**Federal Reserve Bank of Cleveland Working Paper Series**

# Practice Makes Perfect: Learning Effects with Household Point and Density Forecasts of Inflation

James Mitchell, Taylor Shiroff, and Hana Braitsch

Working Paper No. 24-25

November 2024

**Suggested citation:** Mitchell, James, Taylor Shiroff, and Hana Braitsch. 2024. "Practice Makes Perfect: Learning Effects with Household Point and Density Forecasts of Inflation." Working Paper No. 24-25. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202425.

**Federal Reserve Bank of Cleveland Working Paper Series**
ISSN: 2573-7953

# Practice Makes Perfect: Learning Effects with Household Point and Density Forecasts of Inflation*

James Mitchell† Taylor Shiroff‡ Hana Braitsch§

November 5, 2024

**Abstract**

This paper shows how both the characteristics and the accuracy of the point and density forecasts from a well-known panel data survey of households' inflationary expectations – the New York Fed's Survey of Consumer Expectations – depend on the tenure of survey respondents. Households' point and density forecasts of inflation become significantly more accurate with repeated practice of completing the survey. These learning gains are best identified when tenure-based combination forecasts are constructed. Tenured households on average produce lower point forecasts of inflation, perceive less forecast uncertainty, round their uncertainty but not their point forecasts, report unimodal densities, and provide internally consistent point and density forecasts.

**Keywords**: inflation expectations; surveys; forecaster heterogeneity; combination forecasts; density forecasting; learning

**JEL Codes**: C53, D84, E31, E37

*We thank Oli Coibion, Ina Hajdini, Ed Knotek, Mathieu Pedemonte, Michael Weber, and participants at the ISF 2023 and a Cleveland Fed brownbag for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

†Research Department, Federal Reserve Bank of Cleveland, USA; Email address: james.mitchell@clev.frb.org

‡Research Department, Federal Reserve Bank of Cleveland, USA; Email address: taylor.shiroff@clev.frb.org

§Former employee of the Federal Reserve Bank of Cleveland

# 1 Introduction

Recent years have seen the establishment of an increasing number of large panel surveys to measure, on a regular and ongoing basis, the expectations of households or consumers. These household surveys complement long-standing surveys of professional forecasters, such as, in the US, the Blue Chip survey and the Survey of Professional Forecasters (SPF), and existing surveys of households, such as the Michigan Survey of Consumers, that have a much more limited panel structure.[1] Surveys of expectations are designed both to help macroeconomists understand expectation formation and to model and forecast macroeconomic outcomes. Our interest is the utility of household panel survey data when forecasting.

The macroeconomic forecasting literature, albeit often focused on surveys of professional forecasters, has found, although empirical results do vary across time and space, that survey-based forecasts tend to be at least as accurate as model-based forecasts (for example, see Ang et al. (2007)). But, as reviewed in Weber et al. (2022), the expectations of different types of agent can be very different. In turn, papers such as Carroll (2003), Ang et al. (2007), Madeira and Zafar (2015), Trehan (2015), and Mitchell and Zaman (2023) have shown, over different samples and surveys, how the relative utility of different types of survey respondent varies. The literature has tended to conclude that, at least when point forecasting, households forecast less well than professionals, although some types of household have been found to produce more accurate forecasts than others. For example, Binder (2015) finds that high-income, highly educated males produce better forecasts of inflation. This paper contributes to this literature, seeking to understand what drives heterogeneity in households' macroeconomic expectations and, in turn, differences in their forecast accuracy. It does so by emphasizing a new dimension of heterogeneity: the tenure of the survey respondent.

An appealing feature of many of these surveys is often believed to be their panel structure, enabling the answers from a given respondent to be tracked over time. In this paper we focus on drawing out the implications – relevant to macroeconomic forecasters – of this panel structure for households' expectations of inflation, as measured monthly in the US since 2013 by the Federal Reserve Bank of New York in its Survey of Consumer Expectations (SCE). Many other central banks, including the Bank of Canada and the European Central Bank, also now directly measure house-

[1] The Michigan Survey questions around 500 households each month, but as a rotating panel where respondents participate at most twice with at least a 6-month lag between their two survey responses. This longer gap between surveys is likely to lower the size of any tenure-based effects that we go on to analyze in this paper.

holds' expectations of inflation and maintain panel databases. So our results are of relevance beyond the US.

The SCE asks households to report both their subjective point and density expectations of inflation one and three years ahead. Each month, the New York Fed then publishes, as a headline measure of expectations, estimates that take averages over all households present in the SCE that month. The median expectation is emphasized for the point forecasts. But the SCE is a rotating 12-month panel, such that households participate in the survey for up to 12 months in a row. This means that these aggregated measures of inflationary expectations are treating all respondents *equally*, both those that are new to the survey that month and those with "tenure."[2] Bellemare et al. (2020) and Kim and Binder (2023) show how, for panel surveys of household inflation expectations, repeat survey participants in fact have expectations distinct from those of new respondents. Respondents appear to "learn" through the process of repeatedly filling in the survey. This accords with the notion of a "superforecaster," and the Good Judgment Project of Tetlock and Gardner (2015), which shows that training makes for better forecasts from a "crowd" of amateur forecasters.[3]

In this paper, we explore the implications of this apparent "learning-through-survey" for the construction of accurate point and density forecasts from household panel survey data. We propose the use of tenure-based "consensus" point and density forecasts. These consensus forecasts follow the tried-and-tested practice (see Timmermann (2006)) of taking equal-weighted linear combinations of forecasts. But rather than take the average over all forecasters, they take an average only over those forecasts made by households with a specific tenure. When constructing tenure-based consensus density forecasts, this amounts to using the linear opinion pool to combine households' histogram forecasts (see Aastveit et al. (2019)), equal weighting households with a given tenure and zero weighting households with a different tenure. Such an approach follows in the spirit of the recommendation of Engelberg et al. (2011), who, when analyzing professional forecasters in the US SPF, suggest that consensus forecasts should be constructed from sub-panels of fixed (not changing) composition over time. In our case, we fix the tenure of the forecaster since this is one way of acknowledging that "consensus" forecasts are likely based on individual forecasts that are very different.

We find that the accuracy of both the point and density forecasts of households

[2]Tenure refers to how many previous surveys the respondent has undertaken. So a household with tenure=1 is new to the survey. A household with tenure=12 is participating in its last survey (in the SCE), having participated in the previous 11 surveys.

[3]Clements (2021) also finds some evidence of "learning by doing" for professional forecasters in the US SPF.

improves with tenure. This improvement is shown to be associated with changing features of households' forecasts of inflation. With increased tenure, we highlight how households on average: (i) produce lower point forecasts; (ii) perceive less forecast uncertainty; (iii) round their histogram forecasts but not their point forecasts; (iv) reduce their histogram forecasts to two or fewer bins; (v) report unimodal densities; and (vi) report point forecasts that are internally consistent with their density forecasts. In fact, households with tenure (point) forecast as well as professional forecasters. But when we, as is common in previous work, simply take an average over all households, we do indeed find that households forecast less well than professionals.

Our results therefore support the view, made in Kim and Binder (2023), that the very process of repeatedly taking a survey encourages households to pay more attention to inflation and either acquire information about the economy between survey waves or report their expectations more accurately and "sensibly" – so that when subsequently asked the same set of questions, their expectations not only improve in accuracy but also look "better behaved," as captured by the aforementioned six empirical characteristics. Consistent with the wider literature on the heterogeneity of expectations mentioned above, we show that these tenure-based learning effects are heterogeneous. We find faster and larger tenure-based improvements in forecast accuracy for the less-educated, lower-income, low-numeracy, and female respondents.

The remainder of this paper is structured as follows. Section 2 introduces the SCE and explains its panel structure. Section 3 identifies six empirical features that show how households' point and density expectations for inflation vary by tenure. Section 4 then evaluates how the accuracy of these point and density forecasts of inflation depends on tenure. We conclude in Section 5 by emphasizing the benefits, when using panel data on expectations to construct inflation forecasts, of discriminating between the expectations of respondents on the basis of their tenure.

# 2 SCE Panel Data on Inflation Expectations

The SCE is a monthly, nationally representative online survey of household heads, and the underlying micro panel survey data are publicly available (with a lag) via the New York Fed's SCE webpage (https://www.newyorkfed.org/microeconomics/sce#/). The SCE contains rich information about respondents' socio-economic and demographic characteristics. We make use of individual responses from the SCE to questions asking individual households for their one-year- and three-years-ahead expectations of inflation. We consider both individuals' point forecasts and their probabilistic (density) forecasts, as elicited via a separate histogram-type question. This requires households to provide the probabilities that they think inflation, in one or three years time, will fall within prescribed bins.[4] We use monthly SCE data from the beginning of the survey (in June 2013) through December 2022, so that we can evaluate these expectations relative to the subsequent monthly realizations of the annual inflation rate (the 12-month percentage change) through December 2023 (for one-year-ahead expectations) and June 2024 (for three-years-ahead expectations).

As indicated, the SCE has a rotating panel structure, with respondents staying in the survey for up to 12 consecutive months. On average, there are about 1,250 respondents each month, with about 150 to 200 new respondents joining the panel each month. This means that, along with sample attrition as some respondents drop out of the survey, over 85 percent of the around 1,250 survey respondents each month have "tenure," in the sense that they have completed the survey on at least one prior occasion. Close to 50 percent of consumers complete the survey more than four times.

Kim and Binder (2023) emphasize how sample selection occurs due to panel attrition. That is, certain types of respondents are more likely to stay in the panel for the full 12 months. This is indeed the case in the SCE, with, as Table A.1 in the online appendix shows, the tenure 12 group being older, more educated, and more numerically literate than the tenure 1 group. To distinguish tenure effects from attrition effects, in the analysis below we follow Kim and Binder (2023) and present results both for the raw "attrition" sample and when we restrict the sample to "non-attriters," namely, respondents who eventually participate for the maximum (12) number of times. The attrition sample comprises 143,158 responses made by 18,961 unique respondents. The non-attrition sample comprises 71,628 responses made by 5,969 individual households.[5] As expected, approximately one-twelfth of the non-attrition sample is in each of the 12 tenure groups.[6] Importantly, as discussed below, our main conclusions about if and how tenure affects household point and density forecasts of inflation are robust to whether we consider the attrition or non-attrition sample.

To account for outliers at the household level, we trim from the attrition sample (from which the non-attrition sample is in turn extracted) those individuals whose

[4]The bin intervals are: $(\infty, -12\%]$, $[-12\%, -8\ \%]$, $[-8\%, -4\%]$, $[-4\%, -2\%]$, $[-2\%, 0\%]$, $[0\%, 2\%]$, $[2\%, 4\%]$, $[4\%, 8\%]$, $[8\%, 12\%]$, $[12\%, \infty)$. We set $-25\%$ and $25\%$ as the lower and upper bounds in the outermost bins.

[5]In defining these samples, we also drop a small number of additional responses due to data problems. Specifically, we drop 41 respondents who stayed in the survey longer than 12 months; and we drop 307 respondents with incomplete responses to at least one of the four inflation expectations questions. Finally, we exclude those respondents (less than 1 percent of respondents) whose probabilistic responses to the histogram question do not sum to 100 percent.

[6]See Figure A.1 in the online appendix.

responses to the point forecast question are in the top or bottom 5 percent of that month's distribution of responses. Qualitatively, our results are robust to variations in this trimming proportion.

# 3 Tenure-Based Features of Households' Expectations of Inflation

In surveying the recent empirical literature on households' inflation expectations, D'Acunto et al. (2023) document systematic differences by individual characteristics, notably gender, income, and education.[7] In this section we present six empirical facts that reinforce the finding that inflation expectations are heterogeneous across individual households, but that emphasize a new dimension of heterogeneity: tenure in the survey.

Specifically, we delineate six features of households' expectations of inflation that depend on tenure.[8] We compute these features by tenure group, averaging over other individual characteristics, such as age, income, and education, that are also tracked in the SCE. In Section 4, although our focus remains on tenure, we do evaluate households' forecasts across these additional dimensions of heterogeneity. In characterizing these six features, we also pool over time. While there are meaningful temporal changes over our 2013-2023 sample, in particular in the point and uncertainty forecasts that we evaluate below in Section 4, this variation neither obscures nor changes the conclusions about the marked tenure effects.[9]

In summary, Figure 1 illustrates how repeat respondents are more likely to:

1. report lower point forecasts for inflation;

2. report lower uncertainty ranges around their forecasts;

[7]Knotek et al. (2024) and Weber et al. (2024) emphasize the importance of measuring a respondent's information set, which affects expectation formation, showing how information and attentiveness to that information differ systematically across demographic groups and time.

[8]To be clear, the first two of these features are direct updates of estimates reported by Bellemare et al. (2020) and/or Kim and Binder (2023); but we reconsider and subsequently confirm them (on an extended sample and using a complementary, less restrictive measure of uncertainty) given their importance for our focus on a tenure-based assessment of households' point and density forecasts. Similarly, the sixth empirical feature is documented in Zhao (2023).

[9]In the online appendix (Section A.6) we plot the temporal evolution of these six features. As also noted in Armantier et al. (2021), the most notable temporal change is an increase in households' point and histogram mean inflation forecasts, and forecast uncertainty, in the aftermath of the COVID-19 pandemic.

3. round their histogram bin forecasts but not their point forecasts;
4. reduce their histogram forecasts to two or fewer bins;
5. report unimodal density forecasts;
6. report point forecasts that are "consistent" (as defined below) with their density forecasts.

In the remainder of this section we provide details of and motivation for how these six features in Figure 1 are computed. All use the non-attrition sample. The online appendix contains equivalent figures calculated using the attrition sample, as well using subsamples that exclude the inflationary post-COVID-19 period. These robustness exercises confirm that the six features documented in Figure 1 hold over other cuts of the SCE sample. We also emphasize that, in documenting features of the histogram-based forecasts, we favor inference that does not involve first fitting a parametric density to the histograms, a popular strategy when using histogram-based forecasts. As becomes clear as we explain the results below, even fitting a relatively flexible density, like the generalized beta (following Engelberg et al. (2009)), to each household's histogram ("bins") is an assumption rejected for some households. We therefore estimate the moments of each consumer's histogram forecast nonparametrically.[10] Given that we treat the histogram forecasts as histograms, henceforth we shall refer to them as that rather than as "densities."

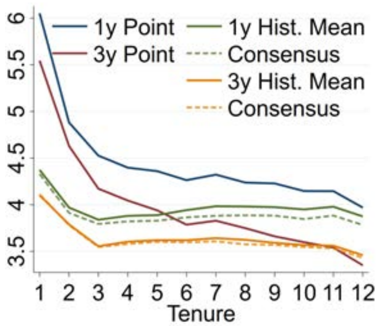
For the histogram forecasts, there is a also question of whether we report their features having first computed the consensus histogram (by taking the average across households in a given tenure group) or whether we first compute the features of interest (such as the uncertainty estimate) for each individual household histogram and then take the average. We consider both strategies.

## 3.1 Expected Inflation Declines

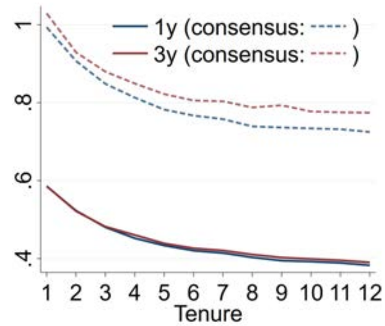
We examine households' point forecasts of inflation and their mean forecasts, as extracted nonparametrically from their histogram forecasts. For the histogram question, we also compare the mean of the consensus histogram forecast with the average (across households) of each household's mean forecast, as extracted from their individual histogram. We should expect these two sets of estimates to be similar, since,

[10]Specifically, we use a standard formula to estimate the first two moments of a histogram nonparametrically; for example, see Clements (2019).

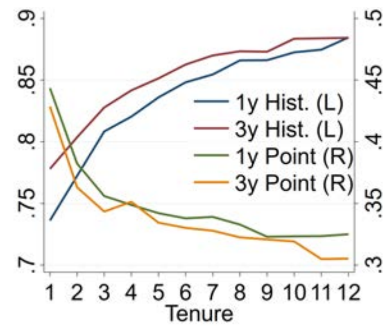
Figure 1: Six Empirical Features of Households' Inflation Expectations by Tenure



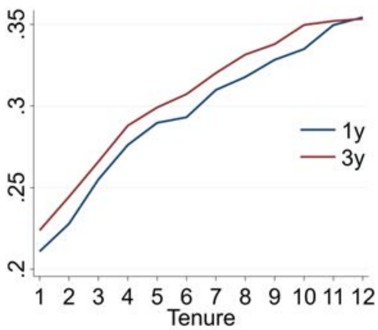

(a) Point and Hist Mean



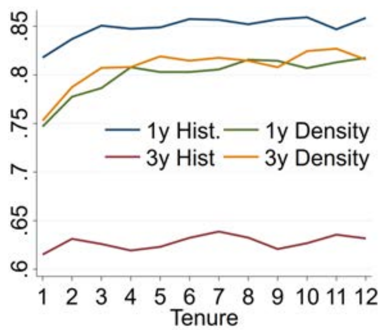

(b) ERPS



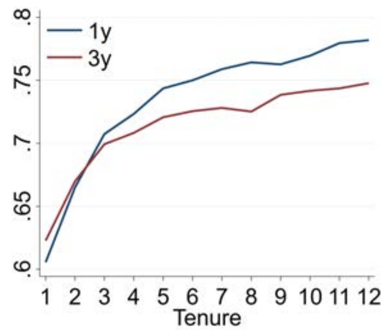

(c) Prop rounding




(d) Prop $\leq$ 2 bins




(e) Prop unimodal ($\geq$ 3 bins)




(f) Prop. consistent

Notes: All panels use the non-attrition sample of 71,628 responses. "1y" denotes one-year-ahead forecasts, "3y" three-years-ahead forecasts. Panel (a) reports by tenure group the mean (consensus) point forecast and the equal-weighted average (across households) mean of each household's histogram forecast ("Hist. Mean"). The mean of the consensus histogram forecast is then shown in the dotted line. Panel (b) reports the mean (across households) expected rank probability score (ERPS), with the ERPS of the consensus histogram forecasts in the dotted line. Panel (c) plots the proportion of respondents rounding their point (right axis) and histogram (left axis) forecasts. Panel (d) reports by tenure group the proportion of respondents replying to the histogram question in fewer than three bins. Panel (e) plots the proportion of respondents reporting unimodal histogram forecasts as either measured directly via the histogram or having first fitted a generalized beta density to each histogram. Panel (f) reports the proportion of respondents whose point expectations for inflation are "consistent" with their histogram forecasts, as defined by Engelberg et al. (2009).

when forecasts are combined linearly, the mean of a (continuous) combined or consensus density forecast equals the average of each individual density's mean forecast (for example, see Wallis (2005)).

Panel (a) of Figure 1 shows that the tenure-based consensus point forecasts for inflation both one-year- and three-years-ahead fall by tenure. With repeated practice of filling in the survey, on average households lower their expectations of inflation. For the point forecasts, over the 12 months of tenure, the forecasts of inflation at both horizons drop by around 2 percentage points. This is not, perhaps, an "economically insignificant" fall, given that the Fed targets 2 percent inflation over the longer run. This result holds when we exclude the higher inflation period observed in the aftermath of the COVID-19 pandemic (see Figure A.3 in the online appendix). The average histogram-based mean forecast and the mean of the consensus histogram forecast also fall by tenure, albeit not as sharply as for the point forecasts. The fact that the histogram question forces respondents to provide their probabilistic forecasts within specified bins is perhaps disciplining their responses by ruling out more extreme forecasts. In Panel (f) of Figure 1, discussed below, we examine the consistency between households' point and histogram forecasts.

## 3.2 Forecast Uncertainty Declines

To assess how households' ex ante perceptions of forecast uncertainty, as measured by their histogram forecasts, vary by tenure, in Panel (b) of Figure 1 we plot by tenure group the expected rank probability score (ERPS), as defined by Krüger and Pavlova (2024). The ERPS is a measure of uncertainty implicit in histogram forecasts.[11] An attraction of using the ERPS to measure uncertainty is that, unlike when estimating the standard deviation, it can be computed even for respondents who report their histograms in fewer than three bins. But we emphasize that, for those respondents who reply using three or more bins, the ERPS and standard deviation deliver the same story.[12] As in Panel (a), we examine both average individual uncertainty (the mean of each individual histogram's ERPS) and the uncertainty (ERPS) of the consensus histogram forecast.

We should expect the uncertainty of the consensus forecast to be higher than average individual uncertainty, since at least when using the variance to measure "uncertainty," the variance of a consensus density forecast can be decomposed into two positive terms: the average individual variance and the variance ("disagreement") of the individual point forecasts (see Wallis (2005)). It is therefore no surprise that

[11] Let $\underline{p}$ be a $K \times 1$ vector containing the probabilities a respondent attaches to inflation falling within each of the $K$ bins. Denote the cumulative probability of inflation falling within the first $k$ bins by $P_k = \sum_{j=1}^{k} p_j$. The ERPS is defined as $ERPS(\underline{p}) = \sum_{k=1}^{K} P_k (1 - P_k)$. Given the format of the SCE bins, see footnote 4, ERPS is bound between 0 and 2.25, with higher values corresponding to greater uncertainty and vice-versa. The ERPS is maximized when $\underline{p} = (0.5, 0, ..., 0, 0.5)'$, i.e., when the outermost bins have equal probabilities of one-half.

[12] See Figure A.2 in the online appendix.

Panel (b) of Figure 1 does show the ERPS of the consensus histogram forecast to be higher. Either way, this panel shows that forecast uncertainty monotonically declines with tenure at both forecast horizons.

## 3.3 More Rounding of the Histogram, Less for Point Forecasts

Respondents are well known to round their numerical answers to survey questions; for example, they might forecast that inflation is going to be 3.5 percent rather than 3.36 percent. Panel (c) of Figure 1 examines the degree to which "rounding" (to a multiple of 5) varies with tenure.[13] Panel (c) shows that, as tenure increases, a lower fraction of households round their point forecasts.

This finding is consistent with previous research, especially when coupled with panel (b) of Figure 1, since it suggests that a forecaster's ex ante impression of uncertainty declines with tenure as they become more confident. Binder (2017) links the rounding of point forecasts to uncertainty. Glas and Hartmann (2022) suggest that rounding is a consequence of forecasts being formed judgmentally rather than from a model, while Manski and Molinari (2010) consider it a consequence of simplifying communication and/or a signal of partial rather than full knowledge. Our results therefore indicate that, with tenure, households become more confident about what will happen to inflation and are better able to provide these expectations in a non-rounded form. At the same time, however, this is true only for their point forecasts. As Panel (c) shows, when respondents report their histogram forecasts, rounding *increases* with tenure. This may be because, with experience gained through tenure, respondents become more aware of what they do not know, and so they round their bin responses to the histogram question. In this sense, as suggested by Boero et al. (2015), it may be better to view rounding of the histogram forecast as an expression of "uncertain uncertainty," that is, Knightian uncertainty, rather than an expression of knowable uncertainty ("risk").

## 3.4 Fewer Histogram Bins Are Used

Panel (d) of Figure 1 documents an increased tendency with tenure for households to use only one or two bins when reporting their histogram forecasts. This finding is consistent with Panel (b): with tenure, households perceive less forecast uncertainty – that is, they become more confident.

[13] We consider a histogram forecast to be "rounded" if *each* of the up to ten bins is rounded. Not counting zero as a rounded response, nearly everyone ($>$ 99% of respondents) rounds at least one bin.

## 3.5 Forecasts Become More Unimodal

Panel (e) looks at the shape of the histogram forecasts for those households who replied using at least three bins. It plots the proportion of households that reported unimodal histograms. The consensus histogram forecast is always unimodal, so we do not discuss it further here.

We measure unimodality in two ways, which also serves to substantiate our preference (stated above) for nonparametric inference of histogram forecasts. First, we examine the unimodality of each household's histogram forecast nonparametrically. This simply involves ascertaining whether the bin responses have just one (interior) peak (mode). Second, we test for unimodality parametrically by fitting a continuous density to each household's histogram. We follow the New York Fed in its analysis of the SCE histograms by fitting a generalized beta distribution. But we do not enforce unimodality on the generalized beta. In their examination of professional forecasters, Engelberg et al. (2009) found that unimodality was supported, so they fitted the generalized beta constraining the two relevant parameters – let us denote them as is customary $\alpha$ and $\beta$ – to be strictly greater than one. We relax this restriction and calculate the proportion of households, by tenure group, where the maximum likelihood estimates for $\alpha$ and/or $\beta$ are greater than unity.

Panel (e) reveals that about a quarter of the households in the SCE produce multimodal histogram forecasts. Moreover, although this trend is modest, we do see a growing proportion of households producing unimodal histograms as tenure increases. But the level of the proportion is sensitive both to whether unimodality is assessed parametrically or nonparametrically and to the forecast horizon.[14]

## 3.6 Forecasts Become More Consistent

Both the point forecast question and the histogram question ask for households' expectations for inflation, but they pose the question in different ways. Despite this, we might still hope that a given household's answers to these two questions tie together. Panel (f) of Figure 1 tests the internal "consistency" between each household's point and histogram forecasts. This involves examining whether a household's point expectation falls within the range that is consistent with their histogram forecast using the nonparametric bounds methods developed by Engelberg et al. (2009).[15] Since

[14] Consistent with our results, Zhao (2023) finds that, on average over their shorter 2013 through 2020 sample, around 85 percent of consumers in the SCE report unimodal densities when measured nonparametrically.

[15] We should note that violations of consistency do not have to imply that households are being "irrational." There are other explanations for forecast inconsistency, such as households forming

it is not known whether a household's point forecast should be interpreted as the mean, mode, or median of their underlying density forecast, we follow Engelberg et al. (2009) and calculate nonparametric bounds from the histogram that apply to the mean, mode, and median. But in Panel (f), as results are little affected, we focus on the mean. Consistency is seen to rise with tenure. This is again suggestive of households learning-by-experience to construct more coherent, even what we might call more *sensible*, forecasts.

# 4 Forecast Accuracy by Tenure

Having established that the properties of households' point, histogram mean, and histogram forecasts for inflation vary by tenure, we turn to an ex post assessment of if and how their accuracy varies by tenure. We focus on measuring forecast accuracy against the headline PCE inflation rate, the Fed's preferred measure. But we emphasize that our conclusions are robust to evaluating the forecasts against realizations of CPI inflation as well.[16]

We compare expectations for inflation over the next 12 (or 36) months, made in month $t$, to the first release of the realized year-over-year inflation rate, $\pi_{t+12}$ (or $\pi_{t+36}$). We extract these first-release realizations from the ALFRED database maintained by the St. Louis Fed. Our sample comprises monthly inflation outturns from June 2014 through December 2023 for the one-year-ahead forecasts and through June 2024 for the three-years-ahead forecasts. For robustness, we again evaluate forecast accuracy over different samples, aware of temporal instabilities in forecast performance.

We proceed by first evaluating – in Section 4.1 – the accuracy of the consensus (by tenure group) point, histogram mean, and histogram forecasts. We then use – in Section 4.2 – panel data methods to test if the average (by tenure group) accuracy of individual households' point, histogram mean, and histogram forecasts varies by tenure. That is, we distinguish between the accuracy of the average ("consensus") forecast and the average accuracy of individual households. The latter can be interpreted as measuring the accuracy of expectations from a randomly drawn single household (for example, see Bomberger (1996)). We might expect forecast accuracy to be higher for the consensus forecasts, given that combining the individual forecasts means that any errors associated with individual-level household expectations

their point expectations of inflation such that the costs of overpredicting are not equal to those of underpredicting; for example, see Clements (2010).

[16] See Figure A.13 in the online appendix.

can cancel each other out; for examples, see Batchelor and Dua (1995) and Lahiri and Sheng (2010). In both cases, we show if and how accuracy varies by tenure. We begin by first presenting results pooled across individual household characteristics, such as age, gender, and education. But we then disaggregate across these and other observed characteristics of the households.

## 4.1 Tenure-Based Consensus Forecasts

We define combination or "consensus" forecasts by tenure group. The forecasting literature has established that combining a subset of individual forecasts can improve on combinations across all forecasters; for a recent review, see Wang et al. (2023). Most commonly, such "trimmed" combination forecasts involve combining only the more accurate forecasts, something that can be identified only once the outturn is realized. Our tenure-based approach, by contrast, is to define subsets of households over which to combine ex ante, based only on knowledge of tenure. Such an approach has the attraction of being implementable in real time, given that in each month's SCE we know every household's tenure.

We obtain tenure-based consensus forecasts by aggregating individual-level point, histogram mean, and histogram forecasts. There are always questions about the appropriate aggregation to use (see Aastveit et al. (2019) for a review); we take equal-weighted linear combinations within each tenure group.

We measure the accuracy of the consensus point and histogram mean forecasts for each tenure group using the root mean squared error (RMSE). To measure the accuracy of the consensus histogram forecasts, acknowledging the reality that the SCE elicits histogram forecasts rather than (continuous) density forecasts, we follow Boero et al. (2011) and use the ranked probability score (RPS) averaged over the 2013 through 2023/2024 sample of up to $T = 115$ months:

$$RPS = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{10}(P_{kt} - D_{kt})^2 \tag{1}$$

where $P_{kt} = \sum_{k=1}^{k} p_{kt}$, $D_{kt} = \sum_{k=1}^{k} d_{kt}$, $d_{kt}$ are a set of indicator variables, which equal 1 if the outturn for time $t$ falls in bin $k$ and 0 otherwise, and $p_{kt}$, $k = 1, ..., K$, are the set of $K = 10$ probability forecasts attached to each bin of the histogram. We re-emphasize that we prefer not to first fit some continuous density to the histogram. Our aim is to evaluate the histogram forecasts "as is," as actually elicited by the SCE. The RPS evaluates the histogram forecast directly, and penalizes bin probability forecasts when the outcome falls into a bin further away from the ones assigned the highest probabilities by the household.

#### 4.1.1 Forecast Accuracy Gains with Tenure

Figure 2 shows that the forecast accuracy of the consensus point, histogram mean, and histogram forecasts all improve with tenure. Household forecasters get better with practice. These improvements in accuracy are statistically significant, as judged via a Diebold and Mariano (1995) test.[17] While the learning is weaker for the one-year-ahead histogram forecasts over the full sample, this is explained by the post-pandemic rise in inflation. When in the bottom panels of Figure 2 we restrict attention to pre-pandemic data, we again see strong learning effects, even for the histogram forecasts.[18] Clements (2021) also finds some evidence of "learning by doing" for professional forecasters in the US SPF.

When we decompose the mean squared forecast error into the sum of variance and bias squared, we see that these tenure-based gains in RMSE are almost entirely explained by households with tenure producing forecasts with lower bias; see Figure A.8 in the online appendix.

Interestingly, given the aformentioned literature that finds households forecast less well than professionals, these forecast accuracy gains mean that, as households accrue tenure, their accuracy matches and then surpasses that of professional forecasters, when evaluated over the same sample. By the time they have completed the survey 6 times, households produce point forecasts as accurate as those of the professional forecasters surveyed by Blue Chip. Only households with tenure 1 produce point forecasts that are statistically less accurate than the professionals, as judged by a Diebold and Mariano (1995) test at 95 percent. Specifically, the Blue Chip consensus one-year-ahead forecast delivers a RMSE of 2.41 over the same sample period as the non-attrition sample, compared to 2.23 for the tenure 12 households. The lack of forecast data for professionals precludes further comparison with the household forecasts from the SCE. Neither Blue Chip nor other surveys of professional forecasters in the US elicit monthly density forecasts.

It is also of note that we find that households' three-years-ahead forecasts tend to be more accurate than their one-year-ahead forecasts, even pre-pandemic. This may be explained by the fact that households' expectations of inflation are generally higher one year ahead than three years ahead. In the pre-pandemic period, with low

[17]In the figure, for space reasons, we only report the Diebold-Mariano test between the tenure 1 and tenure 12 forecasts. Table A.4 in the online appendix reports tests between the tenure 1 forecasts and those with tenure 2 through tenure 12.

[18]These effects are again statistically significant. See the Diebold-Mariano test statistics on the pre-pandemic sample reported in Table A.2 in the online appendix.

inflation realizations, this translates into improved accuracy for the three-years-ahead forecasts. This is consistent with the wider literature that finds households tend to overestimate current inflation when asked for their "perceptions" (for example, see Axelrod et al. (2018)). In contrast, longer-horizon inflation expectations may better reflect households' expectations of monetary policy and the steps they believe the central bank will take to deliver on its inflation target. In contrast, shorter-run expectations of inflation are less affected by monetary policy, given its famous "long-and-variable lags," and as a result are more variable than the longer-run forecasts. As seen in Figure 1, the three-years-ahead forecasts, especially as households accrue tenure, are indeed better "anchored" around low inflation values.

#### 4.1.2 A Deeper Look at Heterogeneity Across Household Forecasts

Kim and Binder (2023) establish that, as far as households' point expectations for inflation go, the tenure-based learning effects are heterogeneous: households that enter the survey with more uncertainty, for example, show faster learning, as do less numerate, lower-income, and less educated individuals. With a Bayesian updating mechanism in mind, Kim and Binder (2023) conjecture that this finding may be because these types of households have weaker priors and are therefore more likely to update them as they learn through the survey. In this section, we extend their analysis to examine heterogeneity in histogram forecast accuracy, as well as the point forecasts, and across a richer set of individual characteristics. We focus here, for space reasons, on the one-year-ahead forecasts, noting that a similar picture emerges three years ahead.[19]

We now define tenure-based consensus forecasts but we also disaggregate households in the non-attrition sample into 10 subgroups. The first five subgroups reflect demographic factors identified in the SCE. The second five subgoups reflect observed characteristics of a household's own forecasts.

In Figure 3, we evaluate tenured-based consensus point, histogram mean, and histogram forecasts for each of five demographic groups that classify households by age, education, income, numeracy, and gender.[20] We find statistically significant tenure-based gains in forecasting accuracy for all five subgroups' point and histogram mean forecasts. The largest gains are seen for the less-educated, lower-income, low-numeracy, and female respondents. These differing speeds of learning reflect important level differences in forecast accuracy between these subgroups, which can be

[19]Equivalent figures for three-years-ahead forecasts are available in the online appendix; see Figures A.9 and A.10.

[20]Numerical literacy is when the respondent is deemed to have "low" or "high" literacy, as evaluated through a set of test questions in the SCE.

clearly seen in Figure 3. Those subgroups that at tenure 1 were worse forecasters generally learn more quickly, with subsequent improvements in forecast accuracy.

Echoing Figure 2, the learning gains in Figure 3 are weaker for the histogram forecasts. But we do observe faster improvements in accuracy for those groups (the less-educated, lower-income, low-numeracy, and female respondents) that provide, at tenure 1, the least accurate histogram forecasts.

Then in Figure 4 we evaluate the tenure-based consensus forecasts by five characteristics of each household's forecast. This involves disaggregating households into five subgroups reflecting whether: (i) and (ii) they entered the SCE panel with a point forecast or uncertainty forecast (as measured by ERPS) in the top 25 percent of all first-time respondents; (iii) their point and histogram forecasts are consistent with one another; (iv) their histogram forecasts are unimodal (defined nonparametrically); and (v) they used more than two bins when replying to the histogram question.

Again we see in Figure 4 that for nearly all of these subgroups there are statistically significant gains in forecasting accuracy across the three forecast metrics seen in the three rows of Figure 4. These gains are strongest for those households that initially made higher point forecasts, expressed more uncertainty, have inconsistent point and histogram forecasts, produced multimodal histogram forecasts, and replied to the histogram question using two or fewer bins. In other words, the observed tenure-based improvements in forecast accuracy are again strongest for those respondents who, on entering the survey, initially produced the least accurate forecasts.

## 4.2 Panel Data Analysis

In this section we test whether the accuracy of a randomly drawn (by tenure group) individual household's point, histogram mean, or histogram forecast also improves with tenure. We do so, in the spirit of Kim and Binder (2023), by estimating panel regressions with tenure dummies, as well as individual fixed effects. Panel data regressions are also used by Clements (2021) when testing for tenure effects with individual professional forecasters, since they can allow for unobserved heterogeneity. Only when forecasters are the same are individual fixed effects unnecessary.

Specifically, we estimate panel regressions with tenure dummies of the form:[21]

[21] Our estimates of the tenure effects are qualitatively unaffected if we also add in time fixed effects. But we do not include time dummies in our main regression, (2), given that this regression, via $\pi_{t+h}$, already controls for time.

$$L\left(\pi_{t+h}, f(\pi^e_{ist})\right) = \sum_{s=2}^{12} \beta_s \tau_s + \alpha_i + \nu_{it} \tag{2}$$

where the dependent variable, $L\left(\pi_{t+h}, f(\pi^e_{ist})\right)$, is the loss function measuring the accuracy of the $h$-year-ahead point, histogram mean, or histogram forecast of household $i$ with survey tenure $s$ made at time $t$, denoted $f(\pi^e_{ist})$, relative to the subsequent inflation outturn, $\pi_{t+h}$. $\tau_s$ is a dummy variable indicating tenure $s$, $s = 1, ..., 12$, and $\alpha_i$ is an individual fixed effect to control for unobserved heterogeneity. For the point and histogram mean forecasts, mimicking Section 4.1, we define $L\left(\pi_{t+h}, f(\pi^e_{ist})\right) = (\pi_{t+h} - \pi^e_{ist})^2$ to be the squared forecast error for the point or histogram mean forecasts. For the histogram forecasts, we define $L\left(\pi_{t+h}, f(\pi^e_{ist})\right) = RPS_{it}$. The $\beta_s$ coefficients thus measure the average (across individual households) learning effect by tenure.

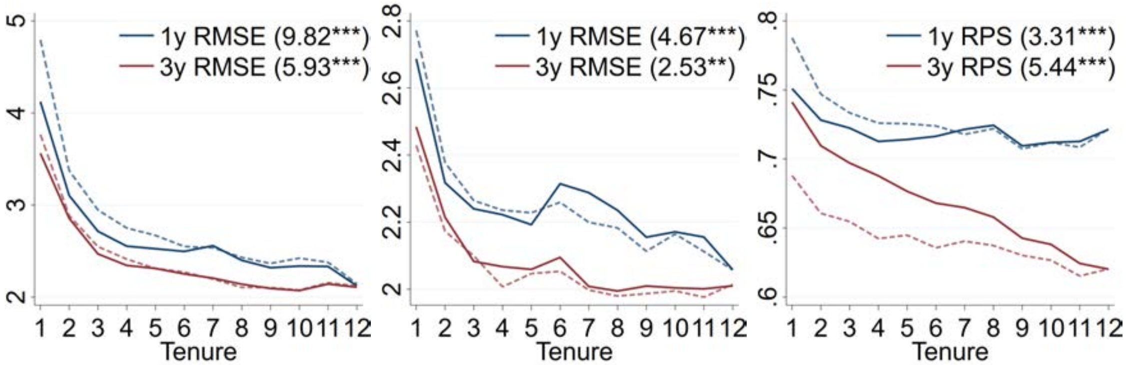
Using the non-attrition sample, Figure 5 plots the estimates of $\beta_s$ from (2), along with 95 percent confidence bands.[22] These show that, as with the consensus forecasts, the accuracy of individual point and histogram-mean forecasts improves with tenure. The negative, and statistically significant, estimates show that, relative to tenure 1 group where $\beta_1 = 0$, the accuracy of households' forecasts increases with tenure. As with the consensus histogram forecasts evaluated in Figure 2, we only see a clear improvement in accuracy for the individual-level histogram forecasts at the three-years-ahead horizon.

We again emphasize that these tenure-based improvements in forecast accuracy are not simply explained by the higher inflation data seen post-pandemic and the greater attention households have been shown to pay to inflation when it is high (cf. Bracha and Tang (2024)). We find similar learning effects when we re-estimate (2) on subsamples that end in December 2019 (see Table A.3 in the online appendix). Finally, we note that when we re-estimate (2) separately across the 10 subgroups considered in Section 4.1.2, we see the same patterns of heterogeneity; see Figures A.6 and A.7 in the online appendix. Less-educated, lower-income, lower numeracy, and female individuals display faster learning, as do those individual households that enter the survey with higher and more uncertain inflation forecasts, that produce inconsistent point and histogram forecasts, that produce multimodal densities, and that characterize forecast uncertainty via fewer histogram bins.

[22] Results are similar using the attrition sample. See Figure A.5 in the online appendix. They are also broadly similar when evaluated against CPI rather than PCE inflation realizations, although the improvements in accuracy for the three-years-ahead histogram forecasts are weaker; see Figure A.14.

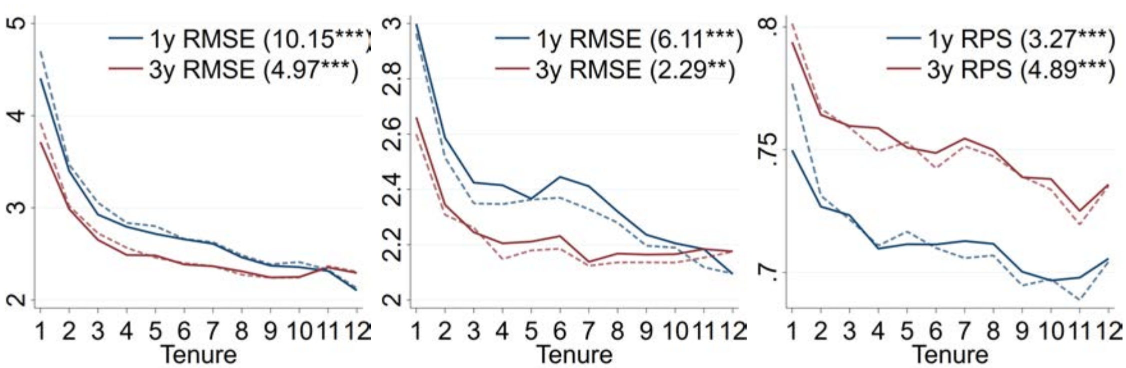
Figure 2: Accuracy of Tenure-Based Consensus Forecasts

Full sample




(a) Point Expectations (b) Histogram Means (c) Histogram

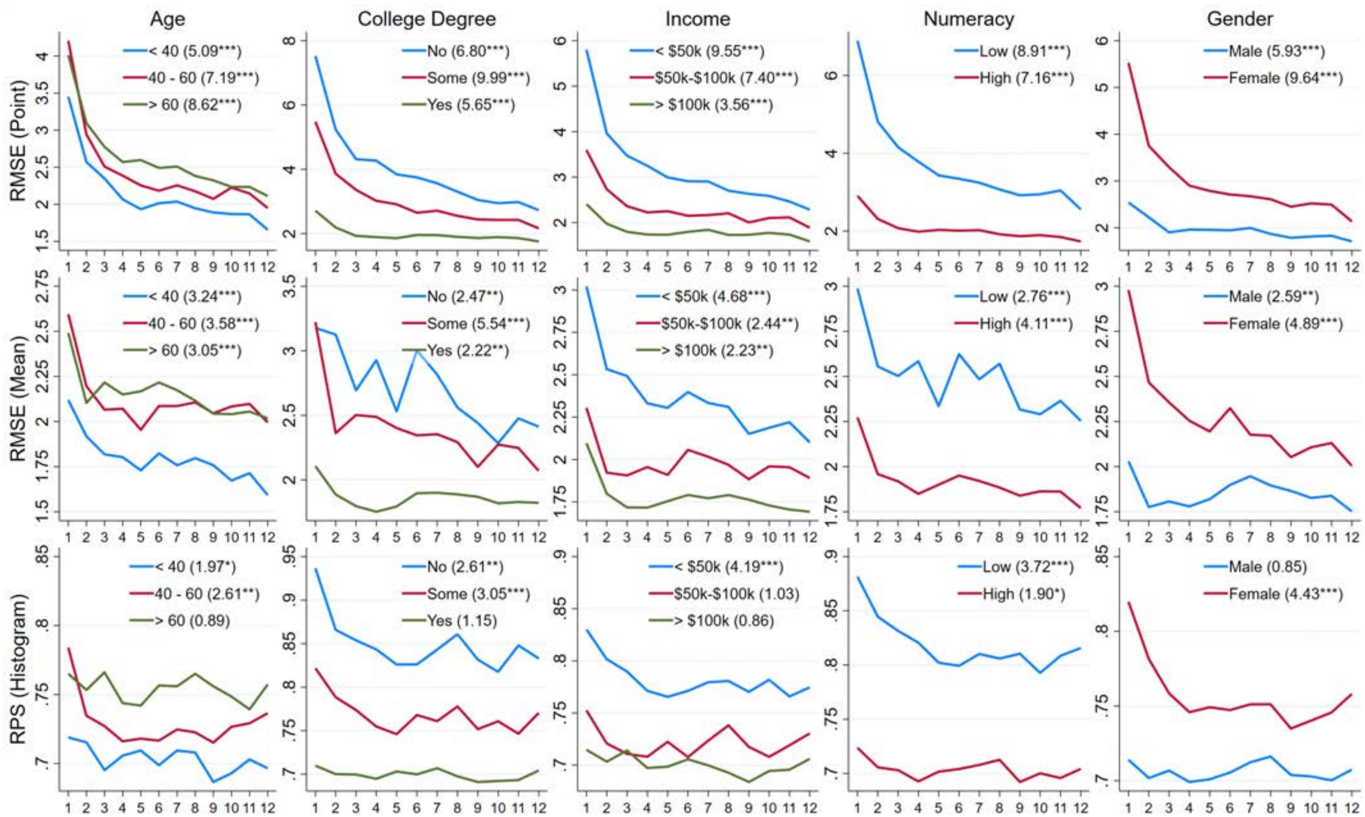
Pre-pandemic sample




d) Point Expectations e) Histogram Means f) Histogram

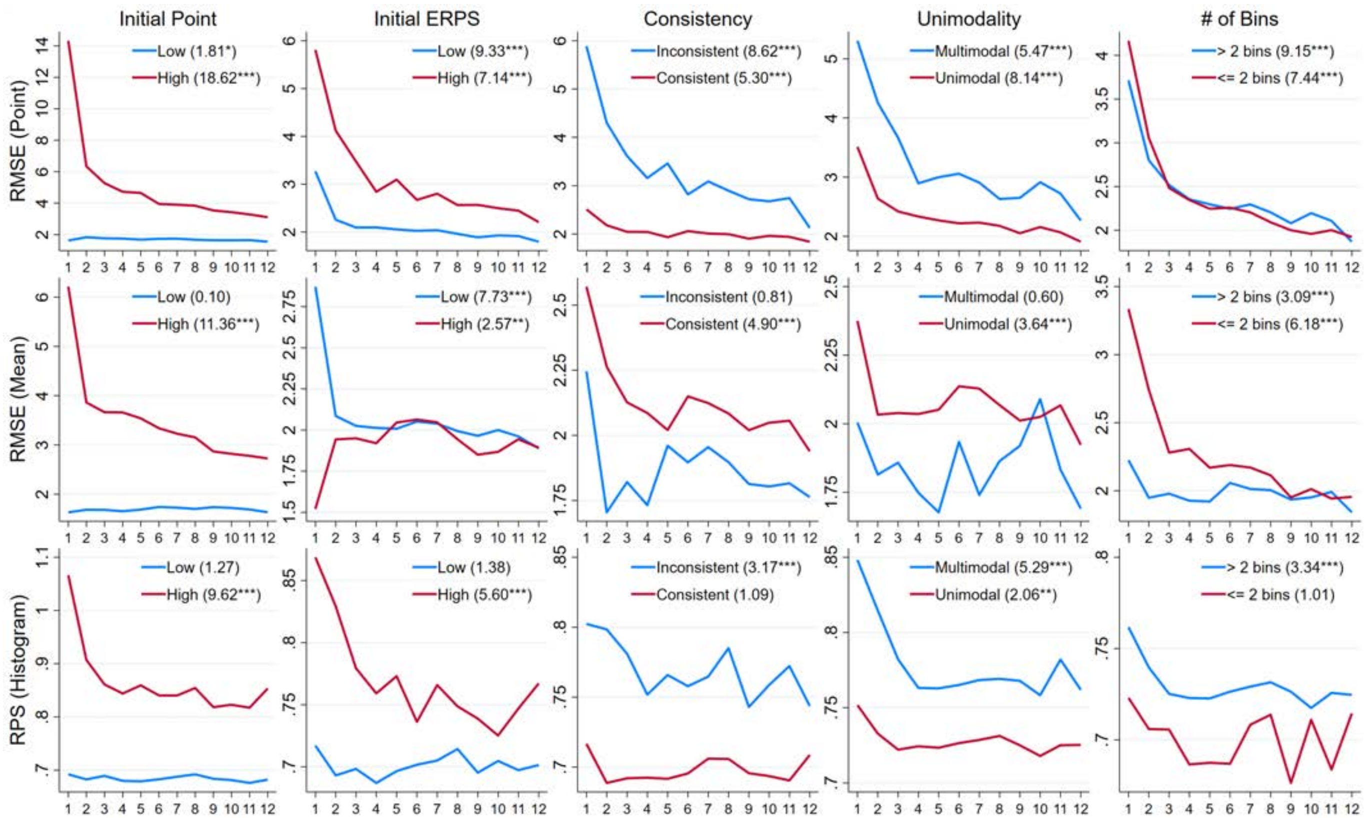
Notes: Full sample evaluates monthly forecasts made from 2013 through 2023. Pre-pandemic sample is monthly from 2013 through 2019m12. Solid and dashed lines indicate non-attrition and attrition samples, respectively. "1y" denotes one-year-ahead, "3y" three-years-ahead forecasts. Numbers in parentheses denote Diebold and Mariano (1995) test statistics for equal forecast accuracy between the tenure 1 and tenure 12 forecasts. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3: Learning Effects for Consensus One-Year-Ahead Forecasts by Respondent Demographics




Notes: Calculated on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and December 2022. The first and second rows show RMSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Diebold-Mariano test statistics comparing tenure 12 consensus forecasts against tenure 1's are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

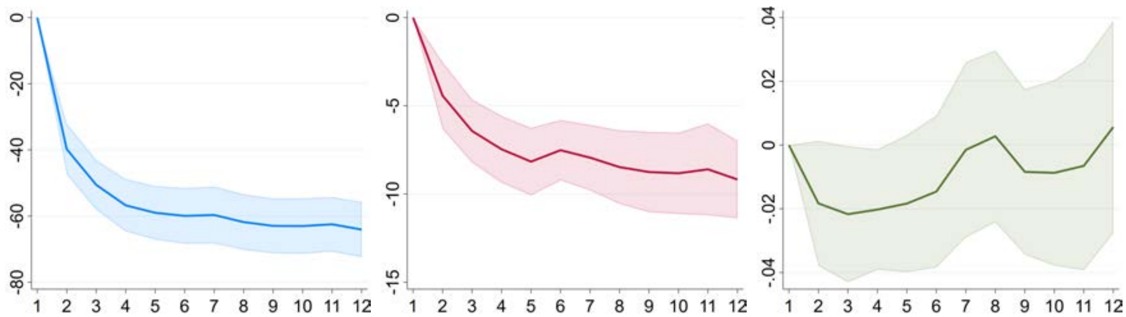
Figure 4: Learning Effects for Consensus One-Year-Ahead Forecasts by Household-Level Forecast Characteristics




Notes: Calculated on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and December 2022. See text for group and subgroup definitions. The first and second rows show RMSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Diebold-Mariano test statistics comparing tenure 12 consensus forecasts against tenure 1's are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

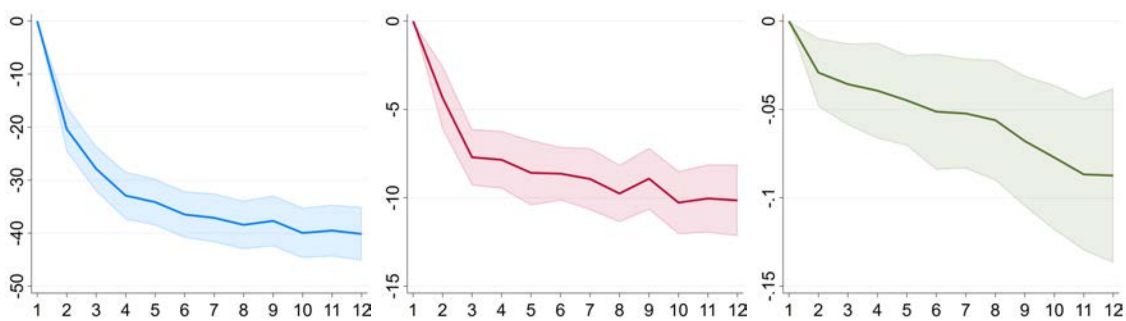
Figure 5: Learning Effects and Individual Forecast Accuracy by Tenure

(a) One-Year-Ahead




i) Point Expectations ii) Histogram Means iii) Histograms

(b) Three-Years-Ahead




i) Point Expectations ii) Histogram Means iii) Histograms

Notes: Panels show estimates of $\beta_s$ from (2) using the non-attrition sample. Shaded areas are 95% confidence intervals using Driscoll-Kraay standard errors with one lag. Forecast errors are defined as the mean squared errors for the point and histogram mean forecasts and the RPS for the histogram forecasts. Underlying sample includes forecasts made between June 2013 and December 2022 (for one-year-ahead forecasts) or June 2021 (for three-years-ahead forecasts).

## 5 Conclusion

This paper demonstrates, using a leading large panel data survey, that as households accrue tenure their point and density forecasts for inflation become more accurate and also look more "sensible." While there remain other well-known important dimensions of heterogeneity in household-level expectations of inflation, notably across socio-economic and demographic factors, our paper emphasizes the importance of constructing survey-based forecasts differentiating between households new to the survey and those with tenure. The very practice of repeatedly completing the survey appears to encourage households to produce better forecasts. By the time they have completed the survey a few times, households' point forecasts match the accuracy of those from professional forecasters. Future research might consider whether such learning effects are also seen in other surveys/countries and for other variables and consider alternative ways of constructing macroeconomic forecasts from panels of changing composition.

# References


Aastveit, Knut Are, James Mitchell, Francesco Ravazzolo, and Herman K. van Dijk (2019). "The Evolution of Forecast Density Combinations in Economics." In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. doi:10.1093/acrefore/9780190625979.013.381.

Ang, Andrew, Geert Bekaert, and Min Wei (2007). "Do macro variables, asset markets, or surveys forecast inflation better?" *Journal of Monetary Economics*, 54(4), pp. 1163–1212. doi:10.1016/j.jmoneco.2006.04.006.

Armantier, Olivier, Gizem Koşar, Rachel Pomerantz, Daphné Skandalis, Kyle Smith, Giorgio Topa, and Wilbert van der Klaauw (2021). "How economic crises affect inflation beliefs: Evidence from the Covid-19 pandemic." *Journal of Economic Behavior and Organization*, 189, pp. 443–469. doi:10.1016/j.jebo.2021.04.036.

Axelrod, Sandor, David E. Lebow, and Ekaterina V. Peneva (2018). "Perceptions and Expectations of Inflation by U.S. Households." Finance and Economics Discussion Series 2018-073, Board of Governors of the Federal Reserve System (U.S.). doi:10.17016/FEDS.2018.073.

Batchelor, Roy and Pami Dua (1995). "Forecaster diversity and the benefits of combining forecasts." *Management Science*, 41(1), pp. 68–75. doi:10.1287/mnsc.41.1.68.

Bellemare, Charles, Rolande Kpekou Tossou, and Kevin Moran (2020). "The Determinants of Consumers' Inflation Expectations: Evidence from the US and Canada." Staff Working Papers 20-52, Bank of Canada. doi:10.34989/swp-2020-52. URL https://ideas.repec.org/p/bca/bocawp/20-52.html.

Binder, Carola C. (2015). "Whose expectations augment the Phillips curve?" *Economics Letters*, 136, pp. 35–38. doi:10.1016/j.econlet.2015.08.013.

Binder, Carola C. (2017). "Measuring uncertainty based on rounding: New method and application to inflation expectations." *Journal of Monetary Economics*, 90, pp. 1–12. doi:10.1016/j.jmoneco.2017.06.001.

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis (2011). "Scoring rules and survey density forecasts." *International Journal of Forecasting*, 27(2), pp. 379–393. doi:10.1016/j.ijforecast.2010.04.003.

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis (2015). "The measurement and characteristics of professional forecasters' uncertainty." *Journal of Applied Econometrics*, 30(7), pp. 1029–1046. doi:10.1002/jae.2400.

Bomberger, William A. (1996). "Disagreement as a measure of uncertainty." *Journal of Money, Credit and Banking*, 28(3), pp. 381–392. doi:10.2307/2077981.

Bracha, Anat and Jenny Tang (2024). "Inflation Levels and (In)Attention." *The Review of Economic Studies*, p. rdae063. doi:10.1093/restud/rdae063.

Carroll, Christopher D. (2003). "Macroeconomic expectations of households and professional forecasters." *Quarterly Journal of Economics*, 118, pp. 269–298. doi:10.1162/003355303360535207.

Clements, Michael P. (2010). "Explanations of the inconsistencies in survey respondents' forecasts." *European Economic Review*, 54(4), pp. 536–549. doi:10.1016/j.euroecorev.2009.10.003.

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

Clements, Michael P. (2021). "Do survey joiners and leavers differ from regular participants? The US SPF GDP growth and inflation forecasts." *International Journal of Forecasting*, 37(2), pp. 634–646. doi:10.1016/j.ijforecast.2020.08.003.

D'Acunto, Francesco, Ulrike Malmendier, and Michael Weber (2023). "What do the data tell us about inflation expectations?" In Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw, editors, *Handbook of Economic Expectations*, chapter 5, pp. 133–161. Academic Press. doi:10.1016/B978-0-12-822927-9.00012-4.

Diebold, Francis X. and Robert S. Mariano (1995). "Comparing predictive accuracy." *Journal of Business & Economic Statistics*, 13(1), pp. 134–144. doi:10.1198/073500102753410444.

Engelberg, Joseph, Charles F. Manski, and Jared Williams (2009). "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters." *Journal of Business & Economic Statistics*, 27, pp. 30–41. doi:10.1198/jbes.2009.0003.

Engelberg, Joseph, Charles F. Manski, and Jared Williams (2011). "Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition." *Journal of Applied Econometrics*, 26(7), pp. 1059–1078. doi:10.1002/jae.1206.

Glas, Alexander and Matthias Hartmann (2022). "Uncertainty measures from partially rounded probabilistic forecast surveys." *Quantitative Economics*, 13(3), pp. 979–1022. doi:10.3982/QE1703.

Kim, Gwangmin and Carola Binder (2023). "Learning-through-survey in inflation expectations." *American Economic Journal: Macroeconomics*, 15(2), pp. 254–78. doi:10.1257/mac.20200387.

Knotek, Edward S., James Mitchell, Mathieu Pedemonte, and Taylor Shiroff (2024). "The effects of interest rate increases on consumers' inflation expectations: The roles of informedness and compliance." Working Papers 24-01, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-202401.

Krüger, Fabian and Lora Pavlova (2024). "Quantifying subjective uncertainty in survey expectations." *International Journal of Forecasting*, 40(2), pp. 796–810. doi:10.1016/j.ijforecast.2023.06.001.

Lahiri, Kajal and Xuguang Sheng (2010). "Measuring forecast uncertainty by disagreement: The missing link." *Journal of Applied Econometrics*, 25(4), pp. 514–538. doi:10.1002/jae.1167.

Madeira, Carlos and Basit Zafar (2015). "Heterogeneous inflation expectations and learning." *Journal of Money, Credit and Banking*, 47(5), pp. 867–896. doi:10.1111/jmcb.12230.

Manski, Charles F. and Francesca Molinari (2010). "Rounding probabilistic expectations in surveys." *Journal of Business and Economic Statistics*, 28(2), pp. 219–231. doi:10.1198/jbes.2009.08098.

Mitchell, James and Saeed Zaman (2023). "The distributional predictive content of measures of inflation expectations." Working paper 23–31, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-202331.

Tetlock, Philip E. and Dan Gardner (2015). *Superforecasting: The Art and Science of Prediction*. Crown Publishers, New York, first edition.

Timmermann, A (2006). "Forecast combinations." In G. Elliott, C.W.J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting Volume 1*, pp. 135–196. North-Holland. doi:10.1016/S1574-0706(05)01004-9.

Trehan, Bharat (2015). "Survey measures of expected inflation and the inflation process." *Journal of Money, Credit and Banking*, 47(1), pp. 207–222. doi:10.1111/jmcb.12174.

Wallis, Kenneth F. (2005). "Combining density and interval forecasts: A modest proposal." *Oxford Bulletin of Economics and Statistics*, 67(s1), pp. 983–994. doi:10.1111/j.1468-0084.2005.00148.x.

Wang, Xiaoqian, Rob J. Hyndman, Feng Li, and Yanfei Kang (2023). "Forecast combinations: An over 50-year review." *International Journal of Forecasting*, 39(4), pp. 1518–1547. doi:10.1016/j.ijforecast.2022.11.005.

Weber, Michael, Bernardo Candia, Hassan Afrouzi, Tiziano Ropele, Rodrigo Lluberas, Serafin Frache, Brent Meyer, Saten Kumar, Yuriy Gorodnichenko, Dimitris Georgarakos, Olivier Coibion, Geoff Kenny, and Jorge Ponce (2024). "Tell me something I don't already know: Learning in low and high-inflation settings." *Econometrica*, Forthcoming. URL https://www.econometricsociety.org/publications/econometrica/forthcoming-papers/0000/00/00/Tell-Me-Something-I-Dont-Already-Know-Learning-in-Low-and-High-Inflation-Settings/file/22764-3.pdf.

Weber, Michael, Francesco D'Acunto, Yuriy Gorodnichenko, and Olivier Coibion (2022). "The subjective inflation expectations of households and firms: Measurement, determinants, and implications." *Journal of Economic Perspectives*, 36(3), pp. 157–84. doi:10.1257/jep.36.3.157.

Zhao, Yongchen (2023). "Internal consistency of household inflation expectations: Point forecasts vs. density forecasts." *International Journal of Forecasting*, 39(4), pp. 1713–1735. doi:10.1016/j.ijforecast.2022.08.008.

# Online Appendix for "Practice Makes Perfect: Learning Effects with Household Point and Density Forecasts of Inflation" by Mitchell, Shiroff, and Braitsch

## A.1 Sample Characteristics

Figure A.1: Average (Across Time) Proportion of Respondents in the SCE Panel by Tenure: Attrition Sample

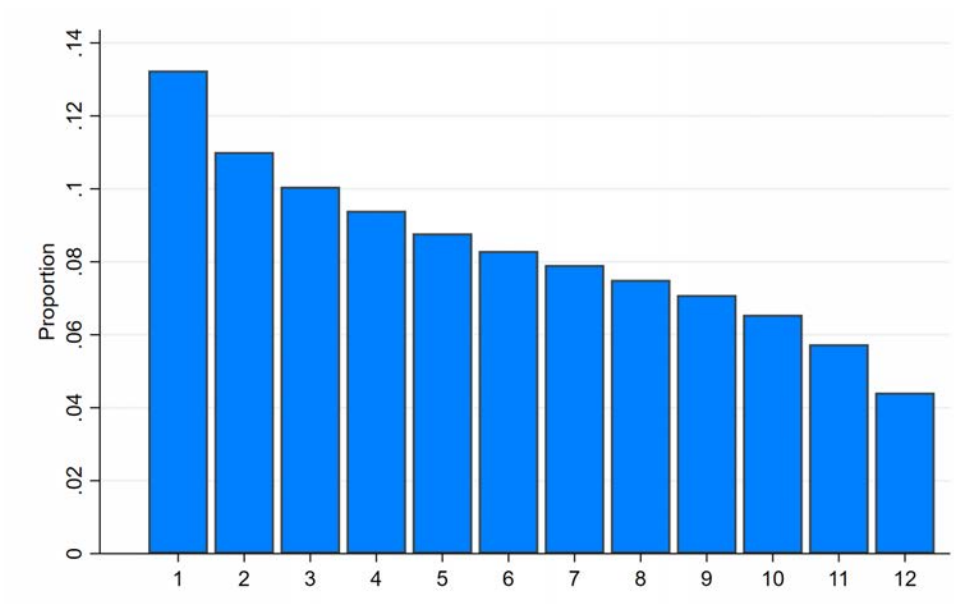
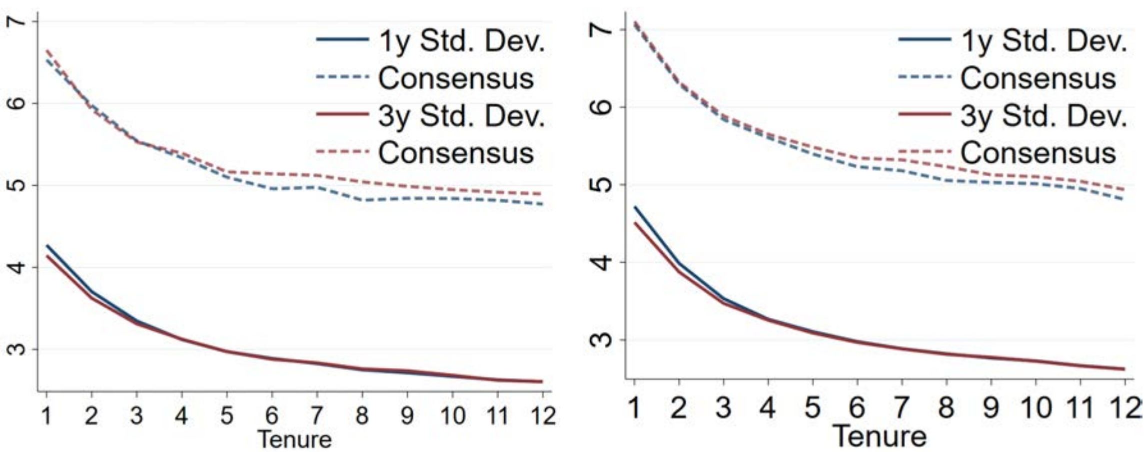


Figure A.2: Standard Deviation of One-Year- and Three-Years-Ahead Histogram Forecasts by Tenure




(a) Non-attrition sample (b) Attrition sample

Notes: Standard deviation computed nonparametrically for those respondents who replied to the histogram question using at least three bins. Dotted lines display the standard deviation of the tenure-based consensus histogram forecasts.

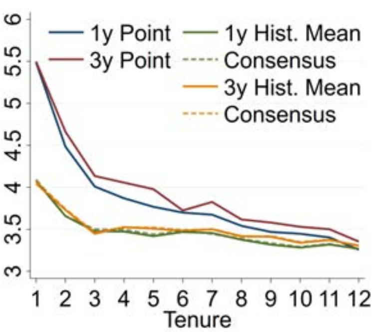
Table A.1: Respondent Characteristics by Tenure: Proportion of Respondents in Each Tenure Group (1 through 12)

| | Attrition Sample (By Tenure) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Characteristic | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Nonattr. |
| Age: Under 40 | 0.30 | 0.30 | 0.30 | 0.30 | 0.29 | 0.29 | 0.28 | 0.28 | 0.28 | 0.27 | 0.26 | 0.25 | 0.25 |
| Age: 40-60 | 0.40 | 0.39 | 0.39 | 0.39 | 0.39 | 0.38 | 0.38 | 0.38 | 0.38 | 0.38 | 0.38 | 0.37 | 0.37 |
| Age: Over 60 | 0.30 | 0.30 | 0.31 | 0.32 | 0.32 | 0.33 | 0.33 | 0.34 | 0.34 | 0.35 | 0.36 | 0.38 | 0.37 |
| Edu.: High School | 0.12 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| Edu.: Some College | 0.34 | 0.33 | 0.33 | 0.33 | 0.33 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.32 | 0.31 | 0.31 |
| Edu.: College | 0.53 | 0.55 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.57 | 0.57 | 0.57 | 0.57 | 0.58 |
| Income: Under $50k | 0.35 | 0.34 | 0.34 | 0.34 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.36 | 0.36 | 0.35 |
| Income: $50k-$100k | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.36 | 0.36 | 0.36 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 |
| Income: Over $100k | 0.29 | 0.30 | 0.30 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.29 | 0.28 | 0.28 | 0.28 | 0.29 |
| Numerical Literacy | 0.69 | 0.71 | 0.72 | 0.72 | 0.72 | 0.72 | 0.72 | 0.72 | 0.72 | 0.72 | 0.73 | 0.74 | 0.75 |
| Consistent 1y | 0.50 | 0.56 | 0.60 | 0.62 | 0.64 | 0.65 | 0.66 | 0.66 | 0.66 | 0.67 | 0.68 | 0.69 | 0.66 |
| Consistent 3y | 0.50 | 0.56 | 0.59 | 0.60 | 0.61 | 0.62 | 0.63 | 0.63 | 0.64 | 0.65 | 0.65 | 0.66 | 0.64 |
| High Initial 1y SD | 0.25 | 0.23 | 0.23 | 0.23 | 0.23 | 0.23 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.21 |
| High Initial 3y SD | 0.25 | 0.24 | 0.23 | 0.23 | 0.23 | 0.23 | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 | 0.21 | 0.21 |
| High Initial 1y Point | 0.27 | 0.27 | 0.26 | 0.26 | 0.26 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.24 | 0.24 |
| High Initial 3y Point | 0.26 | 0.25 | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 | 0.24 | 0.23 | 0.23 | 0.23 |

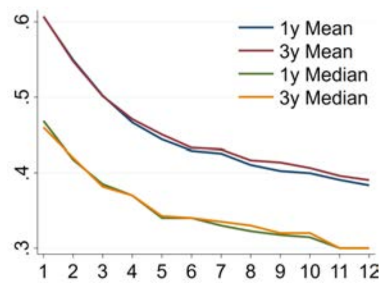
Notes: Proportions across age, education (edu), and income groups may not add to one due to rounding. Numerical literacy is when the respondent is deemed to have "high" literacy, as evaluated through a set of test questions in the SCE. "Consistent" means that the respondent's point and density forecasts are internally consistent. A respondent is classified as having a "high" initial standard deviation (SD) or point expectation if their response is greater than the 75th percentile response among all tenure 1 respondents. "Nonattr." refers to the nonattrition sample.

## A.2 Pre-Pandemic Results

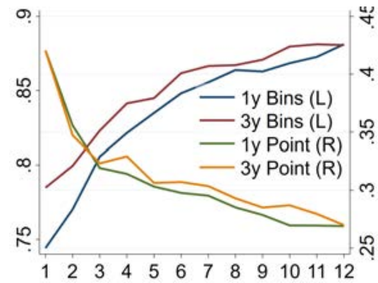
Figure A.3: Six Empirical Features of SCE Inflation Expectations by Tenure
(Pre-Pandemic Subsample: June 2013 - December 2019)



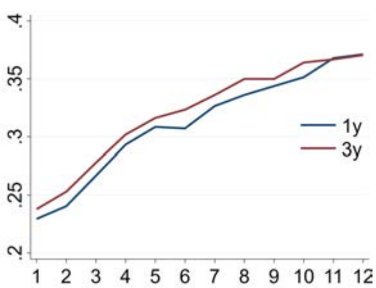

(a) Point and Hist Mean



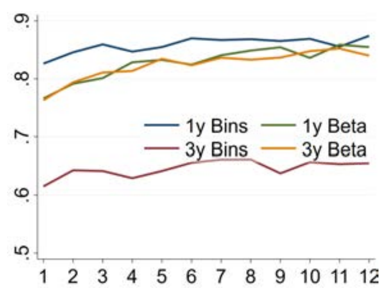

(b) ERPS



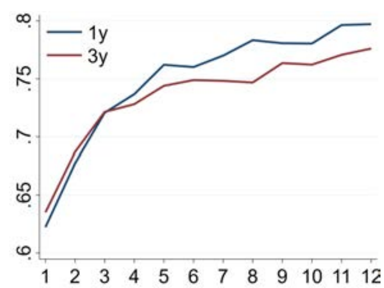

(c) Prop rounding




(d) Prop $\leq$ 2 bins




(e) Prop unimodal ($\geq$ 3 bins)




(f) Prop. consistent

Notes: "Non-attrition" pre-pandemic sample of 48,411 SCE responses. "1y" denotes one-year-ahead forecasts, "3y" three-years-ahead forecasts. Panel (a) reports by tenure group the mean (consensus) point forecast and the mean of each household's histogram forecast. The mean of the consensus histogram forecast is shown in the dotted line. Panel (b) reports the mean expected rank probability score (ERPS), with the ERPS of the consensus histogram forecasts in the dotted line. Panel (c) plots the proportion of respondents rounding their point and histogram forecasts. Panel (d) reports by tenure group the proportion of respondents replying to the histogram question in fewer than three bins. Panel (e) plots the proportion of respondents reporting unimodal histogram forecasts as either measured directly via the histogram or having first fitted a generalized beta density to each histogram. Panel (f) reports the proportion of respondents whose point expectations for inflation are "consistent," in the sense of Engelberg et al. (2009), with their histogram forecasts.

Table A.2: Diebold-Mariano Test Statistics: Pre-Pandemic Sample

| | 1-Year-Ahead | | | 3-Years-Ahead | | |
|---|---|---|---|---|---|---|
| | Point | Histogram Mean | RPS | Point | Histogram Mean | RPS |
| Tenure 2 | -8.84*** | -5.79*** | -6.43*** | -6.51*** | -3.67*** | -5.32*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 3 | -12.35*** | -8.81*** | -7.93*** | -6.60*** | -3.82*** | -5.26*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 4 | -13.00*** | -8.41*** | -9.16*** | -7.28*** | -5.35*** | -6.58*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 5 | -13.78*** | -9.03*** | -8.59*** | -7.44*** | -4.22*** | -6.22*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 6 | -13.20*** | -7.95*** | -8.41*** | -6.87*** | -3.80*** | -5.64*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 7 | -13.96*** | -8.30*** | -8.28*** | -6.90*** | -4.40*** | -5.15*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 8 | -13.98*** | -8.74*** | -8.89*** | -6.73*** | -3.88*** | -4.79*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 9 | -13.37*** | -8.07*** | -11.02*** | -6.73*** | -3.74*** | -5.85*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 10 | -14.50*** | -8.77*** | -8.63*** | -6.72*** | -3.61*** | -6.19*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) |
| Tenure 11 | -13.22*** | -9.26*** | -10.05*** | -5.83*** | -3.17*** | -7.10*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.002) | (0.000) |
| Tenure 12 | -14.02*** | -8.49*** | -6.65*** | -5.75*** | -2.56 | -5.92*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.013) | (0.000) |

Notes: Forecast errors measured using year-over-year PCE inflation realizations. Statistics shown are robust t-statistics testing for equal forecast accuracy of the consensus forecasts of a given tenure group relative to the consensus forecasts of new (tenure=1) households. Sample includes forecasts made by respondents in the non-attrition sample between June 2013 and December 2019, before the pandemic-related inflation surge. $p$ values are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

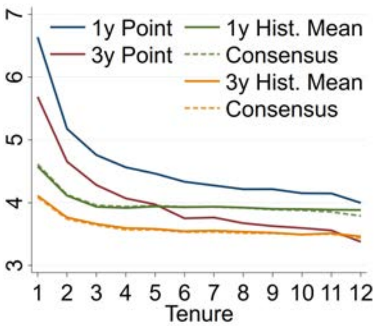
Table A.3: Learning Effects and Forecast Accuracy by Tenure: Pre-Pandemic Sample

| | 1 Year Ahead | | | 3 Years Ahead | | |
|---|---|---|---|---|---|---|
| | Point | Mean | RPS | Point | Mean | RPS |
| $\beta_2$ | -32.74*** | -5.31*** | -0.02* | -19.84*** | -5.20*** | -0.03** |
| | (3.21) | (1.17) | (0.01) | (2.38) | (1.01) | (0.01) |
| $\beta_3$ | -44.74*** | -8.18*** | -0.02** | -27.60*** | -8.45*** | -0.02 |
| | (3.05) | (0.88) | (0.01) | (2.35) | (0.90) | (0.01) |
| $\beta_4$ | -49.32*** | -9.38*** | -0.02** | -32.22*** | -8.42*** | -0.01 |
| | (3.09) | (1.02) | (0.01) | (2.51) | (0.94) | (0.01) |
| $\beta_5$ | -51.32*** | -10.25*** | -0.02** | -34.35*** | -9.36*** | -0.01 |
| | (3.08) | (1.01) | (0.01) | (2.41) | (1.03) | (0.01) |
| $\beta_6$ | -51.41*** | -9.38*** | -0.02* | -36.16*** | -9.03*** | -0.01 |
| | (3.29) | (0.90) | (0.01) | (2.41) | (0.91) | (0.02) |
| $\beta_7$ | -53.37*** | -10.22*** | -0.02 | -37.04*** | -9.87*** | -0.01 |
| | (2.98) | (0.93) | (0.01) | (2.56) | (0.98) | (0.02) |
| $\beta_8$ | -56.24*** | -11.65*** | -0.02 | -38.47*** | -10.50*** | -0.00 |
| | (3.16) | (0.87) | (0.01) | (2.56) | (0.91) | (0.02) |
| $\beta_9$ | -56.94*** | -12.40*** | -0.03** | -38.55*** | -9.60*** | -0.01 |
| | (3.23) | (1.08) | (0.01) | (2.63) | (1.05) | (0.02) |
| $\beta_{10}$ | -57.14*** | -12.72*** | -0.04*** | -39.83*** | -10.59*** | -0.01 |
| | (3.21) | (0.99) | (0.01) | (2.59) | (1.12) | (0.02) |
| $\beta_{11}$ | -57.40*** | -13.06*** | -0.04** | -39.75*** | -10.64*** | -0.02 |
| | (3.35) | (1.18) | (0.01) | (2.74) | (1.19) | (0.02) |
| $\beta_{12}$ | -58.07*** | -12.82*** | -0.03* | -40.18*** | -10.31*** | -0.00 |
| | (3.26) | (0.97) | (0.01) | (2.87) | (1.25) | (0.02) |
| Obs. | 44006 | 43878 | 43878 | 44050 | 43977 | 43977 |
| Adj. $R^2$ | 0.06 | 0.01 | 0.00 | 0.05 | 0.01 | 0.00 |

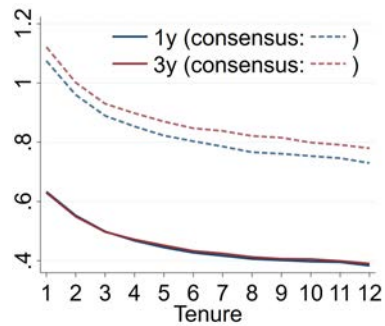
Notes: Estimates from panel regression (2) using the non-attrition sample. Underlying sample from June 2013 through December 2019, excluding the pandemic period, using PCE inflation outturns through December 2022. The dependent variable is the squared point forecast error for the point columns, the squared histogram mean forecast error for the mean columns, and the RPS for the histogram forecasts. Regressions include household-level fixed effects. Driscoll-Kraay standard errors of lag one are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## A.3 Attrition Sample Results

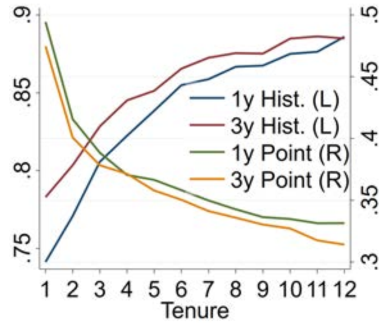
Figure A.4: Six Empirical Features of SCE Inflation Expectations by Tenure



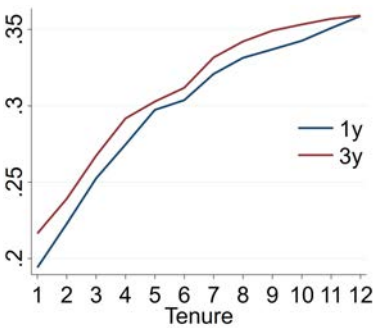

(a) Point and Hist Mean



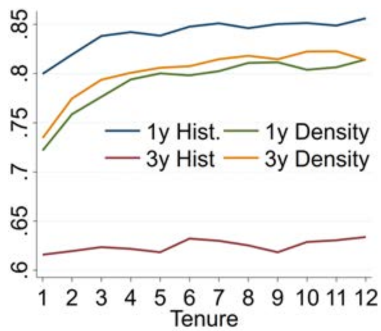

(b) ERPS



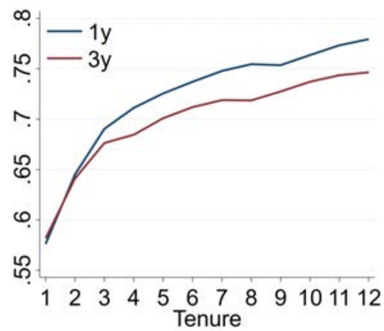

(c) Prop rounding




(d) Prop $\leq$ 2 bins




(e) Prop unimodal ($\geq$ 3 bins)

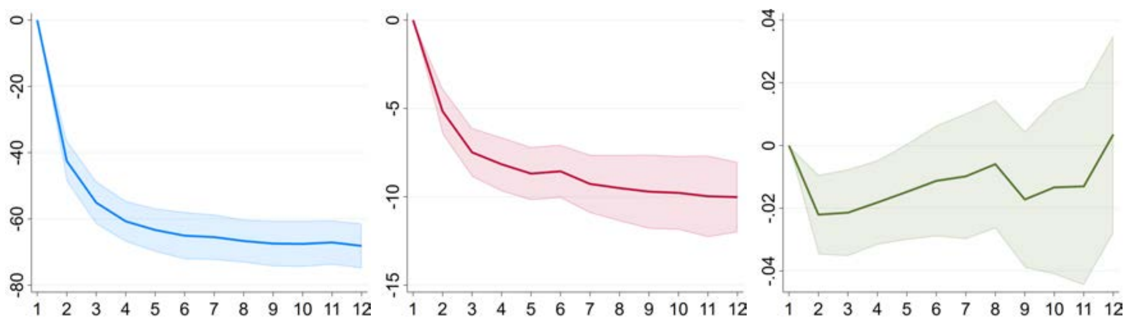



(f) Prop. consistent

Notes: All panels use the "attrition" sample of 143,158 responses. After trimming the top and bottom 5 percent of responses in each tenure group in each month, there are 127,505 and 127,775 observations of one-year- and three-years-ahead forecasts, respectively. "1y" denotes one-year-ahead forecasts, "3y" three-years-ahead forecasts. Panel (a) reports by tenure group the mean (consensus) point forecast and the mean of each household's histogram forecast. The mean of the consensus histogram forecast is shown in the dotted line. Panel (b) reports the mean expected rank probability score (ERPS), with the ERPS of the consensus histogram forecasts in the dotted line. Panel (c) plots the proportion of respondents rounding their point and histogram forecasts. Panel (d) reports by tenure group the proportion of respondents replying to the histogram question in fewer than three bins. Panel (e) plots the proportion of respondents reporting unimodal histogram forecasts as either measured directly via the histogram or having first fitted a generalized beta density to each histogram. Panel (f) reports the proportion of respondents whose point expectations for inflation are "consistent," in the sense of Engelberg et al. (2009), with their histogram forecasts.

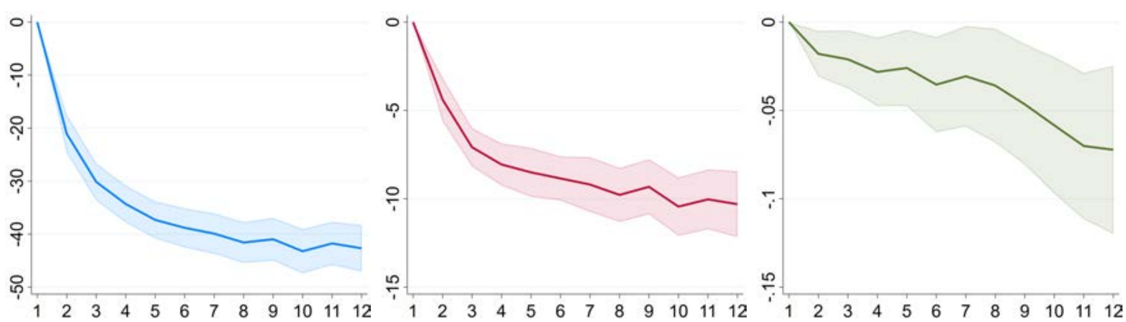
Figure A.5: Learning Effects and Individual Forecast Accuracy by Tenure (Attrition Sample)

(a) One Year Ahead



i) Point Expectations ii) Density Means iii) Density

(b) Three Years Ahead



i) Point Expectations ii) Density Means iii) Density

Notes: Panels show estimates of $\beta_s$ from (2) using the attrition sample. Shaded areas show 95% confidence intervals with Driscoll-Kraay standard errors of lag one. Forecast errors are calculated against first-release year-over-year PCE inflation rates, and reflect the mean squared errors for the point and histogram mean forecasts and the RPS for the histogram forecasts. Underlying sample includes forecasts made between June 2013 and December 2022 (for one-year-ahead forecasts) or June 2021 (for three-years-ahead forecasts).

## A.4 Additional Forecast Accuracy Results

Table A.4: Diebold-Mariano Test Statistics (Consensus Forecasts): Non-Attrition Sample

| | 1-Year-Ahead | | | 3-Years-Ahead | | |
|---|---|---|---|---|---|---|
| | Point | Histogram Mean | RPS | Point | Histogram Mean | RPS |
| Tenure 2 | -6.29*** | -3.66*** | -2.38** | -5.44*** | -2.49** | -2.77*** |
| | (0.000) | (0.000) | (0.019) | (0.000) | (0.014) | (0.007) |
| Tenure 3 | -9.43*** | -4.45*** | -2.64*** | -6.41*** | -3.20*** | -3.49*** |
| | (0.000) | (0.000) | (0.010) | (0.000) | (0.002) | (0.001) |
| Tenure 4 | -9.14*** | -4.65*** | -3.79*** | -6.93*** | -3.47*** | -3.68*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) |
| Tenure 5 | -9.69*** | -5.48*** | -4.31*** | -7.06*** | -3.42*** | -5.40*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) |
| Tenure 6 | -9.44*** | -3.89*** | -3.39*** | -6.82*** | -2.77*** | -4.19*** |
| | (0.000) | (0.000) | (0.001) | (0.000) | (0.007) | (0.000) |
| Tenure 7 | -9.91*** | -4.15*** | -3.01*** | -6.97*** | -3.65*** | -4.15*** |
| | (0.000) | (0.000) | (0.003) | (0.000) | (0.000) | (0.000) |
| Tenure 8 | -10.02*** | -4.79*** | -3.29*** | -6.61*** | -3.32*** | -3.92*** |
| | (0.000) | (0.000) | (0.001) | (0.000) | (0.001) | (0.000) |
| Tenure 9 | -9.77*** | -4.83*** | -4.67*** | -6.80*** | -3.29*** | -4.96*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) |
| Tenure 10 | -9.99*** | -5.29*** | -4.38*** | -6.85*** | -3.01*** | -4.69*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.003) | (0.000) |
| Tenure 11 | -9.38*** | -4.81*** | -4.51*** | -6.08*** | -2.77*** | -5.63*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.007) | (0.000) |
| Tenure 12 | -9.82*** | -4.67*** | -3.31*** | -5.93*** | -2.53** | -5.44*** |
| | (0.000) | (0.000) | (0.001) | (0.000) | (0.013) | (0.000) |

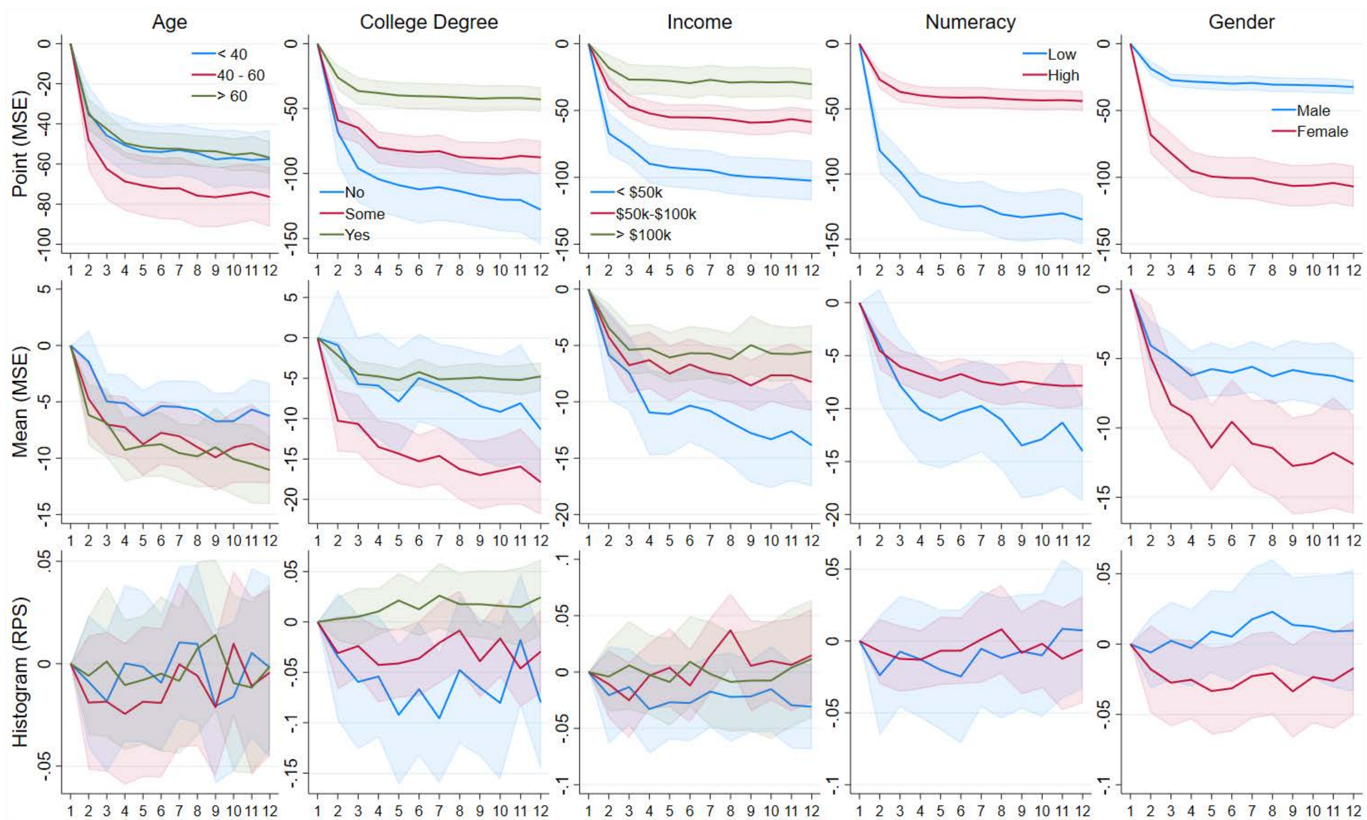
Notes: Forecast errors measured against year-over-year PCE inflation. Shown are robust t-statistics testing for a zero average difference between the monthly consensus forecasts of a given tenure group against the consensus forecasts of new respondents, with their associated $p$ values in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Sample includes forecasts made by respondents in the non-attrition sample between 2013m6 and either 2022m12 (for one-year-ahead forecasts) or 2021m6 (for three-years-ahead forecasts). We repeat this exercise on the attrition sample in Table A.5.

Table A.5: Diebold-Mariano Test Statistics (Consensus Forecasts): Attrition Sample

| | 1-Year-Ahead | | | 3-Years-Ahead | | |
|---|---|---|---|---|---|---|
| | Point | Histogram Mean | RPS | Point | Histogram Mean | RPS |
| Tenure 2 | -8.52*** | -5.87*** | -7.14*** | -6.99*** | -3.41*** | -4.71*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) |
| Tenure 3 | -9.74*** | -7.21*** | -8.76*** | -7.56*** | -4.18*** | -5.00*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 4 | -10.79*** | -7.37*** | -10.70*** | -8.07*** | -5.39*** | -6.70*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 5 | -10.58*** | -7.91*** | -10.55*** | -8.31*** | -4.13*** | -6.24*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 6 | -10.57*** | -6.87*** | -9.97*** | -7.61*** | -3.71*** | -5.66*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 7 | -10.80*** | -7.45*** | -10.33*** | -7.73*** | -4.24*** | -5.47*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 8 | -10.93*** | -7.79*** | -10.25*** | -7.72*** | -4.22*** | -5.11*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 9 | -10.84*** | -7.56*** | -11.99*** | -7.56*** | -3.96*** | -5.91*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 10 | -10.34*** | -6.49*** | -10.79*** | -7.67*** | -3.64*** | -5.92*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 11 | -10.65*** | -6.97*** | -11.57*** | -6.90*** | -3.66*** | -6.72*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Tenure 12 | -10.74*** | -6.70*** | -7.88*** | -6.68*** | -2.73*** | -5.80*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.008) | (0.000) |

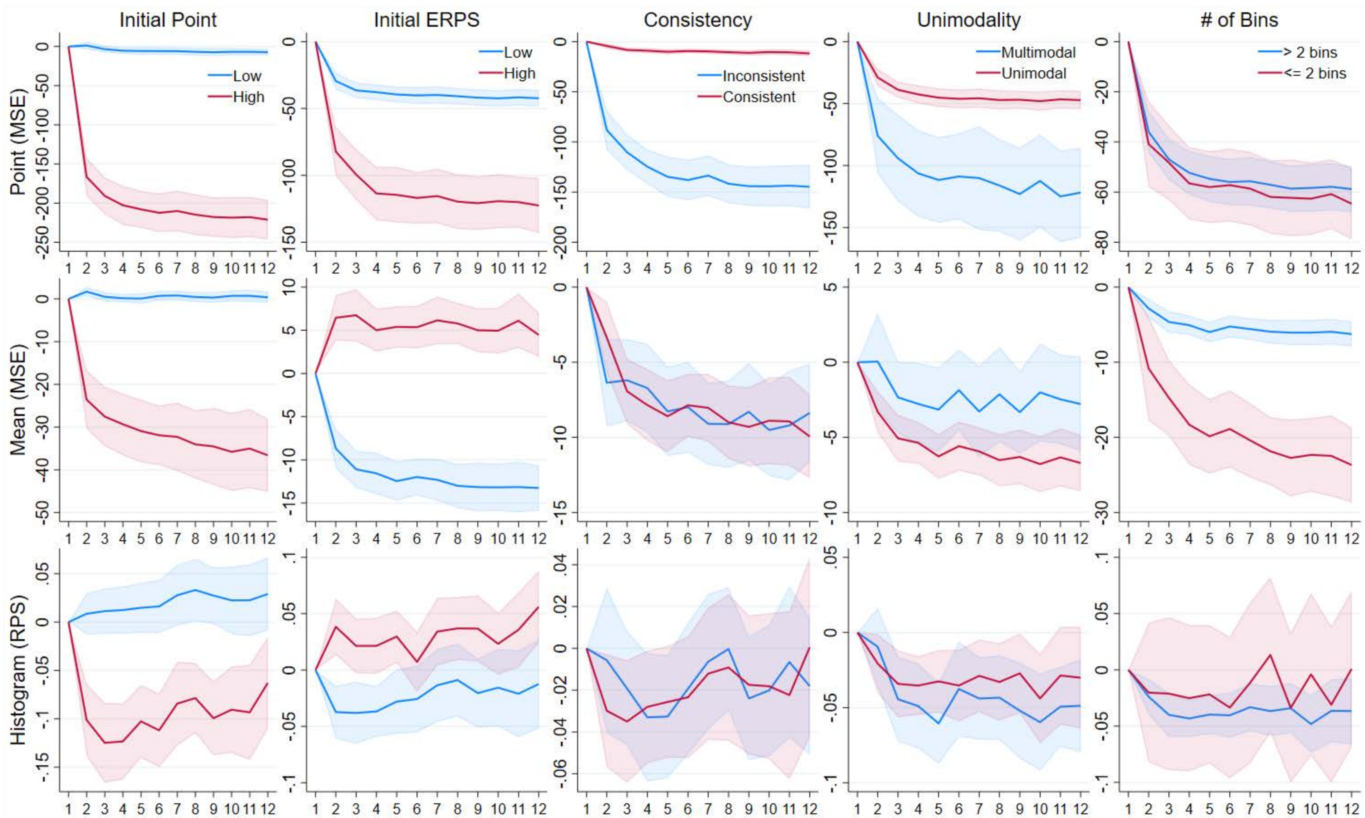
Notes: Forecast errors measured against year-over-year PCE inflation. Shown are robust t-statistics testing for a zero average difference between the monthly consensus forecasts of a given tenure group against the consensus forecasts of new respondents. Sample includes forecasts made by respondents in the attrition sample between June 2013 and either December 2022 (for one-year-ahead forecasts) or June 2021 (for three-years-ahead forecasts). $p$ values are reported in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Figure A.6: Learning Effects in Individual One-Year-Ahead Forecasts by Respondent Demographics



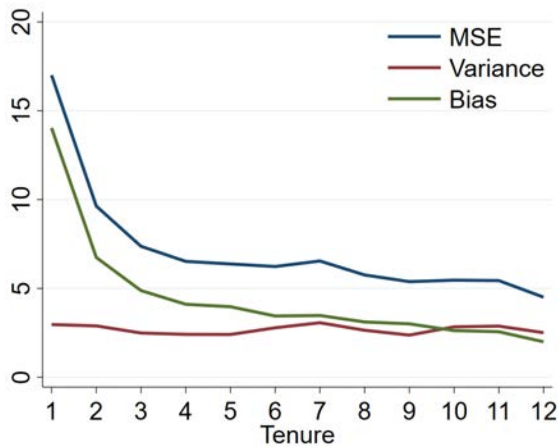

Notes: Estimates of $\beta_s$ from (2) on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and December 2022. The first and second rows show MSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Shaded areas represent 95% confidence intervals using Driscoll-Kraay standard errors of lag 1.

Figure A.7: Learning Effects in Individual One-Year-Ahead Forecasts by Forecast Characteristics

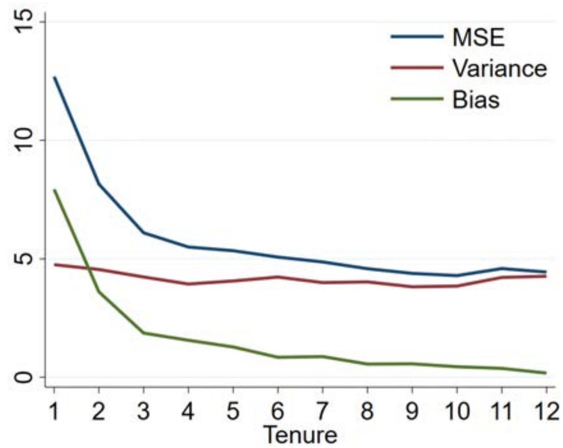



Notes: Estimates of $\beta_s$ from (2) on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and December 2022. See text for group and subgroup definitions. The first and second rows show RMSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Shaded areas represent 95% confidence intervals using Driscoll-Kraay standard errors of lag 1.

Figure A.8: Bias-Variance Decomposition of Mean Squared Forecast Errors



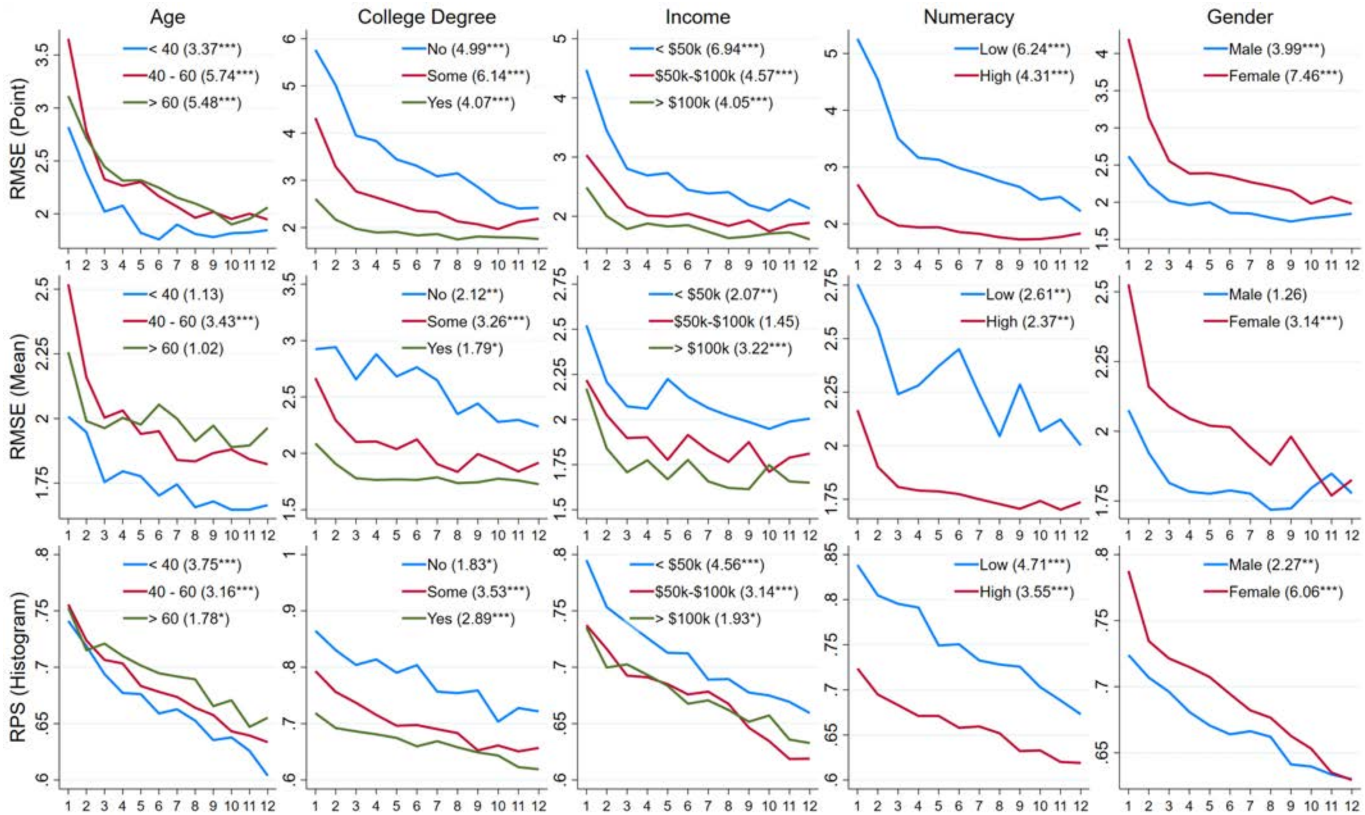

(a) One year ahead




(b) Three years ahead

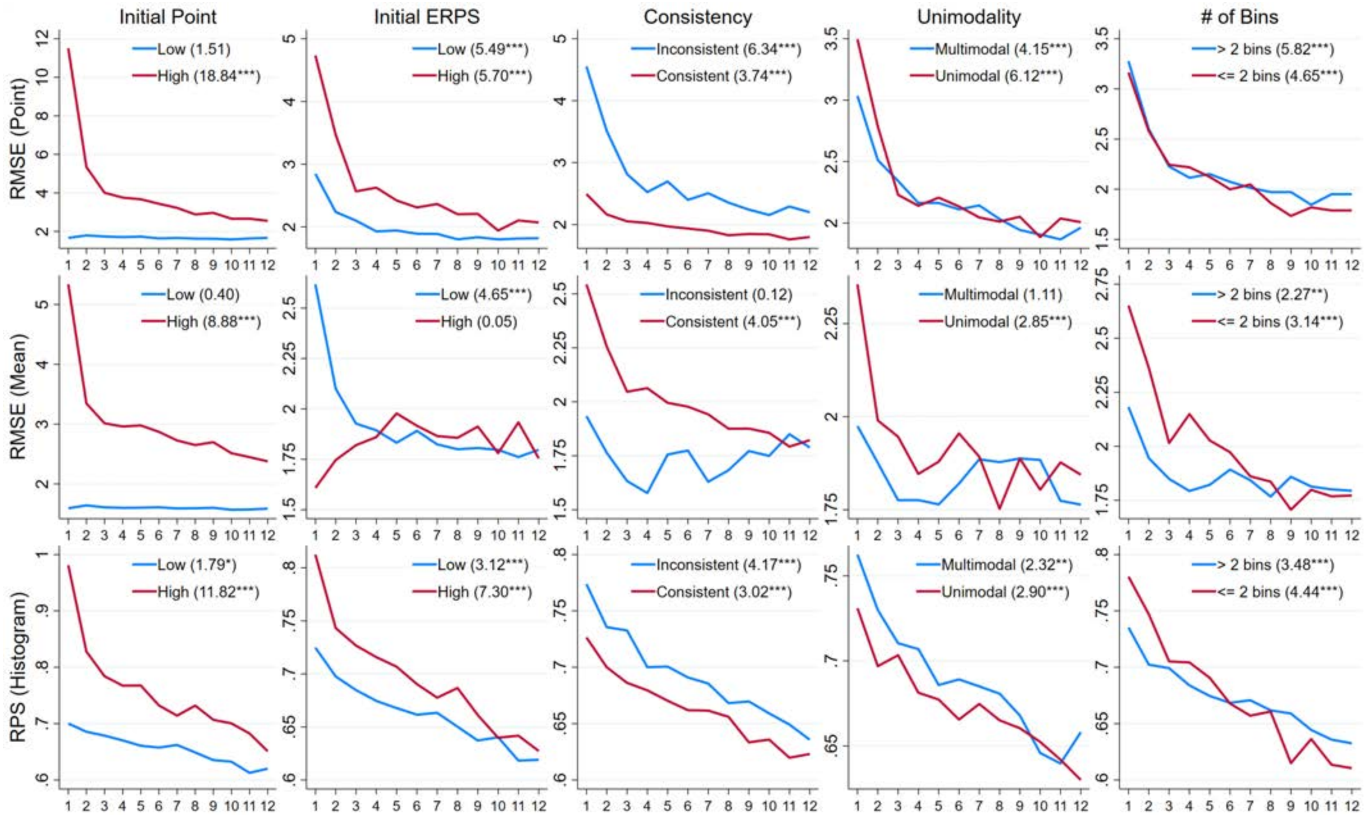
Notes: Calculated on the non-attrition sample.

Figure A.9: Learning Effects in Consensus Three-Years-Ahead Forecasts by Respondent Demographics



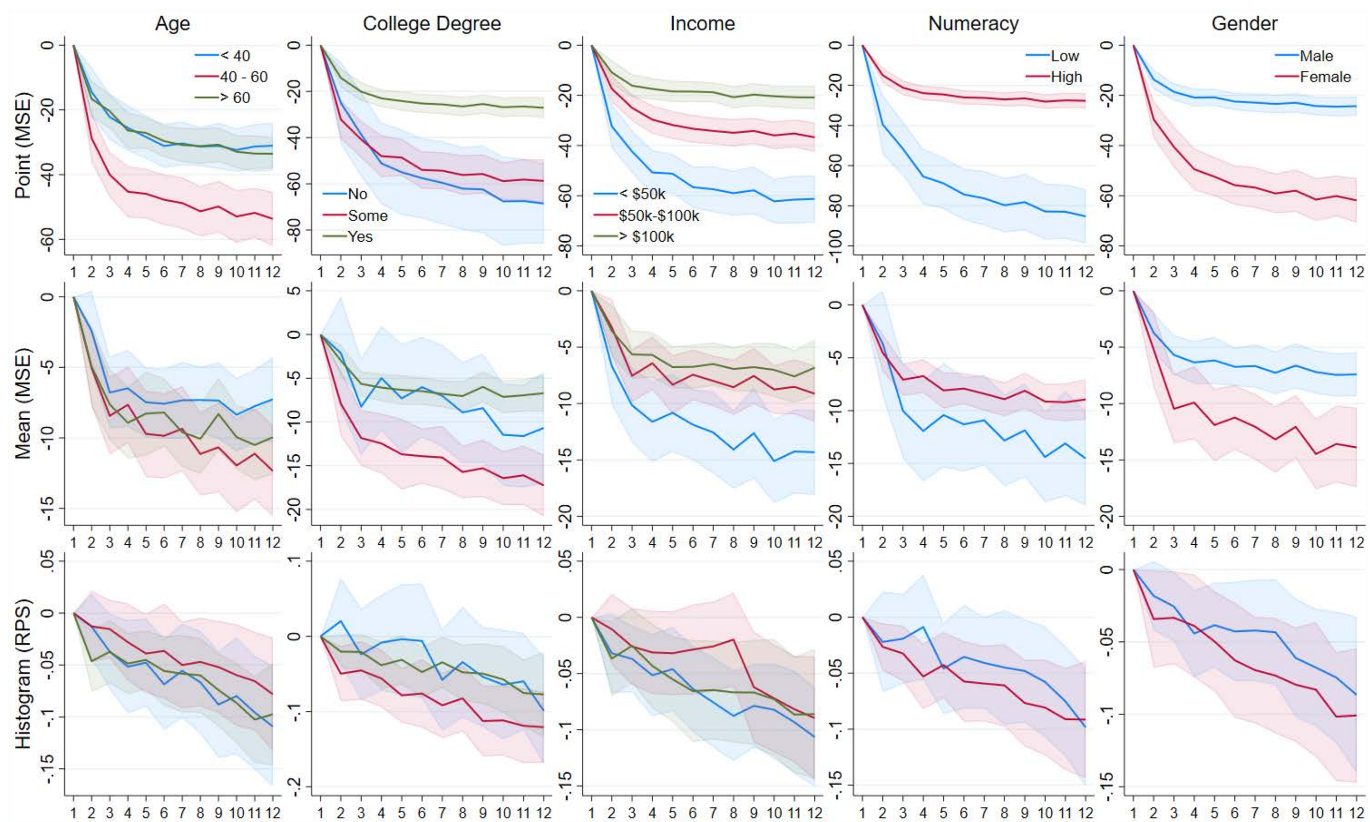

Notes: Calculated on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and June 2021. The first and second rows show RMSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Diebold-Mariano test statistics comparing tenure 12 consensus forecasts against tenure 1's are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.10: Learning Effects in Consensus Three-Years-Ahead Forecasts by Forecast Characteristics



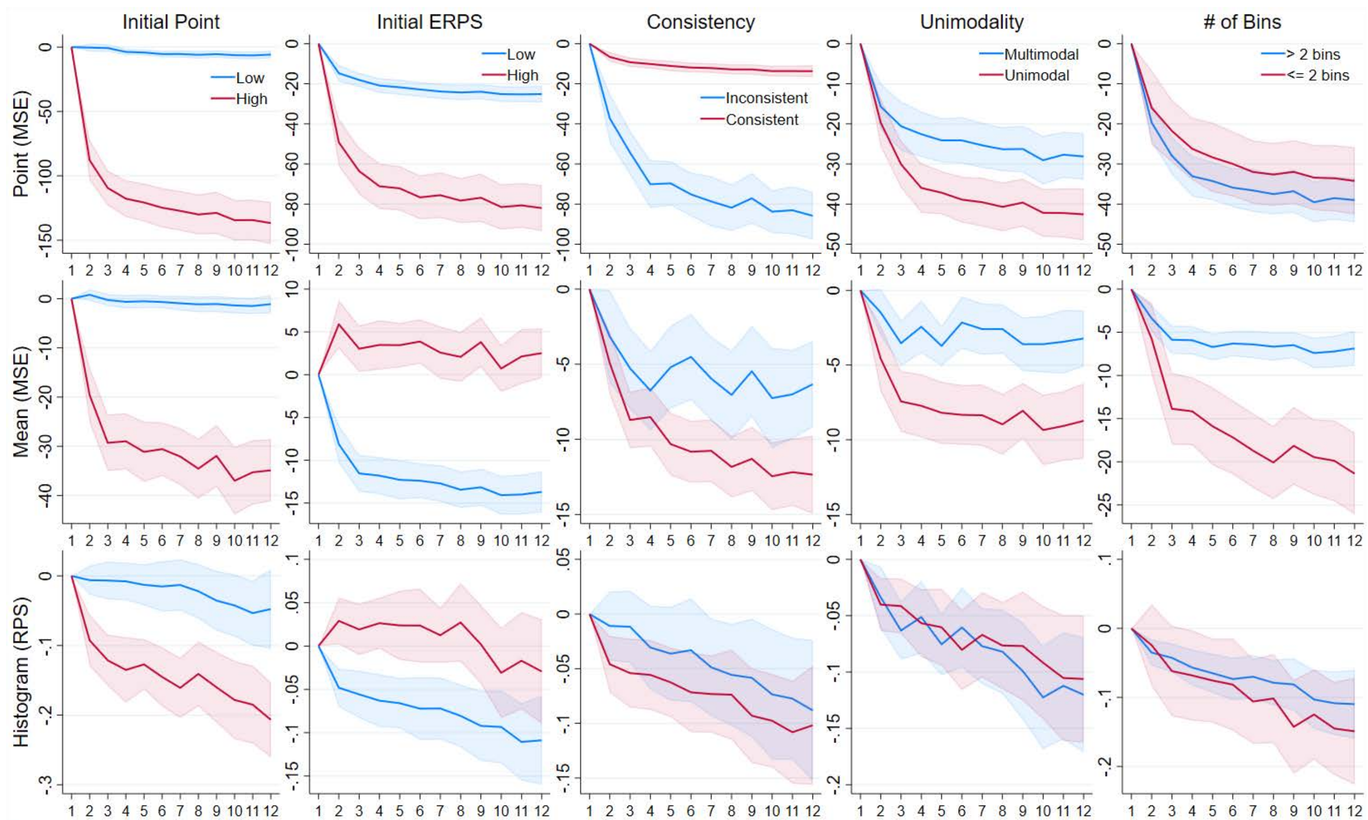

Notes: Calculated on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and June 2021. See text for group and subgroup definitions. The first and second rows show RMSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Diebold-Mariano test statistics comparing tenure 12 consensus forecasts against tenure 1's are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.11: Learning Effects in Individual Three-Years-Ahead Forecasts by Respondent Demographics




Notes: Estimates of $\beta_s$ from (2) on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and June 2021. The first and second rows show MSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Shaded areas represent 95% confidence intervals of Driscoll-Kraay standard errors of lag 1.

Figure A.12: Learning Effects in Individual Three-Years-Ahead Forecasts by Forecast Characteristics



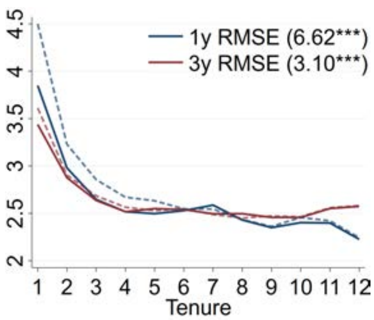

Notes: Estimates of $\beta_s$ from (2) on the non-attrition sample using PCE inflation. Includes forecasts made between June 2013 and June 2021. See text for group and subgroup definitions. The first and second rows show RMSEs of point and histogram mean forecasts, respectively, while the third shows the RPS of the histogram forecast. Shaded areas represent 95% confidence intervals using Driscoll-Kraay standard errors of lag 1.

## A.5 Evaluation Against Realizations of CPI Inflation

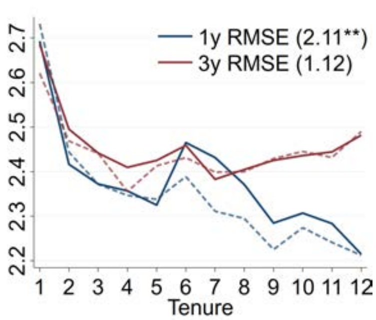
All results in this section calculate forecast errors using CPI rather than PCE inflation, as in the main paper.

Figure A.13: RMSE and RPS of Tenure-Based Consensus Point, Histogram Mean, and Histogram Forecasts Using CPI Inflation Realizations

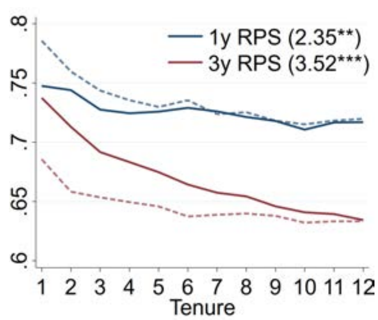
(a) Full Sample




i) Point Expectations

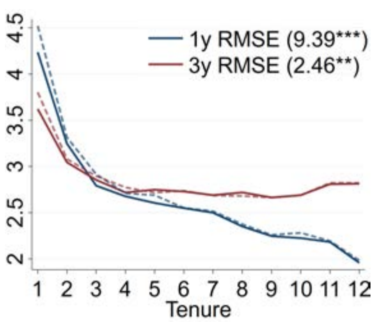



ii) Histogram Means

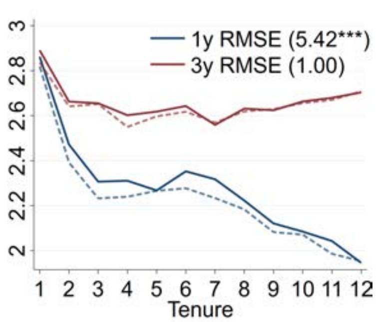



iii) Histogram

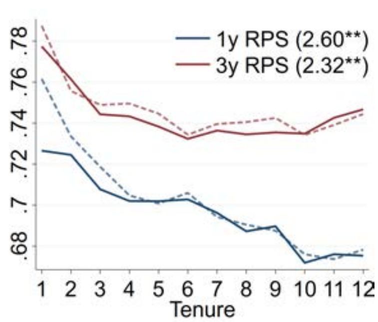
(b) Pre-Pandemic Sample




i) Point Expectations




ii) Histogram Means

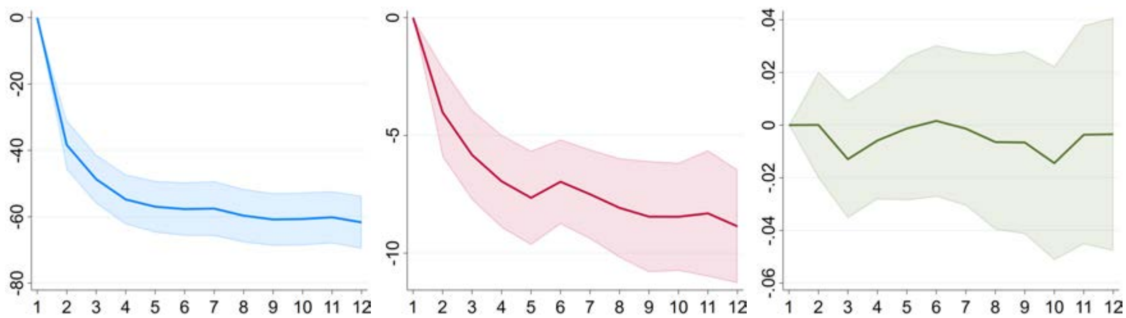



iii) Histogram

Notes: Full-sample evaluates monthly forecasts made from 2013 through 2023. Pre-pandemic sample is monthly from 2013 through 2019m12. Solid and dashed lines indicate non-attrition and attrition samples, respectively. "1y" denotes one-year-ahead, "3y" three-years-ahead. Numbers in parentheses denote Diebold and Mariano (1995) test statistics for equal forecast accuracy between the tenure 1 and tenure 12 forecasts. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

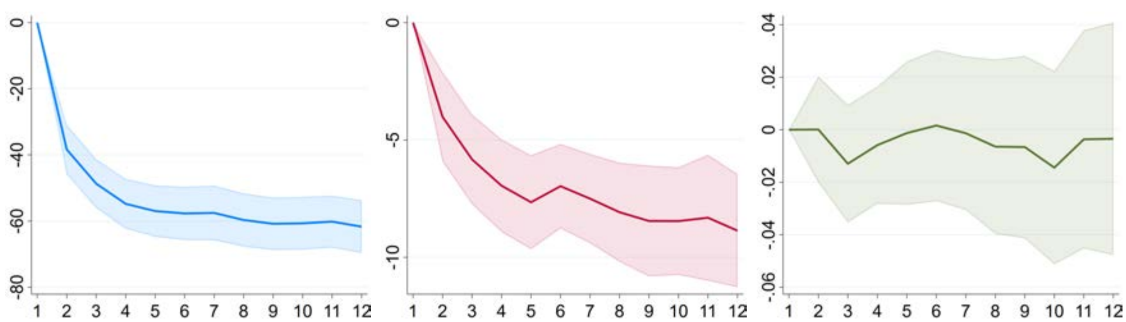
Figure A.14: Learning Effects and Individual Forecast Accuracy by Tenure

(a) One Year Ahead




i) Point Expectations ii) Histogram Means iii) Histogram

(b) Three Years Ahead




i) Point Expectations ii) Histogram Means iii) Histogram

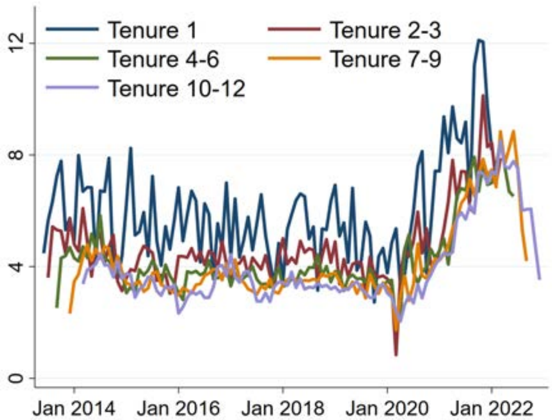
Notes: Panels show estimates of $\beta_s$ from (2) using the non-attrition sample. Shaded areas show 95% confidence intervals with Driscoll-Kraay standard errors of lag one. Forecast errors are calculated against first-release year-over-year CPI inflation rates, and reflect the mean squared errors for the point and histogram mean forecasts and the RPS for the histogram forecasts. Underlying sample includes forecasts made between June 2013 and December 2022 (for one-year-ahead forecasts) or June 2021 (for three-years-ahead forecasts).

## A.6 Temporal Variation

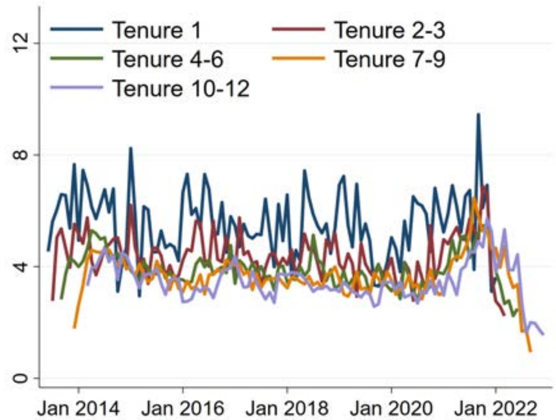
The graphs that follow show the six features of households' expectations shown in Figure 1, but over (not pooled, that is, averaged across) time. See the main paper for further description of how these features are measured and defined.

Figure A.15: Inflation Expectations Over Time, Averaged Across Tenure Groups

(a) Point Expectations

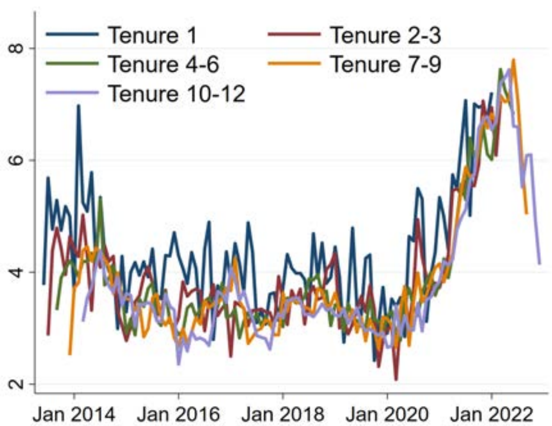



i) One Year Ahead

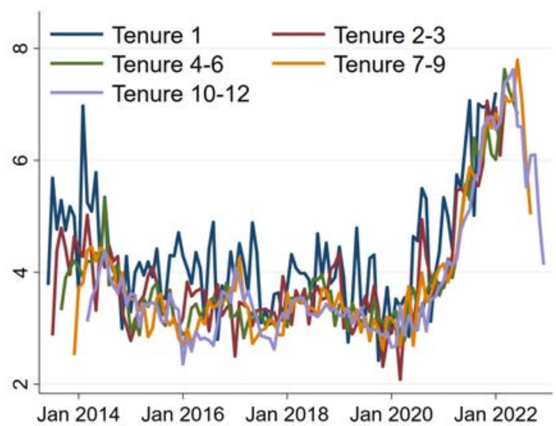



ii) Three Years Ahead

(b) Histogram Means

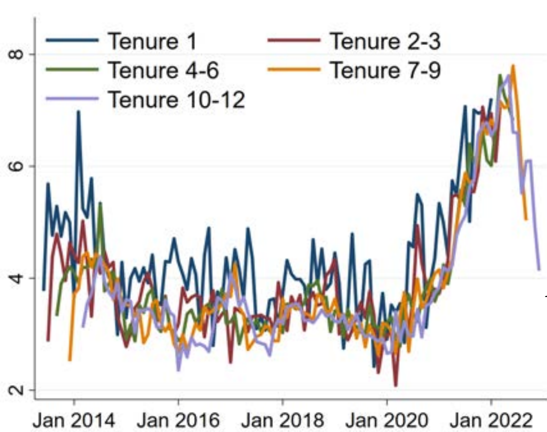



i) One Year Ahead

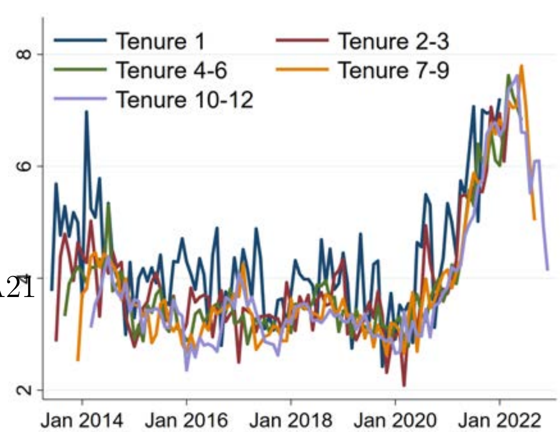



ii) Three Years Ahead

(c) Consensus Histogram Means



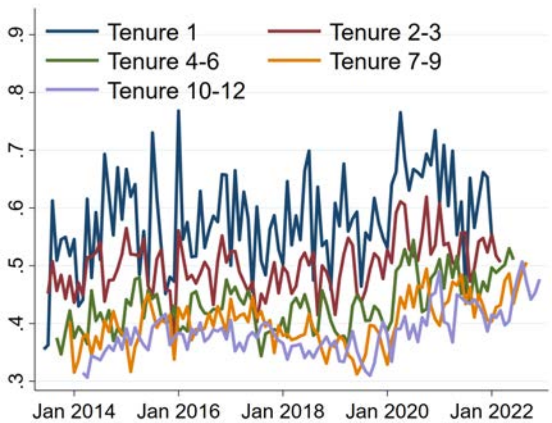

i) One Year Ahead



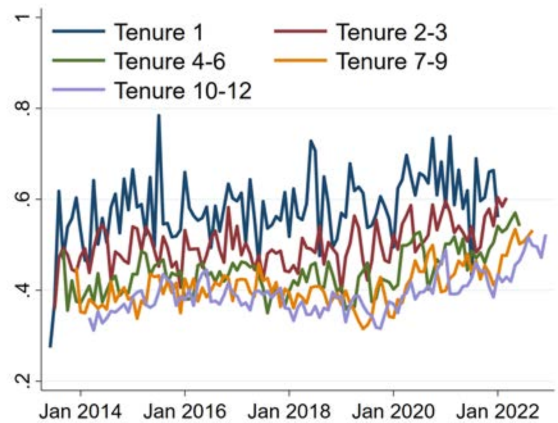

ii) Three Years Ahead

Figure A.16: ERPS Over Time by Tenure

(a) Individual Histograms

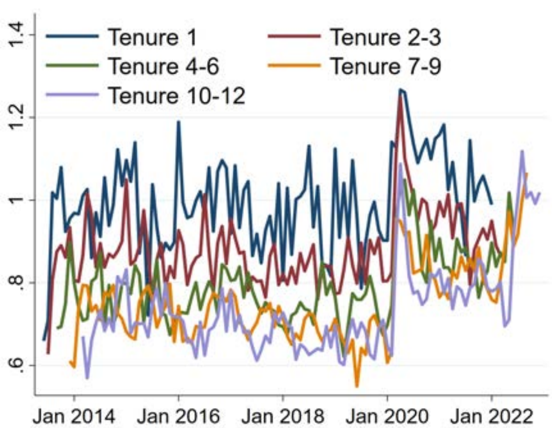



i) One Year Ahead

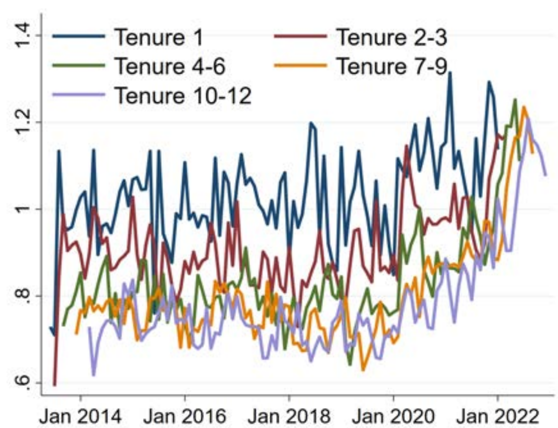



ii) Three Years Ahead

(b) Consensus Histograms



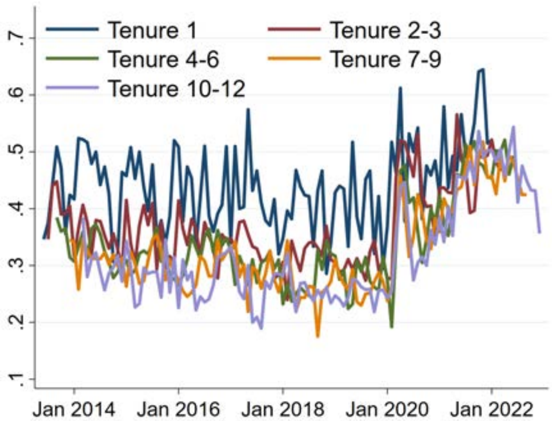

i) One Year Ahead



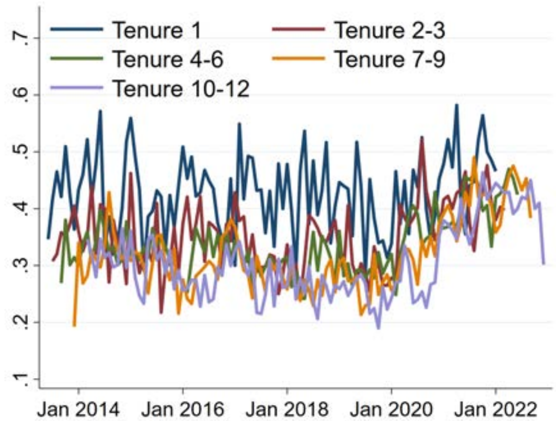

ii) Three Years Ahead

Figure A.17: Rounding Over Time by Tenure

(a) Point Expectations

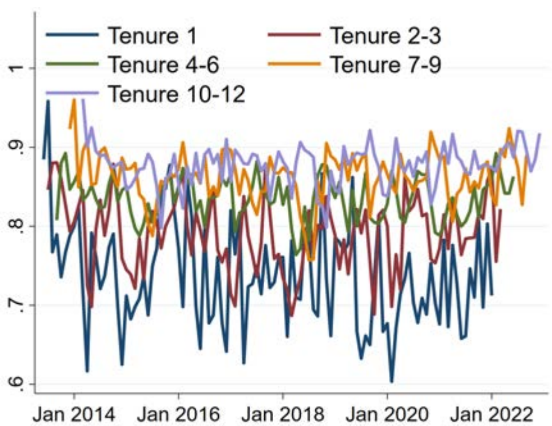



i) One Year Ahead

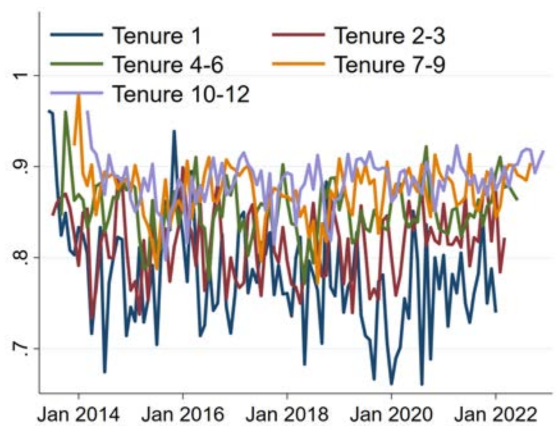



ii) Three Years Ahead

(b) Histogram Bins

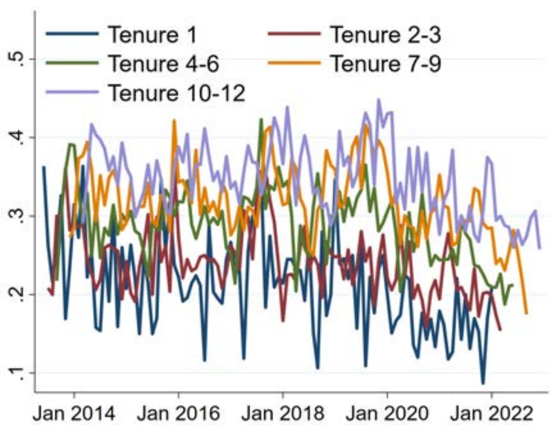



i) One Year Ahead

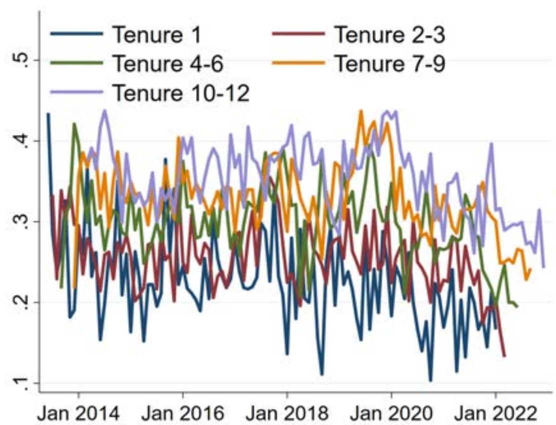



ii) Three Years Ahead

Figure A.18: Use of One or Two Bins Over Time by Tenure



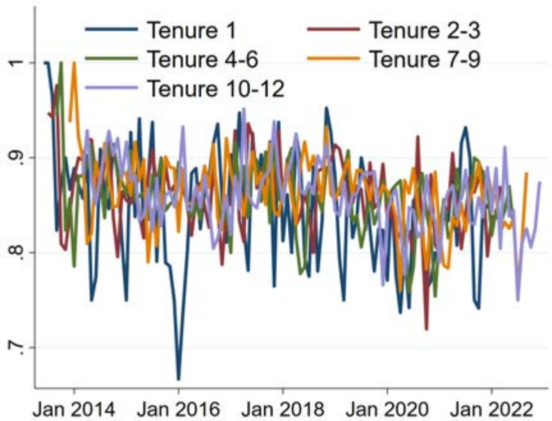

i) One Year Ahead



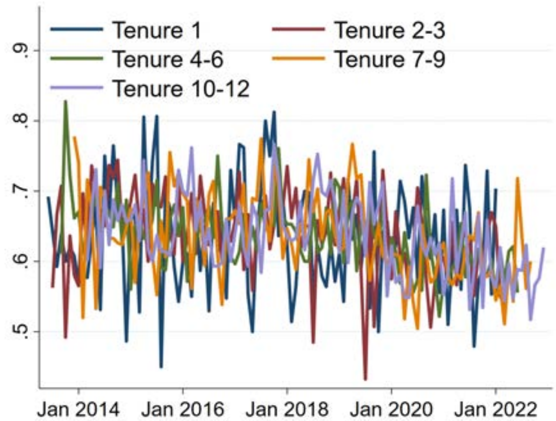

ii) Three Years Ahead

Figure A.19: Unimodality Over Time by Tenure

(a) Histogram

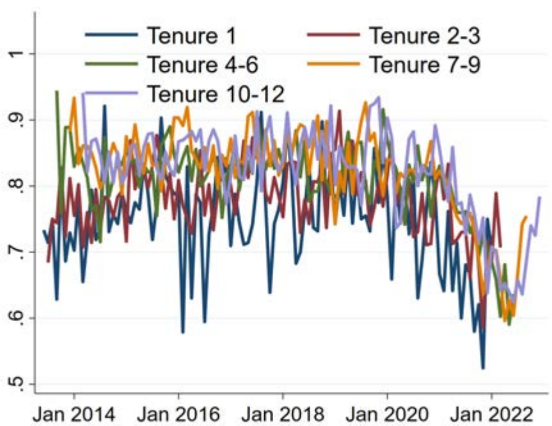



i) One Year Ahead

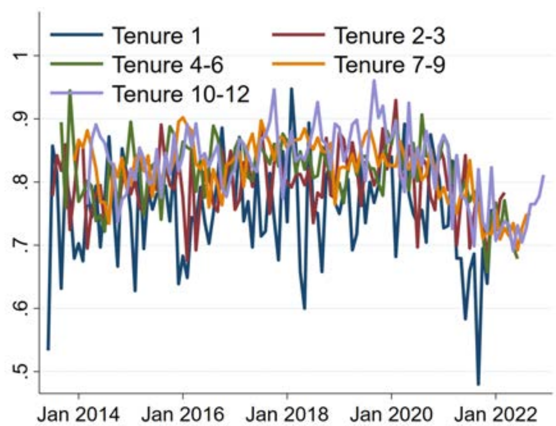



ii) Three Years Ahead

(b) Density

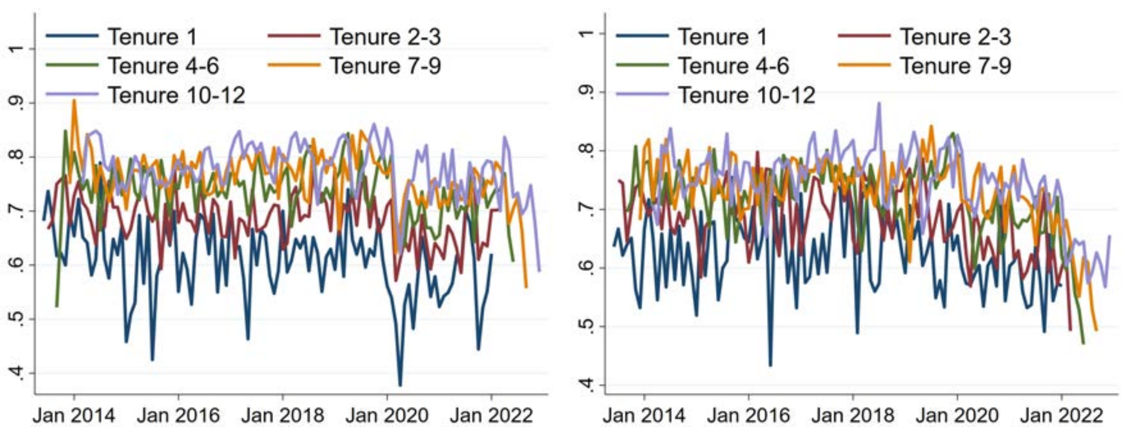



i) One Year Ahead




ii) Three Years Ahead

Figure A.20: Consistency Over Time by Tenure




i) One Year Ahead

ii) Three Years Ahead