

# *DSSR*

Discussion Paper No. 153

**Two Answers, One Belief  
What Household Inflation Expectations Measure**

Wali Ullah

August 19, 2026

**Data Science and Service Research  
Discussion Paper**

---

Center for Data Science and Service Research  
Graduate School of Economic and Management  
Tohoku University  
27-1 Kawauchi, Aobaku  
Sendai 980-8576, JAPAN

# Two Answers, One Belief

## What Household Inflation Expectations Measure

Wali Ullah\*

Department of Economics  
Institute of Business Administration (IBA), Karachi  
E-mail: waliullah@iba.edu.pk

August 19, 2026

### Abstract

The New York Fed Survey of Consumer Expectations asks households for both a point forecast of inflation and a probability distribution over the same horizon. The two responses differ by 3.46 percentage points on average. The analysis shows that much of this gap reflects the artifact of survey design rather than differences in beliefs. The reported standard deviation is mechanically generated by which bins a household ticks for about 42% of respondent-months, and restricting the sample to densities inside the interior bins cuts the average divergence by 89%. What remains is a reporting error rather than a preference, which is two-sided, and its persistence decays to half in about three months. Between 65 and 67 percent of the cross-sectional dispersion in reported point forecasts is measurement error rather than disagreement, and the implied degree of information rigidity is 0.54 using the density means against 0.38 using the point forecasts.

**Keywords:** Inflation expectations; subjective probability distributions; measurement error; rounding; survey design.

**JEL:** C81, D84, E31, E52.

---

\*The author acknowledges the Federal Reserve Bank of New York for making the Survey of Consumer Expectations (SCE) microdata publicly available. The data used in this paper were downloaded from the New York Fed's Center for Microeconomic Data. Source: Survey of Consumer Expectations, Federal Reserve Bank of New York (FRBNY).

# 1 Introduction

Household inflation expectations play an important role in monetary policy. They are used to assess whether expectations are anchored, to capture expectations in Phillips curve model, and to examine whether monetary policy signals are reaching the public (Mankiw et al., 2003; Carroll, 2003; Coibion et al., 2018a; Blinder et al., 2008; Weber et al., 2022). These applications generally take the reported forecast as a direct measure of what households believe. If the survey response differs systematically from the underlying belief, this measurement error can affect the conclusions drawn from the data.

The New York Fed Survey of Consumer Expectations provides a useful setting for studying this issue because it asks households about inflation expectations in two different ways. Respondents first provide a point forecast and then allocate 100 percentage points across ten bins covering the same horizon. The latter responses are used to construct a fitted density and its mean and standard deviation. If both questions capture the same belief, the point forecast and density mean should be close. In the data, however, the two responses differ by 3.46 percentage points on average across 174,013 respondent-months. The gap also increases sharply with uncertainty, from 1.17 points among the most confident respondents to 10.53 points among the least confident.

This difference has usually been interpreted as evidence of behavioral preferences, particularly asymmetric loss. Under such a loss function, the optimal point forecast may differ from the mean of the respondent’s predictive distribution (Granger, 1969; Christoffersen and Diebold, 1997; Newey and Powell, 1987; Elliott et al., 2005; Engelberg et al., 2009).

There is, however, another possibility. The density question uses ten bins with open-ended tails, so the reported mean for respondents using those boundary bins depends partly on how the tails are fitted. In such cases, the difference between the point forecast and density mean may reflect the survey instrument rather than the household’s preferences. This motivates the central question of the paper: how much of the observed divergence is actually generated by the way expectations are elicited?

## 1.1 What the Study Does

The main question is whether the gap between the point forecast and the density mean reflects differences in household beliefs or features of the survey instrument. We address this question in four steps.

- We separate respondents whose reported density remains within the interior of the bin support from those who use an open-ended bin. This allows us to examine whether the point–density gap is linked to the respondent or to the procedure used to construct the density moments.
- Respondents with similar elicited beliefs are compared to identify how the point forecast varies with subjective uncertainty. This approach avoids imposing a parametric relationship between the reported forecast and the underlying belief to avoid the noisy measure of belief.
- Moments in levels and within-respondent autocovariances are used to separate the remaining variation into persistent respondent heterogeneity and reporting error.
- Finally, we examine whether these measurement issues affect aggregate measures of disagreement, forecast accuracy, and commonly used expectations regressions.

We also consider three alternative explanations for the divergence: the impact of open-ended bins, whether point forecasts capture a different distributional feature, and the persistence of the remaining component.

## 1.2 The Results in Brief

The analysis leads to three main findings.

- Reported uncertainty is partly determined by the survey design. For respondents who use only one or two bins, the reported standard deviation is largely determined by the bins they select. In the single-bin case, it equals the bin width divided by  $\sqrt{12}$ . Overall, 42.1% of respondent-months have reported uncertainty that is determined by the selected bins.
- Most of the point–density gap is associated with the open-ended bins. When the reported density remains within the interior of the support, the average divergence falls from 0.891 to 0.099 percentage points, while the corresponding loss-asymmetry measure falls from 0.374 to 0.057. The opposite effects of upper- and lower-tail responses, together with the bin-midpoint counterfactual, indicate that much of the original gap is generated by the procedure used to fit the tails.
- A smaller reporting effect remains after accounting for the bin structure. Holding the elicited belief fixed, a one-unit increase in subjective uncertainty is associated with a 0.186 percentage-point increase in the reported point forecast. The remaining variation is largely explained by reporting error, with rounding accounting for most of the additive component once uncertainty is measured using the unrestricted density responses. The component identified as persistent also declines with the time gap between observations, suggesting persistence over months rather than a permanent respondent characteristic.

These measurement issues also matter for aggregate statistics. Between 65% and 67% of the cross-sectional variation in reported point forecasts appears to reflect reporting components rather than differences in underlying inflation beliefs. The density mean also provides more accurate forecasts of realised inflation than the reported point forecast. Moreover, the estimated degree of informational rigidity changes substantially depending on which measure of expectations is used.

## 1.3 Contribution

The paper contributes to the measurement of household inflation expectations by showing that differences between the point forecast and the density mean cannot be interpreted solely as differences in beliefs or preferences. The survey design itself can generate a substantial part of the observed divergence. The study contributes in three ways.

- **Measurement:** Constructing density moments from interval responses can mechanically determine reported uncertainty for a substantial share of respondents. When a fixed-shape density is fitted over the reported support, its standard deviation is tied to the width of that support. Thus, when respondents place their probability in only one or two bins, much of the variation in measured uncertainty comes from the intervals they choose rather than from differences in the underlying distribution. This feature follows from the survey’s elicitation and fitting procedure.
- **Identification:** We develop a framework that separates this mechanical component from reporting behaviour while accounting for the boundaries of the response bins. This distinction is important when the point-density divergence is normalized by individual uncertainty. The decomposition shows that persistent individual differences account for the smallest component, while the component unrelated to the underlying belief accounts for the largest share.
- **Aggregate Implications:** The measurement problem also has implications for aggregate measures of disagreement and estimates of information rigidity. A non-trivial part of the disagreement commonly attributed to differences in inflation expectations can instead arise from the measurement and reporting process. The estimated degree of information rigidity

also depends on which survey measure is used to construct the consensus, falling from 0.543 to 0.377. The direction of this change is consistent with the attenuation bias implied by the corresponding revision variances.

The estimated relationships should not be interpreted as causal effects. They are conditional correlations, because neither the response format nor subjective uncertainty is randomly assigned. The magnitudes of the reporting components are also survey-specific and depend on the sample and elicitation procedure. The estimated behavioural, transitory, and rounding components should therefore not be expected to carry over unchanged to other surveys.

The broader point is about measurement. When continuous moments of beliefs are constructed from interval responses using a fixed fitting rule, some of the variation in those moments can come from the survey design itself. This needs to be taken into account when interpreting differences between point and density measures of expectations.

## 1.4 Outline

The remainder of the paper is organised as follows. Section 2 reviews the related literature. Section 3 describes the data and the elicitation. Section 4 sets out the framework and the identification results. Section 5 reports the results, taking the restrictions of the framework in the order in which they can be refuted, and then decomposes what remains. Section 6 draws out the consequences for measured disagreement, Section 7 does the same for three aggregate regressions, and Section 8 concludes.

# 2 Related Literature

The literature relevant to this study can be grouped into three strands. The first concerns how household expectations are elicited and measured. The second interprets the gap between point forecasts and probabilistic beliefs as evidence of asymmetric preferences. The third examines reporting error, disagreement, and the use of survey expectations in macroeconomic models. We bring these strands together by asking a more basic measurement question: how much of the observed divergence reflects household beliefs and how much is generated by the survey instrument itself?

## 2.1 What Surveys of Expectations Measure

Household expectations are typically measured in two ways. The traditional approach asks respondents for a single numerical forecast. Surveys such as the Michigan Survey of Consumers and the New York Fed’s Survey of Consumer Expectations have used this format extensively, and their point forecasts form the basis of many aggregate expectations measures (Armantier et al., 2013, 2017; Bryan and Venkatu, 2001).

A second approach asks respondents to describe a probability distribution over possible outcomes. Manski (2004) argued that this provides richer information because it separates beliefs about possible outcomes from the single number a respondent chooses to report. Subsequent research shows that households can answer these questions and that the resulting distributions contain useful information about expectations and behaviour (Manski, 2018; Delavande and Rohwedder, 2011; Bruine de Bruin et al., 2010; Giustinelli et al., 2022). These distributions have also been validated against observed behaviour and other economic characteristics (Armantier et al., 2015; Bachmann et al., 2015; Cavallo et al., 2017; Coibion and Gorodnichenko, 2012).

Our question is different. Rather than asking whether elicited distributions are informative, we ask whether the point forecast and the density mean can be treated as two measures of the same underlying belief. This distinction is important because the two measures are generated by different survey questions and need not contain the same measurement error.

## 2.2 Behavioural Interpretations of the Divergence

The difference between point forecasts and density means is well documented and is usually given a behavioural interpretation. Under asymmetric loss, for example, the optimal point forecast need not equal the mean of the predictive distribution. A respondent who considers one type of forecast error more costly may therefore report a value above or below the mean (Granger, 1969; Newey and Powell, 1987; Christoffersen and Diebold, 1997; Elliott et al., 2005; Patton and Timmermann, 2010; Zellner, 1986).

Engelberg et al. (2009) provided an important method for working with probabilistic survey responses by fitting a continuous density to reported histogram probabilities. Subsequent studies have used this approach to compare density-based measures with point forecasts and to study the role of asymmetric preferences in household expectations (Clements, 2010, 2014; Pesaran and Weale, 2006). Under this interpretation, the divergence between the two measures contains information about the respondent’s preferences, and its average can be used to infer the degree of loss asymmetry.

This interpretation, however, relies on an important measurement assumption that the density mean itself is a reliable measure of the respondent’s underlying belief.

## 2.3 Measurement, Reporting, and This Paper

Several strands of the literature suggest that this assumption should be evaluated carefully. First, survey respondents often round their answers, making reported forecasts partly a feature of the response process rather than only of the underlying belief (Binder, 2017; Manski, 2018; Pfajfar and Žakelj, 2014). Second, reporting behaviour varies across respondents and is related to characteristics such as demographics and financial literacy (Bruine de Bruin et al., 2010; D’Acunto et al., 2021; Malmendier and Nagel, 2016; Kumar et al., 2015; Coibion et al., 2018b). These findings highlight the possibility that differences in survey responses may reflect reporting behaviour as well as differences in beliefs.

A related issue concerns the use of cross-sectional dispersion. Dispersion in reported inflation expectations is commonly interpreted as household disagreement or aggregate uncertainty and is also used to study expectation anchoring (Mankiw et al., 2003; Mankiw and Reis, 2002, 2010; Branch, 2004; Manski and Molinari, 2010). Survey revisions are further used to measure information frictions in models of sticky information and rational inattention (Sims, 2003; Reis, 2006; Maćkowiak and Wiederholt, 2009; Woodford, 2003; Carroll, 2003; Coibion and Gorodnichenko, 2015b,a; Andrade and Le Bihan, 2013; Bordalo et al., 2020; Coibion et al., 2020, 2022, 2018a).

Our contribution is to bring these measurement concerns directly into the comparison of point and density forecasts. We ask whether the divergence between the two responses reflects household preferences, reporting behaviour, or the way the survey converts interval responses into a continuous density. This distinction matters because any measurement error that enters the point forecast or the density mean can also affect aggregate measures of disagreement and the estimated parameters of expectations models.

The central finding is that a substantial part of the divergence is generated by the survey instrument itself. In particular, the treatment of open-ended bins and the density-fitting procedure can create differences between the two measures even when they refer to the same underlying belief. The paper therefore complements the existing literature by shifting the focus from interpreting the divergence to first establishing what the survey responses actually measure.

## 3 Data and Elicitation

### 3.1 The Survey

The public microdata of the Survey of Consumer Expectations (SCE) comprise 186,660 respondent-months on 24,592 individuals over the period 2013m6 to 2025m10. The variable `userid` provides

a stable respondent identifier across survey waves. The SCE follows a rotating-panel design, with respondents typically interviewed monthly for up to twelve months.

We use two survey items. The point forecast, **Q8v2part2**, asks respondents to report their expected inflation or deflation rate over the next twelve months. The density question, **Q9**, asks them to distribute 100 percentage points across ten intervals covering the same horizon. The Bank then uses these responses to construct and publish the mean and variance of a fitted density. Table 1 summarizes the survey items and the measures used in the analysis.

Table 1: Survey Items and Constructed Variable Objects

| Object       | Item                          | Content                                                                                                                                                      |
|--------------|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\pi^e$      | Q8v2part2                     | the point forecast, one number for inflation over the next 12 months                                                                                         |
| $\mu$        | Q9_mean                       | mean of the ten-bin density over the same horizon                                                                                                            |
| $\sigma$     | $\sqrt{\text{Q9\_var}}$       | its standard deviation                                                                                                                                       |
| –            | Q9_bin1–Q9_bin10              | raw bin probabilities; <b>bin1</b> is $\geq +12\%$ and <b>bin10</b> is $\leq -12\%$ . Used to rebuild the moments independently and to score density quality |
| –            | Q9_cent50, Q9_iqr             | density median and interquartile range                                                                                                                       |
| 3-year block | Q9bv2part2, Q9c_mean, Q9c_var | an independent point-plus-density pair covering the 12 months between 24 and 36 months ahead, used for replication                                           |
| Gender       | Q33                           | 1 = female, 2 = male                                                                                                                                         |
| Other        | Q32, Q36, Q47                 | age; education (1–8); household income band (1–11)                                                                                                           |

*Note:* The identifier **userid** and the survey month complete the panel structure.

### 3.2 The Bin Support and the Fitting Rule

The ten bins are, in order, above 12%, 8–12, 4–8, 2–4, 0–2, –2–0, –4––2, –8––4, –12––8, and below –12 %. The first and last bins are open-ended, so the interior support is  $[-12, 12]$ . The eight interior bins have width two percentage points, except for the two bins adjacent to the open tails, which have width four.

The published mean and variance are not calculated directly from the reported bin probabilities. Instead, the Bank fits a continuous density following Engelberg et al. (2009): a generalized beta distribution when three or more bins are used, a symmetric triangular distribution when two bins are used, and a uniform distribution when only one bin is used. This fitting procedure creates two measurement issues that are central to our analysis.

First, when respondents place probability in an open bin, the fitted mean depends on how the unobserved tail is extrapolated beyond the last reported boundary. In our data, 46.3 percent of respondent-months place some probability in an open bin, with an average of 11.2 percentage points allocated to the tails. Thus, for a substantial part of the sample, the gap between the point forecast and the density mean may partly reflect the tail-fitting procedure rather than differences in beliefs. Section 5.1.3 examines this possibility.

Second, even when the density remains within the interior support, the fitted standard deviation can be mechanically determined by the width of the bins used. For respondents who use only one or two bins, the reported uncertainty is therefore partly a consequence of the survey design rather than unrestricted variation in their underlying beliefs. Proposition 5 formalizes this result, while Section 5.1.1 tests it empirically. We examine its implications for the measurement of household uncertainty in Section 5.2.6.

### 3.3 Samples

The baseline sample keeps respondent-months with  $|\pi^e| \leq 30$ ,  $|\mu| \leq 30$ ,  $0.05 \leq \sigma \leq 15$ , and non-missing gender. These restrictions remove extreme observations and incomplete records, leaving 174,013 respondent-months for 23,519 respondents over 149 months. We then define a cleaner interior-support sample by requiring that respondents place no probability in either open-ended bin. This leaves 100,419 respondent-months for 17,044 respondents, or 57.6 percent of the weighted baseline sample. We use the interior-support sample as our preferred specification, while reporting the baseline estimates alongside it to show how much the results change.

This restriction has two advantages over conventional trimming rules. First, it is determined by the structure of the survey itself rather than by an arbitrary cutoff chosen by the researcher. Second, it does not restrict the point forecast, which is important because the point forecast is one of the outcomes whose divergence from the density mean we seek to explain. The restriction therefore avoids selecting directly on the variable of interest. At the same time, because responses in the open tails account for an important part of the observed divergence, Section 5.2.5 examines whether the main findings survive under four alternative sample definitions.

Unless stated otherwise, all estimates use the survey weights, with standard errors clustered at the respondent level.

### 3.4 Constructed Variables

All moments reported in the paper use the survey weights  $\varpi_{it}$ . For any set of observations  $S$ , we define the weighted mean and variance as

$$\widehat{\mathbb{E}}_S[x] = \frac{\sum_S \varpi_{it} x_{it}}{\sum_S \varpi_{it}}, \quad \widehat{\text{var}}_S(x) = \widehat{\mathbb{E}}_S[x^2] - \left(\widehat{\mathbb{E}}_S[x]\right)^2. \quad (1)$$

We construct three sets of variables that are used throughout the analysis.

- **Divergence and subjective uncertainty.** We measure the gap between the reported point forecast and the central tendency of the elicited density in two ways:

$$w_{it} \equiv w_{it}^{\text{pub}} = \pi_{it}^e - \mu_{it}, \quad w_{it}^{\text{mid}} = \pi_{it}^e - \sum_{k=1}^{10} p_{k,it} c_k, \quad (2)$$

where  $p_{k,it}$  is the probability assigned to bin  $k$  and  $\sum_k p_{k,it} = 1$ . The value  $c_k$  is the midpoint of bin  $k$ , with  $c_1 = 14$  and  $c_{10} = -14$  used for the two open-ended bins. The first measure uses the published mean of the fitted density and is our main measure of divergence. The second replaces the fitted mean with the probability-weighted bin midpoint and is used only in Section 5.1.4 to examine the role of the tail-fitting procedure. Subjective uncertainty,  $\sigma_{it}$ , is the published standard deviation of the fitted density.

- **Reporting grid and density characteristics.** The reporting grid is inferred from the point forecast rather than directly observed in the survey. We classify an answer as being on a 10-point grid if it is a multiple of ten, on a 5-point grid if it is not a multiple of ten but is a multiple of five, on a 1-point grid if it is an integer, and on a 0.5-point grid otherwise. In the interior-support sample, these categories account for 6.8, 17.0, 68.8 and 7.4 percent of respondent-months by survey weight. The direction of the classification error should be noted: a respondent using a finer grid may occasionally produce a number that falls on a coarser grid, but the reverse cannot occur. We also construct the first four moments of the elicited density and identify its modal bin directly from the reported bin probabilities, which are used in Section 5.1.5 to evaluate the alternative explanation of the point forecast rather than mean.

- **Realised inflation and forecast performance.** Realised inflation,  $\theta_{t+12}$ , is measured as the twelve-month change in the all-items Consumer Price Index reported by the Bureau of Labor Statistics over the horizon covered by the survey questions. We evaluate forecast performance using the root mean squared error (RMSE), mean absolute error (MAE), and bias:

$$\text{RMSE}(\mathbf{x}) = \left\{ \widehat{\mathbb{E}} \left[ (\mathbf{x}_{it} - \theta_{t+12})^2 \right] \right\}^{1/2}, \quad \text{MAE}(\mathbf{x}) = \widehat{\mathbb{E}}(|\mathbf{x}_{it} - \theta_{t+12}|), \quad \text{Bias}(\mathbf{x}) = \widehat{\mathbb{E}}[\mathbf{x}_{it} - \theta_{t+12}], \quad (3)$$

where  $\widehat{\mathbb{E}}$  is the weighted mean defined in Equation (1), and  $\mathbf{x}$  denotes the candidate forecast measures considered in Tables 20 and 21.

### 3.5 Descriptive Statistics

Table 2 compares the baseline and interior-support samples. The last row reports the average loss asymmetry, measured by the slope from a regression of the divergence on subjective uncertainty through the origin:

$$w_{it} = \bar{h}\sigma_{it} + u_{it}, \quad (4)$$

where the regression is estimated without an intercept and standard errors are clustered at the respondent level. This provides a sample analogue of  $\bar{h} = \mathbb{E}[w]/\mathbb{E}[\sigma]$ , as derived in Proposition 2.

Table 2: Descriptive Statistics on the Two Samples

|                                                   | Baseline      | Interior support |
|---------------------------------------------------|---------------|------------------|
| Respondent-months                                 | 174,013       | 100,419          |
| Respondents                                       | 23,519        | 17,044           |
| Survey months                                     | 149           | 149              |
| Mean own uncertainty $\sigma$                     | 3.065         | 1.289            |
| Mean divergence $w = \pi^e - \mu$                 | 0.891 (0.040) | 0.099 (0.020)    |
| Mean absolute divergence                          | 3.460         | 1.545            |
| Variance of the divergence                        | 42.72         | 11.10            |
| Share with $w < 0$ , per cent                     | 41.6          | 42.7             |
| Share with $w = 0$ exactly, per cent              | 6.5           | 11.1             |
| $\bar{h}$ from $w$ on $\sigma$ through the origin | 0.374 (0.016) | 0.057 (0.018)    |
| Share placing mass in an open bin, per cent       | 46.3          | 0.0              |
| Mean skewness of the elicited density             | 0.160         | 0.281            |
| Mean absolute skewness                            | 0.778         | 0.830            |
| Mean excess kurtosis                              | 0.961         | 0.955            |

*Note:* Moments are computed using Equation (1). The final row estimates Equation (4). Standard errors clustered at the respondent level are reported in parentheses.

The main contrast is clear. Moving from the baseline to the interior-support sample reduces the mean divergence from 0.891 to 0.099 percentage points, an 89 percent decline. The corresponding estimate of  $\bar{h}$  falls from 0.374 to 0.057. The sharp reduction is consistent with the view that much of the original divergence is associated with responses in the open-ended bins rather than with a stable difference between the two expectation measures.

The correction does not eliminate the divergence entirely. Within the interior-support sample, the mean difference remains positive and statistically significant, while the mean absolute divergence is 1.545 percentage points. This remaining variation is the focus of the analysis that follows. The two samples also show substantial non-normality in the elicited densities, with mean absolute skewness of 0.778 and 0.830. This provides motivation for examining alternative interpretations of the point forecast in Section 5.1.5.

## 4 Model and Identification

### 4.1 The Asymmetric-Loss Benchmark

Suppose a respondent has a subjective distribution  $F_i$  over future inflation  $\theta$ , with mean  $\mu_i$  and standard deviation  $\sigma_i$ , but the survey asks them to report only one number. Choosing that number is naturally a decision problem and the optimal reported number depends on how costly it is to be wrong in either direction.

This idea is conceptualized in the model with an asymmetric quadratic loss function in (Christofersen and Diebold, 1997) as:

$$L(e) = ae^2 \mathbf{1}\{e \geq 0\} + be^2 \mathbf{1}\{e < 0\}, \quad e \equiv q - \theta, \quad b \geq a > 0, \quad (5)$$

where  $e$  is the reporting error. Here,  $a$  and  $b$  determine the cost of over- and under-prediction, respectively. We summarize the degree of asymmetry by

$$\kappa_i \equiv \frac{b_i - a_i}{a_i} \geq 0, \quad \text{so that} \quad \frac{b_i}{a_i} = 1 + \kappa_i.$$

A positive  $e$  means that the respondent over-predicts inflation, while a negative  $e$  means that they under-predict it. Since  $b_i \geq a_i$ , the benchmark assumes that under-prediction is at least as costly as over-prediction. The economic intuition is that underestimating inflation can leave a household with insufficient real resources after it has committed to wages, saving, or other nominal decisions, whereas overestimating inflation may primarily result in saving somewhat more than necessary.

**Proposition 1** (Expectile representation). *Under the loss function in Equation (5), the optimal report exists and is unique whenever  $\mathbb{E}_i|\theta| < \infty$ . It satisfies*

$$q_i^* - \mathbb{E}_i[\theta] = \kappa_i \mathbb{E}_i[(\theta - q_i^*)^+]. \quad (6)$$

If  $F_i$  is Gaussian with mean  $\mu_i$  and variance  $\sigma_i^2$ , the optimal report can be written as

$$q_i^* = \mu_i + h(\kappa_i)\sigma_i, \quad h = \kappa\psi(h), \quad \psi(h) = \phi(h) - h\Phi(-h), \quad (7)$$

where  $\phi(h)$  denotes the standard normal density at  $h$ , and  $\Phi(-h)$  denotes the standard normal cumulative distribution function,

$$\Phi(-h) = \int_{-\infty}^{-h} \phi(t) dt.$$

By the symmetry of the standard normal distribution,  $\Phi(-h) = 1 - \Phi(h)$ . The function  $h(\kappa)$  is implicitly defined by  $h = \kappa\psi(h)$  with  $h(0) = 0$  and  $h'(\kappa) > 0$ . Thus, the optimal point forecast differs from the subjective mean  $\mu_i$  by a term proportional to the subjective uncertainty  $\sigma_i$ , with the magnitude of the adjustment determined by  $\kappa_i$ .

The intuition behind is that when under-prediction is more costly, respondents have an incentive to move their forecast above the mean as a form of protection against an unfavorable inflation outcome. The amount of this upward adjustment depends on how much uncertainty surrounds the forecast, measured by  $\sigma_i$ . Under Gaussian beliefs, uncertainty determines the scale of the adjustment, while the number of standard deviations by which the forecast is shifted depends only on the degree of loss asymmetry,  $\kappa$ .

**Corollary 1** (Sign restriction). *Under the loss function in Equation (5), the divergence is necessarily non-negative:*

$$w_{it} = q_i^* - \mathbb{E}_i[\theta] \geq 0.$$

It equals zero only when  $\kappa_i = 0$  or  $\text{var}_i(\theta) = 0$ .

This sign restriction is the key empirical implication of the asymmetric-loss model. Section 5.1.1 shows that it fails for 42.7 percent of the uncensored respondent-month observations. The benchmark can be extended in two ways. One possibility is to allow some households to have  $b_i < a_i$ , implying  $\kappa_i < 0$  and therefore negative divergence. The other is to introduce reporting error, allowing the observed divergence to differ from the underlying optimal report in either direction. With a single cross-section, these explanations are difficult to distinguish. The panel structure provides a way forward: preferences should be relatively stable for a given respondent, whereas reporting errors can vary from one interview to the next. Persistence therefore provides the key empirical distinction between the two mechanisms.

## 4.2 A Three-Component Model of the Reported Forecast

We now allow the reported point forecast to differ from the respondent's subjective mean for three distinct reasons. Focusing on the interior of the bin support, we write

$$\pi_{it}^e = \mu_{it} + h_i \sigma_{it} + \eta_{it} \sigma_{it} + \varepsilon_{it}, \quad (8)$$

where  $h_i = h(\kappa_i)$  captures a respondent-specific component, with mean  $\bar{h}$  and variance  $\sigma_h^2$ . The term  $\eta_{it}$  represents a transitory reporting error whose size varies with subjective uncertainty, with variance  $\omega^2$ , while  $\varepsilon_{it}$  is a scale-independent reporting error with variance  $\varsigma^2$ . We assume that these three components are mutually independent, that the two reporting errors are serially independent, and that  $h_i$  is independent of  $\sigma_{it}$ .

The restriction to the interior support is important. Outside this range, the published  $\mu_{it}$  partly reflects the procedure used to fit the distribution rather than the respondent's reported probabilities alone. In those observations, it is therefore difficult to interpret  $\pi_{it}^e - \mu_{it}$  as a decomposition of respondent behavior. The assumption that  $h_i$  is independent of  $\sigma_{it}$  is also important and relatively strong. As discussed in Section 5.2.2, the model explains the interior data only moderately well, with an  $R^2$  of 0.326 in Equation (9); the estimate of  $\sigma_h^2$  should therefore be interpreted in light of this approximation.

**Proposition 2** (Identification from moments in levels). *Under the model in Equation (8),*

$$\text{var}(w_{it} \mid \sigma_{it} = \sigma) = (\sigma_h^2 + \omega^2)\sigma^2 + \varsigma^2, \quad (9)$$

$$\mathbb{E}[w_{it}w_{i,t+L}] = (\bar{h}^2 + \sigma_h^2)\mathbb{E}[\sigma_{it}\sigma_{i,t+L}], \quad L \geq 1, \quad (10)$$

and

$$\bar{h} = \frac{\mathbb{E}[w_{it}]}{\mathbb{E}[\sigma_{it}]}.$$

Hence, the slope and intercept in Equation (9) identify  $\sigma_h^2 + \omega^2$  and  $\varsigma^2$ , while the slope in Equation (10) identifies  $\bar{h}^2 + \sigma_h^2$ .

The proposition gives a simple way to separate the different sources of divergence. The conditional variance is linear in  $\sigma^2$ , so its slope captures the uncertainty-dependent component while its intercept captures the scale-independent error. The covariance across repeated observations then provides the information needed to distinguish persistent respondent heterogeneity from transitory reporting noise.

Two points are important. First, the moments are written in levels rather than normalized by  $\sigma$ . Dividing by subjective uncertainty would give disproportionate influence to the small number of observations with almost degenerate subjective distributions; Appendix B considers this alternative. Second, Equation (10) holds at different time horizons, which allows us to ask how persistent the respondent-specific component actually is.

**Proposition 3** (The horizon of the persistent component). *Suppose that the respondent component evolves according to  $h_{it} = \varphi h_{i,t-1} + \nu_{it}$ , with  $|\varphi| < 1$ . Then*

$$\mathbb{E}[w_{it}w_{i,t+L}] = (\bar{h}^2 + \varphi^L \sigma_h^2) \mathbb{E}[\sigma_{it}\sigma_{i,t+L}].$$

*Thus, changes in this ratio across  $L$  reveal how quickly the respondent-specific component decays.*

This result is why we report estimates at both short and annual horizons rather than treating persistence as a single fixed quantity. A twelve-month rotating panel cannot tell us whether a persistent component reflects a genuinely fixed preference or a characteristic that simply lasts for several years. Therefore, in Section 5.2.3 we examine the full horizon profile.

**Proposition 4** (Conditional slopes). *Let  $\zeta_{it} = h_i + \eta_{it}$ , so that  $w_{it} = \zeta_{it}\sigma_{it} + \varepsilon_{it}$ . Then*

$$\frac{\partial \mathbb{E}[w \mid \sigma, \zeta > 0]}{\partial \sigma} = \mathbb{E}[\zeta \mid \zeta > 0] > 0, \quad \frac{\partial \mathbb{E}[w \mid \sigma, \zeta < 0]}{\partial \sigma} = \mathbb{E}[\zeta \mid \zeta < 0] < 0.$$

*If  $\zeta$  is approximately symmetric around  $\bar{h}$  and its dispersion is large relative to  $|\bar{h}|$ , the two slopes should be similar in magnitude but opposite in sign.*

This result gives the three reporting states a clear empirical interpretation. Under the asymmetric-loss model,  $\zeta$  is always positive, so the lower state should not occur at all. Under the reporting-error model, both positive and negative states arise naturally. When the transitory component is sufficiently dispersed, the two outer states should also display slopes that are roughly mirror images of one another. This provides a direct and testable distinction between the two explanations.

### 4.3 Bin-Determined Uncertainty

Subjective uncertainty is the key regressor throughout the analysis, and the model in Equation (8) treats it as an observed measure of respondents' beliefs. But the way the density is fitted, as described in Section 3.2, means that this interpretation is not always valid. For some respondents, the reported standard deviation is determined almost entirely by which bins they select.

**Proposition 5** (Bin-determined uncertainty). *Suppose a respondent puts all probability mass in a single bin of width  $\Delta$  and the fitted density is uniform over that bin. Then*

$$\sigma = \frac{\Delta}{\sqrt{12}}.$$

*If the respondent divides the probability equally between two adjacent bins with combined width  $W$ , and the fitted density is symmetric triangular over that range, then*

$$\sigma = \frac{W}{\sqrt{24}}.$$

*In both cases,  $\sigma$  is determined mechanically by the selected bins. It does not depend on the probabilities assigned to those bins, their location within the support, or any other response provided by the respondent.*

The important point is what these formulas imply. For these respondents, the recorded uncertainty contains no information beyond the particular bins selected. More generally, whenever a fixed-shape density is fitted over a reported support, its standard deviation is proportional to the width of that support. The problem is therefore not specific to the particular fitting rule used here; it is a broader feature of this type of elicitation procedure.

Section 5.1.1 empirically tests these implications in the data. Section 5.2.6 then re-estimates the reporting model specified in Equation (8) after excluding the observations for which uncertainty is mechanically pinned down in the manner described above.

## 4.4 Rounding and the Scale-Independent Error

A natural source of the scale-independent error in Equation (8) is simple rounding. Suppose a respondent first forms an optimal forecast  $q^*$  and then reports it using a personal grid with spacing  $\Delta_i$ . The resulting rounding error is defined as  $\varepsilon_{it}$ .

**Proposition 6** (Rounding). *If  $q^*$  is rounded to the nearest multiple of  $\Delta_i$ , the induced error*

$$\varepsilon_{it} = \text{round}_{\Delta_i}(q^*) - q^*$$

*has approximately zero mean and standard deviation  $\Delta_i/\sqrt{12}$ . Because the rounding grid does not depend on  $\sigma_i$ , the error is approximately independent of subjective uncertainty. Consequently,*

$$\varsigma^2 = \frac{\mathbb{E}[\Delta_i^2]}{12},$$

*and respondents using coarser grids should exhibit greater dispersion in  $w$  at every level of  $\sigma$ .*

The intuition is that when  $q^*$  varies smoothly relative to the reporting grid, the rounding remainder is approximately uniform over  $[-\Delta/2, \Delta/2]$ . Its mean is therefore zero and its standard deviation is  $\Delta/\sqrt{12}$ . Most importantly, the rounding process does not depend on the respondent's level of uncertainty.

This proposition gives us two directly testable predictions. The intercept of  $\text{sd}(w \mid \sigma)$  should provide information about  $\varsigma$ , which can be compared with the reporting grids visible in the data. In addition, respondents using wider grids should have systematically larger divergence variance.

## 4.5 Implications for Aggregates

**Proposition 7** (Measured disagreement). *Averaging Equation (8) across respondents within a survey month gives*

$$\bar{\pi}_t^e = \bar{\mu}_t + \bar{h} \bar{\sigma}_t + \text{cov}_i(h_i, \sigma_{it}).$$

*Thus, mean-zero reporting errors disappear from the cross-sectional mean. They do not, however, disappear from the cross-sectional variance, which decomposes as*

$$\text{var}_i(\pi_{it}^e) = \text{var}_i(\mu_{it}) + \text{var}_i(w_{it}) + 2\text{cov}_i(\mu_{it}, w_{it}).$$

*Conditioning on the uncertainty path,  $\text{var}_i(w_{it})$  contains  $\omega^2 \mathbb{E}_i[\sigma_{it}^2] + \varsigma^2$  purely from reporting, so measured dispersion in point forecasts overstates genuine dispersion in beliefs by that amount. The transitory component of  $\text{var}_i(w_{it})$  should therefore be interpreted as measurement noise rather than as genuine dispersion in beliefs.*

The distinction matters when point forecasts are aggregated. Measurement error can leave the average forecast largely unchanged while adding noise to its cross-sectional distribution.

**Proposition 8** (Attenuation in aggregate regressions). *Let  $F_t^\mu$  and  $F_t^{\pi^e}$  denote the cross-sectional means of the elicited density means and reported point forecasts, respectively. Consider regressing an outcome  $y_t$  on the  $k$ -month change  $R_t = F_t - F_{t-k}$  using either series. If the additional variation in  $R_t^{\pi^e}$  relative to  $R_t^\mu$  is unrelated to  $y_t$ , then*

$$\frac{\beta^{\pi^e}}{\beta^\mu} = \frac{\text{var}(R_t^\mu)}{\text{var}(R_t^{\pi^e})}. \quad (11)$$

The implication is standard: additional measurement noise in the point-forecast series increases the variance of the regressor without adding useful information about the outcome, pushing the estimated coefficient toward zero.

**Proposition 9** (Relative accuracy). *Let  $\theta_t$  denote realized inflation. Then*

$$\begin{aligned}\text{MSE}(\pi^e) - \text{MSE}(\mu) &= \mathbb{E}[w^2] + 2\mathbb{E}[(\mu - \theta)w] \\ &= \mathbb{E}[w^2] + 2\text{cov}(\mu - \theta, w) + 2\mathbb{E}[\mu - \theta]\mathbb{E}[w].\end{aligned}$$

*The identity carries the cross-moment; it reduces to the covariance alone when the elicited mean is unbiased for realised inflation or when  $\mathbb{E}[w] = 0$ , neither of which holds in this setup, so we work with the first form (Appendix A). If reporting errors are independent of belief errors, the covariance term becomes  $2\bar{h}\text{cov}(\mu - \theta, \sigma)$ . Since  $\mathbb{E}[w^2 \mid \sigma]$  increases with subjective uncertainty, the difference in forecasting accuracy between the elicited mean and the reported point forecast should become larger as uncertainty rises.*

Together, these results clarify how reporting noise can affect both individual forecasts and aggregate empirical relationships without necessarily changing the average level of reported expectations.

All proofs are provided in Appendix A.

## 5 Results

The results are organized around three questions. Section 5.1 asks whether the divergence can reasonably be viewed as a preference, or whether it can be explained by features of the survey itself. Section 5.2 takes the reporting model that remains after these checks and estimates its three components. Sections 5.3 and 5.4 then ask whether the choice between the two survey measures matters for forecasting outcomes and how broadly the findings can be applied.

### 5.1 The Divergence and Its Sources

The existing literature often interprets the gap between a household’s point forecast and the mean of its reported density as a reflection of uncertainty or household preferences. In particular, under an asymmetric-loss framework, the gap should be non-negative. We therefore examine how the divergence,  $w$ , varies with the respondent’s subjective uncertainty,  $\sigma$ , using both parametric and non-parametric approaches.

#### 5.1.1 Bin-Determined Uncertainty in the Data

All results in this section are based on the interior-support sample, where the reported density does not place probability mass in either open-ended bin. This restriction removes the direct influence of the tail-fitting procedure on the reported mean and variance. We later relax this restriction in Section 5.1.3 to assess how much the boundary observations affect the results.

Under Corollary 1, the asymmetric-loss model implies that the reported point forecast cannot fall below the mean of the underlying belief. In other words,  $w$  should be non-negative. This prediction does not depend on the shape of the belief distribution or on the particular parametric form of the loss function, making it a simple and testable implication of the model.

We begin by examining the empirical distribution of  $w$ . Table 3 reports conditional moments of the divergence for respondents grouped into deciles according to their own subjective uncertainty. Table 3 reveals several patterns that are less visible in the aggregate results.

- **The conditional mean is close to zero.** Across the ten uncertainty deciles, mean divergence ranges from  $-0.190$  to  $0.286$ . Thus, the near-zero overall mean is not simply driven by a few extreme observations; positive and negative divergences occur throughout the distribution.

Table 3: Divergence by Decile of Own Uncertainty

| Decile of $\sigma$ | $n$    | Mean $\sigma$ | Mean $w$ | Mean $ w $ | var( $w$ ) | Share $w < 0$ | Share $w = 0$ |
|--------------------|--------|---------------|----------|------------|------------|---------------|---------------|
| 1                  | 15,115 | 0.561         | +0.201   | 1.182      | 7.48       | 0.312         | 0.277         |
| 2                  | 5,922  | 0.611         | +0.139   | 0.992      | 5.43       | 0.401         | 0.022         |
| 3                  | 16,458 | 0.779         | +0.286   | 1.229      | 8.37       | 0.304         | 0.196         |
| 4                  | 2,688  | 0.871         | +0.104   | 0.748      | 1.95       | 0.453         | 0.000         |
| 5                  | 16,090 | 1.105         | -0.190   | 1.600      | 13.54      | 0.517         | 0.169         |
| 6                  | 4,054  | 1.199         | +0.040   | 1.188      | 6.16       | 0.523         | 0.002         |
| 7                  | 10,045 | 1.364         | +0.111   | 1.209      | 6.65       | 0.482         | 0.013         |
| 8                  | 10,250 | 1.624         | +0.118   | 1.564      | 9.48       | 0.499         | 0.024         |
| 9                  | 9,756  | 2.071         | +0.035   | 1.793      | 11.56      | 0.494         | 0.001         |
| 10                 | 10,041 | 3.017         | +0.161   | 3.245      | 29.39      | 0.449         | 0.000         |

*Note:* The moments calculation is based on the interior-support sample and computed within decile by Equation (1). The deciles are uneven in size because own uncertainty takes a small number of discrete values on this sample, which Table 4 documents.

- **Divergence becomes more dispersed with uncertainty.** Both the mean absolute divergence and its variance increase with  $\sigma$ . This widening is particularly important because the conditional mean remains relatively flat, as shown in Panel (a) of Figure 1. Greater subjective uncertainty therefore appears to be associated mainly with more dispersed reporting rather than with a systematic shift in the mean divergence.
- **Negative divergence becomes less common at higher uncertainty.** Around one-half of respondents have  $\pi^e < \mu$  near the middle of the uncertainty distribution, compared with about 31% at the lower end (Panel (b) of Figure 1). The combination of a nearly flat conditional mean and increasing dispersion is consistent with reporting noise operating around a right-skewed belief distribution.

Two features of the table are less smooth than this broad pattern suggests. Mean divergence is negative only in the fifth decile, while the share of observations with  $w = 0$  falls sharply from 0.277 to 0.022 between the first two deciles before increasing again. These changes reflect the highly discrete distribution of reported uncertainty. Although  $\sigma$  takes 10,403 distinct values in the sample, just three values account for 26.0% of respondent-months.

Tables 4 and 5 examine these mass points in more detail by linking reported uncertainty to the number and configuration of bins used. The three most common values are approximately  $2/\sqrt{12}$ ,  $4/\sqrt{24}$ , and  $4/\sqrt{12}$ . As shown in Table 4, these correspond respectively to a uniform distribution over one two-point bin, a symmetric triangular distribution over two adjacent bins, and a uniform distribution over one four-point bin.

These values are therefore not necessarily measures of genuine differences in household uncertainty. They are, at least in part, mechanically generated by the density-fitting rule and the width of the bins selected by respondents. This discreteness helps explain the sharp movements in the conditional moments of  $w$  and motivates the analysis of reported uncertainty in the next section. The bin-count analysis in Table 5 shows that these predictions hold exactly, rather than only approximately.

- **Single-bin responses.** Among the 17,976 respondent-months in which all probability mass is placed in one bin,  $\sigma$  takes only two values in the sample. Specifically,  $\sigma = \Delta/\sqrt{12}$ , where the bin width  $\Delta$  is either 2 or 4.
- **Two-bin responses.** For 24,266 respondent-months, probability is split evenly between two adjacent bins of equal width. In every such case,  $\sigma = W/\sqrt{24}$ , where  $W$  is the width of the two-bin support.

Table 4: Common Values of Own Uncertainty and Divergence Dynamics

| $\sigma$ | Share | Modal bin pattern                                       | Mean $w$ | $\text{var}(w)$ | Share $w < 0$ | Share $w = 0$ |
|----------|-------|---------------------------------------------------------|----------|-----------------|---------------|---------------|
| 0.577    | 10.73 | all mass in one bin of width two                        | +0.171   | 8.66            | 0.278         | 0.371         |
| 0.816    | 8.05  | half the mass in each of two adjacent bins of width two | +0.366   | 10.83           | 0.229         | 0.351         |
| 1.155    | 7.21  | all mass in one bin of width four                       | -0.409   | 22.28           | 0.521         | 0.310         |

*Note:* The three most common values of own uncertainty in the interior-support sample, their modal bin-probability patterns, and the corresponding divergence are reported. Shares are calculated relative to the full interior-support sample. All moments are computed by Equation (1).

- **Combined implication.** These two groups account for 42.1% of the interior-support sample (45.7% using survey weights), while their reported uncertainty is drawn from only 83 distinct values. For these observations,  $\sigma$  is therefore largely determined by the geometry of the selected bins rather than by differences in underlying beliefs.

Once respondents use three or more bins, the number of distinct values of  $\sigma$  increases sharply: to 1,046 for three-bin responses and to 2,514 and 2,541 for four- and five-bin responses, respectively. With more bins, the fitted generalized beta distribution uses the relative allocation of probability across bins, making the reported standard deviation more informative about the shape and dispersion of the underlying belief. In this sense,  $\sigma$  becomes closer to a continuous measure of subjective uncertainty as respondents provide richer density information.

Table 5: Own Uncertainty by the Number of Bins Used

| Bins used | $n$    | Share | Distinct values of $\sigma$ | Mean $\sigma$ | $\text{var}(w)$ | Predicted exactly |
|-----------|--------|-------|-----------------------------|---------------|-----------------|-------------------|
| 1         | 17,976 | 0.179 | <b>2</b>                    | 0.826         | 14.62           | <b>1.000</b>      |
| 2         | 24,266 | 0.242 | <b>81</b>                   | 0.804         | 8.67            | <b>1.000</b>      |
| 3         | 28,364 | 0.282 | 1,046                       | 1.352         | 6.85            |                   |
| 4         | 16,457 | 0.164 | 2,514                       | 1.751         | 12.14           |                   |
| 5         | 6,752  | 0.067 | 2,541                       | 1.964         | 12.89           |                   |
| 6         | 3,701  | 0.037 | 2,123                       | 2.461         | 17.86           |                   |
| 7         | 1,639  | 0.016 | 1,225                       | 2.821         | 26.97           |                   |
| 8         | 1,264  | 0.013 | 912                         | 3.240         | 27.87           |                   |

*Note:* Calculated using the interior-support sample. The final column indicates the share of respondents in that row for whom the recorded  $\sigma$  equals the exact value predicted by Proposition 5 from the bin widths alone (using a tolerance threshold of one thousandth). All moments are computed by Equation (1).

In the existing literature, household uncertainty, typically measured by  $\sigma^2$ , is widely used to explain the divergence between point forecasts and density means and to motivate behavioural interpretations. Our results suggest that this variable is not always a clean measure of underlying uncertainty. For a substantial share of respondents,  $\sigma^2$  is mechanically determined by the bins selected and the density-fitting rule. It therefore contains measurement error that can affect both the estimated relationship with divergence and its interpretation. Section 5.2.6 addresses this issue by re-estimating the model on the subsample for which Proposition 6 does not bind, and the resulting coefficients change substantially.

The negative mean divergence in the fifth decile of Table 3 provides a useful example. Of the 16,101 respondent-months in this decile, 7,236 place all probability mass in either the 4–8 or 8–12 bin. This produces  $\sigma^2 = 1.155$  (Table 4, third row), with a mean point forecast of 6.354 and

a density mean of 6.764, giving a negative divergence. The remaining 8,865 observations have a positive mean divergence of 0.056 and  $\sigma^2 = 3.61$ . Thus, the negative average in this decile does not appear to reflect a systematic behavioural difference among households. Instead, it is largely driven by the particular bin configurations and the resulting mechanical value of reported uncertainty.

The same mechanism helps explain observations with exactly zero divergence. When a respondent places all probability in a single bin and reports its midpoint as the point forecast, the point forecast and density mean coincide by construction. Such responses are concentrated among respondents using one or two bins and account for roughly one-third of the sample, while this pattern is much less common among respondents providing richer density information.

### 5.1.2 The Sign Restriction

The results in Table 2 for the unconditional moments of  $w$  and certainty based conditional moments of  $w$  in Table 3 show that the divergence is strictly negative for 42.7% of the respondent-months in the interior sample. Furthermore, this share is above 0.44 in the seven out of 10 deciles and remains above 0.30 in all ten deciles of subjective uncertainty. The share of respondents with a negative divergence ranges from 0.229 to 0.521 across the three discrete mass points of uncertainty. The violations are pervasive. They are confined neither to particular regions of the uncertainty distribution nor to particular configurations of bin usage. The restriction is rejected upon the one sample where the elicited mean does not depend upon the choice of tail-fitting procedure, and Section 5.1.3 reports that the conclusion survives the alternative specifications.

The failure of Corollary 1 can arise for two conceptually different reasons:

- **Preference-based explanation.** A household may find overprediction more costly than underprediction. In this case, the loss function implies  $b < a$  and hence  $\kappa < 0$ , so the household's optimal point forecast can lie below its mean belief.
- **Measurement-error explanation.** The reported point forecast may instead contain reporting noise. Unlike a preference-driven wedge, reporting error can be either positive or negative, so it can generate divergence on both sides of zero.

Both mechanisms can therefore produce negative divergence, but they make different predictions about how observations are distributed across the two sides of zero. A single cross-section cannot cleanly distinguish between them. Proposition 5 provides a way to do so by exploiting the different predictions for the lower state,  $w < 0$ .

- **Asymmetric preferences.** If  $\zeta_{it} = h_i + \eta_{it} > 0$ , the asymmetric-loss model predicts no observations in the lower state. In other words,  $w < 0$  should be empty.
- **Reporting error.** If the divergence contains two-sided reporting noise, both  $w > 0$  and  $w < 0$  should be populated. Moreover, after conditioning on subjective uncertainty  $\sigma$ , the relationship should be approximately symmetric across the two sides: the slope for observations with  $w > 0$  should have a corresponding mirror-image slope for  $w < 0$ .

This difference provides a direct empirical test. A sparsely populated or empty lower state would support the preference interpretation, whereas substantial observations on both sides of zero with approximately symmetric patterns in  $\sigma$  would point toward reporting error.

If the second explanation (measurement error) is correct, splitting the data by positive or negative divergence allows to estimate the regime switching framework to compute the conditional slope which helps to distinguish the measurement explanation from preference based explanation. The regime-switching specification is written as follows:

$$w_{it} = \theta_s \sigma_{it} + \delta_t + u_{it} \quad \text{estimated separately on each } \mathcal{S}_s, \quad (12)$$

Table 6: Estimates of State Equations on the Interior-Support Sample

| State                                      | $n$    | Share | Mean $\sigma$ | Slope on $\sigma$ |
|--------------------------------------------|--------|-------|---------------|-------------------|
| Reports above the density mean             | 39,656 | 39.5  | 1.360         | +0.677 (0.040)    |
| Reports the same to within a quarter point | 24,961 | 24.9  | 1.060         | −0.001 (0.002)    |
| Reports below the density mean             | 35,802 | 35.7  | 1.369         | −0.876 (0.051)    |

*Note:* The table presents estimates of Equation (12) using the interior-support sample. Standard errors clustered on the respondent are reported in parentheses.

where the sample partitions are defined as  $\mathcal{S}_+ = \{w_{it} > 0.25\}$ ,  $\mathcal{S}_0 = \{|w_{it}| \leq 0.25\}$ , and  $\mathcal{S}_- = \{w_{it} < -0.25\}$ , and  $\delta_t$  denotes a survey-month fixed effect.

The regime-specific estimates reported in Table 6 suggest that the relationship between divergence and subjective uncertainty is not adequately summarized by a single common slope. Instead, the association changes systematically across regimes. For respondents whose reports exceed their subjective mean, the estimated coefficient is +0.677, while for those reporting below their mean it is −0.876. Thus, the two estimates are close in magnitude but reverse in sign, indicating a pronounced difference in the way divergence is related to uncertainty across the two sides. Panel (c) of Figure 1 presents the corresponding estimates and their confidence intervals.

The similarity in uncertainty across the two outer regimes is also informative. The conditional means are 1.360 and 1.369, respectively, leaving virtually no difference in average subjective uncertainty between respondents classified into the two states. This makes it difficult to interpret the regime-specific pattern as the result of respondents with systematically different levels of uncertainty sorting into one state or the other. In contrast, approximately one-quarter of the observations are concentrated in the central state,  $\mathcal{S}_0$ , where mean subjective uncertainty falls to 1.060. The relatively large mass in this central regime, together with its substantially lower uncertainty, is precisely the type of pattern that can arise when respondents introduce reporting noise around their underlying beliefs.

Taken together, these findings accord with the implications of Proposition 4 for a reporting-error mechanism and pose a challenge to the conventional asymmetric-loss explanation. If asymmetric loss were the sole source of the observed pattern, the lower regime would be expected to be absent rather than comprising 35.7% of the observations. The evidence therefore points toward a two-sided error component in the reported divergence. It does not, however, by itself establish that this heterogeneity is persistent at the respondent level. The same regime structure could emerge from transitory sources of variation that are unrelated to stable individual characteristics. The next two subsections investigate these possibilities in turn.

### 5.1.3 The Bin Support

To decompose the observed divergence into a persistent respondent-specific component ( $h_i$ ), a transitory reporting shock ( $\eta_{it}$ ), and classical measurement error ( $\varepsilon_{it}$ ), we restrict the analysis to observations within the interior support rather than using the full sample. The reason is that the extreme outer bins are subject to an additional complication: the fitting procedure used to construct the published mean ( $\mu_{it}$ ) and uncertainty measure ( $\sigma_{it}$ ) relies partly on extrapolation. At these boundaries, the resulting estimates can become unreliable and may no longer provide an accurate representation of the respondent’s underlying beliefs. Including such observations could therefore introduce an artificial component into the measured divergence.

It is consequently important to determine how much of the observed divergence is mechanically generated by these features of the measurement procedure. This consideration also determines the appropriate estimation sample and provides a direct robustness check for the patterns documented in the existing literature. If the apparent divergence is primarily driven by extrapolation at the tails of the distribution, then restricting the analysis to the interior support should substantially attenuate or eliminate the systematic component of divergence, while leaving the underlying

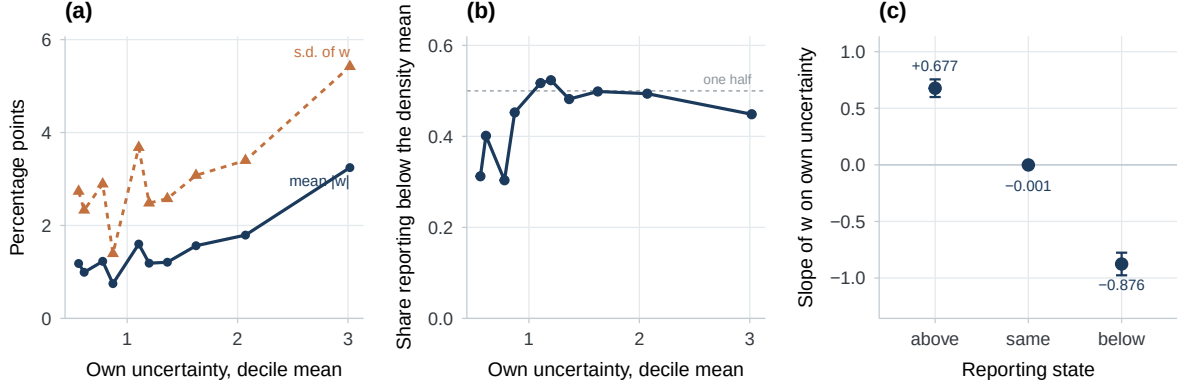


Figure 1: The divergence between the two answers on the interior-support sample. Panel (a) plots the mean absolute divergence and its standard deviation by decile of own uncertainty. Panel (b) plots the share of respondents reporting below their own density mean. Panel (c) plots the slope of the divergence on own uncertainty within each reporting state, with ninety-five per cent intervals.

parameters of the model broadly unchanged.

We examine this issue by separately quantifying the effect of the reporting instrument on divergence ( $w$ ), the contribution of respondents' own uncertainty ( $\sigma_{it}$ ) to divergence, and the magnitude of classical measurement error. To do so, we consider four alternative sample definitions that progressively impose restrictions on the available support. The corresponding estimates are reported in Table 7.

Table 7: Divergence and Estimates Under Successive Sample Restrictions

| Sample                                                   | $n$     | Mean $w$     | $\text{var}(w)$ | Mean $ w $   | Mean $\sigma$ | $\hat{\beta}_\sigma$    | $\hat{\zeta}$ |
|----------------------------------------------------------|---------|--------------|-----------------|--------------|---------------|-------------------------|---------------|
|                                                          |         |              |                 |              |               | by Equ (17)             | by Equ (18)   |
| Baseline                                                 | 174,013 | 0.891        | 42.72           | 3.460        | 3.065         | 0.777<br>(0.024)        | 2.241         |
| No mass in either open bin                               | 100,419 | <b>0.099</b> | <b>11.10</b>    | <b>1.545</b> | <b>1.289</b>  | <b>0.186</b><br>(0.041) | <b>1.862</b>  |
| And point answer within 12                               | 98,780  | -0.001       | 4.49            | 1.219        | 1.279         | 0.180<br>(0.025)        | 1.325         |
| No mass in the four outer bins,<br>point answer within 8 | 71,838  | 0.030        | 2.58            | 1.009        | 1.040         | 0.112<br>(0.029)        | 1.125         |

*Note:* Table reports the divergence and the estimates under successive restrictions to the interior of the bin support. Moments are computed by Equation (1),  $\hat{\beta}_\sigma$  estimates Equation (17) matching the belief to within a quarter of a point, and  $\hat{\zeta}$  estimates Equation (18) at the annual horizon. Standard errors are reported in parentheses.

The estimates reported in Table 7 are broadly stable across the last three sample definitions. We regard the second row as the preferred specification. This restriction is conceptually preferable to the one imposed in the third row because it leaves the point forecast unrestricted. Conditioning directly on an outcome variable whose deviation from the subjective mean is itself the object of interest would otherwise introduce an explicit selection problem. The additional specifications are therefore presented as robustness checks, showing that the main qualitative conclusions do not depend on the precise location of the sample boundary.

Table 7 yields two main findings, while placing an important qualification for a third.

- Excluding observations that place probability mass in the open boundary bins reduces mean

divergence from 0.891 to 0.099. Thus, nearly 89% of the unconditional mean divergence – the central moment used in earlier work to infer substantial loss asymmetry – is generated mechanically by the elicitation boundaries rather than reflecting genuine directional bias in respondents’ beliefs.

- The point estimate on subjective uncertainty,  $\beta_\sigma$ , drops from 0.777 to 0.186 following the boundary restriction, although it remains statistically significant at the 1% level. At the same time,  $\hat{\varsigma}$  declines from 2.241 to 1.862. Removing the boundary observations therefore substantially reduces the dispersion associated with divergence, but it does not eliminate the relationship altogether.

The comparison across rows should nevertheless not be interpreted as a clean quasi-experiment. As the boundary restrictions become progressively tighter, the conditional mean of subjective uncertainty also declines substantially, from 3.065 to 1.289. This matters because assigning probability mass to an open bin is strongly related to subjective uncertainty, with a cross-sectional correlation of 0.645. Hence, part of the decline in  $\text{var}(w)$  across specifications reflects the narrowing of the uncertainty distribution itself, rather than solely the removal of observations affected by tail-fitting problems.

To separate these two channels, we compare respondents within deciles of subjective uncertainty, thereby holding the support of uncertainty approximately fixed. Within these deciles, the variance of divergence is 1.47 times larger among respondents who assign probability mass to open bins, while their mean absolute divergence is 1.15 times larger. The ratios, however, do not increase monotonically across the uncertainty deciles. We therefore place greater weight on the conditional mean of divergence, for which the evidence is considerably stronger, and interpret the evidence on the variance of divergence more cautiously. Panels (a) and (b) of Figure 3 display the corresponding conditional profiles.

Table 8: Conditional Divergence Moments within Deciles of Own Uncertainty

| Decile | Mean $\sigma$ | Observations ( $n$ ) |          | Mean $ w $ |          | $\text{var}(w)$ |          |
|--------|---------------|----------------------|----------|------------|----------|-----------------|----------|
|        |               | Interior             | Boundary | Interior   | Boundary | Interior        | Boundary |
| 2      | 0.754         | 19,493               | 966      | 1.179      | 1.179    | 7.72            | 6.47     |
| 3      | 0.998         | 11,208               | 2,627    | 0.879      | 1.414    | 3.23            | 9.95     |
| 4      | 1.197         | 15,348               | 1,905    | 1.755      | 1.740    | 15.21           | 13.78    |
| 5      | 1.499         | 13,134               | 4,392    | 1.357      | 2.465    | 7.75            | 23.53    |
| 6      | 1.930         | 10,783               | 6,397    | 1.672      | 1.838    | 10.68           | 13.45    |
| 7      | 2.536         | 8,261                | 9,233    | 2.351      | 2.360    | 17.29           | 18.29    |
| 8      | 3.419         | 3,458                | 13,870   | 3.813      | 2.965    | 35.35           | 25.11    |
| 9      | 5.269         | 732                  | 16,661   | 6.439      | 5.883    | 77.18           | 66.21    |

*Note:* Interior denotes respondent-months placing no mass in an open bin and Boundary those placing some. Deciles in which either group has fewer than 200 observations are omitted. This table presents conditional moments calculated via Equation (1), split by the presence of probability mass in open boundary bins. Deciles containing fewer than 200 observations in either partition are omitted for statistical reliability.

The conditional moments details in Table 8 qualify our variance claims, showing that the variance of the divergence is higher for the outer-bin group in only four of the eight uncertainty tiers, and the mean absolute divergence is larger in only three. Controlling for subjective uncertainty reveals that boundary mass allocation does not systematically expand the residual dispersion. The unconditional variance inflation is largely a compositional illusion driven by structural differences in baseline uncertainty across the two groups.

To isolate the mechanical direction of these boundary effects, we estimate the following specification:

$$w_{it} = \gamma_H P_{it}^H + \gamma_L P_{it}^L + \gamma_\sigma \sigma_{it} + \delta_t + u_{it} \quad (13)$$

where  $P_{it}^H$  and  $P_{it}^L$  denote the probability mass allocated to the highest and lowest open boundary bins, respectively, and  $\delta_t$  represents survey-month fixed effects.

As reported in Column 1 of Table 10, mass allocation to the upper boundary bin decreases the divergence by 7.284 points, whereas mass allocation to the lower boundary bin increases it by 5.963 points. Both estimates are statistically significant at the 1% level. The opposite signs of the two coefficients are the relevant feature. The tail-extrapolation rule carries the fitted mean toward whichever open boundary holds the reported mass, and no account based upon household behaviour predicts that the same variable moves the gap in opposite directions according to which boundary is involved.

However, omitting the subjective uncertainty control yields an aggregate boundary coefficient of  $-0.282$  ( $SE = 0.266$ ). Given the opposite signs reported in Column 1, this result is to be expected, because a specification that imposes a single coefficient upon the two tails constrains two effects of opposite sign to be equal. Separating the two boundaries is therefore necessary if the mechanical contribution of the open-ended bins is to be identified.

#### 5.1.4 The Tail-Fitting Rule

The relationship between the divergence gap and boundary mass is clear, but the mechanism behind it requires further identification. We consider two non-mutually exclusive explanations:

- Hypothesis 1 (Behavioral): Respondents expecting inflation to exceed the 12 percent upper bound may process the two elicitation modules differently, generating systematic reporting differences.
- Hypothesis 2 (Mechanical): The parametric assumptions used to extrapolate the open tails may mechanically create a spurious measurement gap.

To distinguish these channels, we conduct a counterfactual exercise that keeps the underlying survey responses unchanged but replaces the aggregation rule used to construct the mean. Under the published procedure,  $\mu_{it}$  is obtained by fitting a continuous parametric density to the ten probability bins. When probability is assigned to an open boundary bin, the fitted tail determines how that probability contributes to the reported mean.

As an alternative, we construct a discrete midpoint measure that assigns each closed bin to its arithmetic midpoint. For the open upper and lower bins, we impose truncation points of 14% and  $-14\%$ , respectively. This provides a transparent benchmark based on the same reported probabilities but without parametric tail extrapolation.

The midpoint measure is not intended to be a superior proxy for latent beliefs, since the  $\pm 14\%$  bounds are arbitrary. Rather, comparing the two measures isolates the portion of the divergence generated by the aggregation procedure itself, holding the underlying survey responses fixed.

The conditional moments reported in Table 9 provide direct evidence that the tail-fitting procedure is the main source of the observed divergence. The comparison is particularly revealing for respondents who place no probability mass in the open boundary bins. For this group, the choice of aggregation method makes virtually no difference: the fitted mean differs from the midpoint-based mean by less than 0.1 percentage points, and mean divergence is only 0.099 under the published procedure compared with 0.006 when the midpoint rule is used.

The picture changes substantially once respondents assign probability mass to one of the boundary categories. In these cases, the two aggregation methods produce markedly different results, with the direction of the difference depending on which boundary is involved.

The divergence is  $+1.379$  for respondents using only the upper boundary bin and  $-2.131$  for those using only the lower boundary bin. In other words, relative to the midpoint benchmark, the fitted mean is systematically drawn toward the open end of the distribution. This is exactly the pattern one would expect when a parametric fitting procedure extrapolates beyond the observed support.

Table 9: Divergence Moments Under Alternative Expectation Aggregation Rules

| Sample Partition       | $n$     | Fitted Mean   | Mean Divergence ( $w$ ) |                      | $\text{var}(w)$ Ratio |
|------------------------|---------|---------------|-------------------------|----------------------|-----------------------|
|                        |         | Less Midpoint | Published               | Midpoint / Published |                       |
| Full Sample            | 174,013 | +0.324        | +0.891                  | +1.215               | 0.96                  |
| Zero Boundary Mass     | 100,419 | −0.092        | +0.099                  | +0.006               | 1.00                  |
| Positive Boundary Mass | 73,594  | +0.806        | +1.809                  | +2.614               | 0.93                  |
| Upper Open Bin Only    | 47,576  | +1.379        | +0.422                  | +1.801               | 0.88                  |
| Lower Open Bin Only    | 2,182   | −2.131        | +0.721                  | −1.410               | 0.80                  |

*Note:* The table uses the baseline sample of 174,013 respondent-months and survey weights, following Equation (1). The midpoint measure replaces the fitted density mean with  $\sum_k p_{k,it} c_k$ , setting  $c_1 = 14$  and  $c_{10} = -14$  for the two open-ended bins. These values are arbitrary; the purpose is to hold the reported probabilities fixed and change only the aggregation rule. The midpoint measure is therefore a diagnostic, not an alternative measure of beliefs.

Fitted Less Midpoint is the published density mean minus the midpoint mean. Mean Divergence reports  $\pi_{it}^e$  minus the published mean and midpoint mean, respectively. The final column is the ratio of the divergence variance under the midpoint rule to that under the published rule; values below one indicate less dispersion under the midpoint rule.

The partitions are not disjoint below the second row. Of the 73,594 respondent-months with probability in an open bin, 47,576 use only the upper bin, 2,182 only the lower bin, and 23,836 use both. The two indented rows therefore do not sum to the row above.

The final column is a ratio of two variances of the divergence, the variance under the midpoint rule in the numerator and the variance under the published rule in the denominator, so an entry below one means the midpoint rule yields the less dispersed divergence.

We next test this mechanism more formally using the regression specification in Equation (13). Rather than relying solely on the unconditional differences in the conditional moments, we estimate the model separately for three related outcomes:

- divergence based on the published mean,
- divergence based on the midpoint mean,
- the direct difference between the two measures of subjective beliefs.

Comparing these specifications allows us to identify whether the relationship is present in the underlying responses themselves or is instead introduced by the procedure used to construct the published subjective mean.

Table 10: Mechanistic Regressions Explaining Core versus Counterfactual Divergence

| Independent Variables           | (1)                           | (2)                          | (3)                          |
|---------------------------------|-------------------------------|------------------------------|------------------------------|
|                                 | Divergence:<br>Published Mean | Divergence:<br>Midpoint Mean | Fitted Less<br>Midpoint Mean |
| Upper Boundary Mass ( $P^H$ )   | −7.284<br>(0.314)             | +0.685<br>(0.282)            | +7.969<br>(0.093)            |
| Lower Boundary Mass ( $P^L$ )   | +5.963<br>(1.090)             | −3.466<br>(1.064)            | −9.429<br>(0.190)            |
| Subjective Uncertainty Controls | Yes                           | Yes                          | Yes                          |
| Survey-Month Fixed Effects      | Yes                           | Yes                          | Yes                          |
| Observations                    | 174,013                       | 174,013                      | 174,013                      |

*Note:* OLS estimates are reported for the full sample of 174,013 respondent-months. The dependent variable is given in the column header. Subjective uncertainty,  $\sigma_{it}$ , and survey-month fixed effects are included in all specifications. Standard errors clustered at the individual respondent level are reported in parentheses.

The regression estimates in Table 10 provide a more direct test of the mechanism and show that the discrepancy is largely generated by the way the reported distribution is converted into a numerical mean. Column 3 is particularly informative because it measures the difference between the official parametric mean and the discrete midpoint benchmark. The estimates imply that the location of the reported probability mass within the open boundary bins has a very large mechanical effect on the resulting mean:

- A one-unit increase in probability mass assigned to the upper boundary bin raises the official mean by 7.969 points relative to the midpoint benchmark.
- The same increase in probability mass assigned to the lower boundary bin lowers the official mean by 9.429 points relative to the midpoint benchmark.

The magnitude and opposite signs of these coefficients leave little room for interpreting the resulting gap as a behavioral asymmetry. Instead, they indicate that the parametric fitting procedure extends the distribution beyond the observed survey support, with the direction of the extrapolation determined mechanically by the boundary at which probability mass is reported.

Columns 1 and 2 show that this mechanical tail extrapolation also affects the estimated parameters of the divergence model. When the official parametric mean is replaced with the counterfactual midpoint-based mean, the coefficient on the upper-boundary indicator changes from  $-7.284$  to  $+0.685$ . Thus, a seemingly substantial coefficient associated with the boundary largely disappears once the numerical integration rule is changed. The estimated coefficient on subjective uncertainty,  $\hat{\beta}_\sigma$ , also moves from  $0.777$  to  $0.895$ , while the estimated respondent component,  $\bar{h}$ , increases from  $0.291$  to  $0.397$ . These changes demonstrate that the choice of aggregation method is not innocuous: it can materially alter the estimated relationships even when the underlying survey responses remain unchanged.

This exercise should not, however, be interpreted as evidence that the discrete midpoint rule provides a better measure of respondents' latent beliefs. In fact, replacing the published mean with the midpoint construction increases average divergence from  $0.891$  to  $1.215$ . The purpose of the exercise is more limited and, we believe, more useful. It shows that the apparent discrepancy between the two elicitation measures cannot be identified from the raw survey responses alone. For the 42% of respondents who place probability mass in the open outer bins, the gap depends critically on how those observations are mapped into a numerical distribution. Part of what appears to be an economically meaningful wedge is therefore generated by the aggregation rule itself rather than by a corresponding difference in household behavior.

Restricting the analysis to the interior support removes this particular source of distortion and provides a cleaner setting for the subsequent analysis. It does not, however, rule out another possible explanation for the remaining divergence. Even when the mechanical effects of the aggregation procedure are removed, respondents may process the two survey questions differently or interpret the elicitation task imperfectly. The next mechanism therefore concerns the behavioral side of reporting rather than the construction of the numerical measure.

### 5.1.5 The Point Answer and the Mode

Proposition 1 assumes that the point forecast can be represented as a linear expectile: the subjective mean plus a term proportional to the standard deviation. The reporting model in Equation (8) then allows for classical reporting noise. Both approaches implicitly assume that respondents are trying to report a measure of central tendency.

But this assumption may not hold. Respondents could instead interpret the question as asking for the most likely outcome, or the mode, of their beliefs. If so, the divergence would not necessarily reflect asymmetric loss or reporting error. It could simply arise because the point forecast and the density question are capturing two different features of the same underlying belief distribution. In that case, the additive error structure in Equation (8) would be misspecified rather than merely noisy.

The evidence so far does not rule out this possibility. A mode-based interpretation can also generate a two-sided divergence that becomes larger as the underlying belief distribution becomes more dispersed. We therefore test this explanation directly.

The elicited density allows us to construct a simple proxy for the mode using the midpoint of the bin receiving the largest probability mass. The difference between this modal midpoint and the published subjective mean provides the divergence predicted by a mode-targeting interpretation. We estimate

$$w_{it} = \lambda(m_{it} - \mu_{it}) + \delta_t + \eta_{b(\mu_{it})} + u_{it}, \quad (14)$$

where  $m_{it}$  is the midpoint of the modal bin,  $\delta_t$  denotes survey-month fixed effects, and  $\eta_{b(\mu_{it})}$  represents fixed effects for bins of the subjective mean. If respondents consistently report the mode, the corresponding restriction is  $\lambda = 1$ .

As a further simplification, we replace the modal gap with the skewness of the elicited distribution:

$$w_{it} = \text{varrho}_{it} + \delta_t + \eta_{b(\mu_{it})} + u_{it}, \quad (15)$$

where  $s_{it}$  is the third standardized moment calculated directly from the reported bin probabilities. A mode-targeting explanation predicts  $\text{varrho} < 0$ : when beliefs are right-skewed, the mode typically lies below the mean, so greater positive skewness should be associated with a larger negative divergence.

Table 11: Evaluating the Mode-Targeting Account Against Empirical Divergence

| <b>Moments and Diagnostics</b>                                                      | <b>(1)<br/>All Densities</b> | <b>(2)<br/>Unimodal</b> | <b>(3)<br/>Interior Support</b> |
|-------------------------------------------------------------------------------------|------------------------------|-------------------------|---------------------------------|
| Respondent-months ( $n$ )                                                           | 152,161                      | 121,677                 | 82,341                          |
| Mean mode gap ( $\mathbb{E}[m_{it} - \mu_{it}]$ )                                   | 1.132                        | 0.162                   | 0.381                           |
| Mean observed divergence ( $\mathbb{E}[w_{it}]$ )                                   | 1.256                        | 0.961                   | 0.147                           |
| Variance ratio ( $\text{var}(m_{it} - \mu_{it})/\text{var}(w_{it})$ )               | 0.354                        | —                       | 0.146                           |
| Cross-sectional correlation ( $\rho$ )                                              | 0.259                        | —                       | 0.099                           |
| <i>Regressions of empirical divergence (<math>w_{it}</math>) on belief features</i> |                              |                         |                                 |
| Mode gap ( $m_{it} - \mu_{it}$ )                                                    | 0.423<br>(0.018)             | —                       | —                               |
| Elicited skewness ( $s_{it}$ )                                                      | −0.193<br>(0.023)            | —                       | —                               |
| Skewness within belief bins                                                         | −0.278<br>(0.021)            | —                       | —                               |
| Mode gap & own uncertainty                                                          | 0.281<br>(0.019)             | —                       | —                               |
| <i>Mean absolute deviation of point answer (<math>\pi_{it}^e</math>) from:</i>      |                              |                         |                                 |
| Elicited Mean ( $\mu_{it}$ )                                                        | 3.508                        | —                       | —                               |
| Elicited Median                                                                     | 3.488                        | —                       | —                               |
| Elicited Mode ( $m_{it}$ )                                                          | 3.401                        | —                       | —                               |

*Note:* The table evaluates predictions from equations Equation (14) and Equation (15). The mode gap is defined as the modal bin midpoint minus the elicited mean. All regression specifications include survey-month fixed effects. The third row regression includes narrow belief-bin fixed effects of width 0.25. Standard errors clustered at the individual respondent level are reported in parentheses.

The estimates in Table 11 do not support the mode-targeting explanation. They do, however, help clarify why the unconditional averages can initially give the impression that the mode and the reported point forecast are closely related.

In the full sample, the average difference between the empirical mode and the parametric mean is 1.132, which is close to the observed divergence gap of 1.256. This similarity disappears once the sample is examined more carefully:

- Unimodal subsample: Among the 121,677 observations with a unique interior mode, the mode-mean difference is only 0.162, while the observed divergence remains much larger at 0.961.
- Interior support: In the preferred interior sample, the mode-based measure actually overstates the observed divergence by roughly a factor of two, with values of 0.381 and 0.147, respectively.
- Residual variation: The mode-mean gap explains only 14.6% of the cross-sectional variance in divergence. Moreover, its correlation with  $w$  is just 0.099, indicating very little association between the two measures.

Taken together, these results leave little support for the idea that respondents systematically report the mode. The mode-mean difference is simply too small, and too weakly related to the observed divergence, to account for the cross-sectional pattern in the data.

Two additional diagnostic checks lead to the same conclusion.

- Point forecasts are, on average, almost equally close to the mean, median, and mode. The corresponding mean absolute distances are 3.508, 3.488, and 3.401, respectively. If respondents were consistently targeting one particular measure of central tendency, that measure should stand out much more clearly.
- The empirical mode performs worse in predicting realized outcomes: within the interior sample, its RMSE is 3.522, compared with 3.266 when the parametric mean is used.

Finally, the unconditional divergence should not be interpreted as a direct measure of behavioral or loss asymmetry. As shown in Section 5.1.3, much of that unconditional gap is generated mechanically by the boundary structure of the elicitation procedure. Once those boundary effects are removed, the remaining divergence is more consistent with reporting heterogeneity than with respondents deliberately optimizing over different statistical functionals. Panel (c) of Figure 3 summarizes these comparative distributions.

## 5.2 Variance Components

Overall, three main conclusions have emerged.

- Divergence occurs on both sides of zero, which is inconsistent with the one-sided prediction implied by the preference-based model in Equation (5).
- Once observations affected by the boundary problem are removed, the remaining variation is not driven by the elicitation boundaries. The divergence therefore appears to reflect meaningful differences across respondents rather than a mechanical feature of the survey design.
- The evidence does not support interpreting the divergence simply as the gap between two different statistical summaries of the same belief distribution. Taken together, these findings favor a reporting-error interpretation over a misspecified-functional explanation.

We therefore retain the representation in Equation (8) for the remainder of the analysis and focus on identifying its three underlying parameters. The baseline estimates are presented in this subsection. All specifications use the interior-support sample defined in Section 3.3; their sensitivity to alternative sample restrictions and filtering rules is examined in Section 5.2.5.

### 5.2.1 The Matched-Belief Estimator

The representation in Equation (8) places a restriction on the conditional mean of the point forecast. Taking expectations conditional on the latent belief and subjective uncertainty, and assuming  $\mathbb{E}[\eta_{it}] = \mathbb{E}[\varepsilon_{it}] = 0$  together with  $\mathbb{E}[h_i | \sigma_{it}] = \bar{h}$ , gives

$$\mathbb{E}[\pi_{it}^e | \mu_{it}, \sigma_{it}] = \mu_{it} + \bar{h}\sigma_{it}. \quad (16)$$

Thus, for two respondents with the same underlying belief, a one-unit increase in subjective uncertainty should raise the point forecast by  $\bar{h}$ . Estimating this slope is useful for two reasons. It provides a direct comparison with the unrestricted OLS specifications commonly used in the literature, while also giving an estimate of  $\bar{h}$  that does not rely on the second-moment restrictions in Proposition 2. It can therefore serve as an independent check on the model.

The main difficulty is that the latent belief is not directly observed. We address this by comparing respondents with similar reported beliefs using the following matching specification:

$$\pi_{it}^e = \beta_\sigma \sigma_{it} + \delta_t + \eta_{b(\mu_{it})} + u_{it}, \quad (17)$$

where  $\delta_t$  denotes survey-month fixed effects and  $\eta_{b(\mu_{it})}$  denotes fixed effects for narrow intervals of the elicited belief. We estimate the model using four different bandwidths.

This approach allows respondents with very similar reported beliefs to be compared without imposing a particular functional form on the relationship between the point forecast and  $\mu_{it}$ . The coefficient on the belief itself is absorbed by the interval fixed effects, avoiding the unit-coefficient restriction imposed by the structural representation.

The approach is not without limitations. Because the sorting variable is measured with error, placing observations into discrete belief intervals can still generate residual grouping bias. Nevertheless, if functional-form misspecification is responsible for the patterns found in conventional specifications, the estimated coefficient should become increasingly stable as the matching intervals become narrower.

Table 12: Non-Parametric Estimates of  $\beta_\sigma$  Across Belief Grid Resolutions

|                                  | (1)              | (2)              | (3)              | (4)              |
|----------------------------------|------------------|------------------|------------------|------------------|
| <b>Elicited Belief Bin Width</b> | <b>1.00</b>      | <b>0.50</b>      | <b>0.25</b>      | <b>0.10</b>      |
| Baseline Estimation Sample       | 0.766<br>(0.023) | 0.774<br>(0.024) | 0.777<br>(0.024) | 0.788<br>(0.024) |
| Interior-Support Sample          | 0.249<br>(0.039) | 0.195<br>(0.040) | 0.186<br>(0.041) | 0.191<br>(0.042) |
| Grid Cells (Interior Support)    | 22               | 42               | 83               | 201              |

*Note:* The table reports OLS estimates of Equation (17) mapping the point forecast  $\pi_{it}^e$  onto subjective uncertainty  $\sigma_{it}$ . All specifications include survey-month fixed effects ( $\delta_t$ ) and alternative sets of narrow belief-bin fixed effects ( $\eta_{b(\mu_{it})}$ ) whose bandwidth configurations are listed in the column headers. Standard errors clustered at the individual respondent level are reported in parentheses.

Table 12 reports the estimates across four progressively narrower matching intervals. In the preferred interior-support sample, the estimated effect of subjective uncertainty is 0.186 (standard error 0.041) when respondents are matched within a quarter of a percentage point of their elicited belief. Narrowing the interval further to one-tenth of a percentage point gives a very similar estimate of 0.191. The fact that the coefficient changes little at these finer resolutions suggests that the matching approach captures the relationship across the underlying belief distribution reasonably well. By comparison, the corresponding estimate in the full sample is 0.777, more than

four times as large. As shown in Section 5.1.3, this difference is attributable to the boundary observations rather than to the underlying relationship. Panel (a) of Figure 2 shows the estimates across bandwidths.

The two matching estimate  $\hat{\beta}_\sigma$  and the unconditional ratio  $\mathbb{E}[w_{it}]/\mathbb{E}[\sigma_{it}]$  from Proposition 2 are designed to capture the same underlying parameter,  $\bar{h}$ . However, within the interior sample, these two methods yield noticeably different results, coming in at 0.186 and 0.076, respectively. The difference is not surprising. The interval fixed effects only control for the elicited belief within a finite range, and subjective uncertainty is positively related to the underlying belief. Consequently, any remaining variation in  $\mu_{it}$  within a cell can be picked up by  $\sigma_{it}$ , pushing  $\hat{\beta}_\sigma$  upward.

This interpretation is consistent with the patterns observed across different bandwidths. As the grid narrows from 1.00 down to 0.25 percentage points, the estimated coefficient drops from 0.249 to 0.186. After that, it stabilizes, shifting only slightly to 0.191 at the narrowest 0.10-point resolution. We therefore view the matching estimate and the unconditional moment ratio as reasonable bounds for  $\bar{h}$ . For the variance decomposition that follows, we use the unconditional ratio because it does not rely on sorting respondents using a noisy proxy and is therefore less vulnerable to within-cell contamination.

### 5.2.2 Decomposition of the Divergence

The first conditional moment identifies only  $\bar{h}$ , so the remaining parameters must be recovered from higher-order moments. We do this using a minimum-distance approach based on Proposition 2. The conditional variance identifies the scale of the uncertainty-related components, while its intercept identifies the unscaled reporting shock. The persistence of respondent-specific heterogeneity is then separated from transitory shocks using the covariance of repeated observations for the same respondent. Together with the first-moment estimate of  $\bar{h}$ , these moments identify the three remaining parameters.

Rather than estimating these relationships directly from the microdata, we match moments calculated within aggregated cells. Specifically, we divide subjective uncertainty into twenty quantile bins, indexed by  $b$ , and estimate

$$\widehat{\text{var}}_b(w) = \varsigma^2 + (\sigma_h^2 + \omega^2)\widehat{\mathbb{E}}_b[\sigma^2] + \xi_b, \quad \text{with weights } \frac{n_b}{\widehat{\text{var}}_b(w)^2}, \quad (18)$$

$$\widehat{\mathbb{E}}_c[w_{it}w_{i,t+L}] = (\bar{h}^2 + \sigma_h^2)\widehat{\mathbb{E}}_c[\sigma_{it}\sigma_{i,t+L}] + \zeta_c, \quad \text{with weights } n_c. \quad (19)$$

Here,  $\widehat{\text{var}}_b$  and  $\widehat{\mathbb{E}}_b$  are the sample moments within uncertainty cell  $b$ . For Equation (19), the cells are formed from twenty quantiles of  $\sigma_{it}\sigma_{i,t+L}$  using respondent pairs separated by exactly  $L$  months; the regression is estimated through the origin. Twenty cells are requested but not always realised. Where few bins are active  $\sigma$  is atomic, since a household placing all of its probability upon one interval has a  $\sigma$  pinned by that interval's width, so quantile cut points coincide and collapse. Upon the interior sample Equation (18) is therefore fitted upon sixteen occupied cells and Equation (19) upon nineteen and twenty at the short and annual horizons; upon the baseline sample all twenty are occupied throughout. The partition is right-closed, and a left-closed one moves blocks of tied  $\sigma$  between cells and returns an intercept of 3.94 in place of 3.47, so a reimplementaion should match the convention. The structural parameters are then recovered from the estimated intercept, the two slopes, and the first-moment restriction  $\bar{h} = \mathbb{E}[w]/\mathbb{E}[\sigma]$ .

The weighting in Equation (18) is important. Conditional variance estimates become much noisier in cells with large variances, particularly in the upper uncertainty quantiles. Without the proposed weighting, these cells would dominate the minimum-distance objective and make the intercept difficult to identify. Standard errors are obtained from a respondent-level block bootstrap with 300 replications.

Column 4 of Table 13 provides our preferred estimates. Three results stand out. First, the additive component accounts for 44.7 per cent of the variance of the divergence, which is close to the largest

Table 13: Minimum Distance Estimates of Structural Variance Components

| Parameter Estimates                                | Baseline Sample |                | Interior-Support Sample |                         |
|----------------------------------------------------|-----------------|----------------|-------------------------|-------------------------|
|                                                    | Short Horizon   | Annual Horizon | Short Horizon           | Annual Horizon          |
| Respondent-months ( $n$ )                          | 174,013         | 174,013        | 100,419                 | 100,419                 |
| Within-person pairs                                | 345,150         | 42,036         | 167,252                 | 11,985                  |
| Slope of $\text{var}(w)$ on $\mathbb{E}[\sigma^2]$ | 1.881           | 1.881          | 1.908                   | 1.908                   |
| Intercept of $\text{var}(w)$ regression            | 5.021           | 5.021          | 3.468                   | 3.468                   |
| First-stage $R^2$                                  | 0.789           | 0.789          | 0.326                   | 0.326                   |
| Slope of paired moment                             | 0.937           | 0.420          | 0.860                   | 0.210                   |
| Moment ratio $\bar{h}$                             | 0.291           | 0.291          | 0.076                   | 0.076                   |
| Uncertainty fixed effect ( $\sigma_h$ )            | 0.923           | 0.579          | 0.924                   | <b>0.451</b><br>(0.100) |
| Transitory shock scale ( $\omega$ )                | 1.014           | 1.243          | 1.026                   | <b>1.305</b>            |
| Unscaled additive error ( $\varsigma$ )            | 2.241           | 2.241          | 1.862                   | <b>1.862</b>            |
| Implied reporting grid ( $\varsigma\sqrt{12}$ )    | 7.76            | 7.76           | 6.45                    | <b>6.45</b>             |
| Persistent variance share (%)                      | 39.4            | 15.5           | 24.8                    | <b>5.9</b>              |
| Scale-dependent share (%)                          | 47.6            | 71.5           | 30.5                    | <b>49.4</b>             |
| Additive floor share (%)                           | 13.0            | 13.0           | 44.7                    | <b>44.7</b>             |
| Total variance explained (%)                       | 90.3            | 90.3           | 69.9                    | <b>69.9</b>             |

*Note:* The table reports structural parameters mapped from the minimum distance conditions Equation (18) and Equation (19). The short-run horizon conditions use panel pairs lagged 1 to 3 months; the annual horizon exploits pairs lagged 9 to 11 months. The first three shares are percentages of the model-implied variance  $\varsigma^2 + (\sigma_h^2 + \omega^2)\mathbb{E}[\sigma^2]$  and sum to 100; the final row reports that model-implied variance as a percentage of the empirical variance of  $w_{it}$ , which is 11.10 in the interior sample. Bootstrap standard errors based on 300 block replications are reported in parentheses for the preferred specification.

of the three components. Its implied scale,  $\hat{\varsigma}\sqrt{12} = 6.45$  percentage points, is also close to the 5.00-point grid that many respondents appear to use. This last comparison is suggestive and we do not rest anything upon it: Section 5.2.7 shows that  $\hat{\varsigma}^2$  is an extrapolated intercept whose value depends on the aggregation rule used to compute it, so the implied grid width inherits that dependence. Under the alternatives reported there the implied width ranges from below zero to about nine points.

Second, persistent respondent-level heterogeneity is relatively small. Although statistically significant, it explains only 5.9% of the variance, with an estimated standard deviation of  $\hat{\sigma}_h = 0.451$  (bootstrap standard error 0.100). Third, the estimated mean parameter  $\bar{h} = 0.076$  implies an asymmetry parameter of 0.211 through the relationship  $h = \kappa\psi(h)$  in Equation (7). This corresponds to under-prediction being about 1.2 times as costly as over-prediction, substantially less asymmetric than the 2.1 ratio implied by the baseline estimate  $\bar{h} = 0.291$ , which maps to  $\kappa = 1.077$ . Panels (b) and (c) of Figure 2 show the model fit and variance decomposition.

The comparisons across columns reinforce the same point. Moving from the baseline sample to the interior-support sample reduces  $\hat{\varsigma}$  from 2.241 to 1.862 and  $\bar{h}$  from 0.291 to 0.076, consistent with the boundary effects documented in Section 5.1.3. Moving from short lags to the annual horizon further reduces persistent heterogeneity, with  $\hat{\sigma}_h$  falling from 0.924 to 0.451, while the scale-dependent component's share rises from 30.5% to 49.4%.

Finally, the maintained components explain 69.9% of the observed divergence variance in the interior sample, compared with 90.3% in the baseline sample. The affine specification also fits the uncensored data considerably better ( $R^2 = 0.789$ ) than the interior sample ( $R^2 = 0.326$ ). Thus, after removing the boundary-driven component, roughly two-thirds of the remaining variation is captured by the model, leaving a substantial share to other sources not included in the maintained structure.

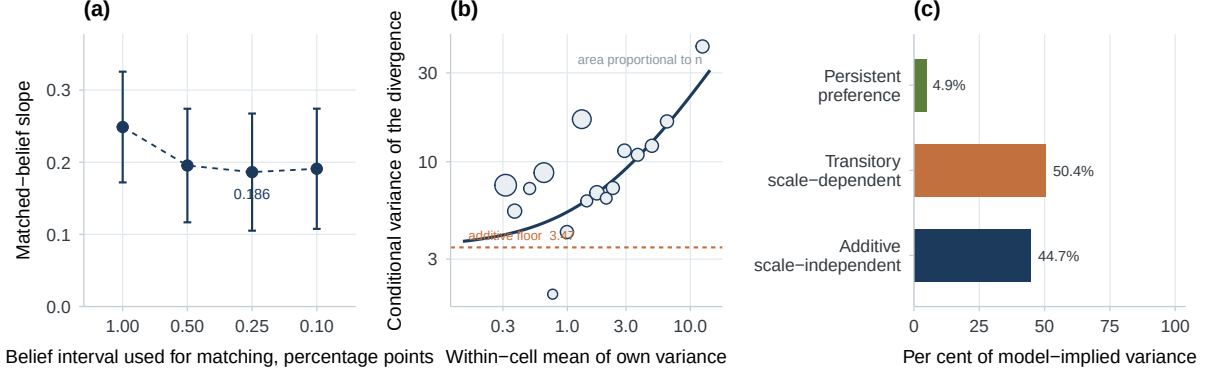


Figure 2: Identification Strategy and Structural Variance Components on the Interior-Support Sample. Panel (a) plots the non-parametric matching coefficient  $\hat{\beta}_\sigma$  against the underlying belief bin resolution alongside 95% confidence intervals. Panel (b) displays the conditional variance of empirical divergence against the within-cell mean of  $\sigma^2$  on logarithmic axes, showcasing the estimated affine function and the additive structural floor. Panel (c) charts the final percentage decomposition across the estimated structural variance shares.

### 5.2.3 The Horizon of the Persistent Component

The previous subsection shows that  $\hat{\sigma}_h$  changes substantially across horizons, falling by about half. This is not simply a matter of statistical robustness; it is precisely what Proposition 3 predicts. If the divergence were mainly driven by a stable loss preference, its effect should not depend on how far apart two interviews are. A systematic decline with the lag instead points to a component that fades over time, such as serially correlated reporting noise. This distinction is important for interpreting the estimates in Table 13, so rather than averaging across horizons, we examine the autocovariance separately at each available lag.

For each lag  $L$ , we use the within-respondent cross-moment

$$R(L) \equiv \frac{\widehat{\mathbb{E}}[w_{it}w_{i,t+L}]}{\widehat{\mathbb{E}}[\sigma_{it}\sigma_{i,t+L}]} = \bar{h}^2 + \varphi^L \sigma_h^2. \quad (20)$$

Here, both moments are calculated using respondent pairs separated by exactly  $L$  months. Under Proposition 3, the resulting profile has a simple interpretation: a flat  $R(L)$  suggests a persistent respondent characteristic, whereas a declining  $R(L)$  indicates that part of the divergence is temporary. Table 14 reports this ratio for every horizon available in the twelve-month rotating panel.

Table 14: Non-Parametric Estimates of the Structural Ratio  $R(L)$  Across Lag Horizons

| Lag Horizon ( $L$ , months)    | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    |
|--------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Panel Pairs ( $\times 1,000$ ) | 71    | 54    | 42    | 33    | 26    | 19    | 14    | 10    | 7     | 4     | 2     |
| Estimated Ratio $R(L)$         | 1.197 | 0.907 | 0.617 | 0.476 | 0.516 | 0.370 | 0.349 | 0.338 | 0.240 | 0.070 | 0.156 |

*Note:* The table reports the empirical ratio  $R(L)$  from Equation (20) estimated on the interior-support sample. Each cell utilizes within-respondent panel observations separated by exactly  $L$  months. Total panel pair counts are listed in thousands.

Table 14 shows a clear decline in the ratio as the time gap between observations increases. It falls from 1.197 at the one-month horizon to 0.156 at eleven months. The pattern is not perfectly monotonic, with somewhat increases at the fifth and eleventh lags, where relatively few within-respondent pairs are available. However, the overall trend is strong as the fitted slope is  $-0.092$  per month (standard error 0.013), shown in Panel (d) of Figure 3.

This pattern changes how the short-run persistence should be interpreted. Much of the variation attributed to stable respondent heterogeneity at short horizons appears to be temporary and serially correlated reporting behavior. Moving from a one-month to an eleven-month horizon reduces  $\hat{\sigma}_h$  from 0.924 to 0.451, indicating that the persistent component is far from fixed over time.

The eleven-month estimate should therefore be viewed as a conservative upper bound on genuinely stable respondent heterogeneity. Any preference component that persists over several years cannot be identified with a twelve-month rotating panel and may instead be absorbed into the remaining transitory component.

The profile declines steeply, from 1.197 at one month to 0.156 at eleven, with a fitted linear slope of  $-0.092$  per month (standard error 0.013) (Table 14). Because Equation (20) implies geometric rather than linear decay, the more informative summary is the implied persistence itself. Regressing  $\log(R(L) - \bar{h}^2)$  upon  $L$  gives  $\hat{\varphi} = 0.798$  per month, a half-life of 3.1 months: roughly four-fifths of whatever divides a household's two answers survives to the following month and half of it is gone within a quarter, against  $\varphi = 1$  under a stable preference. The decline is firmly established over the first six lags, where the ratio is estimated precisely; beyond the seventh the pairs become too few to order the lags, and at the ninth the interval contains zero.

The single  $\sigma_h$  reported in Table 13 should therefore be read as a horizon-specific summary of a decaying object rather than as the variance of a fixed household trait, which is why we report the variance components at two horizons.

#### 5.2.4 The Source of the Additive Component

The additive component is the largest source of variation in the model, and we can examine it independently rather than simply treating  $\varsigma$  as an unexplained residual. Proposition 6 gives us a direct way to do this using the rounding patterns in the data. If respondents report using a grid of width  $\Delta$ , the resulting rounding error has variance  $\Delta^2/12$ . Since the reporting grid can be observed, we can use the distribution of these grids to construct the variance that rounding alone would generate, without introducing any additional parameter.

In the interior-support sample, 6.8% of forecasts are reported at multiples of ten, 17.0% at multiples of five, 68.8% at other integer values, and 7.4% are non-integer. We therefore calculate the variance implied by rounding as

$$\varsigma_{\text{round}}^2 = \sum_g \hat{s}_g \frac{\Delta_g^2}{12}, \quad \Delta_g \in 10, 5, 1, 0.5, \quad (21)$$

where  $\hat{s}_g$  is the share of forecasts associated with each reporting grid, after accounting for finer grids. Table 15 reports these grid shares and the corresponding contribution of each grid to the predicted variance.

The accounting exercise suggests that rounding explains only part of the additive variation. In the interior-support sample, rounding explains only 0.982 of the estimated additive variance of 3.468, which is about 28%. The remaining 72% is not explained by the rounding patterns that is observed in the data. We do not try to assign this remaining component to a specific source without further evidence. Section 5.2.6 instead examines whether part of it can be traced to the survey's elicitation procedure.

The rounding mechanism also gives us a useful additional test. If the point forecast is rounded more coarsely, refining that reporting scale should reduce the divergence. In contrast, changing the precision of the density response should not produce the same effect. We examine this asymmetry by estimating the relationship between divergence and subjective uncertainty among respondents whose point forecasts fall below their density mean:

$$w_{it} = \theta \sigma_{it} + \delta_t + u_{it}, \quad \text{for } \mathcal{S}_- = \{w_{it} < -0.25\}, \quad (22)$$

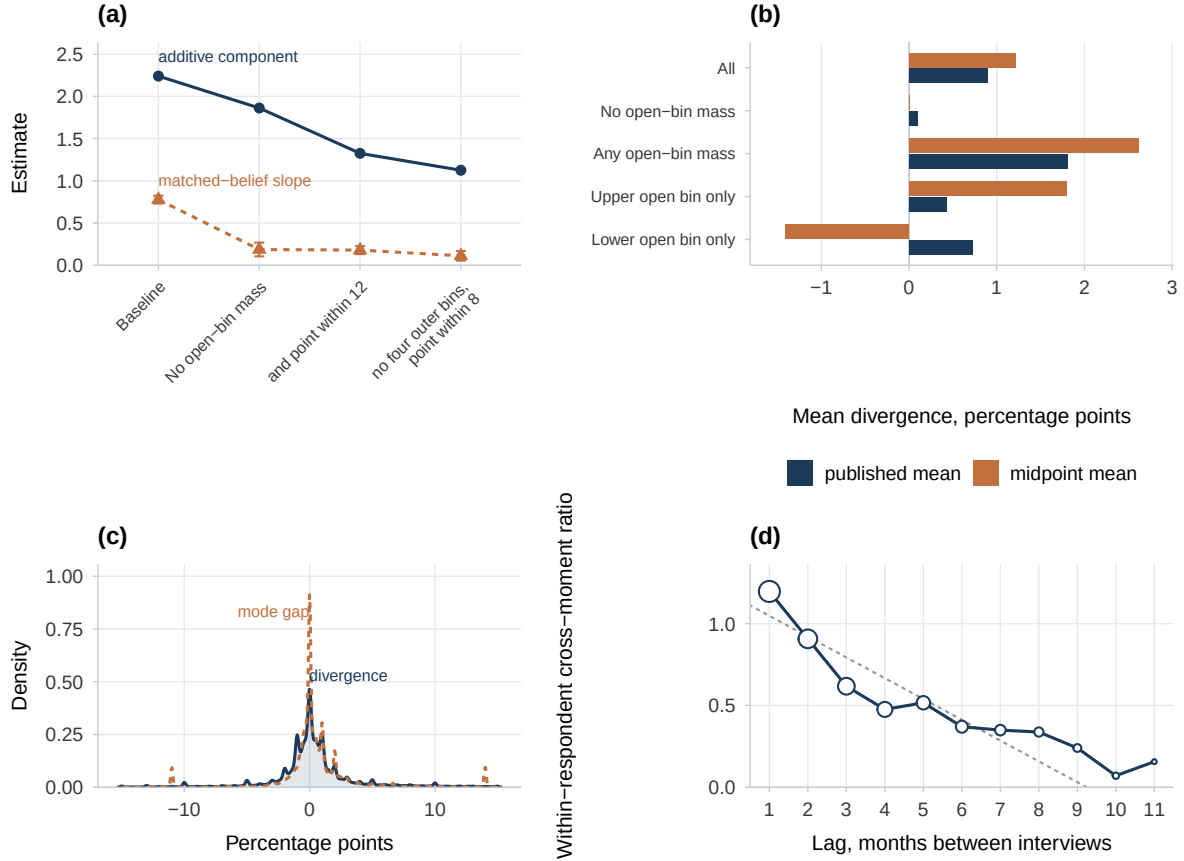


Figure 3: Diagnostic Challenges to Preference-Based Interpretations and Horizon Decay. Panel (a) plots the trajectory of  $\hat{\zeta}$  and  $\hat{\beta}_\sigma$  under nested restrictions expanding the interior-support space. Panel (b) charts the mean absolute divergence within conditional deciles of subjective uncertainty, stratified by the presence of probability mass in open boundary categories. Panel (c) overlays the full probability distribution of the empirical divergence against the counterfactual mode gap. Panel (d) plots the estimated trajectory of the panel cross-moment ratio  $R(L)$  across successive monthly lag horizons.

Table 15: Counterfactual Variance Contribution from Observable Reporting Grids

| Inferred grid $\Delta$                                                     | Share | $\Delta^2/12$ | Contribution |
|----------------------------------------------------------------------------|-------|---------------|--------------|
| 10-percentage-point grid                                                   | 0.068 | 8.333         | 0.570        |
| 5-percentage-point grid                                                    | 0.170 | 2.083         | 0.354        |
| 1-percentage-point grid                                                    | 0.688 | 0.083         | 0.057        |
| 0.5-percentage-point grid                                                  | 0.074 | 0.021         | 0.002        |
| Implied structural variance from rounding ( $\varsigma_{\text{round}}^2$ ) |       |               | <b>0.982</b> |
| Empirical minimum distance intercept ( $\varsigma^2$ ), Table 13           |       |               | 3.468        |

*Note:* The table uses the interior-support sample of 100,419 respondent-months, with survey-weighted shares following Equation (1). Point forecasts are assigned to the coarsest consistent grid (10, 5, 1, or 0.5 percentage points); the 505 forecasts equal to zero are assigned to the one-point grid. Proposition 6 implies a rounding variance of  $\Delta^2/12$ . The implied contribution is 0.982, compared with an estimated additive variance of 3.468, or about 28 percent. This figure is descriptive: the grid classification may overstate coarse reporting, making 0.982 an upper bound, while the denominator is an extrapolated intercept that is sensitive to the aggregation rule (Section 5.2.7).

where  $\delta_t$  denotes survey-month fixed effects. We estimate this relationship across four nested samples, each designed to isolate a different aspect of the reporting process.

Table 16: Conditional Slopes Under Alternative Item-Level Refinement Constraints

| Sample Filtering Restrictions       | $n$     | Regime Share | Conditional Slope ( $\hat{\theta}$ ) |
|-------------------------------------|---------|--------------|--------------------------------------|
| Unconstrained Interior Sample       | 100,419 | 0.357        | −0.876<br>(0.051)                    |
| Contiguous Density Probability Only | 99,477  | 0.356        | −0.766<br>(0.049)                    |
| Non-Integer Point Forecasts Only    | 8,181   | 0.379        | −0.425<br>(0.049)                    |
| Joint Grid and Support Constraints  | 8,135   | 0.378        | −0.369<br>(0.041)                    |

*Note:* The table reports OLS estimates of Equation (22) restricting the sample to the negative divergence regime ( $w_{it} < -0.25$ ) within the interior-support space. All specifications control for survey-month fixed effects ( $\delta_t$ ). Standard errors clustered at the individual respondent level are reported in parentheses.

The results in Table 16 point mainly to the way respondents record their point forecasts. When we restrict the sample to non-integer point forecasts, the conditional slope falls by 51%, from −0.876 to −0.425. By comparison, requiring the probability mass in the density response to be contiguous changes the estimate by only 13%, to −0.766. This suggests that the wedge is driven primarily by the discrete nature of the point forecast rather than by how carefully respondents complete the density question.

There is, however, an important qualification. Only 7.4% of the interior-support sample report a non-integer point forecast, and these respondents are unlikely to be a random subset of the sample. We therefore place more weight on the relative size of the two effects than on their precise magnitudes.

### 5.2.5 Robustness and Sample Selection

The results of the preceding sections depend on how the sample is constructed and how the expectation measures are defined. These choices are not determined by the model itself. It is therefore important to check whether the main estimates change when we use different sample restrictions or alternative ways of measuring the two expectation variables. If the estimates remain broadly similar, the findings are less likely to be an artifact of a particular sample definition or measurement choice.

We therefore re-estimate  $\beta_\sigma$  and the variance components under different sample restrictions and measurement choices. The strongest robustness check comes from the three-year forecast block. Because it uses different survey questions from the one-year measures, it provides a useful test of whether the main findings are specific to the wording or tail-fitting procedure of the baseline questions.

The results in Table 17 suggest that the main parameter is fairly stable under various robustness checks. The independent three-year forecast block gives an uncertainty coefficient of 0.723, close to the baseline estimate of 0.777. This makes it less likely that the result is simply driven by the wording of the original pair of questions. Requiring contiguous density responses also makes almost no difference, with the estimate at 0.776. Using the raw bin midpoints instead gives a somewhat larger estimate of 0.895.

The results are much less stable when we rely on arbitrary trimming rules. Across the four screens, the estimate ranges from 0.466 to 1.190. This sensitivity is precisely why we prefer the interior-support restriction. It is not an arbitrary cutoff selected to obtain a particular result; it follows

Table 17: Sensitivity Analysis of Structural Parameter Estimates Across Selection Criteria

| Sample Selection Protocol             | $n$     | $\hat{\beta}_\sigma$ | $\hat{\varsigma}$ | $\hat{\sigma}_h$ |
|---------------------------------------|---------|----------------------|-------------------|------------------|
| Baseline Selection Screen             | 174,013 | 0.777<br>(0.024)     | 2.241             | 0.923            |
| Contiguous Density Mass Only          | 169,960 | 0.776<br>(0.025)     | 1.997             | 0.941            |
| Exact Probability Summation (=100)    | 174,013 | 0.777<br>(0.024)     | 2.241             | 0.923            |
| Midpoint-Based Moment Aggregation     | 174,013 | 0.895<br>(0.025)     | 2.060             | 0.945            |
| College-Educated Subsample            | 100,514 | 0.518<br>(0.026)     | 1.538             | 0.817            |
| Non-College Subsample                 | 72,884  | 0.810<br>(0.028)     | 2.856             | 0.926            |
| Excluding 2021–2022 Macro Shocks      | 146,508 | 0.740<br>(0.026)     | 2.175             | 0.907            |
| Three-Year Horizon Block              | 174,706 | 0.723<br>(0.026)     | 4.623             | 1.004            |
| <i>Alternative Trimming Schedules</i> |         |                      |                   |                  |
| $ \pi^e  \leq 15\%, \sigma \leq 10\%$ | 157,598 | 0.466<br>(0.016)     | 1.511             | 0.565            |
| $ \pi^e  \leq 30\%, \sigma \leq 15\%$ | 174,013 | 0.777<br>(0.024)     | 2.241             | 0.923            |
| $ \pi^e  \leq 50\%, \sigma \leq 20\%$ | 179,924 | 1.020<br>(0.035)     | 2.844             | 1.254            |
| $ \pi^e  \leq 95\%, \sigma \leq 25\%$ | 182,144 | 1.190<br>(0.044)     | 4.302             | 1.509            |

*Note:* The table reports non-parametric matching parameters  $\hat{\beta}_\sigma$  and minimum distance variance components estimated across alternative sample definitions. Unless specified, all models utilize baseline screening criteria. Models include survey-month fixed effects and narrow belief-bin fixed effects of width 0.25. Standard errors clustered at the individual respondent level are reported in parentheses.

directly from the support of the elicitation instrument and does not introduce an additional tuning parameter.

### 5.2.6 Excluding Bin-Determined Observations

Section 5.1.1 shows that, for 42.1% of the interior-support observations, the reported uncertainty is determined entirely by which histogram bins are selected. This matters because uncertainty is the key regressor in both Equation (17) and Equation (18). The resulting coarsening is not random: it is most common among respondents whose answers reveal relatively little about the shape of their beliefs.

We therefore re-estimate the models after excluding observations for which the binning constraints in Proposition 5 are binding. The remaining sample requires at least three active bins, allowing the fitted generalized beta distribution to be identified from the respondent’s reported probabilities rather than from the bin locations alone. Table 18 presents the resulting estimates using the same estimation procedures as before.

Table 18: Structural Parameter Estimates Excluding Pinned Mass Points

| Structural Parameters                     | (1)<br>Full Interior Support | (2)<br>Excluding Pinned | (3)<br>$\geq$ Three Active Bins |
|-------------------------------------------|------------------------------|-------------------------|---------------------------------|
| Respondent-months ( $n$ )                 | 100,419                      | 76,113                  | 58,177                          |
| Unique Respondents                        | 17,044                       | —                       | 13,616                          |
| Distinct Realizations of $\sigma$         | 10,403                       | 10,400                  | 10,320                          |
| Sample Mean $\mathbb{E}[\sigma]$          | 1.289                        | 1.462                   | 1.689                           |
| Matching Coefficient $\hat{\beta}_\sigma$ | 0.186<br>(0.041)             | 0.211<br>(0.045)        | 0.289<br>(0.057)                |
| Structural Moment Ratio $\bar{h}$         | 0.076                        | 0.088                   | 0.063                           |
| Uncertainty Fixed Effect $\sigma_h$       | 0.451                        | 0.542                   | 0.576                           |
| Transitory Shock Scale $\omega$           | 1.305                        | 1.353                   | 1.470                           |
| Additive Floor Intercept $\varsigma$      | 1.862                        | 1.496                   | <b>0.902</b>                    |
| Implied Rounding Grid (Points)            | 6.45                         | 5.18                    | <b>3.12</b>                     |
| Variance Regression $R^2$                 | 0.326                        | 0.572                   | <b>0.868</b>                    |
| Additive Variance Share (%)               | 44.7                         | 27.2                    | <b>8.5</b>                      |
| Total Variance Explained (%)              | 69.9                         | 81.7                    | <b>88.9</b>                     |
| Estimated Intercept $\varsigma^2$         | 3.468                        | 2.238                   | 0.813                           |
| Analytical Rounded Prediction             | 0.982                        | 0.782                   | 0.797                           |
| Explained Intercept Share                 | 0.28                         | 0.35                    | <b>0.98</b>                     |

*Note:* The table reports matching estimates and minimum-distance variance components for different subsets of the interior-support sample. Column 1 reproduces the preferred specification in Table 13. Pinned observations correspond to the deterministic thresholds implied by Proposition 5, and the rounding benchmarks are recalculated for each partition using Equation (21). Standard errors clustered at the respondent level are reported in parentheses. The “Explained Intercept Share” is a mechanical ratio based on an extrapolated intercept rather than a directly estimated parameter. As shown in Section 5.2.7, it is sensitive to the specification and should therefore be viewed as descriptive, not as an estimate of the share of additive error attributable to rounding.

Table 18 shows how the additive component behaves as the sample is cleaned. In the full interior-support sample the estimated additive component is  $\hat{\varsigma} = 1.862$ , and the variance predicted by the observed rounding grids is about a quarter of it. Once we keep only respondents who use at least three bins,  $\hat{\varsigma}$  falls to 0.902 while the variance predicted by the observed grids is almost unchanged, 0.797 against 0.813 estimated. That the predicted variance is stable is reassuring in itself, because the grid a household uses to write a number should not depend on how many bins it fills in.

It is tempting to conclude from the last two figures that rounding explains about 98% of the additive component in this subsample. We do not make that claim. The ratio 0.797/0.813 is

mechanically correct, but its denominator is not precisely identified. As shown in Section 5.2.7, the ratio changes from 0.77 to 1.40 when the number of quantile cells is changed, and it can even become negative under an alternative aggregation. We therefore do not interpret this ratio as an economically meaningful share.

What the data do support is a more modest conclusion. The rounding variance implied by respondents' reporting grids is stable across the specifications we consider, ranging from about 0.78 to 0.82. At the same time, the estimated additive component falls substantially when respondents with mechanically pinned uncertainty are excluded. These patterns are consistent with rounding being an important source of the remaining divergence, but the data do not allow us to quantify its exact share.

### 5.2.7 The Additive Component

Because much of the interpretation depends on  $\hat{\varsigma}^2$ , we examine how precisely this parameter is identified. It is less precisely determined than the standard errors in Table 13 might suggest, mainly because it is an extrapolated quantity rather than a parameter directly observed in the data.

The parameter is defined as

$$\varsigma^2 = \lim_{x \rightarrow 0} \text{var}(w_{it} \mid \sigma_{it}^2 = x).$$

However, the data do not include observations close to zero uncertainty. The smallest standard deviation in the interior-support sample is 0.433, corresponding to  $\sigma^2 = 0.188$ . Thus, estimating  $\varsigma^2$  requires extrapolating from the observed range to the origin.

To assess how sensitive this extrapolation is, Table 19 keeps the sample and functional form fixed while varying the estimation procedure. This provides a direct check of how much the estimated additive component depends on the extrapolation rather than on the underlying data.

Several features of Table 19 are important.

- The estimate is sensitive to the number of cells. Within our preferred cell-based estimator, changing the number of cells from 15 to 40 moves  $\hat{\varsigma}^2$  from 1.112 to 0.585. The implied rounding share moves from 1.40 to 0.74. Our reported value of 0.813 at 20 cells is therefore a specification choice rather than a uniquely determined estimate.
- The aggregation method matters even more. Using the same linear specification without cells gives an intercept of  $-0.470$  (standard error 0.687). A negative variance has no direct interpretation; instead, this result shows that the fitted relation does not imply a positive intercept when estimated at the individual-observation level. The difference comes from the cell aggregation and weighting. In the cell estimator, a small number of low-uncertainty cells receive most of the weight, and these are also the observations for which  $\sigma$  is least reliably measured.
- Allowing for curvature makes the uncertainty explicit. Once the conditional relationship is allowed to deviate modestly from global linearity, the confidence interval includes zero. Under the mildest curvature bound considered, the interval is  $[-0.85, 4.12]$ . This is not surprising given that  $\varsigma^2$  is extrapolated from the smallest observed  $\sigma^2$  of about 0.19 to zero.

More importantly, these results do not alter the primary insights established in Sections 5.1.3 and earlier. The boundary effects, the bin-midpoint counterfactual, and the decline in mean divergence from  $+0.891$  to  $+0.099$  are identified directly from the data. What remains uncertain is the magnitude of  $\varsigma^2$  and, consequently, any claim that assigns a precise percentage of the additive component to rounding. We therefore treat such quantities cautiously and flag them wherever they enter the analysis.

Several other results do not depend on the magnitude of  $\varsigma^2$  and point in the same direction:

Table 19: Sensitivity of the additive component to the aggregation rule

| Estimator of $\zeta^2$                      | Full interior support |              | $\geq$ three active bins |              |
|---------------------------------------------|-----------------------|--------------|--------------------------|--------------|
|                                             | $\zeta^2$             | Share        | $\zeta^2$                | Share        |
| Cell WLS, 10 cells                          | 3.690                 | 0.279        | 1.067                    | 0.766        |
| Cell WLS, 15 cells                          | 3.283                 | 0.314        | 1.112                    | 0.735        |
| Cell WLS, 20 cells (this paper)             | <b>3.468</b>          | <b>0.297</b> | <b>0.813</b>             | <b>1.006</b> |
| Cell WLS, 30 cells                          | 3.140                 | 0.328        | 1.018                    | 0.803        |
| Cell WLS, 40 cells                          | 2.780                 | 0.371        | 0.585                    | 1.398        |
| Same linearity, no cells                    | 5.211                 | 0.198        | -0.470                   | -1.740       |
| cluster standard error                      | (0.473)               |              | (0.687)                  |              |
| Local linear at $x = 0$ , linearity imposed | 6.959                 | 0.148        | 1.438                    | 0.569        |
| 95% interval                                | [6.10, 7.82]          |              | [0.61, 2.27]             |              |
| Local linear, at $B/4$                      | 7.154                 | 0.144        | 1.634                    | 0.500        |
| 95% interval                                | [3.65, 10.66]         |              | [-0.85, 4.12]            |              |
| Local linear, at $B/2$                      | 7.154                 | 0.144        | 1.634                    | 0.500        |
| 95% interval                                | [1.75, 12.56]         |              | [-1.33, 4.60]            |              |

*Note:* All estimators use the same sample and the same specification,  $\text{var}(w \mid \sigma) = (\sigma_h^2 + \omega^2)\sigma^2 + \zeta^2$ . They differ only in how the data are aggregated before estimating the intercept. Cell WLS” is our preferred estimator, using quantile cells weighted by  $n_b/\widehat{\text{var}}_b^2$ . Same linearity, no cells” estimates the same relation directly at the observation level with survey weights and respondent-clustered standard errors. “Local linear at  $x = 0$ ” estimates the boundary value directly, with confidence intervals following Armstrong and Koles’ar (2018).

The “Share” is the rounding variance implied by the observed reporting grids divided by  $\zeta^2$ . The implied rounding variance is 1.031 for the baseline sample and 0.818 for the interior-support sample. These values differ slightly from Table 15 because of the grid-classification rule.

The “ $B$ ” is the curvature bound of the interval construction. Because  $\zeta^2$  is the value of the conditional variance at  $x = \sigma^2 = 0$ , a point the data do not reach, an interval for it can be valid only uniformly over a class of conditional expectations rather than at one. We take the class whose departure from a straight line is at most  $(B/2)x^2$ , that is  $|f''| \leq B$ . Over that class the worst-case bias of a linear estimator with weights  $a_i$  is  $(B/2) \sum_i |a_i| x_i^2$ , and the reported interval is  $\hat{\theta} \pm cv_{1-\alpha}(\text{bias/se}) \cdot \text{se}$ , where  $cv_{1-\alpha}(t)$  is the  $1 - \alpha$  quantile of  $|N(t, 1)|$ . Since  $\zeta^2$  is a variance, the conditional expectations of both  $w^2$  and  $w$  upon  $x$  are extrapolated and each carries its own bound;  $B$  denotes the pair, set by the rule of thumb of a weighted quartic fit over the lower nine deciles of  $x$  with  $B = \max |f''|$ . This gives (22.819, 1.547) upon the full interior support and (5.093, 0.229) upon the sample with three or more active bins. The rows  $B/4$  and  $B/2$  scale both bounds by those fractions, so they permit a quarter and a half of the curvature a flexible fit detects, while  $B = 0$  imposes exact linearity and reproduces the extrapolation the cell estimator performs implicitly. The bandwidth is re-selected at each bound to minimise the length of the interval.

- The same pattern appears in the model fit. The model’s fit improves sharply, with the  $R^2$  increasing from 0.326 to 0.868 after excluding the pinned observations (Section 5.2.5; Table 18).
- Correcting for the coarsened uncertainty measure raises  $\hat{\beta}_\sigma$  from 0.186 to 0.289 and  $\hat{\sigma}_h$  from 0.451 to 0.576, consistent with the attenuation expected from measurement error (Section 5.2.5; Table 18).

There are two qualifications to keep in mind. The cleaner sample is not simply a random subset: it has higher average uncertainty (1.689 versus 1.289) and contains only 13,616 respondents. Thus, the comparison should not be interpreted as a clean experiment. Also,  $\bar{h}$  changes only slightly, from 0.076 to 0.063, rather than moving with  $\hat{\beta}_\sigma$ . The two estimates therefore continue to provide bounds rather than converging to a single value.

The main takeaway is narrower but important: the measurement-error model fits the data much better once subjective uncertainty is genuinely measured rather than mechanically determined by the reported bins.

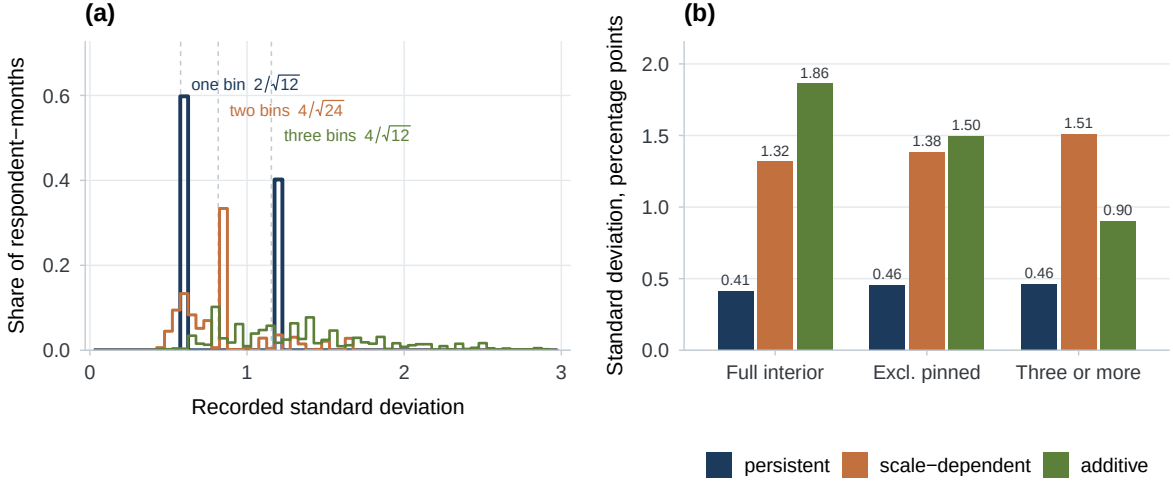


Figure 4: Own uncertainty and the fitting rule on the interior-support sample. Panel (a) plots the distribution of the recorded standard deviation separately for respondents using one, two and three bins, with the three values that Proposition 5 predicts marked. Panel (b) plots the three estimated components upon the whole interior support, upon the sample excluding the respondents to whom the proposition applies, and upon the densities using at least three bins.

### 5.3 Predictive Accuracy

The components estimated above are identified from the relationship between the two survey responses. We next test them against inflation outcomes that were not used in estimation. The relevant question is whether the density mean or the reported point forecast better predicts subsequent inflation.

Proposition 9 provides a direct test. If the divergence between the two measures is mainly due to reporting noise, the elicited density mean should predict realised inflation more accurately than the reported point forecast. The model also predicts that this difference should be larger when respondents report greater uncertainty, because two of the three error components scale with  $\sigma$ . We therefore test two implications:

- **Relative accuracy:** the density mean should have a lower prediction error than the point forecast.

- **Dependence on uncertainty:** the accuracy advantage of the density mean should increase with reported uncertainty.

The purpose is not to assess whether households forecast inflation well in absolute terms. We instead compare the two survey measures as alternative measures of the same underlying expectation. We also report standard forecasting benchmarks to provide a reference for the magnitude of the errors.

Table 20: Out-of-Sample Predictive Accuracy Against Realized Macroeconomic Inflation

| Expectation Measure                        | RMSE (Interior) | MAE (Interior) | Bias (Interior) | RMSE (Baseline) |
|--------------------------------------------|-----------------|----------------|-----------------|-----------------|
| Reported Point Forecast                    | 4.741           | 3.057          | +0.419          | 8.204           |
| Elicited Density Mean                      | <b>3.266</b>    | 2.420          | +0.305          | 5.862           |
| Elicited Density Median                    | 3.271           | 2.413          | +0.270          | 5.908           |
| Elicited Density Mode                      | 3.522           | 2.634          | +0.602          | 5.916           |
| Pooled Average ( $\frac{\pi^e + \mu}{2}$ ) | 3.718           | 2.642          | +0.362          | 6.323           |
| <i>Statistical Forecast Benchmarks</i>     |                 |                |                 |                 |
| Random Walk (Lagged Inflation)             | 2.076           | 1.456          | −0.193          | 2.203           |
| Constant Inflation Floor (2%)              | 2.098           | 1.360          | −0.594          | 2.249           |

*Note:* The forecast performance is evaluated on the interior-support sample against realized CPI inflation over the specific twelve-month horizon targeted by each survey item. Performance criteria map to Equation (3). The final column reports the comparative full-sample baseline RMSE.

Table 20 shows a clear difference in forecasting performance. In the preferred interior-support sample, the elicited density mean has a lower RMSE than the reported point forecast, 3.266 compared with 4.741. The same ranking appears in the full sample, although the boundary observations make the point forecast substantially less accurate.

Table 21 takes the comparison a step further by examining forecast accuracy separately across quintiles of subjective uncertainty.

Table 21: Conditional Predictive Accuracy Across Subjective Uncertainty Quintiles

| Uncertainty Quintile ( $b$ ) | Mean $\sigma_b$ | Point Forecast RMSE | Density Mean RMSE | Predictive Margin |
|------------------------------|-----------------|---------------------|-------------------|-------------------|
| 1                            | 0.573           | 3.50                | 2.40              | +1.10             |
| 2                            | 0.791           | 3.70                | 2.43              | +1.28             |
| 3                            | 1.122           | 5.57                | 4.50              | +1.08             |
| 4                            | 1.497           | 4.27                | 3.03              | +1.24             |
| 5                            | 2.565           | 6.10                | 3.43              | +2.68             |

*Note:* The conditional forecast RMSE is split across five equal-sized quintiles of the subjective uncertainty distribution ( $\sigma_{it}$ ) within the interior-support sample space. The predictive margin is explicitly defined as the point forecast RMSE minus the continuous density mean RMSE.

Across the five uncertainty quintiles in Table 21, the density mean consistently delivers the more accurate forecast. The difference is relatively small for the least uncertain respondents, at 1.10 percentage points, but reaches 2.67 points for the most uncertain group. The intermediate quintiles are somewhat irregular, so we interpret the evidence as supporting the predicted direction and providing weaker support for the exact shape of the relationship.

At the same time, neither survey measure performs particularly well in absolute terms. A simple random-walk forecast based on past-year inflation has an RMSE of 2.076, well below both survey measures. This is important for interpretation: the exercise is about relative measurement quality, not about showing that household forecasts are accurate in an absolute sense.

Finally, households may evaluate inflation using consumption baskets that differ from the official price index. This could affect the levels and biases reported in the table, but it does not change the ranking because every measure is evaluated against the same realised inflation series.

## 5.4 Scope of the Findings

The results above come from a single household survey, so it is important to distinguish between what follows directly from the survey design and what is an empirical finding that may not carry over to other datasets.

The mechanical result is broadly general. The finding in Section 5.1.1 follows from the way interval probabilities are converted into continuous moments. Proposition 5 is therefore a feature of the fitting procedure, not of a particular population. The interpolation method of Engelberg et al. (2009), or closely related approaches, is used in several major inflation-expectation surveys, including the Survey of Professional Forecasters, the European Central Bank’s Consumer Expectations Survey, and the Bank of England’s household survey. Whenever an instrument assigns a fixed density shape within reported bins, respondents who place all their probability in one bin mechanically receive an uncertainty measure determined by that bin’s width.

Thus, the key quantity to check in another dataset is simply:

- the bin boundaries used by the survey;
- the share of respondents with only one active bin; and
- the uncertainty implied by those bins.

In this sample, 42.1% of observations fall into this mechanically determined category.

The estimated variance components are less portable. Unlike Proposition 5, the results in Sections 5.2.2–5.2.6 are empirical estimates. We therefore cannot assume that the additive reporting component will dominate in another survey or that density coarseness will remain the main source of measurement error. These questions depend on the population, questionnaire design, and bin structure.

Replication on other microdata will be valuable and we leave a full cross-survey comparison for future work. The main quantities to compare are:

- conditional mean divergence by boundary-mass status;
- the share of mechanically pinned uncertainty measures; and
- the three variance components evaluated at a common lag.

There is also some evidence of external validity within our data. The three-year block uses a different pair of point-forecast and density questions from the one-year block. Estimating the model on this independent set of questions gives  $\hat{\beta}_\sigma = 0.723$ , close to the baseline estimate of 0.777. The similarity is reassuring because it suggests that the main relationship is not simply a consequence of the wording or tail-fitting procedure of one particular pair of survey questions.

## 6 Measured Disagreement

The spread in reported point forecasts is often used in macroeconomic research to measure consumer disagreement, subjective uncertainty, and the degree to which inflation expectations are anchored. However, Proposition 7 shows that this dispersion also contains reporting noise. We therefore separate the observed variation into three components:

Persistent heterogeneity, captured by differences in  $h_i$  across respondents. Scale-dependent reporting noise, which grows with respondents’ subjective uncertainty. Additive reporting noise, which does not depend on the level of uncertainty.

We measure their contributions using the following variance shares:

$$S^{\text{pers}} \equiv \left[ \frac{\hat{\sigma}_h^2 \hat{\mathbb{E}}_i[\sigma_{it}^2]}{\widehat{\text{var}}_i(\pi_{it}^e)} \right], \quad S^{\text{scale}} \equiv \left[ \frac{\hat{\omega}^2 \hat{\mathbb{E}}_i[\sigma_{it}^2]}{\widehat{\text{var}}_i(\pi_{it}^e)} \right], \quad S^{\text{add}} \equiv \left[ \frac{\hat{\zeta}^2}{\widehat{\text{var}}_i(\pi_{it}^e)} \right], \quad (23)$$

Here, the moments are calculated separately for each survey month, following Equation (1), and then averaged across the 149 months. We report the decomposition under both parameter estimates: the full sample, which corresponds to the published aggregate series, and the interior-support sample, where the boundary problems do not affect identification.

Table 22: Decomposing the Within-Month Cross-Sectional Variance of Point Forecasts

| Variance Shares and Decompositions                            | (1)<br>Baseline Framework | (2)<br>Interior Parameters |
|---------------------------------------------------------------|---------------------------|----------------------------|
| Total Point Forecast Dispersion ( $\text{var}_i(\pi^e)$ )     | 59.49                     | 59.49                      |
| Elicited Central Tendency Variance ( $\text{var}_i(\mu)$ )    | 29.67                     | 29.67                      |
| Empirical Divergence Dispersion ( $\text{var}_i(w)$ )         | 42.96                     | 42.96                      |
| Persistent Individual Characteristic ( $S^{\text{pers}}$ , %) | 10.4                      | 6.3                        |
| Transitory Scale-Dependent Shock ( $S^{\text{scale}}$ , %)    | 47.9                      | 52.8                       |
| Additive Metric Noise ( $S^{\text{add}}$ , %)                 | 8.8                       | 6.1                        |
| <b>Total Spurious Measurement Component (%)</b>               | <b>67.1</b>               | <b>65.2</b>                |
| Residual Structural Variation (%)                             | 32.9                      | 34.8                       |

*Note:* The table reports the structural decomposition of the cross-sectional variance of the published point forecast series according to the share equations defined in Equation (23). Values represent arithmetic averages compiled across the 149 active survey waves. The structural parameters correspond to the annual lag horizon specifications mapped in Table 13.

Table 22 suggests that a large part of the observed dispersion in household inflation expectations is not genuine disagreement about future inflation. Depending on the specification:

- 65.2–67.1% of the cross-sectional variance is associated with reporting noise.
- A further 6.3–10.4% reflects persistent respondent-specific differences that are not directly tied to beliefs.

We report ranges rather than a single number because the two parameter sets are estimated from different samples. Interestingly, the resulting variance shares are much more stable than the individual parameter estimates. This is because the lower additive component in the interior sample is largely offset by its higher scale-dependent component.

These results also suggest a relatively simple way to improve measures of disagreement. Rather than relying on the raw dispersion of point forecasts, researchers can:

- calculate dispersion separately for respondents whose density responses remain within the interior support;
- distinguish respondents using coarse reporting grids from those using finer grids; and
- track the difference between the clean and boundary samples as an indicator of measurement contamination.

The broader implication is important. An increase in the dispersion of reported forecasts does not necessarily mean that households have become more divided in their beliefs. It may instead reflect changes in how respondents report those beliefs. Consequently, measures of anchoring based solely on the dispersion of point forecasts should be interpreted with caution.

## 7 Aggregate Estimands

A measurement problem is important only if it changes the economic conclusions drawn from the data, not simply the size of the standard errors. We therefore take the analysis one step further. We construct a monthly measure of consumer disagreement that removes the reporting

components identified above and compare it with both the raw point-forecast measure and the density-based measure.

The comparison uses three standard exercises from the macroeconomic literature, with the survey month as the unit of observation ( $T = 149$ ).

To make the corrected measure comparable with the published aggregate series, we use the baseline parameter estimates from Column 1 of Table 13. Throughout this section, standard errors are based on Newey–West estimates with a 12-month lag window, allowing for both heteroskedasticity and serial correlation.

## 7.1 A Corrected Disagreement Index

Proposition 7 implies that the observed dispersion of point forecasts contains a predictable amount of reporting noise. In month  $t$ , this noise is given by  $\omega^2 \mathbb{E}_i[\sigma_{it}^2] + \varsigma^2$ . We remove this component to obtain a cleaner measure of the underlying disagreement:

$$D_t^{\text{corr}} \equiv [\text{var}_i(\pi_{it}^e) - \hat{\omega}^2 \mathbb{E}_i(\sigma_{it}^2) - \hat{\varsigma}^2]^{1/2}. \quad (24)$$

Here,  $\mathbb{E}_i$  and  $\text{var}_i$  are survey-weighted cross-sectional moments calculated separately for each month. We compare the corrected series with two familiar measures:

$$D_t^{\pi^e} \equiv [\text{var}_i(\pi_{it}^e)]^{1/2}, \quad D_t^\mu \equiv [\text{var}_i(\mu_{it})]^{1/2}. \quad (25)$$

In simple terms:

- $D_t^{\pi^e}$  is the dispersion in reported point forecasts.
- $D_t^\mu$  measures dispersion in the density-based means.
- $D_t^{\text{corr}}$  removes the reporting noise identified by the model.

The correction is substantial. It removes 67.1% of the variance in the raw published dispersion measure.

## 7.2 Disagreement and the Level of Inflation

Following Mankiw et al. (2003), we examine whether disagreement in inflation expectations tends to increase when inflation itself is higher. We run the same regression for each of the three disagreement measures:

$$D_t = \alpha + \gamma \pi_t + \delta u_t + e_t, \quad (26)$$

Here,  $\pi_t$  is twelve-month inflation and  $u_t$  is the unemployment rate. We estimate the model separately for the raw, corrected, and density-based disagreement measures. We then repeat the exercise in first differences to check whether the relationship reflects longer-run trends. The results are reported in the right-hand columns of Table 23.

Table 23 suggests that the relationship reported by Mankiw et al. (2003) is much clearer when disagreement is measured from the density responses rather than the raw point forecasts. The density-based measure,  $D_t^\mu$ , has a slope of 0.216 and an  $R^2$  of 0.392, compared with 0.169 and 0.180 for  $D_t^{\pi^e}$ . After removing the reporting noise from the point forecasts, the slope falls further to 0.119 and is no longer statistically different from zero. A moving-block bootstrap estimates the difference between the raw and corrected measures at 0.061 (recentered  $p = 0.034$ ).

The results look quite different in first differences. The inflation coefficient becomes negative for all three measures and is statistically significant for two of them. This suggests that the positive relationship is mainly a long-run pattern, rather than a month-to-month relationship. Given that the sample covers only one major inflation episode, this distinction is important when interpreting the result.

Table 23: Time-Series Projections of Consumer Disagreement on Macroeconomic Realizations

| Disagreement Metric ( $D_t$ )             | Level Specifications                 |              | First-Difference Specifications      |       |
|-------------------------------------------|--------------------------------------|--------------|--------------------------------------|-------|
|                                           | Conditional Slope ( $\hat{\gamma}$ ) | $R^2$        | Conditional Slope ( $\hat{\gamma}$ ) | $R^2$ |
| Raw Point Forecast ( $D_t^{\pi^e}$ )      | 0.169<br>(0.049)                     | 0.180        | -0.271<br>(0.140)                    | 0.037 |
| Artifact-Filtered ( $D_t^{\text{corr}}$ ) | 0.119<br>(0.069)                     | 0.105        | -0.371<br>(0.193)                    | 0.033 |
| Continuous Density Mean ( $D_t^\mu$ )     | <b>0.216</b><br>(0.021)              | <b>0.392</b> | -0.032<br>(0.091)                    | 0.011 |

*Note:* The table reports time-series parameter estimates of Equation (26) compiled across  $T = 149$  active survey waves. The independent variable of interest is realized consumer price inflation ( $\pi_t$ ). Models control for the civilian unemployment rate ( $u_t$ ). The left-hand panels estimate the variables in levels; the right-hand panels execute the models in first differences. Newey–West standard errors computed using a twelve-month lag truncation window are reported in parentheses.

### 7.3 Information Rigidity and Inattention

To measure the extent of informational frictions, we use the sticky-information approach of Coibion and Gorodnichenko (2015b). The idea is straightforward: if some households do not update their information regularly, changes in the consensus forecast should contain information about future forecast errors. We estimate

$$\pi_{t+12} - F_t = c + \beta (F_t - F_{t-k}) + u_t, \quad \lambda = \frac{\beta}{1 + \beta}, \quad (27)$$

where  $F_t$  is the consensus forecast, calculated separately from the reported point forecasts and the density means,  $\pi_{t+12}$  is realised inflation over the following twelve months, and  $k$  is the forecast-revision horizon. The parameter  $\lambda \in [0, 1)$  measures the implied share of households that have not updated their information.

Table 24: Information Rigidity Projections and Implied Inattention Parameters

| Forecast Horizon   | Consensus Source | Slope ( $\hat{\beta}$ ) | Std. Error | Rigidity ( $\hat{\lambda}$ ) | $N$ |
|--------------------|------------------|-------------------------|------------|------------------------------|-----|
| One-Year Horizon   | Point Forecasts  | 0.605                   | 0.391      | 0.377                        | 138 |
|                    | Density Means    | 1.188                   | 0.541      | <b>0.543</b>                 | 138 |
| Three-Year Horizon | Point Forecasts  | -0.500                  | 0.236      | <i>Undefined</i>             | 114 |
|                    | Density Means    | 0.701                   | 0.447      | 0.412                        | 114 |

*Note:* The table reports time-series OLS estimates of Equation (27) over a revision window of  $k = 6$  months. The dependent variable is the aggregate forecast error ( $\pi_{t+12} - F_t$ ). The implied inattention parameter is mapped as  $\hat{\lambda} = \hat{\beta}/(1 + \hat{\beta})$ . Newey–West standard errors estimated using a twelve-month lag truncation window are reported in the fourth column.

Table 24 shows a sizeable difference between the two measures of expectations. Using the density responses gives an implied informational rigidity of  $\hat{\lambda} = 0.543$ , compared with only 0.377 when the point forecasts are used. The coefficient is statistically significant at the 5% level only for the density-based measure. A moving-block bootstrap also finds a significant difference between the two estimates, with a gap of 0.583 (standard error 0.148,  $p = 0.024$ ).

This difference is broadly consistent with the attenuation predicted by Proposition 8. The six-month revision is more variable for the point forecasts (0.949) than for the density means (0.446), which implies an attenuation ratio of about 0.470. The observed ratio is 0.509, reasonably close to this prediction. This supports the interpretation that reporting noise in the point forecasts weakens the estimated degree of informational rigidity.

There are, however, two important qualifications. First, because the same aggregate forecast enters both sides of Equation (27), the attenuation result should be viewed as an approximation. Second, the survey horizon moves forward with each wave, so changes in the forecast may reflect both new information and changes in the forecast horizon. The separate three-year block reinforces this concern: its point-forecast specification produces a negative slope, making the implied inattention parameter undefined. Thus, the evidence for informational rigidity is best viewed as suggestive for the one-year horizon, rather than as a general result across horizons.

## 7.4 Pass-Through to Realised Inflation

The final time-series exercise asks whether the household consensus contains useful information about subsequent inflation outcomes. This provides a reduced-form way to assess the aggregate expectations channel discussed by Coibion et al. (2018a). We consider three standard specifications:

$$\pi_{t+12} = a + bF_t + e_t, \quad (28)$$

$$\pi_{t+12} = a + bF_t + cu_t + e_t, \quad (29)$$

$$\pi_{t+12} - \pi_t = a + b(F_t - \pi_t) + e_t. \quad (30)$$

The third specification corresponds to the usual accelerationist relationship. For each specification, we use two alternative measures of the consensus forecast: the reported point forecasts and the density-based means. This allows us to see whether the choice of survey measure affects how much information household expectations contain about future inflation.

Table 25: Reduced-Form Predictive Pass-Through to Realized Inflation

| Specification and Constraints       | Point Forecast Consensus   |       | Density Mean Consensus     |       |
|-------------------------------------|----------------------------|-------|----------------------------|-------|
|                                     | Pass-Through ( $\hat{b}$ ) | $R^2$ | Pass-Through ( $\hat{b}$ ) | $R^2$ |
| Equation (28): Consensus Baseline   | 0.855<br>(0.296)           | 0.274 | 1.094<br>(0.351)           | 0.291 |
| Equation (29): With Unemployment    | 0.866<br>(0.298)           | 0.307 | 1.120<br>(0.345)           | 0.330 |
| Equation (30): Accelerationist Form | 0.907<br>(0.302)           | 0.342 | 0.952<br>(0.277)           | 0.360 |

*Note:* The table reports the estimated pass-through parameter  $\hat{b}$  across  $T = 144$  active survey waves. The dependent macro variable is realized consumer inflation over the targeted twelve-month window ( $\pi_{t+12}$ ). Specifications correlate to equations Equation (28) through Equation (30). Newey–West standard errors calculated using a twelve-month lag truncation window are reported in parentheses.

Table 25 gives a pass-through estimate of  $\hat{b} = 0.855$  when the consensus is based on the raw point forecasts, compared with 1.094 when it is based on the density means. Both estimates are fairly close to the benchmark value of one, so neither is statistically distinguishable from full pass-through on its own. The more relevant comparison is between the two estimates. Here, the moving-block bootstrap gives a difference of 0.240.

The same pattern appears when unemployment is added as a control in Equation (29) and when the accelerationist specification in Equation (30) is used. This suggests that the difference is not simply a consequence of the particular regression specification. The estimates are also shown graphically in Panel (d) of Figure 5.

These results should not be interpreted as evidence that household expectations causally affect subsequent macroeconomic outcomes. The regressions are reduced-form relationships. Our narrower point is that the choice of elicitation method can change the estimated pass-through coefficient in economically meaningful ways.

## 7.5 Statistical Significance of the Differentials

Each comparison above is based on two estimates obtained from the same survey sample and waves. As a result, the two estimates are naturally correlated, so their difference cannot be assessed reliably by simply comparing their individual Newey–West standard errors. We therefore use a moving-block bootstrap to estimate the sampling distribution of the difference directly:

$$\hat{\Delta} \equiv \hat{\beta}^{\mu} - \hat{\beta}^{\pi^e}, \quad \hat{\Delta}^{(r)} \equiv \hat{\beta}^{\mu(r)} - \hat{\beta}^{\pi^e(r)}, \quad r = 1, \dots, 999, \quad (31)$$

For each bootstrap replication, we draw contiguous blocks of twelve survey months with replacement and combine them until the original sample length is restored. Both coefficients are then re-estimated on the same bootstrap sample. Using blocks rather than individual months preserves the serial dependence in the data.

We summarize the bootstrap distribution in two ways.

- The 2.5th and 97.5th percentiles give the 95% confidence interval for the difference.
- We calculate a two-sided  $p$ -value using the recentred bootstrap distribution,  $\hat{\Delta}^{(r)} - \hat{\Delta}$ . Re-centering is important because the bootstrap distribution is centered around the estimated difference rather than zero; the raw bootstrap tails therefore do not provide a direct test of the null hypothesis  $\hat{\Delta} = 0$ . This procedure gives a direct comparison of the two estimates while accounting for their common sampling variation and the time-series dependence in the survey data.

Table 26: Moving-Block Bootstrap Distribution of Elicitation Parameter Gaps

| Comparison                                                         | Density | Point | Gap   | Std. error | 95% percentile interval | $p$ -value |
|--------------------------------------------------------------------|---------|-------|-------|------------|-------------------------|------------|
| Inattention Slope ( $\beta$ in Equation (27), $k = 6$ )            | 1.188   | 0.605 | 0.583 | 0.148      | [0.008, 0.574]          | 0.024      |
| Predictive Pass-Through ( $b$ in Equation (28))                    | 1.094   | 0.855 | 0.240 | 0.170      | [0.079, 0.708]          | 0.099      |
| Disagreement Level ( $\gamma$ in Equation (26), Density vs. Point) | 0.216   | 0.169 | 0.054 | 0.044      | [-0.056, 0.130]         | 0.182      |
| Disagreement Bias ( $\gamma$ in Equation (26), Point vs. Filtered) | 0.169   | 0.119 | 0.061 | 0.028      | [-0.006, 0.101]         | 0.034      |

*Note:* The table reports the moving-block bootstrap distribution of the test statistic defined in Equation (31). Replicates are constructed using a 12-month block window and 999 resampling paths, and the underlying data span  $T = 149$  active survey periods. The standard error is the cross-replicate standard deviation, and the percentile interval runs between the 2.5th and 97.5th quantiles of the replicate distribution. The  $p$ -value is two-sided and computed from the recentred replicates as described in Section 7.5; it is not the raw tail mass of the replicate distribution about zero, which would not be a test of the gap. All four replicate distributions are strongly skewed, with standardised third moments of +1.74, +2.50, -0.85 and -1.46 in row order, so the percentile interval and the recentred  $p$ -value are not equivalent readings and can disagree. Where they disagree both are reported rather than the more favourable one. For the inattention slope the sample gap of 0.583 lies above the 97.5th replicate quantile, which indicates that the block bootstrap is poorly centred for that statistic and that its interval should not be read as a confidence statement.

The bootstrap results in Table 26 provide a more cautious interpretation of the aggregate comparisons. The clearest evidence concerns the inattention measure: the difference between the two estimates is statistically significant under both bootstrap criteria, with a recentred  $p$ -value of 0.024 and a percentile interval that does not include zero.

The evidence for the pass-through difference is weaker. Although its percentile interval, [0.079, 0.708], excludes zero, the recentred  $p$ -value is 0.099. The bootstrap distribution is also noticeably right-skewed, so we do not regard the estimated gap of 0.240 as conclusive evidence of a difference.

For the disagreement measures, the difference between the raw and artifact-corrected series is statistically significant at the 5% level ( $p = 0.034$ ), whereas the difference between the density-based and point-forecast measures is not ( $p = 0.182$ ).

Taken together, these results suggest that the survey format has a meaningful effect on measured disagreement and on the estimated degree of information rigidity. The evidence for an effect on

reduced-form inflation pass-through is considerably less decisive, particularly given the relatively short time series of 149 monthly observations.

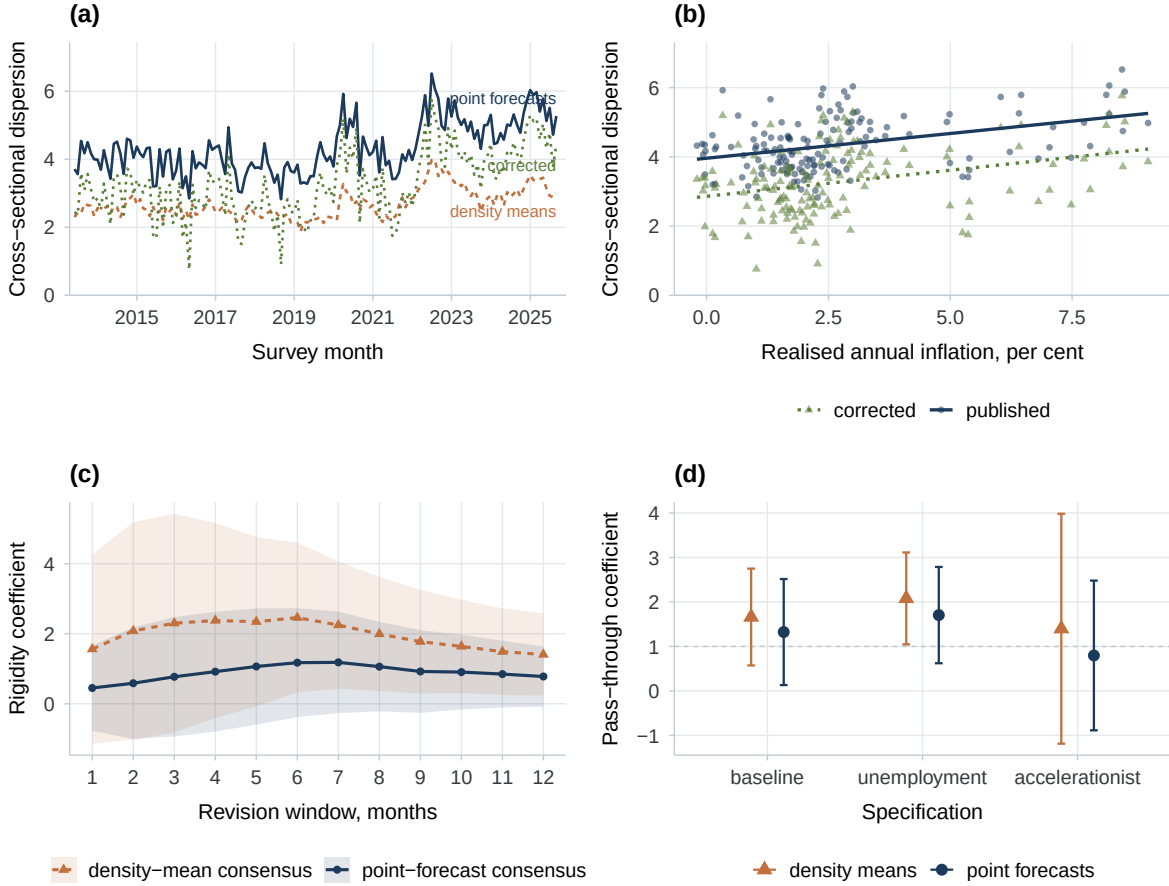


Figure 5: Counterfactual Macroeconomic Aggregates and Resampling Distributions. Panel (a) plots the time-series paths of the three alternative monthly inflation disagreement measures. Panel (b) models each disagreement series against backward-looking realized annual inflation alongside their respective linear profiles. Panel (c) charts the empirical rigidity parameter from Equation (27) across rolling revision windows with 95% confidence bands. Panel (d) displays the estimated reduced-form predictive pass-through coefficients alongside their respective 95% intervals.

## 7.6 The 2021–2022 Episode

Proposition 7 shows that the published point forecast should not be interpreted as a pure measure of the respondent’s underlying belief. It combines the average belief with a reporting component that can vary over time. As discussed in Section 5.1.3, this reporting wedge becomes more pronounced when respondents’ distributions approach the boundaries of the available response range, something that is likely to occur when uncertainty is unusually high.

To see whether this matters during periods of economic stress, Table 27 compares reporting behaviour across two very different episodes: the relatively stable, low-inflation period from 2015–2019 and the sharp inflationary episode during 2021–2022.

Table 27 suggests that the rise in reported inflation expectations during 2021–2022 was somewhat larger in the point forecasts than in the density means. The aggregate point forecast increased by 3.082 percentage points, compared with an increase of 2.660 points in the density mean. The difference of 0.421 percentage points therefore represents about one-seventh of the increase in

Table 27: Expectation Moments and Elicitation Wedges Across Inflation Regimes

| Expectation Moments                         | 2015–2019 Regime | 2021–2022 Shock | Regime Change |
|---------------------------------------------|------------------|-----------------|---------------|
| Respondent-Months ( $n$ )                   | 73,412           | 27,562          | —             |
| Mean Subjective Uncertainty ( $\sigma$ )    | 2.795            | 3.552           | 0.757         |
| Mean Reported Point Forecast ( $\pi^e$ )    | 4.368            | 7.450           | <b>3.082</b>  |
| Mean Continuous Density Mean ( $\mu$ )      | 3.455            | 6.115           | <b>2.660</b>  |
| Mean Empirical Divergence ( $\pi^e - \mu$ ) | 0.914            | 1.335           | <b>0.421</b>  |

*Note:* The table presents weighted cell means computed according to Equation (1) utilizing the baseline full sample. The target cohorts isolate the stable pre-pandemic benchmark (2015–2019) against the high-inflation macro regime shift (2021–2022). The final column provides the exact difference vector.

the point forecasts and can be viewed as the additional movement associated with the reporting wedge.

Subjective uncertainty also increased substantially during this period, rising by 0.757 percentage points. Using the baseline estimate of  $\bar{h} = 0.291$  from Table 13 gives an implied increase in the wedge of about 0.220 points, or roughly 52% of the observed difference.

There is, however, an important qualification. The exercises in Sections 5.1.3 and 5.1.4 suggest that the baseline estimate of  $\bar{h}$  is strongly affected by the way the tails are fitted. It is therefore difficult to interpret the increase as evidence that households became more loss-averse during the pandemic.

A more cautious interpretation is that the sharp increase in inflation and uncertainty pushed more responses toward the open boundary bins, which in turn affected the reported point forecasts. Since uncertainty and boundary-mass allocation are strongly correlated over time (0.645), aggregate data cannot cleanly separate these two mechanisms. We therefore treat the decomposition as an accounting exercise rather than as evidence of a change in household preferences.

## 8 Conclusion

Household inflation expectations are widely used in monetary economics to study inflation dynamics, central-bank credibility, and policy communication. This makes it important to understand what these surveys actually measure.

This study examines the gap between a household’s reported point forecast and the mean of its own reported density. The main finding is that much of this gap comes from the survey design rather than from differences in beliefs or preferences. For respondents whose distributions remain away from the open-ended boundary bins, the average divergence is only +0.099, compared with +0.891 in the full sample. The implied loss-asymmetry index also falls sharply, from 0.374 to 0.057. More importantly, the boundary effects have opposite signs: probability placed at the upper boundary pushes the divergence down, while probability at the lower boundary pushes it up. This pattern is difficult to reconcile with a preference-based explanation alone.

The counterfactual exercises support this interpretation. Using raw bin midpoints leaves almost no divergence for respondents away from the boundaries, while the difference becomes substantial once boundary responses are included. For the roughly 42% of observations affected by the open bounds, the reported divergence therefore depends heavily on how the tails of the distribution are fitted. This is a feature of the elicitation procedure rather than something that can be recovered from household preferences.

A smaller reporting effect remains after removing the boundary problem. Among respondents with comparable beliefs, greater subjective uncertainty is associated with a higher reported point forecast, with an estimated slope of 0.186. The effect also differs across reporting states: the slope is positive for respondents whose point forecast exceeds their density mean and negative for those whose point forecast falls below it. At the annual horizon, the persistent, scale-dependent,

and additive components have standard deviations of 0.451, 1.305, and 1.862 percentage points, respectively. The additive component is the largest, accounting for about 45% of the variation and corresponding to a reporting grid of roughly 6.45 points. This last figure should be read with the caution set out in Section 5.2.7: the additive component is an extrapolated intercept whose value depends on the aggregation rule, and under alternative rules the implied grid ranges from below zero to about nine points. The persistent component also declines substantially as the time gap between observations increases, suggesting that much of the short-run persistence is temporary. These measurement issues matter for aggregate analysis as well. Between 65% and 67% of the cross-sectional variation in reported point forecasts appears to come from reporting components rather than genuine disagreement about future inflation, though this range subtracts  $\hat{\zeta}^2$  and therefore inherits its sensitivity; we report it as an indication of order of magnitude rather than as an estimate. Consistent with this, the density mean forecasts inflation more accurately than the raw point forecasts, with RMSEs of 3.266 and 4.741, respectively. The choice of survey measure also affects subsequent macroeconomic estimates. For example, the estimated informational rigidity is 0.543 using the density mean but only 0.377 using the raw point forecasts. The difference is broadly consistent with the attenuation expected from measurement error.

There are several limitations to keep in mind:

- The additive component parameter  $\zeta^2$  is a boundary limit that the data do not reach. As shown in Section 5.2.7, reasonable changes in aggregation produce estimates ranging from  $-0.47$  to  $+7.15$ , while an interval that allows for modest curvature includes zero. We therefore do not report a percentage share based on  $\hat{\zeta}^2$ . The evidence that rounding is an important part of the residual divergence is more robust: the rounding variance remains between 0.78 and 0.82 across the partitions considered.
- The decomposition is strongest within the interior support. Our model assumes that reporting tendencies are independent of subjective uncertainty. The three components explain 69.9% of the variance in the main interior sample and 88.9% among respondents who use at least three bins. The latter group, however, is selected and reports higher average uncertainty.
- Persistence is measured over a short panel. Respondents are observed for at most twelve months, so the persistent component captures short-run persistence. Longer-lived differences in preferences or reporting behaviour cannot be ruled out.
- The aggregate results are descriptive. The macroeconomic analysis covers 149 months and includes one major inflation episode. Since neither the response format nor subjective uncertainty is randomly assigned, these results should be interpreted as descriptive relationships rather than causal effects.

It is useful to distinguish the findings that are firmly established from those that remain uncertain. The results concerning the survey design are directly supported by the data: the mechanical determination of  $\sigma$  for one- and two-bin responses, the 89% decline in mean divergence after restricting the sample to the interior, the opposite effects of the two open tails, and the bin-midpoint counterfactual. The structural decomposition provides a useful framework, and its qualitative pattern is robust. What the data do not support is a precise percentage for the additive component.

Overall, the evidence points to a simple conclusion: survey design matters considerably for how household inflation expectations are measured. In particular, point forecasts can contain substantial reporting noise that is easily mistaken for differences in beliefs, loss aversion, disagreement, or information rigidity. The algebraic result concerning boundary-based uncertainty should apply more broadly to surveys that construct continuous moments from interval responses, including professional and household expectation surveys. The precise size of the reporting components, however, remains an empirical question for each survey design.

# APPENDICES

## A Proofs

Throughout,  $\mathbb{E}_i$  denotes expectation under the respondent's subjective distribution  $F_i$ ,  $\phi$  and  $\Phi$  are the standard normal density and distribution function, and  $(x)^+ = \max\{x, 0\}$ . In the three-component model Equation (8) the two reporting errors are taken to have mean zero,  $\mathbb{E}[\eta_{it}] = \mathbb{E}[\varepsilon_{it}] = 0$ , which is a normalisation rather than a restriction: any non-zero mean is absorbed into  $\bar{h}$  and into  $\mu_{it}$  respectively.

### Proposition 1

Write  $e = q - \theta$  and define

$$\Psi(q) \equiv a \mathbb{E}_i[(q - \theta)^+] - b \mathbb{E}_i[(\theta - q)^+].$$

Both expectations are finite for every  $q$  whenever  $\mathbb{E}_i|\theta| < \infty$ , since  $0 \leq (q - \theta)^+ \leq |q| + |\theta|$ , so  $\Psi$  is well defined on  $\mathbb{R}$  under the stated moment condition. The map  $q \mapsto \mathbb{E}_i[(q - \theta)^+]$  is convex with right derivative  $F_i(q)$ , and  $q \mapsto \mathbb{E}_i[(\theta - q)^+]$  is convex with right derivative  $-(1 - F_i(q))$ . Hence  $\Psi$  is continuous with right derivative

$$\Psi'(q) = aF_i(q) + b(1 - F_i(q)) \geq \min\{a, b\} > 0,$$

so  $\Psi$  is strictly increasing. Moreover  $\Psi(q) \rightarrow -\infty$  as  $q \rightarrow -\infty$  and  $\Psi(q) \rightarrow +\infty$  as  $q \rightarrow +\infty$ . A unique root  $q_i^*$  therefore exists. When  $\mathbb{E}_i[\theta^2] < \infty$  the objective  $Q(q) = \mathbb{E}_i[L(q - \theta)]$  is finite and strictly convex with  $Q' = 2\Psi$ , so  $q_i^*$  is its unique minimiser; when only  $\mathbb{E}_i|\theta| < \infty$  the same  $q_i^*$  is the unique minimiser of the finite difference  $Q(q) - Q(q_0)$  for any reference point  $q_0$ .

For the representation, add and subtract to obtain the identity

$$\mathbb{E}_i[(q - \theta)^+] - \mathbb{E}_i[(\theta - q)^+] = \mathbb{E}_i[q - \theta] = q - \mathbb{E}_i[\theta].$$

Setting  $\Psi(q_i^*) = 0$  gives  $a\mathbb{E}_i[(q_i^* - \theta)^+] = b\mathbb{E}_i[(\theta - q_i^*)^+]$ , and substituting the identity for the left-hand expectation,

$$a\left\{\mathbb{E}_i[(\theta - q_i^*)^+] + q_i^* - \mathbb{E}_i[\theta]\right\} = b\mathbb{E}_i[(\theta - q_i^*)^+],$$

so that  $a(q_i^* - \mathbb{E}_i[\theta]) = (b - a)\mathbb{E}_i[(\theta - q_i^*)^+]$ . Dividing by  $a$  and using  $\kappa_i = (b_i - a_i)/a_i$  yields Equation (6).

Now let  $F_i$  be  $N(\mu_i, \sigma_i^2)$  and write  $q = \mu_i + h\sigma_i$ . Substituting  $\theta = \mu_i + \sigma_i z$  with  $z$  standard normal,

$$\mathbb{E}_i[(\theta - q)^+] = \sigma_i \int_h^\infty (z - h)\phi(z) dz = \sigma_i [\phi(h) - h(1 - \Phi(h))] = \sigma_i \psi(h),$$

using  $\int_h^\infty z\phi(z)dz = \phi(h)$ . Equation (6) becomes  $h\sigma_i = \kappa_i\sigma_i\psi(h)$ , that is  $h = \kappa_i\psi(h)$ , which is Equation (7).

Finally,  $\psi'(h) = -h\phi(h) - \Phi(-h) + h\phi(h) = -\Phi(-h)$ , so  $G(h, \kappa) \equiv h - \kappa\psi(h)$  has  $\partial G/\partial h = 1 + \kappa\Phi(-h) > 0$ ; since  $G(\cdot, \kappa)$  increases from  $-\infty$  to  $+\infty$  the fixed point is unique. At  $\kappa = 0$  the equation reads  $h = 0$ , so  $h(0) = 0$ . Because  $\psi(h) > 0$  for every  $h$  —  $\psi$  is decreasing with  $\psi(h) \rightarrow 0$  as  $h \rightarrow \infty$  — the implicit function theorem gives

$$h'(\kappa) = -\frac{\partial G/\partial \kappa}{\partial G/\partial h} = \frac{\psi(h)}{1 + \kappa\Phi(-h)} > 0.$$

□

### Corollary 1

By Equation (6),  $w_{it} = q_i^* - \mathbb{E}_i[\theta] = \kappa_i \mathbb{E}_i[(\theta - q_i^*)^+]$ . Both factors are non-negative, so  $w_{it} \geq 0$ . Equality requires either  $\kappa_i = 0$  or  $\mathbb{E}_i[(\theta - q_i^*)^+] = 0$ . In the second case  $\theta \leq q_i^*$  almost surely, while  $w_{it} = 0$  forces  $q_i^* = \mathbb{E}_i[\theta]$ ; a random variable bounded above by its own mean is degenerate, so  $\text{var}_i(\theta) = 0$ . □

### Proposition 2

Write  $w_{it} = h_i \sigma_{it} + \eta_{it} \sigma_{it} + \varepsilon_{it}$ . Condition upon  $\sigma_{it} = \sigma$ . Because  $h_i$ ,  $\eta_{it}$  and  $\varepsilon_{it}$  are mutually independent, each is independent of  $\sigma_{it}$ , and  $\eta$  and  $\varepsilon$  have mean zero,

$$\mathbb{E}[w_{it} \mid \sigma_{it} = \sigma] = \bar{h}\sigma,$$

$$\begin{aligned} \text{var}(w_{it} \mid \sigma_{it} = \sigma) &= \sigma^2 \text{var}(h_i) + \sigma^2 \text{var}(\eta_{it}) + \text{var}(\varepsilon_{it}) \\ &= (\sigma_h^2 + \omega^2) \sigma^2 + \varsigma^2, \end{aligned}$$

which is Equation (9). The conditioning may be dropped from the variances of  $h$ ,  $\eta$  and  $\varepsilon$  precisely because of the independence assumption; this is where that assumption does its work.

For Equation (10), expand the product for  $L \geq 1$ :

$$w_{it} w_{i,t+L} = h_i^2 \sigma_{it} \sigma_{i,t+L} + h_i \sigma_{it} (\eta_{i,t+L} \sigma_{i,t+L} + \varepsilon_{i,t+L}) + (\eta_{it} \sigma_{it} + \varepsilon_{it}) w'_{i,t+L},$$

where  $w'$  collects the remaining terms. Every term containing exactly one factor  $\eta$  or  $\varepsilon$  has expectation zero, because that factor is mean zero and independent of everything multiplying it. Every term containing  $\eta_{it} \eta_{i,t+L}$ ,  $\varepsilon_{it} \varepsilon_{i,t+L}$  or a cross product of the two has expectation zero by serial independence and mean zero, since  $L \geq 1$ . What remains is

$$\mathbb{E}[w_{it} w_{i,t+L}] = \mathbb{E}[h_i^2 \sigma_{it} \sigma_{i,t+L}] = \mathbb{E}[h_i^2] \mathbb{E}[\sigma_{it} \sigma_{i,t+L}] = (\bar{h}^2 + \sigma_h^2) \mathbb{E}[\sigma_{it} \sigma_{i,t+L}],$$

the middle equality using independence of  $h_i$  from the uncertainty path.

Finally  $\mathbb{E}[w_{it}] = \mathbb{E}[h_i \sigma_{it}] + \mathbb{E}[\eta_{it} \sigma_{it}] + \mathbb{E}[\varepsilon_{it}] = \bar{h} \mathbb{E}[\sigma_{it}]$ , which gives  $\bar{h} = \mathbb{E}[w_{it}] / \mathbb{E}[\sigma_{it}]$  whenever  $\mathbb{E}[\sigma_{it}] \neq 0$ . Identification then follows: Equation (9) is affine in  $\sigma^2$  with slope  $\sigma_h^2 + \omega^2$  and intercept  $\varsigma^2$ , and Equation (10) is proportional to  $\mathbb{E}[\sigma_{it} \sigma_{i,t+L}]$  with factor  $\bar{h}^2 + \sigma_h^2$ ; given  $\bar{h}$ , the two slopes separate  $\sigma_h^2$  from  $\omega^2$ . □

*Remark.* The intercept in Equation (9) is a limit as  $\sigma \downarrow 0$  and is therefore identified only if the support of  $\sigma_{it}$  accumulates at zero. It does not: Section 5.2.7 shows that the smallest recorded  $\sigma$  is 0.433. What the data identify is the restriction of Equation (9) to the observed range of  $\sigma^2$ ;  $\varsigma^2$  is recovered only by extrapolating the affine form outside it, and is therefore identified relative to the maintained assumption of exact linearity rather than by the moments alone.

### Proposition 3

Replace the fixed  $h_i$  by the stationary first-order process  $h_{it} = \varphi h_{i,t-1} + \nu_{it}$  with  $|\varphi| < 1$ , where  $\nu_{it}$  is serially uncorrelated with  $\mathbb{E}[\nu_{it}] = (1 - \varphi) \bar{h}$  and  $\text{var}(\nu_{it}) = (1 - \varphi^2) \sigma_h^2$ , so that the stationary distribution has mean  $\bar{h}$  and variance  $\sigma_h^2$  as before. Iterating,

$$h_{i,t+L} = \varphi^L h_{it} + \sum_{j=0}^{L-1} \varphi^j \nu_{i,t+L-j},$$

and since the innovations dated after  $t$  are uncorrelated with  $h_{it}$ ,

$$\text{cov}(h_{it}, h_{i,t+L}) = \varphi^L \text{var}(h_{it}) = \varphi^L \sigma_h^2, \quad \mathbb{E}[h_{it} h_{i,t+L}] = \bar{h}^2 + \varphi^L \sigma_h^2.$$

Repeating the argument of Proposition 2, all terms involving  $\eta$  or  $\varepsilon$  drop out for  $L \geq 1$ , leaving

$$\mathbb{E}[w_{it} w_{i,t+L}] = \mathbb{E}[h_{it} h_{i,t+L} \sigma_{it} \sigma_{i,t+L}] = (\bar{h}^2 + \varphi^L \sigma_h^2) \mathbb{E}[\sigma_{it} \sigma_{i,t+L}],$$

where the factorisation again uses independence of the  $h$  process from the uncertainty path. Proposition 2 is the case  $\varphi = 1$ . Since  $\varphi^L \rightarrow 0$ , the ratio  $\mathbb{E}[w_{it} w_{i,t+L}] / \mathbb{E}[\sigma_{it} \sigma_{i,t+L}]$  declines from  $\bar{h}^2 + \sigma_h^2$  towards  $\bar{h}^2$ , and its profile in  $L$  identifies  $\varphi$ .  $\square$

#### Proposition 4

Let  $\zeta_{it} = h_i + \eta_{it}$ , so  $w_{it} = \zeta_{it} \sigma_{it} + \varepsilon_{it}$ . Both  $\zeta_{it}$  and  $\varepsilon_{it}$  are independent of  $\sigma_{it}$ , and  $\varepsilon$  has mean zero. Hence for any event  $A$  defined on  $\zeta$  alone,

$$\mathbb{E}[w_{it} \mid \sigma_{it} = \sigma, \zeta_{it} \in A] = \sigma \mathbb{E}[\zeta_{it} \mid \zeta_{it} \in A] + \mathbb{E}[\varepsilon_{it}] = \sigma \mathbb{E}[\zeta \mid \zeta \in A],$$

which is linear in  $\sigma$  with slope  $\mathbb{E}[\zeta \mid \zeta \in A]$ . Taking  $A = \{\zeta > 0\}$  and  $A = \{\zeta < 0\}$  gives the two displayed derivatives, and the signs are immediate because a variable conditioned to be positive has positive mean and one conditioned to be negative has negative mean, provided each event has positive probability.

For the comparison of magnitudes, suppose  $\zeta = \bar{h} + s$  with  $s$  symmetric about zero and  $\Pr(s = 0) = 0$ , and write  $m = \mathbb{E}[s \mid s > 0] = -\mathbb{E}[s \mid s < 0] > 0$ . Then

$$\mathbb{E}[\zeta \mid \zeta > 0] = \bar{h} + \mathbb{E}[s \mid s > -\bar{h}], \quad \mathbb{E}[\zeta \mid \zeta < 0] = \bar{h} + \mathbb{E}[s \mid s < -\bar{h}].$$

As  $\bar{h}/\text{sd}(s) \rightarrow 0$  these converge to  $\bar{h} + m$  and  $\bar{h} - m$ , whose sum is  $2\bar{h}$  and whose magnitudes therefore differ by  $O(\bar{h})$  while each is of order  $m$ . The two slopes are thus approximately equal and opposite when the dispersion of  $\zeta$  is large relative to  $|\bar{h}|$ .  $\square$

*Remark.* The conditioning event in the proposition is defined on the latent  $\zeta$ , whereas the empirical counterpart in Section 5.1.1 conditions on the observed sign of  $w$ . The two coincide only in the limit  $\zeta^2/\sigma^2 \rightarrow 0$ ; for finite  $\zeta^2$ , conditioning on  $\text{sign}(w)$  selects on  $\varepsilon$  as well and the conditional slope is attenuated towards zero, by an amount that shrinks as  $\sigma$  grows. The qualitative prediction of opposite signs is unaffected, since  $\varepsilon$  is symmetric about zero and independent of  $\sigma$ .

#### Proposition 5

Suppose the fitted density is supported on an interval of width  $W$  and has the form  $f(x) = W^{-1}g((x - c)/W)$  for a fixed shape  $g$  on  $[0, 1]$  with  $\int_0^1 g = 1$ . Substituting  $x = c + Wu$ ,

$$\mathbb{E}[X] = c + W \int_0^1 u g(u) du, \quad \text{var}(X) = W^2 \left[ \int_0^1 u^2 g(u) du - \left( \int_0^1 u g(u) du \right)^2 \right] \equiv W^2 v_g.$$

The standard deviation is therefore  $W\sqrt{v_g}$ : it is invariant to the location  $c$  and proportional to the width of the reported support, with a constant of proportionality determined by the shape alone.

Two cases specialise this. If all mass lies in one bin of width  $\Delta$  and  $g$  is uniform,  $v_g = 1/12$  and  $\sigma = \Delta/\sqrt{12}$ . If mass is divided equally between two adjacent bins of combined width  $W$  and  $g$  is the symmetric triangular shape on  $[0, 1]$ , then  $v_g = 1/24$  and  $\sigma = W/\sqrt{24}$ ; the symmetric triangular density on an interval of length  $W$  has variance  $W^2/24$ , which follows from  $v_g = \int_0^1 u^2 g - \frac{1}{4}$  with  $g(u) = 4u$  on  $[0, \frac{1}{2}]$  and  $g(u) = 4(1 - u)$  on  $[\frac{1}{2}, 1]$ , giving  $\int_0^1 u^2 g(u) du = \frac{7}{24}$  and  $v_g = \frac{7}{24} - \frac{6}{24} = \frac{1}{24}$ . In both cases the reported probabilities enter only through which bins are selected: in the one-bin case the probability is necessarily one, and in the two-bin case the equal split fixes the shape. The recorded  $\sigma$  therefore carries no information about the respondent's beliefs beyond the identity of the bins used.  $\square$

## Proposition 6

Write  $q^*/\Delta_i = n + u$  with  $n$  an integer and  $u \in [-\frac{1}{2}, \frac{1}{2})$ , so that  $\text{round}_{\Delta_i}(q^*) = \Delta_i n$  and

$$\varepsilon_{it} = \Delta_i n - q^* = -\Delta_i u.$$

If  $u$  is uniform on  $[-\frac{1}{2}, \frac{1}{2})$  and independent of  $\Delta_i$ , then  $\mathbb{E}[\varepsilon \mid \Delta_i] = 0$  and  $\text{var}(\varepsilon \mid \Delta_i) = \Delta_i^2/12$ , whence by the law of total variance

$$\zeta^2 = \mathbb{E}[\varepsilon^2] = \mathbb{E}[\Delta_i^2/12] = \frac{\mathbb{E}[\Delta_i^2]}{12}.$$

The uniformity of  $u$  is exact, rather than approximate, under a band-limitation: if the characteristic function of  $q^*$  vanishes outside  $(-\pi/\Delta, \pi/\Delta)$ , the quantisation residual is exactly uniform on  $[-\Delta/2, \Delta/2)$  and independent of  $q^*$ . This is the quantisation theorem of Widrow et al. (1996). Absent band-limitation the statement is approximate, with an error governed by the aliased tail  $\sum_{m \neq 0} \varphi_{q^*}(2\pi m/\Delta)$ , which is negligible when  $q^*$  is dispersed over many grid points relative to  $\Delta$ . Independence of  $\varepsilon$  from  $\sigma_i$  follows because the grid  $\Delta_i$  is assumed not to depend on  $\sigma_i$  and, given  $\Delta_i$ , the residual  $u$  is uniform irrespective of the location of  $q^*$ . Finally, conditioning on the grid,

$$\text{var}(w_{it} \mid \sigma_{it} = \sigma, \Delta_i = \Delta) = (\sigma_h^2 + \omega^2)\sigma^2 + \frac{\Delta^2}{12},$$

which is strictly increasing in  $\Delta$  at every  $\sigma$ , giving the second testable prediction.  $\square$

## Proposition 7

Average Equation (8) across the  $N_t$  respondents observed in month  $t$ :

$$\bar{\pi}_t^e = \bar{\mu}_t + \frac{1}{N_t} \sum_i h_i \sigma_{it} + \frac{1}{N_t} \sum_i \eta_{it} \sigma_{it} + \frac{1}{N_t} \sum_i \varepsilon_{it}.$$

The last two averages have mean zero and variance of order  $N_t^{-1}$ , so they vanish in probability as  $N_t \rightarrow \infty$ . Writing the cross-sectional average of a product as a covariance plus the product of averages,  $N_t^{-1} \sum_i h_i \sigma_{it} = \bar{h} \bar{\sigma}_t + \text{cov}_i(h_i, \sigma_{it})$ , which gives the first display.

For the second,  $\pi_{it}^e = \mu_{it} + w_{it}$  identically, so the cross-sectional variance decomposes as:

$$\text{var}_i(\pi_{it}^e) = \text{var}_i(\mu_{it}) + \text{var}_i(w_{it}) + 2\text{cov}_i(\mu_{it}, w_{it}),$$

which is an algebraic identity requiring no assumption.  $\square$

*Remark.* It is worth being explicit about what the two displays do and do not say, because the natural reading of the first is wrong. Mean-zero reporting errors vanish from the cross-sectional *mean*, as the first display shows, but they do not vanish from the cross-sectional *variance*: they enter  $\text{var}_i(w_{it})$ , and conditioning on the uncertainty path that term contains  $\omega^2 \mathbb{E}_i[\sigma_{it}^2] + \zeta^2$  purely from reporting. That is precisely why measured disagreement overstates genuine disagreement, which is the use made of the proposition in Section 6.

## Proposition 8

Let  $R_t^\mu = F_t^\mu - F_{t-k}^\mu$  and  $R_t^{\pi^e} = F_t^{\pi^e} - F_{t-k}^{\pi^e}$ , and write  $R_t^{\pi^e} = R_t^\mu + \xi_t$  where  $\xi_t$  is the additional variation carried by the point-forecast series. The population regression coefficients from projecting  $y_t$  upon each regressor are:

$$\beta^\mu = \frac{\text{cov}(y_t, R_t^\mu)}{\text{var}(R_t^\mu)}, \quad \beta^{\pi^e} = \frac{\text{cov}(y_t, R_t^{\pi^e})}{\text{var}(R_t^{\pi^e})}.$$

The maintained condition is  $\text{cov}(y_t, \xi_t) = 0$ , so  $\text{cov}(y_t, R_t^{\pi^e}) = \text{cov}(y_t, R_t^\mu)$  and therefore

$$\frac{\beta^{\pi^e}}{\beta^\mu} = \frac{\text{cov}(y_t, R_t^{\pi^e})}{\text{cov}(y_t, R_t^\mu)} \cdot \frac{\text{var}(R_t^\mu)}{\text{var}(R_t^{\pi^e})} = \frac{\text{var}(R_t^\mu)}{\text{var}(R_t^{\pi^e})},$$

which is Equation (11), valid whenever  $\beta^\mu \neq 0$ . The ratio is less than one, so the coefficient is attenuated, exactly when  $\text{var}(R_t^{\pi^e}) > \text{var}(R_t^\mu)$ ; a sufficient condition is that  $\xi_t$  be uncorrelated with  $R_t^\mu$ , in which case  $\text{var}(R_t^{\pi^e}) = \text{var}(R_t^\mu) + \text{var}(\xi_t)$  and the attenuation factor is  $\text{var}(R^\mu)/[\text{var}(R^\mu) + \text{var}(\xi)]$ .  $\square$

### Proposition 9

Since  $\pi_{it}^e - \theta_t = (\mu_{it} - \theta_t) + w_{it}$  identically,

$$\mathbb{E}[(\pi^e - \theta)^2] = \mathbb{E}[(\mu - \theta)^2] + 2\mathbb{E}[(\mu - \theta)w] + \mathbb{E}[w^2],$$

so that

$$\text{MSE}(\pi^e) - \text{MSE}(\mu) = \mathbb{E}[w^2] + 2\mathbb{E}[(\mu - \theta)w]. \quad (32)$$

*Remark.* The second term is a cross-moment and not a covariance. The two differ by  $\mathbb{E}[\mu - \theta] \mathbb{E}[w]$ , since  $\mathbb{E}[(\mu - \theta)w] = \text{cov}(\mu - \theta, w) + \mathbb{E}[\mu - \theta] \mathbb{E}[w]$ , and they coincide only if the density mean is unbiased for realised inflation or if  $\mathbb{E}[w] = 0$ . Neither holds here, because  $\mathbb{E}[w] = \bar{h}\mathbb{E}[\sigma] > 0$ . Equation (32) is therefore the form we use; the difference is second order for the comparison drawn in Section 5.3, but the covariance form is not an identity and should not be written as one.

For the covariance term itself, substitute  $w = h\sigma + \eta\sigma + \varepsilon$ . Because  $\eta$  and  $\varepsilon$  have mean zero and are independent of  $(\mu - \theta, \sigma)$ ,

$$\text{cov}(\mu - \theta, \eta\sigma) = \mathbb{E}[\eta] \mathbb{E}[(\mu - \theta)\sigma] - 0 = 0, \quad \text{cov}(\mu - \theta, \varepsilon) = 0,$$

while independence of  $h$  from  $(\mu - \theta, \sigma)$  gives

$$\text{cov}(\mu - \theta, h\sigma) = \bar{h} \mathbb{E}[(\mu - \theta)\sigma] - \mathbb{E}[\mu - \theta] \bar{h} \mathbb{E}[\sigma] = \bar{h} \text{cov}(\mu - \theta, \sigma).$$

Hence  $\text{cov}(\mu - \theta, w) = \bar{h} \text{cov}(\mu - \theta, \sigma)$ , as stated.

Finally, by Proposition 2,

$$\mathbb{E}[w^2 \mid \sigma] = \text{var}(w \mid \sigma) + \mathbb{E}[w \mid \sigma]^2 = (\sigma_h^2 + \omega^2 + \bar{h}^2)\sigma^2 + \varsigma^2,$$

which is strictly increasing in  $\sigma$  whenever  $\sigma_h^2 + \omega^2 + \bar{h}^2 > 0$ . The accuracy advantage of the elicited mean over the reported point forecast therefore widens with subjective uncertainty, which is the comparison reported in Section 5.3.  $\square$

## B An Alternative Variance Estimator

Two alternative approaches are commonly used to decompose the divergence, and we report both as a useful comparison with the level-based approach in Proposition 2. The first normalises the divergence by respondents' own uncertainty, defining

$$r_{it} \equiv \frac{w_{it}}{\sigma_{it}}.$$

Under the reporting model in Equation (8), its variance and within-person autocovariance are

$$\text{var}(r_{it}) = \sigma_h^2 + \omega^2 + \varsigma^2 \mathbb{E}[\sigma_{it}^{-2}], \quad \text{cov}(r_{it}, r_{i,t+L}) = \sigma_h^2. \quad (33)$$

The usual implementation effectively sets  $\varsigma^2 = 0$ , allowing the autocovariance to identify  $\sigma_h^2$  and treating the remaining variation as  $\omega^2$ .

The second approach uses the cross-sectional dispersion of the divergence. Specifically, we estimate

$$\widehat{\text{sd}}_b(w) = c_0 + c_1 \widehat{\mathbb{E}}_b[\sigma] + \xi_b, \quad (34)$$

where the intercept is interpreted as  $\varsigma$  and the slope as  $\{\sigma_h^2 + \omega^2\}^{1/2}$ .

Table A1 reports estimates from both approaches for the baseline sample and compares them with the level-based estimates in Table 13. This comparison highlights how the choice of normalization affects the decomposition when additive reporting error is present.

Table A1: Comparison of Alternative Estimators Against Level Moments

|                                           | Normalised<br>Autocovariance | Line Fitted to<br>$\text{sd}(w \mid \sigma)$ | Level Moments<br>(This Paper) |
|-------------------------------------------|------------------------------|----------------------------------------------|-------------------------------|
| $\text{var}(r)$ on the trimmed sample     | 1.9926                       |                                              |                               |
| $\text{cov}(r_{it}, r_{i,t+L}), L \leq 3$ | 0.3965                       |                                              |                               |
| Fitted intercept                          |                              | 1.731                                        | 5.021 in $\sigma^2$           |
| Fitted slope                              |                              | 1.164                                        | 1.881 in $\sigma^2$           |
| $R^2$                                     |                              | 0.969                                        | 0.789                         |
| Implied $\text{sd}(h_i)$                  | 0.630                        |                                              | 0.579                         |
| Implied $\omega$                          | 1.263                        |                                              | 1.243                         |
| Implied $\varsigma$                       | 0 by construction            | 1.731                                        | 2.241                         |
| Implied persistent share, per cent        | 19.9                         |                                              | 15.5                          |

*Note:* This table compares the two common estimators in the literature, Equation (33) and Equation (34), evaluated on the baseline sample against the level moments from Proposition 2. The normalised device requires trimming the tail of  $r = w/\sigma$  at the 1st and 99th percentiles, resulting in a final sample of 170,608 respondent-months.

Overall, Table A1 shows why we do not use either alternative decomposition in the main analysis. The normalised approach attributes 19.9 per cent of the divergence to the persistent component and, by construction, assigns no variation to additive reporting error. The conditional-dispersion approach instead gives  $\varsigma = 1.731$ , compared with 2.241 from the level-based estimates. The main problem with the normalised approach is that, when  $\varsigma^2 > 0$ ,

$$\text{var}(r_{it}) - \text{cov}(r_{it}, r_{i,t+L}) = \omega^2 + \varsigma^2 \mathbb{E}[\sigma_{it}^{-2}].$$

Thus, the difference does not identify  $\omega^2$  alone. The additional term depends on  $\mathbb{E}[\sigma^{-2}]$ , which is driven largely by observations with very small reported uncertainty. Trimming these observations changes both the estimate and the observations available for the autocovariance calculation.

The second approach also has a limitation. Since

$$\text{sd}(w \mid \sigma) = \{(\sigma_h^2 + \omega^2)\sigma^2 + \varsigma^2\}^{1/2}$$

is concave in  $\sigma$ , a linear fit understates the underlying slope:  $\sqrt{1.881} = 1.372$ , compared with the fitted value of 1.164 in the baseline sample.

For these reasons, the main analysis relies on the level-based moments in Equation (9) and Equation (10). They avoid the inverse-uncertainty problem and provide a direct linear relationship in  $\sigma^2$  and  $\mathbb{E}[\sigma_{it}\sigma_{i,t+L}]$ .

## References

- Andrade, P. and Le Bihan, H. (2013), “Inattentive professional forecasters”, *Journal of Monetary Economics*, 60(8): 967–982.
- Armantier, O., Bruine de Bruin, W., Potter, S., Topa, G., van der Klaauw, W. and Zafar, B. (2013), “Measuring inflation expectations”, *Annual Review of Economics*, 5(1): 273–301.

- Armantier, O., Bruine de Bruin, W., Topa, G., van der Klaauw, W. and Zafar, B. (2015), “Inflation expectations and behavior: do survey respondents act on their beliefs?”, *International Economic Review*, 56(2): 505–536.
- Armantier, O., Topa, G., van der Klaauw, W. and Zafar, B. (2017), “An overview of the Survey of Consumer Expectations”, *Federal Reserve Bank of New York Economic Policy Review*, 23(2): 51–72.
- Bachmann, R., Berg, T.O. and Sims, E.R. (2015), “Inflation expectations and readiness to spend: cross-sectional evidence”, *American Economic Journal: Economic Policy*, 7(1): 1–35.
- Binder, C. (2017), “Measuring uncertainty based on rounding: new method and application to inflation expectations”, *Journal of Monetary Economics*, 90: 1–12.
- Blinder, A.S., Ehrmann, M., Fratzscher, M., de Haan, J. and Jansen, D.-J. (2008), “Central bank communication and monetary policy: a survey of theory and evidence”, *Journal of Economic Literature*, 46(4): 910–945.
- Bordalo, P., Gennaioli, N., Ma, Y. and Shleifer, A. (2020), “Overreaction in macroeconomic expectations”, *American Economic Review*, 110(9): 2748–2782.
- Branch, W.A. (2004), “The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations”, *Economic Journal*, 114(497): 592–621.
- Bruine de Bruin, W., van der Klaauw, W., Downs, J.S., Fischhoff, B., Topa, G. and Armantier, O. (2010), “Expectations of inflation: the role of demographic variables, expectation formation, and financial literacy”, *Journal of Consumer Affairs*, 44(2): 381–402.
- Bryan, M.F. and Venkatu, G. (2001), “The curiously different inflation perspectives of men and women”, *Federal Reserve Bank of Cleveland Economic Commentary*.
- Carroll, C.D. (2003), “Macroeconomic expectations of households and professional forecasters”, *Quarterly Journal of Economics*, 118(1): 269–298.
- Cavallo, A., Cruces, G. and Perez-Truglia, R. (2017), “Inflation expectations, learning, and supermarket prices: evidence from survey experiments”, *American Economic Journal: Macroeconomics*, 9(3): 1–35.
- Christoffersen, P.F. and Diebold, F.X. (1997), “Optimal prediction under asymmetric loss”, *Econometric Theory*, 13(6): 808–817.
- Clements, M.P. (2010), “Explanations of the inconsistencies in survey respondents’ forecasts”, *European Economic Review*, 54(4): 536–549.
- Clements, M.P. (2014), “Probability distributions or point predictions? Survey forecasts of US output growth and inflation”, *International Journal of Forecasting*, 30(1): 99–117.
- Coibion, O. and Gorodnichenko, Y. (2012), “What can survey forecasts tell us about information rigidities?”, *Journal of Political Economy*, 120(1): 116–159.
- Coibion, O. and Gorodnichenko, Y. (2015a), “Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation”, *American Economic Journal: Macroeconomics*, 7(1): 197–232.
- Coibion, O. and Gorodnichenko, Y. (2015b), “Information rigidity and the expectations formation process: a simple framework and new facts”, *American Economic Review*, 105(8): 2644–2678.
- Coibion, O., Gorodnichenko, Y. and Kamdar, R. (2018), “The formation of expectations, inflation, and the Phillips curve”, *Journal of Economic Literature*, 56(4): 1447–1491.
- Coibion, O., Gorodnichenko, Y. and Kumar, S. (2018), “How do firms form their expectations? New survey evidence”, *American Economic Review*, 108(9): 2671–2713.
- Coibion, O., Gorodnichenko, Y., Kumar, S. and Pedemonte, M. (2020), “Inflation expectations as a policy tool?”, *Journal of International Economics*, 124: 1–27.

- Coibion, O., Gorodnichenko, Y. and Weber, M. (2022), “Monetary policy communications and their effects on household inflation expectations”, *Journal of Political Economy*, 130(6): 1537–1584.
- D’Acunto, F., Malmendier, U., Ospina, J. and Weber, M. (2021), “Exposure to grocery prices and inflation expectations”, *Journal of Political Economy*, 129(5): 1615–1639.
- Delavande, A. and Rohwedder, S. (2011), “Individuals’ uncertainty about future Social Security benefits and portfolio choice”, *Journal of Applied Econometrics*, 26(3): 498–519.
- Elliott, G., Komunjer, I. and Timmermann, A. (2005), “Estimation and testing of forecast rationality under flexible loss”, *Review of Economic Studies*, 72(4): 1107–1125.
- Engelberg, J., Manski, C.F. and Williams, J. (2009), “Comparing the point predictions and subjective probability distributions of professional forecasters”, *Journal of Business and Economic Statistics*, 27(1): 30–41.
- Giustinelli, P., Manski, C.F. and Molinari, F. (2022), “Tail and center rounding of probabilistic expectations in the Health and Retirement Study”, *Journal of Econometrics*, 231(1): 265–281.
- Granger, C.W.J. (1969), “Prediction with a generalized cost of error function”, *Journal of the Operational Research Society*, 20(2): 199–207.
- Kumar, S., Afrouzi, H., Coibion, O. and Gorodnichenko, Y. (2015), “Inflation targeting does not anchor inflation expectations: evidence from firms in New Zealand”, *Brookings Papers on Economic Activity*, 46(2): 151–225.
- Maćkowiak, B. and Wiederholt, M. (2009), “Optimal sticky prices under rational inattention”, *American Economic Review*, 99(3): 769–803.
- Malmendier, U. and Nagel, S. (2016), “Learning from inflation experiences”, *Quarterly Journal of Economics*, 131(1): 53–87.
- Mankiw, N.G. and Reis, R. (2002), “Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve”, *Quarterly Journal of Economics*, 117(4): 1295–1328.
- Mankiw, N.G. and Reis, R. (2010), “Imperfect information and aggregate supply”, in B.M. Friedman and M. Woodford (eds.), *Handbook of Monetary Economics*, volume 3, chapter 5, Elsevier, 183–229.
- Mankiw, N.G., Reis, R. and Wolfers, J. (2003), “Disagreement about inflation expectations”, *NBER Macroeconomics Annual*, 18: 209–248.
- Manski, C.F. (2004), “Measuring expectations”, *Econometrica*, 72(5): 1329–1376.
- Manski, C.F. (2018), “Survey measurement of probabilistic macroeconomic expectations: progress and promise”, *NBER Macroeconomics Annual*, 32(1): 411–471.
- Manski, C.F. and Molinari, F. (2010), “Rounding probabilistic data”, *Journal of Business and Economic Statistics*, 28(2): 219–231.
- Newey, W.K. and Powell, J.L. (1987), “Asymmetric least squares estimation and testing”, *Econometrica*, 55(4): 819–847.
- Patton, A.J. and Timmermann, A. (2010), “Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion”, *Journal of Monetary Economics*, 57(7): 803–820.
- Pesaran, M.H. and Weale, M. (2006), “Survey expectations”, in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland.
- Pfajfar, D. and Žakelj, B. (2014), “Experimental evidence on inflation expectation formation”, *Journal of Economic Dynamics and Control*, 44: 147–168.
- Reis, R. (2006), “Inattentive producers”, *Review of Economic Studies*, 73(3): 793–821.
- Sims, C.A. (2003), “Implications of rational inattention”, *Journal of Monetary Economics*, 50(3): 665–690.

- Weber, M., D’Acunto, F., Gorodnichenko, Y. and Coibion, O. (2022), “The subjective inflation expectations of households and firms: measurement, determinants, and implications”, *Journal of Economic Perspectives*, 36(3): 157–184.
- Widrow, B., Kollár, I. and Liu, M.-C. (1996), “Statistical theory of quantization”, *IEEE Transactions on Instrumentation and Measurement*, 45(2): 353–361.
- Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press.
- Zellner, A. (1986), “Bayesian estimation and prediction using asymmetric loss functions”, *Journal of the American Statistical Association*, 81(394): 446–451.