ICMA Centre

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Do Survey Joiners and Leavers Differ from Regular Participants? The US SPF GDP Growth and Inflation Forecasts

January 2020

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

Discussion Paper Number:
ICM-2020-01

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

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

Do Survey Joiners and Leavers Differ from Regular Participants? The US SPF GDP growth and Inflation Forecasts.

Michael P. Clements
ICMA Centre,
Henley Business School,
University of Reading,
Reading RG6 6BA
m.p.clements@reading.ac.uk.

January 23, 2020

Abstract

If learning-by-doing is important for macro-forecasting, newcomers might be different to regular, established particants. Stayers may also differ from the soon-to-leave. We test these conjectures for macro-forecasters point predictions of output growth and inflation, and for their histogram forecasts. A bootstrap approach is used to overcome the problems associated with the relatively small numbers of joiners and leavers. Controlling for the numbers of forecasters with the bootstrap approach is required to correctly determine whether there are systematic differences between experienced forecasters and newcomers, and between stayers and leavers.

Journal of Economic Literature classification: C53.

Keywords: Forecast accuracy, experience, learning-by-doing, probability forecasts.

1 Introduction

In this paper we consider whether newcomers or inexperienced forecasters are systematically different from experienced forecasters. There are a number of reasons to suspect the two groups may differ, with a presumption that newcomers may not be as skilled. We may also suspect that forecasters who leave the survey are less able forecasters. These questions have received little attention in the macro-forecasting literature, but bear on a number of issues.

One is whether 'learning by doing' is an important driver of the performance of professional forecasters, and whether it might account for differences between forecasters. There has been much interest in the recent literature regarding the sources of disagreement between forecasters [1]. Recent explanations stress the role of informational rigidities (e.g., Coibion and Gorodnichenko (2012, 2015)), as well as different reporting practices, e.g., Patton and Timmermann (2007), Engelberg, Manski and Williams (2009) and Clements (2009, 2010). The changing composition of the panel of forecasters is seldom explicitly addressed.

Another reason for considering whether there are significant differences between forecasters is the widespread use of aggregates of individual respondents' expectations for a variety of purposes. Aggregates of surveys of macro-forecasters are often used with little attention paid to compositional effects. The aggregate is often referred to as the consensus forecast, even though the 'consensus' may be based on very different views. Aggregates of survey expectations are used in comparisons of model and survey forecasts (e.g., Ang, Bekaert and Wei (2007) or Clements and Galvão (2017)), as sources of expectations shocks in structural VAR analysis of macroeconomic fluctuations (e.g., Leduc and Sill (2013), Clements and Galvão (2019)), and to test theories of expectations formation which make predictions about aggregate expectations (as in the approach to testing for informational rigidities in Coibion and Gorodnichenko (2012, 2015)).

Engelberg, Manski and Williams (2011) advise caution when using aggregate measures of expectations, highlighting the problems involved in interpreting changes in the consensus, especially when the composition of the panel is changing over time. Even with a fixed panel, changes in the consensus may not reflect near-unanimous changes in beliefs about future trends. It may be that all respondents revise their forecasts in the direction indicated by the change in the aggregate, and even change their forecasts by the same magnitude. On the other hand, there might be substantial

[1] See, for example, Zarnowitz and Lambros (1987), Bomberger (1996), Rich and Butler (1998), Capistrán and Timmermann (2009), Lahiri and Sheng (2008), Rich and Tracy (2010) and Patton and Timmermann (2010), *inter alia*.

disagreement even about the direction of change, let alone the magnitude. When the composition of the panel is changing, Engelberg *et al.* (2011) argue that the interpretation of temporal variation in the consensus is potentially even more problematic. Changes in the aggregate may simply reflect joiners having different views than leavers, or may reflect the specific subset of the 'active' participants who happen to respond to the surveys in question.[2] As noted by Engelberg *et al.* (2011, p.1061), changing panel composition in surveys of forecasts could be ignored 'if it were credible to assume that panel members are randomly recruited from a stable population of potential forecasters and that participation in the survey after recruitment is statistically independent of forecasters beliefs about inflation'. These assumptions support missing data being treated as 'missing at random', and allow us to ignore the changing composition of the panel of forecasters. They note however that there is no evidence to justify these assumptions.

Without knowing the processes for recruiting new forecasters to the panel, or the reasons why some participants leave, the key determinant of the *practical* importance of compositional effects on aggregate estimates is whether newcomers (and leavers) are substantively different from incumbents. Engelberg *et al.* (2011, p.1061) suggest the use of fixed composition sub-panels to eliminate compositional effects. But rather than using the subset of forecasters who responded in each of two adjacent periods, they suggest using the union of the forecasters and calculating a bound on the temporal variation in the aggregate from imputing values for the missing forecasts. They suppose the missing values should not generate changes between the two periods which are more extreme than the observed changes (from those who responded to both periods). Although this is a reasonable approach, it doesn't differentiate between (say) first-time respondents, and non-participation in the previous period by an otherwise regular respondent.[3] Moreover, the process used to fill in missing values might be too conservative if we were to find systematic differences between those who responded (in both periods) and those who did not.

The key difficulty in determining whether there are systematic differences between newcomers and experienced forecasters is the small number of newcomers, which is zero for some surveys. The same is true of course when we compare stayers and leavers - there are typically few leavers, if

[2] From a practical perspective, the long-term entry and exit of survey respondents, and the occasional non-response by active participants, typically results in only relatively short unbroken series of expectations by any individual, whereas consensus forecasts constitute long unbroken series that sustain the sorts of analyses described above.

[3] Just as we do not know the reasons for joining and leaving the survey, there appear to be few studies looking at partipication decisions by active forecasters. López-Pérez (2016) suggests an individual's decision to respond to a particular survey may be influenced by the perception of (aggregate) uncertainty. Response rates to the ECB's SPF appear to be lower when uncertainty is higher, consistent with the 'production-cost hypothesis' - the costs of producing a forecast may be greater when the outlook is more uncertain.

any. This situation can be ameliorated to some extent by considering 'relatively' in-experienced forecasters, rather than first-time forecasters, but this attenuates any differences between the new and the incumbents, especially if people learn quickly. We show that ignoring the problem of small numbers can be misleading, falsely suggesting the in-experienced are less skillful. We devise a bootstrap to account for the small numbers of joiners and leavers.

We also investigate whether the complexity of the forecasting task affects the relative advantage that experienced forecasters might enjoy over newcomers. We do this by considering not just point predictions, but also the probability distributions reported in the form of histograms. We use the US Survey of Professional Forecasters (SPF) as our source of macro survey data, because it provides both types of forecast.

The plan of the remainder of the paper is as follows. Section 2 explains why newcomers and experienced (and leavers and stayers) might differ. Section 3 describes the forecast data, and how we define the groups. The groups change over time, and are defined for each survey t. Section 4 considers whether the groups of experienced forecasters and newcomers differ, and similarly for the groups of stayers and leavers, but without explicitly allowing for the small numbers of newcomers and leavers. In section 5 we report the results of a bootstrap approach which corrects for the small numbers of newcomers and leavers. Section 6 offers some concluding remarks.

2 Experienced forecasters and newcomers, and stayers and leavers

Are there good reasons to suppose that newcomers would differ from experienced forecasters? Newcomers have either made no survey returns prior to their participation at time t, or only a small number of responses, if we consider relative newcomers. Experienced forecasters have made a number of responses to earlier surveys. There are several reasons why the two groups may differ. More experienced forecasters might have 'learnt from doing': they might better understand the task and the form of the required survey return; they may be better at processing information; and be less likely to make transcription errors, etc. This might be more likely to be the case when the task at hand is unusual, in the sense that it is not a regular part of the newcomers' activities prior to their signing up. An example might be the generation of the histograms of future inflation and output growth outcomes, which we consider along with the point predictions.[4]

[4] Clements (2019a) finds more evidence that forecasters are not all alike in terms of forecast accuracy when he considers their histogram forecasts, as opposed to their point predictions. But contrary to this paper, his focus is not on comparing joiners (and leavers) to established forecasters.

Recent theories of information rigidities suggest that not all agents are equally attentive to data releases, or update their information sets at each point in time (see, e.g., the noisy and sticky information models reviewed by Coibion and Gorodnichenko (2012, 2015)), and it might be that established survey participants are more attentive to data releases and economic news more generally than agents who until recently have not participated.

Malmendier and Nagel (2016) argue that agents 'learn from experience', such that the inflation expectations of consumers are affected by their life-time experiences of inflation. Individuals who have lived through periods of high inflation typically expect higher inflation, and the expectations of the young are more affected by the recent data than the expectations of older people, as they have shorter lifetime experiences to draw on. Professional forecasters' analyses of current conditions and future prospects might also depend on their past experiences, including their time spent as active survey respondents: similar in this regard to the consumers of Malmendier and Nagel (2016).

In short, there are reasons to suppose newcomers may differ from established forecasters. If so, changes in panel composition due to newcomers replacing more experienced forecasters may affect aggregate summaries, as warned by Engelberg *et al.* (2011). Given our forecast data set we are unable to separate out reasons for any apparent differences - for example, we do not know the age of the respondents, so any effects of age and experience are confounded.[5] Our interest is in whether there are differences in the forecasts of newcomers and experienced forecasters.

What about those who leave the survey? Less accurate forecasters might be expected to cease to work as forecasters, and might leave the survey too, although we know that in some circumstances accuracy is not the only prized attribute in the forecasting industry: see e.g., Ottaviani and Sorensen (2006b, 2006a) and Marinovic, Ottaviani and Sorensen (2013). Coupled with the anonymity of the respondents, it is perhaps simplistic to assume that bad forecasters will necessarily be driven out.[6] Hence we regard as an empirical question whether those soon-to-leave perform worse than stayers.

[5] That is, from the duration of an individual's involvement as a participant, we are unable to determine whether forecast characteristics arise due to on-the-job experience or age.

[6] Bad forecasters might be driven out even if they are anonymous within the SPF if they report essentially the same poor forecasts on other platforms where they are identified.

3 The US SPF data, and Newcomers and Leavers

3.1 Macro-survey data

We use the US Survey of Professional Forecasters (SPF). The SPF is a quarterly survey of macroeconomic forecasters of the US economy that began in 1968, administered by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER). Since June 1990 it has been run by the Philadelphia Fed, renamed as the Survey of Professional Forecasters (SPF): see Croushore (1993). We evaluate the forecasts made over the period 1991:1 to 2014:4. We do not use the very earliest data from the survey because the SPF documentation mentions the suspicion that more recent forecasters may have been given the identifiers once associated with participants to the early surveys who have since left.

We use the US SPF because it provides probability assessments of output growth (real GDP) and GDP deflator inflation, as well as the point predictions for these two variables, over a long historical period.

There are differences in the nature of the point predictions and histogram forecasts. The histogram forecasts are of the annual rates of growth (of GDP, and the GDP deflator) for the year of the survey relative to the previous year (as well as of the next year, relative to the current year). That is, they are 'fixed-event' histogram forecasts. This provides an annual series of (approximately) year-ahead forecasts, if we consider the Q1 surveys, or an annual series of one-quarter ahead forecasts if we consider the Q4 surveys, say. The available series of fixed-horizon forecasts are therefore necessarily rather short. Instead we consider the quarterly series of histogram forecasts, cognizant of the fact that the horizon is changing from one forecast to the next. For the point predictions we have quarterly fixed-horizon series of the current quarter, up to the same quarter a year ahead, and we consider these two series - the shortest and longest horizon quarterly series of forecasts provided by the SPF.

D'Agostino, McQuinn and Whelan (2012) conclude that apparent differences between forecasters in terms of the accuracy of their point predictions may be illusory. However, Clements (2019a) found more evidence that forecasters differed in terms of the accuracy of their probability assessments. This is perhaps unsurprising given that histogram forecasts are more costly to produce, especially if forecasters only produce them for the Survey, and they are discussed much less in the business and news media: a relative novice may not be able to draw on a prevailing view about the probability assessment. The complexity of the task may allow some forecasters to shine, and may magnify the benefits of experience.

3.2 Assessment of the histogram forecasts

Whilst the point predictions are evaluated by squared-error loss, we evaluate the SPF probability distributions using the ranked probability score (RPS: Epstein (1969)). As shown below, RPS is closely related to the quadratic probability score (QPS: Brier (1950)), which is a natural alternative to the log score. QPS is defined by:

$$QPS = \sum_{k=1}^{K} \left(p^k - y^k \right)^2 \tag{1}$$

and RPS by:

$$RPS = \sum_{k=1}^{K} \left(P^k - Y^k \right)^2 \tag{2}$$

for a single histogram with K bins (indexed by the superscript k), where p^k is the probability assigned to bin k. y^k is an indicator variable equal to 1 when the actual value is in bin k, and zero otherwise. In the definition of RPS, P^k is the cumulative probability (i.e., $P^k = \sum_{s=1}^{k} p^s$), and similarly Y^k cumulates y^s. Note that if $y^{s_1} = 1$, then $Y^k = 1$ for all $k \geq s_1$.[7]

RPS and QPS are natural scoring rules to use when, as here, the probability assessments are provided as histograms. Although the log score is perhaps the most popular scoring rule for densities (see, e.g., Winkler (1967)), RPS has the advantage of being calculable directly from the histograms, that is, without making any additional assumptions.[8]

Being based on cumulative distributions, RPS will penalize less severely forecasts with probability close to the bin containing the actual, relative to QPS. For QPS, a given probability outside the bin in which the actual falls has the same cost regardless of how near or far it is from the outcome-bin. For this reason, the RPS seems preferable to QPS.

[7] The SPF definitions of the bin locations are formally given as, e.g., 4 to 5.9, and 6 to 7.9, and so on. Usually the bins are interpreted as $[4, 6]$ and $[6, 8]$, but if the realization is, say, 5.98, interpreting this as falling in the lower bin ($[4, 6]$) might be misleading. In such circumstances we instead assume that $y_t^k = \frac{1}{2}$ for these two bins, and the RPS Y_t^k values for the two bins are $\frac{1}{2}, 1$, respectively.

[8] The calculation of QPS and RPS only requires knowledge of the probabilities assigned to each bin, $\{p^k\}$, provided explicitly by the survey respondents, and a stance on what constitutes the actual value (and therefore y^k). Unlike fitting parametric distributions, no difficulties arise when probability mass is assigned to only one or two bins. For example, to calculate the log score we would need to make an assumption about the distribution of the probability mass within each of the histogram intervals, which might be done by first fitting a density to the histogram, see, e.g., the review by Clements (2019b).

3.3 Defining the groups

One could simply define newcomers (or joiners, J) as those who have not previously responded to the survey (either since its inception, or who have not responded to a given number of earlier surveys). However, for the US SPF this would yield a small number of newcomers $N_{J,t}$ for many t, making comparisons to the experienced forecasters (E) unreliable, or even infeasible for t when $N_{J,t} = 0$. Instead we might define J as forecasters who are relatively inexperienced - those who have made no more than n_r previous responses - and compare these to the relatively experienced - those who have made n_e or more responses. This would be sensible if those who have made a small number of forecasts are still more similar to the first-timers than the experienced respondents.

Our approach means we are comparing the two groups' forecasts over the same periods. This is key to obtaining meaningful comparisons. Comparisons across different target periods are likely to be misleading when economic conditions are very different. An obvious example is that forecasts of outcomes during 2008-10 would be much less accurate than forecasts of outcomes in e.g., 2005-2007.

One might argue that there will be more new entrants when forecasting is easier, or that respondents are more likely to quit during difficult times, when making a forecast requires a greater cost. Then the newcomers will be reporting forecast at times which are more conducive to accurate forecasting. And leavers might tend to incur larger forecast errors in the difficult periods prior to their leaving the survey. But the newcomers will be compared with the forecasts made by experienced forecasters for precisely the same periods. The same is true of exiters - they will not be disadvantaged relative to stayers by forecasting at difficult times.

To illustrate, at a given date t, say 1991:Q1, we identify all the individual respondents who submitted a forecast. We then categorize these forecasters into the two groups, depending on their past participation: $\leq n_r$ and $\geq n_e$. We might then calculate averages (the cross-sectional medians) for each of the two groups (but see below). We then move to the next quarter, and repeat the exercise. As we move through the sample, individuals will move from newcomers J to experienced E (provided they do not quit the survey). Our definitions of the two groups mean they will be in limbo before they are classified as experienced, whenever $n_e > n_r + 1$. Hence the constituents of J_t (and E_t) change over time, and we re-define the groups, E and J anew at each forecast origin (or survey). Hence the size of each group is changing over time, and at times the group of relatively inexperienced forecasters J has few members, especially when it is defined as $n_r = 0$, so that is consists of first-time respondents. When J is the null set the corresponding survey is dropped from the analysis. There is a clear trade-off - a low value of n_r (and a high value of n_e) increases the separation between the two groups in terms of their average degree of experience, but reduces the size of both groups.

To define quitters Q and stayers S at time t, we proceed as follows. If a respondent at time t makes n_s or more responses to the next two years of surveys, the individual is a stayer, S. No more than n_q and the individual is a leaver. Then setting $n_s = 7$ defines the set S as consisting of those who respond to 7 of the next 8 surveys, and $n_q = 0$ defines Q as those who do not respond to any of the next 8 surveys. As for experienced/inexperienced there is a trade-off. Setting $n_q = 2$, say, captures 'infrequent' forecasters over the next 2 years.

We illustrate the assignment of forecasters of groups using the actual survey data recorded in table 1. Consider the response pattern of the first forecaster (id. 65). Recall that 1991:1 is the first period for which evaluate the forecasts. The forecaster records a forecast to the survey in 1991:1, and is classified as an experienced forecaster having already responded on 8 previous occasions. (The table truncates the past record to begin in 1987:1, and shows only the top 12 individuals in terms of number of forecasts made). The forecaster remains an experienced forecaster throughout, up to 2007:3, albeit with non-participation in 2002:3 and 2004:1. As of (and including) 2005:4 onwards, this forecaster is no longer a stayer, making fewer than 7 responses to the next 8 surveys. If we set $n_q = 2$, the forecaster becomes a leaver in 2006:4 (and 2007:2 and 2007:3). If $n_q = 0$, the leaver status is only conferred on the forecaster in 2007:3.

The forecaster with id. 446 is a first-time respondent in 1993:2 (assuming no forecasts have been made in the period prior to that shown in the table). This forecaster is classed as a joiner in that period only, when $n_r = 0$, but is also a joiner in 1993:4 and 1993:4 under the more lax definition of bring 'relatively inexperienced', given by $n_r = 2$. Either way, the forecaster is first classified as experienced when they respond to the 1995:3 survey (having made 8 forecasts at this point).

Across the 96 surveys from 1991:Q1 to 2014:Q4, the average number of respondents was 39.3, the average number of experienced forecasters was 28.7, and the average number of stayers was 22.5.[9] On average there were 1.4 joiners, and the same number of leavers, when $n_r = 0$ and $n_q = 0$, increasing to around 4.3 for both categories when we adopt the looser definitions; $n_r = 2$, $n_q = 2$. Figure 1 shows these averages hide much variation over time. For example, there was an

[9] These are for the current-quarter output growth point forecasts, and will differ a little for other horizons and variables.

increase in the number of joiners at the beginning of the period displayed in the figure following the Philadelphia Fed taking charge of the survey, as well as in 1995:2, 1999:2 and 2005:2. .

4 Are Newcomers and Leavers different?

We test: (i) whether the joiners are more/less accurate than the experienced forecasters, and (ii) whether the quitters are more/less accurate than the stayers, using a Diebold and Mariano (1995) (DM) test of equal predictive ability of the medians of each of the groups.

Our comparisons of forecast accuracy are based on the responses to the near quarter-century of surveys from 1991:Q1 to 2014:4. The choice of start date means we consider only those surveys administered by the Philadelphia Fed, although the responses to earlier surveys are used in the classification of each respondent. For example, we use responses to the previous 11 years of surveys to categorize current respondents as belonging to E or J. The end-date is selected to allow subsequent surveys to determine stayer/leaver status.

Our calculations of forecast loss are based on a vintage of the actual values released soon after the reference quarter,[10] rather than the latest-available vintage at the time of the investigation. This seems preferable to using 'fully-revised' data which will typically include benchmark revisions, rebasings, and other methodological changes to the way the data are collected and measured, as well as regular annual revisions. Fortunately the Real Time Data Set for Macroeconomists (RTDSM) maintained by the Federal Reserve Bank of Philadelphia, see Croushore and Stark (2001), greatly facilitates the use of real-time data.

There are a number of ways of calculating the loss for period t for a group, G. We can calculate the average (we use the median) of the individual forecasts f_{it}, $i \in G$, say $\overline{f}_{t,G}$, and then calculate the loss of this median forecast, which we denote $S_{t,G,mf}$. Or we can calculate the median of the losses of the individual forecasts, which we denote $S_{t,G,ms}$. We know that the median forecast will have a smaller loss than that of an individual forecast drawn at random from the set of forecasts,[11] but it seems unlikely that either approach will adequately account for the small numbers of forecasters in J or Q. In this section we calculate DM tests of equal predictive ability for both approaches for the point predictions. That is, for comparing newcomer and experienced forecaster point predictions, for example, we test the sequences of loss differentials $d_{t,mf} = S_{t,E,mf} - S_{t,J,mf}$,

[10] In fact we use the first estimates (known as the 'advance' estimates).
[11] Manski (2011) reviews results on the 'Algebra of Consensus Forecasts'.

and $d_{t,ms} = S_{t,E,ms} - S_{t,J,ms}$, for a squared error loss function. We find that the results of the two approaches are similar. In the following section we consider whether the results of the DM tests are affected by the small numbers of joiners and leavers.

For the histogram forecasts we report DM test results for the average (median) of the individual histogram RPS scores.

The results are reported in table 3. We give the p-values of the null of equal forecast accuracy of the two groups in the comparison. For example, for 'Experienced v. Newcomers', the loss differential is the squared forecast error for the median of E minus the squared forecast error for the median of J, for column (2), or the median loss differential of E minus the median loss of J, for column (3). Consequently, a p-value less than 0.025 (greater than 0.975) indicates a rejection of the null of equal forecast accuracy in favour of group E (in favour of group J), in a two-sided test at the 5% level. The table is divided into 4 quadrants. The top half refers to current-quarter forecasts. That is, forecasts of the quarter in which the forecasts are elicited. The bottom panel refers to 4-quarter or year-ahead forecasts. These are forecasts of the same quarter of the year as the survey quarter but in the following year. The left panels gives a strict delineation between J and E, and between Q and S. J consists of only first-time forecasters, and Q those who do not issue another forecast in the next two years. The right panel allows members of J to have responded to up to 2 surveys (in the previous 11 years), and members of Q to make up to 2 more responses in the subsequent two years. As well as the p-values, the table records the number of time periods on which the DM test is calculated. The maximum is 96, the number of surveys between 1991:Q1 to 2014:4. A lower number indicates that for some surveys J or Q were empty sets.

Our findings for the current quarter forecasts suggest that experienced forecasters are more accurate than newcomers, for both output growth and inflation. And that this holds for the strict definition of J (left panel) and also when we adopt the more lax definition of J (top right quadrant). The evidence for Q being worse than S is weaker, except for inflation using the strict definition of leavers.

For the year-ahead forecasts (bottom two quadrants) for the most part there is no evidence that E and S are superior to J and Q (respectively), except for inflation, when inexperienced forecasters again appear to be less accurate. That there is less evidence of significant differences between the groups for the longer-horizon forecasts is perhaps not surprising. Forecasting a quarterly growth rate a year ahead is difficult, and our results suggest that being a seasoned survey respondent confers less of an advantage over an inexperienced forecaster.

The results are generally qualitatively the same whether we use the score of the median forecast, or the median of the individual scores (compare columns (2) and (3), and (5) and (6)).

The results for the histogram forecasts are shown in table 4 for the median RPS losses of the groups. The clear message is that inexperienced inflation forecasters are not as good as experienced, but experience does not appear to matter for output growth. This holds whether we consider first-time forecasters, or the relatively inexperienced. These findings for inflation match the findings for the point forecasters.[12] However the results for the histograms also suggest that leavers are worse than stayers (columns (2)) which was not true of the point forecasts. Rather than delving more deeply into possible differences between the results for the two types of forecast, in the following section we consider the reliability of these findings.

5 Bootstrapping the distribution of the test of equal accuracy

A potential problem with the approach in section 4 is that the group medians are calculated over different numbers of forecasters at each t, and that the numbers in J and Q may be low, especially when the groups are defined in a strict fashion. Taking the median of the larger number of forecasts in E or S, relative to J or Q, might of itself result in superior apparent accuracy of the members of E and S over J and Q.

One way of controlling for this is to bootstrap the distribution of the test of equal forecast accuracy under the null that there are no systematic differences between the individual forecasters in the sets J or Q and E or S. We do this ensuring that the small numbers of members of J and Q observed in the actual survey data are preserved in the simulated data. That is, taking the medians of different group sizes is also a feature of the simulated data, and so cannot account for differences between the actual test outcomes and the simulated distributions of the test statistics. Consequently, any differences between the actual and simulated test statistics must reflect inter-group homogeneity not holding in the actual data.

We mirror the unbalanced nature of the SPF panel by simulating a set of imaginary forecasters with the same response patterns as the actual respondents, as in D'Agostino *et al.* (2012, p. 718).

Let f_{it} denote the forecast for individual i in response to survey t, where we abstract from the forecast horizon. For each survey t, the actual forecasts f_{jt}, where $j = 1, \ldots, N_t$ (and j indexes the

[12] Note that the results for the histogram forecasts average across forecast horizons up to one-year ahead. For the point forecasts there was evidence that experience matters at both the current-quarter and year-ahead horizons.

N_t non-missings), are randomly assigned (with replacement) across a set of 'imaginary' forecasters who match the SPF forecasters exactly in terms of participation. For example, if the n^{th} SPF forecaster did not participate at time t, the n^{th} imaginary forecaster's time t forecast will also be missing. If the n^{th} SPF forecaster did participate at time, the matching imaginary forecaster is equally as likely to receive any one of the actual forecasts made at time t. That is, the allocation does not depend on whether the n^{th} forecaster is a newcomer or an established forecaster. Hence by construction there are no systematic differences between the imaginary forecasters in the J and E (or S and Q) groups. We continue for each t, for $t = 1, \ldots, T$. As stressed by D'Agostino *et al.* (2012), forecasters at each t can only be assigned a forecast from another forecaster made at that time.

For each group at each t, we then calculate the loss differential between the loss of the median forecaster for that group (or the difference between the median losses). We then calculate the test of equal accuracy over all valid t, exactly as in table 3. This provides DM^1. We repeat the above another 999 times to obtain the bootstrap distribution $\{DM^1, \ldots, DM^{1000}\}$. We reject the null - that there is no difference in accuracy between the members of the J and E groups - if the actual DM statistic takes an extreme value relative to the bootstrap distribution.

In tables 5 and 6 we record the proportion of the bootstrap replications which yield a DM statistic below the statistic obtained on the actual data. Consider the point predictions. There is no evidence that leavers are worse than stayers, at odds with the results in table 3. For example, consider the test statistic that gave rise to a value of 0.019 in table 3 for comparing Stayers and Levers for current-quarter inflation forecasts. The value of 0.322 in table 5 indicates that nearly a third of the simulated values provided as much evidence in favour of stayers being more accurate. We conclude that the apparent support for stayers over leavers is consistent with no difference in accuracy between the two groups once the small number of leavers is accounted for. The same is true of all the comparisons between stayers and leavers, and the experienced and joiners, for both variables, at the longer, year-ahead horizon. The only significant findings are for the current quarter, and are that experienced forecasters are better than first-time inflation forecasters (but not the relatively inexperienced), and experienced forecasters of GDP growth are better than the relatively-inexperienced (but not the first-timers).

The bootstrap results in table 6 suggest little evidence that there is a significant difference between the groups for GDP growth histograms, but experienced inflation forecasters appear to be better than the relatively inexperienced inflation forecasters.

6 Conclusions

Despite the interest in macroeconomic survey expectations in recent years, and the diverse uses to which aggregate survey quantities have been put, little is known about the wisdom of including inexperienced forecasters (relative newcomers) alongside experienced forecasters. One way of tackling this issue is to ask whether there are systematic differences between inexperienced and experienced forecasters in terms of the accuracy of their forecasts. We argue that attempts to address this question run into problems associated with small numbers of newcomers, or of relatively in experienced forecasters, more generally. We show that the evidence favouring stayers over the soon-to-leave largely disappears when the small numbers of leavers is controlled for. There is little evidence that those who leave the survey do so because they are worst forecasters than the remainers. There is some evidence that there are real differences between newcomers/inexperienced and experienced forecasters, in terms of both their current-quarter point forecasts and histogram forecasts of inflation. Controlling for the small numbers of leavers and joiners is shown to greatly attenuate the apparent advantages that incumbents enjoy over members of both these groups.

References

Ang, A., Bekaert, G., and Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better?. *Journal of Monetary Economics*, **54(4)**, 1163–1212.

Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking*, **28**, 381–392.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **75**, 1–3.

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

Clements, M. P. (2009). Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry. Chapter 8*, pp. 206–226. Oxford: Oxford University Press.

Clements, M. P. (2010). Explanations of the Inconsistencies in Survey Respondents Forecasts. *European Economic Review*, **54(4)**, 536–549.

Clements, M. P., and Galvão, A. B. (2017). Model and survey estimates of the term structure of US macroeconomic uncertainty. *International Journal of Forecasting*, **33**, 591 – 604. DOI: 10.1016/j.ijforecast.2017.01.004.

Clements, M. P. (2019a). Are some forecasters probability assessments of macro variables better than those of others?. Mimeo, ICMA Centre, University of Reading.

Clements, M. P. (2019b). *Macroeconomic Survey Expectations*: Palgrave Texts in Econometrics. Palgrave Macmillan. DOI 10.1007/978-3-319-97223-7.

Clements, M. P., and Galvão, A. B. (2019). Measuring the effects of expectations shocks. Mimeo, ICMA, Henley Business School, Reading.

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

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

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

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

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

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, **13**, 253–263.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*, **27(1)**, 30–41.

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

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, **8**, 985–987.

Lahiri, K., and Sheng, X. (2008). Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics*, **144(2)**, 325–340.

Leduc, S., and Sill, K. (2013). Expectations and Economic Fluctuations: An Analysis Using Survey Data. *The Review of Economics and Statistics*, *95*(4), 1352–1367.

López-Pérez, V. (2016). Does uncertainty affect participation in the European Central Bank's Survey of Professional Forecasters?. mimeo, Universidad Politécnica de Cartagena.

Malmendier, U., and Nagel, S. (2016). Learning from inflation experiences. *Quarterly Journal of Economics*, **131(1)**, 53–87.

Manski, C. F. (2011). Interpreting and combining heterogeneous survey forecasts. In Clements, M. P., and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting, Chapter 16*, pp. 457–472. Oxford: Oxford University Press.

Marinovic, I., Ottaviani, M., and Sorensen, P. N. (2013). Forecasters' Objectives and Strategies, chapter 2. In Elliott, G., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 2A*, pp. 690–720: North Holland.

Ottaviani, M., and Sorensen, P. N. (2006a). Reputational cheap talk. *RAND Journal of Economics*, **37(1)**, 155–175.

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

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

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

Rich, R., and Tracy, J. (2010). The relationships among expected inflation, disagreement, and uncertainty: Evidence from matched point and density forecasts. *Review of Economics and Statistics*, **92(1)**, 200–207.

Rich, R. W., and Butler, J. S. (1998). Disagreement as a measure of uncertainty: A comment on Bomberger. *Journal of Money, Credit and Banking*, **30**, 411–419.

Winkler, R. L. (1967). The quantification of judgement: Some methodological suggestions. *Journal of the American Statistical Associaton*, **62**, 1105–1120.

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

Table 1: Illustration of Forecast Data: Top 12 respondents 1987:1 to 2014:4, Current quarter GDP growth

.	65	84	40	426	421	428	433	20	446	411	407	420
1987.1	0.57	.	0.68	.	.	.	.	0.49	.	.	.	.
1987.2	0.59	0.56	0.35	.	.	.	.	0.67	.	.	.	.
1987.3	0.71	0.74	.	.	.	.	.	.	.	.	.	.
1987.4	.	0.89	0.42	.	.	.	.	0.75	.	.	.	.
1988.1	-0.85	.	0.03	.	.	.	.	1.03	.	.	.	.
1988.2	.	0.66	0.64	.	.	.	.	0.77	.	.	.	.
1988.3	.	0.85	.	.	.	.	.	2.15	.	.	.	.
1988.4	.	0.62	0.42	.	.	.	.	.	.	.	.	.
1989.1	.	1.31	1.02	.	.	.	.	.	.	.	.	.
1989.2	.	0.59	0.37	.	.	.	.	0.78	.	.	.	.
1989.3	.	0.58	.	.	.	.	.	0.44	.	.	.	.
1989.4	-0.05	0.36	0.41	.	.	.	.	.	.	.	.	.
1990.1	.	0.48	0.34	.	.	.	.	0.98	.	.	.	.
1990.2	0.25	0.57	0.60	.	.	.	.	0.84	.	.	.	.
1990.3	-0.12	0.55	.	.	.	.	.	0.93	.	.	0.59	.
1990.4	-0.57	-0.74	-0.46	.	-0.36	.	.	0.25	.	-0.60	-0.30	-0.36
1991.1	-0.39	-0.63	-0.48	-0.48	-0.44	-0.41	-0.34	.	.	-0.35	-0.11	.
1991.2	-0.12	-0.10	-0.19	0.00	-0.15	0.19	-0.05	0.05	.	-0.27	-0.06	-0.02
1991.3	0.53	0.85	.	0.43	0.65	0.56	0.70	.	.	.	0.73	.
1991.4	0.63	0.36	0.24	0.18	0.51	0.33	0.57	0.36	.	0.27	0.36	0.37
1992.1	0.15	0.24	0.06	-0.05	0.18	0.24	0.23	0.39	.	0.11	0.12	1.76
1992.2	0.67	0.47	0.72	0.45	0.94	0.85	0.82	.	.	0.72	0.54	.
1992.3	0.44	0.46	.	0.37	0.50	0.51	0.52	0.64	.	0.50	0.61	0.58
1992.4	0.66	0.51	0.48	0.10	0.09	0.34	0.30	1.36	.	0.31	0.18	0.57
1993.1	0.64	.	0.76	0.52	0.72	0.63	0.77	1.53	.	0.79	0.67	0.53
1993.2	0.69	0.50	.	0.47	0.59	.	0.55	0.66	0.62	0.67	0.43	.
1993.3	0.39	0.83	.	0.52	0.99	0.59	0.72	0.53	.	0.79	0.63	0.67
1993.4	0.59	0.95	.	0.79	0.93	.	0.84	1.09	0.72	0.84	0.89	0.85
1994.1	0.65	0.92	0.78	0.55	0.74	0.75	0.86	1.48	0.82	0.86	0.77	0.72
1994.2	0.43	1.14	0.71	0.84	0.91	0.76	0.96	2.57	0.91	0.98	0.78	0.87
1994.3	0.30	0.56	0.54	0.67	0.37	0.58	0.52	.	0.69	.	0.40	0.82
1994.4	0.50	0.93	0.68	0.69	0.86	0.51	0.74	1.68	0.79	0.78	0.47	0.74
1995.1	0.74	0.76	0.58	0.69	0.45	0.61	0.84	2.03	0.75	0.74	0.67	0.83
1995.2	0.60	0.33	0.55	0.50	0.47	0.42	0.55	1.76	0.42	0.45	0.30	0.52
1995.3	0.72	0.35	0.43	0.42	0.64	0.51	0.55	0.99	0.43	0.46	0.52	0.38
1995.4	0.62	0.69	0.63	0.42	0.62	0.47	0.57	1.98	0.66	0.63	0.59	0.55
1996.1	0.18	0.59	0.47	0.77	0.30	0.33	0.44	0.86	0.29	.	.	0.32
1996.2	0.47	0.89	0.65	0.67	1.01	0.47	0.84	.	0.84	.	0.70	0.58
1996.3	0.58	0.80	0.78	0.45	0.62	0.51	0.64	0.80	0.59	0.64	0.37	.
1996.4	0.50	0.49	.	0.20	0.42	0.43	0.42	.	0.52	0.57	0.50	0.68
1997.1	0.62	0.76	0.54	0.27	0.37	0.46	0.52	0.87	0.55	0.46	0.42	0.60
1997.2	0.75	0.59	0.52	0.62	0.35	0.57	0.69	0.88	0.69	0.41	.	0.60
1997.3	0.62	0.59	.	0.69	0.89	0.61	0.67	.	0.62	0.52	0.42	0.59
1997.4	0.62	0.66	0.59	0.77	0.86	0.85	0.55	0.53	.	0.66	.	0.63
1998.1	0.50	0.63	0.47	0.57	.	0.71	0.57	0.89	0.50	0.47	0.54	0.65
1998.2	0.67	0.87	0.59	.	0.55	0.54	0.64	.	0.55	0.55	0.62	0.74
1998.3	0.50	.	.	0.52	0.59	0.47	0.72	1.24	0.47	0.53	0.53	0.43
1998.4	0.25	0.87	0.61	0.59	0.64	0.46	0.62	1.03	0.60	0.63	0.63	0.83
1999.1	0.74	0.66	0.93	0.55	0.84	0.63	0.79	.	0.69	.	0.78	.
1999.2	0.62	0.95	1.06	0.59	0.84	0.76	0.94	0.93	0.74	.	0.82	0.57
1999.3	0.77	0.86	0.63	0.86	.	0.85	.	0.74	0.55	0.70	0.75	0.85
1999.4	1.23	0.69	0.92	.	1.00	1.06	0.81	1.04	0.99	.	0.74	0.80
2000.1	0.50	0.94	0.78	0.45	0.40	0.98	0.79	1.17	.	0.82	0.56	0.90
2000.2	0.74	1.31	.	1.03	0.75	1.15	1.06	0.73	1.03	.	1.02	0.98
2000.3	0.86	0.73	1.05	0.84	1.08	0.83	0.94	1.29	0.74	.	0.90	0.98
2000.4	0.50	0.70	0.81	1.06	1.00	0.62	0.86	1.07	0.89	0.74	0.61	0.72
2001.1	0.37	0.25	0.47	.	-0.01	0.10	0.12	0.61	0.25	0.19	.	.
2001.2	0.30	0.41	0.54	.	0.49	0.33	0.12	0.15	0.37	0.53	0.36	.
2001.3	0.42	0.39	0.48	0.07	0.65	0.48	0.42	0.28	.	0.47	-0.20	0.30
2001.4	-0.38	-0.14	-0.10	-0.76	.	-0.46	-0.48	0.12	-0.63	-0.75	-0.22	-0.70
2002.1	0.12	1.01	0.17	0.15	1.08	0.33	0.57	0.13	0.45	0.59	0.15	.
2002.2	0.12	1.10	0.81	0.22	0.87	0.57	0.84	0.36	0.74	.	0.51	0.70
2002.3	.	0.84	0.18	0.40	1.00	0.62	0.55	0.47	0.62	0.62	0.65	0.58
2002.4	0.25	0.28	1.30	0.15	0.61	0.36	0.27	0.44	0.45	.	0.34	.

Table 2: Continuation of Table 1

.	65	84	40	426	421	428	433	20	446	411	407	420
2003.1	0.45	.	0.66	0.00	0.90	0.59	0.67	0.04	0.62	0.47	0.48	0.63
2003.2	0.45	0.45	0.45	0.37	0.70	0.43	0.45	0.44	0.42	.	0.70	0.60
2003.3	0.86	1.04	0.97	0.99	1.12	0.80	1.01	0.46	0.86	1.04	0.87	0.98
2003.4	0.99	1.02	1.10	0.86	1.05	0.87	1.06	0.62	.	.	.	1.00
2004.1	.	.	1.07	1.15	0.96	1.06	1.18	1.20	1.11	1.15	.	.
2004.2	1.11	0.97	.	0.84	0.98	1.25	1.11	0.84	1.11	1.06	.	1.20
2004.3	0.86	0.80	0.86	.	0.89	.	0.84	0.68	0.99	.	.	0.95
2004.4	0.62	1.04	.	0.72	0.89	1.06	0.94	0.71	0.84	0.79	.	1.00
2005.1	0.74	.	.	0.96	0.84	0.82	1.01	.	0.89	1.01	.	0.90
2005.2	0.62	0.81	.	.	0.73	0.76	0.69	0.45	0.74	0.73	.	.
2005.3	0.67	1.14	.	0.88	0.70	1.05	1.06	0.53	1.03	1.13	0.74	0.95
2005.4	0.69	0.61	.	0.72	0.65	0.72	0.79	0.95	0.84	0.86	.	0.75
2006.1	0.67	.	.	1.11	0.80	1.15	1.20	0.97	1.01	1.12	1.18	0.95
2006.2	0.64	0.87	.	0.74	0.53	0.90	0.84	1.05	0.89	0.81	.	.
2006.3	0.50	0.70	.	0.81	.	0.70	0.55	1.02	0.67	.	0.79	0.66
2006.4	0.74	0.49	.	0.69	0.60	0.56	.	0.60	0.62	0.61	0.62	0.52
2007.1	.	0.51	.	0.71	0.42	0.58	0.64	0.68	0.64	0.57	0.67	0.63
2007.2	0.59	0.43	.	0.69	0.55	.	.	0.59	0.59	0.62	0.55	0.53
2007.3	0.64	0.51	.	0.64	.	0.66	0.59	0.86	0.69	0.59	.	0.48
2007.4	.	0.36	.	.	0.40	0.41	.	1.12	0.37	0.31	0.16	0.37
2008.1	.	.	.	-0.15	0.22	.	.	.	0.07	0.18	0.12	-0.05
2008.2	.	0.16	.	-0.05	0.25	-0.17	.	0.15	-0.18	0.56	0.25	.
2008.3	.	0.11	.	-0.28	0.78	0.49	.	0.17	0.25	0.28	0.50	0.28
2008.4	.	-0.78	.	-1.02	-0.85	-0.55	-0.86	.	-0.76	-0.95	-0.63	-0.54
2009.1	.	.	.	-1.12	-0.80	-1.27	.	.	-1.33	-1.32	.	-1.18
2009.2	.	-0.47	.	-0.48	0.50	-0.61	-0.48	-0.35	-0.30	-0.14	-0.13	-0.54
2009.3	.	0.43	.	0.09	.	.	0.32	.	0.37	0.73	0.86	0.63
2009.4	.	0.53	.	0.19	0.87	0.78	.	.	0.72	0.84	0.50	.
2010.1	.	.	.	0.51	0.65	0.65	.	.	0.69	0.81	0.69	0.65
2010.2	.	.	.	0.38	0.48	0.71	.	.	0.89	0.90	0.74	.
2010.3	.	.	.	0.31	0.37	0.42	.	.	0.62	0.63	0.74	0.75
2010.4	.	.	.	0.31	0.60	0.52	.	.	0.50	0.42	0.62	0.60
2011.1	.	.	.	0.93	.	1.06	.	.	0.86	0.98	0.50	0.95
2011.2	.	.	.	0.73	.	.	.	.	0.81	0.86	0.74	0.92
2011.3	.	.	.	0.04	.	.	0.57	.	0.62	.	0.72	0.40
2011.4	.	.	.	0.49	0.89	0.68	0.64	.	0.62	0.73	0.67	0.60
2012.1	.	.	.	0.53	0.86	0.72	0.59	.	0.47	.	0.59	0.60
2012.2	.	.	.	0.47	0.81	0.56	0.67	.	0.57	.	0.57	0.50
2012.3	.	.	.	0.30	0.64	0.39	0.40	.	0.52	0.41	0.57	0.40
2012.4	.	.	.	0.44	.	0.45	0.55	.	0.50	0.33	0.30	.
2013.1	.	.	.	0.93	0.55	0.77	0.68	.	0.50	0.52	0.59	0.53
2013.2	.	.	.	0.33	0.72	0.44	0.43	.	0.48	0.42	0.45	0.60
2013.3	.	.	.	0.46	0.59	0.64	0.69	.	0.42	0.58	.	0.65
2013.4	.	.	.	0.48	0.47	0.33	0.42	.	0.46	.	0.57	0.57
2014.1	.	.	.	0.50	0.52	0.62	0.56	.	0.55	0.47	0.42	0.45
2014.2	.	.	.	0.82	1.10	1.35	0.91	.	0.64	.	0.84	0.93
2014.3	.	.	.	0.77	0.77	0.72	0.73	.	0.86	0.74	.	0.82
2014.4	.	.	.	0.66	1.10	0.77	0.60	.	0.74	0.54	.	.

Table 3: Tests of Equal Predictive Accuracy of Point Predictions of Different Groups, Surveys 1991:Q1 to 2014:Q4

(1)	(2) mf	(3) ms	(4) No. Surveys	(5) mf	(6) ms	(7) No. Surveys
Current-quarter forecasts						
	Experienced v. Newcomers			Experienced v. Relative Newcomers		
GDP	0.011	0.014	61	0.000	0.000	86
INF	0.000	0.001	57	0.029	0.009	87
	Stayers v. Leavers			Stayers v. Soon-to-Leave		
GDP	0.044	0.069	74	0.056	0.028	94
INF	0.019	0.016	71	0.120	0.028	95
Year-ahead quarter forecasts						
	Experienced v. Newcomers			Experienced v. Relative Newcomers		
GDP	0.240	0.189	60	0.113	0.098	87
INF	0.019	0.017	57	0.094	0.024	88
	Stayers v. Leavers			Stayers v. Soon-to-Leave		
GDP	0.073	0.045	77	0.891	0.863	93
INF	0.064	0.028	74	0.218	0.113	94

The left panel is when Newcomers (J) and Leavers (Q) are the sets of forecasters who respectively have no responses before t and none after t. Column (4) is the number of surveys (or forecast observations) when the groups are so defined. The right panel is for J and Q defined as no more than 2 responses prior to t, and no more than 2 responses after t, respectively. And column (7) gives the number of surveys for the tests of equal accuracy.

The results in columns (2) to (3), and (5) and (6), are for the comparisons of the squared errors of the medians of the forecasts (for each group) and for the medians of squared forecast errors (of each group). The elements in the table in these columns are the p-values of the tests of equal accuracy of the Experienced versus the Newcomers, and the Stayers versus the Leavers. A p-value < 0.025 (> 0.975) indicates rejection of the null of equal accuracy (Diebold-Mariano test) in favour of the first group (second group) being more accurate at the 5% level in a two-sided test. (The actual values are the advance estimates).

Table 4: Tests of Equal Predictive Accuracy of Histograms for year-on-year growth rates of Different Groups, Surveys 1991:Q1 to 2014:Q4

(1)	(2)	(3)	(4)	(5)
	Experienced v. Newcomers		Experienced v. Relative Newcomers	
GDP	0.126	58	0.182	83
INF	0.001	56	0.000	81
	Stayers v. Leavers		Stayers v. Soon-to-Leave	
GDP	0.001	66	0.055	82
INF	0.002	66	0.077	82

The left panel is when Newcomers (J) and Leavers (Q) are the sets of forecasters who respectively have no responses before t and none after t. Column (3) is the number of surveys (or forecast observations) when the groups are so defined. The right panel is for J and Q defined as no more than 2 responses prior to t, and no more than 2 responses after t, respectively. And column (5) gives the number of surveys for the tests of equal accuracy.
The elements in the table in columns (2) and (4) are the p-values of the tests of equal accuracy of the Experienced versus the Newcomers, and the Stayers versus the Leavers, where accuracy is assessed by RPS. A p-value < 0.025 (> 0.975) indicates rejection of the null of equal accuracy (Diebold-Mariano test) in favour of the first group (second group) being more accurate at the 5% level in a two-sided test. (The actual values are the advance estimates).

Table 5: Tests of Equal Predictive Accuracy (squared error loss) of Point Predictions of Different Groups using the Bootstrap

(1)	(2) mf	(3) ms	(5) mf	(6) ms
	Current-quarter forecasts			
	Experienced v. Newcomers		Experienced v. Relative Newcomers	
GDP	0.143	0.168	0.003	0.003
INF	0.014	0.029	0.409	0.246
	Stayers v. Leavers		Stayers v. Soon-to-Leave	
GDP	0.451	0.652	0.323	0.263
INF	0.322	0.322	0.631	0.397

	Year-ahead quarter forecasts			
	Experienced v. Newcomers		Experienced v. Relative Newcomers	
GDP	0.600	0.572	0.385	0.407
INF	0.245	0.230	0.683	0.432
	Stayers v. Leavers		Stayers v. Soon-to-Leave	
GDP	0.334	0.282	0.970	0.973
INF	0.694	0.545	0.737	0.658

The left panel is when Newcomers (J) and Leavers (Q) are the sets of forecasters who respectively have no responses before t and none after t. Column (4) is the number of surveys (or forecast observations) when the groups are so defined. The right panel is for J and Q defined as no more than 2 responses prior to t, and no more than 2 responses after t, respectively. And column (7) gives the number of surveys for the tests of equal accuracy.
The results in columns (2) to (3), and (5) and (6), are for the comparisons of the median of the forecasts (for each group) and for the medians of squared forecast errors (of each group). The elements in the table in these columns are the proportion of the bootstrap replications which yielded a test statistic value less than the actual test statistic. The tests are of the Experienced versus the (Relative) Newcomers, and the Stayers versus the (Soon-to-be) Leavers.

Table 6: Tests of Equal Predictive Accuracy of Histograms (RPS) for year-on-year growth rates of Different Groups using the Bootstrap

(1)	(2)	(3)
	Experienced v. Newcomers	Experienced v. Relative Newcomers
GDP	0.685	0.790
INF	0.079	0.018
	Stayers v. Leavers	Stayers v. Soon-to-Leave
GDP	0.047	0.323
INF	0.154	0.598

The left panel is when Newcomers (J) and Leavers (Q) are the sets of forecasters who respectively have no responses before t and none after t. Column (3) is the number of surveys (or forecast observations) when the groups are so defined. The right panel is for J and Q defined as no more than 2 responses prior to t, and no more than 2 responses after t, respectively. And column (5) gives the number of surveys for the tests of equal accuracy.

The results in columns (2) to (3), and (5) and (6), are for the comparisons of the median of the forecasts (for each group) and for the medians of squared forecast errors (of each group). The elements in the table in these columns are the proportion of the bootstrap replications which yielded a test statistic value less than the actual test statistic. The tests are of the Experienced versus the (Relative) Newcomers, and the Stayers versus the (Soon-to-be) Leavers.

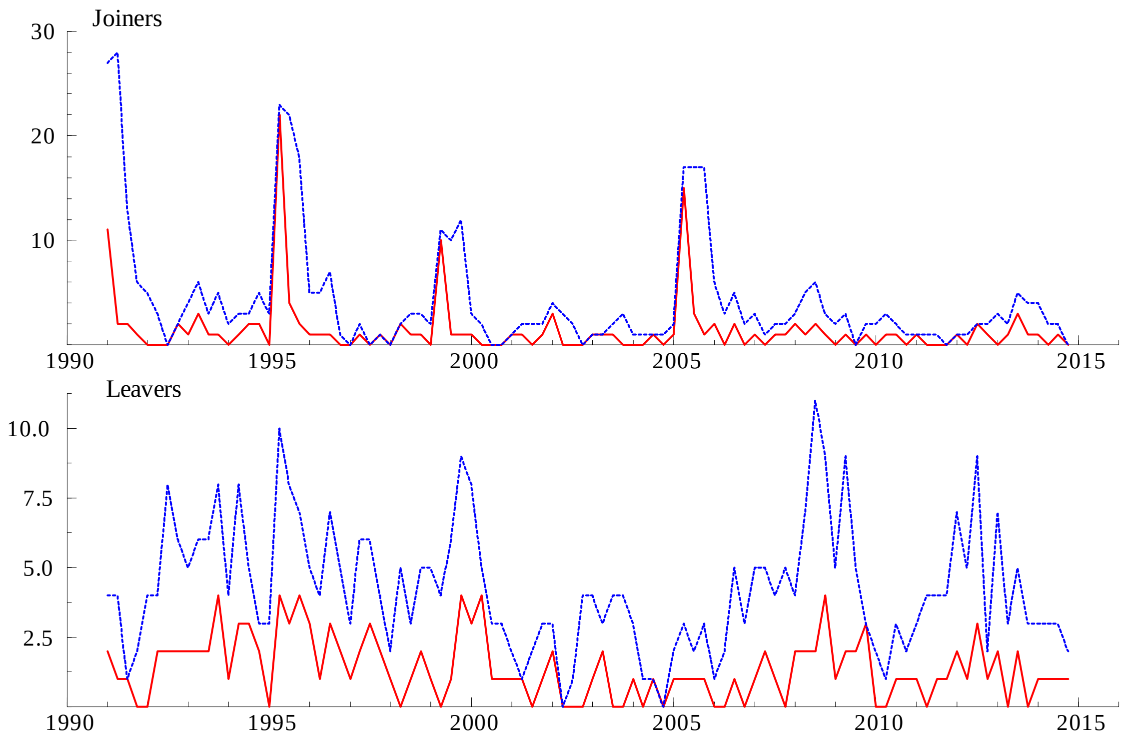

Figure 1: At each survey, 1991:1 - 2014:4, the number of active participants designated as joiners and leavers. The solid line gives the number of first-time forecasters and those making no more subsequent forecasts, in the top and bottom panels, respectively ($n_r = 0$, $n_q = 0$). The dotted lines correspond to $n_r = 2$ (top panel) and $n_q = 2$ (bottom panel).