ICMA Centre

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Assessing Macro-Forecaster Herding: Modelling versus Testing

July 2018

Michael P. Clements
ICMA Centre, Henley Business School, University of Reading

Discussion Paper Number:
ICM-2018-01

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAs AMBA ACCREDITED

admin@icmacentre.ac.uk

www.icmacentre.ac.uk

Assessing Macro-Forecaster Herding: Modelling versus Testing

Michael P. Clements*
ICMA Centre, Henley Business School, University of Reading

July 13, 2018

Abstract

We draw on fixed-event and fixed-horizon survey expectations to better understand macroeconomic forecasting behaviour. Fixed-event forecasts facilitate testing for herding behaviour, while fixed-horizon forecasts lend themselves to modelling the effects of consensus forecasts on individual forecasts. By pursuing these two approaches simultaneously for each individual forecaster, we can determine when the significance of the consensus forecasts in explaining an individual's forecasts is consistent with enhancing forecast accuracy, and when it reflects strategic behaviour in response to other motives.

JEL Classification: E37.

Keywords: macro-forecasting, (anti-)herding, fixed-event forecasts, fixed-horizon forecasts.

1 Introduction

It is generally held that the assumption that all macro-forecasters report their most accurate forecasts of the economic outlook may be naive, and that forecasters' payoffs may be affected by a range of other factors which will therefore affect their behaviour. As noted by Lamont (2002, p. 265), for example, forecaster behaviour may be influenced by goals which include to 'optimize profits or wages, credibility, shock value, marketability, political power...' (see Laster, Bennett and Geoum (1999) and Ottaviani and Sorensen (2006), *inter alia*). Many of these suggest that individuals will consider the views of others when they make their forecasts, and not solely for the

*Michael Clements is Professor of Econometrics at the ICMA Centre, Henley Business School, University of Reading, and an associate member of the Institute for New Economic Thinking at the Oxford Martin School, University of Oxford. Email: m.p.clements@reading.ac.uk. The computations were undertaken using code written in Gauss 14, Aptech Systems.

accuracy-enhancing information such forecasts may carry. A forecaster may not wish to stand too far apart from 'the crowd' for fear of being exposed as second rate. Alternatively, a forecaster may actively choose to distance herself from the consensus in the hope that she might on occasion be the out-and-out winner, and that the attendant acclaim would more than compensate for the more numerous occasions such a strategy results in large forecast errors. Both the forecaster's preferences, and those of the consumers of the forecasts, are likely to influence forecasting strategies. The literature allows for the possibility that forecasters may consider the forecasts of others for 'variety-seeking' or 'consensus-seeking' motivations that are distinct from considerations of forecast accuracy (see, e.g., Batchelor and Dua (1992) for an early contribution, and Clements (2018) for a list of recent papers suggesting 'anti-herding' behaviour). We define herding (anti-herding) as occurring when a forecaster puts more weights on the views of others than is warranted by forecast accuracy considerations, either in terms of moving too far towards the herd, or too far away (anti-herding).[1]

As suggested above, if forecasters have private information, then the forecasts of others will typically be informative if the objective is maximizing forecast accuracy. If forecasters have private information, then the usual practice of aggregating that information by taking a cross-sectional 'mean' might be questionable.[2] If there is little private information, and forecasts can be regarded as equal to the true value plus measurement or idiosyncratic error, then averaging would be a sensible way of aggregating the information in the disparate forecasts.[3] Our focus is not on the best way to aggregate the information provided by the cross-sectional pool of forecasts, and we follow the literature in defining the consensus as an average, namely the median.

There are two conceptually distinct ways of assessing the importance of herding (or anti-herding) behaviour: the modelling and testing approaches to macroeconometric research, although in practice the two may merge. The modelling approach seeks to determine the key factors or explanatory variables which determine the forecasts, and to estimate the quantitative effects of these determinants: the coefficients on the explanatory variables in the equation determining the forecasts. Variables capturing the consensus view, or last period's consensus, either of the target period or of

[1] Banerjee (1992, p.798) describes *herd behaviour* as "everyone doing what everyone else is doing, even when their private information suggests doing something quite different".

[2] In the context of probability forecasting, Satopää (2018) gives an example suggesting an 'average' might not accurately reflect the diversity of information. If a company's marketing and operations experts both independently report a forecast of 0.9, based on their own sources of information, arguably an aggregate probability forecast ought to be more extreme than the individual forecasts: closer to 1 than 0.9.

[3] Satopää (2018) suggests the classic example of guessing the weight of an ox at a fair (Galton (1907). Private information - information known to some individuals but not others - is likely to play only a minor role, as the ox is on public display. The median of the 787 guesses of the weight of the ox was 1207lbs, compared to an actual weight of 1198lbs, an error of less than 1% of the actual weight.

an intermediate time period, may enter the equation. However, depending on how the consensus terms are defined and the form of the equation, it may be unclear whether apparent significance of such terms suggests herding (or anti-herding), or simply reflects the useful information (in terms of improving forecast accuracy) contained in such terms. Whether forecasters put undue weight on the views of other forecasters, from an accuracy-enhancing perspective, may be difficult to determine, and be made more so by relatively high degrees of collinearity between the variables in the model.

The testing approach sidesteps this issue by setting up the problem in such a way that no weight should be given to the consensus under the null of no herding. We are essentially testing whether the forecasts, or more usually some transformation of the forecast, such as the forecast revision, is orthogonal to a term capturing the herding effect.

The testing and modelling approaches are perhaps best exemplified by the rich and diverse literature on aggregate consumers' expenditure, beginning with the 'modelling' paper of Davidson, Hendry, Srba and Yeo (1978) and the 'testing' paper of Hall (1978). The former sought to *explain* aggregate consumers' expenditure in terms of its key determinants, such as income and inflation. Davidson *et al.* (1978) also proposed a progressive research strategy, whereby any model should be able to explain extant findings. Equally as importantly, it should be able to explain what researchers would find when they consider new specifications: otherwise it would be deficient in some respect, as a model of the phenomenon of interest.[4] By way of contrast, Hall (1978) sought to *test* an implication of the rational expectations permanent-income hypothesis, namely, the log of consumers expenditure should (approximately) resemble a random walk, with unpredictable increments. This was tested by regressing the log difference at time t on variables which could reasonably be assumed to be in the agents' information set at time $t-1$. Unlike the modelling approach, the aim is not to model the change in consumers' expenditure between periods t and $t-1$, that is, how current consumption responds to current income, and other possible determinants.[5]

Our approach in this paper considers what can be learnt from the modelling and testing approaches, in terms of understanding herding in macroeconomic forecasting. Relative to the aggregate consumers' expenditure literature, there are a number of key differences. Firstly, Hall's hypothesis had to rely upon the hypothetical notion of the 'representative consumer' in order to identify aggregate consumers' expenditure (and aggregate income) as the consumption and income of a given individual in a decision-setting context (maximizing intertemporal life-time utility), that is, in order that the hypothesis relates to the behaviour of a rational-expectations utility-maximizing

[4] See e.g., Hendry (1995) for a textbook account.

[5] As described by Deaton (1992, p.79), Davidson *et al.* (1978) exemplifies a tradition that stresses "the derivation of a stable predictive equation for use in macroeconometric modelling", while the papers following Hall (1978) were concerned with "discovering a theoretically coherent model of household behaviour".

individual. In our application to herding, this fiction and the difficulties it causes,[6] do not arise, as we consider forecasts, not in aggregate, but by individual survey respondent. Secondly, the forecast data we have has two dimensions: fixed-event and fixed-horizon. As explained in section 2, each lends itself naturally to either modelling (fixed-horizon) or testing (fixed-event), so that a fuller picture of macro-herding can be built up, compared to the use of either alone.

The plan of the remainder of the paper is as follows. Section 2 describes the modelling and testing approaches used to understand macroeconomic forecasting herding behaviour. Section 3 describes the forecast data, and section 4 records the results. Section 5 offers some concluding remarks.

2 Assessing Forecaster Behaviour

2.1 Modelling rolling-event forecaster behaviour

Suppose at each origin forecasts are made for 1 to h-steps ahead. This gives a series of fixed-horizon (or rolling-event) 1-step ahead forecasts, a series of 2-step ahead forecasts, up to a series of h-step ahead forecasts. Our example of modelling fixed-horizon forecasts is based on Bewley and Fiebig (2002), who considered a sequence of 3-month ahead forecasts of the interest rate made every month. For a horizon of h, their general model for addressing forecaster herding is (in our notation):

$$y^i_{t-1+h|t-1} = \alpha + \beta_{1,i} y_{t-1} + \beta_{2,i} y_{t-2} + \beta_{3,i} \overline{y}_{t-2+h|t-2} + \beta_{4,i} y^i_{t-2+h|t-2} + u^i_t. \tag{1}$$

Here, $y^i_{t-1+h|t-1}$ is respondent i's forecast of period $t-1+h$ made at time $t-1$, y_{t-1} is the period $t-1$ actual value, $\overline{y}_{t-2+h|t-2}$ is the consensus (cross-sectional average) of the forecasts made at $t-2$ of the value in $t-2+h$. We explain in detail in section 3 the timing conventions we adopt when we take the model to the data. But in brief, letting t denote the survey quarter, $_{|t-1}$ denotes a forecast from survey quarter t, where the timing denotes that the available (quarterly) data refers to the $t-1$ quarter. y_{t-1} will be known at survey quarter t, or more precisely, the advance estimate will be available, in the case of data which are subject to revision, such as GDP deflator inflation and GDP growth.

The equation is estimated separately for each i, where the variation is over t for a given h: the individual, lagged individual and consensus forecasts are all of length h.

[6] There are numerous critiques. See, for example, Muellbauer (1983). There are a raft of issues to do with the possibility that some individuals may be liquidity constrained. Campbell and Mankiw (1990) have an enterprising way of estimating the proportion of liquidity-constrained consumers in the economy, within the setup of Hall (1978).

Suppose that $h = 2$. In (1), β_1 gives the effect of new information (holding constant the lagged forecasts on the RHS, and the previous value of y, y_{t-2}), and β_3 gives the effect of the previous consensus forecasts of y_t. If we assume that $\overline{y}_{t|t-2}$ is not in the information set of individual i at time $t-2$ (as in Bewley and Fiebig (2002)), then the consensus may contain useful information in an accuracy-enhancing sense relative to $y^i_{t|t-2}$. This information may or may not be superseded by the availability of y_{t-1} at forecast origin $t-1$, depending on the nature of the data generation process.

As noted by Bewley and Fiebig (2002), a major shortcoming of this model of expectations determination is the inability to discriminate between a significant β_3 flagging strategic behaviour, versus the consensus containing useful information in terms of reducing squared-error loss. Nevertheless, (1) does tell us something about the the forecast-generation mechanism, and in response to this shortcoming, we consider (1) alongside a test for herding, for each individual i.

Irrespective of the herding question, $\beta_3 = 0$ has the interpretation that individuals act independently of others. Bewley and Fiebig (2002) term forecasters with $\beta_4 = 0$, and $\beta_3 \neq 0$ strong followers, in that they pay no regard to their own forecasts but do consider the forecasts of others, whereas if both $\beta_3 = 0$ and $\beta_4 = 0$, forecast generation is consistent with y_t following an autoregressive process: only lagged values of the variable itself are deemed relevant. Either $\beta_3 \neq 0$ or $\beta_4 \neq 0$ allows for the possibility that other-variable information (included in past individual or consensus forecasts) is relevant for forecasting y. $\beta_3 = 0$ and $\beta_4 \neq 0$ corresponds to a form of adaptive expectations.

Bewley and Fiebig (2002) suggest imposing the restriction that $\sum_{j=1}^{4} \beta_{i,j} = 1$ for all i, motivated by a consideration of the long-run relationships between the variables. This gives rise to a number of alternative parameterizations. We use the parameterization generated by $\beta_{2,i} = 1 - \beta_{1,i} - \beta_{3,i} - \beta_{4,i}$, i.e.,

$$y^i_{t-1+h|t-1} - y_{t-2} = \alpha + \beta_{1,i}\left(y_{t-1} - y_{t-2}\right) + \beta_{3,i}\left(\overline{y}_{t-2+h|t-2} - y_{t-2}\right) + \beta_{4,i}\left(y^i_{t-2+h|t-2} - y_{t-2}\right) + u^i_t. \tag{2}$$

since β_2 does not have an obvious interpretation, but could be estimated by an alternative parameterization if required.[7] As noted by Bewley and Fiebig (2002), the imposition of the restriction facilitates inference by reducing the collinearity between the variables.

[7]Note the estimates of the parameters are invariant to the parameterisation. Moreover, we use the same symbols for the parameters (and disturbance term) in the unrestricted (1) and restricted models (2) to economise on notation, but formally this requires that the restriction is true.

2.2 Testing fixed-event forecasts

A fixed-event forecasting framework is characterized by multiple forecasts of the same target at different forecast horizons. Then the revision to the forecast of period t, made between times $t-h$ and $t-h+1$, should not be systematically related to information known at the time the longer-horizon forecast was made, $t-h$. If it is, then the forecasts do not make an efficient use of the available information. Hence the null of no herding is that the revision is unrelated to the consensus view known at the time of the individual's first forecast.[8] Rather than modelling forecaster behaviour, this approach amounts to testing an orthogonality condition, and is closely-related to a Mincer-Zarnowitz (Mincer and Zarnowitz (1969), henceforth MZ) test.

To see why, note that MZ suggest running the regression:

$$y_t = \delta_0 + \delta y_{t|t-h} + u_t \tag{3}$$

where the observations range over t for a given h, and the null of optimality is that $\delta_0 = 0$ and $\delta = 1$.[9] In (3), $\{y_{t|t-h}\}$ denotes a sequence of fixed-horizon forecasts. Under the null, the covariance between the forecast error and the forecast is zero:

$$Cov\left(y_t - y_{t|t-h}, y_{t|t-h}\right) = Cov\left(\left(\delta - 1\right) y_{t|t-h} + u_t, y_{t|t-h}\right).$$

Patton and Timmermann (2012) note that the actual value y_t can be replaced by a short-horizon forecast, say, $y_{t|t-h_1}$, to give:

$$y_{t|t-h_1} = \delta_0 + \delta y_{t|t-h_2} + u_t \tag{4}$$

where $h_2 > h_1$. Now the same null ($\delta_0 = 0$ and $\delta = 1$) tests the rationality of both $y_{t|t-h_1}$ and $y_{t|t-h_2}$. We can test against a specific direction, such as herding, by using consensus forecasts in the agent's information set at $t-h_2$, for example, by including a term such as $g'_{t-h_2}\theta$ in (4), where g_{t-h_2} denotes a vector of variables in the agent's information set at time $t-h_2$. The null of rationality is then that H_0: $\delta_0 = 0$ and $\delta = 1$ and $\theta = 0$, and the specific null of no herding (or anti-herding) is that $\theta = 0$. A more parsimonious test comes from testing $\theta = 0$ in (4) but first setting $\delta = 1$ (so the LHS is the revision). This test is possible because we have two forecasts of the

[8] The consensus is often used to measure 'the views of others' but the context may instead suggest using an 'influential' forecaster.

[9] Strictly speaking, $y_{t|t-h}$ should have an i-subscript, $y^i_{t|t-h}$, to denote the forecast of period t made at time $t-h$ by forecaster i. We drop the i-subscript to economize on notation. As we run a separate regression for each i, the parameters should also have an i-script, but this too is dropped for simplicity.

same target. If instead the LHS variable was y_t, as in (3), the test would be of whether the forecast error is systematically-related to the consensus variable(s). This adds a potentially large element of uncertainty to the test. We would need to select a vintage of data to be used as actual values, y_t. US national accounts data are subject to various rounds of revisions, as described by Croushore (2011a, 2011b) and Landefeld, Seskin and Fraumeni (2008) and Fixler, Greenaway-McGrevy and Grimm (2014), *inter alia*.[10] Further, considering the revision between two forecasts controls for any deficiencies common to both horizon forecasts and should result in better tests of the herding null. This is the sense in which fixed-event forecasts facilitate the testing of herding.

We run regressions of the revision to the forecast on the difference between longer-horizon forecast and the one-period-longer consensus forecast (of the same target):

$$y_{t|t-(h-1)} - y_{t|t-h} = \phi_0 + \phi_1 \left(y_{t|t-h} - \overline{y}_{t|t-(h+1)} \right) + u_t. \tag{5}$$

In the absence of herding, the revision between two forecasts of the same target should not be systematically related to information known at the time the longer-horizon forecast was made, here $y_{t|t-h} - \overline{y}_{t|t-(h+1)}$. Here $\phi_1 < 0$ implies that if the individual's $t-h$-origin forecast was greater (less) than the consensus, the forecast would be revised down (up): a manifestation of herding behaviour. $\phi_1 > 0$ implies differences relative to the consensus are exacerbated by the forecast revision - anti-herding or 'variety-seeking'. $\phi = 0$ is no herding (or anti-herding) null. Because both $y_{t|t-h}$ and $\overline{y}_{t|t-(h+1)}$ are in the individual's information set at time $t-h$, absent herding motives they ought not influence the revision. The consensus forecast $\overline{y}_{t|t-(h+1)}$ is made the period before the individual forecasts $y^i_{t|t-h}$ and thus $\overline{y}_{t|t-(h+1)} \in \mathcal{I}^i_{t-h}$, and by construction $y^i_{t|t-h} \in \mathcal{I}^i_{t-h}$, where $\mathcal{I}^i_{t-h}$ is forecaster i's information set (at time $t-h$). A case could be made for supposing in addition $\overline{y}_{t|t-h} \in \mathcal{I}^i_{t-h}$, for all i, if it were the case that respondents knew 'the consensus' from sources other than the SPF (such as reports in the media).

Relative to section 2.1 we are not seeking to model the expectations formation process, but instead test a simple orthogonality condition. Note that (5) is completely uninformative about the determinants of the forecast $y_{t|t-(h-1)}$ (under the null): and in any case is uninformative about the role of information which becomes available after the earlier forecast is made at $t-h$. Of course, we might reject $\phi_1 = 0$ in (5) if the explanatory variable happens to be correlated with any other variables which help explain the forecast revision, but are omitted from the test regression ("omitted variable bias"). Consequently, a rejection of $\phi_1 = 0$ does not necessarily provide strong evidence for (anti-)herding. Our approach is to consider the findings of tests based on (5) alongside

[10] The initial ('advance') estimates are based on partial data, in an attempt to provide timely estimates, and are revised as more complete data becomes available.

evidence based on (2) which includes other potential determinants.

2.3 Using both fixed- and rolling-event forecasts: a composite model

In principle, a composite model could be formed. For $h = 2$ in (1), the terms in the second line add in the testing approach terms, and give a composite model:

$$\begin{aligned} y^i_{t+1|t-1} &= \alpha + \beta_{1,i} y_{t-1} + \beta_{2,i} y_{t-2} + \beta_{3,i} \overline{y}_{t|t-2} + \beta_{4,i} y^i_{t|t-2} + u^i_t \\ &\quad + \phi_0 y^i_{t+1|t-2} + \phi_1 \left(y^i_{t+1|t-2} - \overline{y}_{t+1|t-3} \right). \end{aligned} \tag{6}$$

The model includes the fixed-horizon

and fixed-event analyses terms. Setting $\phi_0 = 1$ would define the LHS variable as the fixed-event revision. However, the model has many parameters for the samples of the size available for many individuals, and is likely to have highly collinear explanatory variables. It seems sensible to keep the analyses separate.

3 Description of Forecast Data

We use the US Survey of Professional Forecasters (SPF) as our source of expectations. It is a quarterly survey of macroeconomic forecasters of the US economy, providing a record of expectations from 1968 to the present day. It began life as the NBER-ASA survey in 1968:4, and since June 1990 has been run by the Philadelphia Fed, renamed as the Survey of Professional Forecasters (SPF): see Zarnowitz (1969) and Croushore (1993). Partly because of its length, it is a popular choice for academic research on expectations. As of February 2018, the Academic Bibliography maintained by the Philadelphia Fed listed 101 research papers based on the SPF forecast data.[11]

SPF respondents provide forecasts of a number of macroeconomic variables, and we analyze the forecasts of real GDP growth and (the GDP-deflator measure of) inflation. We analyze the 181 quarterly surveys from 1968:Q4 to 2013:Q4. Prior to 1981:3 the point predictions for output referred to nominal output, but a series for real output has been imputed (by the Philadelphia Fed) from the forecasts of nominal output and the deflator.

When actual values are used in the modelling of expectations formation, as in the Bewley-Fiebig approach, we use the vintages of the actuals available at the time the LHS forecast is made. These are taken from the Real Time Data Set for Macroeconomists (RTDSM) run by the Federal Reserve

[11] See http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/academic-bibliography.cfm.

Bank of Philadelphia (see Croushore and Stark (2001)), and seems preferable to using a vintage from many years later as this will typically contain revisions and definitional changes (see e.g., Landefeld *et al.* (2008) for a discussion of the revisions to US national accounts data). The survey reporting date is around the middle of the quarter, so that the respondents will have access to the advance (first) estimate for output and inflation in the previous quarter. In terms of (1) (reprinted here):

$$y^i_{t-1+h|t-1} = \alpha + \beta_{1,i} y_{t-1} + \beta_{2,i} y_{t-2} + \beta_{3,i} \overline{y}_{t-2+h|t-2} + \beta_{4,i} y^i_{t-2+h|t-2} + u^i_t. \tag{7}$$

these are the advance estimate of y_{t-1} and the first-quarterly revised value of y_{t-2}. As an illustration, suppose respondent i made forecasts in response to survey quarters 1990Q2 (this is $y^i_{t-2+h|t-2}$) and 1990Q3 (this is $y^i_{t-1+h|t-1}$). So here 't' refers to the survey quarter 1990Q3, and the conditioning of the forecast on $t-1$ indicates that the information set for survey t runs through $t-1$. Hence the LHS forecast is made to the 1990Q3 survey, and is of the current quarter when $h = 1$, and of 1991Q3 when $h = 5$, etc.. The RHS individual forecast is made in response to the 1990Q2 survey, and is of the same length as the LHS forecast. The consensus forecast matches the dating of the RHS individual forecast, but may not be known until the results are published. Irrespective of the forecast horizon, the actual values y_{t-1} and y_{t-2} are the advance estimate of the variable's value in 1990Q2 and second-quarterly estimate of the value in 1990Q1, as these are the latest vintages of these observations available to the forecaster at the time of the 1990Q3 survey.

For the fixed-event analysis, we have 4 separate sets of pairs of forecasts of the same event: the $h = 2$ and $h = 1$ forecasts, the $h = 3$ and $h = 2$ forecasts, the $h = 4$ and $h = 3$ forecasts, and finally, the $h = 5$ and $h = 4$ forecasts. For example, the first pair of $h = 2$ and $h = 1$ forecasts are the 1968:Q4 survey $h = 2$ forecast and the 1969:Q1 survey $h = 1$ forecast, respectively, both of the value of the variable in 1969:Q1. The last pair are the 2012:Q4 $h = 2$ forecasts and the 2013:Q1 $h = 1$ forecasts (both of 2013:Q1). We compare the fixed-event h and $h - 1$ forecasts to the rolling-event h-step forecasts, where $h = 2, \ldots, 5$.

4 Empirical Findings

Our examination of the evidence for herding involves the estimation of the model of forecaster behaviour, equation (2), and testing for forecaster herding, based on equation (5). We compare the two when (5) is estimated with the consensus matching the dating of the RHS individual forecast, and when it is estimated using the lagged consensus. Use of the lagged consensus is the conservative strategy, in that the lagged consensus will be known to agents from the published survey results, but might represent stale information and not act as a focal point for forecasters who herd on more up-to-date information. The use of the current consensus might falsely indicate herding, as defined here, if it was not known when the earlier forecast was made. This is because forecast efficiency in the Mincer-Zarnowitz sense requires that the forecast error is not predictable from information available at the time the forecast is made. The forecast error will naturally be correlated with relevant information not known at the time the forecast was made. In terms of (5), the forecast revision stands for the forecast error, and the current consensus forecast will be relevant new information, if not already in the individual's information set at time $t-h$.

Table 1 reports the (cross-sectional) medians and standard deviations of the estimates of β_1, β_3, and β_4, as well as the percentage of times we reject the null that the population parameters are zero, as well as the rejection frequency of $\beta_3 = 0 \cap \beta_4 = 0$. It is evident that for both output growth and inflation β_1 tends to be small (the medians are in the range -0.02 to -0.06 and 0.00 to 0.12, for output growth and inflation, respectively). Forecasters appear to place relatively little weight on the most recent data releases when they make their forecasts.

However, the estimates of β_3 tend to be relatively large and positive (as evidenced by the median and standard deviation), and statistically significant for the majority of respondents (see the rejection frequencies of the null of $\beta_3 = 0$ at the 5% level). For output growth we reject $\beta_3 = 0$ for in excess of 80% of regressions (and around 90% or more for longer horizons), and rejection rates for inflation are also high. β_4 tends to be smaller in magnitude, but the null that both β_3 and β_4 are zero is rejected in virtually all cases. The consensus term in the regression is an important determinant of the current forecast.

Hence in terms of the Bewley and Fiebig (2002) classification of different forecaster types, there appears to be little evidence that forecasters are purely autoregressive (we reject $\beta_3 = 0 \cap \beta_4 = 0$): they use information besides that contained in past values of the variable. Moreover, there is evidence against adaptive expectations behaviour (we reject $\beta_3 = 0$).

To answer the question of whether the significance of β_3 suggests strategic behaviour, we turn to the tests based on the fixed-event forecasts. Table 2 shows evidence of herding for around half of the forecasters using the lagged consensus, and for in excess of 80% of forecasters when the current consensus is used. These percentages are based on the rejections of the null that $\phi_1 = 0$ in favour of $\phi_1 < 0$ at the 5% level (i.e., a one-sided test). There is some evidence of more herding as the horizon lengthens when the lagged consensus is used as the candidate focal point. Of the 50-60 forecasters, there is only evidence of anti-herding (at the 5% level in a one-sided test) for 1 or 2 forecasters.

The results in table 2 suggest that the rejections of β_3 in the model of forecaster behaviour may indicate strategic behaviour, rather than the consensus providing useful information.[12]

However, for a sharper interpretation, we compare the results of the two testing procedures for each individual. For each respondent for whom we reject $\beta_3 = 0$, we look to see whether we also reject $\phi_1 = 0$. That is, we consider the proportion of respondents for which a rejection of $\beta_3 = 0$ can be interpreted as 'herding', in the sense of strategic behaviour, as indicated by $\phi_1 \neq 0$ in the fixed-event regression. [13]

Using the conservative approach to estimating ϕ_1, column (14) indicates that a half or more of inflation forecasters with $\beta_3 \neq 0$ are in fact acting strategically (i.e., exhibit herding behaviour) on the basis of the test, with a higher fraction at the longer horizons. For output growth the fraction acting strategically is below a half at the shorter two horizons, but around two thirds at the longest horizon. [14]

When we drop the conservative assumption that the most recent consensus view available to the forecasters is from the previous quarterly survey, there are more forecasters for whom significant values of β_3 are consistent with strategic behaviour: the difference is particularly marked for output growth at the two shorter horizons. The fractions are appreciably higher and less sensitive to the forecast horizon, and are generally in excess of three quarters for both variables. It is tempting to put more store on the conservative approach, because the current SPF-survey consensus will not be known at the time an individual submits their forecast. But this is to take a narrow view of the information available to an individual, who will surely be aware of the prevailing view regarding the economic outlook from sources other than the survey itself.

In summary, for output growth forecasts at $h = 2$, 3, the assumption concerning whether the respondents know the 'current' consensus (at the time the longer-horizon forecast is made) is key in the assessment of whether a rejection of $\beta_3 = 0$ in the model points to herding or not. For inflation, this assumption turns out to matter less. Except for $h = 2$, we find that a rejection of $\beta_3 = 0$ signifies herding for at least 70% of the respondent.

Also of interest is whether other characteristics of the forecast generation mechanism given by (2) tend to be correlated with strategic behaviour. Of interest might be the tendency to consider sources of information other than lagged values of the process. However, since we almost always reject $\beta_3 = 0 \cap \beta_4 = 0$, the results for rejecting $\phi_1 = 0$ conditional on this event will be similar to the rejection frequencies in table 2.

[12] A potentially important caveat is that regressions such as equation (5) have been shown to reject the null for reasons other than herding: see Clements (2018).

[13] Formally, the calculation is $\sum_i \left[1\left(\phi_{i,1} \neq 0\right).1\left(\beta_{i,3} \neq 0\right)\right] / \sum_i 1\left(\beta_{i,3} \neq 0\right)$, where $1\left(\beta_{i,3} \neq 0\right) = 1$ when the test of (e.g.,) $\beta_{i,3} = 0$ rejects, and $1\left(\beta_{i,3} \neq 0\right) = 0$ when the test of $\beta_{i,3} = 0$ fails to reject.

[14] When we consider the fixed-event tests using the lagged consensus, the longest horizon forecasts ($h = 5$) are used to generate the lagged consensus, so we lose the $h = 5$ comparison.

However, we can see whether individuals who are influenced by their past history of forecasts (i.e., $\beta_4 \neq 0$) are more or less likely to behave strategically. These results are recorded in columns (15) and (16) of table 1. For output growth, for individuals for whom $\beta_4 \neq 0$, whether strategic behaviour is suggested or not again depends on the assumption about the respondents' information sets. For inflation the results suggest forecasters who pay heed to their own past forecasts are less likely to act strategically compared to those for whom we reject $\beta_3 = 0$.

5 Conclusions

By using survey respondents' fixed-horizon forecasts, and fixed-event forecasts, to both model forecaster behaviour (with a term reflecting potential herding) and to simultaneously test for herding behaviour, we are able to examine whether forecasters tend to herd. We find some evidence that survey respondents' forecasting behaviour differs between the two variables we analyse, (GDP deflator) inflation and output growth. For inflation, the results are relatively unequivocal, in the sense that they do not reply upon whether respondents know the consensus view at the time they make their initial forecasts. For more than a half of the respondents, the consensus is an important determinant of their forecasts *and* suggests herding behaviour.[15] For output growth, the apparent prevalence of herding is more dependent on the assumption about what the respondents know. Under the conservative assumption, for example, fewer than a half of the rejections of $\beta_3 = 0$ denote strategic behaviour for $h = 2$ and $h = 3$.

As noted in the Introduction, the modelling approach seeks to obtain a model that can be used for prediction, and the ability to forecast 'out-of-sample' is often regarded as the acid test of a model (see, e.g., Clements and Hendry (2005) for a discussion). We do not pursue that here given the relatively small numbers of forecasts available for a given individual: no individual is ever-present, and as well as 'late' entry and 'early' exit relative to our sample period (1968-2013), there is non-participation by otherwise active respondents.[16] Hence the results we report are from in-sample regressions, and we do not attempt to estimate models with a hold-out sample which can subsequently be used to assess the out-of-sample 'forecasts of the forecasts'. This seems less important in our context, as it would not be directly informative about whether the consensus term in the model indicates herding, although it might bear on whether the consensus terms should be included in the model.

[15] From table 1, we find $\beta_3 \neq 0$ for a proportion of 0.89 of forecasters at $h = 4$, for inflation, and for 0.82 of these, we reject $\phi_1 = 0$, suggesting that for $0.82 \times 0.89 = 0.73$ there is evidence of herding. This assumes the current consensus is known. Using the lagged consensus, the calculation is $0.70 \times 0.89 = 0.62$.

[16] Engelberg, Manski and Williams (2011) provide an indication of the degree of change in the composition of the US SPF panel.

Appendix: Econometric Issues

Firstly, we can show that the error terms in the fixed-event regression (5) are uncorrelated under the null, obviating the need for autocorrelation-consistent standard errors, so that inference is based on OLS with the standard formulae.

To see this, suppose the time series y_t is written as an infinite-order moving average:

$$y_t = \psi\left(L\right)\varepsilon_t = \sum_{j=1}^{h} \psi_{h-j}\varepsilon_{t-h+j} + \psi_h \varepsilon_{t-h} + \psi_{h+1}\varepsilon_{t-(h+1)} + \sum_{j=0}^{\infty} \psi_{h+2+j}\varepsilon_{t-(h+2+j)}$$

Written this way, it follows immediately that $y_{t|t-h} \equiv E\left(y_t \mid \mathcal{I}_{t-h}\right) = \psi_h \varepsilon_{t-h} + \sum_{j=0}^{\infty} \psi_{h+1+j}\varepsilon_{t-h-1-j}$, and $y_{t|t-(h+1)} \equiv E\left(y_t \mid \mathcal{I}_{t-(h+1)}\right) = \sum_{j=0}^{\infty} \psi_{h+1+j}\varepsilon_{t-h-1-j}$, so that under the null that $\phi_1 = 0$ (and $\phi_0 = 0$), the LHS of (5) is $\psi_h \varepsilon_{t-h}$. The adjacent error in the regression for $t+1$ is $\psi_h \varepsilon_{t+1-h}$, so that the regression equation error terms are uncorrelated by virtue of $\{\varepsilon_t\}$ being white noise. (We have dropped the i subscripts).

Secondly, the errors in the fixed-horizon forecast regression (1) would be expected to be autocorrelated for multi-period horizons even for optimal forecasts. For this reason, the results reported in table are based HAC estimates of the variance-covariance matrix of the estimated parameters, taking care to handle the missing values correctly. Some experimentation suggested the results were largely qualitatively unchanged if standard OLS estimates were used instead.

Table 1: Assessment of Forecaster Behaviour based on Rolling and Fixed-Event Forecasts (equations (2) and (5))

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
h	No. Regns	Results from equation: Eq. (Rolling-Event Forecasts)										Eqs. (Rolling and Fixed)			
		β_1		β_3		β_4		$\beta_1 = 0$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_3 = 0 \cap \beta_4 = 0$	$\phi_1 \neq 0 \mid \beta_3 \neq 0$		$\phi_1 \neq 0 \mid \beta_4 \neq 0$	
		Median	Std.	Median	Std.	Median	Std.								
Output growth															
2	52	-0.03	0.16	0.87	0.31	0.18	0.35	0.19	0.88	0.33	1.00	0.80	0.39	0.82	0.29
3	52	-0.02	0.13	0.73	0.28	0.27	0.23	0.31	0.98	0.46	1.00	0.73	0.41	0.71	0.38
4	52	-0.03	0.19	0.76	0.32	0.29	0.24	0.29	0.96	0.54	1.00	0.78	0.70	0.82	0.78
5	47	-0.06	0.15	0.66	0.30	0.36	0.22	0.30	0.91	0.68	1.00	0.79		0.72	
Inflation															
2	57	0.12	0.17	0.66	0.38	0.29	0.31	0.39	0.77	0.47	0.98	0.77	0.50	0.70	0.52
3	57	0.07	0.16	0.60	0.31	0.25	0.28	0.30	0.86	0.56	0.96	0.78	0.63	0.63	0.47
4	57	0.07	0.23	0.61	0.48	0.33	0.30	0.42	0.89	0.67	1.00	0.82	0.70	0.68	0.61
5	48	0.00	0.29	0.64	0.58	0.35	0.36	0.17	0.90	0.67	1.00	0.91		0.78	

$$y^i_{t-1+h|t-1} - y_{t-2} = \alpha + \beta_{1,i}\left(y_{t-1} - y_{t-2}\right) + \beta_{3,i}\left(\overline{y}_{t-2+h|t-2} - y_{t-2}\right) + \beta_{4,i}\left(y^i_{t-2+h|t-2} - y_{t-2}\right) + u^i_t. \qquad \text{(Rolling-Event Forecasts)}$$

and either(for the current consensus):

$$y^i_{t|t-(h-1)} - y^i_{t|t-h} = \phi_0 + \phi_1\left(y^i_{t|t-h} - \overline{y}_{t|t-h}\right) + u^i_t. \qquad \text{(Fixed-Event Forecasts (current consensus))}$$

or (for the lagged consensus):

$$y^i_{t|t-(h-1)} - y^i_{t|t-h} = \phi_0 + \phi_1\left(y^i_{t|t-h)} - \overline{y}_{t|t-(h+1)}\right) + u^i_t. \qquad \text{(Fixed-Event Forecasts (lagged consensus))}$$

The entries in the table (columns (3) to (8)) are the cross-section medians and standard deviations (Std.) of the parameter estimates. Columns (9) to (12) are the rejection frequencies (at the 5% significance level) of the specified null hypotheses. Column (13) estimates the probability of finding ϕ_1 is non-zero when β_3 is non-zero, when ϕ_1 is estimated using the current consensus. Column (14) estimates the same probability but when ϕ_1 is estimated using the lagged consensus. Columns (15) and (16) estimate the same probabilities, but conditioning on $\beta_4 \neq 0$ (rather than $\beta_3 \neq 0$).
We carry out a test when there are 30 or more pairs of forecasts for an individual respondent.

Table 2: Tests of herding based on ϕ_1 in equation (5)

h	No. Regns	$\phi_1 < 0$	$\phi_2 > 0$
Current consensus			
Output growth			
2	52	0.85	0.02
3	52	0.85	0.02
4	52	0.87	0.02
5	47	0.81	0.02
Inflation			
2	57	0.84	0.02
3	57	0.79	0.02
4	57	0.81	0.02
5	48	0.88	0.02
Lagged consensus			
Output growth			
2	52	0.40	0.02
3	52	0.52	0.04
4	49	0.78	0.02
Inflation			
2	57	0.63	0.02
3	57	0.65	0.02
4	51	0.73	0.02

The test regression is:

$$y^i_{t|t-(h-1)} - y^i_{t|t-h} = \phi_0 + \phi_1 \left(y^i_{t|t-h} - \overline{y}_{t|t-h} \right) + u^i_t$$

when the current consensus is used, and for the lagged consensus is:

$$y^i_{t|t-(h-1)} - y^i_{t|t-h} = \phi_0 + \phi_1 \left(y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} \right) + u^i_t$$

The entries in the table (columns 3 and 4) are the proportion of regressions for which the null is rejected in favour of the specified alternative, when the tests are applied at the 5% level.
We carry out a test when there are 30 or more pairs of forecasts for an individual respondent.

References

Banerjee, A. V. (1992). A simple model of herd behavior. *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**, 169–181.

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

Campbell, J. Y., and Mankiw, N. G. (1990). Permanent income, current income, and consumption. *Journal of Business and Economic Statistics*, **8**, 265–279.

Clements, M. P., and Hendry, D. F. (2005). Evaluating a model by forecast performance. *Oxford Bulletin of Economics and Statistics*, **67**, 931–956.

Clements, M. P. (2018). Do macroforecasters herd?. *Journal of Money, Credit and Banking*, *50*(2-3), 265–292.

Croushore, D. (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review, November*, 3–15.

Croushore, D. (2011a). Forecasting with real-time data vintages, chapter 9. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 247–267: Oxford University Press.

Croushore, D. (2011b). Frontiers of real-time data analysis. *Journal of Economic Literature*, **49**, 72–100.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105(1)**, 111–130.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692.

Deaton, A. (1992). *Understanding Consumption.* Oxford: Clarendon Press.

Engelberg, J., Manski, C. F., and Williams, J. (2011). Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition. *Journal of Applied Econometrics*, *26*(7), 1059–1078.

Fixler, D. J., Greenaway-McGrevy, R., and Grimm, B. T. (2014). The revisions to GDP, GDI, and their major components. *Survey of Current Business*, **August**, 1–23.

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature*, **75**, 450–451.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Evidence. *Journal of Political Economy*, **86**, 971–987.

Hendry, D. F. (1995). *Dynamic Econometrics.* Oxford: Oxford University Press.

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

Landefeld, J. S., Seskin, E. P., and Fraumeni, B. M. (2008). Taking the pulse of the economy. *Journal of Economic Perspectives*, **22**, 193–216.

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

Mincer, J., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In Mincer, J. (ed.), *Economic Forecasts and Expectations*, pp. 3–46. New York: National Bureau of Economic Research.

Muellbauer, J. N. J. (1983). Surprises in the consumption function. *Economic Journal*, **Conference Supplement**, 34–49.

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

Patton, A. J., and Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, *30*(1), 1–17.

Satopää, V. A. (2018). Combining information from multiple forecasters: Inefficiency of central tendency. mimeo, INSEAD, Technology ad Operations Management.

Zarnowitz, V. (1969). The new ASA-NBER Survey of Forecasts by Economic Statisticians. *The American Statistician*, **23, No. 1**, 12–16.