International Journal of Forecasting 39 (2023) 772–790

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Differing behaviours of forecasters of UK GDP growth

Nigel Meade [a,*], Ciaran Driver [b]

[a] The Business School, Imperial College London, SW7 2AZ, UK
[b] School of Finance and Management, SOAS, University of London, WC1H 0XG, UK

ARTICLE INFO

Keywords:
GDP forecasts
Fixed event forecasting
Herding
Cluster analysis
Granger causality

ABSTRACT

The literature suggests that the dispersion of agents' forecasts of an event flows from heterogeneity of beliefs and models. Using a data set of fixed event point forecasts of UK GDP growth by a panel of independent forecasters published by HM Treasury, we investigate three questions concerning this dispersion: (a) Are agent's beliefs randomly distributed or do agents fall into groups with similar beliefs? (b) as agents revise their forecasts, what roles are played by their previous and consensus forecasts? and (c) is an agent's private information of persistent value? We find that agents fall into four clusters, a large majority, a few pessimists, and two idiosyncratic agents. Our proposed model of forecast revisions shows agents are influenced positively by a change in the consensus forecast and negatively influenced by the previous distance of their forecast from the consensus. We show that the forecasts of a minority of agents significantly lead the consensus.

1. Introduction

An important summary statistic for all economies is the percentage growth in gross domestic product per annum (% GDP growth). In the UK, a compendium of forecasts of real GDP growth in year T by around 50 different agents is published monthly by HM Treasury sequentially from early in year $T-1$ to the beginning of year $T+1$. The preparation of forecasts of GDP growth in year T over decreasing horizons is an example of fixed event forecasting. At each horizon, the available forecasts can be summarised by measures of location and dispersion, these measures represent a consensus forecast and the level of disagreement about it. The literature suggests that the dispersion of forecasts of an event is partly due to the heterogeneity of agents' beliefs and models; however, the nature of this heterogeneity is unclear. The literature suggests that several forces are contributing to forecast dispersion: the heterogeneous prior beliefs of the agents; the flow of noisy public and private information; and behavioural effects present in an agent's utility or cost functions. It has proved difficult to disentangle these influences in analysing the pattern of dispersion over different horizons and over economic cycles. For example, there seems no firm view in the literature as to whether dispersion reflects idiosyncratic variation or persistent deviations between sets of forecasters.

One influence that has generated considerable controversy is whether herding behaviour—or its opposite, anti-herding—is a major contributor to dispersion. Herding involves an agent tending to adjust his/her forecast towards the consensus and is thus a possible contributor to a reduction in dispersion. However, a responsiveness to the consensus is a broad definition of herding; behavioural explanations usually make it conditional on the agent's utility being negative in deviation from the consensus. The fuzziness of the term makes for difficulties in testing and in judging which commonly used tests can adequately identify herding (Clements, 2018).

Using our data set of agents' forecasts of UK GDP growth for 1997 to 2019, we measure and explore the differences in the behaviour of these agents, considering these three research questions.

* Corresponding author.
E-mail address: n.meade@imperial.ac.uk (N. Meade).

https://doi.org/10.1016/j.ijforecast.2022.02.005
0169-2070/

Q(a) What is the structure of the heterogeneity of beliefs? Are agent's beliefs randomly distributed or do agents fall into groups with similar beliefs. For example, do some agents persistently exhibit relative optimism or pessimism?

Q(b) Is there evidence of herding behaviour as agents revise their forecasts? What is the extent of the roles played by: the previous consensus forecast; the agent's previous forecast; and the change in consensus forecast? As the horizon shortens, how can the decreasing dispersion be explained?

Q(c) Is the private information available to some agents of persistent value? Do some agents tend to lead or to follow the consensus? Are some clusters of forecasters persistently better than others?

The cross-section variation in forecasts is the primary focus of this article, and we only give some secondary consideration to the issue of forecast accuracy. A recurrent theme in our study is that, although the majority of agents exhibit similar behaviour in each analysis, the behaviour of a small number of agents differs from the majority in a distinctive way.

The structure of the article is as follows: We review the literature on the behaviour of fixed event forecasters and provide the basis for our research questions in Section 2. Relevant approaches to identifying different behaviours are identified, and some hypotheses are proposed in Section 3. In Section 4, we describe the data set, and in Section 5, we make an exploratory analysis of Q(a), looking at the dispersion of forecasts. Section 6 deals with Q(b) considering different models of the forecast revision process. The final question Q(c) concerning whether some agents lead or follow the consensus and the persistence of performance is considered in Section 7. We summarise our findings, draw conclusions, and make suggestions for further work in Section 8.

2. Literature review and development of research questions

Fixed event forecasting is the preparation of a forecast which is subsequently revised over time as the event approaches. After the agent makes the first forecast, revisions are made periodically in the light of new information. For the value of a particular target variable at a specified period, say % GDP growth in year T, a group of agents will publish forecasts with a long horizon, say 24 months. As time passes, agents will revise their forecasts. The set of forecasts with the same target variable and the same horizon is summarised by a consensus forecast. Although consensus can mean a generally agreed value, here, it simply denotes the group mean of the forecasts, about which there is some disagreement.

The heterogeneity of forecasts is receiving increased attention in the literature. The strong rational expectations model tends now to be only regarded as a benchmark. The weak version of it that includes private information, uncorrelated across time and agents, has also been challenged by alternative explanations such as time-invariant prior beliefs, which are possibly more important for long-horizon forecasts.

Increasingly, the interest is in "looking at how [forecasts] are derived instead of simply assuming they are rational" (Weale, 2021). The topic has been boosted by the evidence of persistent heterogeneity across agents (see, for example, Patton and Timmermann (2010); by new theories on how agent disagreement magnifies fluctuations in the economy (Guzman & Joseph E Stiglitz, 2021); and by an extensive debate over whether heterogeneity is a good proxy for uncertainty (see Gallo et al., 2002; Rich & Tracy, 2018; Zarnowitz & Lambros, 1987).

Nevertheless, there has been only limited success in explaining forecast heterogeneity itself.

The literature on the heterogeneity in forecasts discusses differences in forecast accuracy and differences across forecasters. Forecast differences—the primary focus of this article—have been attributed in the literature to different causes:

(i) Private information obtained randomly at any point in time.
(ii) Agents targeting distinct definitions or vintages of the variable of interest.
(iii) Distinct clusters of agents that are differentiated by mind sets, background, identity, acculturation, or competence in accessing or interpreting information.
(iv) Behavioural biases resulting from agents' motivations being other than forecast accuracy, given their competence level.

Models based on (i) generally imply that individual forecasts are efficient, a feature investigated and rejected in Clements (2021) with the finding that approximately half of forecasters do not generate rational forecasts, given their information sets, and that there are unaccounted for persistent patterns of difference across forecasters. It has been argued that "rational inattention" (RI) may cause variation in information across agents, as for example in the sticky information model of Mankiw and Reis (2002) and Clements (2012) where adjustment is discrete rather than continuous. However, RI models have been argued to be most useful when there are no clear limits to what information is available (Maćkowiak et al., 2021) and thus may be less applicable to a well-defined and frequently forecast variable such as national GDP.[1] In an analysis of the behaviour of a large international panel of GDP forecasts, Dovern (2013) found that 40%–50% of forecasters revise each month. Failure to adjust forecasts, resulting in persistent error autocorrelation, is likely to reflect caution in interpreting a structural change; a dramatic illustration of this phenomenon is wage forecasting by the Monetary Policy Committee 2014–20 (Blanchflower, 2019, p.66).

Explanation (ii) could reflect differences in client time horizons which might be at the short end for financial traders and the longer end for policy-makers (Anesti et al.,

[1] Clements (2021) investigates whether the rejection of RI assumptions are not only significant but indicate substantially important and persistent deviation from rationality of some forecasters, with mixed findings.

2020).[2] However, these influences seem too weak to account for heterogeneity in that they affect at most a small minority of forecasters. One study based on US Society of Professional Forecasters data suggested "little evidence against the null that the first estimate is being targeted" (Clements, 2019).

Explanation (iii) is difficult to fully investigate because of the many potential attributes that might influence forecaster behaviour. For example, a banker may be conditioned to favour a particular model or information source and may have access to private client information. Forecasters may have relational contacts with one of the formal large-scale macro-models. An agent, or agents, may follow the forecasts of a forecaster who is perceived to be superior. We use the term "clustering" to describe the time-invariant heterogeneity resulting from the behaviour described in case (iii).

Explanation (iv) implies there are differences in loss functions across agents. For example, an agent may fear to "stand out" as in the model of Scharfstein and Stein (1990). A fear of standing out is a manifestation of herding behaviour, a desire to stand out is termed "anti-herding". Although there are many studies of herding behaviour in forecasters, in the context of finance (e.g., corporate earnings) and in macro-economics (e.g., inflation and GDP), there is no agreed single definition of "herding" in these studies. In Appendix A, we group definitions from ten different studies under four headings: non-specific to forecasting; adjusting an individual forecast towards a consensus view; adjusting an individual forecast towards a consensus view to the detriment of accuracy; and herding as the converse of boldness (referring to forecasts rather than forecasters). An uncontentious definition of herding is given by Bewley and Fiebig (2002): "…the tendency to produce a range of forecasts which is narrower than that which would likely be observed if the forecasts were produced on a strictly independent basis because a forecaster takes the previous consensus mean into account". The arithmetic mean forecast—the consensus—has been found to be more accurate than most individual forecasters (Spiro, 1989), and as a result, the most recently available consensus data may exert an influence on individual forecasts. We explore this hypothesis in our research question Q(b) and investigate the extent to which it might be identified as herding.

The literature broadly agrees that there is heterogeneity of agents' views leading to the disagreement between their forecasts, but little is said about the structure of the heterogeneity. In our research question Q(a), we investigate whether the observed disagreement flows from randomly distributed heterogeneous beliefs or whether there are clusters of similar beliefs.

In Gallo et al. (2002)'s model of imitative forecasting behaviour, an agent's next forecast is essentially a weighted average of their past forecast and the currently available (lagged) consensus. Imitative behaviour is said to occur if the sum of the weights is greater than unity. They conclude that this imitative behaviour may cause forecasting accuracy to deteriorate and forecasts to converge on the wrong value.[3] Testing has also been performed for anti-herding where forecasters seek variety rather than convergence by over-weighting their previous forecasts and under-weighted known forecasts of others (Batchelor & Dua, 1992). The S statistic developed by Bernhardt et al. (2006) has been widely used to detect anti-herding behaviour in forecasts of several market prices.[4]

Much of the evidence for herding and anti-herding is undermined by Clements (2018) who analyses three tests for herding or anti-herding and interprets his own empirical finding. The first two tests (labelled T_1 and T_2) are modifications of one proposed by Gallo et al. (2002), assuming an earlier awareness of the consensus view. The third test (T_3) follows the assumptions of Bernhardt et al. (2006). Using U.S. quarterly survey forecasts of inflation and output growth data 1981–2013, T_1 detected very little evidence of herding behaviour, whereas "the related approach" T_2 suggested that herding was the predominant behaviour, and T_3 suggested that anti-herding predominated. However, Clements concludes from his analysis of the test properties that all tests will reject the no-herding null when there is noisy information and "that the empirical pattern of rejections that we observe across the different tests is consistent with the pattern predicted by differences among forecasters primarily reflecting idiosyncratic errors or reflecting noisy information. Either way, the evidence for (anti-)herding is far from compelling".

In our research question Q(b), we investigate whether there is evidence of behaviour by the macro-economic forecasters in our data set in respect of the influence exerted by the consensus. First, we use the approach of Gallo et al. (2002) and Clements (2018). Second, by modelling the revision of an agent's forecast, we seek an alternative view of the influence of the consensus. In analysing revisions, we take account of how the parameters of the model may change with the length of the forecast horizon. We also revisit the stylised fact that dispersion among agents' forecasts is highest at longer horizons and lower at shorter horizons. Amador and Weill (2010) propose a model of the effect of information releases on the heterogeneity between agents as the horizons shorten. Patton and Timmermann (2010) find that dispersion between the forecasts of both GDP and inflation by private sector forecasters contributing to Consensus Economics Inc. follows this behaviour. They suggest that this dispersion persists through time because of heterogeneity in prior beliefs and models, and they further find that heterogeneity is

[2] Published estimates of UK GDP growth are subject to revision in the short term in measurement and in the longer term in both measurement and definition. For background, see Symons (2001), Office for National Statistics (2020) and Blastland (2019), pp 87–100.

[3] Interestingly, Gallo et al. (2002) do not use the term 'herding' but use 'shrinking to the mean', possibly making a link to shrinkage estimation where an estimate is adjusted by other information such as a group mean (e.g., see Copas, 1983).

[4] See work on: oil prices (Pierdzioch et al., 2010); S&P 500 stock prices (Pierdzioch & Rülke, 2012); inflation in South Africa (Pierdzioch et al., 2016); metal prices (Pierdzioch et al., 2013); foreign exchange rates (Frenkel et al., 2020; Pierdzioch & Stadtmann, 2011; Tsuchiya, 2015).

greater during recessions. Using a subset of the same data, Patton and Timmermann (2011) model the monthly updating of forecasts using an unobserved components model, and they confirm that agents are hampered by measurement errors in real-time GDP growth. In a similar vein, Angeletos et al. (2020) found that agent's updates of forecasts of unemployment and inflation initially under-react to shocks and then, after a delay, over-react. In the third part of research question Q(b), we investigate how our model of an agent's forecast revision may capture dispersion decreasing with horizon.

In research question Q(c), we look for evidence of forecasters leading or following the consensus. Evidence of agents following the consensus may indicate rational inattention because of high search costs, but if persistent, it may reflect distinct levels of resource capacity that would be associated with reduced forecast dispersion.

3. Approaches to modelling forecaster's behaviour for Q(b)

Although Q(a) analysis is mostly exploratory, Q(b) involves testing explicit models. In this section, we establish some hypotheses that underpin our answers to Q(b) in respect of the revisions to fixed event forecasts. Our notation is as follows, a forecast of GDP growth in year T by forecaster i is denoted as $F_{i,T,h}$, where h is the forecast horizon (the number of months before the publication of the official figure). The consensus of forecasts of GDP growth at horizon h is measured as $C_{T,h} = \overline{F}_{T,h}$, and the dispersion around the consensus is the standard deviation $S_{T,h} = SD\left(F_{T,h}\right)$.

3.1. The proposed model of forecast revision

Here, we propose a model of the revision of forecasts for the same target year at k month intervals. The revision is $R_{i,T,h} = \left(F_{i,T,h} - F_{i,T,h+k}\right)$. We consider the revision process from each forecaster's viewpoint, so the consensus used refers to the other forecasters. Thus, for forecaster i, the consensus of forecasts with horizon j, excluding that of forecaster i, is $C_{i,T,j} = mean\left(F^{*}_{r,T,j}, r \neq i\right)$, where $F^{*}_{r,T,j}$ is the latest available forecast by forecaster and r is for year T with a horizon of j. To make this revision, we suggest that the forecaster considers both:

- The new information that has become available since the last forecast was published; this effect is encapsulated by the change in consensus view during the revision—$\left(C_{i,T,h+1} - C_{i,T,h+k}\right)$.
- The distance of the previous forecast from the consensus of other forecasters at that time $\left(F_{i,T,h+k} - C_{i,T,h+k}\right)$.

Our model for the forecast revision is:

$$R_{i,T,h} = \theta_0 + \theta_1\left(C_{i,T,h+1} - C_{i,T,h+k}\right) + \theta_2\left(F_{i,T,h+k} - C_{i,T,h+k}\right) + \xi_{i,h} \quad (1)$$

The intuition underlying this formulation is that the forecast revision consists of a change in location driven by new information, $\left(C_{i,T,h+1} - C_{i,T,h+k}\right)$, and a change in dispersion about the consensus represented by the previous distance from the consensus, $\left(F_{i,T,h+k} - C_{i,T,h+k}\right)$. Three further hypotheses are considered.

H_1: $\theta_0 = 0$; this hypothesis implies no upward or downward bias in the revision process.

H_2: $\theta_1 = 0$; this hypothesis implies that the forecaster ignores the new information that causes the consensus among other forecasters to change, and a positive value for θ_1 suggests the forecaster's adjustment is in the same direction as other forecasters.

H_3: $\theta_2 = 0$; this hypothesis implies that the previous distance from the consensus does not influence the revision. A negative value for θ_2 suggests a pressure to be closer to the consensus. According to our definition (ii), the condition $\theta_2 < 0$ implies herding. Whether the underlying change in the agent's information set is due to extra private information or a desire to be closer to the consensus is likely to be unknowable.

4. Description of the data set

The source of our data is "Forecasts for the UK economy: a comparison of independent forecasts" published by HM Treasury,[5] and we consider monthly issues from October 1997 to July 2019. Each issue contains one or more tables of forecasts of UK % GDP growth and its components for the previous, current, or following year. The original data comprise 20,822 published forecasts from the 262 monthly issues. Once repeated publications of the same forecast are removed, and there are 13,225 forecasts. In some cases, forecasters were active for an interval and then withdrew from publishing forecasts. In other cases, when a name changes because of a takeover or merger, the identity of the forecaster is considered to be the same. There are 77 forecasters in the analysis, of which 52 are classified by HM Treasury as City forecasters and 25 are non-City forecasters. We classify the forecasters further into 33 banks, 11 other asset managers, 24 brokers/consultancies, and 9 public/professional institutions.

Here, we illustrate the evolution of the forecasts for % GDP growth for a given year and the reducing level of disagreement about the consensus forecast as the horizon shortens. We define the horizon of the forecasts as the interval in months between the publication date (forecast origin) of the forecast for year T and June $(T+1)$ when the "actual" % GDP growth is first published. The last data were collected in July 2019; thus, the full range of the subscripts of the forecast $F_{i,T,h}$ in the data set is $i \in \{1, \ldots, 77\}$, $T \in \{1997, \ldots, 2019\}$ and $h \in \{5, \ldots, 32\}$. However, because of forecasters not being active for the whole period and not publishing forecast revisions every month, there are many absent values of $F_{i,T,h}$. To have a reasonable basis for analysis, we consider forecasts with origins ranging from June $(T-1)$ to December (T), equivalent to horizons of 24 months to 6 months, for all years where this range of horizons is

[5] These data are point forecasts, the Bank of England Survey of External Forecasters collects point and density forecasts of GDP growth, see Boero et al. (2008).

available, i.e. $\{F_{i,T,h}: T = 1998, 2018; h = 24, 6\}$. Within this data set, for each horizon for each target year, the mean number of forecasts published is 24.7.

Fig. 1 shows time series of forecasts with horizons decreasing from 24 to 6 months before the publication of the actual % GDP growth. The mean forecast is shown with the dispersion represented by the mean +/- two standard deviations, and the actual % GDP growth (represented by the first quarterly estimate PN2) is shown in the bottom plot.

We choose two years, 1999 and 2009, where the response of the mean forecasts to fresh information is particularly evident, and we show histograms of the forecasts at decreasing horizons in Fig. 2. At a 24-month horizon, the mean forecast for 1999 is near to 2%. For horizons of 18 and 12 months, extra information decreases the mean forecast to below 1%, before reverting to near 2% at 6-month horizon. In 2009, the histograms reflect the growing pessimism as the mean forecast decreases as the horizon shortens. Extra information moves the mean forecast consistently downwards from 1.33% in June 2008 to −1.67% in December 2008 to −3.68% in June 2009 and finally to −4.53% in December 2009.

It is apparent in both Figs. 1 and 2 that the variation among forecasts published observed at 6-month intervals tends to decrease as the forecast horizon shortens. This decrease in the level of disagreement is made clear in Fig. 3, where the standard deviations about the mean forecast are plotted by horizon. The median standard deviation drops more and more sharply as the horizon shortens. It falls by 9% between 24 and 18 months, by 39% between 18 and 12 months, and by 70% from 12 to 6 months.

5. Exploratory analysis of Q(a): the dispersion of agent forecasts over different horizons

We consider Q(a): *What is the structure of the heterogeneity of beliefs? Are agent's beliefs randomly distributed or do agents fall into groups with similar beliefs*. The heterogeneous beliefs of agents are manifested by the dispersion of their forecasts. Here, we investigate the similarities and differences in individual agent behaviour. To facilitate the comparison of behaviours over time, with changing values of the mean forecast and the level of disagreement about it, the forecasts are standardised to t-values. The standardised data are first summarised, and then, cluster analysis is used to investigate similarities in agents' behaviours.

5.1. Summary of forecasts as t-values

We consider forecasts over six-month intervals published at horizons of 24, 18, 12, and 6 months, and if a forecaster does not publish at that horizon, a forecast from one or two months previous is used (if available). The purpose of this choice of forecast frequency is to better isolate deliberate changes in forecasts, and many forecast revisions over higher frequencies, one or two months, are either very small or zero.

For each forecaster, i, for each year, T, we consider horizons of 24, 18, 12, and 6 months, the set of forecasts made is identified as $(F_{i,T,24}, F_{i,T,18}, F_{i,T,12}, F_{i,T,6})$. Each forecast is converted to a t statistic to capture its position in relation to the consensus and take into account the dispersion around it, $t_{i,T,h} = \frac{(F_{i,T,h} - C_{T,h})}{S_{T,h}}$. The t values observed for each forecaster are summarised in Table 1. The summary describes the 56 forecasters for which there are at least 15 forecasts meeting these criteria. The mean t value indicates the average location of this forecaster in the distribution of forecasts; the standard deviation of the forecaster indicates how volatile the location of the forecaster's forecasts is relative to the consensus. A mean t value close to zero with a low standard deviation indicates a consistent consensus forecaster; a negative mean t value indicates a forecaster who is pessimistic relative to the consensus.

The number of forecasts available from these forecasters varies from 15 to 83, and the mean t value of the forecasters over all forecasts has an inter-quartile range of −0.18 to 0.21, indicating that, on average, the middle 50% of forecasters do not stray far from the consensus forecast. The statistics of five forecasters are chosen as examples across the spread of overall mean t values. Forecaster ID26 has a negative mean t value for all four horizons, becoming increasingly negative as the horizon shortens; the standard deviation of the t value is noticeably higher than those of the other examples. Forecasters ID73, ID38, and ID63 are on average close to the consensus for most horizons with comparatively low standard deviations. The forecasts of forecaster ID49 tend, on average, to be higher than the consensus for the 18-, 12-, and 6-month horizons; this forecaster has a noticeably higher standard deviation for all horizons. These examples demonstrate several different types of behaviour exhibited by the forecasters.

5.2. A cluster analysis

We examine the structure of the heterogeneity of agents' beliefs and models manifested by the dispersion of forecasts around the consensus. Are beliefs randomly distributed or do groups of agents hold similar beliefs leading to persistent optimism or pessimism relative to the consensus? We use cluster analysis to address Q(a). The incomplete nature of the data set allows for the construction of a distance matrix permitting the use of hierarchical clustering methods but precludes the use of a partitioning algorithm such as K-means clustering. The element of the distance matrix showing the distance between agent A and agent B is

$$D_{AB} = \frac{\sum_T \sum_j \left| t_{A,T,j} - t_{B,T,j} \right|}{n_{AB}}$$

where n_{AB} is the number of available observations of this distance. A minimum value of $n_{AB} = 10$ is enforced to ensure the distance measures are robust. Sample rows and columns of the resulting distance matrix are shown in Table 2.

Hierarchical clustering involves several arbitrary decisions including the choice of distance metric, the rule

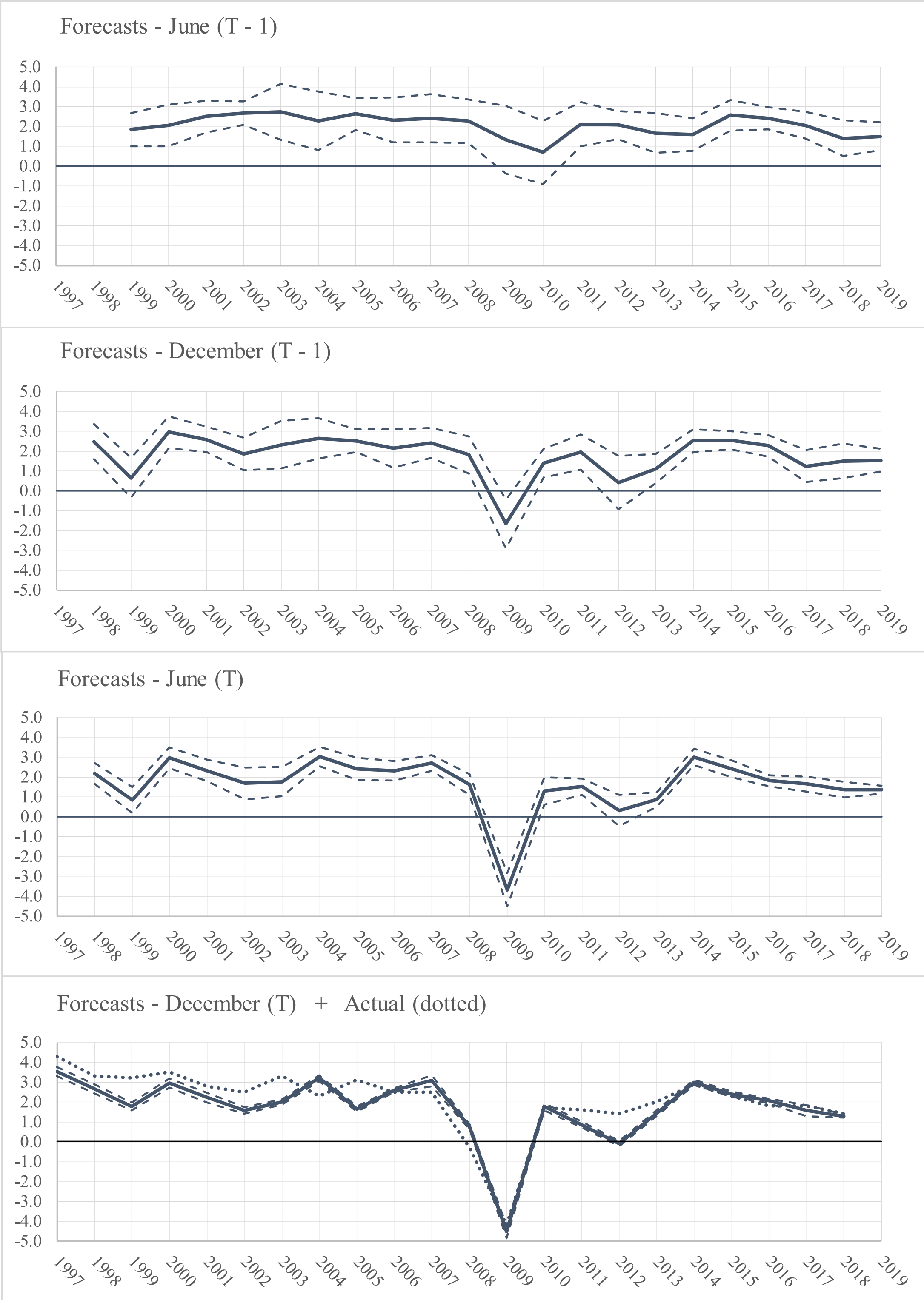

Fig. 1. Forecasts of % GDP growth published in June and December of the previous year and June and December of the target year. The dispersion of forecasts is summarised by the mean (solid line) and the mean +/−2 standard deviations (dashed). Actual values (PN2) are shown as dots in the last panel.

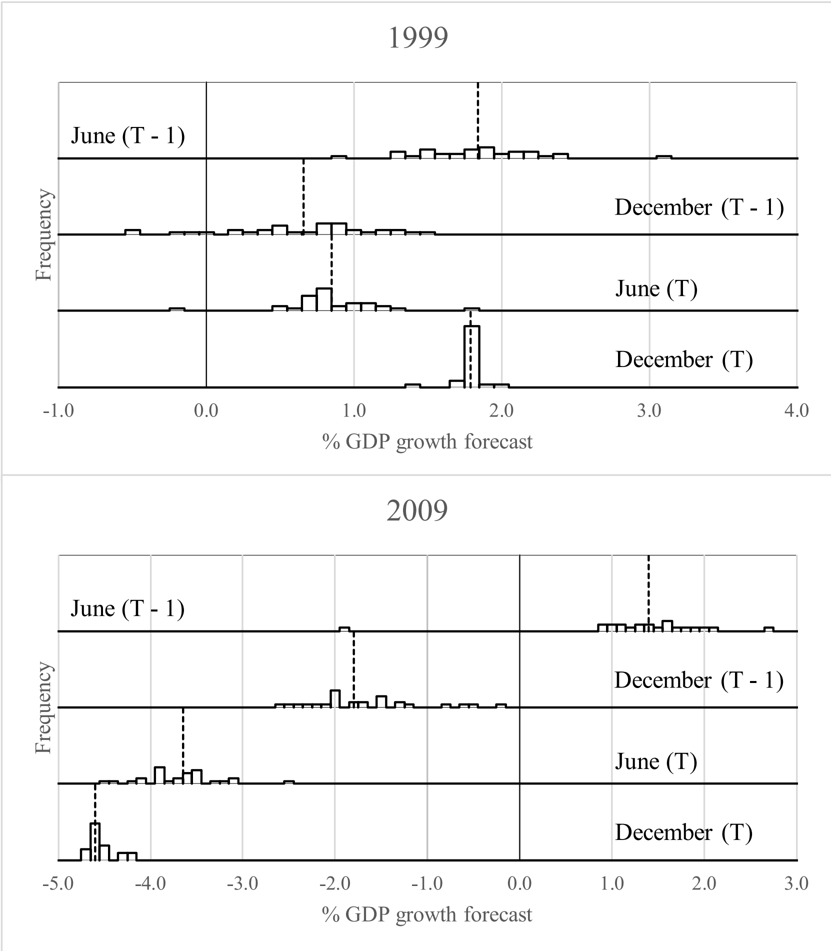

Fig. 2. Two sets of histograms of GDP growth forecasts with horizons of 24, 18, 12, and 6 months for 1999 and 2009 (the mean forecast is shown as a vertical dashed line in each case).

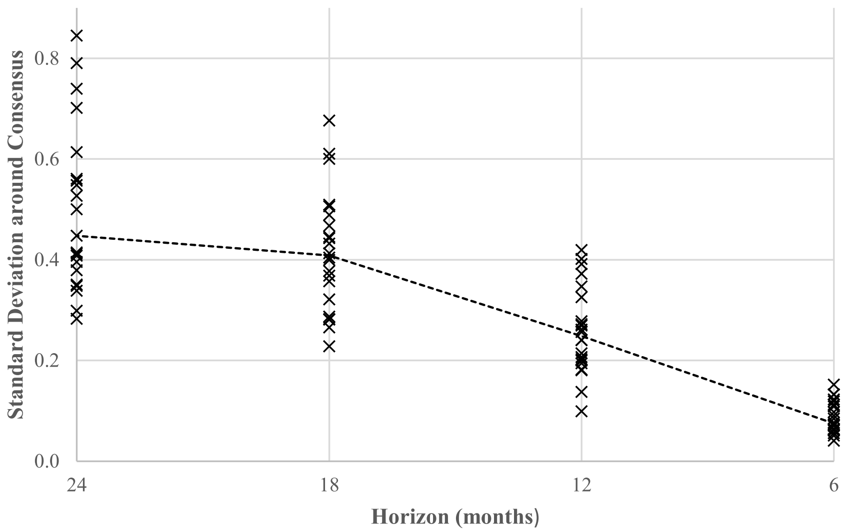

Fig. 3. The standard deviations about the consensus forecast for each year shown for horizons of 24 months decreasing to 6 months. The dotted line connects the median SD for each horizon.

Table 1
For each forecaster, a forecast is standardised by the mean and standard deviation of contemporaneous forecasts as t values, and these t values are summarised by their mean and standard deviation for horizons of 24, 18, 12, and 6 months and overall. Five example forecasters are chosen roughly equally spread according to their overall mean t value. The ranges of these statistics for 56 forecasters with at least 15 forecasts overall are summarised by their quartiles.

Horizon		Forecasters					Lower quartile	Median	Upper quartile
		ID26	ID73	ID38	ID63	ID49			
24 months	Count	21	15	14	21	19	7	14	18
	Mean	−0.31	−0.10	−0.10	0.08	−0.23	−0.22	0.04	0.19
	SD	2.01	1.12	0.57	0.46	2.10	0.61	0.81	1.00
18 months	Count	21	15	15	21	20	7	13	20
	Mean	−1.30	−0.25	−0.04	−0.01	0.57	−0.20	−0.05	0.25
	SD	2.37	0.97	0.58	0.42	1.53	0.58	0.72	1.01
12 months	Count	21	15	14	21	20	7	14	19
	Mean	−2.08	−0.45	−0.01	0.09	0.82	−0.15	0.02	0.19
	SD	2.35	0.95	0.39	0.49	1.35	0.56	0.73	0.88
6 months	Count	20	14	15	20	19	6	12	19
	Mean	−2.71	−0.09	0.04	0.27	0.74	−0.20	0.07	0.29
	SD	1.91	0.59	0.62	0.43	1.07	0.53	0.64	0.85
Overall	Count	83	59	58	83	78	28	56	78
	Mean	−1.59	−0.22	−0.03	0.11	0.48	−0.18	−0.02	0.21
	SD	2.32	0.92	0.54	0.45	1.58	0.63	0.74	0.92

chosen to measure the distance between clusters, and the final number of clusters. After some experimentation, the intuitively reasonable choice is four clusters chosen using the "average" measure of distance between clusters. The output of the hierarchical clustering is a dendrogram, where clusters are formed as the distance measure is increased. In this case, many small clusters aggregate to form the majority cluster, and a few other agents gather into a "pessimists" cluster; at the completion of the clustering process are two agents ID26 and ID49 which are distant from each other and the other clusters. The membership of these clusters is shown in Fig. 4, where the co-ordinates of each agent are $\left(SD\left(t_i\right), \overline{t_i}\right)$ (as in the lowest panel of Table 1) and the clusters are identified by different markers. The largest cluster, "majority" contains 31 agents whose forecasts are similar to each other, and they share the characteristics of mean t values close to zero and relatively low standard deviations. There are three agents in the "pessimists" cluster, and their forecasts are similar and on average lower than the majority. Agent ID26 (shown in Table 1) is an extreme case of comparative pessimism. Agent ID49 (also shown in Table 1) is optimistic and has a larger standard deviation of t values.

The cluster analysis identifies four types of behaviour. The majority group of 31 agents publish similar forecasts and thus are mainly responsible for determining the consensus forecast. A cluster of three agents tends to be more pessimistic than the consensus forecasts. Two agents ID26 and ID49 behave sufficiently idiosyncratically to warrant a cluster each.

The membership of these four clusters is cross-tabulated with both the City and non-City dichotomy and the more detailed categorisation in Table 3. The comparison between City and non-City agents is unrevealing. All the agents classified as banks or asset managers belong to the majority group, as do six of seven of the public/professional institutions. The three pessimists and ID26 are classified as brokers/research/consultancies. ID49 is a public/professional institution. The response to Q(a) is that agents' beliefs are not randomly distributed, and that the heterogeneity of agents' beliefs is due to a majority group following homogeneous beliefs, a small minority with relatively pessimistic beliefs, and two idiosyncratic agents with their own unshared beliefs.

Table 2
A sample of the entries in the distance matrix between the forecasts of the 36 agents for which sufficient data are available.

	ID01	ID04	.	ID26	.	ID76	ID77
ID01	0.00	0.99	.	2.95	.	0.79	1.06
ID04	0.99	0.00	.	2.68	.	0.93	0.85
.	.	.	.	.	.	.	.
ID26	2.95	2.68	.	0.00	.	3.04	3.46
.	.	.	.	.	.	.	.
ID76	0.79	0.93	.	3.04	.	0.00	1.03
ID77	1.06	0.85	.	3.46	.	1.03	0.00

6. Analysis of Q(b): modelling forecasts and forecast revisions

In this section, we address Q(b): *Is there evidence of herding behaviour as agents revise their forecasts? What is the extent of the roles played by: the previous consensus forecast; the agent's previous forecast; and the change in consensus forecast? As the horizon shortens, how can the decreasing dispersion be explained?* First, we estimate a literature-based linear model of the revised *forecasts* $\left(F_{i,T,18}, F_{i,T,12}, F_{i,T,6}\right)$. Second, we consider linear models of the *revision* of forecasts of GDP growth $\left(R_{i,T,18}, R_{i,T,12}, R_{i,T,6}\right)$, as in Section 3.1. The objective is to understand the pressures on the forecaster resulting in a revision. As part of this analysis, we consider the mechanism behind the decrease in the dispersion of forecasts as horizons shorten. Third, we investigate the possible negative association between levels of dispersion and recessions.

6.1. Estimation of Gallo et al.'s model

This analysis considers the linear models used by Gallo et al. (2002) described in (2) and (3); we estimate these

Table 3
A cross-classification of agents by cluster and type of institution.

	Majority	ID26	ID49	Pessimists	Total
City	21	1		1	23
Non-City	10		1	2	13
Bank	14				14
Other asset manager	5				5
Broker/research/consultancy	6	1		3	10
Public/professional institution	6		1		7
Total	31	1	1	3	36

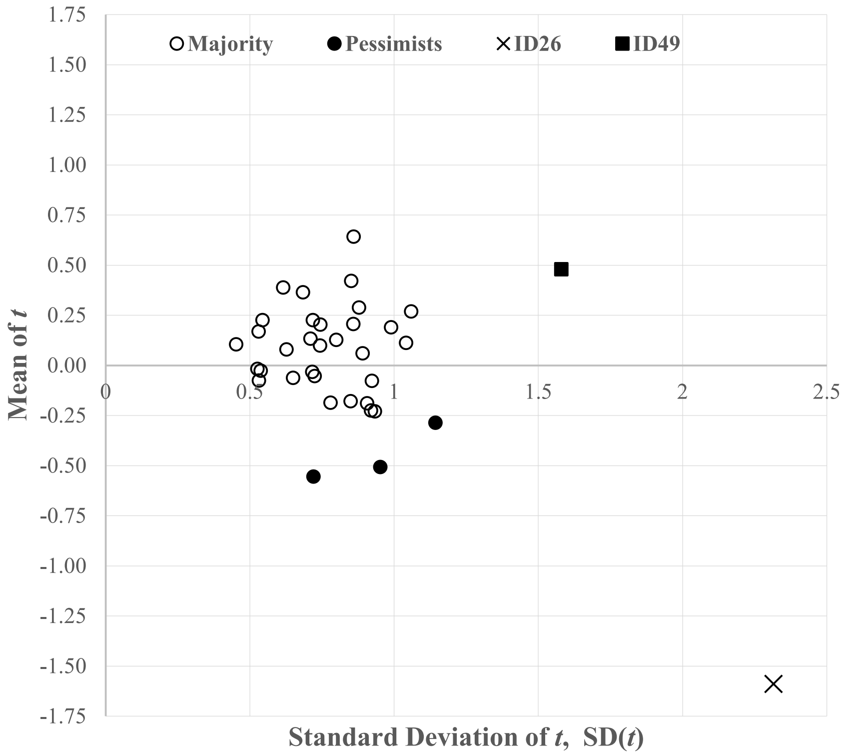

Fig. 4. A scatter plot of agents defined by the mean and standard deviation of their t values over the forecasts available. The clusters shown are derived from hierarchical "average"clustering using absolute difference in t values as the distance metric.

models for $k = 6$ and $h = 18, 12, 6$.

$$F_{i,T,h} = \alpha_0 + \alpha_1 F_{i,T,h+k} + \alpha_2 C_{T,h+k} + \alpha_3 S_{T,h+k} + u_i \tag{2}$$

From (2), we see that the coefficients α_1 and α_2 reflect the relative weighting on the forecast and the consensus forecast k periods previous. Considering the movement of forecasts towards the consensus, the authors re-parameterise their model to:

$$F_{i,T,h} - C_{T,h+k} = \alpha_0 + \alpha_1 \left(F_{i,T,h+k} - C_{T,h+k}\right) + \alpha_4 C_{T,h+k} + \alpha_3 S_{T,h+k} + u_i \tag{3}$$

where $\alpha_4 = \alpha_1 + \alpha_2 - 1$.

To contrast our model in (1) with the Gallo et al. model, we rewrite (1) as

$$F_{i,T,h} = \theta_0 + (1 + \theta_2) F_{i,T,h+k} + (-\theta_1 - \theta_2) C_{i,T,h+k} + \theta_1 C_{i,T,h+1} + \xi_{i,h}.$$

Our main innovation is the inclusion of the consensus view in the month before publication capturing the change in location of the forecasts. Gallo et al. include $\alpha_3 S_{T,h+1}$ in (2) as a control variable to account for decreasing dispersion as the horizon shortens. We bypass this issue by performing regressions for specific values of h and k and examine the effect of h on $V(\xi)$.

These models are estimated for all agents where a minimum of 15 observations are available, and the results are summarised in Table 4. The model is also fitted to the pooled forecasts made by the members of the four clusters identified in 5.2. The individual agents were ranked by the adjusted R^2 for (2), and the results for four agents roughly equally spaced across this ranking are shown; these agents all belong to the majority cluster. The estimated coefficients across all agents are summarised by their medians. The estimated coefficients from the pooled majority cluster have very similar values to the median values across all agents. As described by (2), the behaviour of individual agents varies widely, and the adjusted R^2 ranges from 0.94 to 0.23. We consider four hypotheses concerning the Gallo et al. coefficients:

H_a: $\alpha_1 = 0$; this hypothesis implies no memory of the previous forecast; the coefficient, α_1, measures an agent's adherence to their previous forecast. Of the 31 agents with $\alpha_1 > 0$, 9 coefficients are significantly positive. Of the other 12 agents with $\alpha_1 < 0$, only one case is significant. For all the clusters, the weighting on the previous forecast is significantly positive, with ID26 noticeably higher at 0.6 compared with 0.39 for the majority cluster.

H_b: $\alpha_2 = 0$; this hypothesis implies the previous consensus forecast is ignored. There is evidence of imitation of the previous consensus; α_2 is positive in all but one case and significant at 5% for 16 agents. For the majority, pessimism, and ID49 clusters, the weighting of around 0.7 on the previous consensus is significant, and the weighting for ID26 is noticeably lower at 0.42. Before Clements (2018), these results would be taken as evidence of herding behaviour by the 16 agents, but he suggests that more reliable evidence comes from H_d.

H_c: $\alpha_3 = 0$; this hypothesis implies no notice is taken of the previous level of disagreement. The coefficient, α_3, is negative in all but one case and is significant at 5% for 19 of 43 agents. Thus, the level of disagreement 6 months previously will tend to lower the forecast. However, the effect of disagreement differs between the clusters. The pessimism cluster is most sensitive to disagreement ($\alpha_3 = -1.74$), and the majority cluster is less sensitive ($\alpha_3 = -1.39$). The two individual clusters are even less sensitive with $0 > \alpha_3 > -1$.

H_d: For evidence of herding behaviour, we require $\alpha_2 \neq 0$ and $\alpha_4 < 0$ equivalent to $\alpha_1 + \alpha_2 < 1$, and the estimated values for α_1 and α_2 for both agents and clusters are shown in Fig. 5. Most agents and all the clusters fall into the no-herding region of the plot, just above the line $\alpha_1 + \alpha_2 = 1$. There is only one case where α_4 is significantly negative at 10%. Using this criterion, there is no convincing evidence of herding behaviour.

To compare our analysis with Clements (2018), we look at quarterly updated forecasts at 3-, 6-, and 9-month horizons. The models[6] fitted to individual agents are:

$$F_{i,T,h} = \alpha_0 + \alpha_1 F_{i,T,h+3} + \alpha_2 C_{T,h+3} + u_i \tag{2'}$$

$$F_{i,T,h} - C_{T,h+3} = \alpha_0 + \alpha_1 \left(F_{i,T,h+3} - C_{T,h+3}\right) + \alpha_4 C_{T,h+3} + u_i \tag{3'}$$

Subject to the provisos raised by Clements (2018) mentioned in Section 2, satisfying the conditions $\alpha_2 \neq 0$ and $\alpha_4 < 0$ is evidence of herding. The results shown in Table 5 are very similar over the three horizons considered, and the main conclusion is that there is no evidence of herding behaviour by any of the agents for which there were sufficient data (more than 10 observations per regression).

We look further at the pooled data for the majority cluster, estimating the models separately for horizons of 18, 12, and 6 months. The results are shown in Table 6. The coefficients, α_1, on the previous forecasts decrease as the horizon shortens and are not significant at a 6-month horizon. In contrast, the coefficients, α_2, on the previous consensus increase as the horizon shortens. The negative effect of the level of disagreement around the previous consensus is only significant for the revised forecasts with a 12-month horizon. The coefficient α_4 is non-negative for the separate horizons, so the conclusion of no herding behaviour is unchanged. These findings suggest that taking forecasts over all horizons together masks the changing behaviour of agents as the horizon shortens.

In summary, within the context of Gallo et al.'s (2002) equations (2) and (3), we find that the variation explained by these regressions varies widely between agents, but, according to their criteria, there is no evidence of imitative behaviour.

In model (2), the coefficient α_1 indicating the weighting placed on the previous forecast was negative for about a third of agents and positive for the rest. In contrast, the narrative underlying model (1) is more consistent. All the agents appear to react similarly, and the coefficients θ_1 and θ_2 are of the same sign and in nearly all cases significantly different from zero.

6.2. Estimation of (1)—a model of forecast revision

We continue to address Q(b) by proposing a new model of an agent's revision of their forecast. This model of forecast revision is used to test hypotheses about two stimuli for forecast revision: change in consensus view and distance of previous forecast from the then consensus. Modelling the revision of forecasts of GDP growth for the same target year at six-month intervals, the revision is $R_{i,T,h} = \left(F_{i,T,h} - F_{i,T,h+6}\right)$, and for clarity, (1) is rewritten as

$$\begin{aligned} R_{i,T,h} = {} & \theta_0 + \theta_1 \left(C_{i,T,h+1} - C_{i,T,h+6}\right) \\ & + \theta_2 \left(F_{i,T,h+6} - C_{i,T,h+6}\right) + \xi_{i,h} \end{aligned} \tag{1'}$$

where $h = 18, 12, 6$. Enforcing a minimum of 10 observations for each regression gave sufficient data to fit model (1') for the revisions over the three sets of horizons for 35 agents.[7]

We summarise the results of this analysis in Table 7, the agents were ranked by the adjusted R^2 of the 18- to 12-month horizon model, and four roughly evenly spread agents were chosen from the highest to the lowest R^2. The four agents ID10, ID63, ID38, and ID73 are members of the majority cluster. Estimated coefficients for all agents are summarised by the median. The model is also fitted to the four clusters. The estimated coefficients for the majority cluster are very similar to the median values of the coefficients for the individual agents.

Summary statistics for the revisions show that, on average, the revisions are negative, the magnitude of the mean revision and the standard deviation of revisions decrease as the horizon shortens. The exception is ID26, where the mean revision is positive and does not decrease as the horizon shortens. The revision model explains a large proportion of the variation in the six-monthly forecast revisions, and the median adjusted R^2 is 0.91 for

[6] This not a strict reproduction of Clements T_1 test, using quarterly updated forecasts would introduce $C_{T,h+6}$ and weaken the test.

[7] The results of a RESET test for the functional form of (3) finds no evidence of non-linearity, see Appendix B.

Table 4
Estimations of the model in (2) and (3): examples and summary for the 43 agents with sufficient available data and for the four clusters.

	Example agents				Median	Clusters			
	ID39	ID12	ID33	ID76		Majority	Pessimists	ID26	ID49
n	36	59	58	24	39	1371	140	62	54
Adj. R^2 (2)	0.935	0.814	0.766	0.233	0.792	0.800	0.788	0.719	0.813
$\widehat{SD(u)}$	0.398	0.611	0.686	0.664	0.652	0.651	0.709	0.695	0.577
α_0	0.293	0.305	0.210	0.426	0.261	0.254	0.387	0.310	−0.083
p value	*0.212*	*0.328*	*0.554*	*0.587*		*0.000*	*0.085*	*0.444*	*0.794*
α_1	−0.914	0.822	0.386	−0.489	0.378	0.392	0.344	0.602	0.405
p value	*0.000*	*0.071*	*0.239*	*0.265*		*0.000*	*0.050*	*0.000*	*0.025*
α_2	2.022	0.238	0.691	1.330	0.691	0.692	0.710	0.420	0.730
p value	*0.000*	*0.609*	*0.048*	*0.030*		*0.000*	*0.000*	*0.000*	*0.000*
α_3	−1.456	−1.536	−1.272	−0.526	−1.334	−1.390	−1.742	−0.973	−0.832
p value	*0.001*	*0.006*	*0.050*	*0.664*		*0.000*	*0.000*	*0.199*	*0.120*
Adj. R^2 (3)	0.584	0.191	0.067	−0.023	0.149	0.168	0.203	0.500	0.090
α_4	0.107	0.060	0.078	−0.158	0.078	0.083	0.054	0.021	0.136
p value	*0.076*	*0.455*	*0.385*	*0.561*		*0.000*	*0.334*	*0.849*	*0.122*

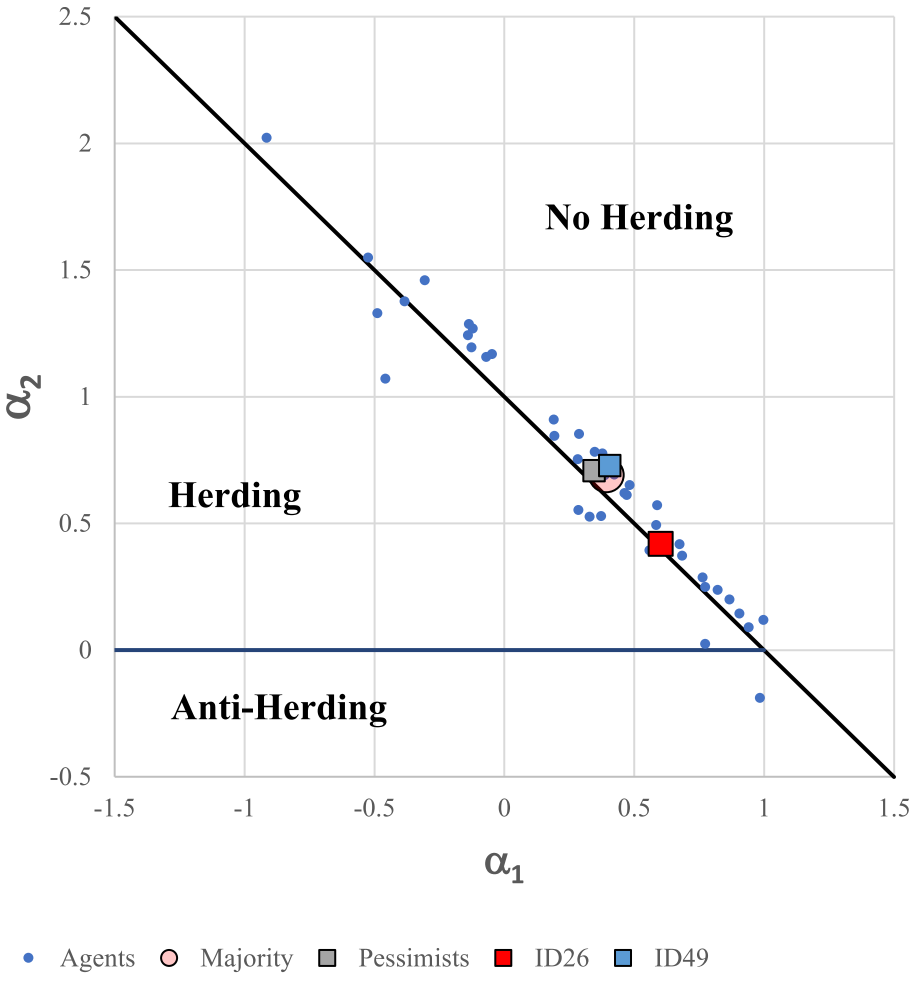

Fig. 5. A plot of estimated (α_1, α_2) for agents and clusters, with the boundaries of implied herding or non-herding regions.

the early revisions between 24 and 18 months, 0.90 for the 18- to 12-month revisions, and then rises to 0.98 for the 12- to 6-month revisions. To isolate the effect of the large revisions that took place during 2009, the analysis was repeated omitting this year. Although the mean and standard deviations of revisions were reduced in magnitude, the estimated regression coefficients and their p values did not change substantially.

For the three hypotheses relevant to Q(b), the conclusions are consistent for three horizons, for the agents,

Table 5
The proportion of agents satisfying the conditions indicated (tests performed at 5% level). For ease of comparison, this table uses the same format as Table 1, Clements (2018).

Horizon	Number	$\alpha_2 \neq 0$	$\alpha_2 \neq 0$ and $\alpha_4 < 0$	$\alpha_4 \neq 0$
9	15	0.47	0.00	0.07
6	17	0.47	0.00	0.00
3	17	0.47	0.00	0.00

Table 6
Summary of the estimations of the model (2) and (3) for the majority agents using separate horizons (and over all horizons for comparison).

	Pooled majority	Horizon (months)		
		18	12	6
n	1371	453	471	462
Adj. R^2 (2)	0.800	0.325	0.913	0.924
$\widehat{SD(u)}$	0.651	0.895	0.434	0.473
α_0	0.254	−0.090	0.585	−0.044
p value	*0.000*	*0.772*	*0.000*	*0.763*
α_1	0.391	0.590	0.288	0.065
p value	*0.000*	*0.000*	*0.000*	*0.498*
α_2	0.693	0.416	0.836	1.015
p value	*0.000*	*0.004*	*0.000*	*0.000*
α_3	−1.390	−0.469	−2.157	−0.365
p value	*0.000*	*0.200*	*0.000*	*0.399*
Adj. R^2 (3)	0.168	0.055	0.543	0.075
α_4	0.083	0.006	0.125	0.080
p value	*0.000*	*0.941*	*0.000*	*0.000*

and the pooled majority. For H_1, the coefficient θ_0 is not significantly different from zero; thus, there is no evidence of a bias in the revisions (minor exceptions are the majority and pessimism clusters for 12- to 6-month revisions, where θ_0 is significant). For H_2, the coefficient θ_1 is significantly positive in all cases. θ_1 is close to unity in nearly all cases, implying that the response to the new information received during the revision period is similar across agents. This result corresponds with, and formalises, the behaviour illustrated in Fig. 2. However, the exceptions are the individual clusters where θ_1 values are noticeable lower than the other clusters.

For H_3, the coefficient θ_2 is significantly negative in all cases. This implies that there is a strong tendency for the agent to revise the forecast in the direction of the consensus. The magnitude of the coefficient for this "correction" increases as the horizon shortens; the median value of θ_2 is −0.64 for the 24- to 18-month revision, −0.72 for the 18- to 12-month revision, and −0.98 for the 12- to 6-month revision.

Here, we consider the single member clusters, agents ID49 and ID26. Although the coefficient estimates for ID49 are broadly similar to the majority agents, the mean 12- to 6-month revisions are comparatively more negative than the other agents, suggesting a regular loss of optimism as the target year develops. The performance of agent ID26 is noticeably different from other agents; the adjusted R^2 is comparatively low, especially for 24- to 18-month revisions, and the magnitudes of θ_1 and θ_2 are lower for all horizons. The mean revisions are comparatively large and positive over each pair of horizons, suggesting a regular gain in optimism as the target year develops.

6.2.1. The mechanism for reducing disagreement as horizons shorten

Continuing to address Q(b), how can the decrease in the level of disagreement as the horizons shorten be explained? We show that the correction process modelled in (1) is potentially the mechanism that leads to the decreasing level of disagreement about the consensus as the forecast horizon shortens, as illustrated in Fig. 3. To simplify the derivation, we consider a median forecaster and drop the suffices i and T. We see from Table 7 that we can further simplify by setting θ_0 to zero and θ_1 to unity, thus rearranging (1) gives

$$F_h - C_{h+1} = (1+\theta_2)\left(F_{h+6} - C_{h+6}\right) + \xi_h. \tag{4}$$

Taking variances of both sides of the equation, we have

$$V\left(F_h - C_{h+1}\right) = (1+\theta_2)^2\, V\left(F_{h+6} - C_{h+6}\right) + V\left(\xi_h\right). \tag{5}$$

If we assume that there is little difference between $V\left(F_h - C_h\right)$ and $V\left(F_h - C_{h+1}\right)$, the variance about the consensus is $(1+\theta_2)^2$ of the variance about the consensus 6 months previously. Since θ_2 is negative, the variances will decrease as the horizon shortens. We may compare the predictions of (5) with the median variances shown (as standard deviations) in Fig. 3.

These comparisons are shown in Table 8, and the match between the median observed variances and those estimated using median values from Table 7 is very close for 6- and 12-month horizons. The match at the 18-month horizon shows an underestimate of 0.11 versus 0.17; however, this can be explained by the positively skewed distribution of 24-month standard deviations apparent in Fig. 3.

6.3. Does the level of disagreement vary with length of horizon or over the cycle with growth expectations

Here, we consider how the level of disagreement, measured by $S_{T,h}$ changes with respect to horizon and with respect to time, or more precisely, the economic cycle.

In Fig. 3, it is clear that the levels of cross-section disagreement are highest at long horizons and decrease as the horizon shortens ($S_{T,h}$ decreases as h decreases). In contrast in Fig. 1, we see that the volatility of the time series of consensus forecasts is low for the long (24-month) horizon and the volatility increases as the horizon shortens ($SD\left(C_{t,h}\right)$ increases as h decreases). Patton and Timmermann (2010) observed a similar phenomenon in the professional forecasts of US GDP and argued that since agents' information is of less value at long horizons, the higher levels of disagreement observed at long horizons are due to heterogeneous prior beliefs.

In the same data set, Patton and Timmermann also observed countercyclical movements in the level of disagreements about GDP growth. Disagreement appeared greatest when growth expectations were low ($S_{T,h}$ increases as $C_{t,h}$ decreases). A simple way of assessing this association, used by Patton and Timmermann, is the correlation between the consensus forecast, $C_{t,h}$, and the cross-section variance of forecasts, $S^2_{T,h}$, for different years at a fixed horizon. We compare our results with those of Patton and Timmermann in Table 9, in nearly all cases

Table 7
A summary of the estimates from fitting the revision model in (1') to the agents' revisions. The model is fitted for each set of horizons. In the left-hand panel, results are shown for a selection of agents and their median values. In the right-hand panel, the model is fitted to revisions from the four clusters, pooled when necessary.

		Example Agents				Median	Clusters			
		ID10	ID63	ID38	ID73		Majority	Pessimism	ID26	ID49
Revisions 24 to 18	n	16	20	13	14	16	453	45	20	18
	Mean Rev.	−0.300	−0.405	−0.446	−0.507	−0.355	−0.351	−0.207	0.270	−0.344
	SD Rev.	0.972	0.994	1.080	1.211	0.945	0.912	0.943	0.991	0.813
	Adj. R^2	0.792	0.911	0.915	0.775	0.909	0.841	0.880	0.305	0.647
	$\widehat{SD(\xi)}$	0.444	0.297	0.315	0.574	0.292	0.363	0.327	0.826	0.483
	θ_0	0.028	0.005	−0.005	−0.154	0.010	0.017	−0.017	−0.003	0.097
	p value	*0.839*	*0.946*	*0.963*	*0.370*		*0.361*	*0.792*	*0.992*	*0.501*
	θ_1	1.207	1.349	1.376	1.599	1.204	1.205	1.309	0.784	0.827
	p value	*0.000*	*0.000*	*0.000*	*0.000*		*0.000*	*0.000*	*0.011*	*0.000*
	θ_2	−0.217	−0.404	−0.708	−0.538	−0.638	−0.654	−0.393	−0.335	−0.518
	p value	*0.471*	*0.190*	*0.040*	*0.364*		*0.000*	*0.002*	*0.058*	*0.024*
Revisions 18 to 12	n	15	21	14	14	16	471	50	21	19
	Mean Rev.	−0.127	−0.190	−0.214	−0.043	−0.185	−0.180	−0.050	0.352	−0.384
	SD Rev.	0.900	0.568	0.530	0.696	0.651	0.658	0.781	0.700	0.757
	Adj. R^2	0.973	0.950	0.899	0.882	0.904	0.900	0.854	0.763	0.793
	$\widehat{SD(\xi)}$	0.148	0.127	0.169	0.239	0.192	0.208	0.299	0.341	0.344
	θ_0	−0.081	−0.028	−0.058	−0.032	−0.013	−0.016	−0.010	0.063	0.010
	p value	*0.072*	*0.344*	*0.243*	*0.672*		*0.099*	*0.844*	*0.531*	0.915
	θ_1	1.045	1.070	0.819	0.973	0.970	0.964	1.054	0.601	0.773
	p value	*0.000*	*0.000*	*0.000*	*0.000*		*0.000*	*0.000*	*0.000*	*0.000*
	θ_2	−0.907	−0.849	−0.508	−0.627	−0.718	−0.741	−0.622	−0.444	−0.705
	p value	*0.000*	*0.000*	*0.098*	*0.000*		*0.000*	*0.000*	*0.000*	*0.000*
Revisions 12 to 6	n	16	21	13	14	16.5	462	47	21	18
	Mean Rev.	−0.063	−0.005	−0.108	−0.029	−0.050	−0.034	0.017	0.281	−0.300
	SD Rev.	0.612	0.479	0.502	0.541	0.538	0.538	0.550	0.690	0.646
	Adj. R^2	0.937	0.980	0.983	0.935	0.975	0.964	0.960	0.944	0.858
	$\widehat{SD(\xi)}$	0.154	0.067	0.066	0.138	0.092	0.101	0.109	0.164	0.243
	θ_0	−0.077	0.014	0.003	−0.027	0.010	0.019	−0.035	0.062	−0.082
	p value	*0.098*	*0.350*	*0.860*	*0.528*		*0.000*	*0.045*	*0.135*	*0.202*
	θ_1	0.962	1.126	1.155	1.114	1.109	1.083	1.109	0.836	0.933
	p value	*0.000*	*0.000*	*0.000*	*0.000*		*0.000*	*0.000*	*0.000*	*0.000*
	θ_2	−1.692	−1.174	−1.199	−1.090	−0.975	−1.000	−0.874	−0.762	−0.785
	p value	*0.000*	*0.000*	*0.000*	*0.000*		*0.000*	*0.000*	*0.000*	*0.000*

Table 8
A comparison between the observed disagreement around the consensus measured as variance with that estimated by (5) using median values for the variables and parameters.

Horizon	Variance about consensus $V(F_h - C_{h+1})$ Median Observed	Median θ_2	$(1+\theta_2)^2$ $V(F_{h+6} - C_{h+6})$	Median Residual Variance $V(\xi_h)$	Estimated Variance $\widehat{V(F_h - C_{h+1})}$	Absolute error
24	0.2004					
18	0.1669	−0.638	0.0262	0.0850	0.1112	0.0556
12	0.0617	−0.718	0.0132	0.0368	0.0500	0.0117
6	0.0056	−0.975	0.0000	0.0084	0.0085	0.0029

there is a negative correlation suggesting greater disagreement is associated with lower expected growth. The full UK results show a lower magnitude of correlation than the US results at horizons of 12 months or less, and the magnitude of correlation is greater for the UK at 18-month horizon. However, if the forecasts for 2009 are omitted, the correlations for 12- and 18-month horizons are lower (in magnitude) and less significant; at a 6-month horizon, the negative correlation disappears. In summary, we find similar evidence of heterogeneous beliefs of forecasters, but less compelling evidence that the level of disagreement is negatively associated with the anticipated rate of growth.

We investigate this relationship further at the agent level. Disagreement is represented by the absolute difference between the forecast and the contemporaneous consensus; the previous month's consensus represents the current view of economic growth; the model tested

Table 9
The correlations *(p values)* between the mean and the variance of forecasts of GDP growth using HM Treasury data for UK at 24- to 6-month horizons, values for US taken from Table 1 of Patton and Timmermann (2010).

Horizon	UK (1997–2019)		UK (1997–2008, 2010–2019)		USA (1991–2008)	
24	−0.36	*0.11*	−0.36	*0.11*	−0.30	*>0.10*
18	−0.55	*0.01*	−0.42	*0.06*	−0.10	*>0.10*
12	−0.55	*0.01*	−0.32	*0.16*	−0.69	*<0.10*
6	−0.30	*0.17*	0.37	*0.10*	−0.48	*<0.10*

Table 10
Estimation of (6) at different horizons for forecasters with more than 10 observations. The test for $\gamma_1 < 0$ is at 5% (one-sided).

Horizon	No forecasters	Mean γ_0	Proportion $\gamma_1 < 0$ and significant at 5%	
			1997–2019	1997–2008, 2010–2019
24	19	0.436	0.16	0.16
18	20	0.411	0.30	0.20
12	18	0.213	0.06	0.00
6	19	0.067	0.26	0.05

is:

$$\left|F_{i,T,h} - C_{T,h}\right| = \gamma_0 + \gamma_1 C_{T,h+1} + \epsilon_i \tag{6}$$

The condition $\gamma_1 < 0$ indicates that a forecaster tends to differ more from the consensus when growth expectations are lower. Table 10 shows summary results for the estimation of (6) at 6-, 12-, 18-, and 24-month horizons, and the value of the mean γ_0 indicates the decreasing magnitude of the disagreement as the horizon shortens, replicating the findings in Fig. 3 and above in Table 8. At the shorter horizons of 6 and 12 months, the proportion of forecasters where $\gamma_1 < 0$ is strongly influenced by the effect of 2009 where $C_{T,h+1}$ was unprecedently low, see Figs. 1 and 2. With 2009 removed, there is no real evidence of an effect. At longer horizons of 18 and 24 months, there is stronger evidence that some forecasters tend to disagree with the consensus more when expected growth is below average (Forecasters ID11 and ID49, a single member cluster, satisfy $\gamma_1 < 0$ at both 18- and 24-month horizons.).

7. Question (c): Persistent distinctions—leaders and followers and accuracy of clusters

Question(c) asks whether there are persistent patterns across forecasters or clusters. We investigate two issues. First, we examine monthly revisions of forecasts to discover whether there is any evidence of an agent leading or following the consensus. Second, we look to see whether there are persistent differences in the forecast accuracy of different clusters.

7.1. Leaders and followers

A Granger causality test is performed with a one-month lag only. The one-month lag is considered for two reasons; first, a reaction time of a month is relevant in practice, and second, a data set with two or more lagged observations would include fewer agents. The following analysis considers 38 agents where there are at least 50 relevant observations available. Given the observations of the forecast of agent i at horizon k, the contemporary consensus, and the corresponding values for the previous month, $\left(F_{i,T,k}, C_{T,k}, F_{i,T,k+1}, C_{T,k+1}\right)$, the following regressions are performed for each agent (observations are pooled over T and k). Note that (7) is a rewriting of (2') using a monthly rather than a quarterly interval.

$$F_{i,T,k} = \varphi_{10} + \varphi_{11} F_{i,T,k+1} + \varphi_{12} C_{T,k+1} + \varepsilon_1 \tag{7}$$

$$C_{i,T,k} = \varphi_{20} + \varphi_{21} F_{i,T,k+1} + \varphi_{22} C_{T,k+1} + \varepsilon_2. \tag{8}$$

If φ_{21} is significantly positive, then agent i tends to lead the consensus. The one-month ahead consensus forecast, using only its previous values, is improved by the previous forecast of agent i. Similarly, if φ_{12} is significantly positive, then agent i tends to follow the consensus. In our analysis, we identify agents where the coefficients are not only significant but noticeably larger than average, specifically $\varphi_{21} > 0.1$ and $\varphi_{12} > 0.5$. Under these criteria, 4 agents (2 banks, 1 asset manager, and 1 consultancy) lead the consensus and 2 agents (1 consultancy and 1 professional institution) follow the consensus. The results are summarised in Table 11 showing the estimated coefficients for these 6 agents and the two single agent clusters are included for comparison purposes. Looking across the rows for φ_{12} and φ_{21}, the contrast between the leaders and followers is clear. The leaders are little affected by the previous consensus, and the consensus is little affected by the followers. For the single cluster agents, ID26 and ID49, their own forecasts are dominated by their own previous forecasts with small weightings on the previous consensus, in return the consensus places little weight on their previous forecasts. To test the robustness of these results, the models were re-estimated excluding forecasts for 2009. Apart from minor changes in the estimates the results in Table 11 stand, with the exception of ID38 where φ_{21} ceases to be significant. However, as 2009 was a year where forecasting was more demanding than usual, it is worth illustrating ID38's leading of the consensus for that year. The forecasts of % GDP in 2009 by ID38 and ID26 at horizon h alongside the consensus forecasts at horizon $h-1$, one month later, are shown in Fig. 6. The contrast is clear: For the crucial year 2009, the forecasts of ID38 closely predict the consensus forecast one month later, and ID26 is initially far more pessimistic than the consensus, but from an 18-month horizon onwards is a little less pessimistic than the consensus.

Since the consensus is the mean value of the available forecasts, it follows that some individual agents will follow a leading agent more strongly. To identify agents who tend to follow leading agents, the following regression is performed for each of the four leaders.

$$F_{j,T,k} = \kappa_{20} + \kappa_{21} F_{Leader,T,k+1} + \kappa_{22} F_{j,T,k+1} + \varepsilon_4 \tag{9}$$

Applying the filters of at least 50 observations for the regression, a significant $\kappa_{21} > 0.3$ for the data with and without forecasts for 2009, we find that:

ID38 is followed by ID10, ID28, ID30, ID48, and ID63;
ID44 is followed by ID02 and ID40;

Table 11
Summary of the Granger causality test showing estimated coefficients for four leaders of the consensus, two followers of the consensus, and two others.

		Leaders				Followers		Others	
		ID38	**ID44**	**ID73**	**ID74**	**ID12**	**ID21**	**ID26**	**ID49**
	n	218	133	81	128	86	94	232	241
(7)	R^2	0.974	0.980	0.993	0.973	0.980	0.988	0.936	0.966
	$\sigma(\varepsilon_1)$	0.249	0.255	0.155	0.270	0.142	0.201	0.309	0.209
	φ_{10}	-0.110	-0.059	-0.071	-0.103	-0.050	-0.071	-0.042	-0.032
	p value	*0.000*	*0.054*	*0.002*	*0.004*	*0.181*	*0.017*	*0.259*	*0.304*
	φ_{11}	0.864	0.967	0.993	0.814	0.445	0.374	0.944	0.837
	p value	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	φ_{12}	0.173	0.059	0.042	0.212	0.594	0.676	0.081	0.182
	p value	*0.104*	*0.318*	*0.253*	*0.043*	*0.000*	*0.000*	*0.001*	*0.000*
(8)	R^2	0.983	0.987	0.992	0.989	0.982	0.994	0.974	0.985
	$\sigma(\varepsilon_2)$	0.192	0.197	0.169	0.171	0.127	0.134	0.186	0.160
	φ_{20}	-0.088	-0.086	-0.066	-0.095	-0.051	-0.072	-0.081	-0.069
	p value	*0.000*	*0.000*	*0.010*	*0.000*	*0.135*	*0.000*	*0.000*	*0.004*
	φ_{21}	0.152	0.136	0.194	0.170	0.071	-0.024	0.033	-0.027
	p value	*0.051*	*0.002*	*0.000*	*0.009*	*0.414*	*0.714*	*0.014*	*0.236*
	φ_{22}	0.878	0.887	0.854	0.855	0.939	1.059	1.007	1.054
	p value	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*

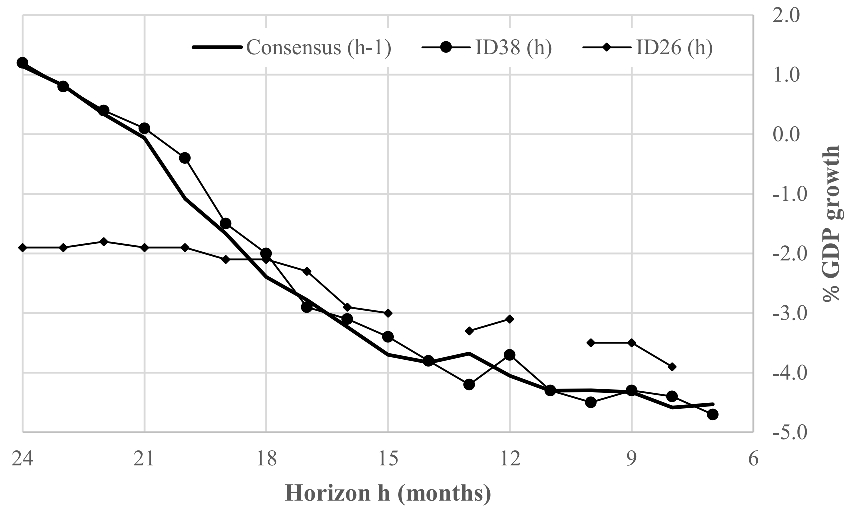

Fig. 6. The evolution of forecasts of 2009% GDP growth showing the forecasts of ID38 and ID26 and the consensus forecast one month later.

ID74 is followed by ID08, ID11, and ID30.

By identifying some agents as leaders, this analysis demonstrates that their information sources are deemed to have some persistent value by the other agents, as evidenced by their influence on the consensus. Further analysis shows that this influence on the consensus is driven by most of the leading agents having subsets of the consensus specifically following them. These subsets of agents are either influenced by the forecasts of the leading agents or receive their information after a delay. By identifying some agents as followers, we find more evidence of herding, and these agents weight the previous consensus forecast more heavily than their own previous forecast.

7.2. Comparative forecasting accuracy of the clusters

D'Agostino et al. (2012) ask "Are some forecasters really better than others?" and found little difference between the forecasters they considered. Here, we compare the forecasting accuracy of the clusters identified in Section 5. The actual % GDP growth (represented by the first quarterly estimate PN2) is denoted A_T. We consider the period 1997 to 2019, omitting 2009. To make our measure of accuracy robust to differing levels of disagreement over the years, we standardise the errors. For agent i's forecast of GDP in year T with a horizon of h months, $F_{i,T,h}$, the error, standardised by the contemporary level of disagreement among the $n_{T,h}$ forecasts available, is

Table 12
The mean standardised errors and the root mean squared standardised errors for the four clusters and the group mean forecasts for horizons of 24 months to 6 months.

Horizon		Cluster				Group mean
		Majority	Pessimists	ID26	ID49	
24	Count	256	26	17	15	20
	MsE	−0.009	0.093	1.684	−0.352	0.034
	RMSsE	0.878	1.159	2.315	1.219	0.737
18	Count	271	28	17	18	21
	MsE	0.288	0.281	1.832	−0.120	0.270
	RMSsE	0.893	1.201	2.384	1.233	0.777
12	Count	248	27	18	15	21
	MsE	0.308	−0.049	0.875	−0.337	0.265
	RMSsE	0.950	1.202	1.597	1.533	0.815
6	Count	275	26	17	18	21
	MsE	0.424	0.375	0.613	0.431	0.420
	RMSsE	1.014	1.013	1.168	1.204	0.938

computed thus:

$$e_{i,T,h} = \frac{\left(A_T - F_{i,T,h}\right)}{\sqrt{\frac{\sum_j \left(A_T - F_{j,T,h}\right)^2}{n_{T,h}}}}$$

For each cluster c, the standardised errors are summarised by the mean standardised error (MsE) and the root mean squared standardised error (RMSsE) for all $e_{i,T,h}$ for $i \in c$, for all available T and for $h = 24, 18, 12, 6$ months. The results for each cluster and for the group mean forecast for each horizon are shown in Table 12.

For each horizon, the pattern is similar. The group mean forecast has the lowest RMSsE. The majority group is more accurate than the pessimists (apart from the 6-month horizon when the RMSsE are similar). The two single member clusters are less accurate than the larger clusters, and ID49 is more accurate than ID26 for all except the 6-month horizon. For all horizons, ID26's MsE is greater than for the other clusters, showing persistent underestimation. Once the superior accuracy of the group mean has been established, it follows that the differences between the clusters formed on the basis of forecaster behaviour will be reproduced in the comparative accuracy of their forecasts. At each horizon, one, two, or three agents had a lower RMSsE than the group mean forecast, but no agent achieved this more than once.

In a study of US Survey Professional Forecasters forecasts, Clements (2021) found that disagreement among forecasters tended to be mainly caused by the same groups of forecasters and that "forecasters who stand out from the crowd do not tend to produce more accurate forecasts". Our results in Table 12 chime with these findings.

8. Summary and conclusions

In our analysis of fixed event forecasts of annual UK GDP growth by 77 agents active during the period 1997 to 2019, we sought to answer three research questions concerning: the heterogeneity of agent's beliefs manifested by the dispersion of forecasts; the factors determining how agents revise their forecasts; and the value of private information available to some agents. We will summarise our findings for each research question and then draw our conclusions.

Q(a) What is the structure of the heterogeneity of beliefs? Are agent's beliefs randomly distributed or do agents fall into groups with similar beliefs. For example, do some agents persistently exhibit relative optimism or pessimism.

The relative position of agents' forecasts to the consensus was transformed to t values, and this allowed cluster analysis to be used to study the persistence of agents' behaviours relative to the consensus or group mean forecast. We found that the heterogeneity was not due to a random distribution of behaviours, but to four groups of agents exhibiting similar behaviour. The largest group tends to produce similar forecasts that dominate the consensus forecast. A group of three agents tend to be more pessimistic than the majority group. The beliefs and models of two agents lead them to produce idiosyncratic forecasts, very different from each other and both other groups of agents. Both agents exhibit greater variability in their forecasts' positions relative to the group mean, ID26 tends to be atypically pessimistic, and ID49 tends to relative optimism.

Q(b) Is there evidence of herding behaviour as agents revise their forecasts? What is the extent of the roles played by: the previous consensus forecast; the agent's previous forecast; and *the change in consensus forecast? As the horizon shortens, how can the decreasing dispersion be explained?*

In the first part of the analysis, we applied the Gallo et al.'s (2002) linear model to all the forecasts over 18-, 12-, and 6-month horizons, and the model explains a forecast in terms of the previous forecast, the previous consensus, and the previous levels of disagreement. The proportion of variation in forecasts explained by this model varies widely between agents from 0.93 to 0.23 with a median of 0.79. The most consistent effect was the predominantly negative sign on the previous level of disagreement (measured by standard deviation), inclusion of this term allows the model to capture decreasing dispersion as the horizon shortens. Pooling the forecasts of agents in the majority group showed that the ratio of the weights on the previous forecast versus the consensus is 0.36 to 0.64. Under the model assumptions of Clements (2018), no convincing evidence of herding behaviour was found.

In the second part of the analysis, we apply our new model to forecast revisions over a six-month interval, separately for three horizons. The model captures the effect of the change in consensus over the period and the effect of the distance between the agent's previous forecast and the then consensus. The fit of this model, the proportion of variation in revisions explained, is consistently strong across agents with a median adjusted R^2 greater than 0.9 for all three horizons. The results show a consistently positive effect of the change in consensus view and a consistently negative effect on the distance of the previous view from the consensus. We show how this negative effect leads to decreasing dispersion as horizons shorten. Patton and Timmermann (2010) find evidence that disagreement between agents (forecasting US GDP and inflation) increases during recessions.

In our exploration of the data, we found some evidence of a negative association between the level of disagreement between agents' forecasts and expected growth in GDP. An agent-based model of disagreement with the group mean forecast shows evidence of this effect for a small minority of agents, including the single member cluster, ID49, at 18- and 24-month horizons.

Q(c) Is the private information available to some agents of persistent value? Do some agents tend to lead or to follow the consensus? Are some clusters of forecasters persistently better than others?

A Granger causality test is used in a comparison of the monthly forecasts of individual agents with the consensus forecast. The forecasts of a minority of agents (4 of 38 considered) are shown to significantly lead the consensus. The forecasts of two agents are shown to follow the consensus by weighting the previous consensus more heavily than their own previous forecast. These results suggest that a minority of agents have superior models or have access to flows of private information before the other agents, allowing them to adjust their forecasts. The other agents either receive the information later or react to the changes in the forecasts of other agents. An analysis of forecasts by clusters of agents shows the overall mean forecast to be most accurate over all horizons, followed by the majority cluster, the single member clusters are least accurate.

The literature suggests that the dispersion of agents' forecasts is due to the heterogeneity of their beliefs and models. We have established that an important driver of the dispersion of forecasts is that the agents fall into four groups; thus, it is reasonable to assume that beliefs within groups are relatively homogeneous and the heterogeneity of agents' beliefs is partially due to differences in beliefs between these groups.

The finding that the revision of forecasts as horizons shorten is due to the change in consensus forecast during the revision interval is intuitively reasonable. The negative effect of the distance between the previous forecast and the consensus is likely to be partially due to the agent's model reacting to reducing uncertainty about the drivers of GDP, and in later stages, publication of early quarterly GDP estimates.

There are several alternative research directions to further explore the background to these findings. To shed more light on the heterogeneity of beliefs, a questionnaire survey of agents may reveal their preferred definition of GDP growth, but it is less likely to reveal sources of private information such as the identity of the covariates they consider pivotal in revising their forecasts. A study of agents' reactions to the publication of data such as quarterly %GDP growth figures or Purchasing Managers' Indices is likely to be more productive.

Appendix A. Definitions of herding behaviour

(i) Non-specific to forecasting

Banerjee (1992): "Herd behavior-everyone doing what everyone else is doing, even when their private information suggests doing something quite different".

Hirshleifer and Teoh (2003): "Herding/dispersing is defined to include any behaviour similarity/dissimilarity brought about by the interaction of individuals".

Meyer and Kunreuther (2019): "Herding behaviour—a tendency to base choices on the observed actions of others".

(ii) Adjusting an individual forecast towards a consensus view

Bewley and Fiebig (2002): "The term herding is used to denote the tendency to produce a range of forecasts which is narrower than that which would likely be observed if the forecasts were produced on a strictly independent basis because a forecaster takes the previous consensus mean into account".

Bernhardt et al. (2006): "In its most basic form, herding amounts to biasing a forecast away from an analyst's best estimate, toward the consensus forecast of earlier analysts; while anti-herding amounts to biasing a forecast away from that consensus".

Tsuchiya (2015): "Herding is a behavior wherein the forecaster biases a forecast away from his best estimate of the posterior mean toward the consensus".

Pierdzioch et al. (2016): "Forecaster herding arises when forecasters do not rely only on private information while forming their inflation forecasts but rather manoeuvre their forecasts in the direction of the forecasts of others".

(iii) Adjusting an individual forecast towards a consensus view to the detriment of accuracy

Clements (2015): "On the notion of herding—whether forecasters take into account the views of others when they produce their forecasts. This may be manifest in forecasters skewing their optimal forecast towards the consensus view, or artificially exaggerating the difference between their forecast and the consensus, where optimal is to be understood in the narrow sense of maximizing the expected accuracy of the forecast (for example, minimizing the expected squared forecast error)".

Clements (2018): "On the notion of herding—whether forecasters put undue weight on the views of others when they produce their forecasts, and either move their forecasts toward, or away from the consensus view, in a way that is detrimental to forecast accuracy".

(iv) Herding as the converse of boldness

Clement and Tse (2005): "We classify forecasts as bold if they are above both the analyst's own prior forecast and the consensus immediately prior to the analyst's forecast, or else below both. We classify all other forecasts (i.e., those that move away from the analyst's own and toward the consensus) as herding forecasts".

Appendix B. Further examination of pooled consensus model

The estimated coefficients of the revision model in (3') applied to the members of the consensus cluster are given in Table 7. Here, we check the validity of the functional form of (3') using Ramsey's RESET test using powers of the estimated values from 2 to 4. The results of the RESET test suggest that there is little non-linearity present; however,

Table A.1
Pooled consensus cluster data: F values from RESET test and estimated coefficients of (3a).

Revisions	RESET test			Adj. R^2		Estimated coefficients for (3a) (*p val.*)				
	F val.	df.	*p* val.	(3a)	(3)	θ'_0	θ'_1	θ'_2	θ'_3	θ'_4
24 to 18	1.304	(3, 452)	*0.272*	0.843	0.841	0.044	1.203	−0.642	−0.007	−0.163
						0.044	*0.000*	*0.000*	*0.756*	*0.018*
18 to 12	2.474	(3, 470)	*0.061*	0.901	0.900	−0.023	1.046	−0.733	0.041	0.020
						0.034	*0.000*	*0.000*	*0.007*	*0.536*
12 to 6	1.203	(3, 461)	*0.308*	0.965	0.964	0.010	1.091	−1.000	0.048	0.000
						0.141	*0.000*	*0.000*	*0.025*	*0.993*

as a further check, the following model including squared covariates is estimated.

$$R_{i,T,h} = \theta'_0 + \theta'_1\left(C_{i,T,h+1} - C_{i,T,h+6}\right) + \theta'_2\left(F_{i,T,h+6} - C_{i,T,h+6}\right) + \theta'_3\left(C_{i,T,h+1} - C_{i,T,h+6}\right)^2 + \theta'_4\left(F_{i,T,h+6} - C_{i,T,h+6}\right)^2 + \xi'_{i,h} \tag{3a}$$

where $h = 18, 12, 6$ and i covers members of the consensus cluster. The results considering revisions over the three sets of horizons are given in Table A.1. For the revisions over the longer horizon from 24 months to 18 months, there is a small negative effect of the squared distance of the previous forecast from the previous consensus. This suggests a possible tendency at the longer horizon to make a greater revision if the previous forecast was above the previous consensus (than if the previous forecast was the same distance below the previous consensus). For the revisions over shorter horizons, 18 to 12 months and 12 to 6 months, there is a small positive effect of the squared change in consensus. This suggests that if the consensus forecast has moved upwards then the magnitude of the effect of this move will be greater (than the magnitude of the effect if the consensus had moved downwards by the same distance).

References

Amador, M., & Weill, P.-O. (2010). Learning from prices: Public communication and welfare. *Journal of Political Economy*, *118*, 866–907.

Anesti, N., Galvão, A. B., & Miranda-Agrippino, S. (2020). *Uncertain kingdom: nowcasting GDP and its revisions*: *bank of england working paper no. 764, Last revised: 5 feb 2020*.

Angeletos, G.-M., Huo, Z., & Sastry, K. A. (2020). Imperfect macroeconomic expectations: Evidence and theory. *NBER Macroeconomics Annual*, *35*.

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, *107*(3), 797–817.

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

Bernhardt, D., Campello, M., & Kutsoati, E. (2006). Who herds? *Journal of Financial Economics*, *80*, 657–675.

Bewley, R., & Fiebig, D. G. (2002). On the herding instinct of interest rate forecasters. *Empirical Economics*, *27*, 403–425.

Blanchflower, D. (2019). *Not working: where have all the good jobs gone?*. New Jersey: Princeton University Press.

Blastland, M. (2019). *The hidden half: how the world conceals its secrets*. London: Atlantic Books.

Boero, G., Smith, J., & Wallis, K. F. (2008). Uncertainty and disagreement in economic prediction: Bank of England survey of external forecasters. *The Economic Journal*, *118*, 1107–1127.

Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. *The Journal of Finance*, *60*(1), 307–341.

Clements, M. P. (2012). Do professional forecasters pay attention to data releases? *International Journal of Forecasting*, *28*, 297–308.

Clements, M. P. (2015). Do US macroeconomic forecasters exaggerate their differences? *Journal of Forecasting*, *34*, 649–660.

Clements, M. P. (2018). Do macroforecasters herd? *Journal of Money, Credit and Banking*, *50*, 265–292.

Clements, M. P. (2019). Do forecasters target first or later releases of national accounts data? *International Journal of Forecasting*, *35*(4), 1240–1249.

Clements, M. P. (2021). Forecaster efficiency, accuracy and disagreement: Evidence using individual-level survey data. *Journal of Money, Credit and Banking*, http://dx.doi.org/10.1111/jmcb.12867, Forthcoming.

Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society. Series B.*, *45*, 311–354.

D'Agostino, A., McQuinn, K., & Whelan, K. (2012). Are some forecasters really better than others? *Journal of Money, Credit and Banking*, *44*(4), 715–732.

Dovern, J. (2013). When are GDP forecasts updated? Evidence from a large international panel. *Economics Letters*, *120*(3), 521–524.

Frenkel, M., M, Mauch., & Rülke, J.-C. (2020). Do forecasters of major exchange rates herd? *Economic Modelling*, *84*, 214–221.

Gallo, G., Granger, C. W. J., & Jeon, Y. (2002). Copycats and common swings: the impact of the use of forecasts in information sets. *IMF Staff Papers*, *49*(1), 1–18.

Guzman, M., & Joseph E Stiglitz, J. E. (2021). Pseudo-wealth and consumption fluctuations. *The Economic Journal*, *131*(633), 372–391.

Hirshleifer, D., & Teoh, S. H. (2003). Herd behaviour and cascading in capital markets: a review and synthesis. *European Financial Management*, *9*(1), 25–66.

Maćkowiak, B., Matějka, F., & Wiederholt, M. (2021). Rational inattention: A review, ECB working paper no. 2021/2570. Available at SSRN: https://ssrn.com/abstract=3871534.

Mankiw, N. G., & Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the new Keynesian Phillips curve. *Quarterly Journal of Economics*, *117*, 1295–1328.

Meyer, R., & Kunreuther, H. (2019). *The ostrich paradox: why we underprepare for disasters*. Philadelphia PA: Wharton School Press.

Office for National Statistics (2020). Improvements to the measurement of UK GDP: an update on progress. published 6 2020.

Patton, A. J., & Timmermann, A. (2010). Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, *57*(7), 803–820.

Patton, A. J., & Timmermann, A. (2011). Predictability of output growth and inflation: A multi-horizon survey approach. *Journal of Business & Economic Statistics*, *29*(3), 397–410.

Pierdzioch, C., Reid, M. B., & Gupta, R. (2016). Inflation forecasts and forecaster herding: Evidence from South African survey data. *Journal of Behavioral and Experimental Economics*, *62*, 42–50.

Pierdzioch, C., & Rülke, J.-C. (2012). Forecasting stock prices: Do forecasters herd? *Economics Letters*, *116*(3), 326–329.

Pierdzioch, C., Rülke, J.-C., & Stadtmann, G. (2010). New evidence of anti-herding of oil-price forecasters. *Energy Economics*, *32*, 1456–1459.

Pierdzioch, C., Rülke, J.-C., & Stadtmann, G. (2013). Forecasting metal prices: Do forecasters herd? *Journal of Banking & Finance*, *37*, 150–158.

Pierdzioch, C., & Stadtmann, G. (2011). Do exchange-rate forecasters herd? *Applied Economics Letters*, *18*(8), 739–741.

Rich, R., & Tracy, J. (2018). *A closer look at the behavior of uncertainty and disagreement: micro evidence from the euro area*: *Working paper*, Federal Reserve Bank of Cleveland.

Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *American Economic Review*, *80*, 464–479.

Spiro, P. S. (1989). Improving a group forecast by removing the conservative bias in its components. *International Journal of Forecasting*, *5*, 127–131.

Symons, P. (2001). Revisions analysis of initial estimates of annual constant price GDP and its components: The pattern of bias in the revisions to the initial estimate of real growth in GDP (and its components) as published in the annual blue book. *Economic Trends*, [ISSN: 00130400] *568*.

Tsuchiya, Y. (2015). Herding behavior and loss functions of exchange rate forecasters over interventions and financial crises. *International Review of Economics and Finance*, *39*, 266–276.

Weale, M. (2021). Do economists expect too much from expectations? *National Institute Economic Review*, *255*, 25–41.

Zarnowitz, V., & Lambros, L. A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, *95*(3), 591–621.