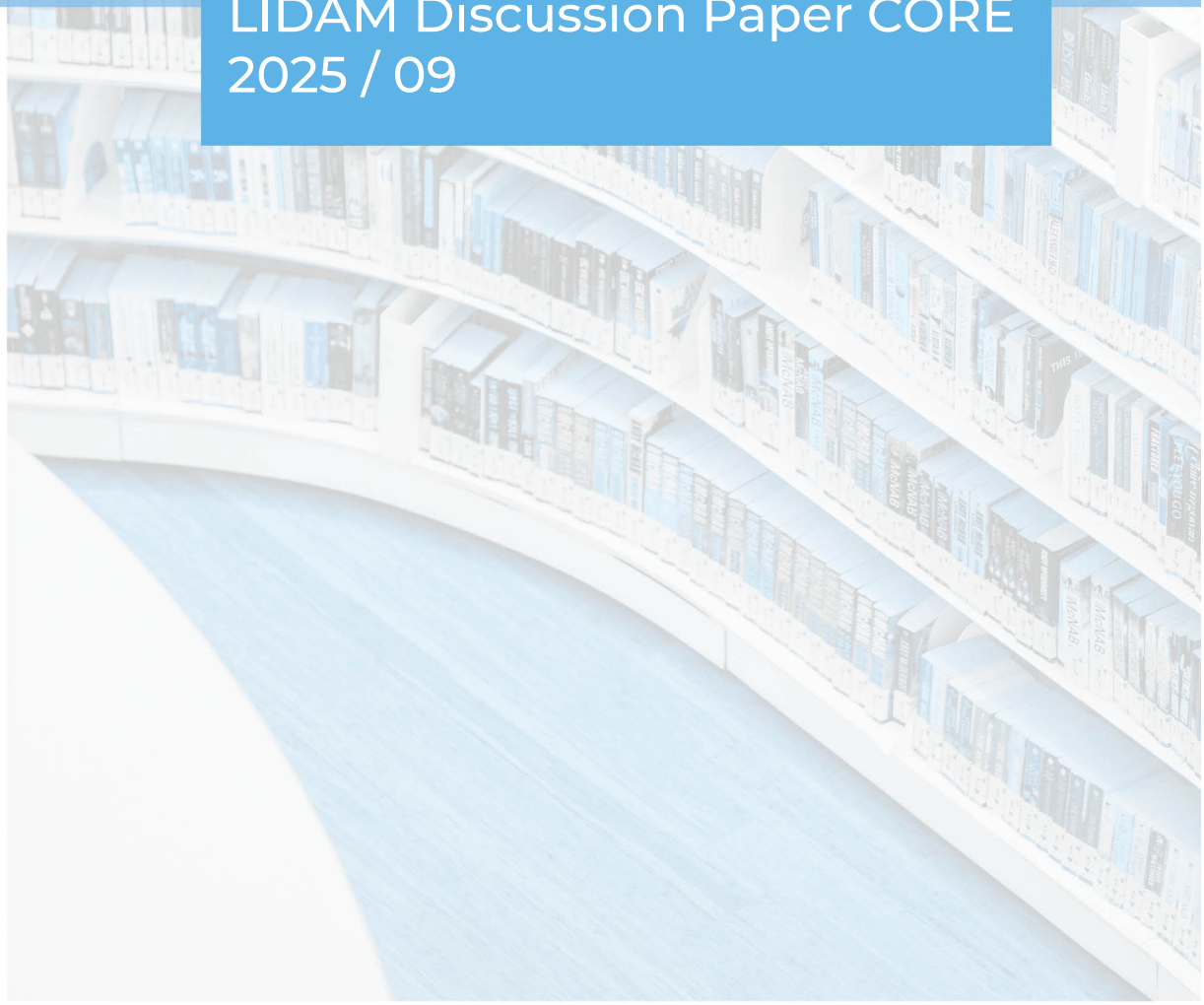
UNCOVERING ATTENTION HETEROGENEITY

Jérémy Boccanfuso, Luca Neri

LIDAM Discussion Paper CORE
2025 / 09



UCLouvain

CORE | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/core-discussion-papers

Uncovering attention heterogeneity*

Jérémy Boccanfuso
University of Bologna

Luca Neri
UCLouvain

April 10, 2025

Abstract

Heterogeneity in individuals' attention to news shapes how aggregate shocks propagate through the economy. We propose a novel method to uncover the full distribution of attention in surveys of consumers and professional forecasters, by developing a finite mixture model to cluster forecasts according to their level of attention. Our results reveal substantial heterogeneity in attention, which, when ignored, leads to an overestimation of average attention. Leveraging our clustering, we construct decade-long panel datasets on attention and provide new insights into its drivers. We document an asymmetric effect of aggregate shocks on the distribution of attention and assess how attention varies with socioeconomic characteristics. Finally, we identify systematic deviations from Bayesian updating and related theories designed to account for behavioral biases. In particular, we find that individuals' attention is less sensitive to changes in uncertainty and signal precision than predicted by these theories. Overall, this paper opens new avenues by providing the first panel datasets on attention and illustrating how they can inform theories of imperfect information through within-individual variations.

JEL codes: D83, D84, E31, C23.

Keywords: attention, information frictions, heterogeneity, expectations.

*Contacts: jeremy.boccanfuso@unibo.it and luca.neri@uclouvain.be. We are grateful to Gaetano Gaballo, Erwan Gauthier, Luca Gemmi, Paul Hubert, Daniel Lewis, Francesca Monti, Jérémi Montornès and seminar participants at Banque de France, UCLouvain, and Unibo for extensive comments. We thank Romain Lebeau for great research assistance.

1 Introduction

Attention shapes how individuals incorporate news into their expectations and respond to changes in the economic environment (Sims, 2003; Maćkowiak et al., 2023). Measuring attention, an unobservable trait, is challenging. This difficulty has spurred a growing literature developing methods to infer attention from forecast disagreement (Coibion and Gorodnichenko, 2012; Andrade and Le Bihan, 2013), forecast sluggishness (Coibion and Gorodnichenko, 2015; Goldstein, 2023), information experiments (Armantier et al., 2016; Haaland et al., 2023), internet trends (Korenok et al., 2023; Pfäuti, 2023), and open-ended surveys (Link et al., 2025). This research has largely focused on estimating average attention, providing little systematic evidence on its cross-sectional heterogeneity.

Documenting this heterogeneity is crucial for several reasons. First, as highlighted by the growing interest in heterogeneous agent models in macroeconomics, cross-sectional heterogeneity hold major implications for the effectiveness of macroeconomic policies. In this regard, we show that attention heterogeneity is no exception, as it shapes the transmission of aggregate shocks to the economy. Second, understanding how this heterogeneity correlates with observable characteristics is essential for policymakers to design targeted communication strategies that reach individuals with varying levels of attention to economic news. Finally, identifying variations in attention both across and within individuals provides valuable insights for theories of expectation formation.

Existing insights into attention heterogeneity have been typically obtained from grouping forecasts based on observable characteristics. For instance, Goldstein (2023) proposes to group forecasts by date to retrieve variation of attention over time, or to group forecasts by forecaster to retrieve variation of attention across individuals.[1] While insightful, this approach requires taking a stance, *ex ante*, on the potential drivers of attention heterogeneity (here, either time or individuals). Furthermore, it cannot uncover the full extent of attention heterogeneity, as it ignores differences within the predefined groups.

We take a different approach. Rather than conditioning on specific observables, we uncover the *unconditional* heterogeneity of attention in the ECB Survey of Professional Forecasters (ECB SPF) and the US Survey of Consumer Expectations (US SCE). To achieve this, we rely on finite mixtures to endogenously cluster forecasts into a finite num-

[1]Similarly, some authors group observations based on shared characteristics, such as gender, education, and geography (e.g., Weber et al., 2023; Gemmi and Mihet, 2024).

ber of homogeneous groups, for which we estimate attention.[2] This clustering approach does not require taking a stance on the conditional drivers of attention heterogeneity. Instead, it leverages parametric assumptions from information models, allowing us to cluster forecasts based on their likelihood of belonging to each group. The clustering and parameter estimation are performed jointly.

Ultimately, this approach allows us to recover the expected heterogeneity of attention across and within individuals. We find that attention heterogeneity is sizable, and that grouping forecasts over time and across individuals ignores about 71% to 94% of this heterogeneity. Moreover, we obtain a forecast-specific measure of attention, which enables a detailed investigation of the drivers of attention heterogeneity *ex post*. To our knowledge, this is the first instance where a panel of attention data has been constructed, allowing us to apply econometric methods for panel data to study the drivers of attention from within-individual variations.

More specifically, we estimate attention from the weight placed on news when revising forecasts. This approach extends the seminal work of Coibion and Gorodnichenko (2015) to individual-level forecasts, and is used in Goldstein (2023) and Gemmi and Valchev (2023). Its advantage lies in being grounded in theories of information frictions (Lucas, 1972; Sims, 2003) and behavioral biases (Bordalo et al., 2020; Angeletos et al., 2021; Broer and Kohlhas, 2022). Our first main contribution is to extend this framework to make it robust to, and capable of estimating, unconstrained heterogeneity in attention.

The standard approach involves obtaining a measure of news from the difference between prior forecasts and the average forecast, which is then regressed onto forecast revisions. In this specification, the average forecast serves as a proxy for common information. As we demonstrate, this method does not consistently estimate mean attention in the population when attention is heterogeneous. The challenge with heterogeneity arises because the average forecast reflects both common information and average attention. In contrast, we avoid using average forecasts and directly estimate common information alongside attention. This allows for heterogeneous attention by estimating multiple slopes, one for each attention level, in an otherwise standard linear regression of forecast revisions on the difference between prior forecasts and common information

[2]See McLachlan et al. (2019) for an introduction to finite mixture models. In macroeconomics, Lewis et al. (2024) rely on finite mixture to uncover latent heterogeneity in the marginal propensity to consume.

(the news)—which we estimate with a Gaussian mixture. Importantly, our approach remains robust to individual-level (time-invariant) unobserved heterogeneity, letting us isolate genuine attention differences.

We find substantial heterogeneity among both consumers and professional forecasters. Using information criteria and K-fold cross-validation, we estimate between 5 and 24 distinct attention groups, depending on the dataset considered.[3] The attention weights placed on news span the entire unit interval. Allowing for heterogeneity also has quantitative implications for estimating average attention. Comparing our average attention estimate to one that assumes homogeneity (a single group), we find that the former is systematically lower—often by a substantial margin—suggesting that ignoring heterogeneity leads to an overstatement of average attention.

Our second main contribution is the development and utilization of decade-long panel data on attention. To achieve this, we compute a measure of expected attention for each forecast by leveraging the posterior probabilities of their assignment to the estimated groups. As we illustrate, this novel dataset enables us to gain new insights into the drivers of attention heterogeneity.

We show that greater attention improves forecast accuracy on average, as measured by absolute forecast errors and the amount of information incorporated into forecasts. These findings confirm that our estimated attention weights capture a learning channel through which agents acquire information about aggregate economic variables.

We investigate the impact of aggregate factors on attention. Consistent with previous research, we find that attention remained relatively stable over recent decades but experienced sharp fluctuations during major economic downturns and significant political events (Coibion and Gorodnichenko, 2015; Goldstein, 2023). Unlike prior studies, our approach enables us to examine changes beyond the average level of attention. Notably, we observe an unprecedented increase in the standard deviation of attention during the pandemic years. Additionally, we document that the rise in attention during the Global Financial Crisis and the COVID recession was not uniform across households but was rather driven by the larger negative skewness of the distribution.

[3]We analyze six datasets from the ECB SPF, covering growth, inflation, and unemployment forecasts for the current and next year, and two datasets from the US SCE, focusing on one- and three-year-ahead inflation.

While previous studies have examined how socioeconomic characteristics correlate with forecast errors (Broer et al., 2022; Goldstein, 2023), less is known about their direct effect on attention. To address this gap, we utilize the socioeconomic variables provided in the core US SCE dataset. In assessing the impact of these variables, we control for various observed and unobserved factors, such as aggregate shocks, geographic variations, and survey tenure. Our findings indicate that consumers with lower incomes, less education, those who are unemployed, and individuals from ethnic minorities tend to exhibit higher attention to inflation. These results support the view that inflation disproportionately affects these groups.[4] Additionally, we find that women are generally more attentive, potentially due to traditional gender roles that expose them more to highly volatile grocery prices (D'Acunto et al., 2021).

Ultimately, aggregate and idiosyncratic factors explain little of the total heterogeneity that we retrieve. Accounting for estimation uncertainty in forecast-specific attention weights, these factors explain jointly about 6% to 12% of the total variance in attention weights in the ECB SPF, and about 23% to 29% in the US SCE. This suggests that attention varies primarily due to individual-specific fluctuations over time rather than systematic aggregate or idiosyncratic factors.

Finally, we illustrate how our panel data on attention can inform theories of expectation formation. Examining the relationship between forecast-specific prior uncertainty, posterior uncertainty, and attention weights, we document systematic deviations from Bayesian updating. Since this tests a core prediction of Bayesian models and relies on self-reported measures of uncertainty, the deviations we identify cannot be attributed to behavioral theories in which agents behave as if they were Bayesian but misperceive the data-generating process, such as overconfidence (Broer and Kohlhas, 2022), and over- and under-extrapolation (Angeletos et al., 2021; Ryngaert, 2024). Similarly, our findings are robust to theories in which the actual gain is proportional to the Bayesian gain, such as diagnostic expectations (Bordalo et al., 2020) and cautious expectations (Kohlhas and Robertson, 2025).

[4]Recently, Gemmi and Mihet (2024) analyzed the effect of socioeconomic variables on attention using the same dataset (inflation, US SCE). Unlike our approach, they group forecasts *ex ante* and find that most of these variables are not individually statistically significant. In contrast, we observe that virtually all variables are significant, despite considering more categories and controlling for unobserved factors. We interpret this as evidence of the increased statistical power provided by our clustering approach.

Instead, we show that this deviation from Bayesian updating and related theories arises from under-adjustments to variations in prior uncertainty and signal precision. Specifically, while Bayesian updating predicts that a 1% increase in prior uncertainty should raise the attention weight by 1%, we find that attention weights increase by only 0.2% to 0.3% in the data. Similarly, an increase in prior signal variance reduces the attention weight by only 0.5% to 0.7%, compared to the 1% decrease predicted by Bayesian updating. These under-adjustments likely reflect the cognitive complexity involved in optimally trading off a signal's informativeness against its noisiness.

Related literature First, we contribute to the literature by estimating attention to public news using survey data. Extending the seminal approach of Coibion and Gorodnichenko (2015), Goldstein (2023), Gemmi and Valchev (2023), and Gemmi and Mihet (2024), propose a method to directly estimate average attention from individual forecasts. Our contribution is to formally extend their framework to be robust to, and estimate, unconditional heterogeneity in attention. We show that attention heterogeneity is sizable and varies over the business cycle. Moreover, ignoring heterogeneity leads to an overestimation of average attention. Finally, we find that previous attempts to address heterogeneity through ex ante conditioning explain only about 10% of total heterogeneity.

Second, we contribute to the literature that seeks to provide causal evidence on the drivers of attention. This body of work largely relies on information experiments, such as randomized controlled trials (RCTs), where some participants are randomly provided with information (Haaland et al., 2023). For instance, this approach is employed to assess the effects of uncertainty on attention (Kumar et al., 2023; Coibion et al., 2024), as we do. The strength of this method lies in its ability to establish causal relationships. However, as Maćkowiak and Wiederholt (2024) highlights, RCTs cannot estimate individuals' attention to public information—the key parameter for macroeconomic models. Instead, we focus on estimating this parameter. We demonstrate how statistical clustering can be used to construct decade-long panel datasets on attention, which allow us to provide quasi-causal evidence on the drivers of attention from within-individual variations.

Third, we contribute to recent work documenting deviations from Bayesian updating in macroeconomics using survey data. This literature has consistently documented that expectation errors are predictable, a fact which has led to the emergence of a variety of

competing explanations: diagnostic expectations (Bordalo et al., 2020), over confidence (Broer and Kohlhas, 2022), over and under extrapolation (Angeletos et al., 2021; Ryngaert, 2024), cautious expectations (Kohlhas and Robertson, 2025), adjustment frictions (Baley and Turén, 2024). Just like this literature, we document deviations from Bayesian updating. Unlike this prior work, however, we do so by assessing if the weight agents place on news is consistent with Bayesian updating, making our approach more direct. We find that attention weights under-adjust to variations in prior uncertainty and signal precision. These under-adjustments challenge the aforementioned explanations.

Lastly, our work aligns with the initiative of Lewis et al. (2024), who also use Gaussian finite mixtures to offer new insights into first-order macroeconomic phenomena.

The paper is organized as follows. Section 2 motivates our focus on attention heterogeneity. Section 3 presents the estimation strategy to uncover attention heterogeneity in survey data. Section 4 presents the results from this estimation, characterizing the distribution of attention in the ECB Survey of Professional Forecasters (ECB SPF) and the US Survey of Consumer Expectations (US SCE). Section 5 documents the effect of aggregate and idiosyncratic factors on attention. Section 6 discusses implications of our findings for theories of expectation formation. The last section concludes.

2 The macroeconomics of attention heterogeneity

This section reviews a simple framework with endogenous attention to illustrate two key insights: first, attention is heterogeneous across agents and over time, and, second, this heterogeneity carries important implications for aggregate macroeconomic dynamics. This section motivates also the measure of attention that we use in this paper.

2.1 A toy model with rational inattention

Setup The framework we consider largely builds on Maćkowiak et al. (2023). We examine the problem of an agent i choosing an action at time t, denoted $z_{i,t}$, in response to an unknown state x_t. The agent's instantaneous loss from not implementing the

optimal action under perfect information at t is given by

$$\mathcal{L}(z_{i,t}, x_t) \;=\; r_{i,t}\;\left(a_{i,t}\;x_t \;-\; z_{i,t}\right)^2 \;. \tag{1}$$

This loss function emerges from a second-order approximation of the agent's utility (or value function) around the optimal action under perfect information. With this interpretation in mind, $r_{i,t}$ represents the stakes from selecting the optimal action, i.e., the second derivative of the utility with respect to the action. Moreover, $a_{i,t}$ is the (local) slope of the optimal action, i.e., the first derivative of the policy function. We explicitly highlight individual- and time-specific indices in (1) to underscore heterogeneity.

Let $\mathcal{I}_{i,t}$ represent agent i's information set available at t. Suppose the prior about the state is Gaussian, $x_t|\mathcal{I}_{i,t-1} \sim \mathcal{N}\Big(F_{i,t-1},\, \sigma^2_{i,t|t-1}\Big)$, with $F_{i,t-1} := \mathbb{E}[x_t|\mathcal{I}_{i,t-1}]$ the conditional expectation and $\sigma^2_{i,t|t-1}$ capturing prior uncertainty.

Attention choice The agent chooses an action to minimize expected loss, subject to an information-processing cost, $\lambda_{i,t} \cdot I(z_{i,t}; x_t)$, where $I(z_{i,t}; x_t)$ is the mutual information (Sims, 2003). Optimal information acquisition involves observing a noisy Gaussian signal $y_{i,t} \;=\; x_t \;+\; \omega_{i,t}$, with $\omega_{i,t} \;\sim\; \mathcal{N}\Big(0,\, \sigma^2_{\omega,i,t}\Big)$.

Denote $F_{i,t} := \mathbb{E}[x_t|y_{i,t}, \mathcal{I}_{i,t-1}]$ the posterior belief after observing $y_{i,t}$. Bayesian updating implies a linear forecast $F_{i,t} \;=\; (1 - k_{i,t})\; F_{i,t-1} \;+\; k_{i,t}\; y_{i,t}$, where

$$k_{i,t} = k^{RI}_{i,t} \quad := \quad \max\left(0\;,\; 1 \;-\; \frac{\lambda_{i,t}}{2\; r_{i,t}\; a^2_{i,t}\; \sigma^2_{i,t|t-1}}\right) \tag{2}$$

is the optimal attention of agent i at time t.[5] Equation (2) clearly illustrates that attention is (endogenously) heterogeneous. This heterogeneity arises from differences in prior uncertainty ($\sigma^2_{i,t|t-1}$), marginal costs of information ($\lambda_{i,t}$), and stakes ($r_{i,t}$, $a_{i,t}$).

Remark 1. *Attention—defined as the weight $k_{i,t}$ agents place on signals—is inherently heterogeneous across individuals and over time.*

[5] A highly attentive agent collects and processes substantial information, forming a precise (mental) signal $y_{i,t}$ about x_t. This corresponds to a low noise variance, $\sigma^2_{\omega,i,t} \to 0$, and a high attention weight, $k^{RI}_{i,t} \to 1$. In contrast, a highly inattentive agent observes a very noisy signal, $\sigma^2_{\omega,i,t} \to \infty$, resulting in a low attention weight, $k^{RI}_{i,t} \to 0$.

Aggregate dynamics To illustrate the importance of attention heterogeneity for the dynamics of aggregates, consider an economy composed of a unit mass of individuals, and let $Z_t = \int_0^1 z_{i,t}, di$ denote the aggregate action. Then we have:

$$Z_t \;=\; \int_0^1 a_{i,t} \left[k_{i,t}\; y_{i,t} \;-\; \Big(1 \;-\; k_{i,t}\Big)\; F_{i,t-1} \right] di\,, \tag{3}$$

where we used the fact that the agent i's optimal action at t is given by $z_{i,t} = a_{i,t} F_{i,t}$.

Consider an aggregate shock S_t of size $s \neq 0$. The instantaneous (relative) effect of this shock on the aggregate action is

$$\frac{Z_{t|S_t=s} - Z_{t|S_t=0}}{s} = \int_0^1 (k_{i,t} \;\cdot a_{i,t})\; di = \mathbb{E}_t[k_{i,t}]\; \mathbb{E}_t[a_{i,t}] + \mathbb{C}ov_t(k_{i,t},\, a_{i,t})\,, \tag{4}$$

where $\mathbb{E}_t[\bullet]$ and $\mathbb{C}ov_t(\bullet)$ are the cross-sectional mean and covariance at time t.

Equation (4) shows that (unpredictable) variations in aggregate variables are shaped by attention heterogeneity. In particular, it reveals that attention heterogeneity matters when:

(i) shifts in the cross-sectional distribution drive average attention over time, $\mathbb{E}_t[k_{i,t}]$;

(ii) cross-sectional heterogeneity of attention correlates with individual policy functions, $\mathbb{C}ov_t(k_{i,t},\, a_{i,t})$.

Conditions (i) and (ii) hold in the RI framework discussed previously. For instance, equation (2) indicates that (i) uncertainty shocks result in time-varying average attention, while (ii) the cross-sectional correlation arises from attention increasing with $a_{i,t}^2$.

More generally, equation (4) indicates that the relevant measure of attention for macroeconomic modeling is the weight agents' place on their signal, $k_{i,t}$. Henceforth, we refer to this measure of attention as an attention weight.

Remark 2. *The cross-sectional distribution of attention weights, $k_{i,t}$, and its variation over time, play a central role in shaping macroeconomic responses to aggregate shocks.*

2.2 Alternative models of attention

While the setup presented thus far focuses on predictions from rational inattention with an entropy-based information cost, alternative theories have been proposed to model

the attention weight $k_{i,t}$. We summarize some of these alternatives in Table 1. Despite differing formulations, all converge on a unifying measure of attention: the weight agents place on their signals. Across these frameworks, heterogeneity naturally arises from differences in how agents weight new signals relative to their priors.

Table 1: Attention weights in the literature

Models	Attention weights ($k_{i,t}$)	References
Noisy information	$k^B_{i,t} := \sigma^2_{i,t\|t-1} / \left(\sigma^2_{i,t\|t-1} + \sigma^2_{\omega,i,t}\right)$	Lucas (1972)
Sticky expectations	$\{0, 1\}$	Mankiw and Reis (2002)
Rational inattention	Equation (2)	Sims (2003)
Over-reaction	$(1 + \theta_{i,t})k^B_{i,t}$	Bordalo et al. (2020)
Over-extrapolation	$\tilde{\sigma}^2_{i,t\|t-1} / \left(\tilde{\sigma}^2_{i,t\|t-1} + \sigma^2_{\omega,i,t}\right)$	Angeletos et al. (2021)
Over-confidence	$\sigma^2_{i,t\|t-1} / \left(\sigma^2_{i,t\|t-1} + \tilde{\sigma}^2_{\omega,i,t}\right)$	Broer and Kohlhas (2022)
Cautious expectations	$\vartheta_{i,t}k^B_{i,t}$	Kohlhas and Robertson (2025)

Notes: Tildes denote perceived variables, as opposed to actual variables. Signal precision differs across specifications (e.g., it is infinite in the sticky expectation model).

Remark 3. *While theories differ in their assumptions about the determinants of attention, they share a unifying measure: the attention weights $k_{i,t}$.*

As seen in equation (4), the attention weights $k_{i,t}$ serve as sufficient statistics for the impact of attention on the dynamics of aggregate variables.

2.3 What we do (and don't do)

In the following, we propose a statistical method to estimate the full, unconditional distribution of attention weights, $k_{i,t}$, allowing for heterogeneity both across and within individuals. Given the wide range of structural assumptions that lead to specific formulations of attention, *we do not* rely on the structural specifications reported in Table 1. Instead, *we do* estimate attention weights using reduced-form specifications that do not require committing to any particular underlying theory.

We focus on estimating attention from the weight agents place on signals, $k_{i,t}$, because this measure is directly relevant for the dynamics of aggregate variables (see Remark 2). In doing so, we follow the seminal approach of Coibion and Gorodnichenko (2015), aiming to provide a practical measure of attention that is directly informative for macroeconomic modeling.

3 Estimating attention heterogeneity

This section motivates and presents our estimation strategy to retrieve the full, unconditional heterogeneity in attention from expectation surveys.

3.1 Model with homogeneous attention

We first recall the framework used to estimate attention from the cross-sectional distribution of forecast revisions under the assumption of homogeneous attention. This approach builds on Coibion and Gorodnichenko (2015), and is used in Goldstein (2023) and Gemmi and Valchev (2023), among others. We extend this framework to be robust and capable of estimating heterogeneous attention weights next.

Environment We consider a general model for individual forecasts of a random variable, denoted x_t. We do not restrict the stochastic process for x_t. Instead, we work directly with the mathematical expectation of its realization h periods ahead given all available information at time t, denoted $x_{t+h|t} := \mathbb{E}_t[x_{t+h}]$. This mathematical expectation is referred to as the Full-Information Rational Expectation (FIRE) in the macroeconomic literature, a terminology that we follow henceforth.

We denote individual's i forecast at time t of x_{t+h} by $F_{i,t} := F_{i,t}x_{t+h}$, where $F_{i,t}$ alone indicates the variable "forecast", whereas $F_{i,t}$ in $F_{i,t}x_{t+h}$ is the forecast operator. This forecast can differ from $x_{t+h|t}$ as individuals might not observe all available information at time t (information friction), and might not process information rationally (behavioral biases). Henceforth, we intend every variable of the model to be related to a horizon h, and omit indexing for horizon h whenever it does not create confusion.

Specifically, we assume that individual i observes a signal

$$y_{i,t} \;=\; x_{t+h|t} \;+\; \omega_{i,t} \tag{5}$$

about the FIRE, $x_{t+h|t}$, which is the optimal belief without informational or behavioral frictions. The noise, $\omega_{i,t} = \eta_{i,t} + \eta_t$, contains a Gaussian idiosyncratic noise, denoted $\eta_{i,t}$, and, potentially, a white noise common to all individuals, denoted η_t. The variance of the noise is given by $\mathbb{V}ar(\omega_{i,t}) = \sigma^2_{\omega,i,t}$.

Forecast revision We assume that individuals adjust their forecast linearly

$$F_{i,t} \;=\; (1-k)\, F_{i,t-1} \;+\; k\, y_{i,t} \tag{6}$$

where k is the weight put on the signal $y_{i,t}$, and $(1-k)$ the weight put on the prior belief $F_{i,t-1}$. In the homogeneous specification, (6), this weight is assumed constant across individuals and time.

Define $R_{i,t} := F_{i,t} - F_{i,t-1}$ the revision in individual forecasts. Combining equations (5) and (6), we have

$$R_{i,t} \;=\; k\,(\,x_{t+h|t} - F_{i,t-1}\,) \;+\; k\,\omega_{i,t} \;=\; k\,(\,\mu_t - F_{i,t-1}\,) \;+\; \varepsilon_{i,t} \tag{7}$$

where $\varepsilon_{i,t} \;:=\; k\,\eta_{i,t}$ is an i.i.d. zero-mean Gaussian error term and $\mu_t :=\; x_{t+h|t} + \eta_t$ is the common part of private signals (an aggregate factor). This linear expression can be taken to the data to estimate the weight k.[6]

Relation to the literature Equation (7) is reminiscent of the specification estimated in Goldstein (2023), Gemmi and Valchev (2023), and Gemmi and Mihet (2024). A notable distinction is that we do not use deviations from average forecasts to control out μ_t. Instead, we estimate these aggregate factors jointly with the weight k. This distinction allows us to consider heterogeneous weights.

As these authors demonstrate, equation (7) characterizes the expectation formation process in theories with rational Bayesian updating, including noisy information (Lucas, 1972) and rational inattention (Sims, 2003). Similarly, this representation is also consistent with many deviations from rational Bayesian updating, such as sporadic adjustments (Mankiw and Reis, 2002; Boccanfuso, 2022; Baley and Turén, 2024), diagnostic expectations (Bordalo et al., 2020), overconfidence (Broer and Kohlhas, 2022), overextrapolation (Angeletos et al., 2021; Ryngaert, 2024), strategic behaviors (Ottaviani and Sørensen, 2006; Gemmi and Valchev, 2023), and systematic biases (Capistrán and Timmermann, 2009; Patton and Timmermann, 2010).[7]

[6] While expression (7) is restrictive in that it assumes homogeneous attention weights, it allows additive unobserved heterogeneity coming from a forecast adjustment model, as in (6), with idiosyncratic bias: $F_{i,t} = (1-k)F_{i,t-1} + ky_{i,t} + \gamma_i$, where γ_i is potentially correlated with $F_{i,t-1}$ and $y_{i,t}$. In fact, γ_i cancels out when working with revisions in the form of first differences of forecasts, i.e., $R_{i,t} := F_{i,t} - F_{i,t-1}$.

[7] As we have illustrated in Table 1, the main effect of these deviations is to lead to an attention weight

Terminology In the following, we refer to k in (6) and (7) as a measure of attention. This terminology arises from recognizing that $\mu_t - F_{i,t-1}$ represents individual-specific news, and k the extent to which forecasts adjust to this news: higher k indicates greater responsiveness of expectation to new information (higher attention), while a lower k reflects a stronger reliance on prior forecast (lower attention and more sluggish updating). As we have seen in the previous section, this measure of attention is directly informative for macroeconomic modeling.

Two features of this measure of attention are worth highlighting. First, the news in (7) is defined relative to common information μ_t. As such, this measure of attention is central to macroeconomic modeling, as it shapes how expectations respond to aggregate information that is potentially available to everyone. However, it is silent on how individuals incorporate idiosyncratic information into their forecasts. Second, this measure of attention is not necessarily indicative of the optimal use of information. For example, if the average respondent deviates from rational Bayesian updating, they might fail to adequately smooth out the signal noise, $\omega_{i,t}$. This could result in excessively sensitive forecasts that we would label as highly attentive.

3.2 Model with heterogeneous attention

Motivation As suggested in section 2, the attention weights from equation (6) are likely to be heterogeneous, rather than constant over time and across individuals. There are at least two reasons to account for such heterogeneity in attention weights. First, this heterogeneity is important in its own right, as it shapes the effects of macroeconomic shocks (Remark 2). Second, as we demonstrate in Online Appendix C, the homogeneous regression (7) does not consistently estimate the average attention weight in the presence of heterogeneity. For these reasons, developing a framework that is both robust to and capable of identifying this heterogeneity is a first-order empirical endeavor.

To circumvent these limitations, some authors extend the homogeneous regression (7) by including interaction terms (e.g., Benhima and Bolliger, 2022; Goldstein, 2023; Gemmi and Mihet, 2024). This approach requires an explicit assumption about which characteristics are relevant for heterogeneity, and which are not. Furthermore, it still

k different from the (MSE-)optimal weight. To the extent that we are interested in directly estimate this weight from data, there is no need for it to be optimal.

yields inconsistent estimates as long as some heterogeneity remains within the guessed shared characteristics.

In contrast, we reverse this logic by assuming the existence of $G < \infty$ homogeneous groups and allowing the data to optimally allocate forecasts into these groups. Consequently, we refrain from taking an explicit stance on the drivers of heterogeneity, instead letting the data speak. This approach provides a consistent estimate of the average attention weight and allows us to recover unconstrained heterogeneity in weights.

Heterogeneity Formally, we introduce heterogeneity in weights by assuming that each observation, indexed by (i, t), belongs to one of G groups. Forecast revisions write

$$R_{i,t} \;=\; \sum_{g=1}^{G} \; d_{i,t}(g) \Big[\, k_g \Big(\, \mu_t - F_{i,t-1} \Big) \Big] \;+\; \varepsilon_{i,t} \;, \qquad \varepsilon_{i,t} | d_{i,t}(g) = 1 \sim \mathcal{N} \left(0, \; \sigma_g^2 \right). \tag{8}$$

where $\mathcal{N}$ denotes the Gaussian density. Thus, we assume that the error term, $\varepsilon_{i,t}$, is independently drawn from the zero-mean Gaussian distribution with group-specific variance σ_g^2—as predicted by the forecasting model (6). The variable $d_{i,t}(g)$ is an indicator that takes value 1 if observation (i, t) belongs to group g, and 0 otherwise. The aggregate factor μ_t, which captures common information, is the same across groups.

We treat group membership, $d_{i,t}(g)$, as an unobservable object to be estimated. Allocation to groups is a discrete process, where each observation effectively belongs to a unique group. Specification (8) allows individuals to switch between groups over time. Similarly, group allocation can correlate with observable, unobservable, constant, and time-varying characteristics of individuals. In principle, this approach permits retrieval of the full distribution of attention weights as G increases.

3.3 Model estimation

Finite mixture of regressions We estimate both group membership and model parameters via a finite mixture of G regressions. This method offers a probabilistic clustering framework capable of detecting unobserved, latent group memberships without relying on a predetermined set of observable characteristics (see McLachlan et al., 2019, for an introduction). By leveraging the parametric assumptions in equation (8), the finite mixture approach computes posterior probabilities that assign each observation to one of the

G latent groups.

More specifically, we estimate the mixture by maximum likelihood (ML). Let denote the sample size by nT, with $i = 1, ..., n$, and $t = 1, ..., T$. Moreover, let π_g be the unconditional prior probability that any observation (i, t) belongs to group g. By definition, $\pi_g = \mathbb{E}[d_{i,t}(g)]$ is the proportion of observations in the population belonging to group g, with $\pi_g > 0$ and $\sum_{g=1}^{G} \pi_g = 1$. As is standard, these prior probabilities are assumed independent of regressors in (8). The complete-data likelihood considers group membership as an observed variable, and reads as[8]

$$L_c\left(\boldsymbol{R}, \boldsymbol{F}, \boldsymbol{d}; \boldsymbol{\theta}_G\right) \;=\; \prod_{i=1}^{n} \prod_{t=1}^{T} \prod_{g=1}^{G} \left[\pi_g \cdot \phi\left(R_{i,t} \,;\, k_g\left(\mu_t - F_{i,t-1}\right),\, \sigma_g^2\right)\right]^{d_{i,t}(g)} \tag{9}$$

where vectors $\boldsymbol{R}$ and $\boldsymbol{F}$ respectively collect revisions, $R_{i,t}$ and priors, $F_{i,t-1}$, and the matrix $\boldsymbol{d}$ collects group membership indicators $d_{i,t}(g)$. The vector of parameters to be estimated, $\boldsymbol{\theta}_G := (\boldsymbol{\pi}_G', \boldsymbol{k}_G', \boldsymbol{\sigma}_G', \boldsymbol{\mu}')'$, collects the G proportions $\boldsymbol{\pi}_G := (\pi_1, \ldots, \pi_G)'$, the heterogeneous regression slopes $\boldsymbol{k}_G := (k_1, \ldots, k_G)'$, the aggregate factors $\boldsymbol{\mu} := (\mu_1, \ldots, \mu_T)'$, and the standard deviations of the error term, $\boldsymbol{\sigma}_G := (\sigma_1, \ldots, \sigma_G)'$. Finally, the function $\phi(x; a, b)$ denotes the Gaussian pdf with mean a and variance b, evaluated at x.

Expected log-likelihood While the complete-data likelihood (9) depends on group membership indicators, these are unobserved. Therefore, direct maximization of (9) is unfeasible. As a result, we consider posterior membership probabilities, denoted $w_g(r \,|\, f;\, \boldsymbol{\theta}_G) := \mathbb{P}\big(d(g) = 1 \,|\, r, f, \boldsymbol{\theta}_G\big)$ and maximize the expected log-likelihood. More specifically, integration of the log of the complete-data likelihood (9) over the distribution of group-memberships, conditional on $(\boldsymbol{R}, \boldsymbol{F})$, gives the expected log-likelihood

$$\begin{aligned}\mathbb{E}\Big[\log L_c(\boldsymbol{R}, \boldsymbol{F}, \boldsymbol{d}; \boldsymbol{\theta}_G) \mid \boldsymbol{R}, \boldsymbol{F}, \boldsymbol{\theta}_G\Big] = &\sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{g=1}^{G} w_g\Big(R_{i,t} | F_{i,t-1}; \boldsymbol{\theta}_G\Big) \left[\log(\pi_g) - \tfrac{1}{2} \log\Big(2\pi\sigma_g^2\Big)\right] \\ &- \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{g=1}^{G} w_g\Big(R_{i,t} \mid F_{i,t-1}; \boldsymbol{\theta}_G\Big) \left\{\frac{\Big[R_{i,t} - k_g(\mu_t - F_{i,t-1})\Big]^2}{2\,\sigma_g^2}\right\},\end{aligned} \tag{10}$$

[8] By convention, we assume $\lim_{a \mapsto 0^+} a \times \log a = 0$ throughout the paper.

where the posterior group-membership probabilities are given, for all g, by

$$w_g\Big(R_{i,t}|\, F_{i,t-1};\boldsymbol{\theta}_G\Big) \;=\; \frac{\pi_g \,\cdot\, \phi\left(R_{i,t};\; k_g\left(\mu_t - F_{i,t-1}\right),\; \sigma_g^2\right)}{\sum_{g=1}^{G} \pi_g \,\cdot\, \phi\left(R_{i,t};\; k_g\left(\mu_t - F_{i,t-1}\right),\; \sigma_g^2\right)}\,. \tag{11}$$

While group membership $d_{i,t}(g)$ is discrete, the posterior group-membership probabilities have continuous support. This differs from "hard" clustering, which assigns binary weights (Bonhomme and Manresa, 2015). The continuous support of the posterior probabilities enables "soft" (probabilistic) clustering, which is arguably more parsimonious. Values closer to the extremes, 0 and 1, indicate high certainty in group assignment, whereas intermediate values reflect the econometrician's uncertainty in assigning these weights.

ECM algorithm Maximization of the log complete-data likelihood given observed data is performed using the Expectation and Conditional Maximization (ECM) algorithm. At the $(\ell+1)$th iteration of the ECM algorithm, the E-step requires taking the expectation of the complete-data likelihood conditional on the observed data, using the current parameter value, $\boldsymbol{\theta}_G^{(\ell)}$, of $\boldsymbol{\theta}_G$. The expected log-likelihood is given by

$$\mathcal{Q}\Big(\boldsymbol{\theta}_G;\boldsymbol{\theta}_G^{(\ell)}\Big) = \\ \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{g=1}^{G} w_g\left(R_{i,t}|F_{i,t-1};\boldsymbol{\theta}_G^{(\ell)}\right) \,\cdot\, \log\Big(\pi_g \;\cdot\; \phi\left(R_{i,t};\; k_g\left(\mu_t - F_{i,t-1}\right),\, \sigma_g^2\right)\Big)\,, \tag{12}$$

where $w_g\left(R_{i,t} \mid F_{i,t-1};\boldsymbol{\theta}_G^{(\ell)}\right)$ are the posterior membership probabilities, given the parameters from the ℓth iteration.

The M-step requires updating the estimates to

$$\boldsymbol{\theta}_G^{(\ell+1)} := \underset{\boldsymbol{\theta}_G\in\mathcal{P}_G}{\arg\max}\;\; \mathcal{Q}\Big(\boldsymbol{\theta}_G;\boldsymbol{\theta}_G^{(\ell)}\Big)\,, \qquad \text{subject to } \pi_g > 0\,,\;\; \text{and} \;\; \sum_{g=1}^{G}\pi_g = 1\,,$$

where $\mathcal{P}_G \subset \mathbb{R}^{\dim\boldsymbol{\theta}_G}$ is the parameter space. First order conditions imply that the membership proportions are updated for all $g \in \{1,\ldots,G\}$ using

$$\pi_g^{(\ell+1)} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_g\left(R_{i,t}\Big|F_{i,t-1};\; \boldsymbol{\theta}_G^{(\ell)}\right)\,. \tag{13}$$

Similarly, the attention weights are

$$k_g^{(\ell+1)} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} w_g\left(R_{i,t} \mid F_{i,t-1}; \boldsymbol{\theta}_G^{(\ell)}\right)\left(\mu_t^{(\ell+1)} - F_{i,t-1}\right) R_{i,t}}{\sum_{i=1}^{n}\sum_{t=1}^{T} w_g\left(R_{i,t} \mid F_{i,t-1}; \boldsymbol{\theta}_G^{(\ell)}\right)\left(\mu_t^{(\ell+1)} - F_{i,t-1}\right)^2}. \tag{14}$$

Updates of standard deviations are given by

$$\sigma_g^{(\ell+1)} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{t=1}^{T} w_g\left(R_{i,t}\middle|F_{i,t-1};\, \boldsymbol{\theta}_G^{(\ell)}\right)\left[R_{i,t} - \left(\mu_t^{(\ell+1)} - F_{i,t-1}\right) k_g^{(\ell+1)}\right]^2}{\sum_{i=1}^{n}\sum_{t=1}^{T} w_g\left(R_{i,t} \mid F_{i,t-1}, \boldsymbol{\theta}_G^{(\ell)}\right)}}. \tag{15}$$

Finally, aggregate factors are updated at each period t from

$$\mu_t^{(\ell+1)} = \frac{\sum_{i=1}^{n}\sum_{g=1}^{G} w_g\left(R_{i,t}|F_{i,t-1}, \boldsymbol{\theta}_G^{(\ell)}\right)\left(R_{i,t} + \kappa_g^{(\ell+1)} F_{i,t-1}\right)}{\sum_{i=1}^{n}\sum_{g=1}^{G} w_g\left(R_{i,t}|F_{i,t-1}, \boldsymbol{\theta}_G^{(\ell)}\right)\kappa_g^{(\ell+1)}}, \tag{16}$$

where, the summation over groups (indexed by g) accounts for the fact that the time factors are common across groups, while the summation over observations at time t (indexed by i) reflects that, respectively, only revisions and forecasts from period t and $t-1$ provide information about the unobserved time-t factor.[9]

Equations (14) and (16) involve updated parameters $\mu_t^{\ell+1}$ and $k_g^{\ell+1}$ on the right-hand side, respectively. As a result, it is not feasible to concentrate out either parameter from the expressions for $k_g^{(\ell+1)}$ or $\mu_t^{(\ell+1)}$.

Following Meng and Rubin (1993), we rely on Conditional Maximization (CM) to update parameters. For given G, we partition $\boldsymbol{\theta}$ into three blocks of parameters, $\boldsymbol{\vartheta}_{1,G} := (\boldsymbol{\pi}_G', \boldsymbol{\mu}_G')'$, $\boldsymbol{\vartheta}_{2,G} := \boldsymbol{k}_G$, and finally $\boldsymbol{\vartheta}_{3,G} := \boldsymbol{\sigma}_G$ with $\boldsymbol{\theta}_G := (\boldsymbol{\vartheta}_{1,G}', \boldsymbol{\vartheta}_{2,G}', \boldsymbol{\vartheta}_{3,G}')'$. At iteration $(\ell+1)$, the ECM algorithm updates $\boldsymbol{\theta}_G^{(\ell+1)}$ by sequentially conditioning on blocks of parameters. That is, $\boldsymbol{\vartheta}_{1,G}^{(\ell+1)}$ is obtained from (13) and (16) conditional on $\left(\boldsymbol{\vartheta}_{2,G}^{(\ell)}, \boldsymbol{\vartheta}_{3,G}^{(\ell)}\right)$, $\boldsymbol{\vartheta}_{2,G}^{(\ell+1)}$ is estimated using (14) conditional on $\left(\boldsymbol{\vartheta}_{1,G}^{(\ell+1)}, \boldsymbol{\vartheta}_{3,G}^{(\ell)}\right)$, and $\boldsymbol{\vartheta}_{3,G}^{(\ell+1)}$ is estimated from the square root of (15) conditional on $\left(\boldsymbol{\vartheta}_{1,G}^{(\ell+1)}, \boldsymbol{\vartheta}_{2,G}^{(\ell+1)}\right)$.

Identification and incidental parameter Estimating the unobserved factors, $\boldsymbol{\mu}$, alongside a finite mixture model is uncommon and introduces potential complications, which we discuss here. First, it may prevent the identification of the model parameters.

[9] Equations (13)–(16) are based on the assumption of a balanced panel. To account for an unbalanced panel, we allow the cross-sectional dimension to vary with time, meaning the summations over the index i now run from 1 to $n(t)$.

Appendix A.1 demonstrates that the time-t aggregate factor is identified from variations in $F_{i,t-1}$ across individuals i for a fixed t.

Second, following Feng and McCulloch (1996) and Cheng and Liu (2001), we show consistency results for the ML estimator of $\boldsymbol{\theta}_G$ in Appendix A.2. ML estimates of $\boldsymbol{\theta}_G$ are subject to the so-called incidental parameter problem and are inconsistent as $T \to \infty$ while the number of individuals per period, n, remains fixed (for a discussion, see Lancaster, 2000). Deviations from consistency are of order $O(n^{-1})$. In our applications, n ranges from approximately 50 to 1,200. Hence, deviations from consistency should be quantitatively negligible, a prediction which is supported by the simulation study that we discuss next.

4 Results

This section presents the main results from estimating the finite mixture using data from ECB's survey of professional forecasters and the US Survey of Consumer Expectations. It shows that there is significant heterogeneity in attention, and that ignoring this heterogeneity often leads to underestimate average attention. We perform a simulation study, which supports our findings.

4.1 Data

Our analysis relies on two surveys covering different types of individuals and geographical regions: the ECB's survey of professional forecasters (ECB SPF), and the US Survey of Consumer Expectations (US SCE).

ECB SPF The ECB SPF has been conducted every quarter since 1999. Andrade and Le Bihan (2013) provide detailed information about the survey characteristics. This survey combines some features which makes it well-suited for our study. In particular, each quarter, respondents are asked to report inflation, GDP growth, and unemployment rate forecasts for the current calendar year, the next calendar year, and the after next calendar year. These fixed-event forecasts allow to observe up to eight revisions of a forecaster's forecast about the same event. To limit the impact of outliers, that may cause overfitting, we trim forecasts and revisions at the 1st and 99th percentiles each

quarter.[10]

US SCE The US SCE is a monthly survey of approximately 1,200 household heads collected by the Federal Reserve Bank of New York since 2013. Armantier et al. (2017) provide detailed information about the survey characteristics. In this survey, respondents are asked to report growth in prices over the next 12 months and 24 to 36 months.

These forecasts contrast to those from the ECB SPF in that they refer to a rolling horizon, rather than a fixed event. For our purpose, we treat them as if they had an overlapping horizon since the horizons of two consecutive forecasts differ by only one month, a short length compared to the overall horizon (12 or 36 months). A similar approach is followed by Gemmi and Mihet (2024). The rotating structure of the survey allows to retrieve up to 11 revisions for a given date. To limit the impact of outliers, that may cause overfitting, we trim forecasts and revisions at the 1st and 99th percentiles each month. The survey also provides socioeconomic information about respondents, which we exploit in Section 5.

Descriptive statistics. Table 2 reports summary information on both surveys, including the total number of observations, (nT), average forecast (mean), and standard deviation (s.d.). The last column indicates the frequency of individuals who report the same forecast in two consecutive waves of the survey ($\mathbb{P}(R_{i,t} = 0)$). Zero revisions are salient features of forecast surveys (e.g., Andrade and Le Bihan, 2013). 16 to 26% of European forecasters do not change their forecast from one quarter to the next, and about 35% of US consumers' forecasts are unchanged.

4.2 Model selection, estimation and inference

Model selection in mixture models is a difficult task (Celeux et al., 2019, for a discussion). Much of the literature on finite mixture relies on information criteria to select G. Under some regularity conditions, the Bayesian Information Criterion (BIC) consistently selects the correct number of groups. Moreover, it is frequently found to be a conservative criteria when these conditions fail. A further concern with large datasets, such as the SCE survey dataset, is overfitting by increasing G to cluster small groups of outliers.

[10]More specifically, we compute deviations from time averages and trim the bottom and top 1% of these deviations.

Table 2: Summary information on the surveys

Variable	Horizon	Forecast $F_{i,t}$			
		nT	mean	s.d.	$\mathbb{P}(R_{i,t}=0)$
ECB SPF - quarterly from 1999q1 to 2024q2					
Growth	Q0-Q3	4,783	1.36	1.77	0.16
	Q4-Q7	4,469	1.83	0.91	0.23
Inflation	Q0-Q3	4,741	1.89	1.29	0.19
	Q4-Q7	4,452	1.72	0.52	0.27
Unempl.	Q0-Q3	4,496	8.99	1.50	0.26
	Q4-Q7	4,124	8.79	1.51	0.24
US SCE - monthly from 2013m6 to 2023m7					
Inflation	12m	121,851	5.24	7.43	0.35
Inflation	36m	122,022	4.57	7.56	0.32

Notes: This table presents descriptive statistics for forecasts from the ECB Survey of Professional Forecasters (ECB SPF) and the US Survey of Consumer Expectations (US SCE). In the ECB SPF, Q0-Q3 denotes forecasts for the current year, while Q4-Q7 refers to forecasts for the next year. In the US SCE, 12m represents one-year-ahead forecasts, and 36m represents three-year-ahead forecasts. nT indicates the number of observations, "mean" the average forecast, "s.d." the standard deviation, and $\mathbb{P}(R_{i,t}=0)$ the proportion of forecasts unchanged from one period to the next.

To limit this concern, our benchmark estimation relies also on $\mathcal{K}$-fold cross-validation to select the number of groups G. As we shall see, this approach is rather conservative.

We proceed as follows. First, for each dataset {Growth (Q0-Q3), ..., Inflation (36m)}, we estimate models with $G = 1, \ldots, 30$ using all available data. To mitigate the impact of initial starting values in the ECM algorithm and to avoid convergence to local maxima, each estimation is performed 150 times with random starting values. The final estimated parameter vector, $\hat{\boldsymbol{\theta}}_G$, is selected as the one yielding the supremum expected log-likelihood.

Second, we employ $\mathcal{K}$-fold cross-validation to determine the optimal number of groups G. We set $\mathcal{K} = 5$. Specifically, we randomly split the forecasters into five equal parts (folds). For each fold and each G, we estimate the model using the four training folds and evaluate the Bayesian Information Criterion (BIC) on the held-out fold. We retain the G that minimizes this criteria.

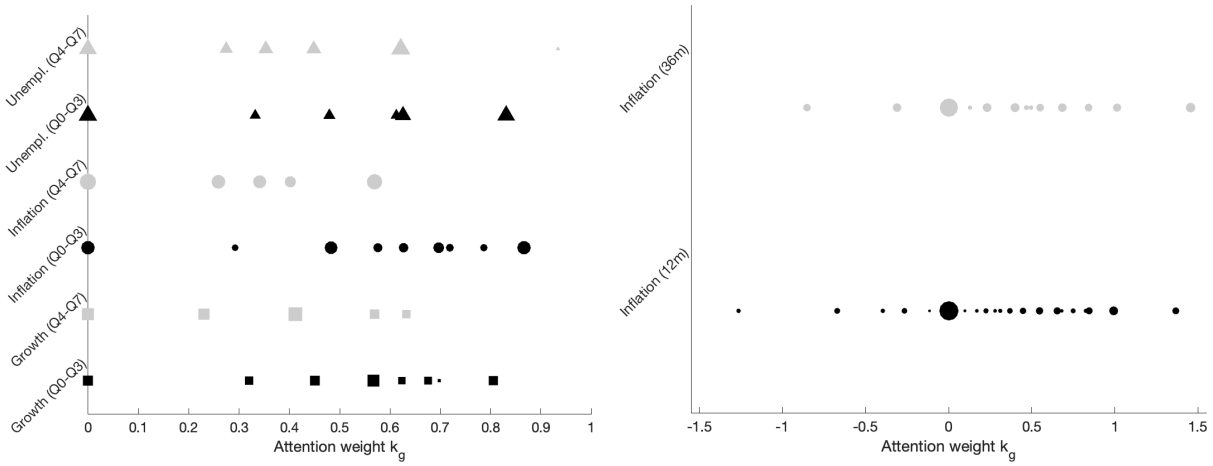
Last, confidence intervals are obtained through cluster wild bootstrap with fixed design, where we draw from the posterior distribution of the data. Given the particular structure of the surveys, wild bootstrap avoids introducing artificial balance and preserves the error structure of the data. For fixed G, we generate $B = 1,000$ bootstrap samples. In

order to circumvent the problem of label switching (identification), we start the ECM algorithm from an initialization matrix assigning posterior membership probabilities using the results from $\hat{\boldsymbol{\theta}}_G$ (O'Hagan et al., 2019). In Table B.2, we report percentile confidence intervals at the 95% confidence level. The confidence intervals that we obtain from this procedure account for both parameter and clustering uncertainty. As a complement, we report also partial confidence intervals. These less conservative intervals take posterior membership probabilities as given, making them more likely to characterize parameter uncertainty only (Lewis et al., 2024, for a similar approach).

4.3 Heterogeneous attention

We estimate a considerable degree of heterogeneity in attention weights. An exhaustive table containing estimated weights, prior membership probabilities, and 95% confidence intervals is relegated to the appendix (Table B.2) to save space. Some general patterns emerge across datasets, which we present in the following.

Figure 1: Estimated attention weights



(a) ECB SPF (b) US SCE

Notes: Estimated attention weights in the ECB survey of professional forecasters (panel (a)) and US survey of consumer expectations (panel (b)). Marker sizes are proportional to population weights (π_g). Each horizontal line refers to a different dataset. See Table B.2 for more information and confidence intervals. Groups representing less than 1% of forecasts are not reported.

Figure 1 presents the estimated attention weights for each dataset, with marker sizes proportional to the membership probabilities, π_g. As shown, attention weights exhibit significant heterogeneity. The number of groups, G, ranges from 5 to 24 across datasets, and attention weights are distributed widely, often extending to or beyond the unit line.

Consumers' attention is more heterogeneous than that of professional forecasters, as

indicated by the number of groups (respectively 19 and 7 on average), the relative size of groups, and the range of attention weights.

The lowest and highest attention weights are negative and larger than one in the US SCE (the estimation is unconstrained). Such weights are difficult to reconcile with theories predicting they should lie on the unit line, such as Bayesian updating. While these extreme gains pertain to only a limited proportion of forecasts in the US SCE, they illustrate a deviation from Bayesian updating—which may be reconcile with some of the theories reported in Table 1.

Finally, we systematically identify a group with an attention weight and standard deviation close to zero. This group corresponds to forecasts which are not revised from one period to the next, and the proportions associated to them correspond to the shares of zero revisions reported in the last column of Table 2.[11]

Overall, our findings highlight highly heterogeneous behaviors in how individuals incorporate news into their forecasts. Our benchmark specification is fairly conservative in this respect. Specifically, the number of groups we selected falls on the lower end of those typically found using other information criteria (Table B.1).

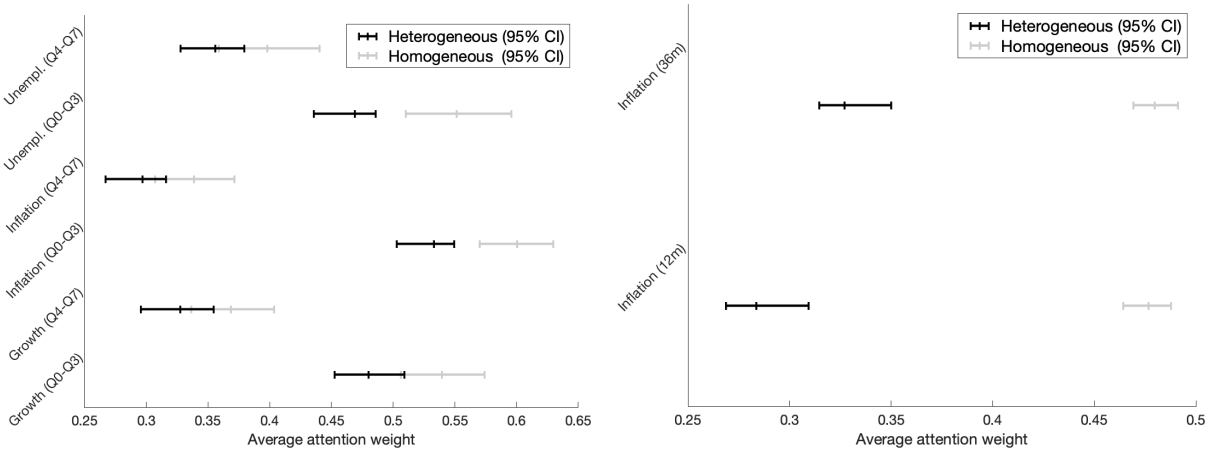
4.4 Average attention

The widespread heterogeneity in attention weights is also crucial for estimating average attention in the population. Figure 2 presents the average attention weights implied by the finite mixture model, computed as the inner product $\boldsymbol{\pi}_G' \boldsymbol{k}_G$, and compares them to those obtained under the homogeneous specification commonly used in the literature (where $G = 1$).

We find that the heterogeneous specification systematically yields a lower average attention than its homogeneous counterpart. This is due to between and within variations that appear in the probability limit of the estimated average attention weight when heterogeneity in the data is neglected (see Proposition 4 in the appendix). While the magnitude of this difference varies across datasets, the homogeneous estimate consistently falls outside the 95% bootstrap confidence intervals of the heterogeneous specification.

[11] The literature generally takes these zero revisions as direct evidence for sporadic forecast adjustments. Instead, as we show in a companion paper (Boccanfuso and Neri, 2025), our finite mixture can be extended to account for forecast rounding and non-adjustments. Doing so, we find that much of the observed zero revisions result, in fact, from forecast rounding.

Figure 2: Average attention



(a) ECB SPF (b) US SCE

Notes: Average attention weight in the ECB survey of professional forecasters (panel (a)) and US survey of consumer expectations (panel (b)). The grey marker is the estimate assuming homogeneous attention weights ($G = 1$). The black marker is the estimated average attention weight from the finite mixture ($\boldsymbol{\pi}_G' \boldsymbol{k}_G$), along it 95% bootstrap confidence interval. Each horizontal line refers to a different dataset.

This difference is not only statistically but also economically significant. For instance, consumers' expectations are about half as responsive to aggregate news as implied by the homogeneous specification.

Consequently, the usual homogeneity assumption on attention does not only overlooks the diversity of attention, but it also generally leads to overestimate the average attention in the population. Figure D.2 in the Online Appendix illustrates the effect of the number of groups, G, for average attention. It shows that the average attention rapidly converges toward a value below its homogeneous counterpart as G increases.

4.5 Forecast-specific attention weights

The distribution characterized by the group-specific weights, $\boldsymbol{k}_G$, and prior membership probabilities, $\boldsymbol{\pi}_G$, is consistently estimated across the population and captures unconditional heterogeneity in attention weights.

Nevertheless, there are several reasons to also analyze the *ex post* heterogeneity retrieved from our clustering approach. For instance, as highlighted in Section 2, the macroeconomic impact of heterogeneous attention may be exacerbated depending on how these rigidities correlate with other factors, such as wealth, financial constraints, uncertainty, state of the economy, etc. Similarly, policymakers might employ communication strategies targeted at different groups of agents depending on their attention: success of

such targeted policies hinges on identifying observable characteristics that correlate with attention. In both cases, novel insights can be gained by analyzing how the *ex post* heterogeneity in attention, which we retrieve from our clustering approach, correlates with aggregate and idiosyncratic factors.

To this end, we compute forecast-specific attention weights, denoted $k_{i,t}$. These weights leverage the posterior membership probabilities that we obtain from our clustering. They are given by[12]

$$\hat{k}_{i,t} \;=\; \sum_{g=1}^{G} w_g\Big(R_{i,t}|F_{i,t-1};\hat{\boldsymbol{\theta}}_G\Big) \;\cdot\; \hat{k}_g. \tag{17}$$

Doing so, we are effectively constructing panel data on attention by attributing an (estimated) attention weight $\hat{k}_{i,t}$ to each observation (i,t) in our datasets.

Figure 3 presents the distribution of these forecast-specific attention weights for each dataset. These *ex post* distributions once again illustrate widespread heterogeneity in attention weights. They deviate notably from Gaussian distributions, particularly because they are typically multimodal and skewed.

The *ex post* heterogeneity revealed by these distributions cannot be attributed to clustering uncertainty. On the contrary, if clustering were entirely uncertain—providing no information—forecast-specific attention weights would remain concentrated at their prior value ($\boldsymbol{\pi}_G'\boldsymbol{k}_G$). Hence, deviations from a single-point distribution reflect the information revealed through clustering rather than merely introducing noise.

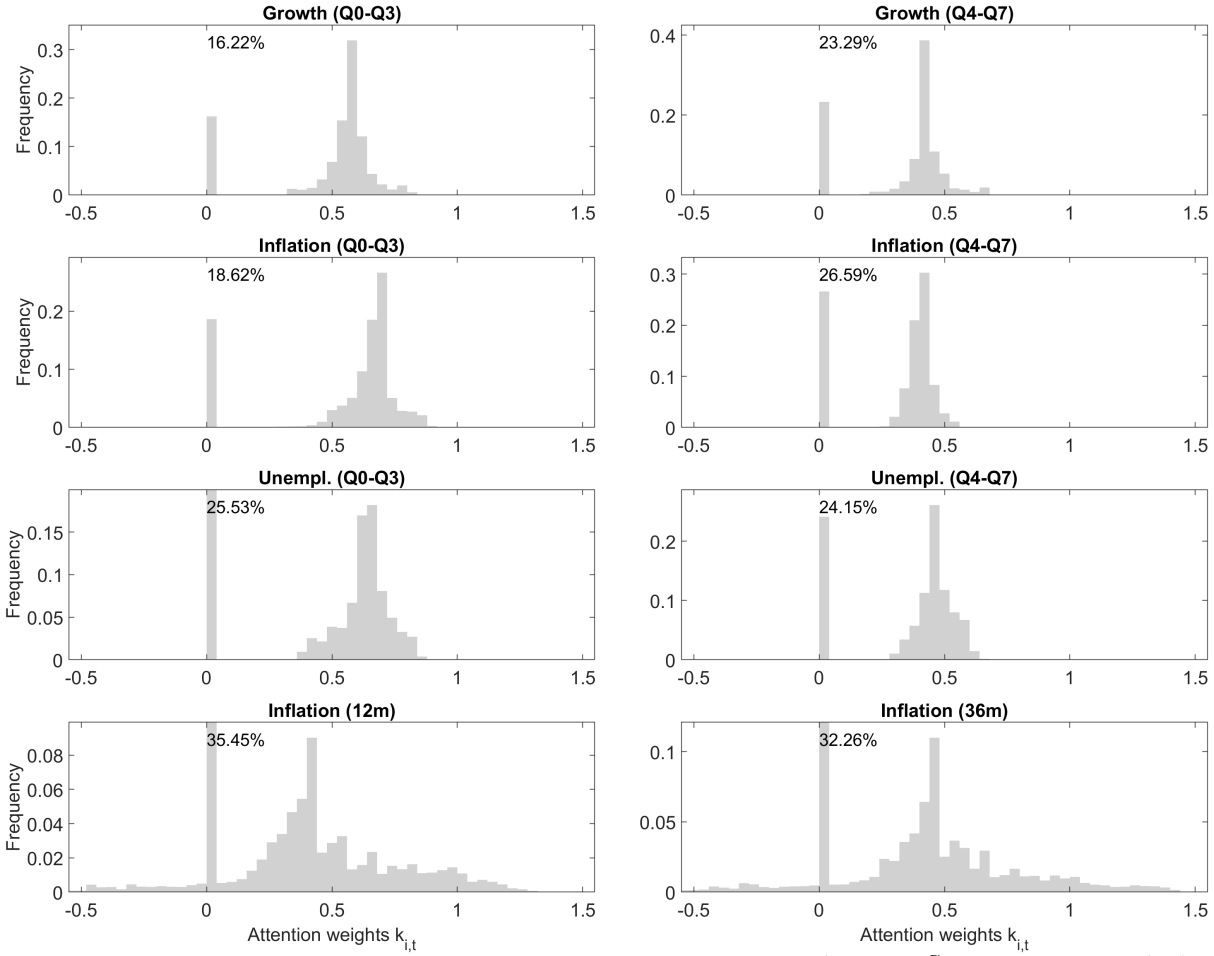
4.6 Attention weights, forecast accuracy, and information

This subsection provides suggestive evidence that our estimated attention weights positively correlate with both forecast accuracy and the extent of information acquisition. These findings support the interpretation of our estimated weights as a reliable measure of how much attention agents pay to incoming news.

Forecast accuracy We begin by examining to what extent our forecast-specific attention weights relate to forecast accuracy, measured by the absolute forecast error,

[12]These forecast-specific attention weights are consistent with our approach of favoring soft clustering to account for clustering uncertainty. Alternatively, we also consider a hard clustering approach where each forecast is associated with its most likely cluster in Online Appendix G.

Figure 3: Distributions of forecast-specific attention weights



Notes: Histograms of (estimated) forecast-specific attention weights, $\hat{k}_{i,t} = \sum_{g=1}^{G} w_g(R_{i,t}|F_{i,t-1};\hat{\boldsymbol{\theta}})\ \hat{k}_g$. Each panel refers to a different dataset. The reported text in the figure indicates the value of the bin with edges containing zero.

$|x_{t+h} - F_{i,t}|$, where, recall $F_{i,t} := F_{i,t}x_{t+h}$. Intuitively, one might be tempted to think of attention as a way for individuals to learn about the realization of economic variables, thus implying that increased attention improves forecast accuracy on average.

To asses if this is the case in our data, we estimate the following regression:

$$\log\left(|x_{t+h} - F_{i,t}|\right) \;=\; \beta_{FA}\,\hat{k}_{i,t} \;+\; \boldsymbol{X}'_{i,t}\boldsymbol{\gamma}_{FA} \;+\; \text{Error}^{FA}_{i,t} \tag{18}$$

where the coefficient β_{FA} is the semi-elasticity of the absolute forecast error with respect to $\hat{k}_{i,t}$. The vector $\boldsymbol{X}_{i,t}$ contains information about socioeconomic characteristics of individuals, dummies corresponding to fixed effects and a constant.

Columns (1) and (2) of Table 3 report estimates of β_{FA} across all dataset, pooling ECB SPF horizons. Except in the case of growth forecasts, the semi-elasticities are negative in all specifications, implying that increased attention is indeed associated with lower

absolute forecast errors. Furthermore, when individual fixed effects are included, these negative coefficients grow larger in absolute value, underscoring the role of unobserved heterogeneity.

Table 3: Regression results: Forecast Accuracy and Information

	Forecast Accuracy		Information	
	(1)	**(2)**	**(3)**	**(4)**
Growth (ECB SPF)				
Attention weight ($\hat{k}_{i,t}$)	0.110	0.126	0.024	0.011
	[0.004, 0.145]	[0.067, 0.229]	[-0.001, 0.037]	[-0.004, 0.033]
Observations	8,842		7,940	
Inflation (ECB SPF)				
Attention weight ($\hat{k}_{i,t}$)	-0.034	-0.073	0.076	0.074
	[-0.114, -0.021]	[-0.207, 0.029]	[0.064, 0.091]	[0.046, 0.109]
Observations	8,823		7,757	
Unempl. (ECB SPF)				
Attention weight ($\hat{k}_{i,t}$)	-0.181	-0.199	0.057	0.038
	[-0.276, -0.133]	[-0.347, -0.068]	[0.047, 0.078]	[0.019, 0.065]
Observations	8,122		7,203	
Inflation (12m, SCE)				
Attention weight ($\hat{k}_{i,t}$)	-0.166	-0.277	0.309	0.335
	[-0.212, -0.160]	[-0.357, -0.240]	[0.280, 0.377]	[0.297, 0.417]
Observations	120,033		117,355	
Inflation (36m, SCE)				
Attention weight ($\hat{k}_{i,t}$)	-0.106	-0.273	0.355	0.404
	[-0.198, -0.074]	[-0.405, -0.191]	[0.269, 0.482]	[0.296, 0.561]
Observations	96,643		117,716	
Ind. Fixed Effects	No	Yes	No	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Tenure Fixed Effects	Yes	Yes	Yes	Yes
Horizon Fixed Effects	SPF	SPF	SPF	SPF
Controls	SCE	SCE	SCE	SCE

Notes: Results for β_{FA} estimated from regression (18) (columns 1–2) and β_I from regression (19) (columns 3–4), using ECB SPF and US SCE data. The 95% bootstrap confidence intervals (in brackets) are robust to the first-stage estimation of $k_{i,t}$ (a generated regressor). Controls include time-invariant socioeconomic variables from the core US SCE (see 5.2 for details).

Because using the estimated attention weights, $\hat{k}_{i,t}$, as regressors ignores their estimation uncertainty, conventional standard errors are typically biased downward. To address this, Table 3 reports 95% confidence intervals obtained by jointly bootstrapping the finite mixture model and equation (18) using cluster wild bootstrap with fixed design. We draw clusters from the posterior distribution of the data. This method is robust to both serial correlation and cross-sectional dependence, and we apply it throughout when attention

weights appear as regressors.

Information We investigate how attention weights relate to the amount of information incorporated into forecasts, measured by the ratio of prior to posterior forecast variances. Under Gaussian prior and posterior, this ratio coincides with mutual information, the measure of information proposed by Sims (2003).

We obtain posterior forecast variances, $\tilde{\sigma}^2_{i,t|t}$, from the probabilistic beliefs reported in the US SCE and ECB SPF. The US SCE directly provides a variance measure for each forecast, while for the ECB SPF we follow D'Amico and Orphanides (2008) to compute posterior forecast variances. These posterior variances represent the forecast's uncertainty; we define prior uncertainty ($\tilde{\sigma}^2_{i,t|t-1}$) as the previous period's posterior uncertainty. We use tildes to emphasize that they refer to perceived variances and may, therefore, differ from those implied by the DGP.

We then examine how forecast-specific attention weights, $\hat{k}_{i,t}$, affect the incorporation of information, hypothesizing that higher attention implies more information. Specifically, we estimate

$$\log\left(\tilde{\sigma}^2_{i,t|t-1}\Big/\tilde{\sigma}^2_{i,t|t}\right) \;=\; \beta_I\,\hat{k}_{i,t} \;+\; \boldsymbol{X}'_{i,t}\,\boldsymbol{\gamma}_I \;+\; \text{Error}^I_{i,t} \tag{19}$$

where β_I captures the semi-elasticity of the information measure with respect to $k_{i,t}$.

Columns (3) and (4) in Table 3 report the estimated β_I. Regardless of whether individual fixed effects are included, β_I is positive across all datasets, indicating that greater attention weights in fact correspond to more information being incorporated. The associated 95% confidence intervals exclude zero in every case except for the ECB SPF growth dataset.

Overall, these results confirm that our estimated attention weights capture a learning channel through which agents acquire information about aggregate economic variables.

4.7 Robustness and simulation study

Robustness to misspecification We conduct several robustness and sensitivity checks to assess the validity of our results. First, we verify that our findings remain broadly consistent when the number of groups, G, is increased. Online Appendix D.1 presents

estimates for $G = 30$. Increasing G effectively partitions existing groups into finer neighboring clusters, with little effect on the overall range and distribution of attention weights.

Second, forecast revisions may exhibit systematic patterns linked to individual characteristics. Since these characteristics could correlate with our measure of news, we re-estimate equation (8) while controlling for socioeconomic variables from the (core) US SCE.[13] Details and estimation results are reported in Online Appendix D.2, which shows that controlling for socioeconomic factors has only marginal effects on the unconditional distribution of attention weights and leaves the average attention weight essentially unchanged.

Third, in Section 3.3, we adopt a parsimonious setup in which prior membership probabilities are assumed to be i.i.d. Online Appendix D.3 relaxes this assumption by allowing membership probabilities to depend on consumers' socioeconomic characteristics. The findings indicate that the i.i.d. assumption in Section 3.3 does not impose an unduly restrictive structure on the results.

Simulation study We also evaluate whether our approach can recover true heterogeneity in attention weights through a simulation exercise. Specifically, we generate 10,000 independent copies of two datasets: one mimicking inflation forecasts from the ECB SPF (Q0–Q3), and another mimicking inflation forecasts from the US SCE (12-month horizon). In each simulated dataset, forecast revisions are drawn from the estimated mixture distributions. We then apply the procedure described in Section 3.3, with results reported in Online Appendix E. The findings show that our estimates exhibit good finite-sample properties.

Finally, one might worry that finite-mixture estimation artificially inflates heterogeneity. To address this concern, we conduct additional simulations to test our model-selection criterion. First, we simulate data generated from a homogeneous regression ($G = 1$), as in equation (7). Our benchmark criterion correctly selects the homogeneous model ($G = 1$) in all 100 simulations for both datasets. Second, we simulate data using the values of G estimated in Section 4. Again, our cross-validation approach selects a parsimonious number of groups that is close to the true G. These findings suggest that our procedure does not spuriously generate excess heterogeneity.

[13] These variables are introduced in the next section.

5 Aggregate and idiosyncratic factors

This section leverages panel data on attention, constructed from forecast-specific attention weights, to establish new empirical facts about the effects of aggregate and idiosyncratic factors on attention. While both factors contribute to shaping attention heterogeneity, the majority of this heterogeneity remains unexplained by these factors alone.

5.1 Aggregate factors

We first investigate how aggregate factors affect the distribution of attention. To save on space, we focus on one-year-ahead inflation forecasts (US SCE). Figure 4, Panel (a), reports the evolution of average news, $\hat{\mu}_t - F_{t-1}$, where $F_{t-1} := n^{-1} \sum_{i=1}^{n} F_{i,t-1}$ is the average previous-period forecast, for one-year-ahead inflation forecasts in the US SCE. This measure of news aligns with significant economic and political events. In particular, it reveals that, at its onset, the expected impact of the pandemic was deflationary, consistent with market views at the time. We also find systematic peaks at three key moments when the media reported that inflation had reached its highest level since 2008, 1990, and 1982, respectively. Finally, news became negative after inflation peaked in June 2022.

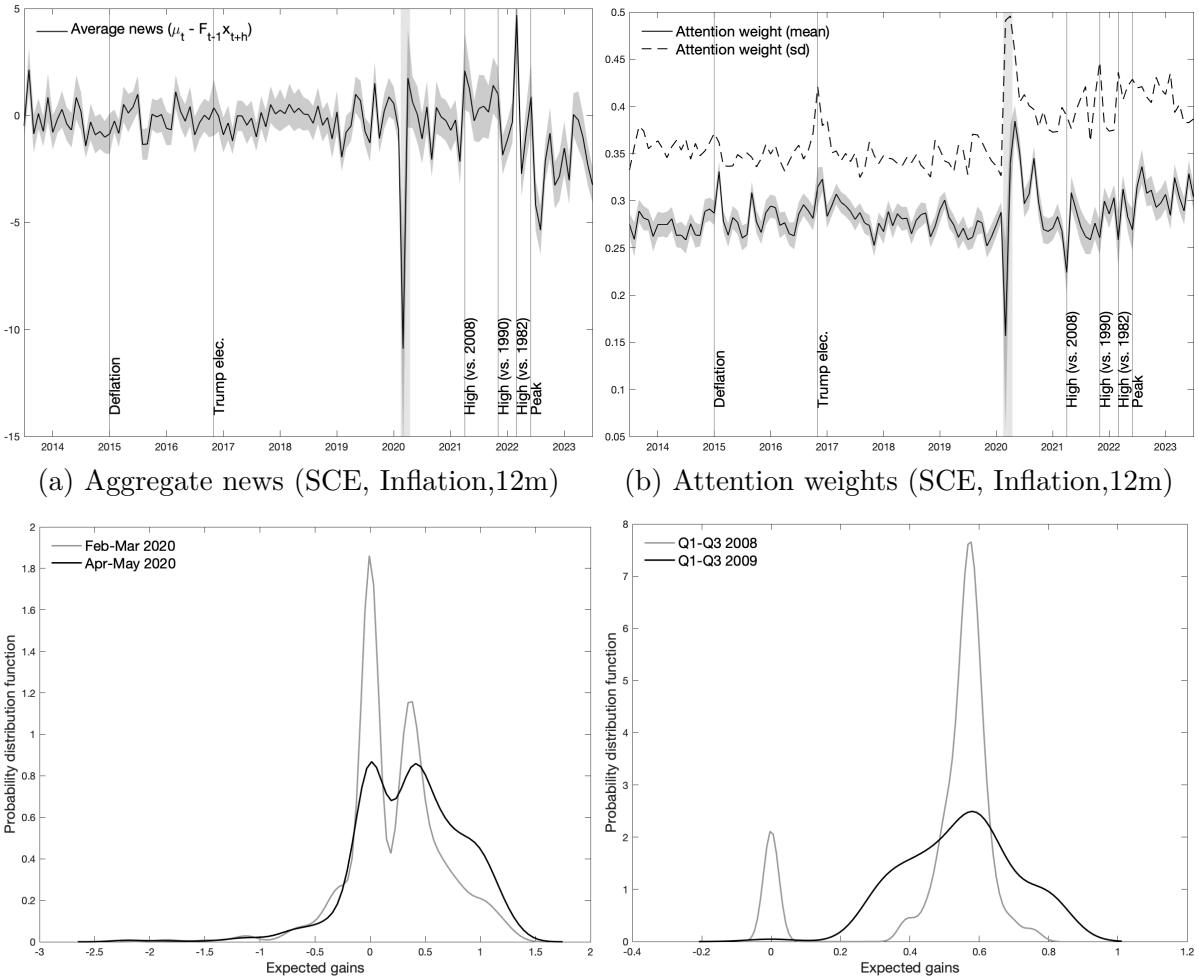
Panel (b) reports the evolution of the average expected attention weight (solid line), $\hat{k}_t := n^{-1} \sum_{i=1}^{n} \hat{k}_{i,t}$, each month.[14] Average attention has remained relatively stable over the last decade. However, certain periods show sudden shifts, often aligning with significant economic and political events. For instance, consumers paid more attention to inflation during the deflationary period of 2015, after Trump's election, and at the end of the COVID recession. Conversely, they were less attentive during the recent inflationary surge. Notably, our data show that consumers were inattentive to media reports stating that inflation had reached its highest level since 2008, 1990, and 1982, respectively.

Having estimated the full distribution of attention weights, our approach allows us to go beyond an analysis of its first moment. Panel (b) also reports the evolution of the standard deviation of attention weights. Like the average, it responds to economic and political events. In particular, it increased at the onset of the pandemic and has remained high since then.

Panels (c) and (d) further illustrate how aggregate shocks alter the full distribution

[14] See Figure F.10 in the Online Appendix for other datasets.

Figure 4: Attention over time



(a) Aggregate news (SCE, Inflation,12m)

(b) Attention weights (SCE, Inflation,12m)

(c) COVID (SCE, Inflation,12m)

(d) Financial crisis (SPF, Growth, Q0-Q3)

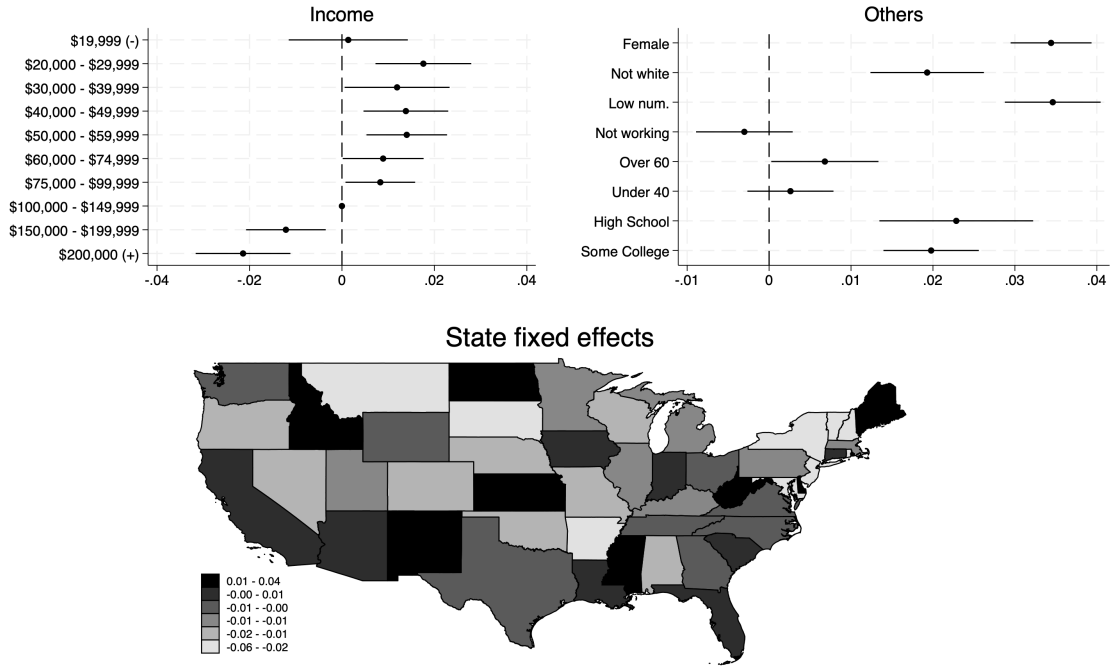
Notes: Panel (a) – Estimated average news, $\mu_t - F_{t-1}$, for one-year ahead inflation in the US SCE (SCE, Inflation,12m). The dark grey area represent 95% confidence intervals (clustered wild bootstrap). Vertical light grey areas indicate recessionary periods, as defined by the NBER. Panel (b) – Average attention weight (plain line) and standard deviation (dashed), $\hat{k}_{i,t}$ for each t, for one-year ahead inflation in the US SCE (SCE, Inflation,12m). The dark grey area represent 95% confidence intervals (clustered wild bootstrap). Panel (c) – Probability distributions of gains, $\hat{k}_{i,t}$, before and during COVID for one-year ahead inflation in the US SCE (SCE, Inflation,12m). Panel (d) – Probability distributions of gains, $\hat{k}_{i,t}$, before and during the 2008 financial crisis for current year growth in the ECB SPF (SPF, Growth, Q0-Q3).

of attention weights, focusing on the COVID and Global Financial Crises, respectively. Specifically, we estimate distributions prior to and at the time of the increase in average attention in April and May 2020 and in the first to third quarters of 2009, respectively. These increases in attention are not uniform across individuals. Instead, they result from a flattening of the right tail of the distribution, with more individuals becoming highly attentive and a missing mass of individuals with attention weights close to zero.

5.2 Socioeconomic factors

We investigate how socioeconomic characteristics affect consumers' attention weights. To this end, we rely on the information available in the core US SCE. Table F.3 reports regression results for our measures of attention to inflation (12m and 36m). These regressions estimate the effects of household pre-tax earnings, education, age, gender, employment conditions, numeracy scores, and race, while controlling for the month of the survey wave (time fixed effects), the number of times a respondent appeared in the survey (tenure fixed effects), and the state of residence (state fixed effects).

Figure 5: Attention and consumer characteristics



Notes: Coefficients from the regression of expected gains, $k_{i,t}$, for inflation (12m) on consumer characteristics reported in the US SCE (core). 95% confidence intervals based on Driscoll and Kraay standard errors with two lags. Low num. indicates a low numeracy score. Tenure fixed effects for the number of months in the survey, the month of the survey wave, and the state of the principal residence. The state fixed effects are reported on the map. See Table F.3 for details and results for other datasets.

Figure 5 summarizes the findings for one-year-ahead inflation forecasts. The top left panel shows that income correlates negatively with attention weights. Consumers reporting an annual income between \$20,000 and \$60,000 have an attention weight 0.03 higher than those with an income above \$200,000. The right panel indicates that female, non-white and individuals with a lower numeracy score tend to have higher attention weights. Finally, education has a negative effect on attention weights, with college educated having lower attention weights on average.

Overall, these effects suggest that poorer individuals may be the most sensitive to aggregate news about inflation, consistent with the view that inflation disproportionately affects them because they spend a larger share of their income on necessities that are sensitive to price changes (Erosa and Ventura, 2002). Similarly, women's higher attention may result from traditional gender roles, such as an increased exposure to highly volatile grocery prices (D'Acunto et al., 2021).

The bottom panel shows that attention heterogeneity also correlates with geographic variations. In particular, consumers tend to have higher attention weights in the South. Among large-population states, attention weights were relatively high in California, Texas, Florida, and Washington over the past decade. These states have also experienced relatively higher inflation rates during the same period.

5.3 Explanatory power

A striking finding is that aggregate and observable idiosyncratic factors explain little of the overall heterogeneity we retrieved.[15] To show this, we decompose the total sum of squares (SST) from the estimated forecast-specific attention weights, $\hat{k}_{i,t}$, and compute the explained sum of squares due to individual effects (SSE_i) and time effects (SSE_t). The quantity $(SST - SSE_i - SSE_t)/SST$ is what is left unexplained by either individual or time effects alone, which could depend on time-varying individual characteristics.

Table 4: Explanatory power: idiosyncratic and aggregate factors

	SSE_i/SST	SSE_t/SST	$\bar{R}^2$	$\mathbb{V}ar(\hat{k}_{i,t})/\mathbb{V}ar(k_{i,t})$	Corrected $\bar{R}^2$
Growth (Q0-Q3)	0.044	0.102	0.106	1.084	0.118
Growth (Q4-Q7)	0.045	0.086	0.088	1.076	0.098
Inflation (Q0-Q3)	0.033	0.104	0.097	1.050	0.104
Inflation (Q4-Q7)	0.044	0.062	0.062	1.040	0.066
Unempl. (Q0-Q3)	0.054	0.095	0.107	1.061	0.116
Unempl. (Q4-Q7)	0.065	0.073	0.092	1.067	0.101
Inflation (12m)	0.240	0.005	0.122	1.396	0.234
Inflation (36m)	0.258	0.004	0.142	1.478	0.287

Notes: Decomposition of the total sum of squares (SST) for $k_{i,t}$ between the explained sum of squares due to individual (SSE_i) and time (SSE_t) effects. $\bar{R}^2$ is the adjusted R-squared. $\mathbb{V}ar(\hat{k}_{i,t})/\mathbb{V}ar(k_{i,t})$ is the multiplicative correction to account for measurement error (see Appendix A.4) and used to compute the corrected adjusted R-squared.

The results of this decomposition are reported in the second and third columns of Table 4. Our analysis shows that heterogeneity in the ECB SPF primarily occurs over

[15] For instance, the R-squared (R^2) obtained from the regressions in Table F.3 is about 3%.

time, whereas within-individual variation is the main factor driving differences in the US SCE. Variations in attention weights across respondents explain approximately between 3% and 26% of the total variation in the ECB SPF and US SCE, respectively. Variations across time explain approximately between less than 10% and 1% of the total variation, respectively.[16] Thus, even when accounting for both sources of variation, the vast majority of the observed heterogeneity is driven by within-individual variations over time. This within-individual heterogeneity is overlooked when conditioning, *ex ante*, on individual- and time-specific factors, as in, e.g., Goldstein (2023).[17]

As we further discuss in Appendix A.4, our decomposition underestimates the actual explanatory power that would result from using the actual forecast-specific attention weights, $k_{i,t}$, rather than their noisy estimates retrieved from the mixture model, $\hat{k}_{i,t}$. Therefore, we propose a back-of-the-envelope correction to assess whether these low R^2 values are consequences of measurement errors. Specifically, we show that the R^2_k from $k_{i,t}$ is given by

$$R^2_k \;=\; R^2_{\hat{k}} \;\cdot\; \frac{\mathbb{V}ar(\hat{k}_{i,t})}{\mathbb{V}ar(k_{i,t})} = R^2_{\hat{k}} \;\cdot\; \frac{\mathbb{V}ar(\hat{k}_{i,t})}{\mathbb{V}ar(\hat{k}_{i,t}) - \sigma^2_e} \tag{20}$$

where $\sigma^2_e := \mathbb{V}ar(e_{i,t})$ is the variance of the estimation error, $e_{i,t} := \hat{k}_{i,t} - k_{i,t}$, which is estimated through our bootstrap procedure. Ratios in (20) are reported in the fifth column of Table 4. Finally, in the last column of Table 4, we compute the corrected adjusted R^2 (Corrected $\bar{R}^2$). This correction indicates that measurement errors are unlikely to be the primary reason for the small R^2 values.

6 Implications for theories of attention

This section investigates the determinants of our estimated attention weights, showing that they depart in systematic ways from the predictions of Bayesian updating and other

[16] Systematic differences in the explanatory power of idiosyncratic and time factors between the ECB SPF and US SCE are, at least partially, attributable to differences in survey design. For instance, time factors capture variations in the forecast horizon in the ECB SPF, whereas the horizon is fixed in the US SCE.

[17] The estimation of individual attention weights through *ex ante* conditioning is largely impractical in the US SCE, as each individual is observed revising their forecast an average of seven times. This difficulty is evident from the comparison between the R^2 (sum of the second and third columns) and the adjusted-R^2 ($\bar{R}^2$) in Table 4. Instead, what is feasible, is to group individuals based on shared characteristics. As we have seen, socioeconomic variables explain only about 3%.

recent theoretical frameworks in macroeconomics reported in Table 1.

6.1 Deviations from Bayesian updating

The attention weights we estimate are derived from reduced-form equations without structural restrictions on their determinants. As a result, they provide a natural setting to test theories that endogenize these weights. A key example is Bayesian updating—as in noisy information and rational inattention models (Lucas, 1972; Sims, 2003).

Testing for Bayesian updating Under Gaussian priors and posteriors with respective variances $\sigma^2_{i,t|t}$ and $\sigma^2_{i,t|t-1}$, the Bayesian attention weight is

$$k^B_{i,t} \;=\; \frac{\sigma^2_{i,t|t-1} - \sigma^2_{i,t|t}}{\sigma^2_{i,t|t-1}}\,. \tag{21}$$

To investigate whether our estimated attention weights, $\hat{k}_{i,t}$, align with this Bayesian benchmark, we estimate:

$$\log\left(\tilde{\sigma}^2_{i,t|t-1} - \tilde{\sigma}^2_{i,t|t}\right) \;= \\ \beta_{BU}\log\left(\hat{k}_{i,t}\right) \;+\; \delta_{BU}\,\log\left(\tilde{\sigma}^2_{i,t|t-1}\right) \;+\; \boldsymbol{X}'_{i,t}\,\boldsymbol{\gamma}_{BU} \;+\; \text{Error}^{BU}_{i,t} \tag{22}$$

where we substituted prior and posterior variances by their self-reported measures of uncertainty ($\tilde{\sigma}^2_{i,t|t-1}$ and $\tilde{\sigma}^2_{i,t|t}$, respectively). We derive (22) by replacing $k^B_{i,t}$ with $\hat{k}_{i,t}$ in (21), taking logs, and rearranging terms to respect Bayesian timing assumptions,[18] then adding controls and an error term. Under strict Bayesian updating, $\beta_{BU} = \delta_{BU} = 1$.

Table 5, columns (1) and (2), show the estimates of β_{BU} and δ_{BU}. Across specifications, β_{BU} lies between -0.01 and 0.11, while δ_{BU} often exceeds one. The confidence intervals associated with δ_{BU} are very narrow, especially for the SPF and excluding individual fixed effects. An F-test strongly rejects the joint hypothesis $H_0 : \beta_{BU} = \delta_{BU} = 1$ in favor of the alternative, H_1 : "not H_0", at all levels of confidence. In columns (3) and (4) of Table 5, we further impose $\delta_{BU} = 1$, yet the confidence intervals associated to β_{BU} do not include 1, ruling out sampling variability as a driving factor. We therefore conclude that our attention weights deviate from the strict Bayesian benchmark.

[18] In these models, individuals begin the period with a prior (state), observe the precision of the signal, set $k^B_{i,t}$ (choice), and form their posterior (outcome).

Table 5: Regression results: Bayesian updating

	Unrestricted		Restricted	
	(1)	**(2)**	**(3)**	**(4)**
Growth (ECB SPF)				
Attention weight ($\hat{k}_{i,t}$)	-0.001	-0.004	-0.001	-0.004
	[-0.002, -0.001]	[-0.022, 0.013]	[-0.002, -0.001]	[-0.021, 0.012]
$\tilde{\sigma}^2_{i,t\|t-1}$	0.991	1.044	1.00	1.00
	[0.991, 0.991]	[1.035, 1.052]	-	-
F-stat ($H_0 : \beta_{BU} = \delta_{BU} = 1$)	14,743.13	15,525.37		
Observations		4,390		
Inflation (ECB SPF)				
Attention weight ($\hat{k}_{i,t}$)	-0.011	-0.013	-0.010	-0.012
	[-0.011, -0.010]	[-0.035, 0.009]	[-0.011, -0.010]	[-0.031, 0.007]
$\tilde{\sigma}^2_{i,t\|t-1}$	1.024	1.187	1.00	1.00
	[1.024, 1.024]	[1.181, 1.193]	-	-
F-stat ($H_0 : \beta_{BU} = \delta_{BU} = 1$)	16,597.45	17,741.86		
Observations		4,170		
Unempl. (ECB SPF)				
Attention weight ($\hat{k}_{i,t}$)	-0.006	-0.010	-0.007	-0.011
	[-0.006, -0.005]	[-0.030, 0.009]	[-0.007, -0.006]	[-0.034, 0.011]
$\tilde{\sigma}^2_{i,t\|t-1}$	0.779	0.878	1.00	1.00
	[0.779, 0.779]	[0.871, 0.884]	-	-
F-stat ($H_0 : \beta_{BU} = \delta_{BU} = 1$)	21,198.13	24,465.03		
Observations		3,691		
Inflation (12m, SCE)				
Attention weight ($\hat{k}_{i,t}$)	0.094	0.051	0.118	0.058
	[0.093, 0.095]	[0.050, 0.052]	[0.117, 0.120]	[0.058, 0.059]
$\tilde{\sigma}^2_{i,t\|t-1}$	1.910	1.413	1.00	1.00
	[1.902, 1.911]	[1.408, 1.415]	-	-
F-stat ($H_0 : \beta_{BU} = \delta_{BU} = 1$)	36,992.65	49,077.78		
Observations		55,085		
Inflation (36m, SCE)				
Attention weight ($\hat{k}_{i,t}$)	0.109	0.055	0.130	0.061
	[0.108, 0.110]	[0.054, 0.055]	[0.130, 0.132]	[0.061, 0.062]
$\tilde{\sigma}^2_{i,t\|t-1}$	1.876	1.379	1.00	1.00
	[1.867, 1.878]	[1.372, 1.379]	-	-
F-stat ($H_0 : \beta_{BU} = \delta_{BU} = 1$)	35,831.41	49,461.00		
Observations		55,325		
Ind. Fixed Effects	No	Yes	No	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Tenure Fixed Effects	Yes	Yes	Yes	Yes
Horizon Fixed Effects	SPF	SPF	SPF	SPF
Controls	SCE	SCE	SCE	SCE

Notes: Results for β_{BU} and δ_{BU} estimated from regression (22), using ECB SPF and US SCE data. Under each coefficient, the 95% bootstrap confidence intervals (in brackets) are robust to the first-stage estimation of $k_{i,t}$ (a generated regressor). The F statistic tests the null hypothesis $H_0 : \beta_{BU} = \delta_{BU} = 1$ against the alternative H_1 : not H_0. In columns (3) and (4) δ_{BU} is restricted to 1.00, thus confidence intervals are not available. Controls include time-invariant socioeconomic variables from the core US SCE.

A further sign of non-Bayesian behavior is that roughly half of the sample exhibits higher posterior than prior uncertainty, which cannot occur under Bayesian updating. The smaller sample used in Table 5 excludes these observations. Moreover, as we have previously documented, some attention weights extend beyond the unit interval, highlighting yet another deviation from Bayesian updating.

Beyond Bayesian updating A key advantage of specification (22) is that it allows testing for strict Bayesian updating as well as for some deviations that have been recently proposed in the macroeconomic literature (see Table 1). As we explain below, this generality arises from the fact that it relies on (i) self-reported measures of uncertainty rather than data-generating-process-consistent variances, $\sigma^2_{i,t|t}$ and $\sigma^2_{i,t|t-1}$, and (ii) a log-transformation of the attention weights.

Recently, many macroeconomic theories have attributed deviations from Bayesian updating to individuals' misperception of signal precision or the underlying process of the variable being forecasted. In these theories, individuals act as if they were Bayesian but hold erroneous perceptions of the fundamentals driving Bayesian weights. Since specification (22) relies on self-reported prior and posterior uncertainties, deviations of β_1 and β_2 from 1 are difficult to reconcile with these theories, which include, but are not limited to, overconfidence (Broer and Kohlhas, 2022), and over- or under-extrapolation (Angeletos et al., 2021; Ryngaert, 2024).[19]

Similarly, since it employs a log transformation, specification (22) also encompasses theories in which the actual gain is proportional to the Bayesian gain. This is the case, in particular, for diagnostic expectations (Bordalo et al., 2020), where, for some constant θ, the attention weight is given by

$$k^d_{i,t} \;=\; (1+\theta)\; k^B_{i,t}, \tag{23}$$

where $k^B_{i,t}$ is defined in equation (21).

[19]For instance, when individuals are overconfident about the signal precision, the attention weight is $k^o_{i,t} = 1 - \tilde{\sigma}^2_{i,t|t}/\sigma^2_{i,t|t-1}$, where $\tilde{\sigma}^2_{i,t|t}$ is the *perceived* posterior uncertainty, which is lower than the actual posterior uncertainty $\sigma^2_{i,t|t}$. Since we use self-reported uncertainties, these already coincide with $\tilde{\sigma}^2_{i,t|t}$.

6.2 Generalized attention weights

While equation (22) tests a central implication of Bayesian updating (and close variants), it does not reveal why these models fail. To address this question, we next investigate how $\hat{k}_{i,t}$ responds to perceived prior uncertainty and signal noise.

Generalized attention weights We propose the following generalization of the Bayesian weight formula, allowing for flexible elasticities with respect to the perceived prior variance and the perceived signal variance, as well as a multiplicative disturbance $\theta_{i,t}$ (in the spirit of (23)):

$$k^G_{i,t} \;=\; (1+\theta_{i,t}) \cdot \frac{\left[\tilde{\sigma}^2_{i,t|t-1}\right]^{\eta_1}}{\left[\tilde{\sigma}^2_{i,t|t-1} \,+\, \tilde{\sigma}^2_{\omega,i,t}\right]^{\eta_2}} \tag{24}$$

where $\tilde{\sigma}^2_{\omega,i,t}$ is the perceived variance of the signal noise, and η_1 and η_2 measure how strongly attention responds to perceived prior uncertainty and perceived signal variance, $(\tilde{\sigma}^2_{i,t|t-1} + \tilde{\sigma}^2_{\omega,i,t})$, respectively. If $\eta_1 = \eta_2 = 1$ (and $\theta_{i,t} = 0$), this recovers the standard Bayesian form. Deviations from unity let agents under-adjust (when < 1) or over-adjust (when > 1) to uncertainty and signal precision.

These generalized weights also allow for potential disturbances, represented by $\theta_{i,t}$. We interpret these as a way to allow for under-reaction to news when $\theta_{i,t} < 0$, which can arise due to factors such as conservatism (Edwards, 1968), and over-reaction to news when $\theta_{i,t} > 0$, which can result from base rate neglect and diagnostic expectations (Tversky and Kahneman, 1974; Bordalo et al., 2020). In the following, we treat $\theta_{i,t}$ as nuisances, rather than coefficients to estimate. To the extent that we control for individual and other fixed effects, these do not have to average to zero and may shift the ratio in (24).

Perceived signal precision To bring (24) to the data, we compute a forecast-specific measure of perceived signal noise variance $\tilde{\sigma}^2_{\omega,i,t}$, which is coherent with our attention weights ($k_{i,t}$). Under the linear forecast rule (6), we have

$$\tilde{\sigma}^2_{\omega,i,t} \;=\; \left(\tilde{\sigma}^2_{i,t|t} \,-(1-k_{i,t})^2\,\tilde{\sigma}^2_{i,t|t-1}\right)\Big/ k^2_{i,t}. \tag{25}$$

By plugin principle, we use this expression together with $\hat{k}_{i,t}$ and the self-reported uncertainties to compute $\tilde{\sigma}^2_{\omega,i,t}$ for each forecast.[20]

Elasticities Define by $\hat{\tilde{\sigma}}^2_{\omega,i,t}$ the estimated noise variance, we estimate η_1 and η_2 from the regression

$$\log(\hat{k}_{i,t}) \;=\; \eta_1 \log\left(\tilde{\sigma}^2_{i,t|t-1}\right) \;-\; \eta_2 \log\left(\tilde{\sigma}^2_{i,t|t-1} \;+\; \hat{\tilde{\sigma}}^2_{\omega,i,t}\right) + \boldsymbol{X}'_{i,t}\,\boldsymbol{\gamma}_{GW} \;+\; \text{Error}^{GW}_{i,t} \qquad (26)$$

Table 6 reports the results. Elasticity η_2 is read by flipping the sign of the coefficient associated with $\log\left(\tilde{\sigma}^2_{i,t|t-1,t} \;+\; \hat{\tilde{\sigma}}^2_{\omega,i,t}\right)$ in Table 6. Across all datasets, η_1 and η_2 are positive and statistically significant—that is, individuals do respond to changes in prior uncertainty and the signal perceived variance, as predicted by Bayesian updating. However, both coefficients are estimated to be less than one, indicating an under-adjustment relative to the Bayesian benchmark (where $\eta_1 = \eta_2 = 1$). The R^2 suggests that simple generalized attention weights account for a significant share of the variance of $\hat{k}_{i,t}$.

Table 6: Elasticity to prior uncertainty and signal precision

	ECB SPF			US SCE	
	Growth	Inflation	Unempl.	Inf. (12m)	Inf. (36m)
	(1)	(2)	(3)	(4)	(5)
$\log(\tilde{\sigma}^2_{i,t\|t-1})$	0.316***	0.300***	0.284***	0.226***	0.210***
	(0.029)	(0.047)	(0.048)	(0.008)	(0.009)
$\log(\tilde{\sigma}^2_{i,t\|t-1} + \hat{\tilde{\sigma}}^2_{\omega,i,t})$	-0.674***	-0.657***	-0.691***	-0.547***	-0.549***
	(0.009)	(0.009)	(0.012)	(0.001)	(0.001)
Tenure FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.8825	0.8807	0.8827	0.8945	0.8977
Observations	7101	6876	6412	81369	83931

Notes: Coefficients η_1 and $-\eta_2$ from regression (26). Driscoll and Kraay standard errors with two lags are reported in parentheses. In the ECB SPF, individual units are defined as the interaction between a forecaster and forecast horizon (current-year or next-year).

Taking stocks The evidence presented in this section shows that while our estimated attention weights move in directions that qualitatively resemble Bayesian updating (see Table 5), they do so less strongly than a fully rational Bayesian model would predict. A 1% increase in prior uncertainty raises the attention weight by only about 0.2–0.3%

[20] We disregard observations for which the implied variance is negative. These cases represent about 10% of observations in the ECB SPF and 25% in the US SCE.

(rather than 1%), and a 1% increase in perceived signal variance lowers the attention weight by about 0.5–0.7% (rather than 1%). Thus, although individuals become more attentive when facing higher uncertainty and do smooth out noisy signals, they do so to a lesser extent than predicted by the Bayesian benchmark. These under-adjustments may reflect the cognitive complexity involved in optimally trading off a signal's informativeness against its noisiness.

7 Conclusions

This paper introduces a novel approach to uncover attention heterogeneity in survey of consumers (US SCE) and professional forecasters (ECB SPF) using finite mixture models. We find that attention heterogeneity is substantial, and attempts to retrieve it by conditioning *ex ante* on observable characteristics miss much of the total heterogeneity.

Furthermore, we demonstrate how statistical clustering can be employed to construct panel datasets on attention. As we illustrate, such dataset can be used to study the impact of aggregate shocks on the full distribution of attention, how attention correlates with socio-economic characteristics, as well as to inform theories of expectation formation.

The tools developed in this paper are broadly applicable and can be used to provide empirical evidence for better understanding how individuals form expectations about macroeconomic variables from decade-long surveys. More generally, we hope this paper motivates further use of finite mixture models to construct panel datasets on parameters that must be estimated and behaviors that are not directly observable.

References

Andrade, P. and Le Bihan, H. (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, 60(8):967–982.

Angeletos, G.-M., Huo, Z., and Sastry, K. A. (2021). Imperfect macroeconomic expectations: Evidence and theory. *NBER Macroeconomics Annual*, 35(1):1–86.

Armantier, O., Nelson, S., Topa, G., Van der Klaauw, W., and Zafar, B. (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics*, 98(3):503–523.

Armantier, O., Topa, G., Van der Klaauw, W., and Zafar, B. (2017). An overview of the survey of consumer expectations. *Economic Policy Review*, 23-2:51–72.

Baley, I. and Turén, J. (2024). Lumpy forecasts. Technical report, Mauscript.

Benhima, K. and Bolliger, E. (2022). Do local forecasters have better information? *Available at SSRN 4294565*.

Boccanfuso, J. (2022). Consumption response heterogeneity and dynamics with an inattention region. Technical report, Quaderni-Working Paper DSE.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–2782.

Broer, T., Kohlhas, A., Mitman, K., and Schlafmann, K. (2022). Expectation and wealth heterogeneity in the macroeconomy. Technical report, Manuscript.

Broer, T. and Kohlhas, A. N. (2022). Forecaster (mis-) behavior. *Review of Economics and Statistics*, pages 1–45.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Capistrán, C. and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, 41(2-3):365–396.

Celeux, G., Frühwirth-Schnatter, S., and Robert, C. P. (2019). Model selection for mixture models–perspectives and strategies. In *Handbook of mixture analysis*, pages 117–154. Chapman and Hall/CRC.

Cheng, R. C. H. and Liu, W. B. (2001). The Consistency of Estimators in Finite Mixture Models. *Scandinavian Journal of Statistics*, 28(4):603–616.

Coibion, O., Georgarakos, D., Gorodnichenko, Y., Kenny, G., and Weber, M. (2024). The effect of macroeconomic uncertainty on household spending. *American Economic Review*, 114(3):645–677.

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

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–2678.

D'Amico, S. and Orphanides, A. (2008). Uncertainty and disagreement in economic forecasting. Technical report, FRB working paper.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

D'Acunto, F., Malmendier, U., and Weber, M. (2021). Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences*, 118(21):e2008534118.

Edwards, W. (1968). Conservatism in human information processing. *New York: Wiley& Sons*.

Erosa, A. and Ventura, G. (2002). On inflation as a regressive consumption tax. *Journal of Monetary Economics*, 49(4):761–795.

Feng, Z. D. and McCulloch, C. E. (1996). Using Bootstrap Likelihood Ratios in Finite Mixture Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(3):609–617.

Garrigos, G. and Gower, R. M. (2024). Handbook of convergence theorems for (stochastic) gradient methods.

Gemmi, L. and Mihet, R. (2024). Household belief formation in uncertain times. *Swiss Finance Institute Research Paper*.

Gemmi, L. and Valchev, R. (2023). Biased surveys. Technical report, National Bureau of Economic Research.

Goldstein, N. (2023). Tracking inattention. *Journal of the European Economic Association*, 21(6):2682–2725.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *Journal of economic literature*, 61(1):3–40.

Hennig, C. (2000). Identifiablity of models for clusterwise linear regression. *Journal of Classification*, 17(2):273–296.

Kohlhas, A. N. and Robertson, D. (2025). Cautious expectations. *Journal of Monetary Economics*, page 103759.

Korenok, O., Munro, D., and Chen, J. (2023). Inflation and attention thresholds. *Review of Economics and Statistics*, pages 1–28.

Kumar, S., Gorodnichenko, Y., and Coibion, O. (2023). The effect of macroeconomic uncertainty on firm decisions. *Econometrica*, 91(4):1297–1332.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of econometrics*, 95(2):391–413.

Lewis, D., Melcangi, D., and Pilossoph, L. (2024). Latent heterogeneity in the marginal propensity to consume. Technical report, National Bureau of Economic Research.

Link, S., Peichl, A., Roth, C., and Wohlfart, J. (2025). Attention to the macroeconomy.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of economic theory*, 4(2):103–124.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273.

Maćkowiak, B. and Wiederholt, M. (2024). Rational inattention during an RCT. Technical report, Manuscript.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual review of statistics and its application*, 6:355–378.

McLachlan, G. J. and Peel (2000). *Finite Mixture Models*. John Wiley & Sons.

Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2):267–278.

Ottaviani, M. and Sørensen, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics*, 81(2):441–466.

O'Hagan, A., Murphy, T. B., Scrucca, L., and Gormley, I. C. (2019). Investigation of parameter uncertainty in clustering using a Gaussian mixture model via jackknife, bootstrap and weighted likelihood bootstrap. *Computational Statistics*, 34(4):1779–1813.

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.

Pfäuti, O. (2023). The inflation attention threshold and inflation surges. *arXiv preprint arXiv:2308.09480*.

Ryngaert, J. (2024). What do (and don't) forecasters know about us inflation? *Journal of Money, Credit and Banking*.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Spall, J. C. (2005). *Introduction to stochastic search and optimization: estimation, simulation, and control*. John Wiley & Sons.

Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. Applied section. Wiley.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.

Weber, M., Candia, B., Afrouzi, H., Ropele, T., Lluberas, R., Frache, S., Meyer, B. H., Kumar, S., Gorodnichenko, Y., Georgarakos, D., et al. (2023). Tell me something I don't already know: Learning in low and high-inflation settings. Technical report, National Bureau of Economic Research.

Wu, C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11(1):95–103.

Yakowitz, S. J. and Spragins, J. D. (1968). On the Identifiability of Finite Mixtures. *The Annals of Mathematical Statistics*, 39(1):209 – 214.

Yao, W. and Xiang, S. (2024). *Mixture Models: Parametric, Semiparametric, and New Directions*. Chapman and Hall/CRC, 1 edition.

A Mathematical Appendix

A.1 Conditions for identification of the parameters

This appendix provides necessary and sufficient conditions for identification of the model parameters. We consider the Gaussian mixture of regressions, that is the univariate conditional density of the observations is given by $R_{i,t}|F_{i,t-1} \sim f(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G)$ with $f := \sum_{g=1}^{G} \pi_g f_g(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G)$ and $f_g(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G) := \mathcal{N}(k_g(\mu_t - F_{i,t-1}), \sigma_g^2)$. A necessary and sufficient condition for the identifiability of $f(\bullet)$ is that the set of all G distributions is a linearly independent set over the set of real numbers, $\mathbb{R}$ (Yakowitz and Spragins, 1968). Thus, components f_g are identifiable if f is identifiable and if $(k_g(\mu_t - F_{i,t-1}), \sigma_g^2)_{g=1}^{G}$ is a collection of distinct pairs of conditional means and variances (Titterington et al., 1985). Consider now fixed $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, Hennig (2000) notes that, other than identifiability of f, a sufficient condition for identification of $(\boldsymbol{k}_G', \boldsymbol{\sigma}_G', \boldsymbol{\pi}_G')'$ is that observations do not lie (entirely) on any union of G arbitrary lines. If observations did happen to be covered by such a union, given $\boldsymbol{\mu}_0$, there could be multiple ways to choose $(\boldsymbol{k}_G', \boldsymbol{\sigma}_G', \boldsymbol{\pi}_G')'$ that fit the data exactly the same way, preventing identification.

In the following, we relax the assumption that $\boldsymbol{\mu}$ is fixed at $\boldsymbol{\mu}_0$ and consider identification of the full vector of parameters, $\boldsymbol{\theta}_G := (\boldsymbol{k}_G', \boldsymbol{\sigma}_G', \boldsymbol{\pi}_G', \boldsymbol{\mu}')'$. In practice, the common

factor $\{\mu_t\}_{t=1}^T$ appears as a latent process that is potentially correlated with the past forecast, $F_{i,t-1}$, and here is treated as a parameter to be estimated.

The next definition provides the concept of identifiability, used to prove identifiability of $\boldsymbol{\theta}_G$.

Definition 1 (Identifiability of $\boldsymbol{\theta}_G$)**.** *Consider the Gaussian mixture of regressions, $R_{i,t}|F_{i,t-1} \sim f(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G)$. For a given $F_{i,t-1}$ and number of groups G, $\boldsymbol{\theta}_G$ is said to be identifiable if for any two parameters $\boldsymbol{\theta}_G$, and $\tilde{\boldsymbol{\theta}}_G$,*

$$\sum_{g=1}^{G} \pi_g \mathcal{N}\Big(k_g(\mu_t - F_{i,t-1}),\, \sigma_g^2\Big) = \sum_{g=1}^{G} \tilde{\pi}_g \mathcal{N}\Big(\tilde{k}_g(\tilde{\mu}_t - F_{i,t-1}),\, \tilde{\sigma}_g^2\Big)$$

for each $i = 1, ..., n$, $t = 1, ..., T$ implies $\boldsymbol{\theta}_G = \tilde{\boldsymbol{\theta}}_G$ up to permutation.

The following proposition establishes identification of $\boldsymbol{\theta}_G$. It is separated in two parts, one with $G = 1$ and one with $G > 1$. While the case $G > 1$ entails standard assumptions for identification of finite mixture of regressions, the case $G = 1$ requires to satisfy an additional condition. In the paper, this allows us to depart from the existing literature on expectations formation models, identifying homogeneous attention weights and aggregate factors jointly.

Assumption 1. *The following conditions hold*

(i) $\Big(k_g,\, \sigma_g^2\Big)_{g=1}^{G}$ *forms a collection of distinct pairs of parameters*

(ii) $\sum_{g=1}^{G} \pi_g = 1$, *and* $\sigma_g, \pi_g > 0$ *for* $g = 1, ..., G$

(iii) $G < H$, *where H is the number of distinct affine lines $\mathcal{L}_1, ..., \mathcal{L}_H \subset \mathbb{R}^2$ such that all $(R_{i,t}, \mu_t - F_{i,t-1})$ of $\{(R_{i,t}, \mu_t - F_{i,t-1})\}_{i=1,t=1}^{n,T}$ lie in the union $\cup_{h=1}^{H} \mathcal{L}_h$ (Hennig, 2000).*

Proposition 1 (Identification conditions for $\boldsymbol{\theta}_G$)**.** *Under Assumption 1, we have*
(i) If $G = 1$, A necessary and sufficient condition for the point identification of $\boldsymbol{\theta}_1$ is

$$\mathbb{V}ar_t(F_{i,t-1}) > 0 \;\; \textit{for some } t, \tag{A.1}$$

where $\mathbb{V}ar_t(\bullet)$ is the conditional variance operator.

(ii) If $G > 1$, $\boldsymbol{\theta}_G$ is point identified up to label permutation.

Proof. The proof of (i) is in two parts. *Necessity* is proved by contradiction. We show that if condition (A.1) is violated, parameters $(\boldsymbol{k}_1, \boldsymbol{\mu})$ are not identified and therefore $\boldsymbol{\theta}_1$ is not identified. Assume that $\boldsymbol{\theta}_1$ is identified, yet $\mathbb{V}ar_t(F_{i,t-1}) = 0$ for all t, hence $F_{t-1} \equiv F_{i,t-1}$. Then, there exist rotations $o_1 \in \mathbb{R}, o_1 \neq 0$, such that $R_{i,t}|F_{i,t-1} \sim f(R_{i,t}, F_{i,t-1}; \tilde{\boldsymbol{\theta}}_G)$ with $f(R_{i,t}, F_{i,t-1}; \tilde{\boldsymbol{\theta}}_G) := \mathcal{N}(\tilde{k}_1(\tilde{\mu}_t - F_{t-1}), \sigma_1^2)$, $\tilde{k}_1 = k_1/o_1$, and $\tilde{\mu}_t = (1 - o_1)F_{t-1} + o_1\mu_t$, for which $f(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G) = f(R_{i,t}, F_{i,t-1}; \tilde{\boldsymbol{\theta}}_G)$ and $(k_1, \sigma_1, \boldsymbol{\mu}) \neq (\tilde{k}_1, \sigma_1, \tilde{\boldsymbol{\mu}})$. Therefore, the collection of parameters $(k_1, \sigma_1, \boldsymbol{\mu})$ is not identifiable according to Definition 1. This contradicts the statement that $(k_1, \sigma_1, \boldsymbol{\mu})$ is identified, hence proving that $\mathbb{V}ar_t(F_{i,t-1}) > 0$ for some t is a necessary condition for identification.

Sufficiency requires to demonstrate that if condition (A.1) is satisfied, the parameters $(k_1, \sigma_1, \boldsymbol{\mu})$ are jointly identified. It is enough to show that if $\mathbb{V}ar_t(F_{i,t-1})$ for some t, any attempt to apply a non-trivial rotation leads to different models. Consider the rotation applied to μ_t, $\tilde{\mu}_{i,t} = (o_g - 1)F_{i,t-1} + o_g\mu_t$. This has to be common along the cross-sectional dimension (index i). Namely, for any two indices, say i, j, $\tilde{\mu}_{i,t}$ must be equal to $\tilde{\mu}_{j,t}$, which holds only for $o_g = 1$. Finally, $(k_1, \sigma_1, \boldsymbol{\mu})$ has the same density as $(\tilde{k}_1, \sigma_1, \tilde{\boldsymbol{\mu}})$ at the unique point $(k_1, \sigma_1, \boldsymbol{\mu}) = (\tilde{k}_1, \sigma_1, \tilde{\boldsymbol{\mu}})$, since the only admissible o_g is restricted to unity.

This is contrast to the linear regression model, where the sufficient condition for identification of the slope is that the unconditional variance (as opposed to the conditional variance) of the covariates is positive definite.

Part (ii) follows straightforwardly from part (i) acknowledging that for $G > 1$, the only possible rotation o_g that for which $f(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G) = f(R_{i,t}, F_{i,t-1}; \tilde{\boldsymbol{\theta}}_G)$ implies $(k_1, \sigma_1, \boldsymbol{\mu}) = (\tilde{k}_1, \sigma_1, \tilde{\boldsymbol{\mu}})$ is $o_g = 1$ for all $g = 1, ..., G$. □

A.2 MLE Consistency

Under standard regularity conditions, the maximum-likelihood estimator (MLE) exists, is consistent, and admits an asymptotic normal distribution at standard rates for $\hat{\boldsymbol{k}}_G^{MLE}, \hat{\boldsymbol{\sigma}}_G^{MLE}$, $\hat{\boldsymbol{\mu}}^{MLE}$, and $\hat{\boldsymbol{\pi}}_G^{MLE}$ (Feng and McCulloch, 1996; Cheng and Liu, 2001). We formalize the consistency of the MLE in Proposition 2, where we also include the convergence rates of the MLE for the various sub-parameters of $\boldsymbol{\theta}_G$.

Assumption 2. *Assume*

1. $\log L_0(\boldsymbol{R}, \boldsymbol{F}, \boldsymbol{d}, \boldsymbol{\theta}_G) := \mathbb{E}[\log L(\boldsymbol{R}, \boldsymbol{F}, \boldsymbol{d}, \boldsymbol{\theta}_G)] := \mathbb{E}\{\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{g=1}^{G}\log[\pi_g f_g(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G)]\}$

is uniquely maximized at $\boldsymbol{\theta}_{G,0} := (\boldsymbol{k}'_{G,0}\,,\,\boldsymbol{\sigma}'_{G,0}\,,\,\boldsymbol{\pi}'_{G,0}\,,\,\boldsymbol{\mu}'_0\,)'$;

2. $\mathcal{P}_G$ *is a closed convex set;*

3. $\log L(\boldsymbol{R},\boldsymbol{F},\boldsymbol{d},\boldsymbol{\theta}_G) \xrightarrow{\mathbb{P}} \log L_0(\boldsymbol{R},\boldsymbol{F},\boldsymbol{d},\boldsymbol{\theta}_G)$ *for all* $\boldsymbol{\theta}_G \in \mathcal{P}_G$.

Proposition 2. *Under the mixture of regressions model* (8)*, the conditions of Proposition 1, the regularity Assumptions (a)–(c) of Cheng and Liu (2001), and Assumption 2, then*

$$d(\hat{\boldsymbol{\theta}}_G^{MLE}, \mathcal{P}_{G_0}) \to 0 \qquad \textit{with probability 1}$$

as $n, T \to \infty$*, where* $\mathcal{P}_{G_0} := \{\boldsymbol{\theta}_G \in \mathcal{P}_G : f(u,v;\boldsymbol{\theta}_G) = f(u,v;\boldsymbol{\theta}_{G,0})\}$*, and* $d(u_n, \mathcal{P}_G) \to 0$ *if and only if there exists a sequence* $\{v_n\}$ *of points in* $\mathcal{P}_G$ *such that* $||u_n - v_n|| \to 0$ *as* $n, T \to \infty$*, where* $||\bullet||$ *is the Euclidian distance.*

Moreover, with $G = G_0$*, the parameters converge at the following rates:*

$$\begin{aligned} \hat{\boldsymbol{\mu}}^{MLE} - \boldsymbol{\mu}_0 &= O_{\mathbb{P}}(n^{-1/2})\,, & \hat{\boldsymbol{\pi}}_G^{MLE} - \boldsymbol{\pi}_{G,0} &= O_{\mathbb{P}}((nT)^{-1/2})\,, \\ \hat{\boldsymbol{k}}_G^{MLE} - \boldsymbol{k}_{G,0} &= O_{\mathbb{P}}((nT)^{-1/2})\,, & \hat{\boldsymbol{\sigma}}_G^{MLE} - \boldsymbol{\sigma}_{G,0} &= O_{\mathbb{P}}((nT)^{-1/2})\,. \end{aligned} \tag{A.2}$$

Proof. The proof is omitted as it is a straightforward application of Theorem 1 of Cheng and Liu (2001), allowing for $G > G_0$. In the case of $G = G_0$, the proposition nests the consistency result of Theorem 1 of Feng and McCulloch (1996), which holds for parameters at the boundary of $\mathcal{P}_G$. For $G = G_0$ we also report convergence rates by standard central limit theorem arguments, acknowledging that $\boldsymbol{\mu}$ is estimated over the time-t cross-section of size n, hence the $\sqrt{n}$-convergence (the partial derivatives with respect to μ_t involve n summands). □

A.3 Convergence of the ECM algorithm

The EM algorithm is an iterative procedure that monotonically increases the observed-data log-likelihood and converges to a stationary point (Dempster et al., 1977; Wu, 1983). Whether that stationary point is the global maximum depends on initialization and the shape of the likelihood. Meng and Rubin (1993) show that the ECM algorithm shares the convergence properties of the EM algorithm.

Proposition 3 (Consistency of the ECM estimator $\hat{\boldsymbol{\theta}}_G$)**.** *Suppose the ECM algorithm*

converges to a global maximum of the observed-data log-likelihood. Then,

$$\hat{\boldsymbol{\theta}}_G \xrightarrow{\mathbb{P}} \hat{\boldsymbol{\theta}}_G^{MLE},$$

hence $\hat{\boldsymbol{\theta}}_G$ inherits the consistency properties (and asymptotic normality) of the MLE (see Proposition 2), $d(\hat{\boldsymbol{\theta}}_G, \mathcal{P}_{G_0}) \to 0$ w.p.1, as $n, T \to \infty$.

Proof. Since the ECM inherits the properties of the EM algorithm (Meng and Rubin, 1993), we sketch the proof by means of the EM. By Dempster et al. (1977), each EM iteration increases or leaves unchanged the observed-data log-likelihood, defined as

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{g=1}^{G}\log[\pi_g f_g(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G)].$$

By Wu (1983), the sequence $\{\boldsymbol{\theta}_G^{(\ell)}\}_\ell$ converges to some stationary point $\bar{\boldsymbol{\theta}}_G$. In finite mixture models, multiple local maxima can occur. If the initial guess or restarting strategy allows the EM iterates to attain the global maximum, then $\bar{\boldsymbol{\theta}}_G = \hat{\boldsymbol{\theta}}_G^{MLE}$. Thus, with high probability $\hat{\boldsymbol{\theta}}_G := \bar{\boldsymbol{\theta}}_G$. By Proposition 2, $d(\hat{\boldsymbol{\theta}}_G, \mathcal{P}_{G_0}) \to 0$ w.p.1, as $n, T \to \infty$. Therefore, for $G = G_0$, if ECM converges to the global maximum, $\hat{\boldsymbol{\theta}}_G \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_{G,0}$. In practice, in order to mitigate the risk of local maxima, we use 150 multiple random initializations and pick the iterate with the highest log-likelihood (see, e.g., Yao and Xiang, 2024). □

A.4 Estimation error in expected attention weights

In this subsection, we provide details for the robust variance decomposition in Table 4.

In Section 5 we estimate the following linear regression

$$\hat{k}_{i,t} = \boldsymbol{\beta}' \boldsymbol{X}_{i,t} + \xi_{i,t} \tag{A.3}$$

where $\xi_{i,t}$ is an error term, assumed to be i.i.d., and $\boldsymbol{X}_{i,t}$ is a vector of dummy variables capturing the socioeconomic status of the survey respondents, aggregate factors, (potentially) a constant, and fixed effects.

Let $k_{i,t} := \sum_{g=1}^{G} w_g(R_{i,t}|F_{i,t-1};\, \boldsymbol{\theta})\, k_g$ be the actual expected attention weights that we

would like to have in (A.3) in place of their estimates

$$\hat{k}_{i,t} = \sum_{g=1}^{G} w_g(R_{i,t}|F_{i,t-1};\hat{\boldsymbol{\theta}})\,\hat{k}_g\,. \tag{A.4}$$

Moreover, we assume that the estimation error arising from (A.4), $e_{i,t} := \hat{k}_{i,t} - k_{i,t}$, is (i) additive separable, and (ii) i.i.d. with finite variance. That is,

$$\hat{k}_{i,t} = k_{i,t} + e_{i,t}, \qquad e_{i,t} \sim iid(0, \sigma_e^2). \tag{A.5}$$

The estimation error can be factorized as the sum of two elements: the estimation error in component-specific attention weights, k_g, and the allocation error of observations into the G components. Namely,

$$e_{i,t} = \underbrace{\sum_{g=1}^{G} w_g(R_{i,t}|F_{i,t-1};\hat{\boldsymbol{\theta}})\big(\hat{k}_g - k_g\big)}_{\text{error in } \hat{k}_g} + \underbrace{\sum_{g=1}^{G}\Big[w_g\big(R_{i,t}|F_{i,t-1};\hat{\boldsymbol{\theta}}\big) - w_g\big(R_{i,t}|F_{i,t-1};\boldsymbol{\theta}\big)\Big]\,k_g}_{\text{allocation error}}\,. \tag{A.6}$$

In addition, we make the assumption that the estimation error is independent of regressors, $e_{i,t} \perp \boldsymbol{X}_{i,t}$. The R-squared ($R^2$) of regression (A.3) is given by $R^2 = \frac{\mathbb{V}ar(\hat{\boldsymbol{\beta}}'\boldsymbol{X}_{i,t})}{\mathbb{V}ar(\hat{k}_{i,t})}$ where $\mathbb{V}ar(\hat{k}_{i,t}) = \mathbb{V}ar(k_{i,t}) + \mathbb{V}ar(e_{i,t})$ from (A.5). Clearly, the presence of estimation error, $e_{i,t}$, pushes the R^2 of regression (A.3) towards zero, and thus the unexplained variance, $1 - R^2$, towards one therefore biasing interpretation of the unexplained variance measure. We apply Algorithm 1 to correct the effect of estimation error in the computation of the unexplained variance, $1 - R^2$.

Algorithm 1 is computationally inexpensive, as the bootstrap is already implemented for the construction of confidence intervals of $\hat{\boldsymbol{\theta}}_G$, and $\hat{\boldsymbol{\theta}}_G^{(b)}$ are outputs readily available to the researcher. Thus, the bootstrap estimates of $k_{i,t}$, $k_{i,t}^{(b)}$, provide synthetic observations for the estimated attention weights. Interestingly, the result in line 6 of Algorithm 1 delivers a consistent estimate of the observation-specific estimation error variance because of Equation (A.5), where $k_{i,t}$ is a constant and the observation-specific variance of $\hat{k}_{i,t}$, computed across the bootstrap samples, is 0, i.e. $\mathbb{V}ar_{i,t}(k_{i,t}) = 0$. Where $\mathbb{V}ar_{i,t}(\bullet)$ denotes the variance operator conditional on respondent i and time t.

Algorithm 1 Unexplained variance

Fix G, estimate $\hat{\boldsymbol{\theta}}_G$
Obtain R^2 from estimation of (A.3)
Run cluster wild bootstrap drawing from the posterior, obtain $\hat{\boldsymbol{\theta}}_G^{(b)}$ with $b = 1, ..., B$
for each tuple (i, t), with $i = 1, ..., n$, and $t = 1, ..., T$ **do**
    Compute $\bar{\hat{k}}_{i,t} = B^{-1} \sum_{b=1}^{B} \hat{k}_{i,t}^{(b)}$
    Compute $\hat{\sigma}^2_{e,\,i,t} = (B-1)^{-1} \sum_{b=1}^{B} \left(\hat{k}_{i,t}^{(b)} - \bar{\hat{k}}_{i,t} \right)^2$
end for
Compute $\hat{\sigma}_e^2 = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\sigma}^2_{e,\,i,t}$ and $\bar{\hat{k}} = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{k}_{i,t}$
Compute $\widehat{\mathbb{V}ar(\hat{k}_{i,t})} = (nT-1)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (\hat{k}_{i,t} - \bar{\hat{k}})^2$ and $\widehat{\mathbb{V}ar(k_{i,t})} = \widehat{\mathbb{V}ar(\hat{k}_{i,t})} - \hat{\sigma}_e^2$
Compute $R^2_\star = R^2 \cdot \left[\widehat{\mathbb{V}ar(\hat{k}_{i,t})} \Big/ \widehat{\mathbb{V}ar(k_{i,t})} \right]$
Return $1 - R^2_\star$

B Complementary tables and figures

Table B.1: Model Selection - Number of groups G

	BIC	BIC (k-fold)	AIC	AIC (k-fold)	ICL	ICL (k-fold)
Growth (Q0-Q3)	15	8	23	13	30	29
Growth (Q4-Q7)	8	5	8	8	28	28
Inflation (Q0-Q3)	9	9	11	9	30	26
Inflation (Q4-Q7)	6	5	6	6	28	25
Unempl. (Q0-Q3)	7	6	7	6	30	27
Unempl. (Q4-Q7)	8	6	8	7	30	27
Inflation (12m)	23	24	23	28	26	26
Inflation (36m)	28	13	28	30	30	30

Notes: Number of groups depending on various information criteria: Bayesian information criterion (BIC), Bayesian information criterion with 5-fold cross-validation (BIC k-fold), Akaike information criterion (AIC), Akaike information criterion with 5-fold cross-validation (AIC k-fold), Integrated Completed Likelihood criterion (ICL), and Integrated Completed Likelihood criterion with 5-fold cross-validation (ICL k-fold). We rely on the results from BIC-cv for our main specification.

Table B.2: Mixture parameters

	g=1	g=2	g=3	g=4	g=5	g=6	g=7	g=8	g=9	g=10	g=11	g=12	g=13	g=14	g=15	g=16	g=17	g=18
Growth (Q0-Q3)																		
π_g =	0.23	0.16	0.15	0.14	0.11	0.10	0.09	0.02										
	[0.17,0.27]	[0.16,0.16]	[0.13,0.22]	[0.08,0.15]	[0.10,0.15]	[0.06,0.13]	[0.07,0.13]	[0.01,0.03]										
k_g =	0.57	0.00	0.45	0.81	0.32	0.68	0.62	0.70										
	[0.53,0.62]	[0.00,0.00]	[0.43,0.50]	[0.77,0.89]	[0.29,0.35]	[0.63,0.75]	[0.58,0.69]	[0.27,1.26]										
	(0.56,0.57)	(0.00,0.00)	(0.44,0.46)	(0.80,0.82)	(0.31,0.33)	(0.67,0.68)	(0.62,0.63)	(0.66,0.73)										
σ_g =	0.11	0.00	0.09	0.20	0.20	0.08	0.09	0.57										
	[0.09,0.12]	[0.00,0.00]	[0.08,0.10]	[0.17,0.22]	[0.17,0.22]	[0.07,0.13]	[0.07,0.14]	[0.40,0.60]										
	(0.11,0.11)	(0.00,0.00)	(0.09,0.09)	(0.19,0.20)	(0.19,0.20)	(0.08,0.08)	(0.09,0.09)	(0.52,0.57)										
Growth (Q4-Q7)																		
π_g =	0.30	0.23	0.20	0.15	0.11													
	[0.18,0.34]	[0.23,0.23]	[0.12,0.24]	[0.11,0.27]	[0.08,0.17]													
k_g =	0.41	0.00	0.23	0.57	0.63													
	[0.33,0.45]	[0.00,0.00]	[0.19,0.26]	[0.47,0.63]	[0.55,0.72]													
	(0.40,0.42)	(0.00,0.00)	(0.22,0.24)	(0.56,0.57)	(0.62,0.65)													
σ_g =	0.14	0.00	0.22	0.15	0.39													
	[0.12,0.16]	[0.00,0.00]	[0.18,0.24]	[0.12,0.18]	[0.33,0.41]													
	(0.14,0.15)	(0.00,0.00)	(0.21,0.22)	(0.15,0.15)	(0.37,0.39)													
Inflation (Q0-Q3)																		
π_g =	0.19	0.18	0.17	0.12	0.09	0.09	0.06	0.06	0.05									
	[0.19,0.19]	[0.10,0.19]	[0.15,0.23]	[0.09,0.18]	[0.05,0.13]	[0.07,0.14]	[0.04,0.10]	[0.02,0.09]	[0.02,0.06]									
k_g =	0.00	0.87	0.48	0.70	0.63	0.58	0.72	0.79	0.29									
	[0.00,0.00]	[0.83,0.93]	[0.45,0.51]	[0.67,0.74]	[0.59,0.67]	[0.55,0.61]	[0.59,0.82]	[0.75,0.85]	[0.19,0.37]									
	(0.00,0.00)	(0.86,0.88)	(0.48,0.49)	(0.69,0.70)	(0.62,0.63)	(0.57,0.58)	(0.70,0.73)	(0.78,0.79)	(0.27,0.32)									
σ_g =	0.00	0.12	0.13	0.06	0.06	0.09	0.28	0.04	0.19									
	[0.00,0.00]	[0.11,0.15]	[0.11,0.15]	[0.05,0.09]	[0.03,0.09]	[0.05,0.12]	[0.25,0.30]	[0.02,0.08]	[0.13,0.21]									
	(0.00,0.00)	(0.12,0.12)	(0.13,0.13)	(0.06,0.07)	(0.06,0.06)	(0.09,0.09)	(0.27,0.28)	(0.04,0.05)	(0.18,0.19)									
Inflation (Q4-Q7)																		
π_g =	0.27	0.24	0.19	0.18	0.13													
	[0.27,0.27]	[0.12,0.24]	[0.12,0.21]	[0.19,0.33]	[0.09,0.17]													
k_g =	0.00	0.57	0.26	0.34	0.40													
	[0.00,0.00]	[0.52,0.65]	[0.23,0.29]	[0.31,0.39]	[0.31,0.51]													
	(0.00,0.00)	(0.55,0.57)	(0.25,0.27)	(0.33,0.35)	(0.39,0.42)													
σ_g =	0.00	0.15	0.18	0.09	0.31													
	[0.00,0.00]	[0.14,0.17]	[0.14,0.24]	[0.09,0.12]	[0.27,0.31]													
	(0.00,0.00)	(0.15,0.15)	(0.18,0.18)	(0.09,0.10)	(0.30,0.31)													
Unempl. (Q0-Q3)																		
π_g =	0.26	0.24	0.20	0.11	0.10	0.10												

Continued on next page

Table B.2: Mixture parameters

	g=1	g=2	g=3	g=4	g=5	g=6	g=7	g=8	g=9	g=10	g=11	g=12	g=13	g=14	g=15	g=16	g=17	g=18
	[0.26,0.26]	[0.14,0.23]	[0.16,0.25]	[0.11,0.21]	[0.05,0.09]	[0.10,0.17]												
k_g =	0.00	0.83	0.63	0.48	0.61	0.33												
	[0.00,0.00]	[0.80,0.92]	[0.60,0.71]	[0.46,0.54]	[0.51,0.89]	[0.30,0.38]												
	(0.00,0.00)	(0.82,0.84)	(0.62,0.64)	(0.47,0.49)	(0.58,0.65)	(0.32,0.34)												
σ_g =	0.00	0.10	0.06	0.07	0.32	0.13												
	[0.00,0.00]	[0.09,0.13]	[0.06,0.08]	[0.06,0.08]	[0.30,0.35]	[0.12,0.17]												
	(0.00,0.00)	(0.10,0.11)	(0.06,0.07)	(0.07,0.08)	(0.30,0.31)	(0.13,0.14)												
Unempl. (Q4-Q7)																		
π_g =	0.27	0.24	0.17	0.17	0.13	0.02												
	[0.20,0.30]	[0.24,0.24]	[0.14,0.23]	[0.12,0.21]	[0.09,0.17]	[0.00,0.09]												
k_g =	0.62	0.00	0.45	0.35	0.27	0.93												
	[0.56,0.68]	[0.00,0.00]	[0.41,0.49]	[0.29,0.41]	[0.25,0.31]	[0.67,1.17]												
	(0.61,0.63)	(0.00,0.00)	(0.44,0.45)	(0.34,0.37)	(0.27,0.28)	(0.88,0.95)												
σ_g =	0.20	0.00	0.10	0.32	0.12	0.20												
	[0.17,0.22]	[0.00,0.00]	[0.09,0.13]	[0.28,0.33]	[0.09,0.14]	[0.06,0.28]												
	(0.20,0.20)	(0.00,0.00)	(0.10,0.11)	(0.30,0.32)	(0.12,0.12)	(0.19,0.21)												
Inflation (12m)																		
π_g =	0.35	0.08	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	[0.35,0.35]	[0.07,0.08]	[0.05,0.06]	[0.05,0.06]	[0.05,0.06]	[0.04,0.05]	[0.04,0.05]	[0.03,0.04]	[0.03,0.03]	[0.03,0.03]	[0.03,0.03]	[0.02,0.02]	[0.03,0.03]	[0.02,0.02]	[0.02,0.02]	[0.02,0.02]	[0.01,0.02]	[0.02,0.02]
k_g =	0.00	0.99	0.55	0.65	0.84	1.37	0.45	-0.67	-0.27	0.37	1.00	0.23	0.75	0.23	-0.40	-1.26	0.31	0.82
	[-0.00,0.00]	[0.99,1.03]	[0.54,0.55]	[0.65,0.67]	[0.84,0.86]	[1.35,1.47]	[0.44,0.45]	[-0.70,-0.65]	[-0.28,-0.26]	[0.36,0.38]	[0.99,1.01]	[0.22,0.23]	[0.75,0.78]	[0.22,0.23]	[-0.41,-0.36]	[-1.35,-1.22]	[0.29,0.32]	[0.62,0.75]
	(-0.00,0.00)	(0.99,1.00)	(0.55,0.55)	(0.65,0.66)	(0.84,0.85)	(1.35,1.38)	(0.45,0.45)	(-0.68,-0.66)	(-0.27,-0.26)	(0.37,0.37)	(1.00,1.00)	(0.22,0.23)	(0.75,0.75)	(0.22,0.23)	(-0.40,-0.39)	(-1.28,-1.24)	(0.31,0.31)	(0.81,0.83)
σ_g =	0.00	3.90	1.80	1.87	2.53	9.13	1.61	2.79	1.35	1.43	1.64	1.19	2.09	1.19	1.55	4.00	1.26	10.11
	[0.00,0.00]	[3.89,4.20]	[1.76,1.89]	[1.82,1.99]	[2.43,2.70]	[8.52,9.37]	[1.56,1.68]	[2.73,3.12]	[1.39,1.55]	[1.40,1.53]	[1.56,1.75]	[1.21,1.30]	[2.07,2.36]	[1.21,1.30]	[1.43,1.80]	[3.90,4.54]	[1.26,1.39]	[9.49,10.57]
	(0.00,0.00)	(3.88,3.91)	(1.81,1.82)	(1.88,1.90)	(2.54,2.56)	(8.99,9.09)	(1.62,1.63)	(2.81,2.84)	(1.36,1.38)	(1.44,1.45)	(1.66,1.69)	(1.21,1.22)	(2.11,2.12)	(1.21,1.22)	(1.56,1.58)	(4.03,4.09)	(1.28,1.29)	(9.96,10.04)
Inflation (36m)																		
π_g =	0.32	0.09	0.08	0.08	0.07	0.07	0.07	0.06	0.06	0.05	0.03	0.02	0.02					
	[0.32,0.32]	[0.07,0.09]	[0.07,0.09]	[0.06,0.08]	[0.06,0.08]	[0.05,0.07]	[0.06,0.07]	[0.04,0.07]	[0.06,0.07]	[0.04,0.05]	[0.02,0.04]	[0.02,0.06]	[0.01,0.02]					
k_g =	0.00	1.46	0.40	0.23	0.69	-0.31	1.01	0.84	0.55	-0.85	0.47	0.49	0.13					
	[-0.00,0.00]	[1.52,1.60]	[0.38,0.40]	[0.22,0.24]	[0.68,0.70]	[-0.37,-0.31]	[1.01,1.03]	[0.84,0.87]	[0.54,0.56]	[-0.97,-0.90]	[0.36,0.54]	[0.48,0.89]	[-0.20,0.19]					
	(-0.00,0.00)	(1.47,1.50)	(0.40,0.40)	(0.23,0.23)	(0.68,0.69)	(-0.32,-0.31)	(1.01,1.02)	(0.84,0.85)	(0.55,0.55)	(-0.88,-0.86)	(0.45,0.52)	(0.49,0.50)	(0.12,0.16)					
σ_g =	0.00	6.63	1.48	1.37	1.73	1.73	1.62	2.05	1.50	3.83	18.59	3.80	11.67					
	[0.00,0.00]	[6.26,6.73]	[1.47,1.61]	[1.35,1.48]	[1.58,1.82]	[1.81,1.99]	[1.53,1.78]	[1.66,2.21]	[1.49,1.65]	[3.70,4.09]	[16.90,18.58]	[3.78,4.51]	[1.46,11.43]					
	(0.00,0.00)	(6.51,6.58)	(1.49,1.50)	(1.38,1.39)	(1.74,1.76)	(1.74,1.76)	(1.65,1.67)	(2.07,2.09)	(1.51,1.52)	(3.83,3.88)	(18.33,18.49)	(3.80,3.82)	(11.46,11.57)					

Notes: The table reports estimates for population shares, π_g, group-specific attention weights, k_g, and standard deviations, σ_g, for each dataset. Groups are ordered based on their estimated size, π_g. Confidence intervals at the 95% level are reported in brackets. These intervals account for group-allocation and parameter uncertainty. They are obtained from estimating the mixtures on 1,000 bootstrapped samples. Partial 95% confidence intervals are reported in parenthesis. These intervals account for parameter uncertainty only. They are obtained from estimating the mixtures on 1,000 bootstrapped samples imposing fixed group-membership posterior probabilities. To save on space, g is truncated to 18 (for US SCE inflation, 12m).

Online Appendix

Not for publication

Contents of the Online Appendix

C Bias from neglecting heterogeneity in attention weights

In this section, we calculate the asymptotic bias in average attention arising when the true DGP is given by the heterogeneous attention model (8), but the researcher instead estimates the homogeneous model (7). Specifically, we consider the case where the researcher employs a within (fixed-effects) estimator on (7) in order to eliminate the aggregate factor μ_t. Consider $d_{i,t}(g)$ as given in (8). Let us provide the following definitions

$$F_{t-1} := \frac{1}{n}\sum_{i=1}^{n} F_{i,t-1}, \qquad p_t(g) := \frac{1}{n}\sum_{i=1}^{n} d_{i,t}(g),$$
$$\text{and} \qquad F_{t-1}^{(g)} := \frac{\sum_{i=1}^{n} d_{i,t}(g)\, F_{i,t-1}}{\sum_{i=1}^{n} d_{i,t}(g)}.$$

The following proposition establishes that ignoring heterogeneity in (8) leads to an inconsistent estimate of the average attention weight k.

Proposition 4 (Inconsistency of the within estimator)**.** *Suppose that as $n, T \to \infty$ the following conditions hold:*

(i). $\frac{1}{T}\sum_{t=1}^{T} p_t(g)\left(\mu_t - F_{t-1}\right)\left(F_{t-1} - F_{t-1}^{(g)}\right) \xrightarrow{\mathbb{P}} B(g),$

(ii). $\frac{1}{T}\sum_{t=1}^{T} p_t(g)\,\frac{\sum_{i=1}^{n} d_{i,t}(g)\left(F_{t-1} - F_{i,t-1}\right)^2}{\sum_{j=1}^{n} d_{j,t}(g)} \xrightarrow{\mathbb{P}} W(g),$

(iii). $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(F_{t-1} - F_{i,t-1}\right)^2 \xrightarrow{\mathbb{P}} V,$

where $-\infty < B(g) < \infty$ and $0 < V, W(g) < \infty$ for all $g = 1, \ldots, G$.

Then, under these conditions, the within estimator k_{FE} of the average attention weight k—obtained from a regression that ignores heterogeneity (as in (8))—is inconsistent. In particular,

$$k_{FE} \xrightarrow{\mathbb{P}} k + \frac{\sum_{g=1}^{G}(k_g - k)\left[W(g) + B(g)\right]}{V} \quad \text{as } n, T \to \infty. \tag{C.1}$$

Therefore, the within estimator of the average attention is generally inconsistent, as it includes two non-vanishing terms that correspond to the between, denoted $B(g)$, and the within group variations, denoted $W(g)$. The between variation, $B(g)$, does not vanish if there is systematic correlation between the average news, $(\mu_t - F_{t-1})$, and the difference between aggregate and group-specific forecasts, $(F_{t-1} - F_{t-1}^{(g)})$. The within variation, $W(g)$, vanishes if respondents in the g-group have priors that match the overall aggregate prior, F_{t-1}.

Proof. The proof of Proposition 4 consists in showing that the probability limit of the homogeneous within estimator of the average attention weight k converges to k plus a non-negligible part, as in (C.1).

Let us fix h and $G < \infty$. The individual revision regression model reads

$$R_{i,t} = k\left(\mu_t - F_{i,t-1}\right) + \sum_{g=1}^{G} d_{i,t}(g)\left[\left(k_g - k\right)\left(\mu_t - F_{i,t-1}\right)\right] + \varepsilon_{i,t}, \tag{C.2}$$

where the researcher, willing to estimate k, ignores heterogeneity in k, such that $\sum_{g=1}^{G} d_{i,t}(g)$ $\Big[\Big(k_g - k\Big)\Big(\mu_t - F_{i,t-1}\Big)\Big]$ enters the error term in estimation.

Define cross-sectional averages by $R_t := \frac{1}{n}\sum_{i=1}^{n} R_{i,t}$, $F_{t-1} := \frac{1}{n}\sum_{i=1}^{n} F_{i,t-1}$, and $\varepsilon_t := \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{i,t}$. Now, let $p_t(g) := \frac{1}{n}\sum_{i=1}^{n} d_{i,t}(g)$ be the fraction of respondents at time t who belong to group g, and define the group-g average of the forecasts (at $t-1$) as $F_{t-1}^{(g)} := \frac{\sum_{i=1}^{n} d_{i,t}(g) F_{i,t-1}}{\sum_{i=1}^{n} d_{i,t}(g)}$. Then, the model for aggregated (cross-sectional average) revisions is given by

$$\begin{aligned} R_t = \; & k\Big(\mu_t - F_{t-1}\Big) \; + \; \frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G} d_{i,t}(g)\Big[\Big(k_g - k\Big)\Big(\mu_t - F_{i,t-1}\Big)\Big] \; + \; \varepsilon_t, \\ = \; & k\Big(\mu_t - F_{t-1}\Big) \; + \; \sum_{g=1}^{G} p_t(g)\Big[\Big(k_g - k\Big)\Big(\mu_t - F_{t-1}^{(g)}\Big)\Big] \; + \; \varepsilon_t, \end{aligned} \tag{C.3}$$

and the within regression model reads,

$$\begin{aligned} R_{i,t} - R_t = k\Big(F_{t-1} - F_{i,t-1}\Big) + \Big(\varepsilon_{i,t} - \; \varepsilon_t\Big) + \sum_{g=1}^{G} d_{i,t}(g)\Big[\Big(k_g - k\Big)\Big(\mu_t - F_{i,t-1}\Big)\Big] \\ - \sum_{g=1}^{G} p_t(g)\Big[\Big(k_g - k\Big)\Big(\mu_t - F_{t-1}^{(g)}\Big)\Big]. \end{aligned} \tag{C.4}$$

The within estimator of k, k_{FE}, is given by

$$k_{FE} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(R_{i,t} - R_t)(F_{t-1} - F_{i,t-1})}{\sum_{i=1}^{n}\sum_{t=1}^{T}(F_{t-1} - F_{i,t-1})^2}. \tag{C.5}$$

The numerator of (C.5) is

$$\begin{aligned} \sum_{i=1}^{n}\sum_{t=1}^{T}(R_{i,t} - R_t)(F_{t-1} - F_{i,t-1}) = \; & k\sum_{i=1}^{n}\sum_{t=1}^{T}(F_{t-1} - F_{i,t-1})^2 \\ & + \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\sum_{g=1}^{G} d_{i,t}(g)(k_g - k)(\mu_t - F_{i,t-1})\right](F_{t-1} - F_{i,t-1}) \\ & + \; O_{\mathbb{P}}(n^{-1/2}T^{-1/2}) \end{aligned} \tag{C.6}$$

because

$$\sum_{i=1}^{n}\sum_{t=1}^{T}(\varepsilon_{i,t} - \varepsilon_t)(F_{t-1} - F_{i,t-1}) = O_{\mathbb{P}}(n^{-1/2}T^{-1/2}) - 0 \tag{C.7}$$

and

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\sum_{g=1}^{G}p_t(g)(k_g-k)(\mu_t-F_{t-1}^{(g)})\right](F_{t-1}-F_{i,t-1})$$
$$=\sum_{t=1}^{T}\left[\sum_{g=1}^{G}p_t(g)(k_g-k)(\mu_t-F_{t-1}^{(g)})\right](nF_{t-1}-\sum_{i=1}^{n}F_{i,t-1})=0. \tag{C.8}$$

Furthermore, we can rewrite the second addend of (C.6) as

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\sum_{g=1}^{G}d_{i,t}(g)(k_g-k)(\mu_t-F_{i,t-1})\right](F_{t-1}-F_{i,t-1})$$
$$=\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{g=1}^{G}d_{i,t}(g)(k_g-k)(\mu_t-F_{t-1})\right\}(F_{t-1}-F_{i,t-1})$$
$$+\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{g=1}^{G}d_{i,t}(g)(k_g-k)(F_{t-1}-F_{i,t-1})\right\}(F_{t-1}-F_{i,t-1}) \tag{C.9}$$

where

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{g=1}^{G}d_{i,t}(g)(k_g-k)(\mu_t-F_{t-1})\right\}(F_{t-1}-F_{i,t-1})$$
$$=n\sum_{g=1}^{G}(k_g-k)\sum_{t=1}^{T}(\mu_t-F_{t-1})\,p_t(g)\left[F_{t-1}-F_{t-1}^{(g)}\right] \tag{C.10}$$

and

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{g=1}^{G}d_{i,t}(g)(k_g-k)(F_{t-1}-F_{i,t-1})\right\}(F_{t-1}-F_{i,t-1})$$
$$=n\sum_{g=1}^{G}(k_g-k)\sum_{t=1}^{T}p_t(g)\sum_{i=1}^{n}\frac{d_{i,t}(g)(F_{t-1}-F_{i,t-1})^2}{\left(\sum_{j=1}^{n}d_{j,t}(g)\right)}. \tag{C.11}$$

Finally, ignoring negligible terms, the within estimator is given by

$$k_{FE} = k+\frac{\frac{1}{T}\sum_{g=1}^{G}(k_g-k)\sum_{t=1}^{T}p_t(g)(\mu_t-F_{t-1})\left[F_{t-1}-F_{t-1}^{(g)}\right]}{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(F_{t-1}-F_{i,t-1})^2}$$
$$+\frac{\frac{1}{T}\sum_{g=1}^{G}(k_g-k)\sum_{t=1}^{T}p_t(g)\sum_{i=1}^{n}\frac{d_{i,t}(g)(F_{t-1}-F_{i,t-1})^2}{\left(\sum_{j=1}^{n}d_{j,t}(g)\right)}}{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(F_{t-1}-F_{i,t-1})^2} \tag{C.12}$$

where the two numerators are the between groups and the within groups variations, respectively.

Assume that as $n, T \to \infty$, for all $g = 1, ..., G$,

$$\frac{1}{T}\sum_{t=1}^{T} p_t(g)(\mu_t - F_{t-1})[F_{t-1} - F_{t-1}^{(g)}] \xrightarrow{\mathbb{P}} B(g) \tag{C.13}$$

$$\frac{1}{T}\sum_{t=1}^{T} p_t(g)\frac{\sum_{i=1}^{n} d_{i,t}(g)(F_{t-1} - F_{i,t-1})^2}{\left(\sum_{j=1}^{n} d_{j,t}(g)\right)} \xrightarrow{\mathbb{P}} W(g) \tag{C.14}$$

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(F_{t-1} - F_{i,t-1})^2 \xrightarrow{\mathbb{P}} V \tag{C.15}$$

where $-\infty < B(g) < \infty$ and $0 < V, W(g) < \infty$. Then as $n, T \to \infty$, we can show the result of Proposition 4,

$$k_{FE} \xrightarrow{\mathbb{P}} k + \frac{\sum_{g=1}^{G}(k_g - k)\left[W(g) + B(g)\right]}{V}. \tag{C.16}$$

□

D Robustness and sensitivity

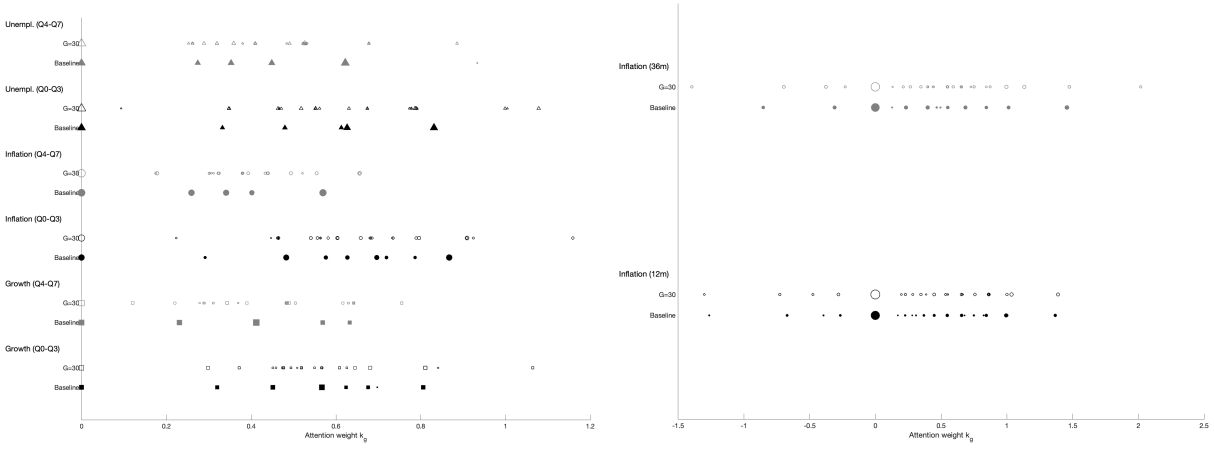
This section presents the details of the various robustness and sensitivity analyses mentioned in Section 4.

D.1 Sensitivity to the number of groups G

The number of groups selected in our benchmark estimation may be considered conservative compared to that chosen by other information criteria (see Table B.1). This subsection examines how our findings change when considering a larger number of groups.

Group-specific weights Figure D.1 presents the group estimates, comparing our baseline estimation with that for $G = 30$—the maximum number of groups considered. As shown, the primary effect of increasing G is to divide existing groups into smaller, neighboring groups. Notably, the range of attention weights remains virtually unchanged in most datasets.

Figure D.1: Group-specific weights when $G = 30$



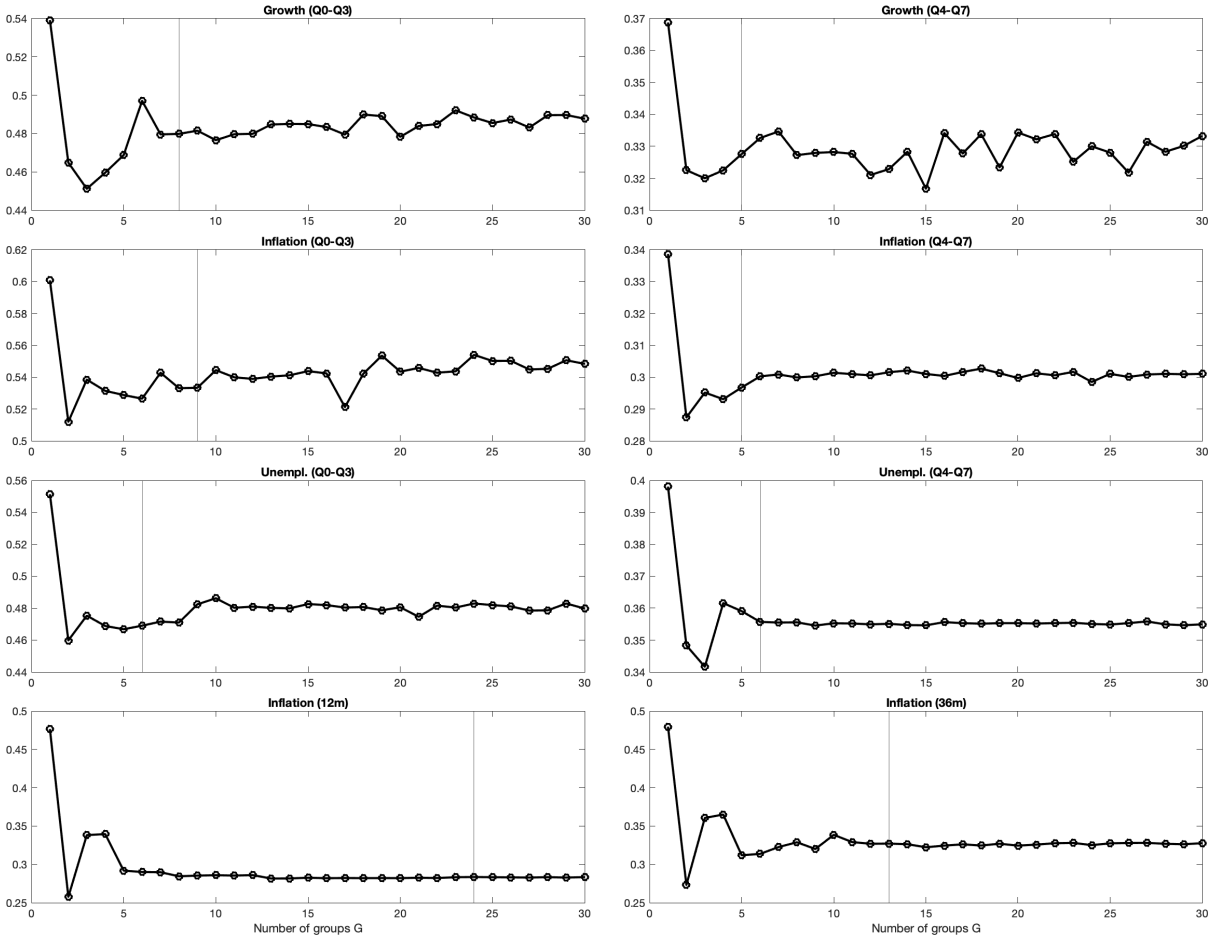
(a) ECB SPF (b) US SCE

Notes: Estimated attention weights in the ECB survey of professional forecasters (panel (a)) and US survey of consumer expectations (panel (b)). Marker sizes are proportional to population weights (π_g). Each horizontal line refers to a different dataset. Filled markers refer to the baseline selection of G, while unfilled markers corresponds to $G = 30$. Groups representing less than 1% of forecasts are not reported.

Average attention weight Figure D.2 illustrates how the average attention weights, $\boldsymbol{\pi}_G' \boldsymbol{k}_G$, evolve as the number of groups, G, increases. As stated in the main text, they rapidly converge to a value below that of the homogeneous case ($G = 1$). The vertical line marks the number of groups chosen in the baseline estimation. Since this selection

yields an average weight comparable to that obtained for much larger G, it suggests that the baseline finite mixture estimate is robust to unconstrained heterogeneity.

Figure D.2: Average attention and group selection



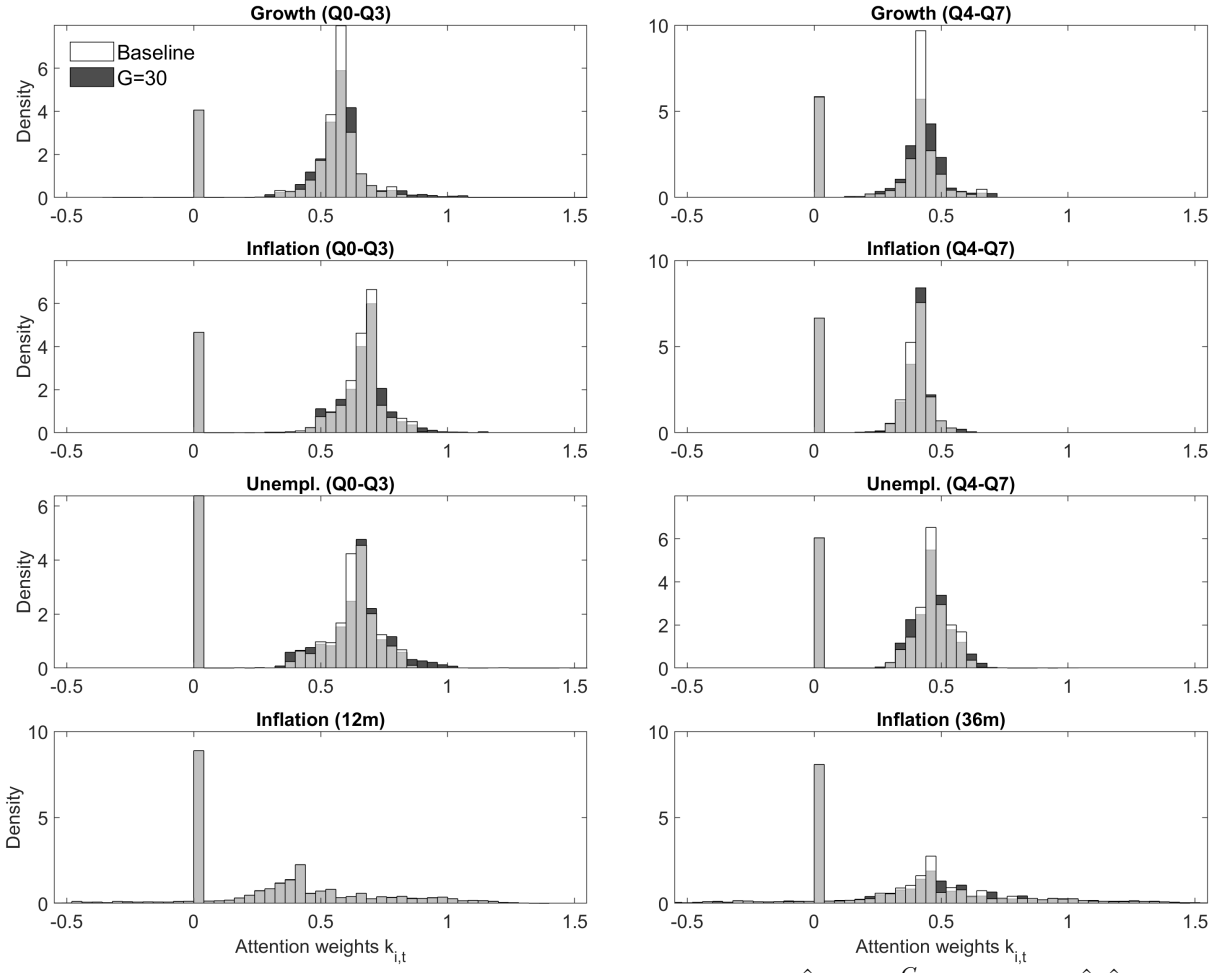
Notes: Estimated average attention weight, $\boldsymbol{\pi}_G'\boldsymbol{\theta}$, as the number of groups, G, increases. Each panel refers to the dataset mentioned in the top. The plain vertical lines corresponds to the G selected in the benchmark estimation.

Forecast-specific attention weights Figure D.3 presents the *ex post* distributions of forecast-specific attention weights, comparing our baseline estimation with that for $G = 30$. The gray areas indicate where the two distributions overlap. As shown, the distributions predicted by both specifications are similar. As expected, the main difference is that the distribution with more groups is slightly less concentrated in general.

D.2 Exogenous regressors in revisions

To limit concerns about omitted variable biases, we introduce covariates in the main regression equation. We focus on the US SCE data, for which we observe socioeconomic

Figure D.3: Distributions of forecast-specific attention weights with $G = 30$



Notes: Histograms of (estimated) forecast-specific attention weights, $\hat{k}_{i,t} = \sum_{g=1}^{G} w_g(r_{i,t}|z_{i,t};\hat{\boldsymbol{\theta}})\,\hat{k}_g$. Each panel refers to a different dataset. White histograms correspond to distributions obtained from our baseline estimation, while the black histograms correspond to distributions obtained with $G = 30$. The gray area is where the two overlap.

characteristics. The regression model is:

$$R_{i,t} \;=\; \sum_{g=1}^{G} d_{i,t}(g)\left[k_g\left(\mu_t - F_{i,t-1}\right)\right] \;+\; \boldsymbol{X}_{i,t}'\boldsymbol{\psi} \;+\; \varepsilon_{i,t}\,, \tag{D.17}$$

with $\varepsilon_{i,t}|d_{i,t}(g) = 1 \sim \mathcal{N}\left(0,\ \sigma_g^2\right)$, and where $\boldsymbol{X}_{i,t}$ is a p-dimensional vector possibly containing tenure and individual fixed effects, socioeconomic characteristics and a constant. We assume that $\boldsymbol{\psi}$ is common across all respondents and time.

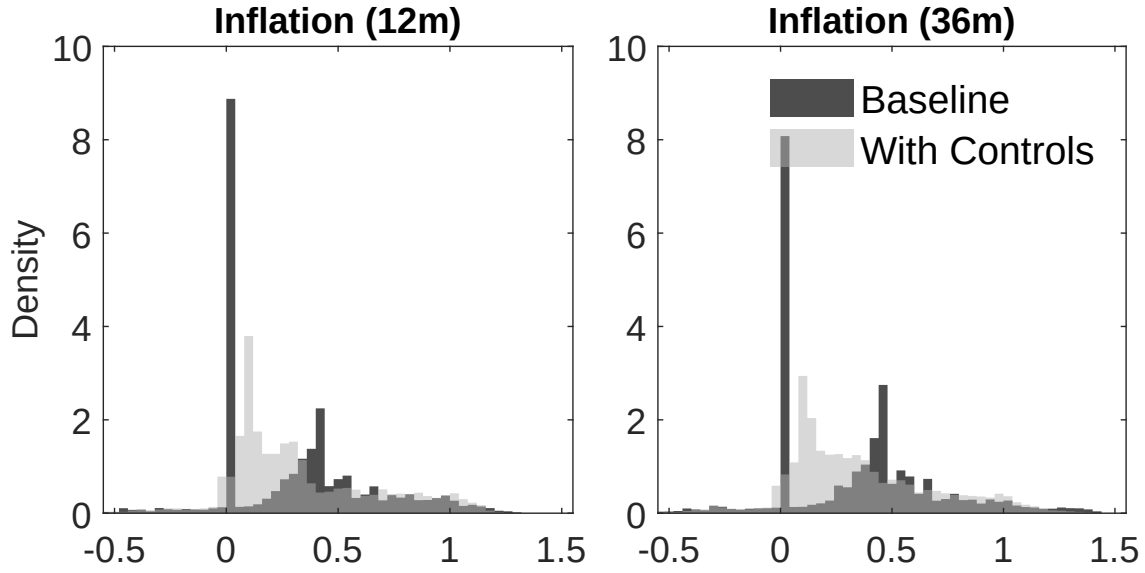
Define $x^{\perp} = x - Proj(x|\boldsymbol{X}_{i,t})$ where $Proj(\bullet|\boldsymbol{X}_{i,t})$ denotes projection onto $\boldsymbol{X}_{i,t}$. By noting that $\mathbb{E}[\mu_t X_{i,t,j}] = 0$ for all $j = 1, ..., p$, $i = 1, ..., n$ and $t = 1, ..., T$ we adapt the ECM algorithm by replacing $R_{i,t}$ and $F_{i,t-1}$ in equations (13)–(16) by $R_{i,t}^{\perp}$ and $F_{i,t-1}^{\perp}$, respectively, while leaving the rest unchanged.

The distribution from estimating (D.17) is shown in Figure D.4, alongisde the baseline

distributions. Figure D.4 highlights the similarity between the two specifications. While the patterns are qualitatively analogous, the model with exogenous regressors displays a notable quantitative difference. The primary distributional impact of estimating (D.17) instead of (8) arises around the zero counts: conditioning on socioeconomic characteristics reduces the frequency of revisions close to zero, thereby decreasing the number of observations with attention weights at 0.

We interpret Figure D.4 as evidence that the correlations between attention weights and socioeconomic characteristics documented in Section 5.2 are, if anything, lower bounds. When these factors are explicitly included as regressors in the estimation (light gray histograms), the distribution shifts away from zero, suggesting that a portion of the variation previously attributed to noise or unobserved heterogeneity is in fact systematically related to observed characteristics.

Figure D.4: Distributions of attention weights with controls



Notes: Histograms of baseline attention weights (dark gray), and attention weights estimated including controls, as in (D.17) (light gray).

D.3 Exogenous regressors in prior membership probabilities

In this section, we allow socioeconomic characteristics to explicitly influence the unconditional distribution of attention weights. Specifically, we relax the assumption of iid prior memberships π_g, and instead let them depend on p individual-specific exogenous

covariates, $\boldsymbol{X}_{i,t}$. That is, we consider the following mixture specification:

$$R_{i,t}|F_{i,t-1} \sim \sum_{g=1}^{G} \pi_g(\boldsymbol{X}_{i,t}) f_g(R_{i,t}, F_{i,t-1}; \boldsymbol{\theta}_G), \tag{D.18}$$

where the (conditional) prior probabilities $\pi_g(\boldsymbol{X}_{i,t})$ are modeled via the softmax function

$$\pi_g(\boldsymbol{X}_{i,t}) = \frac{\exp\{\boldsymbol{X}_{i,t}'\boldsymbol{\lambda}_g\}}{1 + \sum_{j=1}^{G-1} \exp\{\boldsymbol{X}_{i,t}'\boldsymbol{\lambda}_j\}}, \quad \text{for } 1 \leq g < G, \tag{D.19}$$

where $\boldsymbol{\lambda}_g \in \mathbb{R}^p$ denoting the group-specific coefficient and the Gth group serving as the reference category, therefore $\pi_G(\boldsymbol{X}_{i,t}) = (1 + \sum_{j=1}^{G-1} \exp\{\boldsymbol{X}_{i,t}'\boldsymbol{\lambda}_j\})^{-1}$. As a result, the ECM algorithm described in Section 3.3 is extended to include the maximization of the log-likelihood of the softmax probabilities with respect to $\boldsymbol{\lambda}_g$ (for all $g \leq G$) during the M-step. The likelihood function associated with (D.18) is non-convex in the parameters $\boldsymbol{\Lambda}_G$, due to the mixture structure and the softmax transformation of covariates. As is typical in finite mixture models with covariate-dependent prior probabilities, the objective involves a log-sum expression, which admits multiple local maxima and lacks global concavity (cf. McLachlan and Peel, 2000).

In the US SCE data, we have $p = 70$ covariates—representing socio-economic and geographical factors—and estimate $G = 24$ groups for one-year ahead inflation expectations (see Section 4). This adds 1,610 additional parameters, collected in the vector $\boldsymbol{\Lambda}_G := (\boldsymbol{\lambda}_1', ..., \boldsymbol{\lambda}_G')'$. Since no closed-form solution is available, the joint estimation of $\boldsymbol{\theta}_G$ and $\boldsymbol{\Lambda}_G$ is computationally demanding.

Let us collect all parameters in $\breve{\boldsymbol{\theta}}_G := (\boldsymbol{k}_G', \boldsymbol{\sigma}_G', \boldsymbol{\mu}_G', \boldsymbol{\Lambda}_G')'$. We include the M-step for $\boldsymbol{\Lambda}_G$ in the ECM algorithm; we proceed as follows: for computational efficiency, every 10 ECM cycles, we implement a stochastic gradient descent (SGD) step with $E = 200$ epochs and a learning rate of 0.1.[21] To further accelerate convergence, we impose multiple stopping criteria based on: (i) the sup-norm of the parameter updates: $||\boldsymbol{\lambda}_g^{(\ell+1)} - \boldsymbol{\lambda}_g^{(\ell)}||_\infty$; (ii) the evolution of log-likelihood (cross-entropy loss) $[\log \mathcal{L}^{(\ell+1)} - \log \mathcal{L}^{(\ell)}]/[\max(\log \mathcal{L}^{(\ell)}, \epsilon)]$

[21] Estimating $\boldsymbol{\Lambda}_G$ every 10 ECM cycles does not materially affect the final estimates, since $\breve{\boldsymbol{\theta}}_G$ converges to a local maximum. We rely on SGD because it is well suited to high-dimensional M-steps, especially when the objective function is non-convex or lacks a closed-form solution (see, e.g., Spall, 2005; Goodfellow et al., 2016; Garrigos and Gower, 2024). While the dimension of $\breve{\boldsymbol{\theta}}_G$ is conventionally large, we are not properly in the high-dimensional realm.

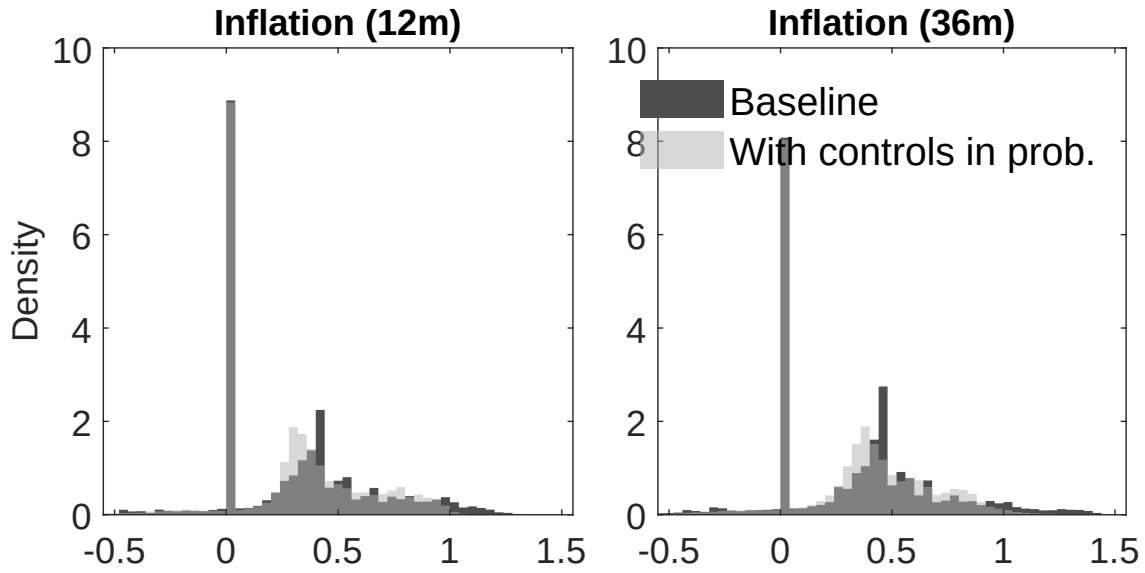
where $\epsilon > 0$ is a small scalar and $\log \mathcal{L}$ is given by:

$$\log \mathcal{L} = \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{g=1}^{G} w_g(R_{i,t}|F_{i,t-1}; \breve{\boldsymbol{\theta}}_G) \log \Big(\frac{\exp\{\boldsymbol{X}_{i,t}' \boldsymbol{\lambda}_g\}}{1 + \sum_{j=1}^{G-1} \exp\{\boldsymbol{X}_{i,t}' \boldsymbol{\lambda}_j\}} \Big), \quad \text{with } 1 \leq g < G;$$

(iii) and the sup-norm of the gradient: $||\nabla_{\boldsymbol{\Lambda}_G} \log \mathcal{L}||_\infty$. For references on the use of cross-entropy and multinomial logit in econometrics and statistical learning, see Cameron and Trivedi (2005) and Goodfellow et al. (2016), respectively. In practice, these convergence criteria often allow us to stop the SGD routine after just a quarter of the total epochs.

Figure D.5 reports the resulting distribution of attention weights when incorporating exogenous predictors in the prior group probabilities. While numerical differences arise compared to our baseline model, the overall distributions are remarkably similar. We take this as evidence that the iid assumption on π_g in Section 3 is not unduly restrictive.

Figure D.5: Distributions of attention weights with $\pi_g(\boldsymbol{X}_{i,t})$



Notes: Histograms of baseline attention weights (dark gray), and attention weights estimated with $\pi_g(\boldsymbol{X}_{i,t})$, as in (D.18) (light gray).

E Simulation study

This section presents Monte Carlo experiments designed to investigate the finite-sample properties of our estimator, $\hat{\boldsymbol{\theta}}_G$, as well as the behavior of model selection criteria. Across exercises, we simulate data drawn from the posterior distribution conditional on empirical datasets, aiming to closely replicate real-world data structures. We focus on two datasets related to inflation expectations:

1. ECB SPF: short-run inflation expectations;

2. US SCE: one-year-ahead inflation expectations.

E.1 Finite-sample properties of $\hat{\boldsymbol{\theta}}_G$.

We first examine the finite-sample performance of the estimator $\hat{\boldsymbol{\theta}}_G$. In all designs, we conduct $S = 10,000$ Monte Carlo replications. The simulations are constructed to mirror key features of the empirical datasets. For a given value of G, we use the ECM algorithm introduced in Section 3, and generate posterior predictive draws from the model using the empirical estimate $\hat{\boldsymbol{\theta}}_G$ as the data-generating parameter.

Table E.1 reports results for the first design, which mimics the ECB SPF dataset. The panel consists of $nT = 4{,}741$ observations over $T = 101$ periods, and is kept unbalanced to reflect the structure of the real data. The baseline parameter vector is set to $\boldsymbol{\theta}_{9,0} = \hat{\boldsymbol{\theta}}_9$, as estimated in Section 4. The table summarizes the finite-sample bias (Mean) and dispersion (IQrange) of the estimator across simulations.[22]

Table E.2 shows analogous results for the second design, based on the US SCE dataset, with $nT = 121{,}851$ and $T = 121$. We set the true model order to $G_0 = 24$, based on the empirical results reported in Section 4.

In both designs, we find that the mean squared error of the parameter estimates is primarily driven by sampling variance rather than bias. This indicates that the estimator is approximately unbiased, and that precision improves with larger panel sizes or more stable component structures.

[22]The interquartile range (IQrange) is defined as the difference between the 75th and 25th percentiles of the Monte Carlo estimates. For the jth parameter in $\boldsymbol{\theta}_G$, it is given by $IQrange_j = F^{-1}_{\boldsymbol{\theta}_{G,j}}(0.75) - F^{-1}_{\boldsymbol{\theta}_{G,j}}(0.25)$, where $F_{\boldsymbol{\theta}_{G,j}}(x) = \mathbb{P}_{\boldsymbol{\theta}_{G,j}}(X \leq x)$ denotes the empirical cumulative distribution function.

Table E.1: Finite sample properties of $\hat{\boldsymbol{\theta}}_9$, sample size: $nT = 4,741$, $T = 101$

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
Truth	0.6267	0.8669	0.5761	0.6963	0.4828	0.0000	0.7193	0.7868	0.2916
Mean	0.6328	0.8836	0.5802	0.7058	0.4849	0.0000	0.6794	0.7945	0.2728
IQrange	0.0174	0.0256	0.0171	0.0200	0.0153	0.0000	0.0564	0.0260	0.0573
	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9
Truth	0.0575	0.1211	0.0885	0.0629	0.1299	0.0000	0.2829	0.0382	0.1894
Mean	0.0595	0.1188	0.0869	0.0663	0.1241	0.0000	0.2696	0.0454	0.1677
IQrange	0.0089	0.0107	0.0143	0.0091	0.0126	0.0000	0.0248	0.0109	0.0339
	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9
Truth	0.0928	0.1835	0.0851	0.1190	0.1666	0.1862	0.0641	0.0559	0.0466
Mean	0.0973	0.1558	0.0978	0.1263	0.1740	0.1862	0.0770	0.0443	0.0412
IQrange	0.0152	0.0257	0.0173	0.0193	0.0192	0.0000	0.0198	0.0157	0.0120

Table E.2: Finite sample properties of $\hat{\boldsymbol{\theta}}_{24}$, sample size: $nT = 121,851$, $T = 121$

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
Truth	1.00	0.17	0.31	-0.67	0.37	0.55	-0.40	1.37	0.75	0.84
Mean	1.00	0.17	0.32	-0.69	0.37	0.55	-0.41	1.42	0.75	0.85
IQrange	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.05	0.02	0.01
	k_{11}	k_{12}	k_{13}	k_{14}	k_{15}	k_{16}	k_{17}	k_{18}	k_{19}	k_{20}
Truth	0.10	0.45	0.00	-0.12	-2.28	0.23	0.65	0.68	0.99	0.82
Mean	0.10	0.45	-0.00	-0.12	-2.29	0.22	0.66	0.74	1.01	0.65
IQrange	0.01	0.01	0.00	0.01	0.09	0.01	0.01	0.09	0.02	0.04
	k_{21}	k_{22}	k_{23}	k_{24}						
Truth	0.23	0.28	-1.26	-0.27						
Mean	0.23	0.29	-1.28	-0.27						
IQrange	0.01	0.01	0.04	0.01						
	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9	σ_{10}
Truth	1.64	1.27	1.26	2.79	1.43	1.80	1.55	9.13	2.09	2.53
Mean	1.73	1.17	1.22	2.52	1.41	1.80	1.45	8.66	2.14	2.52
IQrange	0.13	0.13	0.10	0.14	0.09	0.09	0.12	0.43	0.18	0.12
	σ_{11}	σ_{12}	σ_{13}	σ_{14}	σ_{15}	σ_{16}	σ_{17}	σ_{18}	σ_{19}	σ_{20}
Truth	1.38	1.61	0.00	1.32	5.44	1.19	1.87	20.68	3.90	10.11
Mean	1.35	1.57	0.00	1.16	5.00	1.13	1.92	20.09	3.88	10.22
IQrange	0.16	0.09	0.00	0.16	0.45	0.07	0.10	1.12	0.15	0.73
	σ_{21}	σ_{22}	σ_{23}	σ_{24}						
Truth	1.19	1.18	4.00	1.35						
Mean	1.13	1.08	3.80	1.26						
IQrange	0.06	0.11	0.23	0.07						
	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
Truth	0.030	0.014	0.017	0.034	0.032	0.052	0.019	0.048	0.024	0.049
Mean	0.029	0.012	0.015	0.032	0.030	0.056	0.018	0.044	0.027	0.055
IQrange	0.003	0.002	0.002	0.002	0.003	0.004	0.002	0.005	0.002	0.004
	π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	π_{16}	π_{17}	π_{18}	π_{19}	π_{20}
Truth	0.009	0.041	0.350	0.008	0.008	0.019	0.051	0.013	0.075	0.017
Mean	0.009	0.041	0.350	0.007	0.006	0.016	0.056	0.014	0.078	0.024
IQrange	0.002	0.003	0.000	0.001	0.001	0.001	0.004	0.002	0.005	0.006
	π_{21}	π_{22}	π_{23}	π_{24}						
Truth	0.026	0.014	0.019	0.032						
Mean	0.021	0.013	0.018	0.028						
IQrange	0.002	0.001	0.001	0.002						

E.2 Model Selection

We now evaluate the performance of 5-fold cross-validated (CV) model selection criteria. Due to computational demands, we restrict the Monte Carlo replications to $S = 100$.

We use as true model order G_0 the number of components selected in the empirical exercise (Section 4): $G_0 = 9$ for the ECB SPF-like dataset, and $G_0 = 24$ for the US SCE-like dataset.

In Figures E.6 (ECB SPF) we report the selection frequencies across a range of $G \in \{2, \ldots, 16\}$. Instead, in figure E.7 (US SCE), we report the selection frequencies across the range of $G \in \{16, \ldots, 30\}$. We consider three commonly used criteria: the 5-fold CV Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Integrated Completed Likelihood (ICL). As noted in Section 4, BIC is used in the empirical application.

Figures E.6 and E.7 show that, with the exception of ICL in Figure E.6, all criteria are able to approximate the true degree of heterogeneity in the simulated data. BIC and AIC lead to a relatively accurate selection of the number of components.

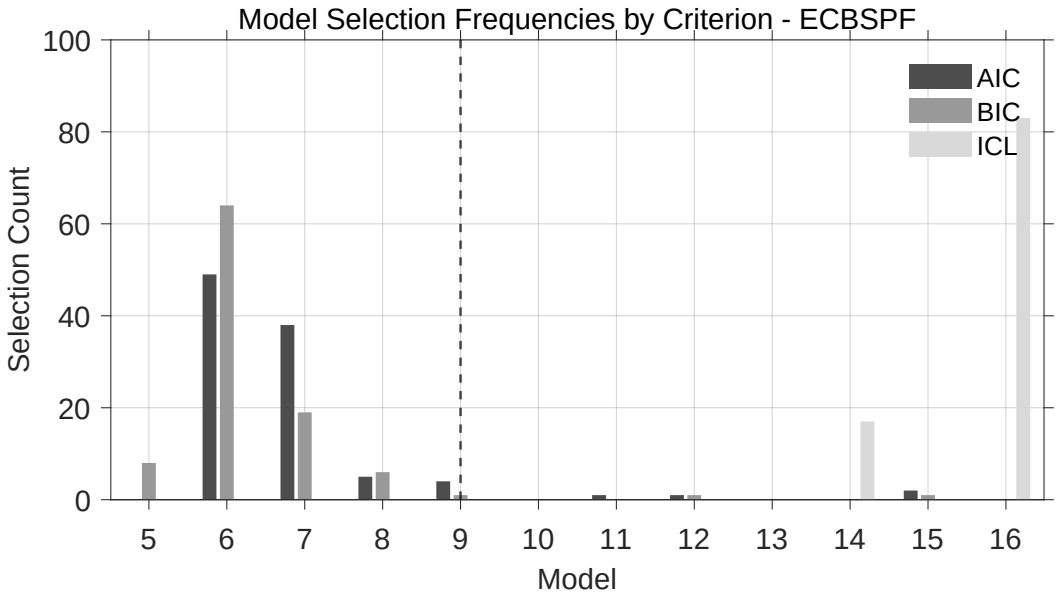


Figure E.6: 5-fold CV Model Selection frequencies, $G_0 = 9$, 100 Monte Carlo simulations.

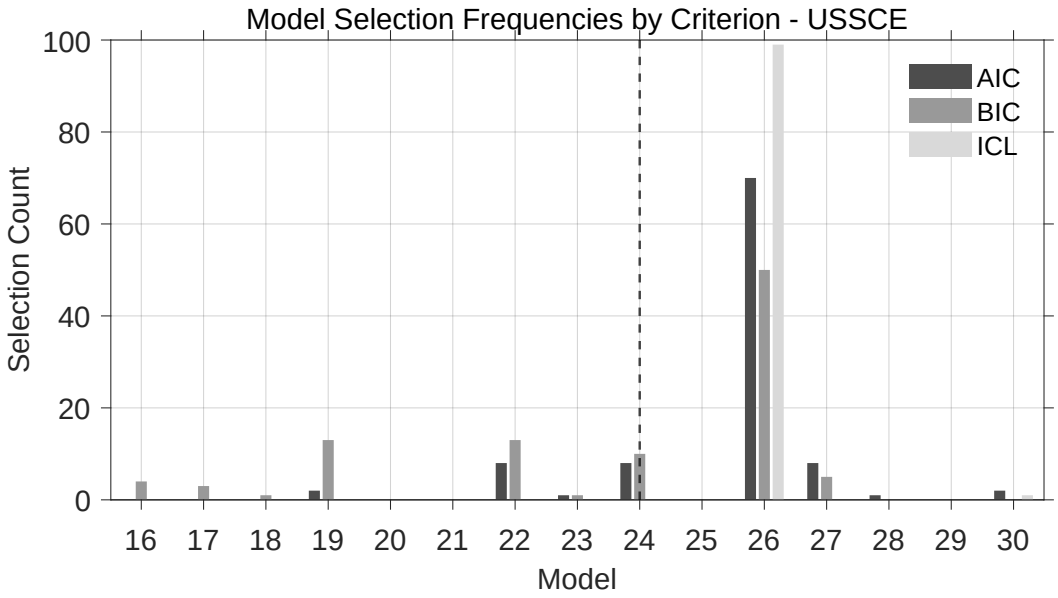


Figure E.7: 5-fold CV Model Selection frequencies, $G_0 = 24$, 100 Monte Carlo simulations.

E.3 Spurious heterogeneity

Finally, we assess the possibility that the estimation procedure might spuriously detect heterogeneity when the true data-generating process is homogeneous. To this end, we simulate data under the null of homogeneity, i.e., $k_{i,t,0} = k_0$ for all i and t, and apply the same model selection procedures described above.

Figures E.8 and E.9 report the frequency of selected models for $G \in \{1, \ldots, 15\}$ under $G_0 = 1$. The results indicate that both BIC and AIC show strong performance in avoiding spurious detection of heterogeneity. This suggests that the heterogeneity identified in the empirical datasets is unlikely to be an artifact of overfitting or noise.[23]

[23] As in Section 4, we use BIC to select G in the empirical analysis.

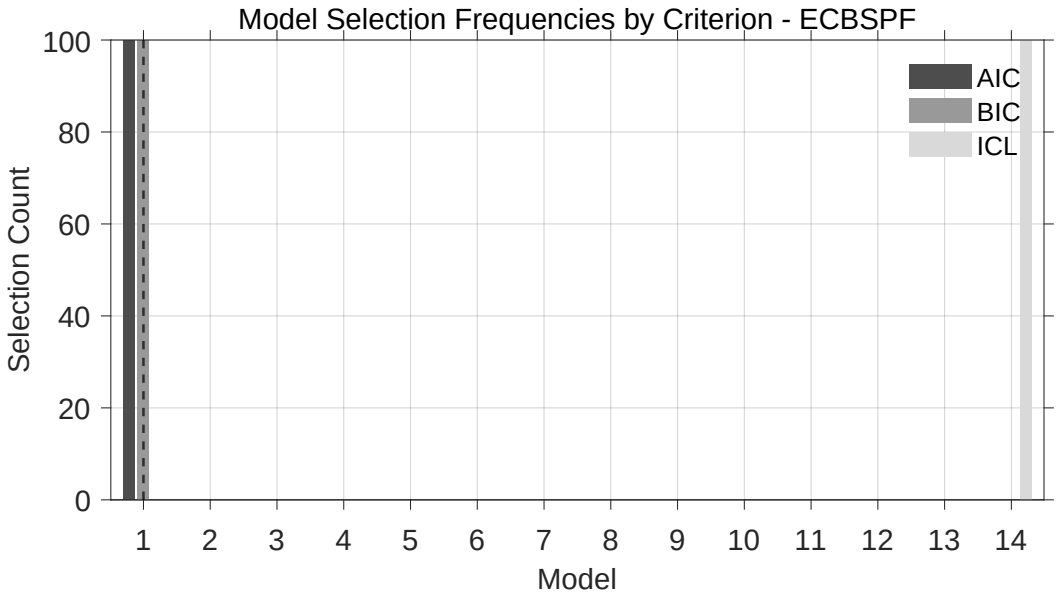


Figure E.8: 5-fold CV Model Selection frequencies, $G_0 = 1$, 100 Monte Carlo simulations.

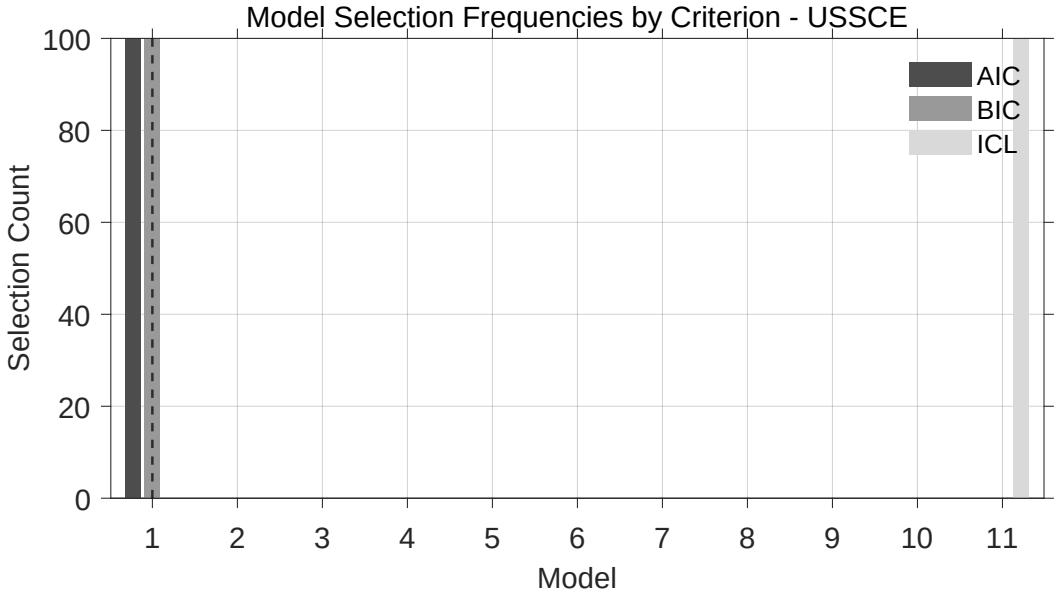
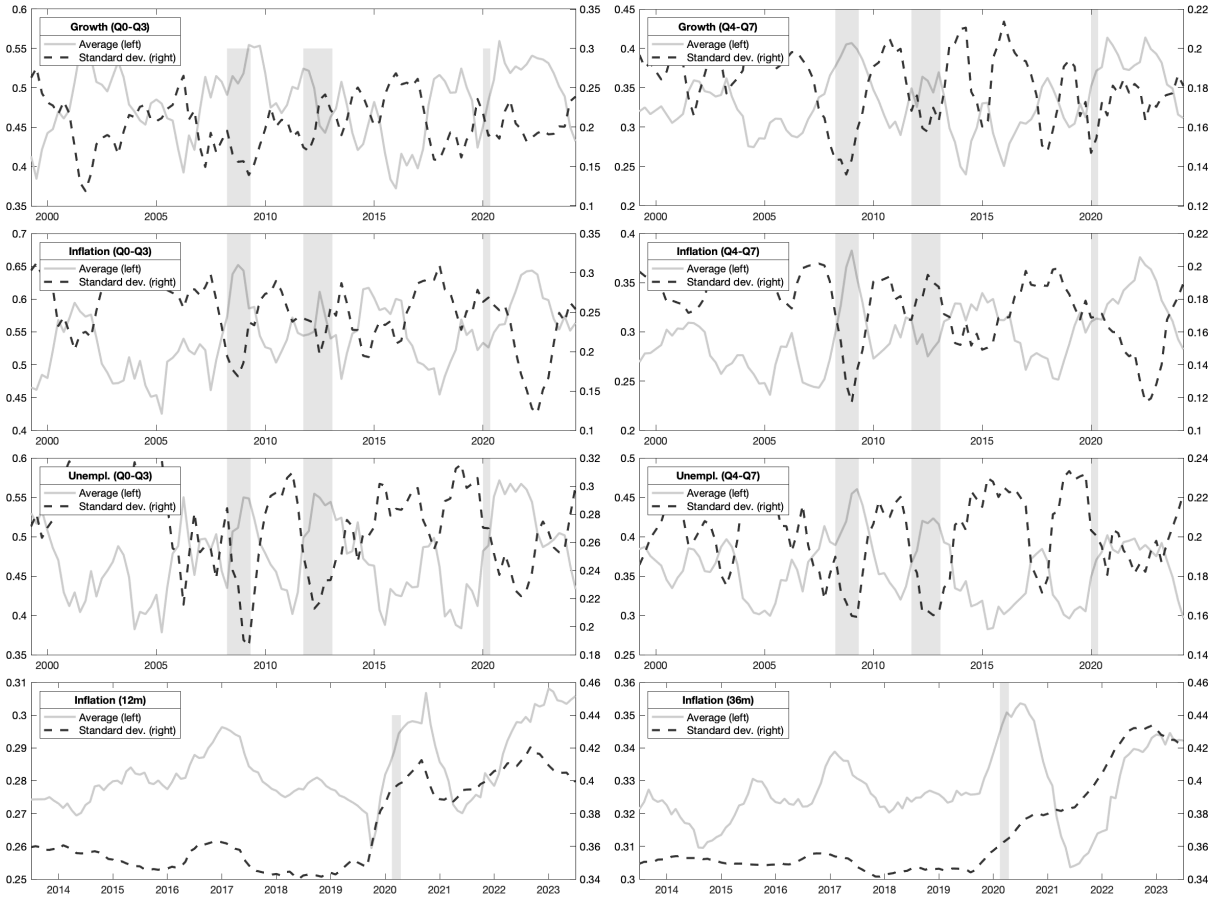


Figure E.9: 5-fold CV Model Selection frequencies, $G_0 = 1$, 100 Monte Carlo simulations.

F Aggregate and idiosyncratic factors: complements

Figure F.10: Attention over time: all datasets



Notes: Average expected gain, $k_{i,t}$, over time (solid line) and standard deviation (dashed line). One year and 6 months moving averages for the ECB SPF and US SCE datasets respectively. Each panel refers to the dataset mentioned in the top left corner. Grey areas stands for recessionary periods (following the NBER and CEPR committees).

Table F.3: Attention and individual characteristics

	Inflation (12m)		Inflation (36m)	
	(1)	(2)	(3)	(4)
Female	0.035***	0.034***	0.038***	0.037***
	(0.003)	(0.003)	(0.002)	(0.002)
Not white	0.019***	0.019***	0.022***	0.023***
	(0.004)	(0.004)	(0.003)	(0.003)
Low num.	0.036***	0.035***	0.040***	0.039***

Continued on next page

Table F.3: Attention and individual characteristics

	Inflation (12m)		Inflation (36m)	
	(0.003)	(0.003)	(0.003)	(0.003)
Employment				
Not working	-0.003	-0.003	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Age				
Over 60	0.004	0.007**	0.015***	0.018***
	(0.003)	(0.003)	(0.003)	(0.003)
Under 40	0.005*	0.003	0.002	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
Education				
High School	0.023***	0.023***	0.021***	0.021***
	(0.005)	(0.005)	(0.005)	(0.005)
Some College	0.021***	0.020***	0.027***	0.025***
	(0.003)	(0.003)	(0.002)	(0.002)
$19,999 (-)	0.001	0.001	0.008	0.010
	(0.007)	(0.007)	(0.006)	(0.006)
$20,000 - $29,999	0.015***	0.018***	0.026***	0.029***
	(0.005)	(0.005)	(0.005)	(0.005)
$30,000 - $39,999	0.011*	0.012**	0.015***	0.017***
	(0.006)	(0.006)	(0.005)	(0.005)
$40,000 - $49,999	0.013***	0.014***	0.012**	0.013***
	(0.005)	(0.005)	(0.005)	(0.005)
$50,000 - $59,999	0.013***	0.014***	0.017***	0.019***
	(0.005)	(0.004)	(0.006)	(0.006)
$60,000 - $74,999	0.009**	0.009**	0.015***	0.016***
	(0.004)	(0.004)	(0.004)	(0.004)
$75,000 - $99,999	0.007*	0.008**	0.008**	0.009***

Continued on next page

Table F.3: Attention and individual characteristics

	Inflation (12m)		Inflation (36m)	
	(0.004)	(0.004)	(0.004)	(0.004)
\$150,000 - \$199,999	-0.011**	-0.012***	-0.007	-0.008*
	(0.004)	(0.004)	(0.005)	(0.005)
\$200,000 (+)	-0.021***	-0.021***	-0.020***	-0.021***
	(0.005)	(0.005)	(0.005)	(0.005)
Constant	0.240***	0.245***	0.273***	0.291***
	(0.005)	(0.005)	(0.005)	(0.005)
R^2	0.009	0.028	0.012	0.031
Tenure FE	No	Yes	No	Yes
Time FE	No	Yes	No	Yes
State FE	No	Yes	No	Yes
Observations	120338	120033	120495	120183

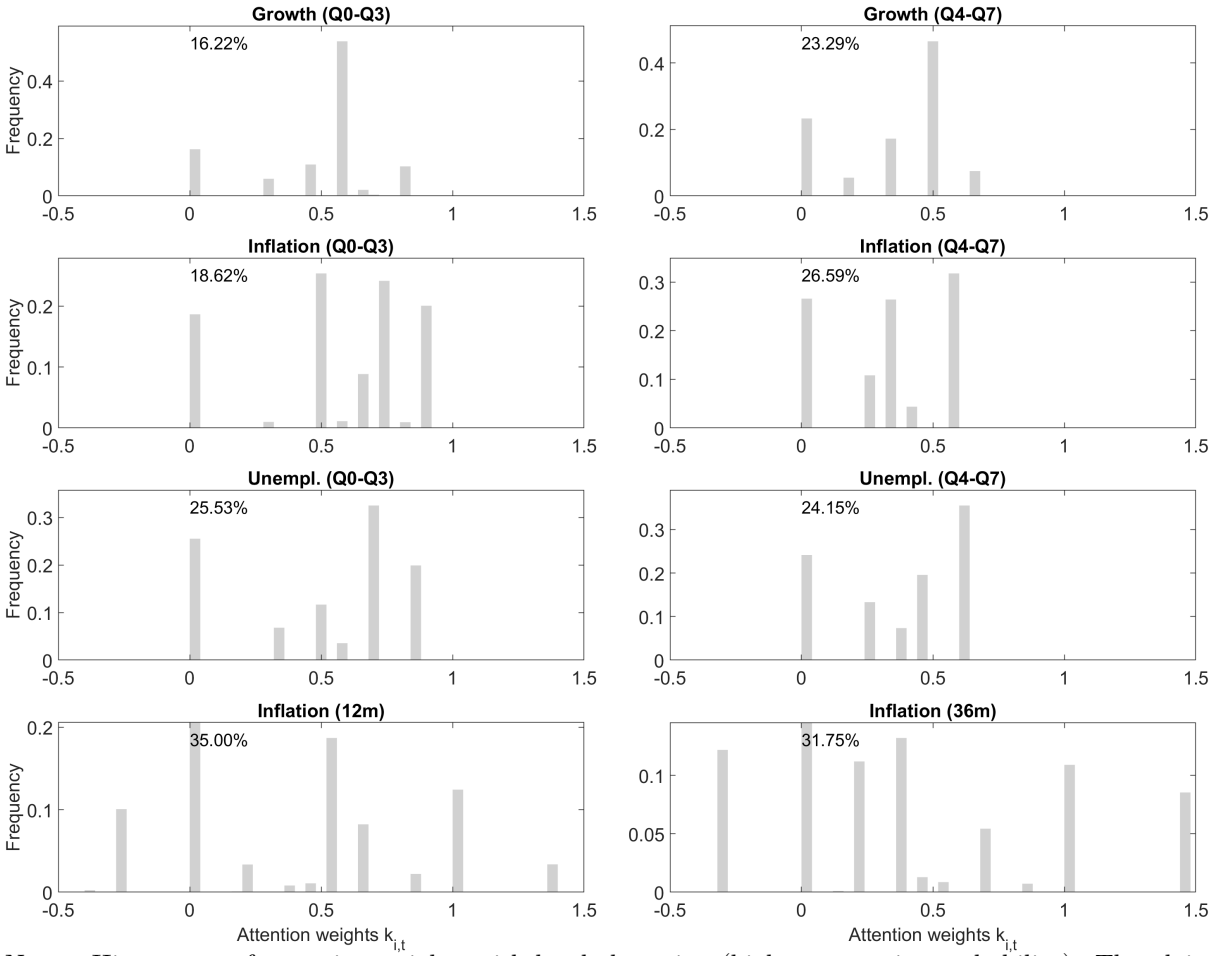
Notes: Regressions of expected attention weights, $k_{i,t}$, on consumers' characteristics from the US SCE (core). Driscoll and Kraay standard errors with two lags in parenthesis. Low num. indicates a low numeracy score. The baseline level refers to a white male, with a high numeracy score, working full time, with a college degree, whose household earnings are between \$100,000 and \$149,999, responding for the first time to the survey in 2013m7, and living in California. Tenure fixed effects for the number of months in the survey (the last category corresponds to ≥ 11 months), the month of the survey wave, and the state of the principal residence. ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

G Hard clustering: highest posterior probability

This online appendix reports the estimated distribution of attention weights based on hard clustering (maximum probability).

$$\hat{k}_{i,t} = k_{g^{\star}_{i,t}} \quad \text{s.t.} \quad g^{\star}_{i,t} = \underset{g=\{1,\dots,G\}}{\arg\sup} \; \mathbb{E}_{i,t}\left[w_g(r_{i,t}|z_{i,t};\boldsymbol{\theta})\right] \tag{G.20}$$

Figure G.11: Distributions of attention weights – highest posterior probability



Notes: Histograms of attention weights with hard clustering (highest posterior probability). The plain vertical line reports the average gain across forecasts (i, t).