan, the Netherlands, Spain, Sweden, the United Kingdom, and the United States—and the data span forecasts made between 1995 and 2017. The dataset is a monthly survey of forecasters who are asked to provide forecasts for the current and next year's annual growth rates. Hence, the forecasts are best characterized as a set of fixed target forecasts where multiple predictions are made for the same outcome at a decreasing forecast horizon. The primary analysis focuses on forecasts for GDP growth and inflation (consumer prices). The data offers a dynamic panel of 591 forecasters. Using this data, it is found that forecasters *underuse* competitor information in their forecasts for current and next year's annual GDP growth and

inflation. The results suggest that forecasters put a sub-optimal high weight on their own lagged forecast and too low a weight on their competitors' work. The empirical results also show that the forecasters behave differently at different forecast horizons. As the forecast horizon decreases, when approaching the realization of the outcome, the forecasters increase the attention to their competitors. However, they are still far from an optimal level.

In line with the empirical analysis, a baseline model with noisy and private information predicts that, for a forecaster to minimize forecast errors, it is optimal to pay considerable attention to competitors' work and, consequently, put a high weight on the forecasts from others. The suggestion that it is optimal to pool information from many forecasters has been raised numerous times and is well-known in the forecasting literature (see, e.g., Elliott et al., 2008; Hendry and Clements, 2004; Zarnowitz, 1984). The typical empirical finding is that consensus (average) forecasts are more accurate over time than individual forecasts. The exercise in this paper should not be seen as a test of whether or not it is optimal to pool forecasts. Instead, it is a test of whether or not forecasters adopt this well-known result in their work.[1] It is important to note that, even if forecasters are efficient and optimally use competitor information, it is still optimal for end-users, such as firms, households, and policymakers, to pool from multiple forecasters due to private information that each forecaster holds. Hence, the well-known pooling result should hold regardless.

Two versions of the model are considered: one with fully informed forecasters and one with "rule of thumb" forecasters. In the fully informed model, the forecaster perfectly filters out signals from competitors to produce the (theoretically) most efficient forecast. The rule of thumb model is used to impose limitations on the forecasters' ability. The forecasters' rule of thumb is to use lagged aggregate forecasts instead of filtering out individual signals. This model, hence, is more in line with the empirical strategy.

The two models differ in efficiency, but both models provide the same prediction regarding relative attention to own versus others' lagged forecasts. Both models are noisy information models that introduce a signal extraction problem for Bayesian updating forecasters.[2] The models' information structure matches the specific structure of fixed target forecasts used in the empirical analysis. The baseline models do not have any form of strategic trade-off and

[1] An ongoing example of incorporating the well-known pooling result is the work by Nate Silver and the FiveThirtyEight polling aggregation website, which focuses on predicting outcomes in politics and sports. See also the book *The Signal and the Noise* by Silver (2012).

[2] In this type of model (see, e.g., Lucas, 1973; Sims, 2003; Woodford, 2001), agents continuously update their information sets and their beliefs via signal extraction.

assume that forecasters' objective is pure forecast error minimization.

In order to explain the observed actual (seemingly inefficient) behavior, the baseline model is extended. The extensions are made to consider alternatives to pure inefficiency. First, the paper rules out one pure cognitive bias by considering forecasters to be overconfident in their ability (signal precision), leading to a behavior response of overweighting private information. It is unlikely that overconfidence can be the sole explanation for the observed behavior. A back-of-the-envelope calculation shows that for overconfidence to explain the observed behavior, forecasters need to think that the precision in their private signal is 100 times more precise than the private signal received by any competitor.

As an alternative, an explicit cost from forecast revisions to capture a credibility loss from making substantial changes is added to the model (see, e.g., Ehrbeck and Waldmann, 1996). Given that forecast smoothing is observed in the data, it is intuitive to consider a backward-looking extension to the model. The extension introduces a trade-off between minimizing the forecast error and minimizing the revisions. However, the revision cost model can only match the observed horizon heterogeneity if it is assumed that the revision cost and the cost from forecast errors have asymmetric horizon discounting.

The main contribution of this paper relates to the literature on the behavior of macroeconomic forecasters. First, the documentation of the *relative* use between forecasters' own lagged forecasts and those of others and how this relationship differs between horizons. The empirical analysis is related to work by, for example, Batchelor and Dua (1992), who find that most U.S. forecasters in the Blue Chip panel could have improved their forecasts if they were less attached to their past forecast and more attached to any new information when making forecast revisions. The analysis is also related to, for example, Lahiri and Sheng (2008), who estimate a Bayesian learning model with heterogeneity aimed at explaining forecast disagreement and its evolution over horizons. More recent related work includes, among others, Coibion and Gorodnichenko (2015), Dovern et al. (2015), Broer and Kohlhas (2018) and Bordalo et al. (2020), who all estimate how forecasters react to new information in general.[3] It is important to keep in mind that forecasters can be efficient (or inefficient) in using competitor information and still under- or overreact to new private information, or different forms of new public information, as

[3] Other related papers include Ager et al. (2009), who study accuracy and efficiency of consensus forecasts, and Isiklar et al. (2006), who study forecast revision to see how quickly forecasters incorporate news in a cross-country setting. Deschamps and Ioannidis (2013) study forecast revisions and find that forecasters underreact to new information. In addition, Chen and Jiang (2006) find that corporate analysts place larger than efficient weights on their private information when they forecast corporate earnings.

is found in the papers cited above. Second, this paper shows in a noisy information framework with private information that this behavior cannot be explained by overconfidence alone. It also shows that the introduction of asymmetric horizon discounting is crucial for revision costs to explain the observed behavior.

The notion of strategic behavior, in addition to forecast-error minimization, among macroeconomic forecasters is not novel. For example, Laster et al. (1999) introduces a trade-off problem for the forecasters, where they have to balance accuracy and publicity. Ottaviani and Sørensen (2006) and Marinovic et al. (2013) generate strategic behavior by considering a forecast contest where forecasters benefit from being the only correct one. Cipullo and Reslow (2019) propose an alternative trade-off where forecasters trade accuracy with electoral incentives. Although strategic behavior is well established in the literature, this paper highlights that, given the observed horizon heterogeneity, strategic behavior is likely to be horizon dependent.

The paper is also related to the work on herd behavior by, for example, Banerjee (1992), but with some differences.[4] In the Banerjee (1992) model, herding is doing what everyone else is doing, even if private information suggests doing something different. In this paper's framework, forecasters still put positive weight on private information and do not explicitly aim to herd. Note that herd behavior would result in the opposite of what is found in this paper. Herd behavior would display inefficient *overuse* of information from others and *underuse* of own information.

The paper proceeds as follows. Section 2 describes the data at hand, and Section 3 outlines the empirical strategy. Section 4 presents the results, and Section 5 investigates potential heterogeneities. A baseline model is established in Section 6 to further support the interpretation of underuse of information and suggested optimal behavior. Section 7 extends the model to understand the observed (seemingly inefficient) behavior. Section 8 concludes.

2 Data

This paper utilizes high-frequency forecaster firm-level forecast data from a monthly survey by Consensus Economics.[5] The data used in this paper covers ten countries—Canada, France, Germany, Italy, Japan, the Netherlands, Spain, Sweden, the United Kingdom, and the United

[4] See, for example, Lamont (2002), Gallo et al. (2002), Bernhardt et al. (2006) and Rülke et al. (2016) for more on herd behavior among forecasters.

[5] The data is copyright protected against redistribution. The data can, however, be purchased from Consensus Economics. See `http://www.consensuseconomics.com` for further details.

States—and span forecasts made between 1995 and 2017. The forecasts refer to the current and next year's annual growth rates. For example, the forecasters were asked each month during 2017 to forecast the 2017 and 2018 annual growth rates. The main analysis is done using forecasts for two variables: GDP growth and inflation (consumer prices).[6]

In order to calculate forecast errors, the actual realizations are needed. The outcome data is collected from the OECD database. For inflation, the latest available numbers are used. For GDP growth, the outcome is given by the first available release (often referred to as real-time data). GDP is often subject to large revisions, and real-time data is often used when evaluating forecast performance for variables that are subject to large revisions.[7]

Table 1 reports firm number descriptive statistics. The sample consists of 591 forecasting firms.[8] Some firms produce a forecast for more than one country. In these cases, the firms are treated as separate entities and are defined at the country level. Hence, the country label identifies the target outcome that is forecasted. Nevertheless, the correlation between the target country and the origin country is very high. Many of the 591 firms are not observed during the entire sample period, and the panel is, therefore, unbalanced. The columns *Mean*, *Median*, *Min*, and *Max* in Table 1 report the firm number statistics from the monthly survey in such a way that the total mean of 19.8 is an indication that each survey month has on average 19.8 forecasters. The firms in the survey are almost exclusively private firms such as financial institutions (banks) and private research or consulting institutions. However, some public institutions are also present.

Table 2 reports descriptive statistics for the forecasts, while Tables A1 and A2 in the appendix report descriptive statistics for the outcomes and forecast errors. From the data, a total of 179,310 observations can be used. From Table 2, we see that the number of observations differs between countries; most observations are from Germany and the United States, and the fewest are from the Netherlands. The number of observations for GDP growth and inflation is similar in total and for each country. The average GDP growth forecast is 1.9 percent, while the average inflation forecast is 1.8 percent. Tables A1 and A2 in the appendix reveal a slight overestimation of both GDP growth and inflation. The average GDP growth outcome targeted by a forecast is 1.7 percent, while the average inflation outcome is also 1.7 percent.

[6] For the U.K., the inflation forecasts refer to both RPIX and HICP (CPI). The RPIX covers the entire period, but toward the end of the sample many forecasters produce a forecast for both measures while others change entirely to HICP. All available price forecasts are used.

[7] Real-time GDP growth data is constructed using the OECD revision database.

[8] Some firm name entries in the raw data have been corrected due to spelling and naming errors.

Table 1: Firm Number Statistics

Country	N	Mean	Median	Min	Max
Canada	46	14.6	15	8	19
France	57	18.7	18	10	26
Germany	63	26.8	27	19	32
Italy	55	14.5	14	4	21
Japan	72	19.5	19	10	25
Netherlands	42	9.9	9	5	15
Spain	45	14.2	14	5	19
Sweden	42	12.9	13	7	17
UK	94	24.4	24	16	37
USA	75	25.0	25	15	32
Total	591	19.8	20	4	37

Notes: N refers to the total number of firms in the entire sample, while Mean, Median, Min, and Max refer to firms by survey month. *Source*: Consensus Economics.

Table 2: Forecast Statistics

		GDP growth			Inflation		
Country	Obs.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Canada	14,872	7,457	2.4	1.0	7,415	1.9	0.5
France	18,600	9,331	1.6	1.0	9,269	1.4	0.6
Germany	27,481	13,739	1.6	1.2	13,742	1.6	0.6
Italy	14,336	7,228	1.1	1.3	7,108	1.9	1.0
Japan	18,450	9,409	1.2	1.5	9,041	0.2	0.9
Netherlands	9,586	4,959	1.7	1.3	4,627	1.9	0.8
Spain	13,927	7,011	2.0	1.7	6,916	2.2	1.2
Sweden	12,885	6,469	2.4	1.3	6,416	1.6	0.9
UK	23,663	12,539	2.0	1.2	11,124	2.7	0.8
USA	25,510	12,778	2.6	1.1	12,732	2.2	0.9
Total	179,310	90,920	1.9	1.4	88,390	1.8	1.0

Notes: Each month, forecasters are asked to forecast the current and next year's annual average growth rate. This results in 24 potential forecast origins (and horizons) for each year's outcome between 1995 and 2017.

3 Empirical Strategy

3.1 Theoretical Underpinnings

It is well established in the forecasting literature that pooling information from multiple forecasters is beneficial (see, e.g., Elliott et al., 2008; Hendry and Clements, 2004; Zarnowitz, 1984). The typical finding is that consensus (average) forecasts are more accurate over time than individual forecasts. The takeaway from this result is that economic agents are advised to rely on forecasts from multiple sources instead of a single forecaster.

To set ideas, let x_t be the target outcome and F_{ith} the forecast for x_t made by forecaster i, at a horizon of h periods before the realization of x_t.[9] Given this notation, the forecast F_{ith+1} is

[9] For example, a forecast made in January 2019, for the 2020 outcome will have $t = 2020$ and $h = 24$, while

released one period *before* the release of the forecast F_{ith}. Further assume that the forecasters in each period have access to a private signal $S_{ith} = x_t + \epsilon_{ith}$, where $\epsilon_{ith} \sim N(0, \sigma_\epsilon^2)$ is a random error, as well as some prior knowledge μ. The forecasters release to the public a forecast given by $F_{ith} = \mathbb{E}(x_t|\mu, S_{ith}) = (1-w)\mu + wS_{ith}$, where w is given by optimal Bayesian updating.[10] In a repeated forecast setting where a forecaster releases multiple forecasts for each x_t, the prior can be expressed as $\mu = F_{ith+1}$. Hence, the prior is the forecast released in the previous period. The rationale behind this assumption is that the previous forecast should summarize all previously known information in an optimal way. A forecast at horizon h is then given by the forecast at horizon $h+1$ updated with the new private signal. In the case of infinite many forecasters, the consensus (average) forecast can be expressed as $\overline{F}_{th} = \int_0^1 F_{ith} di$. In this simple model, $\overline{F}_{th} = \int_0^1 F_{ith} di = x_t$ represents a perfect forecast since the noise in the signal, ϵ_{ith}, cancels out when aggregating over the infinite amount of forecasters.

Even though the assumption of an infinite amount of forecasters is unrealistic, it illustrates an understanding of the well-known pooling result. Over time, consensus forecasts are more accurate than individual forecasts since the error in the noisy information is mitigated. Hence, $\overline{F}_{th}$ is expected to be a better forecast than F_{ith}. This well-known result prompts third-party users of forecasts, such as firms, voters, and policymakers, to collect information from multiple sources. The same advice should also hold for the forecasters themselves. Going back to the expression of the prior knowledge in the forecaster's updating function, it was assumed above that $\mu = F_{ith+1}$. However, given that F_{ith+1} at horizon h is public knowledge, and that we established that $\overline{F}_{th}$ is expected to be a better forecast than F_{ith}, the optimal prior should be $\mu^* = \overline{F}_{th+1}$. Hence, the optimal prior should then be a weighted average of the own lagged forecast as well as the lagged forecast of the competitors.

3.2 Estimation

Nordhaus (1987) presents a useful concept of weak and strong forecast efficiency to assess the efficiency in forecasts. Define *strong efficiency* to be if $\mathbb{E}_{ith}[(x_t - F_{ith})^2 | I_{ith}]$ is minimized, where I_{ith} is all information available to individual i, h periods before the realization of x_t. From the minimization argument, it also holds that the forecasts are unbiased: $\mathbb{E}_{ith}[x_t - F_{ith}|I_{ith}] =$

a forecast made in December 2020 for the 2020 outcome will have $t = 2020$ and $h = 1$, assuming a monthly setting.

[10] In order to simplify the model, it is assumed that there is no public signal. This assumption does not impact the main takeaway from the example.

0. The overall intuition of efficiency is simple: available information should not explain the forecast errors.[11] Nordhaus (1987) then argues that a weaker concept, where instead of the full information set I_{ith}, only the forecaster's own past forecast should be included. Define *weak efficiency* to be if $\mathbb{E}_{ith}[(x_t - F_{ith})^2 | F_{ith+1}]$ is minimized, where F_{ith+1} is the forecast produced in the previous period. Weak efficiency is nested in strong efficiency and is, of course, a much weaker concept. However, an appealing feature about the weaker test is that it can be hard to define the information set I_{ith} in the case of strong efficiency. Claiming that the past forecast was in the information set is not that strong of an assumption.

This paper continues to assume, as in Nordhaus (1987), that forecasters know their own past forecasts and further add the assumption that they know the past (average) forecasts of their competitors, $Z_{ith+1} = \frac{1}{n-1} \sum_{j \neq i} F_{jth+1}$.[12] Hence, Z_{ith+1} is the leave-out mean for each forecaster i. This results in an augmented weak efficiency definition of minimizing $\mathbb{E}_{ith}[(x_t - F_{ith})^2 | F_{ith+1}, Z_{ith+1}]$, which implies that $\mathbb{E}_{ith}[x_t - F_{ith} | F_{ith+1}, Z_{ith+1}] = 0$. Estimating

$$x_t - F_{ith} = \beta_0 + \beta_1 F_{ith+1} + \beta_2 Z_{ith+1} + \varepsilon_{ith} \tag{1}$$

should then yield $\beta_0 = \beta_1 = \beta_2 = 0$ to not rule out efficiency.[13] $\beta_2 > 0$ can be interpreted as underuse of competitors' information, while $\beta_2 < 0$ can be interpreted as overuse of competitors' information. To understand this interpretation, decompose equation (1) in two equations

$$x_t = \gamma_0 + \gamma_1 F_{ith+1} + \gamma_2 Z_{ith+1} + \epsilon_{ith} \tag{2}$$

$$F_{ith} = \alpha_0 + \alpha_1 F_{ith+1} + \alpha_2 Z_{ith+1} + \xi_{ith}, \tag{3}$$

where equation (2) tells us the ex-post optimal behavior that would have minimized the forecast error and equation (3) tells us the actual behavior. The difference between the coefficients in equation (2) and the coefficients in (3) is the coefficients found in equation (1) such that $\beta = \gamma - \alpha$.

The estimation rests on a couple of timing assumptions. Consider especially the estimation of equation (3). First of all, there can be no reverse causality since a forecast released in a certain month (the dependent variable) can neither impact the own released forecast in the previous

[11] This has a direct link to the rational expectations literature (Muth, 1961) and the Efficient Markets Hypothesis literature (Fama, 1970).

[12] Due to the unbalanced nature of the data, the mean of others—those that are available in each survey month—is used instead of each separate forecast. n is the number of forecasters, including forecaster i.

[13] $\beta_0 = \beta_1 = \beta_2 = 0$ is necessary for strong and weak efficiency but not sufficient for claiming strong efficiency.

month nor the forecast released in the previous month by competitors. Second, the timing of the estimation also rules out, by assumption, the existence of simultaneity. The assumption is that there is no information in period h that simultaneously impacts the dependent variable (the current month's forecast) and the regressors (the previous month's forecasts). Hence, the error term in period h is assumed to be uncorrelated with any of the regressors (dating to the last month) as it, by assumption, captures new information received in the current month (horizon h). In contrast, the regressors capture all previously known information.

It is also assumed that a forecast released today cannot impact the forecast released today by another forecaster. The rationale behind this assumption is that forecasters are competitors and are assumed to not share information before releasing it to stakeholders and the public. This assumption rules out the possibility that F_{ith+1} directly impacts Z_{ith+1} and vice versa. The impact rather goes via one lag, such that F_{ith+1} impacts Z_{ith} and Z_{ith+1} impacts F_{ith}.[14] Furthermore, the lagged forecast represents all the previously known information since the past forecasts should summarize all past information. F_{ith+1} and Z_{ith+1} are hence going to be collinear since they are determined, partially, by the same information set. The information set that determines F_{ith+1} is $\{I_{th+1}, S_{ith+1}\}$ while the information set that determines Z_{ith+1} is $\{I_{th+1}, \sum_{j \neq i} S_{jth+1}\}$, where I_{th+1} is all publicly available information at the time of forecast release—including all previously released forecasts—and S_{ith+1} is the private signal of forecaster i.[15] Hence, the information sets differ by the private information in time $th+1$, resulting in multicollinearity, with estimations likely suffering from inflated standard errors.

Due to the information structure of annual average growth rates and the assumption that forecasters have access to each other's lagged estimates, the forecasts are likely to be serially correlated and correlated across firms within the same survey month. To tackle these issues, inference is based on two-way cluster robust standard errors (Cameron et al., 2011; Cameron and Miller, 2015) at the forecaster (firm) level and the country-by-survey month level.

[14] Hence, there should not be any concern regarding reflection problems (Manski, 1993).

[15] Note that a regression of the form $F_{ith} = \alpha + \beta Z_{ith} + error_{ith}$ would likely yield $\beta = 1$ due to the fact that all forecasts today are determined to a large extent by the same information set. However, as Angrist (2014) points out, most often, this cannot be interpreted as a causal relationship. Individual outcomes are often highly correlated with group average outcomes without a causal peer-effect. The same would hold in this paper, since, by assumption, forecaster i at horizon h does not have access to Z_{ith}. They do, however, have access to Z_{ith+1}.

4 Results

Table 3 reports the results from estimating equations (1), (2) and (3). In columns (1) and (4), the dependent variable is the forecast error, and under the hypothesis that forecasters use competitor's forecasts efficiently, the expectation is for these coefficients to be zero. A positive coefficient on *Lagged Others* is interpreted as underuse of the competitors' forecasts, while a negative coefficient is interpreted as overuse of competitors' forecasts. Hence, the coefficient -0.777 in column (1) is interpreted as overuse of the lagged own forecast. In contrast, the coefficient 0.735 is interpreted as underuse of the lagged forecast from others.

Columns (2) and (3) report the estimated coefficients from equations (2) and (3) using GDP growth forecasts while columns (5) and (6) correspond to the inflation forecasts. The coefficients in columns (2) and (5) can be interpreted as the suggested optimal behavior—that is, what they should have done. The estimates suggest that it would be optimal to put very high attention (0.926) on *Lagged Others* while putting very low attention on the own *Lagged Forecast* (≈ 0.05). The estimates suggest that the mean of others' forecasts should get around 17 times higher weight than the own lagged forecast. Columns (3) and (6) report the estimated actual behavior where it can be seen that the own *Lagged Forecast* gets a very high weight (≈ 0.8), while *Lagged Others* gets a relatively small weight (≈ 0.2). The difference between columns (2) and (3) and between (5) and (6) provide the error found in columns (1) and (4), respectively. For example, for *Lagged Others*, for inflation, $0.926 - 0.219 = 0.707$, which indicates underuse of forecasts from competitors. Given the results in Table 3, the forecasts appear to be too persistent, and forecasters do not seem to internalize the importance of pooling information from others.

Table 3: Decomposing the Efficiency Error

	GDP growth			Inflation		
Dep. Variable:	(1) Forecast Error	(2) Outcome	(3) Forecast	(4) Forecast Error	(5) Outcome	(6) Forecast
Lagged Forecast	-0.777***	0.054	0.831***	-0.738***	0.055**	0.792***
	(0.033)	(0.037)	(0.010)	(0.026)	(0.027)	(0.007)
Lagged Others	0.735***	0.926***	0.192***	0.707***	0.926***	0.219***
	(0.036)	(0.039)	(0.011)	(0.028)	(0.029)	(0.007)
Observations	90,920	90,920	90,920	88,390	88,390	88,390
R^2	0.055	0.499	0.968	0.100	0.655	0.963

Notes: Standard errors robust to two-way clustering at the forecaster (firm) level and the country-by-survey month level. Estimated equations are (1), (2) and (3). *,**,*** represent the 10%, 5%, 1% significance levels.

Sensitivity analysis regarding the actual behavior estimates is performed to ensure that the estimation results in Table 3 are robust. First, columns (1) and (4) present results for when forecaster (firm) fixed effects are added to equation (3). The firm fixed effects control for any time-invariant differences in general optimism or pessimism among the forecasters. It is important to note that the estimation refers to a dynamic panel when firm fixed effects are included. One concern using a dynamic panel is the so-called Nickell bias (Nickell, 1981). However, given the reasonably large number of time periods, the estimations' bias should be neglectable. The stability of the coefficients presented in Table 4 mitigate the concerns regarding potential bias in the estimates. Second, there might be some information (for example, actual raw data) that is released in a certain month and observed by everyone. To control for and capture these public signals, country-by-survey month fixed effects are included in the estimations presented in columns (2) and (4). These effects absorb all common information in each period. From the estimates in Table 4, it is clear that the estimated actual behavior is very robust.[16]

Table 4: Robustness of Actual Behavior

	GDP growth		Inflation	
	(1)	(2)	(3)	(4)
Dep. Variable:	Forecast	Forecast	Forecast	Forecast
Lagged Forecast	0.808***	0.805***	0.772***	0.771***
	(0.011)	(0.009)	(0.007)	(0.007)
Lagged Others	0.217***	0.228***	0.242***	0.258***
	(0.012)	(0.010)	(0.008)	(0.009)
Observations	90,920	90,920	88,390	88,390
R^2	0.969	0.978	0.964	0.972
Firm	✓	✓	✓	✓
Country × Survey Month		✓		✓

Notes: Standard errors robust to two-way clustering at the forecaster (firm) level and the country-by-survey month level. *,**,*** represent the 10%, 5%, 1% significance levels.

5 Heterogeneity in Actual Behavior?

The results in the previous section suggest that forecasters are inefficient in their use of competitors' forecasts. The forecasters pay too little attention to their competitors, resulting in forecasts that are too persistent. Specifically, forecasters allocate an 80 to 20 relative weight between their own and others' lagged forecasts in the forecasting process. The results presented in Table 3 are obtained using pooled data over countries, years, and forecast horizons. This

[16] It would make little sense to replicate the same exercise for optimal behavior since, in that case, the interest is in the counterfactual optimal behavior from using lagged forecasts.

section investigates potential heterogeneity in the actual behavior.

5.1 Country and Year

The first heterogeneities investigated are country and year. Forecasters might behave differently for different countries, and forecasters might have changed their behavior over time. The graphs in Figures 1 and 2 plot the estimated coefficients for the different heterogeneities. The estimation is done by restricting the sample selection to each country and year, respectively. The solid lines correspond to the pooled estimates from Table 3 and act as reference points. From the graphs, there is no evidence of a significant difference in the behavior between countries, and the behavior is stable over time.

(a) GDP growth (b) Inflation

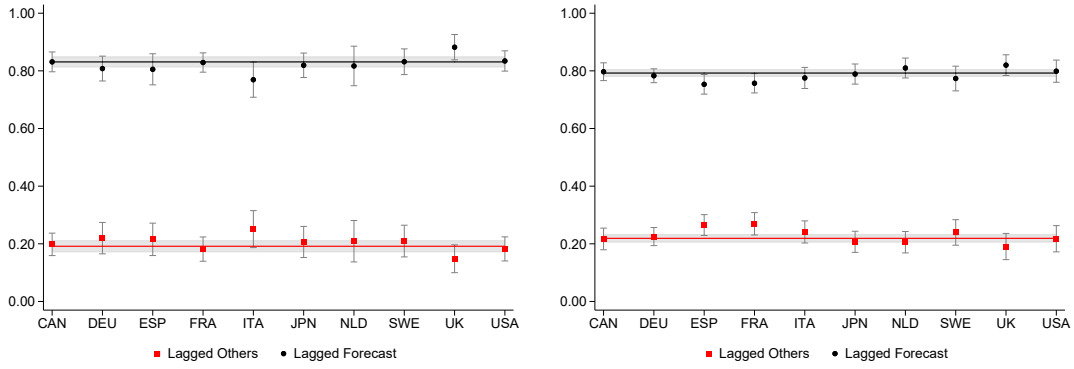


Figure 1: Country Analysis

Notes: Solid lines refer to the estimates in Table 3. The 95% confidence intervals are constructed using standard errors robust to two-way clustering at the forecaster and the country-by-survey month level.

(a) GDP growth (b) Inflation

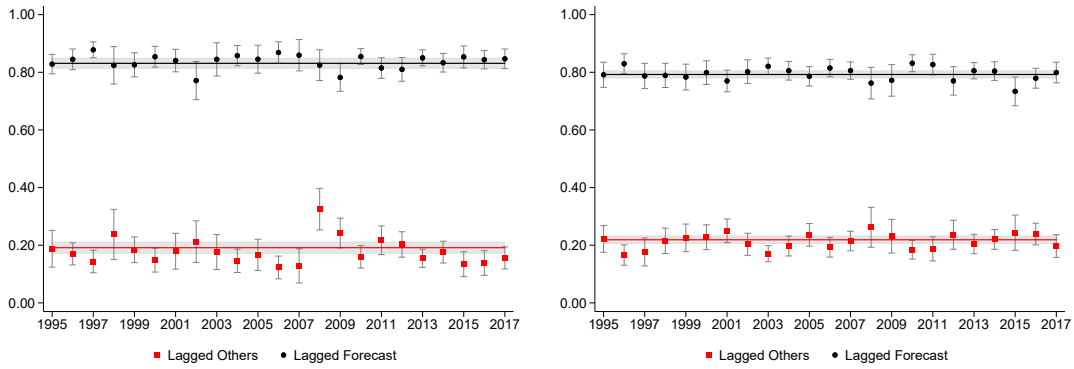


Figure 2: Year Analysis

Notes: Solid lines refer to the estimates in Table 3. The 95% confidence intervals are constructed using standard errors robust to two-way clustering at the forecaster and the country-by-survey month level.

5.2 Variable

From Table 3, the behavior of forecasters is similar using either GDP growth or inflation forecasts. The obtained dataset from Consensus Economics contains many more variables than GDP growth and inflation, allowing for an easy extension to also consider the actual behavior regarding additional variables. Equation (3) is estimated using subsamples based on several other variables.[17] Tables 5 and 6 present the results and confirm that the behavior is very consistent since forecasters for all variables put high weight on their own lagged forecast and low weight on their competitors' lagged forecasts. The *Lagged Forecast* has a coefficient of around 0.8, while *Lagged Others* receives a coefficient of around 0.2.[18]

Table 5: Variable Heterogeneity, Part 1

	Consumpt.	Wholesale	Invest.	Equip. Invest.	Indust. Prod.	Manuf. Prod.	Mining Prod.	Corp. Profits
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable:	Forecast	Forecast	Forecast	Forecast	Forecast	Forecast	Forecast	Forecast
Lagged Forecast	0.82***	0.79***	0.82***	0.79***	0.76***	0.82***	0.85***	0.81***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Lagged Others	0.19***	0.24***	0.21***	0.25***	0.27***	0.20***	0.20***	0.23***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	90,368	8,020	66,795	20,123	54,559	13,971	1,970	21,957
R^2	0.96	0.92	0.95	0.93	0.95	0.94	0.87	0.90

Notes: Standard errors robust to two-way clustering at the forecaster (firm) level and the country-by-survey month level. Estimated equations are (3). *,**,*** represent the 10%, 5%, 1% significance levels.

Table 6: Variable Heterogeneity, Part 2

	Empl.	Unempl.	Wages	New Car	Car Sales	Housing Starts	Current Account	Producer Prices
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable:	Forecast	Forecast	Forecast	Forecast	Forecast	Forecast	Forecast	Forecast
Lagged Forecast	0.81***	0.79***	0.85***	0.93***	0.79***	0.82***	0.67***	0.81***
	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.03)	(0.07)	(0.01)
Lagged Others	0.19***	0.21***	0.15***	0.02	0.22***	0.18***	0.33***	0.23***
	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.03)	(0.06)	(0.01)
Observations	6,320	69,425	48,653	1,595	10,541	24,100	75,068	36,438
R^2	0.94	1.00	0.97	0.91	0.98	1.00	0.95	0.92

Notes: Standard errors robust to two-way clustering at the forecaster (firm) level and the country-by-survey month level. Estimated equations are (3). *,**,*** represent the 10%, 5%, 1% significance levels.

[17] Some variables refer to levels instead of annual growth rates.

[18] The variables differ in terms of sample size. The difference is due to several factors. First, fewer forecasters in each country do forecasts for many of these variables. Some of these variables are also only observed in a subset of the countries, and other are only present during a short period of the sample. Many of them change definitions over time.

5.3 Horizon

The results in Table 3 also pool over the forecast horizon. The main specification estimates an average behavior over horizons 1 to 23 months. However, results from Lahiri and Sheng (2008) suggest that there is heterogeneity in forecast behavior over horizons. As before, the main specification of the actual behavior is estimated using a sample selection for each horizon.

The results are presented as dots in Figure 3. Again, the solid lines correspond to the pooled results from Table 3. From Figure 3, it is evident that, on longer horizons, the weight on *Lagged Others* is small and the weight on the own *Lagged Forecast* is high. As the horizon decreases, the weights converge and end up being equal at horizon one. Hence, as the horizon decreases, forecasters behave more optimally, but still far from optimal, since optimal would be a high weight on *Lagged Others*. The estimated optimal weight on the own *Lagged Forecast* is around 0.05 and hence far from the actual behavior on both long horizons and short horizons, even though the error is smaller on shorter horizons.

The fact that forecasters behave differently at different horizons might suggest some strategic decisions at play. It might be that alternative motives are more pronounced far away from the realization of the outcome. As forecasters then approach the realization of the outcome, they switch to concerns regarding the size of the forecast error. This reasoning is extended further in Section 7.

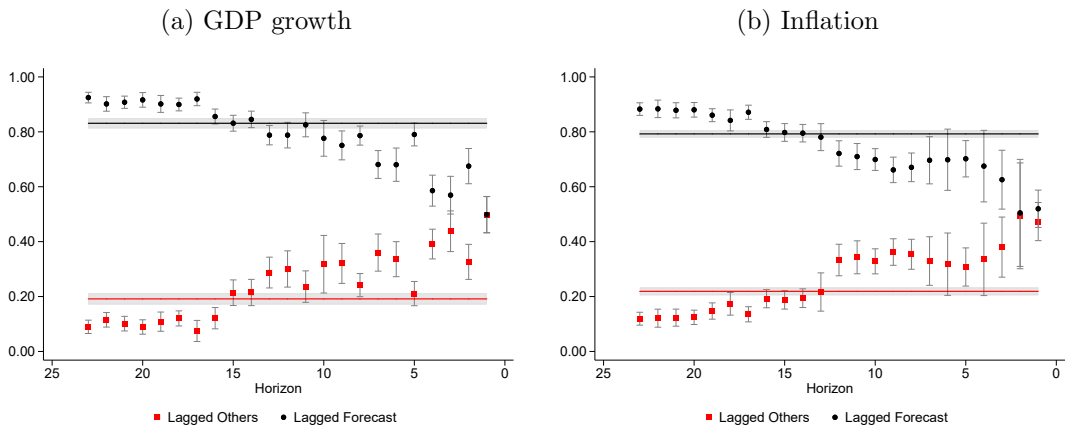


Figure 3: Horizon Analysis

Notes: Solid lines refer to the estimates in Table 3. The 95% confidence intervals are constructed using standard errors robust to two-way clustering at the forecaster and the country-by-survey month level.

6 Baseline Model to Understand Optimal Behavior

The empirical results suggest inefficient behavior among forecasters. Specifically, the results suggest that it is optimal to put a relatively high weight on the lagged forecast of others (the mean of the competitors) compared to the forecasters' own lagged forecast. The decomposition performed in this paper tells us that, on average, the optimal ratio between others' forecasts and own forecasts should be around 17.

This section presents a baseline model to further help us understand these empirical predictions of the suggested optimal behavior. Hence, the model is used to ensure that the empirical prediction of optimal behavior is reasonable. Two versions of the model will be considered: one that has fully informed forecasters who perfectly filter out all the signals and one with forecasters who follow a simplified rule of thumb. The rule of thumb will be such that forecasters will utilize the lagged actual forecasts as reduced form aggregates, instead of filtering out signals. Hence, the rule of thumb model is more in line with the empirical approach and the simple framework presented in Section 3.1.

6.1 A Model with Fully Informed Forecasters

Consider a three-period model where nature draws x at the dawn of time. Assume that we have n forecasters, indexed by i, where all try to forecast x. Each forecaster will produce a forecast in each of the three periods such that they have three attempts at predicting x.[19] At the end of time, the true value is revealed, and each forecaster faces the loss

$$L_i = (x - F_{i3})^2 + (x - F_{i2})^2 + (x - F_{i1})^2, \tag{4}$$

where F_{ih} is forecaster i's forecast of x at horizon h (h periods before the realization of x). Each forecaster minimizes the loss with respect to F_{i3}, F_{i2} and F_{i1}, where from the first-order conditions we have that $F_{ih} = \mathbb{E}_{ih}[x]$. The forecasters have a common prior about x that follows the true distribution $x \sim N(\mu, \sigma_x^2)$. In each period, each forecaster receives a private noisy signal $S_{ih} = x + \epsilon_{ih}$, where $\epsilon_{ih} \sim N(0, \sigma_{\epsilon_h}^2)$ is a random error. The level of noise $\sigma_{\epsilon_h}^2$ is assumed to be the same for everybody.

In period one, i.e., at horizon 3, each forecaster receives the private signal and is called upon

[19] Note that the model considers just one realization of x such that the t indexation from the empirical exercise is omitted.

to declare a forecast. The best each forecaster can do is to forecast the expected value given the information set (containing the prior and the signal). Hence, the forecast is

$$F_{i3} = \mathbb{E}_{i3}[x|\mu, S_{i3}] = \frac{p_\mu}{p_\mu + p_{s_3}}\mu + \frac{p_{s_3}}{p_\mu + p_{s_3}}S_{i3}, \tag{5}$$

where we have used the definition $p_v \equiv \frac{1}{\mathbb{V}[v]}$, such that p is a precision parameter defined as the inverse of the variance. As a result, $p_\mu = \frac{1}{\sigma_x^2}$ and $p_{s_3} = \frac{1}{\sigma_{\epsilon_3}^2}$.[20] To simplify the notation, rewrite the forecast as

$$F_{i3} = w_{\mu|3}\mu + w_{s_{3|3}}S_{i3}, \tag{6}$$

such that w is a relative weight where the weights sum to one.[21] At horizon 2, each forecaster receives a new private signal and observes the horizon 3 forecasts of their competitors. The forecaster filters out the signals from the competitors and produces a new forecast:[22]

$$F_{i2} = w_{\mu|2}\mu + w_{s_{3|2}}S_{i3} + \sum_{j\neq i} w_{s_{3|2}}S_{j3} + w_{s_{2|2}}S_{i2}. \tag{7}$$

Following the horizon 2 structure, we will have the horizon 1 forecast

$$F_{i1} = w_{\mu|1}\mu + \sum_{q=2}^{3} w_{s_{q|1}}S_{iq} + \sum_{j\neq i}\sum_{q=2}^{3} w_{s_{q|1}}S_{jq} + w_{s_{1|1}}S_{i1}, \tag{8}$$

where we see that a forecast is a weighted average of the prior, the own lagged signals, the lagged signals of others, and the own new private signal.[23]

Allowing for an arbitrary number of horizons (H) and outcomes (t), we can rewrite any forecast in a compressed form as

$$F_{ith} = w_{\mu|h}\mu + \sum_{q=h+1}^{H}\sum_{j=1}^{n} w_{s_{q|h}}S_{jtq} + w_{s_{h|h}}S_{ith}. \tag{9}$$

Given the information structure under fixed target forecasts, the information set will grow over horizons such that, as the forecast horizon becomes shorter, the error is expected to be smaller.[24]

[20] Equation (5) follows from Bayesian updating.
[21] Hence, $w_{\mu|3}$ is the prior weight at horizon 3 and $w_{s_{3|3}}$ is the weight on the horizon 3 signal at horizon 3.
[22] The forecaster is assumed to have full knowledge about the prior structure and the precision in all signals. Hence, fully informed forecasters have the ability to filter out the signal from the observed forecasts.
[23] The weights follow the structure presented in Equation (5).
[24] See, for example, Andersson et al. (2017) for an analysis of the expected size of forecast errors at different

From (9), a forecast in the model is a weighted average of all signals and the prior. The model does not provide an expression of the form that is used in the empirical exercises,

$$F_{ith} = \beta_0 + \beta_1 F_{ith+1} + \beta_2 Z_{ith+1} + \varepsilon_{ith}, \tag{10}$$

where the forecast directly depends on the lagged forecasts and not the signals. However, the model does provide a prediction regarding the relationship between β_1 and β_2 in the reduced form estimation.

Consider a world with only three forecasters (a, b and c) such that $n = 3$. Then write the horizon 2 forecast from equation (7) for forecaster a in terms of lagged forecast instead of lagged signals,

$$F_{a2} = \omega_\mu \mu + \omega_{f3} F_{a3} + \omega_{z3} Z_{a3} + \omega_{s2} S_{i2}, \tag{11}$$

where $Z_{a3} = \frac{1}{2}\big(F_{b3} + F_{c3}\big)$. To replicate the individual signal weights, it is then needed that $\omega_{f3} = \frac{p_\mu + p_{s_3}}{P_2}$, $\omega_{z3} = 2\frac{p_\mu + p_{s_3}}{P_2}$, $\omega_{s2} = \frac{p_{s_2}}{P_2}$ and $\omega_\mu = -2\frac{p_\mu}{P_2}$, where $P_2 = p_\mu + 3p_{s_3} + p_{s_2}$. Hence, $\omega_{z3} = (n-1)\omega_{f3}$ is such that the relationship between the own lagged forecast and the lagged forecast of the others is proportional in the number of forecasters.[25]

As a result of this, we should expect $\beta_2 = (n-1)\beta_1$ in the estimation such that the relative weight between the own forecast and that of others should be dependent on n (the number of forecasters). In the empirical estimation, the optimal behavior was estimated to have a ratio of 17 between others' forecasts and the own forecasts, implying in the model that we have around 18 forecasters. This lines up very well with the firm number statistics in Table 1 of an average of 19.8 forecasters.

6.2 Forecasters with Rule of Thumb

The baseline model above assumed highly advanced forecasters with full information about the signal structure. This section presents an alternative model that imposes some form of bounded ability to ensure that we do not ask too much from our forecasters. Follow the same setup as in the fully informed model, but assume now that forecasters cannot, or at least do not, filter

horizons.

[25] Expanding this to more horizons would require including all lagged forecasts and not only the latest. The per-horizon relationship between the own forecasts and those of others, however, stay the same. The relationship always follows $\omega_{zh} = (n-1)\omega_{fh}$.

out the individual signals. Hence, forecasters use a simplifying rule of thumb where they make a weighted average of the lagged forecasts (own and others') and the new private signal.

Assume the following structure. At horizon 3, each forecaster receives the private signal and declares a forecast just as before:

$$F_{i3} = \mathbb{E}_{i3}[x|F_{i3}, Z_{i3}, S_{i2}] = \frac{p_\mu}{p_\mu + p_{s_3}}\mu + \frac{p_{s_3}}{p_\mu + p_{s_3}}S_{i3}. \tag{12}$$

In period two, each forecaster receives a new private signal and observes the horizon 3 consensus forecast of the competitors (Z_{i3}). The information set is now the own lagged forecast, the lagged consensus forecast of others, and the new private signal. Hence, this setup treats Z_{i3} as an aggregate signal such that the Horizon 2 forecast will be

$$F_{i2} = \frac{p_{f_3}}{p_{f_3} + p_{z_3} + p_{s_2}}F_{i3} + \frac{p_{z_3}}{p_{f_3} + p_{z_3} + p_{s_2}}Z_{i3} + \frac{p_{s_2}}{p_{f_3} + p_{z_3} + p_{s_2}}S_{i2}, \tag{13}$$

where, again, we simplify the notation to

$$F_{i2} = w_{f2}F_{i3} + w_{z2}Z_{i3} + w_{s2}S_{i2}. \tag{14}$$

The horizon 1 forecast will then be

$$F_{i1} = w_{f1}F_{i2} + w_{z1}Z_{i2} + w_{s1}S_{i1}. \tag{15}$$

The optimal relationship between the weights will, again, depend on the number of forecasters such that $w_{zh} = (n-1)w_{fh}$.[26] This is the same prediction as in the fully informed model.

However, if a forecaster follows this process by not filtering out the actual signals, the forecast will be less efficient compared to the fully informed forecast. This is because a forecast using aggregate information such as lagged forecasts will have nested information where old information is weighted too much compared to new private information.[27]

[26] As before, it is assumed that everyone has the same signal precision in the private information. It is also assumed that each forecaster assumes that everyone is equally good.

[27] A simple example to consider is the usage of a common prior. If all forecasters have the same prior and then put a weight on each other's forecasts, the prior will be weighted too much relative to private information. This prior problem is noted by Kim et al. (2001), where the authors show analytically that consensus forecasts are inefficient since too much weight is attributed to analysts' common information relative to their private information. This is further highlighted in Crowe (2010), where the author argues that consensus forecasts are inefficient due to over-weighting older information at the expense of new private information.

7 Extending the Model to Understand the Actual Behavior

From the empirical exercise, two facts are of interest. First, the forecasters pay little attention to their competitors. Second, the forecasters change their behavior over forecast horizons. The results show that the actual behavior of forecasters is such that they put high weight on their own lagged forecast, and this weight is different at different horizons. The empirical exercise also provides relative use between others' versus their own lagged forecasts. This is of good use since the model framework provides a prediction for that relationship. The estimated actual behavior is a ratio (others' over own) of 0.25, while the optimal was estimated to be 17. This section aims to extend the model to understand why forecasters seem to behave inefficiently compared to the baseline model. It could be that the assumed loss function of forecasters is incorrect, resulting in incorrect inefficiency judgment.[28]

The fact that forecasters overuse their own signal might suggest that they are overconfident in their own ability. This is reasonable to believe since this behavior bias often is called "the mother of all biases." Overconfidence is well studied in psychology, and De Bondt and Thaler (1995) state that: "perhaps the most robust finding in the psychology of judgment is that people are overconfident."[29] However, it is unlikely that overconfidence is the only explanation since this would require unreasonable levels of overconfidence. A simple back of the envelope example can provide some intuition about the level.

Assume a forecaster is overconfident about their own signal. The beliefs are such that their own precision is perceived to be κp_i, where $\kappa > 1$.[30] Continue to assume that forecasters have correct beliefs about others and that they correctly can filter out the signals from their competitors. This will result in a higher relative weight on their own signals (and forecasts).

To understand the level of κ, assume 20 forecasters. A non-overconfident forecaster will put $1/20 = 0.05$ relative weight on the own signal and $19/20 = 0.95$ on the others' combined. If the forecaster has a κ of 2, reflecting that the own ability is twice that of competitors, the weight structure changes to 0.095 on the own forecast and 0.905 on the others' forecasts. With $\kappa = 10$, the resulting weights are 0.34 on the own and 0.66 on the others', and $\kappa = 100$ yield 0.84 on the own forecast and 0.16 on the others'. Hence, with many forecasters, κ needs to be very

[28] However, foreasters are still going to be inefficient in minimizing forecast errors, but this might not be their only objective.

[29] See, for example, Kahneman and Tversky (1972) and Tversky and Kahneman (1974) for more on this cognitive bias.

[30] See, for example, Daniel et al. (1998) for similar modeling of overconfidence in the securities markets.

high to replicate the relationships found in the data. Judging this number as high or not is, of course, not straightforward. However, consider the following example; assume the true value of 2 to be predicted. Assume a true signal that offers a 95% confidence interval equal to 2 ± 1. A "hundred times" overconfident forecaster would judge the own signal to be in the 95% confidence interval of 2 ± 0.01. In contrast, each of the competitors is judged to have a signal in the 2 ± 1 interval. Overconfidence is then most likely not a reasonable (sole) explanation since the κ levels needed seem to be a bit extreme. It is also the case that overconfidence does not provide any strategic trade-off that produces horizon dynamics as those found in the data.

To explain both the general underuse of competitors' forecasts and the horizon dynamics in this behavior, the introduction of a revision cost is considered (see, e.g., Coibion and Gorodnichenko, 2015; Ehrbeck and Waldmann, 1996). The idea behind the revision cost is as follows. Assume that a forecaster releases a first estimate of the GDP growth in 2018 of 1 percent. After one month, the forecaster observes the forecasts of its competitors and realizes that the consensus view of the others is a GDP growth of 3 percent. When the forecaster releases the next estimate, it would be optimal to revise up to the 3 percent mark, subject to the new private signal. This would imply a forecast revision of 2 percentage points. A revision that large would need substantial explanation since no big news has been uncovered. Telling the public that the forecast last month was wrong after seeing the forecasts of others would probably be a big hit to the forecaster's credibility.[31] What the forecaster could do instead is to slowly revise toward the others to preserve credibility and hide the big revision in many small steps.[32] It is important to note that the revision cost can be a product of forecasters' responses to the demand side of the forecasting business. Hence, the end-user of the forecasts might demand, or reward, consistency and thus punish inconsistent forecasters (Nordhaus, 1987).

Introducing a revision cost gives rise to a trade-off between loss from forecast errors and loss from revisions, resulting in a higher degree of forecast smoothing. However, in order to explain the horizon heterogeneity, the cost function needs to be horizon discounted, so that the trade-off also introduces a horizon trade-off. The following section outlines the technical details of a revision cost extension.

[31] This follows some of the reasoning in Ehrbeck and Waldmann (1996), who present a similar mechanism in a repeated game where there is a trade-off between minimizing forecast errors and looking good before the outcome is realized.

[32] See Nordhaus (1987) for similar reasoning.

7.1 Revision Cost to Preserve Credibility

In the baseline model, it was assumed that forecasters' sole purpose was to minimize the forecast error. This section adds an explicit revision cost to the loss function. The baseline model consisted of three periods (horizons). This section extends the model to 23 horizons to understand the horizon dynamics.[33] Assume that the forecaster faces the following minimization problem:[34]

$$\min_{F_{ih}} \quad \sum_{h=1}^{23} g(h)(\mathbb{E}_{ih}[x] - F_{ih})^2 + r(h)(F_{ih} - F_{ih+1})^2. \tag{16}$$

At each horizon, the forecaster faces a trade-off between the minimization of the forecast error and the forecast revision.[35] The forecast error is horizon discounted by the function $g(h)$ such that $g(h) > g(h+1)$ to account for the fact that it is harder to forecast at longer horizons.[36] The revision cost is also horizon discounted such that $r(h) > r(h+1)$. The relation between $g(h)$ and $r(h)$ then determines the relative cost between forecast errors and revisions. From the first-order condition of the minimization problem, we have that a forecast at horizon h is given by

$$\begin{aligned} F_{ih} = \frac{g(h)}{g(h) + r(h) + r(h-1)} \mathbb{E}_{ih}[x] + \frac{r(h)}{g(h) + r(h) + r(h-1)} F_{ih+1} \\ + \frac{r(h-1)}{g(h) + r(h) + r(h-1)} \mathbb{E}_{ih}[F_{ih-1}]. \end{aligned} \tag{17}$$

Note that $\mathbb{E}_{ih}[x]$ is formed as in the baseline model, where it is a weighted average of the prior and the signals. Solving backwards, the horizon 1 forecast is[37]

$$F_{i1} = \frac{g(1)}{g(1) + r(1)} \mathbb{E}_{i1}[x] + \frac{r(1)}{g(1) + r(1)} F_{i2}. \tag{18}$$

Solving the horizon 2 forecasts then yields

$$F_{i2} = \frac{g(2)}{g(2) + r(2) + r(1)} \mathbb{E}_{i2}[x] + \frac{r(2)}{g(2) + r(2) + r(1)} F_{i3} + \frac{r(1)}{g(2) + r(2) + r(1)} \mathbb{E}_{i2}[F_{i1}]. \tag{19}$$

[33] It is important to note that the predictions from the baseline model do not depend on the number of horizons.

[34] This is similar to a loss function presented in, for example, Coibion and Gorodnichenko (2015), with the difference of having horizon dependent trade-off weights.

[35] In the first period, there is no lagged forecast. Hence, the solution in the first period is trivial and follows $F_{i24} = \mathbb{E}_{i24}[x]$.

[36] See, for example, Andersson et al. (2017) for an analysis of the expected size of forecast errors at different horizons. In the baseline model, the discounting does not have any implications at all since it would drop out from the first-order conditions.

[37] Note that by assumption $F_{i0} = x$ and $r(0) = 0$ since the outcome is known at horizon zero.

Substitute (18) into (19) and rearrange to get[38]

$$F_{i2} = \frac{g(2) + r(1)m_1}{g(2) + r(1)m_1 + r(2)} \mathbb{E}_{i2}[x] + \frac{r(2)}{g(2) + r(1)m_1 + r(2)} F_{i3}, \tag{20}$$

where $m_1 = \frac{g(1)}{g(1)+r(1)}$. Simplify the notation to

$$F_{i2} = m_2 \, \mathbb{E}_{i2}[x] + (1 - m_2) F_{i3}. \tag{21}$$

Then, for a general horizon h forecast, we have that

$$F_{ih} = \frac{g(h) + r(h-1)m_{h-1}}{g(h) + r(h-1)m_{h-1} + r(h)} \mathbb{E}_{ih}[x] + \frac{r(h)}{g(h) + r(h-1)m_{h-1} + r(h)} F_{ih+1}, \tag{22}$$

and simplify any expression of this form as

$$F_{ih} = m_h \, \mathbb{E}_{ih}[x] + (1 - m_h) F_{ih+1}, \tag{23}$$

where $(1 - m_h)$ can be seen as a smoothing parameter. In the empirical results, it was found that m_h was decreasing in h such that $m_1 > m_2 > ... > m_{23}$. Hence, the two functions $g(h)$ and $r(h)$ need to be decreasing in h at different rates. Specifically, $g(h)$ needs to decrease faster than $r(h)$ as the horizon increases. The intuition is that forecast errors need to be more costly in relation to revisions on short relative to long horizons. It then becomes apparent that it is the relative size, and not the actual size, of $g(h)$ and $r(h)$ that is important, such that one could simplify and define $g(h) = 1 - r(h)$. Different functions of $g(h)$ (and/or $r(h)$) could then provide the dynamics needed to explain the empirical results.

8 Concluding Remarks

From the empirical exercise, forecasters are deemed inefficient since they underuse information from their competitors given an assumption of pure forecast error minimization. The empirical results are robust for several heterogeneities, such as country, year, and variable. Moreover, a horizon decomposition revealed that forecasters behave differently at different forecast horizons. Specifically, they pay more attention to their competitors as the forecast horizon decreases and hence behave more efficiently as they approach the realization of the outcome, but still far from

[38] Note that $\mathbb{E}_{i2}[x] \neq \mathbb{E}_{i1}[x]$, $\mathbb{E}_{i2}[\mathbb{E}_{i1}[x]] = \mathbb{E}_{i2}[x]$ and that $\mathbb{E}_{i2}[F_{i2}] = F_{i2}$.

an optimal level.

A simple model of noisy and private information provides a framework to analyze the behavior of forecasters. Under the assumption of forecast error minimization, the model and the empirical exercise predicted that it is optimal for the individual forecaster to pay close attention to what the other forecasters do. This result originates from the fact that it is assumed that each forecaster has private information such that there is a benefit from pooling information from many forecasters. The result regarding the importance of pooling is no big surprise and is in line with standard findings in the forecasting literature. Perhaps more surprising is the finding that forecasters do not incorporate this in their revision process.

Previous studies have found that forecasters appear to put high weight on their own lagged information such that they are too slow to incorporate new information. In the empirical part of this paper, the focus is on the estimation of the relative use of own lagged forecasts and the lagged forecasts of competitors. The results show that forecasters put four times higher relative weight on their own lagged forecast compared to the forecasts of competitors. In comparison, optimal is rather 17 times higher weight on the competitors' forecasts compared to the own lagged forecast.

In order to explain the observed (seemingly) inefficient behavior, the baseline model is extended. First, it is unlikely that the cognitive bias of overconfidence is the only source of the observed behavior. For overconfidence to explain the observed behavior, forecasters need to think that they are around 100 times better than their competitors. It is, of course, possible (and likely) that some degree of overconfidence is present as a partial explanation to the findings. Second, introducing a revision cost can only explain the observed horizon dynamics under the assumption that the horizon discounting is relatively different for the cost from forecast errors and the cost from forecast revisions.

It is well known that the average of multiple forecasts often outperforms individual estimates when compared over time. This well-known result prompts third-party users of forecasts, such as consumers, investors, voters, and policymakers, to collect information from multiple sources. However, the results in this paper show that professional forecasters themselves do not sufficiently internalize this well-established result.

References

Ager, P., Kappler, M., and Osterloh, S. (2009). The Accuracy and Efficiency of the Consensus Forecasts: A Further Application and Extension of the Pooled Approach. *International Journal of Forecasting*, 25(1):167–181.

Andersson, M. K., Aranki, T., and Reslow, A. (2017). Adjusting for information content when comparing forecast performance. *Journal of Forecasting*, 36(7):784–794.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Banerjee, A. (1992). A Simple Model of Herd Behaviour. *The Quarterly Journal of Economics*, 107(3):797–817.

Batchelor, R. and Dua, P. (1992). Conservatism and consensus-seeking among economic forecasters. *Journal of Forecasting*, 11(2):169–181.

Bernhardt, D., Campello, M., and Kutsoati, E. (2006). Who Herds? *Journal of Financial Economics*, 80(3):657–675.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020). Overreaction in Macroeconomic Expectations. Forthcoming, American Economic Review.

Broer, T. and Kohlhas, A. (2018). Forecaster (Mis-)Behavior. CEPR Discussion Papers 12898, C.E.P.R. Discussion Papers.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

Chen, Q. and Jiang, W. (2006). Analysts' Weighting of Private and Public Information. *Review of Financial Studies*, 19(1):319–355.

Cipullo, D. and Reslow, A. (2019). Biased Forecast to Affect Voting Decisions? The "Brexit" Case. Working Paper 364, Sveriges Riksbank.

Coibion, O. and Gorodnichenko, Y. (2015). Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review*, 105(8):2644–78.

Coibion, O., Gorodnichenko, Y., and Kumar, S. (2018). How Do Firms Form Their Expectations? New Survey Evidence. *American Economic Review*, 108(9):2671–2713.

Crowe, C. W. (2010). Consensus Forecasts and Inefficient Information Aggregation. Technical report, International Monetary Fund Working Paper 10/178.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor Psychology and Security Market Under- and Overreactions. *The Journal of Finance*, 53(6):1839–1885.

De Bondt, W. F. and Thaler, R. H. (1995). Chapter 13 Financial Decision-making in Markets and Firms: A Behavioral Perspective. In *Finance*, volume 9 of *Handbooks in Operations Research and Management Science*, pages 385–410. Elsevier.

Deschamps, B. and Ioannidis, C. (2013). Can Rational Stubbornness Explain Forecast Biases? *Journal of Economic Behavior and Organization*, 92:141–151.

Dovern, J., Fritsche, U., Loungani, P., and Tamirisa, N. (2015). Information rigidities: Comparing average and individual forecasts for a large international panel. *International Journal of Forecasting*, 31(1):144–154.

Ehrbeck, T. and Waldmann, R. (1996). Why Are Professional Forecasters Biased? Agency versus Behavioral Explanations. *The Quarterly Journal of Economics*, 111(1):21–40.

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

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Gallo, G. M., Granger, C. W. J., and Jeon, Y. (2002). Copycats and Common Swings: The Impact of the Use of Forecasts in Information Sets. *IMF Staff Papers*, 49(1):4–21.

Hendry, D. F. and Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31.

Isiklar, G., Lahiri, K., and Loungani, P. (2006). How quickly do forecasters incorporate news? Evidence from cross-country surveys. *Journal of Applied Econometrics*, 21(6):703–725.

Kahneman, D. and Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, 3:430–454.

Kim, O., Lim, S. C., and Shaw, K. W. (2001). The Inefficiency of the Mean Analyst Forecast as a Summary Forecast of Earnings. *Journal of Accounting Research*, 39(2):329–336.

Lahiri, K. and Sheng, X. (2008). Evolution of Forecast Disagreement in a Bayesian Learning Model. *Journal of Econometrics*, 144(2):325–340.

Lamont, O. A. (2002). Macroeconomic Forecasts and Microeconomic Forecasters. *Journal of Economic Behavior & Organization*, 48(3):265–280.

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

Lucas, R. (1973). Some International Evidence on Output-Inflation Tradeoffs. *American Economic Review*, 63(3):326–34.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Marinovic, I., Ottaviani, M., and Sørensen, P. N. (2013). Chapter 12 – Forecasters' Objectives and Strategies. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 690–720. Elsevier.

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

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Nordhaus, W. D. (1987). Forecasting Efficiency: Concepts and Applications. *The Review of Economics and Statistics*, 69(4):667–674.

Ottaviani, M. and Sørensen, P. N. (2006). The Strategy of Professional Forecasting. *Journal of Financial Economics*, 81(2):441–466.

Rülke, J.-C., Silgoner, M., and Wörz, J. (2016). Herding behavior of business cycle forecasters. *International Journal of Forecasting*, 32(1):23–33.

Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*. Penguin Publishing Group.

Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3):665–690.

Tanaka, M., Bloom, N., David, J. M., and Koga, M. (2019). Firm performance and macro forecast accuracy. *Journal of Monetary Economics*.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131.

Woodford, M. (2001). Imperfect Common Knowledge and the Effects of Monetary Policy. Technical report, National Bureau of Economic Research.

Zarnowitz, V. (1984). The Accuracy of Individual and Group Forecasts from Business Outlook Surveys. *Journal of Forecasting*, 3(1):11–26.

A Appendix

A.1 Tables

Table A1: Outcome Statistics

		GDP growth			Inflation		
Country	Obs.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Canada	14,872	7,457	2.2	1.4	7,415	1.8	0.6
France	18,600	9,331	1.3	1.1	9,269	1.3	0.8
Germany	27,481	13,739	1.4	1.7	13,742	1.4	0.6
Italy	14,336	7,228	0.5	1.6	7,108	1.9	1.1
Japan	18,450	9,409	1.0	2.1	9,041	0.2	0.9
Netherlands	9,586	4,959	1.6	1.8	4,627	1.8	0.8
Spain	13,927	7,011	2.0	2.0	6,916	2.2	1.4
Sweden	12,885	6,469	2.3	2.0	6,416	1.1	1.1
UK	23,663	12,539	1.8	1.6	11,124	2.8	0.9
USA	25,510	12,778	2.5	1.4	12,732	2.2	1.0
Total	179,310	90,920	1.7	1.8	88,390	1.7	1.2

Notes: Actual outcome data for annual average GDP growth and inflation between 1995 and 2018. Each outcome is repeated for each corresponding forecast observation in the monthly survey by Consensus Economics between 1995 and 2017.

Table A2: Error Statistics

		GDP growth			Inflation		
Country	Obs.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Canada	14,872	7,457	-0.2	1.1	7,415	-0.1	0.6
France	18,600	9,331	-0.3	0.8	9,269	-0.1	0.6
Germany	27,481	13,739	-0.2	1.3	13,742	-0.2	0.6
Italy	14,336	7,228	-0.6	1.1	7,108	-0.0	0.6
Japan	18,450	9,409	-0.2	1.8	9,041	-0.1	0.6
Netherlands	9,586	4,959	-0.1	1.3	4,627	-0.0	0.5
Spain	13,927	7,011	0.0	1.0	6,916	0.0	0.9
Sweden	12,885	6,469	-0.1	1.7	6,416	-0.4	0.9
UK	23,663	12,539	-0.2	1.1	11,124	0.1	0.8
USA	25,510	12,778	-0.0	1.0	12,732	-0.0	0.8
Total	179,310	90,920	-0.2	1.3	88,390	-0.1	0.7

Notes: Data refer to ex-post forecast errors where errors are defined as the outcome minus the forecast. In each month, forecasters are asked to forecast the current and next year's annual average growth rate. This results in 24 potential forecast origins (and horizons) for each year's outcome.