Do macro-forecasters herd?

Article

Accepted Version

Clements, M. P. ORCID: https://orcid.org/0000-0001-6329-1341 (2018) Do macro-forecasters herd? Journal of Money, Credit and Banking, 50 (2-3). pp. 265-292. ISSN 1538-4616 doi: https://doi.org/10.1111/jmcb.12460 Available at https://centaur.reading.ac.uk/72352/

Do Macro-Forecasters Herd?

Michael P. Clements*

ICMA Centre, Henley Business School, University of Reading

August 30, 2017

Abstract

We show that typical tests of whether forecasters herd will falsely indicate herding behaviour for a variety of types of behaviour and forecasting environments that give rise to disagreement amongst forecasters. We establish that forecasters will appear to herd if differences between them reflect noise as opposed to private information, or if they arise from informational rigidities. Noise can have a behavioural interpretation, and if so will depend on the behavioural model under consideration. An application of the herding tests to US quarterly survey forecasts of inflation and output growth data 1981-2013 does not support herding behaviour.

JEL Classification: E37.

*Michael Clements is Professor of Econometrics at the ICMA Centre, Henley Business School, University of Reading, and a 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. I am grateful for constructive comments from two reviewers which led to marked improvements in the paper. This paper has also benefitted from tthe comments of participants at the ECB-CFS-Bundesbank Joint Lunchtime Seminar, and the Warwick-Monash Workshop on Macro Forecasting. An earlier version of this paper was circulated as 'Assessing the Evidence for Macro-Forecaster Herding: Forecasts of Inflation and Output Growth'.

Keywords: macro-forecasting, (anti-)herding, private information, idiosyncratic errors.

1 INTRODUCTION

The recent literature suggests that forecasters may have incentives other than to produce the most accurate forecast possible. That is, the traditional assumption that the reported forecast has the minimum expected squared forecast error neglects other motivations that may also affect the forecasters' payoffs. As noted by Lamont (2002), for example, forecasters may set their forecasts to 'optimize profits or wages, credibility, shock value, marketability, political power...' (see Laster et al. (1999) and Ottaviani and Sorensen (2006), *inter alia*). Some of the empirical literature on assessing the influence of reputation and related factors on the determination of agents' forecasts rests 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 towards, or away from the consensus view, in a way which is detrimental to forecast accuracy.

A review of the recent macro-forecasting literature suggests that herding behaviour (including anti-herding behaviour[1]) is commonplace: see Table 1. At the same time, in recent years there has been much innovative work on expectations formation, and various theories of forecaster behaviour have been proposed to explain the observed heterogeneity in expectations, which of course is not consistent with the full-information rational expectations (FIRE) hypothesis, in which there is no place for disagreement amongst forecasters.

The question we ask is whether forecasters predominantly herd (or anti-herd) or whether instead the positive results from testing for herding in the literature are due to other (non-herding) types of behaviour which falsely show up as herding. Hence we investigate the properties of tests of herding in the presence of some of the (non-herding) forecaster behaviours which have been proposed in the recent literature. The tests we consider are essentially reduced form, in that they are not linked to specific types of behaviour, but are formulated such that the null of no (anti-)herding is a simple exclusion restriction in a regression. The tests we analyze are based on or adapt extant approaches to testing forecaster herding.

It is clear that testing for herding behaviour is unlikely to be straightforward, and might falsely signal herding. Individuals' forecasts will tend to cluster together for reasons other than herding, including for example a sharing of common information about the likely future evolution of the variable of interest. We show that other forms of forecasters behaviour will also (falsely) indicate herding.

We begin with a simple framework which sets out the characteristics of forecaster behaviour which determine the outcomes of the tests of forecasting we consider. We then describe some of the leading theory models in the recent literature which give rise to disagreement among forecasters, and show how these models relate to the simple framework. Hence we are able to distill the various theories of behaviour into a small number of characteristics which determine the outcome of the tests of herding. This means that the herding-detection implications of any new theories that might be proposed can be deduced from this framework. The theory models of forecaster behaviour include informational rigidities, agents with heterogeneous degrees of asymmetry of their loss functions, agents differing in terms of their priors or long-term expectations, in terms of their forecasting models, and in terms of their signals and interpretation of those signals.

The key aspects or characteristics which determine the herding test outcomes are whether forecaster disagreement reflects private information, or 'noise', and whether the history of the variable is observed or remains unobserved and latent. Hence our simple model features private information and noise or idiosyncratic errors. None of the forecasters pay attention to the views of others, so there is no herding. Noise is not intended to simply signify reporting error. It is a source of forecast dispersion which has a behavioural interpretation within the context of the model of expectations formation under consideration. The term noise is a catch-all for differences between forecasters which are unrelated to private information, and consequently the individual forecast errors and forecasts are correlated for forecasts which contain noise components, but not for individual forecasts driven by private information. Private information is assumed to be handled rationally in that forecast errors and forecasts are not correlated for forecasts embedding private information. One might wonder whether macro-forecasters are privy to useful information not available to others, and this suggests that there is a *prima facie* case that differences between macro-forecasters reflect what we term noise, and so are likely to falsely indicate herding. Regardless of one's view on this, the simple model captures the essential features of the forecasting environment in terms of predicting when the herding tests will reject the no-herding null.

We then illustrate our findings with the US Survey of Professional Forecasters (SPF) output growth and inflation forecasts. We deduce from the pattern of testing results we obtain that there is little evidence of concerted (anti-)herding behaviour. The forecasting environment appears to be characterized by agents whose forecasts primarily differ because of idiosyncratic factors, or because of private information with the past values of the variables being forecasting remaining unobserved. We are unable to distinguish between the different theories of forecaster behaviour, as many give rise to the same underlying factors which determine herding test outcomes. However choosing between the different theories is not our purpose: many of the papers extolling the different forms of behaviour assess the evidence for the superiority of their ideas (see, e.g., Coibion and Gorodnichenko (2012, 2015), Giacomini et al. (2015)).

Some of the recent papers focus on the implications for (cross-section) aggregate measures, for example, the responses of mean errors and forecaster dispersion to shocks. In part this is necessary in that some of the implications of the theories are emergent properties in that they are only apparent at the aggregate level.[2] We choose to analyze the individual forecasters rather their behaviour *en masse* in terms of aggregate quantities. This allows us to make statements such as, say, $x\%$ of forecasters appear to herd (using test T_1, say) rather than testing a hypothesis that forecasters taken together behave in a certain way.

The plan of the paper is as follows. Section 2 describes a stylized model with heterogeneity resulting from private information or idiosyncratic errors. A number of theories of forecaster behaviour are then discussed (section 2.2), and we show how some of these models can be subsumed within the stylized model, at least in terms of whether they will give rise to apparent herding behaviour. Section 3 describes the tests of forecaster herding, and derives the population outcomes under different models of behaviour using the stylized model of section 2.1. Section 4 records the results of applying the tests of herding to the US SPF output growth and inflation expectations, and weighs the evidence for herding based on the analyses of section 3. Section 5 concludes. A separate appendix provides the technical details.

2 MODELLING FORECASTER DISAGREEMENT

We first present a simple framework consisting of a stylized model which captures the essence of some of the behavioural models (discussed in section 2.2) in terms of determining the outcomes of the herding tests described in section 3.1. The model of heterogeneous forecasters allows roles for private information, and idiosyncratic error or noise.

2.1 *Stylized Model Containing Key Features of Various Theory Models*

The variable y_t is assumed to be generated by a first-order autoregressive process:

$$y_t = \delta y_{t-1} + w_t. \tag{1}$$

At $t-1$, each agent observes y_{t-1} and a private signal s_{it}, where $s_{it} = y_t + \varepsilon_{it}$. Hence the optimal forecast (in a minimum mean-squared error sense) is given by:

$$y^i_{t|t-1} = (1-\lambda_i)\,\delta y_{t-1} + \lambda_i s_{it}$$

which can be written alternatively as:

$$y^i_{t|t-1} = \lambda_i w_t + \delta y_{t-1} + \lambda_i \varepsilon_{it}$$

where $\lambda_i = \sigma^2_w / \left(\sigma^2_w + \sigma^2_{\varepsilon_i}\right)$. Here, $\sigma^2_w = E\left(w^2_t\right)$, $\sigma^2_{\varepsilon_i} = E\left(\varepsilon^2_{it}\right)$, and $E\left(w_t\right) = 0$, $E\left(\varepsilon_{it}\right) = 0$. Below we assume that $\sigma^2_{\varepsilon_i} = \sigma^2_{\varepsilon}$, for all i, so that $\lambda_i = \lambda$ for all i. We also assume that the ε_{it}'s are serially uncorrelated, are uncorrelated across individuals at all leads and lags, and are uncorrelated with the w_t's at all leads and lags. In addition, we are implicitly assuming that the agent's $t-1$ information set includes y_{t-1} (and additional lags if the data were generated by a higher-order autoregressive process). This is a simple way of generating disagreement, and is consistent with agents behaving rationally, although the source or substance of the private information in a macro-forecasting context is unclear.

We assume that agents only receive a signal about the next period (relative to the forecast origin). Hence the 2-step forecast iterates forward the 1-step forecast using equation (1), i.e.,

$$y^i_{t|t-2} = \delta y^i_{t-1|t-2}.$$

with:

$$y^i_{t-1|t-2} = (1-\lambda)\,\delta y_{t-2} + \lambda s_{it-1}.$$

When calculating consensus forecasts, we assume $N^{-1}\sum_i \varepsilon_{it} = 0$, so that e.g.,

$$\overline{y}_{t|t-1} = \lambda w_t + \delta y_{t-1}.$$

Much of the recent literature suggests a role for an idiosyncratic error in explaining forecaster behaviour.[3] We alternatively refer to this as 'noise'. We use noise as a catch-all term to encompass the implications of a broad range of forecasting behaviours, as outlined in section 2.2. The noise error v is specific to the forecaster i, the target t, and the forecast horizon h, denoted $v_{i,t|t-h}$ for an h-step ahead forecast of y_t. Hence the 1-step ahead forecast by agent i becomes:

$$y^i_{t|t-1} = \lambda w_t + \delta y_{t-1} + \lambda \varepsilon_{it} + v_{i,t|t-1} \tag{2}$$

where $E\left(v^2_{i,t|t-h}\right) = \sigma^2_v$ for all i, t and h, and $E\left(v_{i,t|t-h}, v_{i_1,t_1|t_1-h_1}\right) = 0$ whenever one of $i_1 \neq i$, $t_1 \neq t$, $h_1 \neq h$ is true. We also assume that ε and v are uncorrelated across individuals and time, including $E\left(\varepsilon_{it} v_{i,t|t-1}\right) = 0$. We assume $N^{-1}\sum_i v_{it,t-h} = 0$ so that the consensus is unaffected by noise.

Under these assumptions, the h-step ahead forecast of y_t by individual i is given by their conditional expectation of y_t given the available information $\mathcal{I}^i_{t-h} = (y_{t-h}, s_{i,t-h+1})$, plus the 'noise', $v_{i,t|t-h}$, i.e.,

$$y^i_{t|t-h} = E\left(y_t \mid \mathcal{I}^i_{t-h}\right) + v_{i,t|t-h},$$

where

$$\begin{aligned} E\left(y_t \mid \mathcal{I}^i_{t-h}\right) &= \delta^{h-1} y^i_{t-h+1|t-h} \\ &= \delta^{h-1}\left(\lambda w_{t-h+1} + \delta y_{t-h} + \lambda \varepsilon_{it-h+1}\right). \end{aligned}$$

In this model, $\overline{y}_{t|t-h} \notin \mathcal{I}^i_{t-h}$, because $\mathcal{I}^i_{t-h}$ does not include $\left\{s_{j,t-h+1|t-h}\right\}$ $j = 1, \ldots, N$ except for $j = i$.

To summarize: in our simple model forecaster disagreement can be driven by private information or noise. A key distinction between the two is that the former is consistent with forecaster efficiency, in the sense that the forecasts are uncorrelated with the forecast errors,[4] whereas noise results in forecasts which violate this property. For example, the 1-step ahead forecast error is:

$$\begin{aligned} y_t - y^i_{t|t-1} &= \delta y_{t-1} + w_t - \left(\lambda w_t + \delta y_{t-1} + \lambda \varepsilon_{it} + v_{it|t-1}\right) \\ &= w_t\left(1-\lambda\right) - \lambda \varepsilon_{it} - v_{it|t-1} \end{aligned}$$

so that:

$$\begin{aligned} Cov\left(y_t - y^i_{t|t-1}, y^i_{t|t-1}\right) &= E\left[\left(w_t\left(1-\lambda\right) - \lambda \varepsilon_{it} - v_{it|t-1}\right)\left(\lambda w_t + \delta y_{t-1} + \lambda \varepsilon_{it} + v_{it|t-1}\right)\right] \quad (3) \\ &= -\sigma^2_v, \end{aligned}$$

which is zero in the absence of noise. This derivation uses $\lambda = \sigma_w^2 / (\sigma_w^2 + \sigma_\varepsilon^2)$. Hence we interpret noise in a wide sense to denote any element of the forecast which does not enhance the (squared-error) accuracy of the forecast, or alternatively, is correlated with the forecast error. That is, the noise term captures deviations from rational forecasts (where rational is defined relative to a squared-error loss function). This gives a precise definition of the noise term in terms of its properties. [5] This definition does not preempt the noise term having a behavioural interpretation. As shown in the following section, the source, or interpretation, of the noise term will depend on the behavioural model being considered. Hence the noise term is not to be interpreted as solely reflecting 'reporting error'.

We next consider models of forecaster behaviour which have been proposed in the literature, and show how these fit within our stylized model.

2.2 *Theory Models of Forecaster Behaviour*

We briefly review the theory models that generate forecaster disagreement in sections 2.2.1 to 2.2.4. The salient features of these models from the perspective of testing for herding can be cast in the stylized model framework of section 2.1. In section 2.2.5 we consider two models of informational rigidities which do not fit neatly within our framework.

2.2.1 *Private information and non-optimal weights*

In the stylized model, we assume that the agent i's forecast (equation (2)) is generated by choosing λ to optimally weight private information. Hence the forecasts are formed efficiently: the corresponding individual forecast errors and forecasts are uncorrelated. However, non-optimal weighting (using e.g., $\widetilde{\lambda} \neq \lambda$) would result in forecaster disagreement generated in part by noise. Formally, the 1-step forecast (2) becomes (assuming no initial idiosyncratic error, $\sigma_v^2 = 0$):

$$y^i_{t|t-1} = \widetilde{\lambda} w_t + \delta y_{t-1} + \widetilde{\lambda}\varepsilon_{it} \tag{4}$$

$$= \lambda w_t + \delta y_{t-1} + \lambda \varepsilon_{it} + \overbrace{\left(\widetilde{\lambda} - \lambda\right) w_t + \left(\widetilde{\lambda} - \lambda\right)\varepsilon_{it}}^{\widetilde{v}_{i,t|t-1}} \tag{5}$$

where $\widetilde{v}_{i,t|t-1}$ is the 'induced' noise from the use of non-optimal weights. The importance of the noise component will depend on the magnitude of the divergence $\left|\widetilde{\lambda} - \lambda\right|$.

We have assumed forecasters do not weigh private information optimally, but a closely related situation would be the differential interpretation of public information, as in e.g., Kandel and Zilberfarb (1999) and Manzan (2011)).

2.2.2 *Heterogeneous beliefs about long-run outcomes*

Patton and Timmermann (2010) suggest the observed disagreement among forecasters may in part reflect the influence of different views about the long-run values of variables like inflation and output growth. Suppose that agent i's forecast based on recent information is $E\left(y_t \mid \mathcal{I}^i_{t-1}\right)$, and we sharpen the analysis by assuming, say $E\left(y_t \mid \mathcal{I}^i_{t-1}\right) = \mu_{t|t-1}$, for all i. This forecast is weighted by each forecaster with their prior for the long-run growth rate, μ_i, to give agent i's reported forecast as:

$$y^i_{t|t-1} = \omega\mu_i + (1-\omega)\,\mu_{t|t-1}.$$

In the stylized model, we suppose that agents observe y_{t-1} when forming their forecasts $y^i_{t|t-1}$, then from equation (1) $\mu_{t|t-1} \equiv \delta y_{t-1}$.[6] We further suppose $\mu_i = E\left(y_t\right) + \varepsilon_i$, $E\left(\varepsilon_i\right) = 0$ and $E\left(\varepsilon_i^2\right) = \sigma_\varepsilon^2$ for all i, where for $y_t = \delta y_{t-1} + w_t$ we have $E\left(y_t\right) = 0$. This captures the idea that agents have different long-run expectations (which are assumed to be time invariant). However it then follows that the optimal weight on μ_i in agent i's forecast is $\omega = 0$, in terms of minimizing squared-error loss. Hence forecaster heterogeneity arising from $\omega \neq 0$ will constitute idiosyncratic error or noise.

2.2.3 *Heterogeneous forecasting models*

Giacomini et al. (2015) allow for heterogeneity in models, interpreted in a wide sense to include the use of different statistical models to generate forecasts as well as judgment or incentive-driven adjustments. They implement this idea by assuming individual specific intercepts in the forecasting models. More generally, suppose in addition the autoregressive parameter of the assumed AR(1) differs across forecasters. Then in terms of our stylized model, absent private information and other forms of idiosyncratic error:

$$y^i_{t|t-1} = E\left(y_t \mid \mathcal{I}^i_{t-1}\right) = c_i + \delta_i y_{t-1}.$$

Hence model heterogeneity generates disagreement. We can write $y^i_{t|t-1} = \delta y_{t-1} + c_i + (\delta_i - \delta)\, y_{t-1}$, where the term $c_i + (\delta_i - \delta)\, y_{t-1}$ generates a non-zero correlation between the forecast error and forecast (compare to equation (3)), and plays the role of idiosyncratic error in terms of the herding test results derived in section 3.2.

2.2.4 *Heterogeneous degrees of loss asymmetry*

Capistrán and Timmermann (2009) provide an explanation for forecaster heterogeneity in terms of agents having asymmetric loss functions characterized by differing degrees of asymmetry. Consider the 'LINEX' (LINear-EXponential) loss function[7] defined on the forecast error e:

$$C(e, \varphi_i) = b\left[\exp(\varphi_i e) - \varphi_i e - 1\right], \quad \varphi_i \neq 0,\ b \geq 0$$

where for $\varphi_i > 0$, loss is approximately linear for $e < 0$ (over-predictions), and exponential for $e > 0$ (under-predictions). Suppose inflation is conditionally Gaussian (i.e., given information in the previous period), $y_{t|t-1} \sim N\left(\mu_{t|t-1}, \sigma^2_{t|t-1}\right)$, then it follows that the optimal predictor for agent i is given by:

$$y^i_{t|t-1} = \mu_{t|t-1} + \frac{\varphi_i}{2}\sigma^2_{t|t-1}. \tag{6}$$

Hence even full-information rational expectations forecasters will make biased forecasts and will disagree in terms of their reported forecasts. Optimal forecasts under heterogeneous degrees of loss asymmetry will manifest as forecasts with idiosyncratic errors or noise, when evaluated using a squared error loss function, so will tend to (falsely) indicate (anti-)herding behaviour. In terms of the stylized model, we can relate (6) to the 1-step ahead forecast given by equation (2) by noting that $\mu_{t|t-1} \equiv \delta y_{t-1}$, and $\frac{\varphi_i}{2}\sigma^2_{t|t-1}$ plays the role of $v_{i,t|t-1}$ (and in addition there is no private information, so $\lambda = 0$). The cross-sectional average of the v_i's will be zero (as in the stylized model) assuming the φ_i's are symmetrically distributed about zero, but the direct equivalence between $\frac{\varphi_i}{2}\sigma^2_{t|t-1}$ and $v_{i,t|t-1}$ fails unless we assume $E\left(\frac{\varphi_i}{2}\sigma^2_{t|t-1}\frac{\varphi_j}{2}\sigma^2_{t|t-1}\right) = 0$, for $i \neq j$..

When the disturbances in the underlying model are conditionally homoscedastic, as in the stylized model, i.e., $E\left(w_t^2 \mid \mathcal{I}_{t-1}\right) = \sigma^2_w$, then the individuals forecasts will be scattered about $\mu_{t|t-1} = \delta y_{t-1}$ but the deviations will not vary with t (because now $\frac{\varphi_i}{2}\sigma^2_{t|t-1} = \frac{\varphi_i}{2}\sigma^2_w$).

2.2.5 *Information Rigidities*

The theory models discussed hitherto have the propery that at time $t-h$, an agent's information set includes the history of the variable through y_{t-h}: $\mathcal{I}_{t-h}^{i} = \left\{ y_{t-h}, y_{t-(h+1)}, \ldots \right\}$, in addition to any private or public signals. As a consequence, the framework we have adopted implies that $\left\{ y_{t|t-(h+s)}^{j} \right\}$ for all j and for $s = 1, 2, \ldots$, does not constitute relevant information (in terms of minimizing squared error loss) for an agent forecasting y_t at time $t-h$. Consequently, aggregate summaries of that information, such as the consensus forecast defined by $\overline{y}_{t|t-(h+s)} = N^{-1} \sum_i y_{t|t-(h+s)}^{i}$, $s = 1, 2, \ldots$, have no role to play either. But when there is either of the two types of informational rigidity discussed below, $\overline{y}_{t|t-(h+s)}$ remains valuable in an accuracy–enhancing sense.

Sticky Information Sticky information assumes that in each period, each agent updates their information (relative to the previous period) with probability $1-\lambda$. When they do update, they acquire full information, and form expectations rationally: see *inter alia* Mankiw and Reis (2002) and Mankiw et al. (2003), and Coibion and Gorodnichenko (2012). In terms of forecasting y_t h-steps ahead, the mean forecast across individuals can be shown to be given by:

$$\overline{y}_{t|t-h} = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k E\left(y_t \mid \mathcal{I}_{t-(h+k)}^{i} \right)$$

where $E\left(y_t \mid \mathcal{I}_{t-s}^{i} \right)$ is the full-information rational expectations forecast made at time $t-s$ (where $s \geq h$).

Whereas we establish in section 3.2 that our tests of herding will not falsely show up as (anti-)herding when there is private information, that result is predicated on (e.g.,) $\overline{y}_{t|t-(h+1)}$ providing no useful information. This is not the case for sticky information forecasters.

For those forecasters who have not updated recently, and whose forecasts are given by $E\left(y_t \mid \mathcal{I}^i_{t-(h+k)}\right)$, for $k > 1$, the consensus contains useful information, and the tests of herding will be shown to falsely reject the no-herding null hypothesis. This is discussed further in section 3.

Noisy Information Noisy information assumes agents base their forecasts on the latest information, but only ever observe noisy signals about economic fundamentals: see Woodford (2002), Sims (2003) and Coibion and Gorodnichenko (2012), *inter alia.* Unlike sticky information models, agents are assumed to base their forecasts on the latest information, but this never reveals fundamentals (such as the inflation rate and real output growth).

The model in Coibion and Gorodnichenko (2012) assumes an AR(1) process for the unobserved state variable:

$$y_t = \delta y_{t-1} + w_t, \quad w_t \sim iidN\left(0, \sigma^2_w\right)$$

and that each agent receives a signal common to all as well as a specific signal, i.e., $z_{it} = [s_{it},\, s_t]'$ where:

$$\begin{aligned} s_{it} &= y_t + \varepsilon_{it} \\ s_t &= y_t + \eta_t, \end{aligned}$$

with $\varepsilon_{it} \sim iidN\left(0, \sigma^2_\varepsilon\right)$, $\eta_t \sim iidN\left(0, \sigma^2_\eta\right)$, $E\left(\eta_t \varepsilon_{is}\right) = 0\ \forall i, t, s$. The agent makes optimal forecasts of y_t h-steps ahead given his assumed information set, i.e., $y^i_{t|t-h} = E\left(y_t \mid z_{it-h}, z_{it-(h+1)}, \ldots\right)$, and also $z^i_{t|t-h} = E\left(z_{it} \mid z_{it-h}, z_{it-(h+1)}, \ldots\right)$. Using the Kalman filter, it follows that:

$$y^i_{t|t} = (1 - PH)\,\delta y^i_{t-1|t-1} + PHy_t + P_\varepsilon \varepsilon_{it} + P_\eta \eta_t \tag{7}$$

where $H = [1\ 1]'$, and $P = [P_\varepsilon, P_\eta] = \left[\frac{\Psi\sigma^2_\eta}{\Psi\left(\sigma^2_\varepsilon+\sigma^2_\eta\right)+\sigma^2_\varepsilon\sigma^2_\eta}, \frac{\Psi\sigma^2_\varepsilon}{\Psi\left(\sigma^2_\varepsilon+\sigma^2_\eta\right)+\sigma^2_\varepsilon\sigma^2_\eta}\right]$ with $P_\varepsilon, P_\eta \in (0, 1)$ and Ψ the variance-covariance matrix of the one-step ahead forecast error for y_t (see Coibion and Gorodnichenko (2012) for details). Note that as $\sigma^2_\varepsilon \to 0$ the degree of information rigidity regarding the private signal lessens and $P_\varepsilon \to 1$.

The forecasters have private information, in the form of private signals, but the key difference relative to the model in section 2.1 is that the history of $\{y_t\}$ remains unobserved: each forecaster i's information set consists only of his set of (private and public) signals $\{z_{it}, z_{it-1}, \ldots\}$. Consequently, the consensus forecast will always contain valuable information and this will cause the tests to (wrongly) indicate imitative behaviour.

One could envisage 'private' forecasts being based on private signals, as in equation (7), but suppose that professional forecasters would not forego the opportunity to improve forecast accuracy by 'adapting' their forecasts to the consensus, or lagged consensus. However, this would go against the grain of the noisy information model considered in the literature.

3 PROPERTIES OF APPROACHES TO TESTING FOR HERDING

3.1 *Tests*

There are a number of papers that *model* an agent's forecasts, using the consensus forecast as an explanatory factor: e.g., Batchelor and Dua (1992), Bewley and Fiebig (2002) and Gallo et al. (2002). As an example, consider Gallo et al. (2002) (GGJ), who suggest testing for herding based on the properties of successive individual forecasts of the same fixed event (y_t) and consensus forecasts of the same event, using the following regression:

$$y^i_{t|t-h} = \beta_0 + \beta_1 y^i_{t|t-(h+1)} + \beta_2 \overline{y}_{t|t-(h+1)} + u_t, \tag{8}$$

where $y^i_{t|t-h}$ denotes the forecast by individual i made at $t-h$ about y_t, and a bar (and omission of the i-subscript) denotes a consensus forecast. (For simplicity, we omit the i-subscript from the parameters). This is their equation (1), but omitting their additive term in the group variance of the $h+1$ periods ahead forecast. GGJ offer an alternative re-parameterisation, given by:

$$\begin{aligned} y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} &= \beta_0 + \beta_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)} \right) + (\beta_1 + \beta_2 - 1)\overline{y}_{t|t-(h+1)} + u_t \\ &= \beta_0 + \gamma_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)} \right) + \gamma_2 \overline{y}_{t|t-(h+1)} + u_t. \end{aligned} \tag{9}$$

Based on equations (8) and (9), they argue that γ_1 is expected to be positive ('the choice of forecast ... can be expected to be consistently on the same side of the sample mean...', p.12), and that for an 'imitation effect' we require $\beta_2 \neq 0$ (in equation (8)), and for an imitation effect and 'shrinkage to the mean' (i.e., herding), we require $\beta_2 \neq 0$ *and* $\gamma_2 = \beta_1 + \beta_2 - 1 < 0$ (in equation (9)).

However, $\overline{y}_{t|t-(h+1)}$ will not be known at the time $y^i_{t|t-(h+1)}$ is made, and may provide useful information (in an accuracy-enhancing sense). In order to obtain a cleaner test of herding, we replace $\overline{y}_{t|t-(h+1)}$ by $\overline{y}_{t|t-(h+2)}$ in equation (8) (and therefore also equation (9)) so that the consensus is known at time $t-(h+1)$ and so ought not influence the LHS forecast revision if forecasts are efficient. In the results section we check whether the inference we make regarding herding is sensitive to the omission of the forecast dispersion term.

We estimate the regressions for each individual for a given h using variation over t. The parameters are not explicitly subscripted by individual, but we stress that all the regressions we run are for a given pair of i and h, using variation over t. This allows for individuals to exhibit heterogeneous behaviour, and for that behaviour to be horizon specific: a given forecaster may consider others' forecasts at some horizons, but not at others.

Hence we estimate:

$$y^i_{t|t-h} = \beta_0 + \beta_1 y^i_{t|t-(h+1)} + \beta_2 \overline{y}_{t|t-(h+2)} + u_t \tag{10}$$

$$y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} = \beta_0 + \gamma_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)} \right) + \gamma_2 \overline{y}_{t|t-(h+2)} + u_t \tag{11}$$

for each i separately, and for each of $h = 1, 2$ and 3, using variation over t. We call this testing approach T_1, and we are interested in β_2 in equation (10) and γ_2 in equation (11).

A related approach is simply to test whether $y^i_{t|t-h} - y^i_{t|t-(h+1)}$ is negatively correlated with $y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)}$, so that if last quarter's forecast exceeds the consensus the current forecast is lowered relative to last period's. This corresponds to $\phi_1 < 0$ in:

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \phi_0 + \phi_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)} \right) + u_t \tag{12}$$

and is referred to as test T_2. While $\phi_1 < 0$ indicates herding, $\phi_1 > 0$ indicates anti-herding. This can be motivated as a test for forecaster efficiency in a fixed-event setting: see, e.g., Nordhaus (1987) and Clements (1995, 1997). Clearly, the *future* forecast revision $y^i_{t|t-h} - y^i_{t|t-(h+1)}$ (from the standpoint of the $h+1$-step forecast $y^i_{t|t-(h+1)}$) should not be predictable from information available at that time if information is being used efficiently. It is a test for herding because we test in the direction of the revision being related to the divergence of the earlier forecast from the consensus. Tests based on equation (12) can also be derived by re-parameterising (10) as:

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \beta_0 + (\beta_1 - 1) \left(y^i_{t|t-(h+1)} - \frac{\beta_2}{1-\beta_1} \overline{y}_{t|t-(h+2)} \right) + u_t \tag{13}$$

and then imposing $\beta_1 + \beta_2 - 1 = 0$, such that ϕ_1 from equation (12) corresponds to $(\beta_1 - 1)$.

The assumption that at $t-(h+1)$ the latest consensus forecast known by the forecaster is $\overline{y}_{t|t-(h+2)}$ is conservative. Although each forecaster will typically not know the current forecasts of all the others in response to the particular survey in question, it may be reasonable to assume that professional forecasters effectively have this information assuming they are cognizant of the prevailing view of the forecasting community about the outlook for key macro-aggregates such as GDP growth and inflation. Hence one might proceed as if the current consensus ($\overline{y}_{t|t-(h+1)}$) were known. Use of the lagged consensus $\overline{y}_{t|t-(h+2)}$ perhaps assumes too little - that a forecaster's view of the consensus is anchored to what it was a quarter of a year ago (in the case of the quarterly US Survey of Professional Forecasters). Such an approach may fail to reject (anti-)herding when present simply because forecasters consider the current consensus, and not the previous quarter's consensus. In the empirical work we will also consider the less conservative assumption, and in place of equation (12) we estimate:

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \phi_0 + \phi_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)} \right) + u_t, \tag{14}$$

as well as equations (8) and (9) in place of equations (10) and (11).

The final approach (T_3) is based on Bernhardt et al. (2006) (henceforth, BCK). Their approach supposes that if a forecaster's current *reported* forecast exceeds the consensus, then a forecaster who herds (anti-herds) will have moved his/her forecast towards (away from) the consensus, making it more likely that the error associated with that forecast will be positive (negative). The approach was developed in the context of assessing the behaviour of professional financial analysts, but has since been applied in a number of contexts (see, Pierdzioch et al. (2010) and Pierdzioch and Rülke (2012), *inter alia*). The approach is generally used to assess all the forecasters *en masse*, although in principle it can be applied to individual forecasters, to allow that some forecasters may herd, others anti-herd, and others do neither. Because the test is non-parametric and relies on the large-sample distribution of the test statistic, formal application of the approach to the relatively small samples of forecasts typically available for an individual forecaster may prove unreliable. For this reason we instead report a parametric, regression-based implementation of their approach. This strongly points toward anti-herding, as opposed to herding, in our sample of professional forecasters.

The test is implemented in the regression:

$$y_t - y^i_{t|t-h} = \rho_0 + \rho_1 \left(y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} \right) + w_t \tag{15}$$

The *reported* forecast of a respondent is supposed to have been moved towards the consensus relative to the respondent's *private* (undisclosed) forecast when a respondent herds. By 'private' is simply meant the forecast which would be made in the absence of any (anti-)herding effects. Hence if the private h-step forecast exceeded the $h+1$-step consensus, the reported forecast $y^i_{t|t-h}$ will still exceed the consensus, but if the private forecast was an unbiased forecast, the reported forecast will fall short of the outcome on average. Consequently, we would expect to see a positive association between the LHS and RHS variables, and $\rho_1 > 0$. When the reported forecast is moved further from the consensus relative to the private forecast, the respondent is said to exaggerate his/her differences, or anti-herd, and is indicated by $\rho_1 < 0$.

The non-parametric test of Bernhardt et al. (2006) is based on the sum of two conditional probabilities: the probability that the reported forecast exceeds the outcome conditional on the forecast exceeding the consensus, $\mathrm{CP}_1 = \Pr\left(y_t < y^i_{t|t-h} \mid \overline{y}_{t|t-(h+1)} < y^i_{t|t-h} \right)$, and the probability that the reported forecast is less than the outcome conditional on the forecast being less than the consensus, $\mathrm{CP}_2 = \Pr\left(y_t > y^i_{t|t-h} \mid \overline{y}_{t|t-(h+1)} > y^i_{t|t-h} \right)$. If we set $\gamma^+_t = 1$ if $y^i_{t|t-h} > \overline{y}_{t|t-(h+1)}$, and $\gamma^-_t = 1$ if $y^i_{t|t-h} < \overline{y}_{t|t-(h+1)}$, and define the joint events as $\delta^+_t = 1$ if $y^i_{t|t-h} > \overline{y}_{t|t-(h+1)}$ and $y^i_{t|t-h} > y_t$, and $\delta^-_t = 1$ if $y^i_{t|t-h} < \overline{y}_{t|t-(h+1)}$ and $y^i_{t|t-h} < y_t$, then their test statistic S is calculated as:

$$S = \frac{1}{2} \left[\frac{\sum_t \delta^+_t}{\sum_t \gamma^+_t} + \frac{\sum_t \delta^-_t}{\sum_t \gamma^-_t} \right]. \tag{16}$$

This is asymptotically normally distributed $N\left(0.5, \frac{1}{16}\left[\left(\sum_t \gamma_t^+\right)^{-1} + \left(\sum_t \gamma_t^-\right)^{-1}\right]\right)$ under the null of no (anti-)herding. Given the relatively small numbers of forecasts available for each respondent and the asymptotic justification of the test, we choose to use the test statistic S as an indicator of a tendency to herd or anti-herd rather than a formal hypothesis test.

3.2 *Population Values of the Parameters in the Regression-based Tests of Herding for the Stylized Model*

Based on moment calculations in the Appendix for the model of forecaster behaviour detailed in section 2.1, we present the population values of the key parameters underlying the tests of (anti-)herding. For simplicity, we assume $h = 1$ but expressions are valid for any h.

3.2.1 T_1: *Adapted from Gallo et al. (2002)*

Firstly, consider the approach of GGJ. In the Appendix we calculate the relevant moments for the model in section 2.1. For equations (10) and (11) we find:

$$\beta_2 = \frac{\sigma_v^2\left(\delta^6\sigma_y^2 + \delta^4\lambda\sigma_w^2\right)}{\left(\delta^4\sigma_y^2 + \delta^2\lambda\sigma_w^2 + \sigma_v^2\right)\left(\delta^6\sigma_y^2 + \delta^4\lambda^2\sigma_w^2\right) - \left(\delta^6\sigma_y^2 + \delta^4\lambda\sigma_w^2\right)^2}, \quad \gamma_2 = 0.$$

Hence $\beta_2 = 0$ when there is no idiosyncratic error, but otherwise $\beta_2 > 0$, suggesting either herding or anti-herding. However, $\gamma_2 = 0$ even in the presence of noise or idiosyncratic error.

We can investigate the implications of falsely assuming the forecasters know the forecasts of others. If we were to incorrectly assume agents knew $\overline{y}_{t|t-2}$ at time $t-2$, and instead estimated β_2 in equation (8) and γ_2 in equation (9), we can show that:

$$\widetilde{\beta}_2 = \frac{\sigma_v^2 \left(\delta^4 \sigma_y^2 + \delta^2 \lambda \sigma_w^2\right)}{\left(\delta^4 \sigma_y^2 + \delta^2 \lambda^2 \sigma_w^2\right) \left(\delta^2 \sigma_\varepsilon^2 \lambda^2 + \sigma_v^2\right)}, \quad \widetilde{\gamma}_2 = \frac{\lambda (1-\lambda)}{\delta^2 \left(1-\delta^2\right)^{-1} + \lambda^2}.$$

Now the incorrect informational assumption results in $\widetilde{\gamma}_2 > 0$, suggesting anti-herding behaviour when we mistakenly assume forecasters know more about the forecasts of others than they do.

3.2.2 T_2: *Forecast revisions*

Next, in the Appendix we establish that the population value of the slope parameter ϕ_1 in equation (12) is:

$$\phi_1 = \frac{-\sigma_v^2}{\sigma_w^2 \delta^2 \left(\lambda + \delta^2 (1-\lambda)^2\right) + \sigma_v^2}. \tag{17}$$

Hence $-1 < \phi_1 \leqslant 0$. ϕ_1 will equal zero when the noise or idiosyncratic error is zero, but will approach -1 as the variance of the idiosyncratic error gets large relative to the variance of the underlying shocks, σ_w^2.

To investigate the effects of assuming forecasters know the contemporaneous forecasts of others, we consider regression equation (14) instead of equation (12). The population value of ϕ_1 in equation (14) is:

$$\widetilde{\phi}_1 = -1 \tag{18}$$

regardless of whether or not there is noise. Assuming forecasters know more than they do will manifest as herding behaviour.

3.2.3 T_3: *Based on Bernhardt et al. (2006)*

Consider the regression equivalent of Bernhardt et al. (2006).

$$y_t - y^i_{t|t-1} = \rho_0 + \rho_1 \left(y^i_{t|t-1} - \overline{y}_{t|t-2} \right) + w_t. \tag{19}$$

In the Appendix we establish that:

$$\rho_1 = \frac{-\sigma_v^2}{\sigma_w^2 \left(\lambda + (1-\lambda)^2 \delta^2 \right) + \sigma_v^2}.$$

This behaves similarly to ϕ_1: ρ_1 is zero when there is no noise, but approaches -1 as the noise component becomes large relative to the variance of the underlying disturbances.

Assuming agents know more than they do in this context would amount to replacing $\overline{y}_{t|t-2}$ by $\overline{y}_{t|t-1}$ in (19). Then it is straightforward to show that:

$$\widetilde{\rho}_1 = -1,$$

indicating herding even in the absence of idiosyncratic error.

3.3 *Population Values of the Parameters in the Regression-based Tests of Herding for the Noisy Information Model*

In this section we provide the outcomes (in population) of the herding tests when there is noisy information.

Firstly, for T_1 we calculate the population value of γ_2 in equation (11). To simplify the algebra, we have ignored the public signal s_t, and suppose agents only receive s_{it}. Hence we assume that $\sigma_\eta^2 = \infty$, and P_ε is re-defined accordingly and we denote this by P, and $H = 1$. In the Appendix we detail the calculation of:

$$\gamma_2 = \frac{(1-P)\left(1-(1-P)\,\delta^2\right)^{-1}\sigma_\varepsilon^2}{\left[\left(\sigma_w^2+\sigma_\varepsilon^2\right)\left(1-\delta^2\right)^{-1}\left[1+2\left(1-(1-P)\delta^2\right)^{-1}(1-P)\,\delta^2\right]-\sigma_w^2\delta^2\,(1-P)^2\left(1-(1-P)\,\delta^2\right)^{-2}\right]}$$

and establish that $\gamma_2 > 0$. Further, $\gamma_2 \to 0$ when $\sigma_\varepsilon^2 \to 0$ (and $P \to 1$ indicating declining informational rigidity).

Next, we consider the T_2 test of herding based on forecast revisions, and calculate the population value of ϕ_1 for regression equation (12), which uses the lagged consensus:

$$\phi_1 = \frac{PH\left[1-(1-PH)^2\,\delta^2\right]^{-1}\left((1-PH)\,PH\sigma_w^2 - P_\varepsilon^2\sigma_\varepsilon^2 + \delta^2 PH\,(1-PH)\,P_\eta^2\sigma_\eta^2\right) - PHP_\eta^2\sigma_\eta^2}{\left[1-(1-PH)^2\,\delta^2\right]^{-1}\left(P_\varepsilon^2\sigma_\varepsilon^2 + (PH)^2\,\sigma_w^2 + \delta^2 P_\eta^2\sigma_\eta^2\right) + P_\eta^2\sigma_\eta^2}. \tag{20}$$

A simpler expression results from again ignoring the public signal s_t. Then equation (20) simplifies to:

$$\phi_1 = \frac{\left((1-P)\,\sigma_w^2 - P\sigma_\varepsilon^2\right)}{\left(\sigma_\varepsilon^2+\sigma_w^2\right)}.$$

One can show that $\phi_1 < 0$, and $\phi_1 \to 0$ when $\sigma_\varepsilon^2 \to 0$ (and $P \to 1$, the signal becomes increasingly informative and the degree of informational rigidity declines) and when $\sigma_\varepsilon^2 \to \infty$ (the signal becomes uninformative and the degree of informational rigidity increases): see Appendix. Hence for intermediate values of the rigidity parameter, T_2 will indicate herding when there is noisy information.

Although the use of the $t-2$ consensus does not provide a valid test of herding, were we to apply such a test using equation (14) when there is noisy information (private and public), we would obtain:

$$\phi_1 = \frac{Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-2}\right)}{Var\left(y^i_{t|t-2} - \overline{y}_{t|t-2}\right)} = -PH \tag{21}$$

Since $PH \in (0,1)$, and $PH \to 1$ as rigidity lessens, this has a simple intuitive explanation. The evidence for herding will increase ($\phi_1 \to -1$) as rigidity lessens, because in response to less noisy signals individual forecasters place more weight on their private signals, s_{it}. The consensus forecast aggregates the private information, and so the (negative) correlation between $y^i_{t|t-1} - y^i_{t|t-2}$ and $y^i_{t|t-2} - \overline{y}_{t|t-2}$ increases, and with it the evidence of herding.

In terms of the T_3 test, in the Appendix we derive the following result under the null of noisy information (and assuming no public information, for simplicity):

$$\rho_1 = \frac{-P\sigma^2_\varepsilon\left[1-(1-P)^2\delta^2\right]^{-1} + \sigma^2_w\left[1-(1-P)\delta^2\right]^{-1}\left\{1-P\left[1-(1-P)^2\delta^2\right]^{-1}\right\}}{P\left(\sigma^2_w+\sigma^2_\varepsilon\right)\left[1-(1-P)^2\delta^2\right]^{-1}}.$$

We can show that $\rho_1 < 0$, and so T_3 will indicate anti-herding when there is noisy information. As for the other tests, $\rho_1 \to 0$ as $\sigma^2_\varepsilon \to 0$.

For the sticky information model of section 2.2.5, we do not provide formal derivations of the population values of the test regressions for the individual forecaster, as the implications of the theory are most readily applicable at the aggregate level (e.g., in terms of the properties of the aggregate forecast and forecast errors, etc.). However it follows immediately from the argument in section 2.2.5 that for individual forecasters who have not updated recently the consensus will contain useful information, and will help predict subsequent (updated) forecasts made by those individuals. Hence the tests of herding will falsely reject the no-herding null hypothesis.

3.4 *Summary of the Properties of the Tests of (Anti-)Herding*

We collect the results concerning the properties of the tests in the following propositions.

Proposition 1. When forecaster heterogeneity is driven by noise (idiosyncratic errors), the T_2 test (H_0: $\phi_1 = 0$ in equation (12)) will tend to suggest herding, because $-1 < \phi_1 < 0$, and $\phi_1 \to -1$ as the importance of the idiosyncratic errors increases relative to the variance of the true innovations to the process. The T_2 test will also indicate herding when forecaster heterogeneity is driven by noisy information.

Proposition 2. When forecaster heterogeneity is driven by noise, the T_3 test (H_0: $\rho_1 = 0$ in equation (15)) will suggest anti-herding, because the population value of ρ_1 tends to -1 as the importance of the idiosyncratic errors increases relative to the variance of the true innovations to the process. The T_3 test will also indicate anti-herding when forecaster heterogeneity is driven by noisy information.

Proposition 3. Under forecaster heterogeneity driven by noise, the T_1 testing procedure performs as follows: the test of H_0: $\beta_1 = 0$ (in equation (10)) will reject the null of (anti-)herding, but the test of H_0: $\gamma_2 = 0$ in the re-parameterised equation (11) will not reject. However, the test of H_0: $\gamma_2 = 0$ will reject in favour of $\gamma_2 > 0$ when disagreement is driven by noisy information.

Proposition 4. If agents have access to less information than is assumed in the specification of the test regression, all the testing approaches will reject the (anti-)herding null (irrespective of whether there are idiosyncratic errors).

Propositions 1 – 3 indicate that differences between forecasters due to idiosyncratic errors will falsely indicate (anti-)herding for all the tests considered other than the test of the null that $\gamma_2 = 0$ of T_1. This suggests that one component of the T_1 testing procedure (namely, that $\gamma_2 = 0$) can be used as a reliable test of (anti-)herding, in that it will not falsely reject the no-herding null due to idiosyncratic error.[8] Note that all tests (including H_0: $\gamma_2 = 0$) will reject the no-herding null when there is noisy information.

Proposition 4 asserts that even tests based on $\gamma_2 = 0$ will falsely indicate herding when we incorrectly attribute to agents more information than they have. It might appear that ensuring the consensus forecast is known to all forecasters (by taking the consensus made far enough in the past, for example) would guard against falsely finding herding based on testing $\gamma_2 = 0$. However, this is not true for some recent theories of expectations formation involving informational rigidities, such as the noisy and sticky information models discussed here. Explicit expressions for the test parameters are given for the noisy information model, and are shown to not equal zero, suggesting these tests will indicate herding.

4 EMPIRICAL ANALYSIS

4.1 *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 August 1 2014, the Academic Bibliography maintained by the Philadelphia Fed listed 87 research papers based on the SPF forecast data (see http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/academic-bibliography.cfm.)

SPF respondents are asked to provide forecasts of a number of macroeconomic variables, and we choose to analyze the forecasts of real GDP growth and (the GDP-deflator measure of) inflation. We analyze the forecasts of the survey quarter, the next quarter, and each of the next three quarters. We have 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.

We call forecasts of the current-quarter[9] horizon $h = 1$ forecasts. Then we have 3 separate sets of pairs of forecasts of the same event: the $h = 2$ and $h = 1$ forecasts, the $h = 3$ and $h = 2$ forecasts, and finally the $h = 4$ and $h = 3$ forecasts (with the $h = 5$ forecasts being used to construct the consensus in this last case). For example, the first pair of $h = 1$ and $h = 2$ forecasts are the 1969:Q1 survey $h = 1$ forecast and the 1968:Q4 survey $h = 2$ forecast, both of the value of the variable in 1969:Q1. The last pair are the 2013:Q1 $h = 1$ and the 2012:Q4 $h = 2$ forecasts (both of 2013:Q1).

These samples of pairs of fixed-event forecasts can be used to construct the tests of (anti-)herding described in section 3. When actual values are required, as in the tests based on BCK, we use the vintage value two-quarters after the reference quarter, taken from the Real Time Data Set for Macroeconomists (RTDSM) run by the Federal Reserve Bank of Philadelphia (see Croushore and Stark (2001)). So, for example, the forecasts of 1969:Q1 would be compared to the actual value for 1969:Q1 recorded in the 1969:Q3 quarterly vintage. This seems preferable to using a vintage from many years later as this will contain revisions and definitional changes (see e.g., Landefeld et al. (2008) for a discussion of the revisions to US national accounts data).

What we refer to as the lagged consensus forecast is from the individual forecasts of the target in the survey one quarter before the earlier of the two forecasts being compared. To make matters concrete, consider again the example above, of the $h = 1$ forecast made in response to the 1969:Q1 survey, and the $h = 2$ forecast submitted to the 1968:Q4. These are both of the same target period: the value of the variable in 1969:Q1. The lagged consensus forecast corresponding to this pair of (individual) forecasts would be based on the $h = 3$ forecasts (of 1969:Q1) submitted to the 1968:Q3 survey. The results of a survey are made known shortly after the survey responses have been filed, so will be known well before the next survey. In the above example, the current consensus would be the consensus of the $h = 2$ forecasts submitted to the 1969:Q4 survey. Because it is contemporaneous with the individual's forecast, the individual will not know the SPF consensus. However, it may be reasonable to assume a professional forecaster will effectively know the consensus from other sources. Our central case uses the lagged consensus, and so is uncontroversial in terms of what the forecasters know.

4.2 *Empirical Findings*

Table 2 reports the evidence for (anti-)herding based on equations (10) and (11) from test T_1, and on equation (12) for test T_2. We adopt the conservative assumption that forecasters only know the lagged consensus. We run regressions (10), (11) and (12) on all the forecasts of a given horizon for each individual who reported a sufficient number of forecasts. We require 10 or more forecasts of a given horizon. This requirement was satisfied by around 150 individuals for each horizon. We subsequently check the sensitivity of the results to this requirement. For T_1 we report the proportion of regressions for which (i) we rejected the null that $\beta_2 = 0$ in equation (10), (ii) we rejected the null that $\beta_2 = 0$ (in equation (10)) and $\gamma_2 = 0$ (in equation (11)) in favour of $\beta_2 \neq 0$ and $\gamma_2 < 0$, and (iii) $\gamma_2 = 0$ against $\gamma_2 \neq 0$ (in equation (11)). For T_2, the Table reports: (i) the rejection frequency of a one-sided test of $\phi_1 = 0$ versus $\phi_1 < 0$ (indicating herding), (ii) the rejection frequency of $\phi_1 = 0$ versus $\phi_1 > 0$ (signifying anti-herding), and (iii) the proportion for which neither is found.[10]

In terms of the interpretation of T_1 of GGJ, the results indicate (anti-)herding ($\beta_2 \neq 0$) for around one quarter of the forecasters for inflation and for output growth. However, the results of (ii) indicate that rarely do we find herding (as defined by GGJ) with $\beta_2 \neq 0$ and $\gamma_2 < 0$. Moreover, (iii) indicates that $\gamma_2 = 0$ is rejected for relatively few forecasters. Remember that the equations we have estimated are an adaptation of GGJ, and instead we interpret $\gamma_2 = 0$ as the no-herding null when there is idiosyncratic error.

The findings are little affected by the forecast horizon. Recall that the results for $h = 1$ are based on the revision between the forecasts of the current survey quarter value, and the forecasts of that target made in the previous survey. The results for $h = 2$ relate to the forecasts of the next quarter, and the forecast of that quarter made in the previous quarter, and so on. In short, Table 2 provides evidence of (anti-)herding ($\beta_2 \neq 0$) but little evidence of herding ($\beta_2 \neq 0$ and $\gamma_2 < 0$).

If we include the forecast dispersion term (namely, the cross-sectional variance of the forecasts used to calculate the consensus) in the regressions, the results of the tests for herding are largely unchanged. This could be interpreted as suggesting that the degree of uncertainty (as proxied by the group variance) does not have a measurable effect on whether this approach suggests herding.

The T_2 test has $\phi_1 = 0$ rejected in favour of $\phi_1 < 0$ for around 60% of inflation, with the proportion declining to around one third for output growth at the shortest horizon. There are virtually no instances where $\phi_1 = 0$ is rejected in favour of $\phi_1 > 0$.

We report the results of T_3 tests based on equation (15) in Table 3. When we reject $\rho_1 = 0$ it is mostly in the direction of $\rho_1 < 0$. This occurs for around half of the inflation forecasters, and around one third of the output growth forecasters. This evidence of anti-herding is backed up by the estimates of S (equation (16)), which exceed one half for in excess of 80% of forecasters for inflation, with only slightly lower proportions for the output growth forecasts.

Care has to be taken to ensure the tests are based on appropriate estimates of the standard error of the coefficient estimates. The t-tests of $\rho_1 = 0$ in Table 3 are based on autocorrelation and heteroscedasticity consistent standard errors because of the usual issue of overlapping forecasts. When $h > 1$, a forecast is made before the actual value corresponding to the previous forecast becomes known, engendering correlation in sequences of multi-step forecast errors even for optimal forecasts. For $h = 1$ only a correction for heteroscedasticity is made. The tests in Table 2 do not involve actual values, and only a correction for possible heteroscedasticity is required, even for $h > 1$. To understand why this is the case, consider for example two adjacent observations on the left-hand-side of equation (12): $y^i_{t|t-h} - y^i_{t|t-(h+1)}$ and $y^i_{t-1|t-1-h} - y^i_{t-1|t-1-(h+1)}$, say. Under the null the revisions are serially uncorrelated even when $h > 1$.[11]

Table 4 reports results when we require each forecaster to make a minimum of 30 forecasts (compared to the requirement of 10 or more assumed hitherto). There are only around as third as many respondents as previously, but the results are not changed in any important respect. Hence the relatively small number of observations in the regressions underlying Tables 2 and 3 are not driving the results.

Tables 5 and 6 report results for the current consensus: based on T_1 equations (8) and (9), and T_2 equation (14). [12] Generally there is more evidence of (anti-)herding, but otherwise the pattern of results is little changed. Recall that we established in section 3 that the use of the current consensus when this is unwarranted (not known to the individual forecasters) increases the evidence of apparent (anti-)herding. As discussed earlier, it is a moot point whether SPF respondents effectively know the current consensus via other sources.

In sum, if the tests are taken at face value, their application to the SPF inflation and output growth forecast data suggests the following: the T_2 tests indicate around one in every two forecasters herds; radically different findings result from T_1 tests on the regression parameters of β_2 and γ_2, where there is little evidence of herding, or compared to the evidence for anti-herding based on T_3. However, based on our analysis of the properties of these tests, the pattern of results suggests an absence of herding behaviour, and points to the differences between forecasters primarily reflecting noise or idiosyncratic error, emanating from any one of the theory-based explanations of forecast behaviour described in section 2.2. The evidence against (anti-)herding behaviour is based on the relatively lower rejections rates of the no-herding null using the T_1 test of $\gamma_2 = 0$, compared to the evidence for herding based on T_2, and for anti-herding based on T_3. Our analysis indicates that the T_1 test of $\gamma_2 = 0$ is robust to noise, whereas T_2 and T_3 will indicate herding and anti-herding, respectively, when there is noise. These predictions are found in the forecast data.

This pattern of findings is also broadly consistent with information rigidities, such as noisy information, as shown in section 3.3. Noisy information would also suggest herding based on T_2, and anti-herding based on T_3. The main point is that rejection of the null using any of these testing procedures does not indicate macro-forecasters herd (or anti-herd).

5 CONCLUSIONS

There is much interest in the literature concerning whether forecasters report their 'best' forecasts, where best is narrowly defined to mean most accurate, or whether forecasters are unduly influenced by others, as might be the case if their payoffs depend on the positioning of their forecasts in relation to the forecasts of others. A natural way of making this idea testable is via the notion of herding or anti-herding, and to consider whether the changes in an individual's forecasts of y_t are systematically related to the *past* consensus.

In recent years a number of models of forecaster behaviour have been proposed to explain forecaster dispersion, recognizing that the full-information rational expectations hypothesis is unable to explain this empirical regularity. The models differ in terms of the assumed environment in which agents operate, and in terms of the sorts of information they are privy to, and the models they use, etc. However, in terms of determining the outcomes of the tests of herding we consider, we show that what matters is whether the differences between forecasters reflect private information or noise, and whether past values of summaries of the forecast distribution (such as the consensus) remain informative because the past history of the underlying series remains hidden. We show how the theory models fit within this general framework.

Our application of the herding tests to US SPF individual quarterly inflation and output growth forecasts for the period 1981 to 2013 taken at face value suggests a significant proportion of professional US macroeconomists appear to be influenced by their fellow forecasters, with one test suggesting respondents deliberately exaggerate their differences, while another suggests herding (or consensus-seeking) is the dominant type of behaviour. However, given our analysis of the properties of the herding tests, we are able to conclude that the empirical pattern of rejections that we observe across the different tests is consistent with the pattern predicted by differences amongst forecasters primarily reflecting idiosyncratic errors, or reflecting noisy information. Either way, the evidence for (anti-)herding is far from compelling.

The finding that apparent herding behaviour may result from forecaster heterogeneity due to informational rigidities such as noisy information is interesting, given the recent attention paid to testing such models (see, e.g., Coibion and Gorodnichenko (2012, 2015), Andrade and Le Bihan (2013) and Giacomini et al. (2015)). This arises from the noisy-information model assumption that the underlying variable is only ever observed with error, and the assumption that forecasters do not make use of the consensus forecast, even though it would be beneficial to do so in terms of improving forecast accuracy. The lagged consensus forecasts will be related to the individual's forecast (or forecast revision) under noisy information, and the herding tests considered in this paper will falsely indicate herding. Conceptually herding behaviour and noisy information behaviour are quite different. Noisy-information forecasters are rational, squared-error-loss minimizing agents, given the informational constraints they face, whereas (anti-)herding forecasters are not rational in terms of minimizing mean-squared error. We leave for future research the possibility of developing an approach able to discriminate between the two.

An implication of our paper is that the growing literature suggesting macro forecasters often anti-herd (see, e.g., some of the later entries in Table 1) may simply reflect heterogeneity induced by idiosyncratic errors (and hence any one of a number of theory model explanations) or noisy information. Our findings suggest that some of the recent literature indicating herding behaviour may be misleading.

References

Andrade, Philippe and Hervé Le Bihan (2013), "Inattentive professional forecasters." *Journal of Monetary Economics*, 60, 967–982.

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

Ashiya, Masahiro and Takero Doi (2001), "Herd behavior of Japanese economists." *Journal of Economic Behavior & Organization*, 46, 343–346.

Batchelor, Roy and Pami Dua (1992), "Conservatism and consensus-seeking among economic forecasters." *Journal of Forecasting*, 11, 169–181.

Bernhardt, Dan, Murillo Campello, and Edward Kutsoati (2006), "Who herds?" *Journal of Financial Economics*, 80, 657–675.

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

Capistrán, Carlos and Allan Timmermann (2009), "Disagreement and biases in inflation expectations." *Journal of Money, Credit and Banking*, 41, 365–396.

Clements, Michael P. (1995), "Rationality and the role of judgement in macroeconomic forecasting." *Economic Journal*, 105, 410–420.

Clements, Michael P. (1997), "Evaluating the rationality of fixed-event forecasts." *Journal of Forecasting*, 16, 225–239.

Clements, Michael P. (2015), “Do US Macroeconomic Forecasters Exaggerate their Differences?” *Journal of Forecasting*, 34, 649–660.

Coibion, Olivier and Yuriy Gorodnichenko (2012), “What can survey forecasts tell us about information rigidities?” *Journal of Political Economy*, 120, 116 – 159.

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

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

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

Davies, Antony and Kajal Lahiri (1995), “A new framework for analyzing survey forecasts using three-dimensional panel data.” *Journal of Econometrics*, 68, 205–227.

Davies, Antony, Kajal. Lahiri, and Xuguang Sheng (2011), “Analyzing three-dimensional panel data of forecasts.” In *The Oxford Handbook of Economic Forecasting* (Michael P Clements and David F. Hendry, eds.), 473–495, Oxford University Press.

Elliott, Graham, Ivana Komunjer, and Allan Timmermann (2005), “Estimation and testing of forecast rationality under flexible loss.” *Review of Economic Studies*, 72, 1107–1125.

Elliott, Graham, Ivana Komunjer, and Allan Timmermann (2008), “Biases in macroeconomic forecasts: Irrationality or asymmetric loss.” *Journal of the European Economic Association*, 6, 122–157.

Frenkel, Michael, Jan-C. Rülke, and Lilli Zimmermann (2012), “Do current account forecasters herd? evidence from the euro area and the g7 countries.” *Review of International Economics*, 20, 221–236.

Frenkel, Michael, Jan-C. Rülke, and Lilli Zimmermann (2013), “Do private sector forecasters chase after IMF or OECD forecasts.” *Journal of Macroeconomics*, 37, 217–229.

Gallo, Giampiero M., Clive W.J. Granger, and Yongil Jeon (2002), “Copycats and Common Swings: The Impact of the Use of Forecasts in Information Sets.” *IMF Staff Papers*, 49, 4–21.

Giacomini, Raffaella, Vasiliki Skreta, and Javier Turen (2015), “Models, Inattention and Expectation Updates.” Discussion Papers 1602, Centre for Macroeconomics (CFM).

Granger, Clive W.J. (1969), “Prediction with a generalized cost of error function.” *Operations Research Quarterly*, 20, 199–207.

Kandel, Eugene and Ben-Zion Zilberfarb (1999), “Differential interpretation of information in inflation forecasts.” *The Review of Economics and Statistics*, 81, 217–226.

Lahiri, Kajal and Fushang Liu (2009), “On the use of density forecasts to identify asymmetry in forecasters' loss function.” *Business and Economic Statistics Section - JSM*, 2396–2408.

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

Landefeld, J. Steven, Eugene P Seskin, and Barbara M Fraumeni (2008), “Taking the pulse of the economy.” *Journal of Economic Perspectives*, 22, 193–216.

Laster, David, Paul B. Bennett, and In Sun Geoum (1999), “Rational bias in macroeconomic forecasts.” *The Quarterly Journal of Economics*, 114(1), 293–318.

Mankiw, N. Gregory and Ricardo Reis (2002), “Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips Curve.” *Quarterly Journal of Economics*, 117, 1295–1328.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers (2003), “Disagreement about inflation expectations.” mimeo, National Bureau of Economic Research, Cambridge MA.

Manzan, Sebastiano (2011), “Differential interpretation in the Survey of Professional Forecasters.” *Journal of Money, Credit and Banking*, 43, 993–1017.

McNees, Stephen K. (1989), “Why do forecasters differ?” *New England Economic Review*, 1/2, 42–54.

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

Nordhaus, William D. (1987), “Forecasting efficiency: Concepts and applications.” *Review of Economics and Statistics*, 69, 667–674.

Ottaviani, Marco and Peter N. Sorensen (2006), “The strategy of professional forecasting.” *Journal of Financial Economics*, 81, 441–466.

Papastamos, Dimitrios, George Matysiak, and Simon Stevenson (2015), “Assessing the accuracy and dispersion of real estate investment forecasts.” *International Review of Financial Analysis*, 42, 141 – 152.

Patton, Andrew J. and Allan Timmermann (2007), "Testing forecast optimality under unknown loss." *Journal of the American Statistical Association*, 102, 1172–1184.

Patton, Andrew J. and Allan Timmermann (2010), "Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion." *Journal of Monetary Economics*, 57(7), 803–820.

Pierdzioch, Christian and Jan-Christoph Rülke (2012), "Forecasting stock prices: Do forecasters herd?" *Economics Letters*, 116, 326–329.

Pierdzioch, Christian, Jan Christoph Rülke, and Georg Stadtmann (2010), "New evidence of anti-herding of oil-price forecasters." *Energy Economics*, 32, 1456–1459.

Pierdzioch, Christian, Jan-Christoph Rülke, and Georg Stadtmann (2012a), "A Note on Forecasting Emerging Market Exchange Rates: Evidence of Anti-herding." *Review of International Economics*, 20, 974–984.

Pierdzioch, Christian, Jan-Christoph Rülke, and Georg Stadtmann (2012b), "Housing Starts in Canada, Japan, and the United States: Do Forecasters Herd?" *The Journal of Real Estate Finance and Economics*, 45, 754–773.

Pons-Novell, Jordi (2003), "Strategic bias, herding behaviour and economic forecasts." *Journal of Forecasting*, 22, 67–77.

Rülke, Jan-Christoph (2012), "Do Private Sector Forecasters Desire to Deviate From the German Council of Economic Experts?" *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 232, 414–428.

Rülke, Jan-Christoph, Maria Silgoner, and Julia Woerz (2016), "Herding behavior of business cycle forecasters." *International Journal of Forecasting*, 32, 23 – 33.

Sims, Christopher A (2003), "Implications of rational inattention." *Journal of Monetary Economics*, 50, 665–690.

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

Tsuchiya, Yoichi and T. Kato (2015), "Asymmetric loss and herding behavior of exchange rate forecasters: Evidence from South Africa." *Applied Economics*, 47, 5841–5852.

Woodford, Michael (2002), "Imperfect common knowledge and the effects of monetary policy." In *Knowledge, Information, and Expectations in Modern Macroeconomics: In honor of Edmund Phelps* (P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds.), 25–58, Princeton University Press.

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

Zellner, Arnold (1986), "Biased predictors, rationality and the evaluation of forecasts." *Economics Letters*, 21, 45–48.

Notes

[1]'Anti-herding' refers to the behaviour of deliberately exaggerating the differences between one's forecasts and those of others, while 'herding' has its natural interpretation.

[2]For example, Coibion and Gorodnichenko (2015) show that there is a correlation between the mean forecast error and the revision to fixed-event aggregate forecasts of the target in a noisy information model. This is an emergent property - it does not hold at the individual level assuming rational updating in response to the receipt of the new information. As shown below, the implications for herding can be analyzed at the individual level.

[3]For example, in the recent modelling of three-dimensional panels of forecasters, which allows for multiple forecasters, multiple targets and multiple forecast horizons: see e.g., Davies and Lahiri (1995), and the review article by Davies et al. (2011) for further discussion.

[4]See, for example, the realization-forecast regressions of Mincer and Zarnowitz (1969).

[5]The model of Davies and Lahiri (1995) allows forecasters to disagree because of fixed individual-specific biases and 'idiosyncratic errors', where the latter would appear to correspond to our v's. However, (Davies and Lahiri, 1995, p. 209) suggest that these capture ' 'other factors' (e.g., private information, measurement error, etc.) '. However, this conflates noise and private information, which we keep separate because of their different impacts on tests for herding.

[6]Patton and Timmermann (2010) assume agents receive a noisy public signal, and in that case $E\left(y_t \mid \mathcal{I}_{t-1}^i\right)$ is the optimal forecast using the Kalman Filter, as discussed in section 2.2.5.

[7]See, e.g., Granger (1969) and Zellner (1986) for early contributions on asymmetric loss, and more recently Elliott et al. (2005), Elliott et al. (2008), Patton and Timmermann (2007) and Lahiri and Liu (2009).

[8]Note that GGJ envisaged a different use of the test of $\gamma_2 = 0$, namely that, conditional on rejecting $\beta_2 = 0$, finding $\gamma_2 < 0$ would indicate herding. However, $\beta_2 = 0$ would be rejected because of idiosyncratic error, so the suggestion of GGJ would likely result in the finding of anti-herding. However, we stress that their implementation is not the same as ours: we lag the consensus by one period relative to the longer-horizon individual forecast and we omit the measure of forecast dispersion.

Similar comments apply to T_3. Our implementation is a regression-based adaptation of the original non-parametric test of Bernhardt et al. (2006).

[9]That is, a forecast of 2012:Q1 made in response to the 2012:Q1 survey. This is termed 1-step, because when the forecast of the survey-quarter value is made the most recent GDP growth and (GDP deflator) inflation figures will be the advance estimates of the previous quarter.

[10]All tests were undertaken at the 5% level.

[11]This follows immediately if we consider a time series x_t written as an infinite-order moving average

$$x_t = \psi(L)\varepsilon_t = \sum_{j=1}^{h} \psi_{h-j}\varepsilon_{t-h+j} + \psi_h\varepsilon_{t-h} + \sum_{j=1}^{\infty} \psi_{h+j}\varepsilon_{t-h-j}.$$

Then $x_{t|t-h} - x_{t|t-(h+1)} = \psi_h\varepsilon_{t-h}$. Moreover, from

$$x_{t-1} = \psi(L)\varepsilon_{t-1} = \sum_{j=1}^{h} \psi_{h-j}\varepsilon_{t-1-h+j} + \psi_h\varepsilon_{t-1-h} + \sum_{j=1}^{\infty} \psi_{h+j}\varepsilon_{t-1-h-j},$$

we have $x_{t-1,t-1-h} - x_{t-1,t-1-(h+1)} = \psi_h\varepsilon_{t-1-h}$, and so

$$Cov\left(x_{t|t-h} - x_{t|t-(h+1)}, x_{t-1,t-1-h} - x_{t-1,t-1-(h+1)}\right) = 0.$$

[12]In these implementations, we now have '$h = 4$' results, as these only require $h = 5$ forecasts, which are available in the SPF.

Table 1: Literature on Macro Forecasters and Herding

Study	Finding
McNees (1989)	Herding
Batchelor and Dua (1992)	'Variety-seeking behaviour' (anti-herding)
Ashiya and Doi (2001)	Herding
Lamont (2002)	Herding based on age
Bewley and Fiebig (2002)	Herding
Gallo et al. (2002)	Imitation (herding) for the US forecasters
Pons-Novell (2003)	Herding
Ashiya (2009)	Make extreme forecasts to generate publicity
Pierdzioch et al. (2010)	Anti-herding
Pierdzioch et al. (2012a)	Anti-herding
Pierdzioch and Rülke (2012)	Anti-herding
Pierdzioch et al. (2012b)	Anti-herding
Rülke (2012)	Anti-herding
Frenkel et al. (2012)	Anti-herding
Frenkel et al. (2013)	Anti-herding
Clements (2015)	Evidence of herding and anti-herding depending on forecast horizon
Papastamos et al. (2015)	Herding
Tsuchiya (2015)	Herding and anti-herding
Tsuchiya and Kato (2015)	Neither herding or anti-herding
Rülke et al. (2016)	Anti-herding

The table includes papers from a search on Thomson Reuters Web of Science, of forecaster herding. We omit papers outside the scope of macro and business forecasting, and papers not reporting results of tests of herding. The papers are presented in chronological order.
There are a number of points to be borne in mind.
1. The definition and meaning of herding and anti-herding is not the same in all cases.
2. A variety of approaches to testing for herding are used.
For example, Lamont (2002) tests 'whether the pattern of forecast herding /scattering varies significantly over the professional lifetime of the forecaster' but does not consider whether they are 'placing too much/little weight on the forecasts of others' (p.268), and hence does not test for herding as defined in this paper. McNees (1989) shows that judgmental adjustments to model-based forecasts sometimes make the forecasts more similar to those of others, but again this might simply be a rational response to common information not already accounted for by the model. Bewley and Fiebig (2002) find evidence of herding for around a half of their sample of interest rate forecasters, but their setup does not allow for a distinction between those who are following 'the pack' and those making use of 'important information contained in the previous consensus mean that was not available to the individual forecaster at the time' (p.3).
In some cases the findings are more nuanced than herding/anti-herding (or neither): for example, Batchelor and Dua (1992) find evidence of 'conservatism' (individuals giving too much weight to their own past forecasts) but also that too little weight is placed on the forecasts of others.

Table 2: Tests of herding: T_1 and T_2 (based on the lagged consensus)

h	Proportion of regressions for which we find:					
	T_1			T_2		
	$\beta_2 \neq 0$	$\beta_2 \neq 0$ and $\gamma_2 < 0$	$\gamma_2 \neq 0$	$\phi_1 < 0$	$\phi_1 > 0$	Neither
	Inflation					
3	0.25	0.05	0.18	0.59	0.00	0.41
2	0.31	0.08	0.18	0.67	0.00	0.33
1	0.24	0.05	0.17	0.57	0.00	0.43
	Output growth					
3	0.29	0.01	0.13	0.59	0.00	0.41
2	0.27	0.02	0.20	0.47	0.06	0.46
1	0.20	0.04	0.15	0.34	0.06	0.60

The test regressions are:

$$y^i_{t|t-h} = \beta_0 + \beta_1 y^i_{t|t-(h+1)} + \beta_2 \overline{y}_{t|t-(h+2)} + u_t$$

and

$$y^i_{t|t-h} - \overline{y}_{t|t-(h+2)} = \gamma_0 + \gamma_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)} \right) + \gamma_2 \overline{y}_{t|t-(h+2)} + u_t,$$

for the left-hand-side panel, and for the right-hand-side panel:

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \phi_0 + \phi_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)} \right) + u_t.$$

The entries in the table 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 require 10 or more pairs of forecasts for an individual respondent. This gives around 150 regressions at each h.

Table 3: Tests of herding: T_3 (adaptation of Bernhardt et al. (2006))

h	Proportion of regressions			Proportion of regressions	
	$\rho_1 < 0$	$\rho_1 > 0$	Neither	$S < 0.5$	$S > 0.5$
			Inflation		
4	0.58	0.01	0.42	0.14	0.84
3	0.62	0.01	0.37	0.15	0.83
2	0.60	0.00	0.40	0.14	0.82
1	0.66	0.02	0.32	0.12	0.85
			Output growth		
4	0.58	0.02	0.40	0.18	0.80
3	0.38	0.03	0.59	0.26	0.71
2	0.30	0.03	0.68	0.21	0.78
1	0.37	0.03	0.60	0.20	0.77

The test regression is:

$$y_t - y^i_{t|t-h} = \rho_0 + \rho_1 \left(y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} \right) + u_t$$

The entries in the table 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.
The remaining two columns are the proportion of regressions for which the S-statistics is less than / greater than one half.
We require 10 or more pairs of forecasts for an individual respondent. This gives around 150 regressions at each h.

Table 4: Herding Tests: Summary Table for a minimum of 30 forecasts

h	T_1 (GGJ)			T_2 (Forecast Revisions)			T_3 (BCK regression test)		
	$\beta_2 \neq 0$	$\beta_2 \neq 0$, $\gamma_2 < 0$	$\gamma_2 \neq 0$	$\phi_1 < 0$	$\phi_1 > 0$	Neither	$\rho_1 < 0$	$\rho_1 > 0$	Neither
				Inflation					
3	0.45	0.08	0.18	0.73	0.00	0.27	0.74	0.00	0.26
2	0.39	0.09	0.19	0.66	0.00	0.34	0.74	0.00	0.26
1	0.23	0.05	0.14	0.63	0.00	0.37	0.79	0.02	0.19
				Output Growth					
3	0.46	0.04	0.24	0.78	0.00	0.22	0.52	0.02	0.46
2	0.40	0.00	0.21	0.53	0.04	0.43	0.37	0.00	0.63
1	0.19	0.00	0.11	0.42	0.00	0.58	0.42	0.00	0.58

The test regressions are:

GGJ test regressions

$$y^i_{t|t-h} = \beta_0 + \beta_1 y^i_{t|t-(h+1)} + \beta_2 \overline{y}_{t|t-(h+2)} + u_t$$

and:

$$y^i_{t|t-h} - \overline{y}_{t|t-(h+2)} = \gamma_0 + \gamma_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)} \right) + \gamma_2 \overline{y}_{t|t-(h+2)} + u_t.$$

Forecast Revision regressions

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \phi_0 + \phi_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)} \right) + u_t$$

BCK regressions

$$y_t - y^i_{t|t-h} = \rho_0 + \rho_1 \left(y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} \right) + u_t$$

The entries in the table 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 require 30 or more pairs of forecasts for an individual respondent. This gives 50 to 60 regressions at each h.

Table 5: Tests of herding: T_1 (adapted from Gallo, Granger and Jeon (2002)). *Current Consensus*

h	$\beta_2 \neq 0$	$\beta_2 \neq 0$ and $\gamma_2 < 0$	$\gamma_2 \neq 0$
		Inflation	
4	0.44	0.05	0.16
3	0.42	0.06	0.09
2	0.46	0.06	0.11
1	0.42	0.08	0.13
		Output growth	
4	0.35	0.01	0.11
3	0.45	0.02	0.14
2	0.52	0.02	0.23
1	0.45	0.02	0.15

The test regressions are:

$$y^i_{t|t-h} = \beta_0 + \beta_1 y^i_{t|t-(h+1)} + \beta_2 \overline{y}_{t|t-(h+1)} + u_t$$

and

$$y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} = \beta_0 + \gamma_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)} \right) + \gamma_2 \overline{y}_{t|t-(h+1)} + u_t.$$

The entries in the table (columns 2 to 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 10 or more pairs of forecasts for an individual respondent. This gives around 150 regressions at each h.

Table 6: Tests of herding: T_2 (based on the forecast revision). *Current Consensus*

h	Proportion of regressions		
	$\phi_1 < 0$	$\phi_1 > 0$	Neither
	Inflation		
4	0.79	0.00	0.21
3	0.75	0.00	0.25
2	0.78	0.00	0.22
1	0.74	0.01	0.26
	Output growth		
4	0.74	0.00	0.26
3	0.70	0.00	0.30
2	0.69	0.01	0.30
1	0.62	0.03	0.36

The test regression is:

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

The entries in the table (columns 2 to 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 10 or more pairs of forecasts for an individual respondent. This gives around 150 regressions at each h.