Do survey joiners and leavers differ from regular participants?

Article

Accepted Version

Clements, M. P. ORCID: https://orcid.org/0000-0001-6329-1341 (2021) Do survey joiners and leavers differ from regular participants? International Journal of Forecasting, 37 (2). pp. 634-646. ISSN 0169-2070 doi: 10.1016/j.ijforecast.2020.08.003 Available at https://centaur.reading.ac.uk/92753/

To link to this article DOI: http://dx.doi.org/10.1016/j.ijforecast.2020.08.003

Publisher: Elsevier

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.

August 14, 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. Histogram forecasts of inflation by both joiners and leavers are found to be less accurate, especially if we suppose that joiners take time to learn. For GDP growth, there is no evidence for differences between the groups in terms of histogram forecast accuracy, although leavers GDP point forecasts are less accurate. These findings are predicated on forecasters being homogeneous within groups. Allowing for individual fixed effects suggests fewer differences, including leavers' inflation histogram forecasts being no less accurate.

Journal of Economic Literature classification: C53.

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

*Helpful comments from two referees of this journal are gratefully acknowledged.

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.[1] There has been much interest in the recent literature regarding the sources of disagreement between forecasters.[2] 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

[1]This is conceptually distinct from the literature on panel conditioning effects, which considers how having responded to previous round(s) of the survey might affect the current forecast. See, e.g., the application to consumers' inflation expectations by Binder (2019).

[2]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*.

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 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.[3] 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.[4] 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.

We investigate whether the complexity of the forecasting task affects the relative advantage

[3] 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.

[4] 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.

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.

One of the difficulties 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 any. This situation can be ameliorated by considering 'relatively' in-experienced forecasters, rather than just first-time forecasters, but this will attenuate any differences between the new and the incumbents when people learn quickly. We devise a bootstrap to account for the small numbers of joiners and leavers, but generally our results are broadly the same whether we use standard inference or the bootstrap. Our findings can be summarized as follows. Histogram forecasts of inflation by both joiners and leavers are found to be less accurate, especially if we suppose that joiners take time to learn (that is, learning is not instantaneous). For GDP growth, however, there are few instances of differences between the groups in terms of histogram forecast accuracy, although leavers GDP point forecasts are less accurate than those of remainers. However, these findings are predicated on forecasters being homogeneous within groups. If we allow individual-specific fixed effects (in addition to time fixed effects), the statistical evidence for differences between the groups is broadly unchanged, except that the inflation histogram forecasts of leavers are no longer clearly worse than those of remainers.

The plan of the remainder of the paper is as follows. Section 2 considers why joiners and leavers might differ from the experienced remainers. Section 3 describes the forecast data, and how we score the accuracy of the two types of forecast. Section 4 sets out the econometric approach which underpins all the empirical results. Section 5 presents the results. 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 made either no survey returns prior to their participation at time t, or only a small number of responses, if we consider relative newcomers. 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 producing histograms of future inflation and output growth outcomes, which we consider along with the point predictions.

Histogram forecasts are discussed much less in the business and news media: there is less likely to be a prevailing view that the novice is able to draw on. The complexity of the task may allow some forecasters to shine, and may magnify the benefits of experience.

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. 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]

[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

Hence we regard as an empirical question whether those soon-to-leave perform worse than stayers.

3 The US SPF data, and the evaluation of forecast accuracy

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). 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 when we consider 'learning' the responses to earlier surveys are used. 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. 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,[7] 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

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

same poor forecasts on other platforms where they are identified. Alternatively, the direction of causation might be reversed - forecasters planning to leave may expend less effort producing forecasts.

[7] In fact we use the first estimates (known as the 'advance' estimates). We also experimented with using the vintage available two quarters after the reference quarter, and found the results were qualitatively unchanged.

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. In some of the analyses we consider whether the results depend on the forecast horizon. 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.

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)), defined by:

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

for a histogram with K bins (indexed by the superscript k). Suppose p^k is the probability assigned to bin k, and 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 $k \geq s_1$.[8] RPS is a natural scoring rule to use when, as here, the probability assessments are provided as histograms. It is calculable directly from the histograms, that is, without making any additional assumptions.[9] Being based on cumulative distributions, RPS will penalize less severely forecasts with probability close to the bin containing the actual, relative to the commonly-used quadratic probability score, 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. See Boero, Smith and Wallis (2011) for a

[8] 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.

[9] The calculation of 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). No difficulties arise when probability mass is assigned to only one or two bins. Were we to calculate the log score, for example, we would need to make an assumption about the distribution of the probability mass within each of the histogram intervals, which might be achieved by fitting a density to the histogram.

discussion of the RPS, and its comparison to other scoring rules.

In section 5 we will make use of the variances of the histograms. When probabilities are assigned to three or more histogram intervals, we calculate the variances directly assuming the probability is located at the midpoints of the intervals. When probability is assigned to only one or two intervals, we calculate the variances after fitting triangular (isosceles) distributions, as described by Engelberg *et al.* (2009, pp. 37-8). (See Clements (2019, ch. 3) for a discussion of calculating moments from histograms).

4 Econometric Approach

We use a panel approach to determine whether there are differences between the different groups of forecasters, such as between joiners and experienced forecasters, for example. In our base case, newcomers (or joiners, J) are 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). As described below, we can allow that the transition from first-timer to experienced forecaster takes place over a number of periods, as the respondent learns. We define leavers (or quitters) at time t as those who do respond to any of the next 8 surveys. We might wish to allow that a leaver's forecasts might differ not just in time t, but in the period leading up to leaving, so in periods $t-n, \ldots, t-1$ too (for n a small positive integer).

If s_{it} is the score (either the squared forecast error for the point forecasts, or the RPS value for the histogram forecasts), the generic regression is:

$$s_{it} = \alpha_t + \beta d_{it} + u_{it} \tag{2}$$

where $d_{it} = 1$ if individual i is a joiner (say) at time t, and is zero for all other periods t for individual i. The coefficient β measures the average effect on the score of being a joiner. As indicated by the time-subscripted α_t in (2), we allow for time fixed effects to account for the changes in economic conditions which impact on the difficulty of forecasting (e.g., during recessions as opposed to during less volatile periods). As is evident from Figure 1, which shows the number of joiners and leavers at each point in time, joiners in particular arrive in waves. For example, there was an 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. Across the 96 surveys from 1991:Q1 to 2014:Q4, the average number of respondents was 39.3.[10] On average there were 1.4 joiners, and the same number of leavers each period. Time effects capture the possibility that economic conditions might happen to be easier (harder) than average for forecasting across these episodes, which would otherwise make joiners appear to be better (worse) than experienced forecasters.

Of course we do not observe newcomers (defined as first-time respondents) before they are 'newcomers', that is, if $d_{it_1} = 1$, we do not observe i for $t < t_1$. And similarly for leavers: we only observe leavers for the periods up to their departure.

When we include time-fixed effects, we can analyze the effect of there only being a small number of joiners by considering a simplified, stylized setting. Suppose there are N respondents in each of the T periods, and in each period there is one joiner, and one forecaster leaves. We model this as $d_{it} = 1$ if $i = t + N - 1$, and $d_{it} = 0$ if $i \neq t + N - 1$. That is, at time t, the joiner is respondent $i = t + N - 1$. Straightforward calculations give the fixed-effects estimator of (2) with time effects, under our assumptions on entry and exit, as:

$$\hat{\beta} = \frac{\sum_{t=1}^{T} \sum_{i=t}^{N+t-1} \left(s_{it} - \overline{s}_t\right) \left(d_{it} - \overline{d}_t\right)}{\sum_{t=1}^{T} \sum_{i=t}^{N+t-1} \left(d_{it} - \overline{d}_t\right)^2}. \tag{3}$$

Because the cross-section averages $\overline{d}_t = \frac{1}{N}$, then $d_{it} - \overline{d}_t = \frac{N-1}{N}$ when $i = t + N - 1$, and $d_{it} - \overline{d}_t = \frac{-1}{N}$ otherwise, (3) can be shown to simplify to:

$$\begin{aligned} \hat{\beta} &= \frac{1}{T} \sum_{t=1}^{T} \left(s_{N+t-1,t} - \frac{1}{(N-1)} \sum_{i=t}^{N+t-2} s_{it} \right) \\ &= \frac{1}{T} \sum_{t=1}^{T} s_{N+t-1,t} - \frac{1}{T(N-1)} \sum_{t=1}^{T} \sum_{i=t}^{N+t-2} s_{it} \end{aligned} \tag{4}$$

That is, $\hat{\beta}$ is the difference between the average scores of the first-timers, and the average across all t and i of the experienced forecasters. (Or equivalently, as the average over t of the difference between the joiner score and cross-section mean of the experienced forecasters scores at each t: first line of (4)). When the number of joiners is small (e.g., one each period in our example), the accuracy of the estimation of their effect clearly depends on the number of surveys, T. In our

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

empirical work $T = 96$ (the surveys from 1991:1 to 2014:4, inclusive). We find that a bootstrap approach to determining whether the estimated effect β is statistically significant gives similar results to standard inference.

4.1 A bootstrap

We use a bootstrap to estimate a reference distribution for $\hat{\beta}$ under the assumption of no systematic difference between the accuracy (i.e., the s-values) of the members of the groups being compared.[11] Hence we can assess whether the empirical estimate $\hat{\beta}$ is consistent with the assumption of equal predictive ability across the members of the groups. We ensure that the small numbers of members of J (say) in the actual survey data are a feature of the simulated data, by having exactly the same numbers of respondents in each group at each point in time as in the actual data. That is, we generate artificial forecasters who have the same response/non-response patterns as the actual forecasters. Each respondent at time t is randomly allocated with equal probability (and with replacement) one of the time t scores, whether the recipient is a joiner or experienced forecaster. The pool of time t scores from which the draws are made consists of the scores from joiners and experienced respondents. We construct a bootstrap sample for t, for $t = 1, \ldots, T$, where $T = 96$. On the bootstrap sample, we estimate (2) with time fixed-effects, and record $\hat{\beta}_1^*$. We repeat another 999 times to obtain the bootstrap distribution $\left\{\hat{\beta}_1^*, \ldots, \hat{\beta}_{1000}^*\right\}$. We reject the null - that there is no difference in accuracy between the members of the groups - when $\hat{\beta}$ takes an extreme value relative to the bootstrap distribution. We report the proportion of the bootstrap distribution that exceeds $\hat{\beta}$:

$$P_{\beta^*} = \frac{1}{1000} \sum_{r=1}^{1000} 1\left(\hat{\beta} < \hat{\beta}_r^*\right) \tag{5}$$

where $1(A) = 1$ when A is true, and zero otherwise. Hence in a one-sided test that joiners are less accurate than experienced forecasters, for example, we would reject at the 5% level if $P_{\beta^*} < 0.05$, and would reject using a two-sided test at the 5% level if either $P_{\beta^*} < 0.025$ or $P_{\beta^*} > 0.975$. We use P_{β^*} as given in (5) because it is reasonable to suppose that joiners might be less accurate than experienced forecasters, so that H_0: $\beta = 0$ versus H_1: $\beta > 0$.

Inference using the simulated reference distribution will not be affected by small group sizes at some (or all) t. If $\hat{\beta}$ is 'extreme' relative to the simulated distribution this can only be because

[11] This extends the approach of D'Agostino, McQuinn and Whelan (2012, p. 718), who considered whether some forecasters are better than others.

of inter-group differences in the actual data. A possible caveat with the bootstrap is if individual forecasters exhibit persistent optimism or pessimism, as documented by Batchelor (2007). We consider this possibility below.

As described below, the panel regression in (2) will be adapted to address particular issues. Other approaches could be used. For example, we could calculate the average (either the median or mean) of the scores of the forecasters belonging to a particular group, and we could compare these average scores. For example, for a group G (which might be first-time respondents), we could calculate $\overline{s}_{t,G} = \frac{1}{n_{t,G}} \sum_{i \in G} s_{it}$, where $n_{t,G}$ is the number of forecasters in G at time t, and taking the average to be the mean. Then we could compare joiners J to experienced forecasters E, say, by defining the sequence of loss differentials $d_{t,E-J} = \overline{s}_{t,E} - \overline{s}_{t,J}$, and applying the Diebold and Mariano (1995) test of equal predictive accuracy, or a related test (see, e.g., Giacomini and Rossi (2009, 2010)). Under the Diebold-Mariano null of equal accuracy, $E\left(d_{t,E-J}\right) = 0$. Within this approach issues to do with the small number of members of some groups (such as J) are again pertinent, and may necessitate bootstrapping. For example, for each bootstrap sample we would calculate the Diebold-Mariano test statistic, to give the distribution $\{DM_1^*, \ldots, DM_{1000}^*\}$, against which the value $\widehat{DM}$ calculated for the reported forecast data would be compared. We use the panel regression model as a unified approach to exploring the ways in which joiners and experienced (or leavers and stayers) might differ.

4.2 Alternative modelling assumptions

There is strong case for allowing time fixed effects. The case for individual is not settled: D'Agostino *et al.* (2012) conclude that apparent differences between forecasters in terms of the accuracy of their point predictions may be illusory, although Clements (2020) finds that forecasters are not all alike in terms of histogram forecast accuracy, and Batchelor (2007) reports persistent differences in point forecasts in terms of optimism/pessimism. Our base set of results does not include individual fixed effects. These results also assume the models' disturbances are serially uncorrelated within panels, which may be suspect for multi-step forecasts, and we later relax this to allow (identical) first-order autocorrelation in each panel. Allowing for cross-sectional correlations in the disturbances is complicated when there are few time periods in common between some time periods and in some instances will be infeasible, and we do not attempt this.

Other approaches may also be fruitful, for example, we could decompose forecast errors into common and idiosyncratic shocks, and add a 'joiner error', extending the approach of Lahiri and

Sheng (2010).[12] We leave this for future research. We regard allowing for individual fixed effects and autocorrelated errors as the obvious extensions to the panel model with time fixed effects.

5 Empirical results

5.1 Base results - time fixed effects

5.1.1 Are newcomers and leavers different in terms of forecast accuracy?

We consider whether newcomers and leavers differ from experienced remainers along a number of dimensions. Firstly, we consider whether newcomers are less good than experienced respondents, where we define 'less good' in terms of expected squared error loss for the point forecasts, and in terms of RPS for the histogram forecasts. And relatedly, whether respondents who leave are less good than remainers.

For the point forecasts, we have short and medium horizon forecasts, so we can investigate whether the relative accuracy of the forecasts depends on the horizon. The availability of the different types of forecasts allows us to address the impact of the complexity of the forecasting task on the relative disadvantage of being a newcomer. That is, do newcomers fare relatively worse when the task is to provide a histogram forecast compared to a point forecast?

Table 1 shows the results for the point forecasts, and table 2 the results for the histogram forecasts. Consider the point forecasts. Joiners are defined as first-time respondents, and leavers are defined by $d_{it} = 1$ when t is the last period for which respondent i makes a forecast. There is no evidence that joiners are less accurate for either GDP growth or inflation, for short or medium term forecasts, and whether we use inference based on conventionally calculated t-statistics, or bootstrap inference based on $\hat{\beta}$. However, leavers are less accurate on average at forecasting GDP, but not inflation. This is true for the current-quarter forecasts for conventional and bootstrap inference. For the 4-quarter forecasts we would only reject $\beta = 0$ against a one-sided alternative ($\beta > 0$) at the 10% using conventional inference, but we do so at the 5% level using the bootstrap. In this instance, conventional inference partially masks the significance of the effect. At the short horizon, the average squared error is nearly half as large again as the base line level.[13]

[12] I am grateful to a referee for suggesting this possibility.

[13] Because we use time fixed effects, we calculate the baseline level of forecast loss is the average of the coefficients on the time dummies. In the instance we are referrring to, this is given as 0.169 in the table. The increase due to leaving is apporximately half this at 0.088, the value of $\hat{\beta}$.

When we consider the histogram forecasts we adapt (2) to allow the coefficient β on the dummy variable d_{it} to depend on the quarter of the year in which the survey is held:

$$s_{it} = \alpha_t + \sum_{q=1}^{4} \beta_q \left(d_{it} \times 1_{qt} \right) + u_{it} \tag{6}$$

where $1_{qt} = 1$ when $t \in q$, and zero otherwise, and q indexes the quarters of the year. Whereas the point forecasts are fixed-horizon, the histogram forecasts are fixed-event (see, e.g., Nordhaus (1987)). The forecast target is the same for the four quarters of the year - it is the annual growth rate for the current year relative to the previous year (of either GDP or its deflator), and so the forecast horizon shortens throughout the year. Moreover recent research suggests that the properties of the survey respondents histograms may depend on the survey quarter. Clements (2014) shows that the mismatch between the histogram variances and *ex post* forecast uncertainty varies with the forecast horizon. To allow for the possibility that the effect of joining (or leaving) might also depend on the forecast horizon (equivalently, the quarter of the year of the survey response), for the histogram forecasts we replace (2) by (6).

Table 2 shows no differences between joiners and extant respondents for output growth, mirroring the finding for the point forecasts. However, leavers' Q3 GDP histograms are worse. This is apparent from standard testing and the bootstrap: the one-sided bootstrap p-value is 0.019. For inflation there is some evidence that joiners are worse at the longer two horizons. Leavers whose last inflation forecasts are made in response to anything but Q1 are less accurate on RPS, howsoever inference is conducted. The conclusions of inference based on conventional testing and bootstrapping $\hat{\beta}$ generally point in the same direction, and differences in p-values are generally small. For the remainder of our paper we only report the standard tests of $\beta = 0$.

5.1.2 Do newcomers learn-by-doing?

The formulation in equations (2) and (6) with $d_{it} = 1$ when i is a first-time forecaster at time t supposes that 'learning' is instantaneous, because next period $d_{i,t+1} = 0$ and the erstwhile newcomer is assumed to be as good as the incumbents on average. This may be a reasonable assumption if learning is rapid, but might give a misleading picture if learning is slow. In order to keep the estimation of the model simple, we investigate (2) for three selected rates of learning.[14] That is, we

[14] For the histograms we suppose that the response to learning does not depend on the quarter. This makes the findings for the histograms readily comparable to those for the point forecasts, and avoids a number of additional complications. For example, whether learning for Q1 surveys (say) should only depend on earlier Q1 surveys.

do not attempt to estimate the rate of learning endogenously, but consider whether the newcomer effect depends on the assumed rate.

We set $d_{it} = \phi^{\tau_{it}-1}$, where τ_{it} is a count of the number of forecasts made by respondent i. d_{it} is undefined for $t < t_i$, when t_i is the period when i first responded, and observations $\{s_{it}, \tau_{it}\}$ are missing values for $t < t_i$.[15] For $t = t_i$, $\tau_{it} = 1$, and thereafter τ_{it} increments by one each time i makes a forecast. If we set $\phi = 0.1$, then the cost of being a relative newcomer decreases by a factor of 10 each period, denoting rapid learning. The first-time effect ($\tau_{it} = 1$) is β, next time it is $\frac{\beta}{10}$, and so on. We also consider $\phi = 0.5$ and $\phi = 0.9$, where the latter corresponds to slow learning.

Table 3 for point forecasts suggests that joiners are significantly worse for forecasting inflation, when we do not suppose that learning is near-instantaneous, that is, when $\phi = 0.5, 0.9$. For output growth the results are as in table 1: allowing for slower rates of learning does not suggest newcomers are worse than incumbents.

For the GDP histogram forecasts allowing learning has no impact on the findings for joiners (table 4): joiners are no worse than the experienced forecasters. However, allowing for learning does now clearly suggest that joiners are less capable inflation histogram forecasters. The rejection of the null that $\beta = 0$ does not depend on the rate of learning, ϕ.

In this section we have allowed the transition of first-time forecasters to experienced forecasters to be gradual. We also investigate the possibility that the performance of those who leave at t may change before survey t. Specifically, we define $D_{it-1} = 1$ and $D_{it} = 1$ if individual i's last response is in period t. The results are shown in tables 5 and 6 for point and histogram forecasts respectively. For the point forecasts, the effect of leaving remains statistically significant for the GDP forecasts at both horizons, although the effect size is reduced for the current-quarter forecasts. The main change for the point forecasts is for inflation: leavers are now worse for the short-horizon forecasts. For the histogram forecasts of inflation (table 6) leavers are worse at all horizons.

5.1.3 Forecast efficiency

As well as testing whether point forecast accuracy differs across different groups using (2), in principle one might consider testing whether forecaster rationality or efficiency varies across groups. The main approach is due to Mincer and Zarnowitz (1969) (henceforth, MZ), who estimate the

[15] When the respondent i was an active particpant prior to the beginning of the sample (that is before 1980), we set $d_{it} = 0$.

regression:

$$y_t = \delta_0 + \delta y_{i,t|t-h} + u_{i,t} \tag{7}$$

for a particular h, on all the forecasts made by respondent i.[16] The MZ test of the null of 'weak efficiency' is a joint test that $\delta_0 = 0$ and $\delta = 1$. Under the null, the forecast error has a zero mean, and is not correlated with the forecast. Such a correlation could be exploited to generate a superior forecast. A number of authors have argued that pooled cross-section time-series regressions, or such regressions with fixed effects, are not an appropriate vehicle for testing for rationality: see, e.g., Zarnowitz (1985) and Bonham and Cohen (2001). For this reason, we do not consider weak efficiency by testing the null that H_0: $\delta_0 = 0$ and $\delta = 1$ in (7), but do consider a more restricted notion, that is, whether δ_0 is different for different groups, having imposed $\delta = 1$. We estimate (2) but with $s_{it} = y_t - y_{i,t|t-h}$ (to impose the equivalent of $\delta = 1$ in (7)). Hence β in (2) is the average difference in the bias for joiners (say) compared to experienced forecasters.

Table 7 suggests that the average bias (the average of the coefficients on the time dummies) is small for both variables for both horizons. For example, for inflation 4-quarters ahead it is -0.12: on average inflation forecasts were just over a tenth of a percentage point too high over the period 1991 – 2014. The estimates $\hat{\beta}$ are generally small and insignificant. The exception is for current-quarter GDP leavers, for which the null $\beta = 0$ is borderline significant at the 5% level (one-sided test). The increase in the bias is one thirtieth of a percentage point (0.036), up from a baseline level of 0.050. Hence the findings for bias are similar to the findings for squared-error loss in table 1.

5.1.4 Histogram forecast variances

The results for the evaluation of the histogram forecasts in table 2 could in principle be driven by differences in the abilities of the different groups of forecasters to predict the central tendency (the mean, or point forecast). That is, experienced forecasters might have an edge over joiners in terms of predicting the central tendency of the distribution, and this might translate into superior histogram RPS scores. For this reason we formally assess the histogram forecast variances to see whether there are differences in the groups perceptions of uncertainty.

Clements (2014, 2018) finds that the US SPF respondents are generally not able to accurately gauge the uncertainty they face (see also, Kenny, Kostka and Masera (2014), Knüppel and Schultefrankenfeld (2019)). To consider whether joiners and leavers differ from experienced remainers

[16] The dependence of δ_0 and δ on i and h is suppressed in the notation.

in this regard, we estimate (6) with the dependent variable s_{it} set to the variance of individual i's histogram in response to survey t. In section 3.2 we explain how the variances of the histograms are calculated. The $\hat{\beta}$ estimates are positive and statistically significant for joiners for both output and inflation for three of the four forecast horizons. Joiners tend to have higher perceptions of forecast uncertainty at the two longer-term horizons for both variables. The increase in the Q4 inflation forecast variance of joiners is half as much again as the baseline value (0.24, compared to 0.49, averaging over horizons). Hence joiners account for some of the increase in *ex ante* uncertainty over *ex post* uncertainty for forecasters as a whole at short horizons, as documented by Clements (2014) (but only a small percentage given the small number of joiners). Leavers' forecast-error variances were higher for both variables for third-quarter surveys.

5.2 Alternative modelling assumptions - two-way fixed effects and autocorrelation

Results of the panel regressions with two-way fixed effects, or autocorrelation, are given in tables 9 and 10 for the current-quarter output growth and inflation point forecasts, in tables 11 and 12 for the year-ahead quarter output growth and inflation point forecasts, and in tables 13 and 14 for the histogram forecasts.

Relative to allowing time fixed effects only, the current-quarter output growth point forecasts of leavers remain significantly worse when we use two-way fixed effects (tables 9, column (2)), but the difference becomes insignificant when we allow for autocorrelation (column (7)).

The inflation point forecasts of leavers allowing for learning remain worse under two-way fixed effects (tables 10, column (5) and 12, columns (4) and (5)).

Under two-way fixed effects, the output growth histograms of leavers in the third quarter of the year surveys remain less accurate than those of stayers. The chief change from including individual fixed effects is that the evidence for leavers' inflation histogram forecasts being less accurate vanishes (table 14 column (2)), and the evidence for joiners' inflation histogram forecasts being less accurate is weaker (only statistically significant for $\phi = 0.9$: see table 14, column (5)).

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 forecast performance of recent joiners to the survey, relative to that of experienced forecasters, or whether those who leave have performed less well on average than those who remain. We address this for the US SPF, the longest-running source of expectations of professional forecasters containing both point forecasts and histogram forecasts.

The findings depend to an extent on whether we allow individual fixed effects (in addition to time fixed effects). When we do not allow for unobserved heterogeneity, we find differences between joiners, leavers, and the rest. Namely, joiners are relatively more disadvantaged at forecasting inflation than GDP growth when it comes to histogram forecasting, especially if we allow that learning takes time. Experienced forecasters have acquired knowledge by doing which translates into more accurate forecasts. The findings for whether leaving-forecasters are less accurate - as might be the case if they expend less effort - are somewhat nuanced. Leavers' GDP point forecasts are less accurate on average, but there is little evidence that their histogram of GDP growth are worse. For inflation, leavers' histogram forecasts are less accurate, whether we suppose the effect comes into play the survey before the last response is made, or only affects the last response.

Once we allow for unobserved heterogeneity by including individual fixed effects, these findings remain largely unchanged except that leavers' inflation histogram forecasts no longer appear to be less accurate than those of the remainers. If we suppose forecasters are essentially the same (aside from the joiner / experienced distinction, say), then individual fixed effects are unwarranted, and without the between-variation, estimates of β may be less precisely determined.

We surmise that the dependence of some of the findings on the modelling assumption partly reflects the small number of joiners, averaging only 1.4 per survey, making it difficult to disentangle individual-specific effects, and joiner/leaver effects.

The findings we have presented relate to a single survey of macroeconomic expectations, albeit one of the most important surveys of professional forecasters expectations. Future research might usefully consider other surveys, especially if such surveys had more joiners and leavers (as distinct from occasional non-response by otherwise active participants).

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.

Batchelor, R. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, *23*(2), 189–203.

Binder, C. C. (2019). Panel Conditioning in the Survey of Consumer Expectations. Discussion paper. Available at SSRN: https://ssrn.com/abstract=3330376.

Boero, G., Smith, J., and Wallis, K. F. (2011). Scoring rules and survey density forecasts. *International Journal of Forecasting*, **27(2)**, 379–393.

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

Bonham, C., and Cohen, R. (2001). To aggregate, pool, or neither: Testing the rational expectations hypothesis using survey data. *Journal of Business and Economic Statistics*, **190**, 278–291.

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. (2014). Forecast Uncertainty - Ex Ante and Ex Post: US Inflation and Output Growth. *Journal of Business & Economic Statistics*, **32(2)**, 206–216. DOI: 10.1080/07350015.2013.859618.

Clements, M. P. (2018). Are macroeconomic density forecasts informative?. *International Journal of Forecasting*, **34**, 181–198.

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. (2019). *Macroeconomic Survey Expectations*: Palgrave Texts in Econometrics.

Palgrave Macmillan. DOI 10.1007/978-3-319-97223-7.

Clements, M. P. (2020). Are some Forecasters' Probability Assessments of Macro Variables Better Than those of Others?. *Econometrics*, **8, 16**.

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.

Giacomini, R., and Rossi, B. (2009). Detecting and predicting forecast breakdown. *Review of Economic Studies*, **76**, 669–705.

Giacomini, R., and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, **25**, 595–620.

Kenny, G., Kostka, T., and Masera, F. (2014). How Informative are the Subjective Density Forecasts of Macroeconomists?. *Journal of Forecasting*, *33*(3), 163–185.

Knüppel, M., and Schultefrankenfeld, G. (2019). Assessing the uncertainty in central banks' inflation outlooks. *International Journal of Forecasting*, **35**, 1748–1769.

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

Lahiri, K., and Sheng, X. (2010). Measuring forecast uncertainty by disagreement:the missing link. *Journal of Applied Econometrics*, **25**, 514–538.

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.

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.

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

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

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.

Zarnowitz, V. (1985). Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics*, **3(4)**, 293–311.

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

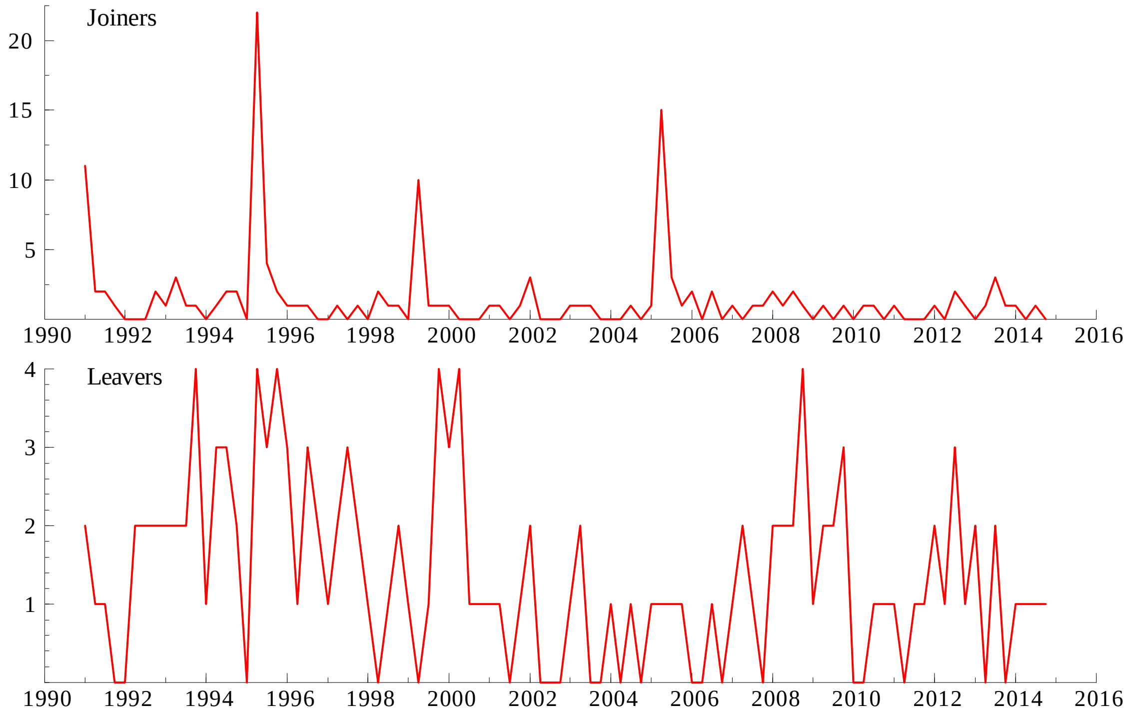

Figure 1: At each survey, 1991:1 - 2014:4, the number of active participants designated as first-time joiners and leavers (those making no subsequent forecasts).

Table 1: Point Forecasts, Squared-error loss

	GDP		Inflation	
	Current-quarter			
	Joiners	Leavers	Joiners	Leavers
$\widehat{\beta}$	-0.0091	0.0878	0.0069	0.0156
p-value	0.6260	0.0006	0.3580	0.1914
BS p-value	0.6030	0.0140	0.2670	0.1990
Constant	0.172	0.169	0.092	0.093
	4-quarters ahead			
	Joiners	Leavers	Joiners	Leavers
$\widehat{\beta}$	-0.0068	0.0581	0.0201	-0.0078
p-value	0.5613	0.0808	0.1715	0.6503
BS p-value	0.4850	0.0470	0.1080	0.6100
Constant	0.322	0.320	0.109	0.110

The table shows the estimates of eqn. (2) where the dummy variable d_{it} is for first-time joiners, or leavers, and the dependent variable is the squared forecast error.
The p-value is for a one-sided test of H_0: $\beta = 0$ versus H_1: $\beta > 0$. That is, it is the probability of obtaining a test statistic at least as large as the observed value if the null hypothesis is true.
BS p-value is the proportion of the bootstrap distribution that exceeds $\hat{\beta}$.
The Constant is the average of the estimated coefficients on the time series dummies. $\widehat{\beta}$ measures the change in the dependent variable relative to this level when $D_{it} = 1$.

Table 2: Current year-on-year Histogram Forecasts. RPS

	Joiners				Leavers			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	GDP							
$\widehat{\beta}$	0.0402	-0.0788	0.0737	-0.0047	0.0227	-0.0503	0.1159	0.0560
p-value	0.2482	0.9549	0.1103	0.5263	0.3272	0.8143	0.0112	0.1346
BS p-value	0.2420	0.9870	0.0710	0.4480	0.3060	0.8540	0.0190	0.1710
Constant	0.311				0.309			
	Inflation							
$\widehat{\beta}$	0.0903	0.0857	0.0477	0.0673	0.0109	0.1658	0.0730	0.1165
p-value	0.0537	0.0187	0.1827	0.1397	0.4059	0.0003	0.0465	0.0054
BS p-value	0.0580	0.0240	0.1480	0.1320	0.3890	0.0020	0.0410	0.0110
Constant	0.220				0.219			

The table shows the estimates of eqn. (6) where the dummy variable d_{it} is for first-time joiners, or leavers, and the dependent variable is the histogram RPS.
See notes to table 1.

Table 3: Point Forecasts, Squared-error loss, 'Learning-by-Doing' for Joiners

	$\phi = 0.1$	$\phi = 0.5$	$\phi = 0.9$	$\phi = 0.1$	$\phi = 0.5$	$\phi = 0.9$
		GDP			Inflation	
			Current-quarter			
$\widehat{\beta}$	-0.0076	0.0165	0.0260	0.0120	0.0392	0.0225
p-value	0.6042	0.2648	0.0728	0.2629	0.0119	0.0282
			4-quarters ahead			
$\widehat{\beta}$	-0.0056	0.0070	0.0027	0.0220	0.0362	0.0365
p-value	0.5504	0.4316	0.4616	0.1514	0.0320	0.0030

The table shows the estimates of eqn. (2) where the dummy variable d_{it} is for joiners adapted to allow for learning (see main text), and the dependent variable is the squared forecast error.
ϕ is the rate at which learning is supposed to occur, where $\phi = 0.1$ is near instantaneous, and $\phi = 0.9$ is slow.
The p-value is for a one-sided test of H_0: $\beta = 0$ versus H_1: $\beta > 0$. That is, it is the probability of obtaining a test statistic at least as large as the observed value if the null hypothesis is true.

Table 4: Histogram Forecasts, RPS, 'Learning-by-Doing' for Joiners

	$\phi = 0.1$	$\phi = 0.5$	$\phi = 0.9$	$\phi = 0.1$	$\phi = 0.5$	$\phi = 0.9$
		GDP			Inflation	
$\widehat{\beta}$	-0.0108	-0.0158	-0.0090	0.0816	0.0938	0.0527
p-value	0.6466	0.7257	0.6892	0.0008	0.0000	0.0006

The table shows the estimates of eqn. (2) where the dummy variable d_{it} is for joiners adapted to allow for learning (see main text), and the dependent variable is the histogram RPS.
ϕ is the rate at which learning is supposed to occur, where $\phi = 0.1$ is near instantaneous, and $\phi = 0.9$ is slow.
The p-value is for a one-sided test of H_0: $\beta = 0$ versus H_1: $\beta > 0$. That is, it is the probability of obtaining a test statistic at least as large as the observed value if the null hypothesis is true.

Table 5: Point Forecasts, Squared-error loss, Leavers Two-quarter Effect

	GDP		Inflation	
	Current	4-quarter	Current	4-quarter
$\widehat{\beta}$	0.0412	0.0395	0.0275	-0.0021
p-value	0.0176	0.0920	0.0153	0.5578
BS p-value	0.0350	0.0460	0.0350	0.4950

The table shows the estimates of eqn. (2) where the dummy variable $d_{it} = 1$ for time t-leavers, and for $t-1$. The dependent variable is the squared forecast error.
See notes to table 1.

Table 6: Current year-on-year Histogram Forecasts, RPS, Leavers Two-quarter Effect

	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	GDP				Inflation			
$\widehat{\beta}$	-0.0247	-0.0304	0.0547	0.0637	0.0822	0.0847	0.0786	0.0653
p-value	0.7360	0.7911	0.0632	0.0508	0.0080	0.0051	0.0055	0.0280
BS p-value	0.6790	0.8480	0.0470	0.0810	0.0240	0.0110	0.0090	0.0300

The table shows the estimates of eqn. (6) where the dummy variable $d_{it} = 1$ for time t-leavers, and for $t-1$. The dependent variable is the histogram RPS.
See notes to table 1.

Table 7: Point Forecasts, Bias

	GDP		Inflation	
	Current-quarter			
	Joiners	Leavers	Joiners	Leavers
$\widehat{\beta}$	0.0084	0.0360	0.0009	-0.0068
p-value	0.3506	0.0432	0.4832	0.6357
Constant	0.050	0.049	-0.062	-0.062
	4-quarters ahead			
	Joiners	Leavers	Joiners	Leavers
$\widehat{\beta}$	0.0236	0.0273	0.0078	0.0187
p-value	0.1324	0.0873	0.3393	0.1481
Constant	-0.081	-0.081	-0.123	-0.123

The table shows the estimates of eqn. (2) where the dummy variable $d_{it} = 1$ for first-time joiners, and leavers, and the dependent variable is the forecast error.
See notes to table 1.

Table 8: Histogram Forecast Variances

	Joiners				Leavers			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
				GDP				
$\widehat{\beta}$	0.6725	0.4915	0.3523	-0.0629	0.0497	0.1670	0.2987	0.0129
p-value	0.0000	0.0000	0.0084	0.6416	0.3473	0.1141	0.0084	0.4588
Constant			0.647				0.658	
				Inflation				
$\widehat{\beta}$	0.2507	0.4347	0.1740	0.2416	0.0314	0.2175	0.2278	0.0872
p-value	0.0193	0.0000	0.0632	0.0361	0.3757	0.0197	0.0079	0.1898
Constant			0.487				0.492	

The table shows the estimates of eqn. (6) where the dummy variable $d_{it} = 1$ for first-time joiners, and leavers, and the dependent variable is the forecast variance of the histograms.
See notes to table 1.

Table 9: Point Forecasts. Output growth, Current-quarter, Additional Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			2-way FEs			Time FEs, AR(1)	
	Joiners	Leavers	Joiners	Joiners	Joiners	Joiners	Leavers
$\widehat{\beta}_J$	-0.0655					0.0104	
	(-2.24)					(0.14)	
$\widehat{\beta}_Q$		0.0625					0.0430
		(2.18)					(1.43)
$\phi = 0.1$			-0.0717				
			(-2.41)				
$\phi = 0.5$				-0.0656			
				(-2.25)			
$\phi = 0.9$					-0.0473		
					(-1.62)		
Constant	0.169	0.168	0.169	0.168	0.164	0.0584	0.0577
	(4.79)	(4.77)	(4.78)	(4.75)	(4.62)	(1.38)	(1.37)
Observations	3768	3768	3768	3768	3768	3598	3598

t statistics in parentheses

Columns (1) and (2) give the coefficients from regressions on the joiner and leaver dummies ($\widehat{\beta}_J$ and $\widehat{\beta}_{QJ}$) with two-way fixed effects. Columns (3) to (5) are for joiners with 3 different rates of learning, with two-way fixed effects. Columns (6) and (7) include time effects only and the errors are allowed to exhibit AR(1) correlation within panels.

Table 10: Point Forecasts. Inflation, Current-quarter, Additional Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			2-way FEs			Time FEs, AR(1)	
	Joiners	Leavers	Joiners	Joiners	Joiners	Joiners	Leavers
$\widehat{\beta}_J$	-0.00951					-0.0970	
	(-0.48)					(-1.93)	
$\widehat{\beta}_Q$		-0.00379					-0.00252
		(-0.20)					(-0.12)
$\phi = 0.1$			-0.00553				
			(-0.27)				
$\phi = 0.5$				0.0292			
				(1.47)			
$\phi = 0.9$					0.0362		
					(1.79)		
Constant	0.0870	0.0873	0.0871	0.0883	0.0920	-0.112	-0.103
	(3.66)	(3.67)	(3.66)	(3.71)	(3.85)	(-3.94)	(-3.66)
Observations	3617	3617	3617	3617	3617	3453	3453

t statistics in parentheses

See notes to table 9.

Table 11: Point Forecasts. Output growth, 4-quarter ahead, Additional Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			2-way FEs			Time FEs, AR(1)	
	Joiners	Leavers	Joiners	Joiners	Joiners	Joiners	Leavers
$\widehat{\beta}_J$	-0.0539					0.0408	
	(-1.11)					(0.24)	
$\widehat{\beta}_Q$		0.0481					0.0121
		(1.02)					(0.24)
$\phi = 0.1$			-0.0517				
			(-1.04)				
$\phi = 0.5$				-0.0370			
				(-0.76)			
$\phi = 0.9$					-0.00305		
					(-0.06)		
Constant	0.301	0.301	0.301	0.301	0.302	-0.214	-0.217
	(4.97)	(4.96)	(4.97)	(4.96)	(4.95)	(-2.32)	(-2.36)
Observations	3638	3638	3638	3638	3638	3470	3470

t statistics in parentheses

See notes to table 9.

Table 12: Point Forecasts. Inflation, 4-quarter ahead, Additional Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			2-way FEs			Time FEs, AR(1)	
	Joiners	Leavers	Joiners	Joiners	Joiners	Joiners	Leavers
$\widehat{\beta}_J$	0.0157					-0.0114	
	(0.72)					(-0.15)	
$\widehat{\beta}_Q$		-0.00684					0.000402
		(-0.32)					(0.02)
$\phi = 0.1$			0.0174				
			(0.78)				
$\phi = 0.5$				0.0362			
				(1.64)			
$\phi = 0.9$					0.0559		
					(2.47)		
Constant	0.0690	0.0689	0.0691	0.0701	0.0760	-0.0906	-0.0899
	(2.54)	(2.54)	(2.54)	(2.58)	(2.79)	(-2.16)	(-2.15)
Observations	3523	3523	3523	3523	3523	3361	3361

t statistics in parentheses

See notes to table 9.

Table 13: Histogram Forecasts. Output growth, Additional Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			2-way FEs			Time FEs, AR(1)	
	Joiners	Leavers	Joiners	Joiners	Joiners	Joiners	Leavers
$\widehat{\beta}_{J,1}$	0.0297					0.0594	
	(0.46)					(0.31)	
$\widehat{\beta}_{J,2}$	-0.0739					0.172	
	(-1.48)					(0.88)	
$\widehat{\beta}_{J,3}$	0.000873					-0.0143	
	(0.01)					(-0.05)	
$\widehat{\beta}_{J,4}$	-0.0525					-0.0512	
	(-0.69)					(-0.17)	
$\widehat{\beta}_{L,1}$		-0.00312					-0.00719
		(-0.06)					(-0.13)
$\widehat{\beta}_{L,2}$		0.0114					0.00630
		(0.18)					(0.09)
$\widehat{\beta}_{L,3}$		0.102					0.0953
		(1.80)					(1.57)
$\widehat{\beta}_{L,4}$		0.0338					0.0133
		(0.62)					(0.23)
$\phi = 0.1$			-0.0371				
			(-1.17)				
$\phi = 0.5$				-0.0462			
				(-1.48)			
$\phi = 0.9$					-0.0584		
					(-1.82)		
Constant	0.341	0.340	0.340	0.340	0.335	0.0942	0.0913
	(8.13)	(8.10)	(8.12)	(8.10)	(7.96)	(1.52)	(1.48)
Observations	3491	3491	3491	3491	3491	3322	3322

t statistics in parentheses

As notes to table 9, except that in columns (1), (2), (6) and (7) the joiner and leaver dummies are multiplied by quarterly dummies for the quarter of the year the survey falls in ($\widehat{\beta}_{J,i}$ and $\widehat{\beta}_{L,i}$).

Table 14: Histogram Forecasts. Inflation, Additional Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			2-way FEs			Time FEs, AR(1)	
	Joiners	Leavers	Joiners	Joiners	Joiners	Joiners	Leavers
$\widehat{\beta}_{J,1}$	0.0115					0.0133	
	(0.21)					(0.09)	
$\widehat{\beta}_{J,2}$	0.0109					0.0402	
	(0.26)					(0.27)	
$\widehat{\beta}_{J,3}$	0.0146					0.0959	
	(0.25)					(0.41)	
$\widehat{\beta}_{J,4}$	0.00557					-0.902	
	(0.09)					(-4.01)	
$\widehat{\beta}_{L,1}$		0.00247					-0.00662
		(0.06)					(-0.14)
$\widehat{\beta}_{L,2}$		0.0682					-0.00527
		(1.31)					(-0.10)
$\widehat{\beta}_{L,3}$		0.0171					0.0236
		(0.38)					(0.49)
$\widehat{\beta}_{L,4}$		-0.00783					-0.0360
		(-0.17)					(-0.75)
$\phi = 0.1$			0.0124				
			(0.47)				
$\phi = 0.5$				0.0225			
				(0.87)			
$\phi = 0.9$					0.0504		
					(1.90)		
Constant	0.131	0.131	0.131	0.132	0.136	-0.0248	-0.0171
	(3.87)	(3.86)	(3.87)	(3.88)	(4.01)	(-0.52)	(-0.36)
Observations	3423	3423	3423	3423	3423	3258	3258

t statistics in parentheses

As notes to table 9, except that in columns (1), (2), (6) and (7) the joiner and leaver dummies are multiplied by quarterly dummies for the quarter of the year the survey falls in ($\widehat{\beta}_{J,i}$ and $\widehat{\beta}_{L,i}$).