ICMA Centre

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Assessing the Evidence of Macro-Forecaster Herding: Forecasts of Inflation and Output Growth

October 2014

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

Discussion Paper Number:
ICM-2014-12

The aim of this discussion paper series is to disseminate new research of academic distinction. Papers are preliminary drafts, circulated to stimulate discussion and critical comment. Henley Business School is triple accredited and home to over 100 academic faculty, who undertake research in a wide range of fields from ethics and finance to international business and marketing.

admin@icmacentre.ac.uk

www.icmacentre.ac.uk

Assessing the Evidence of Macro-Forecaster Herding: Forecasts of Inflation and Output Growth.

Michael P. Clements*

ICMA Centre, Henley Business School, University of Reading

October 9, 2014

Abstract

We consider a number of ways of testing whether macroeconomic forecasters herd or anti-herd, i.e., whether they shade their forecasts towards those of others or purposefully exaggerate their differences. When applied to survey respondents' expectations of inflation and output growth the tests indicate conflicting behaviour. We show that this can be explained in terms of a simple model in which differences between forecasters are primarily due to idiosyncratic factors or reporting errors rather than imitative behaviour. Models of forecaster heterogeneity that stress informational rigidities will also falsely indicate imitative behaviour.

JEL Classification: E37.

Keywords: macro-forecasting, imitative behaviour, private information, idiosyncratic errors.

*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.

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 payoff. As noted by Lamont (2002, p. 265), for example, forecasters may set their forecasts to 'optimize profits or wages, credibility, shock value, marketability, political power...' (see Laster, Bennett and Geoum (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.

It is clear that testing for herding behaviour is unlikely to be straightforward, as individuals' forecasts will tend to cluster together for reasons other than imitation effects, including for example a sharing of common information about the likely future evolution of the variable of interest. Moreover, to the extent that the consensus forecast (or more generally, another individual's forecast) contains relevant information not possessed by a given forecaster, moving toward (or away from) the consensus forecast may be consistent with minimizing a traditional squared error loss function, and so does not constitute imitative behaviour. Hence care needs to be taken to ensure that the consensus calculated by the econometrician is in the individual agents' information sets.[1]

In this paper we provide a critique of two approaches to testing for herding or anti-herding.[2] The first includes a number of variants of an approach suggested by Gallo, Granger and Jeon (2002) (henceforth, GGJ). Under one of these, imitation occurs when the revisions to fixed-event forecasts - the change from the $h+1$-step ahead forecast of the target y_t to the h-step ahead forecast of y_t - are systematically related to the extent to which the $h+1$-step ahead forecasts exceeds or falls short of the consensus view. If the change is in the

[1]Some theoretical models of forecaster heterogeneity have the property that the consensus forecast will contain relevant information not possessed by the individual irrespective of the timing convention we adopt. In such models agents are assumed to pay no attention to the consensus view (or the views of others more generally) albeit that forecast accuracy would improve by doing so. We discuss such models when we analyze information rigidities as a possible source of forecaster disagreement.

[2]'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. Throughout we will use 'imitative' behaviour to cover both herding and anti-herding.

direction of reducing the divergence between the forecaster's $h+1$-forecast and the previous forecast, the forecaster is said to herd. If the forecaster positions himself further from (last period's) consensus they are said to anti-herd. Other variants do not explicitly consider the forecast revision, but are broadly similar. When we apply this approach to the forecasts of inflation and real output growth made by the respondents to the US Survey of Professional Forecasters, we find around a half of all respondents are influenced by the forecasts of others: whether they herd or anti-herd depends on the variant of the test used.

The second approach is based on Bernhardt, Campello and Kutsoati (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, Rülke and Stadtmann (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 for heterogeneity in that some forecasters may herd, others anti-herd, and others do neither. Because the test is non-parametric and relies on large-sample distributions 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. However, we show that the test points toward anti-herding, as opposed to herding, in our sample of professional forecasters of inflation and output growth.

In an attempt to understand these conflicting findings we present a simple model of a group of forecasters which allows for disparate private information and idiosyncratic reporting errors, but in which none of the forecasters pay attention to the views of others, that is, there is no imitation. The model captures the essential features of the forecasting environment in a stylized way, and is simple enough to allow us to relate the population values of the key parameters underpinning the tests in GGJ and BCK to the features of the environment. We show that the pattern of empirical results we obtain across the different tests is consistent with the forecasting environment being characterised by agents' forecasts primarily differing because of idiosyncratic factors or reporting errors (as opposed to private information) and with little concerted imitative behaviour. We then consider how the tests perform when forecaster disagreement is driven by informational rigidities, as stressed in the recent literature.

The plan of the paper is as follows. Section 2 describes the tests of GGJ and BCK, and the related tests that we consider. Section 3 describes the forecast data we analyze for evidence of imitative behaviour, and section 4 records the results. Section 5 describes the model of forecaster heterogeneity we use to examine the properties of the tests, and assesses how well the analytical results match the empirical findings. Section 6 considers the effects of different sources of forecaster heterogeneity, including types of information rigidities, on the outcomes of tests for imitative behaviour. Section 7 concludes.

2 Approaches to Testing for Herding

Gallo *et al.* (2002) (GGJ) exemplify 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. Their test for imitation is based on the following regression[3]:

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

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). GGJ offer an alternative re-parameterisation, given by:

$$\begin{aligned} y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} &= \beta_0 + \beta_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)} \right) + (\beta_1 + \beta_2 - 1)\overline{y}_{t|t-(h+1)} + u_t \\ &= \beta_0 + \gamma_1 \left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)} \right) + \gamma_2 \overline{y}_{t|t-(h+1)} + u_t. \end{aligned} \tag{2}$$

Based on (1) and (2), they argue that γ_1 is expected to be positive,[4] and that for an 'imitation effect' we require $\beta_2 \neq 0$, and for an imitation effect and 'shrinkage to the mean' (i.e., herding), we require $\beta_2 \neq 0$ *and* $\gamma_2 = \beta_1 + \beta_2 - 1 < 0$.

An alternative test for 'shrinkage to the mean' can be obtained by re-parameterising (1)

[3] This is their equation (1), but omitting their additive term in the group variance of the $h+1$ periods ahead forecast.

[4] Gallo *et al.* (2002, p.12): 'the choice of forecast ... can be expected to be consistently on the same side of the sample mean...'.

as:

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \beta_0 + (\beta_1 - 1)\left(y^i_{t|t-(h+1)} - \frac{\beta_2}{1-\beta_1}\overline{y}_{t|t-(h+1)} \right) + u_t \tag{3}$$

and then imposing $\beta_1 + \beta_2 - 1 = 0$. Then a simple test for herding is whether $y^i_{t|t-h} - y^i_{t|t-(h+1)}$ is negatively correlated with $y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)}$, 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:

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \phi_0 + \phi_1\left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+1)} \right) + u_t. \tag{4}$$

While $\phi_1 < 0$ indicates herding, $\phi_1 > 0$ indicates anti-herding. Although the test of ϕ_1 in (4) has been derived from a restricted version of (1), it can also be motivated from testing 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.

These regressions require that $\overline{y}_{t|t-(h+1)}$ is known to each forecaster when they make their forecasts $y^i_{t|t-(h+1)}$, which in turn requires that each forecaster knows the forecasts of the others at that point in time. Otherwise $\overline{y}_{t|t-(h+1)}$ will contain useful information about y_t if the forecasters have private information. If that is the case, then the revision $y^i_{t|t-h} - y^i_{t|t-(h+1)}$ will be systematically related to the explanatory variable, and the population values will be $\beta_2 \neq 0$ and $\phi_1 \neq 0$ even when forecasters do not imitate each other. Empirically, we can either take a stand on whether or not $\overline{y}_{t|t-(h+1)} \in \mathcal{I}^i_{t-(h+1)}$ for all i (where $\mathcal{I}^i_t$ denotes individual i's information set at time t), or more pragmatically, we can report results for different assumptions about agents' information sets. For example, instead of (4) we can estimate:

$$y^i_{t|t-h} - y^i_{t|t-(h+1)} = \phi_0 + \phi_1\left(y^i_{t|t-(h+1)} - \overline{y}_{t|t-(h+2)} \right) + u_t \tag{5}$$

where $\overline{y}_{t|t-(h+1)}$ has been replaced by $\overline{y}_{t|t-(h+2)}$. The only information requirement is that forecasters know the forecasts of others made in the previous period.

Although each forecaster may not know the current forecasts of all the others, 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 we may proceeds as if they knew the current consensus. The conservative approach of using 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.

The second approach to testing for imitation considers the forecast errrors, for example:

$$y_t - y^i_{t|t-h} = \rho_0 + \rho_1 \left(y^i_{t|t-h} - \overline{y}_{t|t-(h+1)} \right) + w_t \tag{6}$$

Equation (6) is a regression-based implementation of the non-parametric test of Bernhardt *et al.* (2006), who suppose that the *reported* forecast of a respondent will be moved towards the consensus relative to the respondent's *private* forecast when a respondent herds.[5] Hence if the private h-step forecast exceeded the $h+1$-step consensus, the reported forecast $y^i_{t|t-h}$ will still exceed 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.

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, $\text{CP}_1 = \Pr\left(y_t < y^i_{t|t-h} \mid \overline{y}_{t|t-(h+1)} < y^i_{t|t-h} \right)$, and the probability that the reported forecast is less than the outcome conditional on the forecast being less than the consensus, $\text{CP}_2 = \Pr\left(y_t > y^i_{t|t-h} \mid \overline{y}_{t|t-(h+1)} > y^i_{t|t-h} \right)$. If we $\gamma^+_\tau = 1$ if $y^i_{t|t-h} > \overline{y}_{t|t-(h+1)}$, and $\gamma^-_\tau = 1$ if $y^i_{t|t-h} < \overline{y}_{t|t-(h+1)}$, and define the joint events as $\delta^+_\tau = 1$ if $y^i_{t|t-h} > \overline{y}_{t|t-(h+1)}$ and $y^i_{t|t-h} > y_t$, and $\delta^-_\tau = 1$ 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:

$$S = \frac{1}{2} \left[\frac{\sum_\tau \delta^+_\tau}{\sum_\tau \gamma^+_\tau} + \frac{\sum_\tau \delta^-_\tau}{\sum_\tau \gamma^-_\tau} \right]. \tag{7}$$

This is asymptotically normally distributed $N\left(0.5, \frac{1}{16} \left[\left(\sum_\tau \gamma^+_\tau \right)^{-1} + \left(\sum_\tau \gamma^-_\tau \right)^{-1} \right] \right)$ under the null of no imitation. Given the relatively small numbers of forecasts available for each respondent and the asymptotic justification of the test, it is not clear how informative the test will be. Nevertheless, the value of S provides an indication of the tendency of an individual to herd (when $S < \frac{1}{2}$) or to anti-herd (when $S > \frac{1}{2}$).

[5] By 'private' is simply meant the forecast which would be made in the absence of any imitation or herding effects.

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.[6] Partly because of its length, it is a popular choice for academic research on expectations. As of April 21 2014, the Academic Bibliography maintained by the Philadelphia Fed listed 79 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.[7]

We call forecasts of the current-quarter[8] horizon $h = 1$ forecasts. Then 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 $h = 1$ 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 imitation described in section 2. 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 typically contain

[6] It began life as began 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).

[7] 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. Some experimentation suggested the results are qualitatively unchanged if we only use the surveys from 1981:Q3 onwards.

[8] 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.

revisions and definitional changes (see e.g., Landefeld, Seskin and Fraumeni (2008) for a discussion of the revisions to US national accounts data).

4 Empirical Findings

Table 1 reports the evidence for imitation based on equations (1) and (2) from the approach suggested by Gallo *et al.* (2002). We assume that at time t each individual knows the consensus forecast based on the forecasts made at time t by all the other forecasters. We run regressions (1) and (2) on all the forecasts of a given horizon for each individual who reported a sufficient number of forecasts.[9] We report the proportion of regressions for which i) we rejected the null that $\beta_2 = 0$ in eqn. (1), (ii) we rejected the null that $\beta_2 = 0$ (in eqn. (1)) and $\gamma_2 = 0$ (in eqn. (2)) in favour of $\beta_2 \neq 0$ and $\gamma_2 < 0$, and (iii) $\gamma_2 = 0$ against $\gamma_2 \neq 0$ (in eqn. (2)).[10] The results indicate an imitation effect ($\beta_2 \neq 0$) for in excess of two-fifths of the forecasters for inflation, and for a third to one half for output growth. However, the results of (ii) indicate that rarely do we find herding as defined by Gallo *et al.* (2002) ($\beta_2 \neq 0$ and $\gamma_2 < 0$). Moreover, (iii) indicates that $\gamma_2 = 0$ is rejected for relatively few forecasters.

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.

Table 1 provides evidence of imitation ($\beta_2 \neq 0$) but little evidence of herding ($\beta_2 \neq 0$ and $\gamma_2 < 0$). Table 2 reports the results of tests based on restricted versions of (1) and (2), such as (4). It gives: (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. The remaining three columns record the average S statistic values (7) across all those respondents for which we found $\phi_1 < 0$, $\phi_1 > 0$ and neither, where the average is the median.

We find that $\phi_1 = 0$ is rejected in favour of $\phi_1 < 0$ for around three quarters of forecasters, for each horizon, for inflation, whereas for output growth the proportion is a little lower at

[9] We require 10 or more forecasts of a given horizon. This reqirement was satisfied by around 150 individuals for each horizon.

[10] All tests were undertaken at the 5% level, including each of the two tests under (ii).

the shortest horizon. There are virtually no instances where $\phi_1 = 0$ is rejected in favour of $\phi_1 > 0$. There does not appear to be a systematic relationship between finding herding based on $\phi_1 < 0$ and a low value of S. For both variables the median value of S is similar for those forecasters for which we reject $\phi_1 = 0$ in favour of $\phi_1 < 0$, and for those forecasters for whom we fail to reject $\phi_1 = 0$.

We report the results of tests based on (6) 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 (eqn. (7)), which exceed one half for in excess of 80% of forecasters for inflation, with only slightly lower proportions for the output growth forecasts. The non-parametric test S generates fewer rejections of the null than the regression implementation (see the last column of the table).

The t-tests of $\rho_1 > 1$ in table 3 are based on autocorrelation and heteroscedasticity consistent standard errors because of the usual issue of overlapping forecasts, whereas for $h = 1$ only a correction for heteroscedasticity is made. The tests in tables 1 and 2 require only a correction for possible heteroscedasticity, even for $h > 1$. To understand why this is the case, consider for example two adjacent observations on the left-hand-side of (4): $y^i_{t|t-h} - y^i_{t|t-(h+1)}$ 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 for $h = 1$.[11]

In sum, between a half and three quarters of all SPF respondents appear to herd if we consider individual-level regressions of forecast revisions on the difference between last period's forecast and the consensus forecast. Radically different findings result from tests on the GGJ regression values of β_2 and γ_2, or compared to the evidence for anti-herding based on the approach of Bernhardt *et al.* (2006).

Table 4 records the key findings when we instead use the lagged consensus (as in (5) and suitably adapted versions of (1) and (2)) reflecting the more conservative assumption that at time t forecasters only know the $t - 1$ consensus.[12] Generally there is less evidence of

[11] This follows immediately if we consider a time series x_t which can be written as $x_t = \psi(L)\varepsilon_t = \sum_{j=1}^{h} \psi_{h-j}\varepsilon_{t-h+j} + \psi_h\varepsilon_{t-h} + \sum_{j=1}^{\infty} \psi_{h+j}\varepsilon_{t-h-j}$. 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 lose the '$h = 4$' results, as these would require $h = 6$ forecasts, and the longest quarterly forecasts in the SPF are $h = 5$.

imitative behaviour - which is of course consistent with last quarter's consensus constituting out-dated information which has less influence on forecaster behaviour. That said, the results are qualitatively unchanged in that: i) tests based on (5) continue to suggest herding (around 60% of inflation forecasters herd, with smaller numbers for output growth), and ii) tests based on GGJ (β_2 and γ_2) suggest a fifth to a third of forecasters imitate, but that few of these herd. There is no need to consider a lagged version of BCK because the implementation reported in the tables already has the consensus lagged one quarter relative to the individual forecast being analyzed.

Hence the information assumption does not drive the divergent findings apparent in tables 1, 2 and 3. Under the conservative assumption, the results based on the forecast revision (equation (5)) suggest around a half of forecasters herd, and those based on GGJ suggest little herding behaviour. Those based on BCK point towards anti-herding.

5 A Simple Model of Forecaster Heterogeneity

In an attempt to understand the conflicting empirical findings in section 4, we consider a model of heterogeneous forecasters, in which there is private information, and forecaster reporting error or noise. In section 6 we discuss the implications of recent models that emphasise informational rigidities.

5.1 Actual and forecast data generating processes

We extend the simple model of Engle (1983) to allow for idiosyncratic error or reporting error. Engle (1983) used the following model to analyze the relationship between the cross-section variance of heterogeneous expectations and ARCH estimates of uncertainty, but it will also serve as the foundation of our model. The variable y_t is generated by a first-order autoregressive process:

$$y_t = \delta y_{t-1} + w_t \tag{8}$$

but each of the N forecasters is privy to private information, so that the 'shock' w_t can be decomposed as $w_t = \eta_t + \sum_{i=1} \alpha_i \varepsilon_{i,t|t-1}$, and it is assumed that $\varepsilon_{i,t|t-1}$ is known by i when forecasting y_t at time $t-1$. So i's forecast is given by $y^i_{t|t-1} = \delta y_{t-1} + \alpha_i \varepsilon_{i,t|t-1}$, where we assume $E\left(\varepsilon^2_{i,t|t-1}\right) = 1$ for all i, which is without loss of generality given that the ε's are multiplied by the α_i's. We also assume that the ε's are serially uncorrelated, are uncorrelated across individuals at all leads and lags, and are uncorrelated with the η's at all leads and lags.

The recent modelling of three-dimensional panels of forecasters[13] suggests a role for an idiosyncratic error in explaining forecaster behaviour. We add an error which is specific to the forecaster i, the target t, and the forecast horizon h. For 1-step ahead, i's forecasts are given by:

$$y^i_{t|t-1} = \delta y_{t-1} + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} \tag{9}$$

where $E\left(v^2_{i,t|t-1}\right) = 1$ all i.

By backward iteration of (8) we obtain:

$$y_t = \delta^h y_{t-h} + \sum_{s=0}^{h-1} \delta^s \eta_{t-s} + \sum_{s=0}^{h-1} \delta^s \sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-s|t-1-s}.$$

Then the h-step ahead forecast of y_t by individual i is given by their conditional expectation of y_t given the available information, plus the idiosyncratic error or 'noise', $\gamma_i v_{i,t|t-h}$, i.e.,

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

The individual's information set $\mathcal{I}^i_{t-h}$ contains public information y_{t-h} and private information $\alpha_i \varepsilon_{i,t-h+1|t-h}$. Then the forecast is given by:

$$y^i_{t|t-h} = \delta^h y_{t-h} + \delta^{h-1} \alpha_i \varepsilon_{i,t-h+1|t-h} + \gamma_i v_{i,t|t-h}. \tag{10}$$

The consensus forecast is assumed to be the simple average of the individual forecasts. Hence:

$$\overline{y}_{t|t-h} = \delta^h y_{t-h} + \delta^{h-1} \frac{1}{N} \sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-h+1|t-h} + \frac{1}{N} \sum_{i=1}^{N} \gamma_i v_{i,t|t-h}.$$

In this model, $\overline{y}_{t|t-h} \notin \mathcal{I}^i_{t-h}$, because $\mathcal{I}^i_{t-h}$ does not include $\left\{\alpha_i \varepsilon_{j,t-h+1|t-h}\right\}$ $j = 1, \ldots, N$ except for $j = i$. This means that when testing for imitation using (1), (2) or (4) the consensus forecasts must be lagged one more period than the longer horizon individual

[13] That is, allowing for multiple forecasters, multiple targets and multiple forecast horizons: see e.g., Davies and Lahiri (1995), and the review article by Davies, Lahiri and Sheng (2011) for further discussion.

forecast. That is, $\mathcal{I}_{t-h}^{i}$ includes $\overline{y}_{t|t-(h+1)}$ but not $\overline{y}_{t|t-h}$. Regressing the revision $y_{t|t-h}^{i} - y_{t|t-(h+1)}^{i}$ on (say) $y_{t|t-(h+1)}^{i} - \overline{y}_{t|t-(h+1)}$ would yield a negative coefficient and falsely indicate herding, simply because $\overline{y}_{t|t-(h+1)}$ contains relevant information (about y_t) not contained in $y_{t|t-(h+1)}^{i}$. In this case the appropriate regression is $y_{t|t-h}^{i} - y_{t|t-(h+1)}^{i}$ on $y_{t|t-(h+1)}^{i} - \overline{y}_{t|t-(h+2)}$, as analyzed below.

We consider $h = 1$ in the analysis below to make matters concrete, but the analysis holds for any h.

The 1-step forecast is given by (9), and the 2-step and 3-step ahead forecasts from (10) as:

$$y_{t|t-2}^{i} = \delta^2 y_{t-2} + \delta\alpha_i\varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2}, \tag{11}$$

and:

$$y_{t|t-3}^{i} = \delta^3 y_{t-3} + \delta^2\alpha_i\varepsilon_{i,t-2|t-3} + \gamma_i v_{i,t|t-3}. \tag{12}$$

The consensus forecasts are the cross-section average of (12) when we consider tests based on (1), (2) or (4).

5.2 Population values of the parameters in the regression-based tests of imitation

5.2.1 Tests of forecast revisions

Firstly, the population value of the slope parameter ϕ_1 in (5) is defined by:

$$\phi_1 = \frac{Cov\left(y_{t|t-1}^{i} - y_{t|t-2}^{i}, y_{t|t-2}^{i} - \overline{y}_{t|t-3}\right)}{Var\left(y_{t|t-2}^{i} - \overline{y}_{t|t-3}\right)}. \tag{13}$$

In the Appendix we calculate the relevant moments for the model in section 5.1 and show that:

$$\phi_1 \simeq \frac{-\gamma_i^2}{\gamma_i^2 + \delta^4\sigma_\eta^2 + \delta^4\alpha^2}, \tag{14}$$

where we have made the simplifying assumptions that i) N is large and ii) the news component is approximately equal across forecasters, and $\alpha_i = \alpha/\sqrt{N}$, so that the importance of news (α_i^2) for each individual is inversely related to the number of forecasters. This assumption ensures that the relative magnitudes of the components η_t and $\sum_{i=1}\alpha_i\varepsilon_{i,t|t-1}$ of the overall variance of w_t remain fixed as N increases.

From (14) it is apparent that $-1 < \phi_1 \leq 0$, and will equal zero when the reporting error is zero. Alternatively, if the variance of the reporting error is large relative to the average news or the variance of the underlying shocks, ϕ_1 will tend to -1.

We can investigate the implications of falsely assuming the forecasters know the forecasts of others. In our framework, this amounts to running the regression (4) instead of (5). Then the population value of ϕ_1 in (4) is:

$$\widetilde{\phi}_1 = \frac{Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-2}\right)}{Var\left(y^i_{t|t-2} - \overline{y}_{t|t-2}\right)} \simeq -1, \tag{15}$$

so that assuming forecasters known more than they do will manifest as herding behaviour.

5.2.2 Gallo *et al.* (2002) approach

Next, consider the approach of GGJ. For the setup in section 5.1, we need to amend (1) and (2) to:

$$y^i_{t|t-1} = \beta_0 + \beta_1 y^i_{t|t-2} + \beta_2 \overline{y}_{t|t-3} + u_t \tag{16}$$

$$y^i_{t|t-1} - \overline{y}_{t|t-3} = \beta_0 + \gamma_1\left(y^i_{t|t-2} - \overline{y}_{t|t-3}\right) + \gamma_2 \overline{y}_{t|t-3} + u_t. \tag{17}$$

Then in the Appendix we show that under the simplifying assumptions we have made we obtain:

$$\beta_2 \simeq \frac{\gamma_i^2}{\gamma_i^2 + \delta^4\left(\sigma_\eta^2 + \alpha^2\right)}, \;\; \gamma_2 \simeq 0.$$

Suppose we incorrectly use $\overline{y}_{t|t-2}$ in these regressions, and estimate β_2 in (16) and γ_2 in (17). Then we can show that:

$$\widetilde{\beta}_2 \simeq 1, \;\; \widetilde{\gamma}_2 \simeq 0.$$

Thus the incorrect informational assumption results in $\widetilde{\beta}_2$ being non-zero and positive even in the absence of individual idiosyncratic errors, although $\widetilde{\gamma}_2 \simeq 0$. Hence the GGJ approach is also susceptible to indicating imitative behaviour when we mistakenly assuming forecasters know more about the forecasts of others than they do.

5.2.3 Bernhardt *et al.* (2006) approach

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

$$y_t - y^i_{t|t-1} = \rho_0 + \rho_1 \left(y^i_{t|t-1} - \overline{y}_{t|t-2} \right) + w_t. \quad (18)$$

Straightforward calculations result in:

$$\rho_1 \simeq \frac{-\gamma_i^2}{\gamma_i^2 + \delta^2\sigma_\eta^2 + \delta^2\alpha^2}.$$

5.3 A re-consideration of the empirical results

The results in section 5.2 suggest that in the absence of private information (so that forecaster disagreement is driven solely by idiosyncratic reporting errors):

i)

$$\phi_1 \simeq \frac{-\gamma_i^2}{\gamma_i^2 + \delta^4\sigma_\eta^2} < 0$$

and will be close to -1 if $\delta^4\sigma_\eta^2$ is small relative to γ_i^2;

ii)

$$\beta_2 \simeq \frac{\gamma_i^2}{\gamma_i^2 + \delta^4\sigma_\eta^2}, \quad \gamma_2 \simeq 0$$

so that β_2 will be of equal size but the opposite sign to ϕ_1, and γ_2 will be close to zero; and,

iii)

$$\rho_1 \simeq \frac{-\gamma_i^2}{\gamma_i^2 + \delta^2\sigma_\eta^2}$$

so that $-1 < \phi_1 < \rho_1 < 0$ (because the implicit assumption that y_t is integrated of order zero implies that $-1 < \delta < 1$).

On the other hand, when the differences between forecasters are solely due to private information (and reporting errors are absent), it follows immediately from the expressions in section 5.2 that none of the approaches will suggest imitative behaviour, provided the consensus forecast is appropriately defined (i.e., such that it is known to all).

Recall that $\phi_1 < 0$ ($\phi_1 > 0$) indicates herding (anti-herding), while $\rho_1 < 0$ ($\rho_1 > 0$) indicates anti-herding (herding). Conditional on $\beta_2 \neq 0$, $\gamma_2 < 0$ indicates herding. Thus the pattern of results in section 4 - namely, herding based on equation (5), evidence of imitative behaviour based on $\beta_2 \neq 0$ but little evidence of $\gamma_2 < 0$ (GGJ approach), and anti-herding from BCK - is consistent with an absence of herding behaviour[14] *and* the differences between forecasters primarily reflecting 'noise' or reporting errors, as opposed to information that would reduce forecast error.

6 Alternative models of forecaster heterogeneity

In our simple model forecaster disagreement can be driven by private information or noise. A key distinction between the two is of course that the former is consistent with forecaster efficiency, in the sense that the forecasts are uncorrelated with the forecast errors,[15] whereas noise results in forecasts which violate this property. For example, the 1-step ahead forecast error is:

$$y_t - y^i_{t|t-1} = \eta_t + \sum_{j \neq i} \alpha_j \varepsilon_{j,t|t-1} - \gamma_i v_{i,t|t-1}$$

so that:

$$Cov\left(y_t - y^i_{t|t-1}, y^i_{t|t-1}\right) = -\gamma_i^2,$$

which is zero in the absence of noise. The model of Davies and Lahiri (1995) only allows forecasters to disagree because of fixed individual-specific biases and 'idiosyncratic errors', where the latter correspond to our $v_{i,t|t-h}$. Davies and Lahiri (1995, p. 209) suggest that these capture ' 'other factors' (e.g., private information, measurement error, etc.) ' but notice they cannot be interpreted as private information in their setup, as commonly understood, because they are correlated with the forecast errors.

Over the last decade or so a number of models have been proposed to explain forecaster heterogeneity, some of which suppose the existence of informational rigidities. In this section we explore the implications of these models for testing for forecaster herding. When these theoretical models can be mapped into private information or idiosyncratic errors, the implications for testing for herding follow immediately. Section 6.1 and 6.2 consider two types of information rigidities, section 6.3 forecaster heterogeneity due to agents having different loss functions, and section 6.4 heterogeneity arising from forecasters having different priors.

---

[14] If herding behaviour were present it would be signalled by the GGJ approach, i.e., rejection of $\gamma_2 = 0$.

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

6.1 Information Rigidities I: Sticky Information

We begin with the first of two models of informational rigidities: sticky information. See *inter alia* Mankiw and Reis (2002) and Mankiw, Reis and Wolfers (2003), and Coibion and Gorodnichenko (2012). 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. Working backwards, a proportion λ of the pool of forecasters will not use the latest information, and of these, a fraction $1-\lambda$ will have updated in the previous period, and so on. Hence in terms of forecasting y_t h-steps ahead, the mean forecast across individuals is given by:

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

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

It might appear that sticky information is akin to forecasters possessing private information, with the immediate implication that forecaster heterogeneity due to sticky information will not falsely show up as imitative behaviour using the tests described in section 2. To see why, notice that a proportion $(1-\lambda)$ of forecasters use the information set $\mathcal{I}^i_{t-h} = \{y_{t-h}, y_{t-h-1}, \ldots\}$, a proportion $\lambda(1-\lambda)$ use the information set $\mathcal{I}^i_{t-(h+1)} = \{y_{t-h-1}, \ldots\}$ etc., so that the first group can be thought of as possessing 'private information' y_{t-h}, relative to the second, and so on. Knowledge of y_{t-h} is clearly information in that it reduces forecast error assuming y_t follows an autoregression. However also relevant is whether the consensus forecast used in the herding test provides useful information compared to the longer-horizon individual forecast. If we allow that individuals are heterogeneous in terms of the frequency with which they update their forecasts, and consider an individual with a high λ (i.e., low average frequency of updating), then the mean forecast will contain useful information and the herding test will typically indicate imitative behaviour (as shown by the results for $\widetilde{\phi}_1$ and $\widetilde{\beta}_2$ in section 5.2).

6.2 Information Rigidities II: Noisy Information

The second model of informational rigidities assumes agents base their forecasts on the latest information, but only ever observe noisy signals about economic fundamentals: see Woodford (2001), 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:

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

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:

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

with $v_{it} \sim iidN\left(0, \sigma_v^2\right)$, $\eta_t \sim iidN\left(0, \sigma_\eta^2\right)$, $E\left(\eta_t v_{is}\right) = 0\ \forall i, t, s$. 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:

$$y^i_{t|t} = \left(1 - PH\right) \delta y^i_{t-1,t-1} + PHy_t + P_v v^i_t + P_\eta \eta_t \tag{19}$$

where $H = [1\ 1]'$, and $P = [P_\eta, P_v]$ is the gain of the filter.

The forecasters have private information, in the form of private signals, but notice that the consensus forecast will always contain valuable information relative to an individual's longer-horizon forecasts used in the herding tests, so that noisy information will (wrongly) show up as imitative behaviour. This is a consequence of the assumption that each forecaster i's information set consists only of his set of signals $\{z_{it}, z_{it-1}, \ldots\}$. Both contemporaneous and lagged forecasts of others contain useful information.

One could envisage 'private' forecasts being based on private signals, as in (19), 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, Coibion and Gorodnichenko (2012) find that the noisy information model generally does a good job of explaining the observed responses of mean inflation forecast errors and disagreement to shocks for professional forecasters (as well as for other agents).

We calculate the population value of ϕ_1 for regression equation (5) to illustrate the effect of noisy information (details in the appendix), and show the rejection of the null hypothesis of no imitation. Using the consensus dated $t-2$ we find:

$$\phi_1 = \frac{Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-2}\right)}{Var\left(y^i_{t|t-2} - \overline{y}_{t|t-2}\right)} = -PH \tag{20}$$

and using the $t-3$-consensus:

$$\begin{aligned} \phi_1 &= \frac{Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-3}\right)}{Var\left(y^i_{t|t-2} - \overline{y}_{t|t-3}\right)} \\ &= \frac{PH\left[1-(1-PH)^2\delta^2\right]^{-1}\left((1-PH)PH\sigma^2_w - P^2_v\sigma^2_v + \delta^2 PH(1-PH)P^2_\eta\sigma^2_\eta\right) - PHP^2_\eta\sigma^2_\eta}{\left[1-(1-PH)^2\delta^2\right]^{-1}\left(P^2_v\sigma^2_v + (PH)^2\sigma^2_w + \delta^2 P^2_\eta\sigma^2_\eta\right) + P^2_\eta\sigma^2_\eta} \end{aligned} \tag{21}$$

Note that $PH \in (0,1)$, and that the degree of information rigidity declines as $PH \to 1$ (because more weight is placed on z_{it}, and consequently the innovation in the inflation process w_t). Hence (20) indicates that the false rejection of the null of no imitation will be exacerbated the lower the degree of information rigidity. Intuitively, individual forecasters are placing more weight on their private signals, y_{it}, so the consensus forecast which aggregates the private information is a more valuable predictor of inflation, and so $y^i_{t|t-1} - y^i_{t|t-2}$ is more highly correlated with $y^i_{t|t-2} - \overline{y}_{t|t-2}$. The expression (21) using the $t-3$-consensus becomes more interpretable if we assume only private signals (that is, we abstract form the common signal s_t). This amounts to setting $P_v = P$, $H = 1$, $P^2_\eta = 0$, $\sigma^2_\eta = 0$. In that case:

$$\phi_1 = \frac{\left((1-P)\sigma^2_w - P\sigma^2_v\right)}{\left(\sigma^2_v + \sigma^2_w\right)}.$$

Again, $P \in (0,1)$, and with low informational rigidity $\phi_1 < 0$, as when the current consensus is used. Only when the private signal is noiseless ($\sigma^2_v = 0$) so that all observe y_t will ϕ_1 equal zero, but in this case there is no heterogeneity.

6.3 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[16] 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:

$$y^i_{t|t-1} = \mu_{t|t-1} + \frac{\varphi_i}{2}\sigma^2_{t|t-1}. \tag{22}$$

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 reporting errors, when evaluated using a squared error loss function, so will tend to (falsely) indicate imitative behaviour.[17]

6.4 Heterogeneous beliefs about long-run outcomes

Patton and Timmermann (2010, 2011) 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 can sharpen the analysis by assuming, say $E\left(y_t \mid \mathcal{I}^i_{t-1}\right) = \mu_{t|t-1}$, for all i. But this is weighted by each forecaster with their prior for the long-run growth rate, μ_i, to give agent i's reported forecast as:

$$\widetilde{y}_{t|t}(i) = \omega\mu_i + (1-\omega)\,\mu_{t|t-1}.$$

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

[17] Elliott *et al.* (2008) and Clements (2014) present evidence for and against survey respondents having asymmetric loss, respectively.

The cross-sectional distribution of μ_i's will again show up as idiosyncratic errors or reporting errors, and as suggested by the analysis in section 5, may (falsely) indicate imitative behaviour.

7 Conclusions

There is much interest in the literature in 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. It is important that the past consensus is known to the individual, otherwise it would typically constitute a useful source of information and would influence the forecasts of a forecaster solely motivated by maximizing forecast accuracy.

We consider approaches to testing forecaster herding based on this and related ideas, and find that a significant proportion of professional US macroeconomists appear to be influenced by their fellow forecasters, but that some approaches suggest respondents deliberately exaggerate their differences, while others suggest herding is the dominant type of imitative behaviour. We use a simple model of forecaster heterogeneity to show that if differences amongst forecasters primarily arise because of private information, then imitative behaviour will not be flagged when forecasts do not imitate. However, if differences arise because of noise or reporting errors, then all the tests will suggest imitative behaviour even when it is absent. Note that noise can be understood in a wide sense of denoting differences which are not attributable to news or information: formally, components of forecasts which are perfectly correlated with the forecast error.

The pattern of empirical results we obtain across the different tests of herding is shown to be consistent with differences between forecasters arising primarily from idiosyncratic factors or reporting errors rather than imitative behaviour.

We consider a number of models of forecaster heterogeneity which have recently been proposed, including various informational rigidities and differences in loss functional and priors concerning long-run values of variables. Informational rigidities will tend to show up as herding because the consensus forecast - assumed absent from individuals' information sets - constitutes useful information for predicting the object of interest. The recent empirical support for the noisy information model reported by Coibion and Gorodnichenko (2012) suggests the tests for herding we analyze may falsely signal herding when in fact noisy information is the explanation. It is also straightforward to show that forecaster heterogeneity arising from forecasters with loss functions characterized by differing degrees of asymmetry, as well as forecasts which incorporate different prior beliefs, may falsely indicate imitative behaviour.

Table 1: Tests of herding based on Gallo, Granger and Jeon (2002)

h	No. Regns	$\beta_2 \neq 0$	$\beta_2 \neq 0$ and $\gamma_2 < 0$	$\gamma_2 \neq 0$
		Inflation		
4	153	0.44	0.05	0.16
3	160	0.42	0.06	0.09
2	160	0.46	0.06	0.11
1	160	0.42	0.08	0.13
		Output growth		
4	151	0.35	0.01	0.11
3	161	0.45	0.02	0.14
2	161	0.52	0.02	0.23
1	161	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 are the proportion of regressions for which the null is rejected in favour of the specified alternative.

Table 2: Tests of herding based on the forecast revision

h	No. Regns	Proportion of regressions			Median S across respondents		
		$\phi_1 < 0$	$\phi_1 > 0$	Neither	given $\phi_1 < 0$	given $\phi_1 > 0$	given Neither
				Inflation			
4	146	0.79	0.00	0.21	0.62	.	0.62
3	150	0.75	0.00	0.25	0.61	.	0.61
2	149	0.78	0.00	0.22	0.61	.	0.63
1	151	0.74	0.01	0.26	0.64	0.34	0.64
				Output growth			
4	148	0.74	0.00	0.26	0.62	.	0.61
3	153	0.70	0.00	0.30	0.59	.	0.55
2	155	0.69	0.01	0.30	0.61	0.56	0.55
1	156	0.62	0.03	0.36	0.60	0.49	0.59

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 are the proportion of regressions for which the null is rejected in favour of the specified alternative.

Table 3: Tests of herding based on Bernhardt *et al.* (2006)

h	No. Regns	Proportion of regressions			Proportion of regressions		
		$\rho_1 < 0$	$\rho_1 > 0$	Neither	$S < 0.5$	$S > 0.5$	NP Rejects
				Inflation			
4	146	0.57	0.01	0.42	0.14	0.84	0.25
3	150	0.61	0.03	0.37	0.15	0.83	0.17
2	149	0.58	0.01	0.42	0.14	0.82	0.16
1	151	0.68	0.02	0.30	0.12	0.85	0.25
				Output growth			
4	148	0.59	0.01	0.39	0.18	0.80	0.22
3	153	0.39	0.03	0.58	0.26	0.71	0.14
2	155	0.32	0.03	0.66	0.21	0.78	0.20
1	156	0.33	0.04	0.63	0.20	0.77	0.19

The test regression is:

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

Table 4: Tests of herding based on the lagged consensus

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

The test regressions are:

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

and:

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

and

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

The entries in the table are the proportion of regressions for which the null is rejected in favour of the specified alternative.

References

Bernhardt, Dan., Campello, Murillo., and Kutsoati, Edward. (2006). Who herds?. *Journal of Financial Economics*, *80*(3), 657–675.

Capistrán, Carlos., and Timmermann, Allan. (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. (2014). US inflation expectations and heterogeneous loss functions, 1968–2010. *Journal of Forecasting*, **33(1)**, 1–14.

Coibion, Olivier., and Gorodnichenko, Yuriy. (2012). What can survey forecasts tell us about information rigidities?. *Journal of Political Economy*, *120*(1), 116 – 159.

Croushore, Dean. (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review*, **"November"**, 3–13.

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

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

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

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

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

Engle, Robert. F. (1983). Estimates of the variance of U.S inflation based upon the ARCH model. *Journal of Money, Credit and Banking*, **15**, 286–301.

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

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

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

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

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

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

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

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

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

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

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

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

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

Pierdzioch, Christian., and Rülke, Jan-Chtristophe. (2012). Forecasting stock prices: Do forecasters herd?. *Economics Letters*, *116*(3), 326–329.

Pierdzioch, Christophe., Rülke, Jan Christophe., and Stadtmann, Georg. (2010). New evidence of anti-herding of oil-price forecasters. *Energy Economics*, *32*(6), 1456–1459.

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

Woodford, Michael. (2001). Imperfect common knowledge and the effects of monetary policy. In Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M. (eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In honor of Edmund Phelps*, pp. 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.

Assessing the Evidence of Macroeconomic Forecaster Herding: An Application to Forecasts of Inflation and Output Growth **Appendix**.

Michael P. Clements
ICMA Centre,
Henley Business School,
University of Reading,
Reading RG6 6BA

October 9, 2014

1 Appendix

1.1 Test of forecast revisions

The population value of the slope parameter ϕ_1 in:

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

is defined by:

$$\phi_1 = \frac{Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-3} \right)}{Var\left(y^i_{t|t-2} - \overline{y}_{t|t-3} \right)}. \tag{2}$$

We have:

$$y^i_{t|t-1} - y^i_{t|t-2} = \delta \left(y_{t-1} - \delta y_{t-2} \right) + \alpha_i \varepsilon_{i,t|t-1} - \delta \alpha_i \varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-1} - \gamma_i v_{i,t|t-2}$$

and substituting for $y_{t-1} - \delta y_{t-2}$ from $y_t = \delta y_{t-1} + \eta_t + \sum_{i=1} \alpha_i \varepsilon_{i,t|t-1}$ gives:

$$\begin{aligned} y^i_{t|t-1} - y^i_{t|t-2} &= \delta \eta_{t-1} + \delta \sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} + \alpha_i \varepsilon_{i,t|t-1} - \delta \alpha_i \varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-1} - \gamma_i v_{i,t|t-2} \\ &= \delta \eta_{t-1} + \delta \sum_{j \neq i}^{N} \alpha_j \varepsilon_{j,t-1|t-2} + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} - \gamma_i v_{i,t|t-2}. \end{aligned} \tag{3}$$

The term $y^i_{t|t-2} - \overline{y}_{t|t-3}$ can be written as:

$$\begin{aligned} y^i_{t|t-2} - \overline{y}_{t|t-3} &= \delta^2 y_{t-2} + \delta\alpha_i\varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2} \\ &\quad - \left(\delta^3 y_{t-3} + \delta^2 \frac{1}{N}\sum_{i-1}^{N} \alpha_i\varepsilon_{i,t-2|t-3} + \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3}\right) \\ &= \delta^2\eta_{t-2} + \delta^2\sum_{j=1}^{N}\alpha_j\varepsilon_{j,t-2|t-3} + \delta\alpha_i\varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2} - \delta^2\frac{1}{N}\sum_{i=1}^{N}\alpha_i\varepsilon_{i,t-2|t-3} - \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3} \end{aligned}$$

Assuming that the ε's, v's and η's are independent at all leads and lags, and $Var(\varepsilon) = 1$ and $Var(v) = 1$ for all i, t, we obtain:

$$Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-3}\right) = -\gamma_i^2$$

The denominator in (2) is:

$$\begin{aligned} Var\left(y^i_{t|t-2} - \overline{y}_{t|t-2}\right) &= \delta^4\sigma_\eta^2 + \delta^4\left(\frac{N-1}{N}\right)^2\sum_{i=1}^{N}\alpha_i^2 + \delta^2\alpha_i^2 + \gamma_i^2 + \frac{1}{N^2}\sum_{i=1}^{N}\gamma_i^2 \\ &\simeq \delta^4\sigma_\eta^2 + \delta^4\alpha^2 + \gamma_i^2 \end{aligned}$$

where the second line assumes i) N is large and ii) the news component is approximately equal across forecasters, and $\alpha_i = \alpha/\sqrt{N}$, which we assume throughout. Hence we drop terms which are of order N^{-1}, such as α_i^2, for example. Hence:

$$\phi_1 = \frac{-\gamma_i^2}{\delta^4\sigma_\eta^2 + \delta^4\alpha^2 + \gamma_i^2}. \tag{4}$$

When we falsely assume $\overline{y}_{t|t-2}$ is in the forecasters' information sets, the population value of the slope parameter ϕ_1 is:

$$\widetilde{\phi}_1 = \frac{Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-2}\right)}{Var\left(y^i_{t|t-2} - \overline{y}_{t|t-2}\right)}. \tag{5}$$

The forecast revision is given by (3), and $y^i_{t|t-2} - \overline{y}_{t|t-2}$ can be written as:

$$y^i_{t|t-2} - \overline{y}_{t|t-2} = \delta\left(\alpha_i\varepsilon_{i,t-1|t-2} - \frac{1}{N}\sum_{i=1}^{N}\alpha_i\varepsilon_{i,t-1|t-2}\right) + \left(\gamma_i v_{i,t|t-2} - \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-2}\right) \tag{6}$$

so that the numerator of $\widetilde{\phi}_1$ is:

$$\begin{aligned}
Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-2}\right) &= -Cov\left(\delta \sum_{j\neq i}^{N} \alpha_j \varepsilon_{j,t-1|t-2}, \delta \frac{1}{N} \sum_{i=`}^{N} \alpha_i \varepsilon_{i,t-1|t-2}\right) \\
&\quad -Cov\left(\gamma_i v_{i,t|t-2}, \gamma_i v_{i,t|t-2} - \frac{1}{N}\sum_{i=1}^{N} \gamma_i v_{i,t|t-2}\right) \\
&= -\delta^2 \frac{1}{N} \sum_{j\neq i}^{N} \alpha_j^2 - \gamma_i^2 \left(\frac{N-1}{N}\right) \\
&\simeq -\gamma_i^2
\end{aligned}$$

The denominator in (5) is:

$$\begin{aligned}
Var\left(y^i_{t|t-2} - \overline{y}_{t|t-2}\right) &= \delta^2 \left[\frac{1}{N^2} \sum_{j\neq i}^{N} \alpha_j^2 + \left(\frac{N-1}{N}\right)^2 \alpha_i^2\right] + \frac{1}{N^2} \sum_{j\neq i}^{N} \gamma_j^2 + \gamma_i^2 \left(\frac{N-1}{N}\right)^2 \\
&\simeq \delta^2 \frac{1}{N^2} \left(\frac{N-1}{N}\right) \alpha^2 + \delta^2 \left(\frac{N-1}{N}\right)^2 \frac{\alpha^2}{N} + \frac{1}{N^2} \sum_{j\neq i}^{N} \gamma_j^2 + \gamma_i^2 \left(\frac{N-1}{N}\right)^2 \\
&\simeq \gamma_i^2
\end{aligned}$$

Hence:

$$\phi_1 \simeq -1.$$

1.2 GGJ tests of imitation

The tests are based on:

$$y^i_{t|t-1} = \alpha + \beta_1 y^i_{t|t-2} + \beta_2 \overline{y}_{t|t-3} + u_{t,h}. \tag{7}$$

$$y^i_{t|t-1} - \overline{y}_{t|t-3} = \alpha + \gamma_1 \left(y^i_{t|t-2} - \overline{y}_{t|t-3}\right) + \gamma_2 \overline{y}_{t|t-3} + u_{t,h}. \tag{8}$$

where relative to GGJ we have lagged the consensus forecast by an extra period because of the assumed information structure.

Consider generic regression, $y = \alpha + \beta_1 x + \beta_2 z + error$. Then:

$$\beta_2 = \frac{V(x)\, C(y,z) - C(x,z)\, C(y,x)}{V(x)\, V(z) - (C(x,z))^2} \tag{9}$$

Consider β_2 in (7).

$$y = y^i_{t|t-1} = \delta y_{t-1} + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1}$$

$$x = y^i_{t|t-2} = \delta^2 y_{t-2} + \delta \alpha_i \varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2}$$

$$z=\overline{y}_{t|t-3}=\delta^3 y_{t-3}+\delta^2\frac{1}{N}\sum_{i=1}^{N}\alpha_i\varepsilon_{i,t-2|t-3}+\frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3}$$

To calculate (9) the following moments are required:

$$V(x)=\delta^4 V_y+\delta^2\alpha_i^2+\gamma_i^2$$

$$\begin{aligned} C(y,z) &= \delta^4 C_{y_{13}}+\delta^3\frac{1}{N}C(y_{t-1},\sum_{i=1}^{N}\alpha_i\varepsilon_{i,t-2|t-3}) \\ &= \delta^4 C_{y_{13}}+\delta^4\frac{1}{N}\sum_{i=1}^{N}\alpha_i^2 \end{aligned}$$

where $C_{ij}=Cov\left(y_{t-i},y_{t-j}\right)$,

$$\begin{aligned} C(x,z) &= \delta^5 C_{12}+\frac{1}{N}\delta^4 Cov\left(y_{t-2},\sum_{i=1}^{N}\alpha_i\varepsilon_{i,t-2|t-3}\right) \\ &= \delta^5 C_{12}+\frac{1}{N}\delta^4\sum_{i=1}^{N}\alpha_i^2 \end{aligned}$$

$$C(y,x)=\delta^3 C_{12}+\delta^2\alpha_i^2$$

$$V(z)=\delta^6 V_y+\delta^4\frac{1}{N^2}\sum_{i=1}^{N}\alpha_i^2+\frac{1}{N^2}\sum_{i=1}^{N}\gamma_i^2$$

Substitution into (9) gives:

$$\beta_2=\frac{\left(\delta^4 V_y+\delta^2\alpha_i^2+\gamma_i^2\right)\left(\delta^4 C_{y_{13}}+\delta^4\frac{1}{N}\sum\limits_{i=1}^{N}\alpha_i^2\right)-\left(\delta^5 C_{12}+\frac{1}{N}\delta^4\sum\limits_{i=1}^{N}\alpha_i^2\right)\left(\delta^3 C_{y_{12}}+\delta^2\alpha_i^2\right)}{\left(\delta^4 V_y+\delta^2\alpha_i^2+\gamma_i^2\right)\left(\delta^6 V_y+\delta^4\frac{1}{N^2}\sum\limits_{i=1}^{N}\alpha_i^2+\frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right)-\left(\delta^5 C_{12}+\frac{1}{N}\delta^4\sum\limits_{i=1}^{N}\alpha_i^2\right)^2}.$$

Substituting $V_y=Var\left(y_t\right)=\delta^2 Var\left(y_{t-1}\right)+\sigma_\eta^2+\sum_{i=1}^{N}\alpha_i^2$, so $V_y=\left(1-\delta^2\right)^{-1}\left(\sigma_\eta^2+\sum\limits_{i=1}^{N}\alpha_i^2\right)$, $C_{12}=\delta V_y$, $C_{13}=\delta^2 V_y$ gives:

$$\beta_2=\frac{\left(\delta^4 V_y+\delta^2\alpha_i^2+\gamma_i^2\right)\left(\delta^6 V_y+\delta^4\frac{1}{N}\sum\limits_{i=1}^{N}\alpha_i^2\right)-\left(\delta^6 V_y+\frac{1}{N}\delta^4\sum\limits_{i=1}^{N}\alpha_i^2\right)\left(\delta^4 V_y+\delta^2\alpha_i^2\right)}{\left(\delta^4 V_y+\delta^2\alpha_i^2+\gamma_i^2\right)\left(\delta^6 V_y+\delta^4\frac{1}{N^2}\sum\limits_{i=1}^{N}\alpha_i^2+\frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right)-\left(\delta^6 V_y+\frac{1}{N}\delta^4\sum\limits_{i=1}^{N}\alpha_i^2\right)^2}.$$

Dropping terms which vanish in N:

$$\beta_2\simeq\frac{\gamma_i^2}{\gamma_i^2+\delta^4 V_y-\delta^6 V_y}\simeq\frac{\gamma_i^2}{\gamma_i^2+\delta^4\left(\sigma_\eta^2+\alpha^2\right)}.$$

Consider γ_2 in (8). Using the generic regression, $y = \alpha + \beta_1 x + \beta_2 z + error$, let:

$$\begin{aligned} y &= y^i_{t|t-1} - \overline{y}_{t|t-3} = \delta\left(y_{t-1} - \delta^2 y_{t-3}\right) + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} - \delta^2 \frac{1}{N}\sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-2|t-3} - \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3} \\ &= \delta^2 \eta_{t-2} + \delta \eta_{t-1} + \delta^2 \left(\frac{N-1}{N}\right)\sum_{i=1}^{N}\alpha_i \varepsilon_{i,t-2|t-3} + \delta \sum_{i=1}^{N}\alpha_i \varepsilon_{i,t-1|t-2} + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} - \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3} \end{aligned}$$

$$\begin{aligned} x &= y^i_{t|t-2} - \overline{y}_{t|t-3} = \delta^2\left(y_{t-2} - \delta y_{t-3}\right) + \delta\alpha_i \varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2} - \delta^2 \frac{1}{N}\sum_{i=1}^{N}\alpha_i \varepsilon_{i,t-2|t-3} - \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3} \\ &= \delta^2 \eta_{t-2} + \delta^2\left(\frac{N-1}{N}\right)\sum_{i=1}^{N}\alpha_i \varepsilon_{i,t-2|t-3} + \delta\alpha_i \varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2} - \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3} \end{aligned}$$

$$z = \overline{y}_{t|t-3} = \delta^3 y_{t-3} + \delta^2 \frac{1}{N}\sum_{i=1}^{N}\alpha_i \varepsilon_{i,t-2|t-3} + \frac{1}{N}\sum_{i=1}^{N}\gamma_i v_{i,t|t-3}.$$

Then the required moments are:

$$V(x) = \delta^4 \sigma^2_\eta + \delta^4 \left(\frac{N-1}{N}\right)^2 \sum_{i=1}^{N}\alpha_i^2 + \delta^2\alpha_i^2 + \gamma_i^2 + \frac{1}{N^2}\sum_{i=1}^{N}\gamma_i^2$$

$$C(y,z) = \delta^4\left(\frac{N-1}{N^2}\right)\sum_{i=1}^{N}\alpha_i^2 - \frac{1}{N^2}\sum\nolimits_{i=1}^{N}\gamma_i^2$$

$$C(x,z) = \delta^4\left(\frac{N-1}{N^2}\right)\sum_{i=1}^{N}\alpha_i^2 - \frac{1}{N^2}\sum\nolimits_{i=1}^{N}\gamma_i^2$$

$$C(y,x) = \delta^4 \sigma^2_\eta + \delta^4\left(\frac{N-1}{N}\right)^2 \sum_{i=1}\alpha_i^2 + \delta^2\alpha_i^2 + \frac{1}{N^2}\sum_{i=1}^{N}\gamma_i^2\alpha_v^2$$

$$V(z) = \delta^6 V_y + \delta^4 \frac{1}{N^2}\sum_{i=1}^{N}\alpha_i^2 + \frac{1}{N^2}\sum\nolimits_{i=1}^{N}\gamma_i^2$$

and substitution into the expression for γ_2 gives:

$$\begin{aligned} \gamma_2 &= \frac{\mathbf{A}\left[\left(\delta^4\sigma^2_\eta + \delta^4\left(\frac{N-1}{N}\right)^2 \sum\limits_{i=1}^{N}\alpha_i^2 + \delta^2\alpha_i^2 + \gamma_i^2 + \frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right) - \left(\delta^4\sigma^2_\eta + \delta^4\left(\frac{N-1}{N}\right)^2\sum\limits_{i=1}^{N}\alpha_i^2 + \delta^2\alpha_i^2 + \frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right)\right]}{\left(\delta^4\sigma^2_\eta + \delta^4\left(\frac{N-1}{N}\right)^2\sum\limits_{i=1}^{N}\alpha_i^2 + \delta^2\alpha_i^2 + \gamma_i^2 + \frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right)\left(\delta^6 V_y + \delta^4\frac{1}{N^2}\sum\limits_{i=1}^{N}\alpha_i^2 + \frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right) - \mathbf{A}^2} \\ &= \frac{\mathbf{A}\gamma_i^2}{\left(\delta^4\sigma^2_\eta + \delta^4\left(\frac{N-1}{N}\right)^2\sum\limits_{i=1}^{N}\alpha_i^2 + \delta^2\alpha_i^2 + \gamma_i^2 + \frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right)\left(\delta^6 V_y + \delta^4\frac{1}{N^2}\sum\limits_{i=1}^{N}\alpha_i^2 + \frac{1}{N^2}\sum\limits_{i=1}^{N}\gamma_i^2\right) - \mathbf{A}^2} \end{aligned}$$

where $\mathbf{A} = \left(\delta^4 \left(\frac{N-1}{N^2} \right) \sum_{i=1}^{N} \alpha_i^2 - \frac{1}{N^2} \sum_{i=1}^{N} \gamma_i^2 \right)$. Dropping terms $1/N$ and smaller:

$$\gamma_2 \simeq 0.$$

When the GGJ tests incorrectly use as the consensus forecasts $\overline{y}_{t|t-2}$, the population value of β_2 is given by (9):

$$\beta_2 = \frac{V(x)\, C(y,z) - C(x,z)\, C(y,x)}{V(x)\, V(z) - (C(x,z))^2}. \tag{10}$$

where y, x and z are:

$$y = y_{t|t-1}^{i} = \delta y_{t-1} + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1}$$

$$x = y_{t|t-2}^{i} = \delta^2 y_{t-2} + \delta \alpha_i \varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2}$$

$$z = \overline{y}_{t|t-2} = \delta^2 y_{t-2} + \delta \frac{1}{N} \sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} + \frac{1}{N} \sum_{i=1}^{N} \gamma_i v_{i,t|t-2}.$$

The required moments are:

$$V(x) = \delta^4 V_y + \delta^2 \alpha_i^2 + \gamma_i^2$$

$$\begin{aligned} C(y,z) &= \delta^3 C_{y_{12}} + \delta^2 \frac{1}{N} C(y_{t-1}, \sum_{i=1} \alpha_i \varepsilon_{i,t-1|t-2}) \\ &= \delta^3 C_{y_{12}} + \delta^2 \frac{1}{N} \sum_{i=1}^{N} \alpha_i^2 \end{aligned}$$

$$C(x,z) = \delta^4 V_y + \frac{1}{N} \delta^2 \alpha_i^2 + \frac{1}{N} \gamma_i^2$$

$$C(y,x) = \delta^3 C_{y_{12}} + \delta^2 \alpha_i^2$$

$$V(z) = \delta^4 V_y + \delta^2 \frac{1}{N^2} \sum_{i=1}^{N} \alpha_i^2 + \frac{1}{N^2} \sum\nolimits_{i=1}^{N} \gamma_i^2.$$

Substitution into (10) gives:

$$\beta_2 = \frac{\left(\delta^4 V_y + \delta^2 \alpha_i^2 + \gamma_i^2\right) \left(\delta^3 C_{y_{12}} + \delta^2 \frac{1}{N} \sum_{i=1}^{N} \alpha_i^2 \right) - \left(\delta^4 V_y + \frac{1}{N} \delta^2 \alpha_i^2 + \frac{1}{N} \gamma_i^2\right) \left(\delta^3 C_{y_{12}} + \delta^2 \alpha_i^2\right)}{\left(\delta^4 V_y + \delta^2 \alpha_i^2 + \gamma_i^2\right) \left(\delta^4 V_y + \delta^2 \frac{1}{N^2} \sum_{i=1}^{N} \alpha_i^2 + \frac{1}{N^2} \sum_{i=1}^{N} \gamma_i^2 \right) - \left(\delta^4 V_y + \frac{1}{N} \delta^2 \alpha_i^2 + \frac{1}{N} \gamma_i^2\right)^2}.$$

Dropping terms which vanish in N:

$$\begin{aligned} \beta_2 &\simeq \frac{\left(\delta^4 V_y + \gamma_i^2\right) \delta^4 V_y - \delta^4 V_y \delta^4 V_y}{\left(\delta^4 V_y + \gamma_i^2\right) \delta^4 V_y - \delta^8 V_y^2} \\ &= 1. \end{aligned}$$

Consider γ_2 in (8), but when the GGJ tests incorrectly use as the consensus forecasts $\overline{y}_{t|t-2}$. Then y, x and z are:

$$\begin{aligned} y &= y^i_{t|t-1} - \overline{y}_{t|t-2} = \delta\left(y_{t-1} - \delta y_{t-2}\right) + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} - \delta \frac{1}{N}\sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} - \frac{1}{N}\sum_{i=1}^{N} \gamma_i v_{i,t|t-2} \\ &= \delta\eta_{t-1} + \delta \sum_{i=1} \alpha_i \varepsilon_{i,t-1|t-2} + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} - \delta \frac{1}{N}\sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} - \frac{1}{N}\sum_{i=1}^{N} \gamma_i v_{i,t|t-2} \end{aligned}$$

$$x = y^i_{t|t-2} - \overline{y}_{t|t-2} = \delta \alpha_i \varepsilon_{i,t-1|t-2} + \gamma_i v_{i,t|t-2} - \delta \frac{1}{N}\sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} - \frac{1}{N}\sum_{i=1}^{N} \gamma_i v_{i,t|t-2}$$

$$z = \overline{y}_{t|t-2} = \delta^2 y_{t-2} + \delta \frac{1}{N}\sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} + \frac{1}{N}\sum_{i=1}^{N} \gamma_i v_{i,t|t-2}$$

with the relevant moments given by:

$$V(x) = \delta^2 \alpha_i^2 + \gamma_i + \delta^2 \frac{1}{N^2}\sum_{i=1}^{N} \alpha_i^2 - 2\delta^2 \frac{1}{N}\alpha_i^2 + \frac{1}{N^2}\sum_{i=1}^{N} \gamma_i^2 - \frac{2}{N}\gamma_i^2$$

$$C(y,z) = \delta^2 \frac{1}{N}\sum_{i=1}^{N} \alpha_i^2 - \delta^2 \frac{1}{N^2}\sum_{i=1}^{N} \alpha_i^2 - \frac{1}{N^2}\sum_{i=1}^{N} \gamma_i^2$$

$$C(x,z) = \delta^2 \frac{1}{N}\alpha_i^2 - \delta^2 \frac{1}{N^2}\sum_{i=1}^{N} \alpha_i^2 + \frac{1}{N}\gamma_i^2 - \frac{1}{N^2}\sum_{i=1}^{N} \gamma_i^2$$

$$C(y,x) = \delta^2 \alpha_i^2 - \delta^2 \frac{1}{N}\alpha_i^2 - \frac{1}{N}\gamma_i^2 - \delta^2 \frac{1}{N}\sum_{i=1}^{N} \alpha_i^2 + \delta^2 \frac{1}{N^2}\sum_{i=1}^{N} \alpha_i^2 + \frac{1}{N^2}\sum\nolimits_{i=1}^{N} \gamma_i^2$$

$$V(z) = \delta^4 V_y + \delta^2 \frac{1}{N^2}\sum_{i=1}^{N} \alpha_i^2 + \frac{1}{N^2}\sum\nolimits_{i=1}^{N} \gamma_i^2.$$

Dropping terms $1/N$ and smaller, we find:

$$\gamma_2 \simeq 0$$

1.3 Regression implementation of BCK test of imitation

$$y_t - y^i_{t|t-1} = \rho_0 + \rho_1 \left(y^i_{t|t-1} - \overline{y}_{t|t-2}\right) + w_t \tag{11}$$

Let:

$$\rho_1 = \frac{C(x,y)}{V(x)}$$

where:

$$y = y_t - y^i_{t|t-1} = \eta_t + \sum_{j \neq i} \alpha_j \varepsilon_{j,t|t-1} - \gamma_i v_{i,t|t-1}$$

and:

$$
\begin{aligned}
x &= y^i_{t|t-1} - \overline{y}_{t|t-2} = \delta\left(y_{t-1} - \delta y_{t-2}\right) + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} - \delta \frac{1}{N}\sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} - \frac{1}{N}\sum_{i=1}^{N} \gamma_i v_{i,t|t-2} \\
&= \delta \eta_{t-1} + \delta \sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} + \alpha_i \varepsilon_{i,t|t-1} + \gamma_i v_{i,t|t-1} - \delta \frac{1}{N}\sum_{i=1}^{N} \alpha_i \varepsilon_{i,t-1|t-2} - \frac{1}{N}\sum_{i=1}^{N} \gamma_i v_{i,t|t-2}.
\end{aligned}
$$

The moments are:

$$C\left(y, x\right) = -\gamma_i^2$$

$$V\left(x\right) = \delta^2 \sigma_\eta^2 + \delta^2 \sum_{i=1}^{N} \alpha_i^2 + \alpha_i^2 + \gamma_i^2 + \delta^2 \frac{1}{N^2}\sum_{i=1}^{N} \alpha_i^2 + \frac{1}{N^2}\sum_{i=1}^{N} \gamma_i^2 - 2\delta^2 \frac{1}{N}\sum_{i=1}^{N} \alpha_i^2$$

and dropping terms of order $\frac{1}{N}$:

$$\rho_2 \simeq \frac{-\gamma_i^2}{\gamma_i^2 + \delta^2 \sigma_\eta^2 + \delta^2 \alpha^2}$$

1.4 Population value of ϕ_1 under noisy information using the $t-2$-dated consensus

Recall ϕ_1 is defined by:

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

We have:

$$
\begin{aligned}
y^i_{t|t} &= \left(1 - PH\right)\delta y^i_{t-1|t-1} + PH y_t + P_v v^i_t + P_\eta \eta_t \\
&= \sum_{k=0}^{\infty} \left(1 - PH\right)^k \delta^k \left[PH y_{t-k} + P_v v^i_{t-k} + P_\eta \eta_{t-k}\right] \\
&= \left[1 - \left(1 - PH\right)\delta L\right]^{-1} \left[\left(1 - \delta L\right)^{-1} PH w_t + P_v v^i_t + P_\eta \eta_t\right] \qquad (12)
\end{aligned}
$$

from which the consensus forecast follows directly as:

$$\overline{y}_{t|t} = \left[1 - \left(1 - PH\right)\delta L\right]^{-1} \left[\left(1 - \delta L\right)^{-1} PH w_t + P_\eta \eta_t\right]$$

because the cross-sectional average of the v^i_t is assumed to be zero.

Hence

$$
\begin{aligned}
y^i_{t|t} - \overline{y}_{t|t} &= \left(1 - PH\right)\delta\left(y^i_{t-1,t-1} - \overline{y}_{t-1|t-1}\right) + P_v v^i_t \\
&= \sum_{k=0}^{\infty} \left(1 - PH\right)^k \delta^k P_v v^i_{t-k} \\
&= \left[1 - \left(1 - PH\right)\delta L\right]^{-1} P_v v^i_t.
\end{aligned}
$$

From $y^i_{t|t-2} = \delta^2 y^i_{t-2|t-2}$, and $\overline{y}_{t|t-2} = \delta^2 \overline{y}_{t-2|t-2}$ we obtain:

$$y^i_{t|t-2} - \overline{y}_{t|t-2} = \delta^2 \left[1 - (1 - PH)\,\delta L\right]^{-1} P_v v^i_{t-2} = \delta^2 \sum_{k=0}^{\infty} (1 - PH)^k \delta^k \left[P_v v^i_{t-2-k}\right]. \tag{13}$$

Consider now the forecast revision $y^i_{t|t-1} - y^i_{t|t-2}$. From (12) we have:

$$y^i_{t|t-1} = \delta y^i_{t-1|t-1} = \delta \left[1 - (1 - PH)\,\delta L\right]^{-1} \left[(1 - \delta L)^{-1} PHw_{t-1} + P_v v^i_{t-1} + P_\eta \eta_{t-1}\right] \tag{14}$$

$$y^i_{t|t-2} = \delta^2 y^i_{t-2|t-2} = \delta^2 \left[1 - (1 - PH)\,\delta L\right]^{-1} \left[(1 - \delta L)^{-1} PHw_{t-2} + P_v v^i_{t-2} + P_\eta \eta_{t-2}\right]. \tag{15}$$

Then $y^i_{t|t-1} - y^i_{t|t-2}$ can be written as:

$$\begin{aligned}
y^i_{t|t-1} - y^i_{t|t-2} &= \delta \sum_{k=0}^{\infty} (1 - PH)^k \delta^k PHw_{t-1-k} \\
&\quad + \delta \left(P_v v^i_{t-1} + P_\eta \eta_{t-1}\right) - \delta^2 PH \sum_{k=0}^{\infty} (1 - PH)^k \delta^k \left(P_v v^i_{t-2-k} + P_\eta \eta_{t-2-k}\right) \\
&= \delta PHw_{t-1} + \delta^2 (1 - PH) \sum_{k=0}^{\infty} (1 - PH)^k \delta^k PHw_{t-2-k} \\
&\quad + \delta \left(P_v v^i_{t-1} + P_\eta \eta_{t-1}\right) \\
&\quad - \delta^2 PHP_\eta \eta_{t-2} - \delta^3 PH (1 - PH) \sum_{k=0}^{\infty} (1 - PH)^k \delta^k \left(P_\eta \eta_{t-3-k}\right) \\
&\quad - \delta^2 PH \sum_{k=0}^{\infty} (1 - PH)^k \delta^k \left(P_v v^i_{t-2-k}\right).
\end{aligned} \tag{16}$$

Because (13) only depends on $\left\{v^i_{t-2}, v^i_{t-3}, \ldots\right\}$, when we calculate the covariance in the numerator of ϕ_1, the only relevant terms in $y^i_{t|t-1} - y^i_{t|t-2}$ are the terms in $\left\{v^i_{t-2}, v^i_{t-3}, \ldots\right\}$: these are given in the bottom line of (16). Consequently:

$$\begin{aligned}
Cov\left(y^i_{t|t-2} - \overline{y}_{t|t-2}, y^i_{t|t-1} - y^i_{t|t-2}\right) &= Cov\left(\delta^2 \sum_{k=0}^{\infty} (1 - PH)^k \delta^k \left[P_v v^i_{t-2-k}\right], -PH\delta^2 \sum_{k=0}^{\infty} (1 - PH)^k \delta^k \left[P_v v^i_{t-2-k}\right]\right) \\
&= -PH\delta^4 \sum_{k=0}^{\infty} (1 - PH)^{2k} \delta^{2k} P_v^2 \sigma_v^2 \\
&= -PH\delta^4 \left[1 - (1 - PH)^2 \delta^2\right]^{-1} P_v^2 \sigma_v^2.
\end{aligned}$$

The variance of $y^i_{t|t-2} - \overline{y_{t|t-2}}$ is:

$$\begin{aligned}
Var\left(y^i_{t|t-2} - \overline{y}_{t|t-2}\right) &= \delta^4 \sum_{k=0}^{\infty} (1 - PH)^{2k} \delta^{2k} P_v^2 \sigma_v^2 \\
&= \delta^4 \left[1 - (1 - PH)^2 \delta^2\right]^{-1} P_v^2 \sigma_v^2,
\end{aligned}$$

from which:

$$\phi = \frac{-PH\delta^4 \left[1-(1-PH)^2\delta^2\right]^{-1} P_v^2\sigma_v^2}{\delta^4 \left[1-(1-PH)^2\delta^2\right]^{-1} P_v^2\sigma_v^2} = -PH.$$

1.5 Population value of ϕ_1 under noisy information using the $t-3$-dated consensus

Consider now the use of the $t-3$-dated consensus:

$$\phi_1 = \frac{Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-3}\right)}{Var\left(y^i_{t|t-2} - \overline{y}_{t|t-3}\right)}.$$

From:

$$y^i_{t-j|t-j} = \sum_{k=0}^{\infty} (1-PH)^k \delta^k \left[PHy_{t-j-k} + P_v v^i_{t-j-k} + P_\eta \eta_{t-j-k}\right]$$

we obtain:

$$\begin{aligned}
y^i_{t|t-2} - \overline{y}_{t|t-3} &= \delta^2 \left(\sum_{k=0}^{\infty} (1-PH)^k \delta^k \left[PHy_{t-2-k} + P_v v^i_{t-2-k} + P_\eta \eta_{t-2-k}\right]\right) \\
&\quad -\delta^3 \left(\sum_{k=0}^{\infty} (1-PH)^k \delta^k \left[PHy_{t-3-k} + P_\eta \eta_{t-3-k}\right]\right) \\
&= \delta^2 \left(\sum_{k=0}^{\infty} (1-PH)^k \delta^k \left[PHw_{t-2-k} + P_v v^i_{t-2-k}\right]\right) \\
&\quad +\delta^2 P_\eta \eta_{t-2} - \delta^3 PH \left(\sum_{k=0}^{\infty} (1-PH)^k \delta^k P_\eta \eta_{t-3-k}\right) \qquad (17)
\end{aligned}$$

and thus:

$$Var\left(y^i_{t|t-2} - \overline{y}_{t|t-3}\right) = \delta^4 \left[1-(1-PH)^2\delta^2\right]^{-1} \left(P_v^2\sigma_v^2 + (PH)^2\sigma_w^2 + \delta^2 P_\eta^2\sigma_\eta^2\right) + \delta^4 P_\eta^2\sigma_\eta^2. \qquad (18)$$

Using expressions (17) and (16) we obtain:

$$\begin{aligned}
& Cov\left(y^i_{t|t-1} - y^i_{t|t-2}, y^i_{t|t-2} - \overline{y}_{t|t-3}\right) \\
= \;& \delta^4 PH \left[1-(1-PH)^2\delta^2\right]^{-1} \left((1-PH)PH\sigma_w^2 - P_v^2\sigma_v^2 + \delta^2 PH(1-PH)P_\eta^2\sigma_\eta^2\right) - \delta^4 PHP_\eta^2\sigma_\eta^2 \qquad (19)
\end{aligned}$$

and (19) and (18) together give:

$$\begin{aligned}
\phi_1 &= \frac{\delta^4 PH\left[1-(1-PH)^2\delta^2\right]^{-1}\left((1-PH)PH\sigma_w^2 - P_v^2\sigma_v^2 + \delta^2 PH(1-PH)P_\eta^2\sigma_\eta^2\right) - \delta^4 PHP_\eta^2\sigma_\eta^2}{\delta^4\left[1-(1-PH)^2\delta^2\right]^{-1}\left(P_v^2\sigma_v^2 + (PH)^2\sigma_w^2 + \delta^2 P_\eta^2\sigma_\eta^2\right) + \delta^4 P_\eta^2\sigma_\eta^2} \\
&= \frac{PH\left[1-(1-PH)^2\delta^2\right]^{-1}\left((1-PH)PH\sigma_w^2 - P_v^2\sigma_v^2 + \delta^2 PH(1-PH)P_\eta^2\sigma_\eta^2\right) - PHP_\eta^2\sigma_\eta^2}{\left[1-(1-PH)^2\delta^2\right]^{-1}\left(P_v^2\sigma_v^2 + (PH)^2\sigma_w^2 + \delta^2 P_\eta^2\sigma_\eta^2\right) + P_\eta^2\sigma_\eta^2}.
\end{aligned}$$